FUJITSU Software PRIMECLUSTER

Installation and Administration Guide 4.3

Oracle Solaris

J2S2-1586-05ENZ0(04)
July 2018

Preface

This manual serves as your starting point for using PRIMECLUSTER. It explains the workflow of the series of operations from installation to operation management of the PRIMECLUSTER system. Since the PRIMECLUSTER system comprises multiple features, there are several other manuals besides this one for each feature. However, by reading this manual first, you will be able to perform the series of operations because this manual refers readers to other manuals that contain feature-specific information that is necessary for the operations.

This manual also provides a functional overview of products that are supported by the PRIMECLUSTER system and describes operation procedures.

This manual only covers the basic operation of PRIMECLUSTER. For operations using different hardware and software configurations, see "Related Documentation."

The table below shows the operation flow from PRIMECLUSTER installation to the start of operation management and indicates the reference location in this manual for each operation.

Flow from PRIMECLUSTER system installation to operation management

PRIMECLUSTER system operation flow		Reference location in this manual
1	Understanding the flow of PRIMECLUSTER system building and designing the PRIMECLUSTER system	Part 1 Planning
2	Installing the PRIMECLUSTER system	Part 2 Installation
3	Monitoring the operation status of the PRIMECLUSTER system	Part 3 Operations
4	Changing the PRIMECLUSTER system configuration after system operation has been started	Part 4 System Configuration Modifications
5	Backing up and restoring the operation environment of the PRIMECLUSTER system	Part 5 Maintenance
6	Building PRIMECLUSTER systems in virtualized environments	Part 6 Virtualized Environments
7	Installing products for the PRIMECLUSTER system	Part 7 PRIMECLUSTER Products

For detailed procedural explanations, refer to the reference manuals that are indicated in the target location of each part.

Target Readers

This manual is intended for all users who use PRIMECLUSTER 4.3 and perform cluster system installation and operation management. It is also intended for programmers who develop applications that operate on PRIMECLUSTER.

Configuration of This Documentation

This manual consists of seven parts, appendices, and a glossary. The contents of each part are described below.

Part 1 Planning

Audience: System administrators who build PRIMECLUSTER systems
Contents: This part describes the overall workflow from installation to operation of the PRIMECLUSTER system.

Part 2 Installation

Audience: System administrators who build PRIMECLUSTER systems
Contents: This part describes operations for software installation, cluster configuration, and application building.

Part 3 Operations

Audience: System administrators who manage system operations
Contents: This part describes operation methods for operations such as monitoring the PRIMECLUSTER system and investigating failure causes.

Part 4 System Configuration Modification

Audience: System administrators who build PRIMECLUSTER systems
Contents: This part describes necessary work items for additions, modifications, and deletions to the PRIMECLUSTER system configuration.

Part 5 Maintenance

Audience: System administrators who build PRIMECLUSTER systems
Contents: This part describes the necessary work items for maintaining the operation environment of the PRIMECLUSTER system.

Part 6 Virtualized Environments

Audience: All users who use PRIMECLUSTER systems in virtualized environments
Contents: This part describes configurations and building procedures when using PRIMECLUSTER systems in virtualized environments

Part 7 PRIMECLUSTER Products

Audience: Users who operate PRIMECLUSTER products on PRIMECLUSTER systems
Contents: This part describes the versions and levels of products that are supported by the PRIMECLUSTER system and provides a functional overview of those products.

Appendix A PRIMECLUSTER System Design Worksheets

Audience: Users who design PRIMECLUSTER systems
Contents: This appendix contains the PRIMECLUSTER System Design Worksheet.

Appendix B Manual Pages

Audience: All users who use PRIMECLUSTER systems
Contents: This appendix describes the online manual pages that are used by the individual features of the PRIMECLUSTER system.

Appendix C Troubleshooting

Audience: All users who use PRIMECLUSTER systems
Contents: This appendix describes corrective actions for problems that may occur in the PRIMECLUSTER system. It also explains how to collect data when requesting a problem investigation.

Appendix D Using SynfinityCluster Products in PRIMECLUSTER

Audience: All users who used the SynfinityCluster system
Contents: This appendix describes feature differences and terminology differences between SynfinityCluster and PRIMECLUSTER.

Appendix E Registering, Changing, and Deleting State Transition Procedure Resources for SynfinityCluster Compatibility

Audience: All users who use SynfinityCluster-compatible resources
Contents: This appendix describes procedures for registering, changing, and deleting procedure resources when the cluster applications use procedure resources.

Appendix F SafeCLUSTER Compatibility Function Settings

Audience: All users who are moving from SafeCLUSTER to PRIMECLUSTER and using the clgettrigger command with a state transition procedure executed from the procedure resources
Contents: This appendix describes the settings for when acquiring the cause of state transition to the cluster service for which the clgettrigger command was used.

Appendix G Startup Scripts and Startup Daemons in PRIMECLUSTER

Audience: System administrators who build PRIMECLUSTER systems
Contents: This appendix describes scripts and daemons that are started by PRIMECLUSTER.

Appendix H SMF Services and Port Numbers in PRIMECLUSTER

Audience: System administrators who build PRIMECLUSTER systems
Contents: This appendix describes SMF services and startup daemons for PRIMECLUSTER, and port numbers being used.

Appendix I Cloning the Cluster System Environment

Audience: System administrators who clone PRIMECLUSTER systems
Contents: This appendix describes the procedures for cloning the PRIMECLUSTER system.

Appendix J Using Remote Power Distribution Unit

Audience: All users who use the Remote Power Distribution Unit in the PRIMECLUSTER system environment
Contents: This appendix describes the procedures when using the Remote Power Distribution Unit in the PRIMECLUSTER system environment.

Appendix K Changes in Each Version

Audience: All users who have used PRIMECLUSTER 4.2A00, 4.3A10, or 4.3A20
Contents: This appendix describes the changes made to the specifications of PRIMECLUSTER 4.3A40.

Appendix L Release Information

Audience: All users who use PRIMECLUSTER systems
Contents: This appendix describes the main changes of this manual.

Glossary

Audience: All users who use PRIMECLUSTER systems
Contents: This section explains terms used to describe the PRIMECLUSTER system.

Related Documentation

Refer to the following manuals as necessary when setting up the cluster:

- *PRIMECLUSTER Concepts Guide*
- *PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*
- *PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*
- *PRIMECLUSTER Global Disk Services Configuration and Administration Guide*
- *PRIMECLUSTER Global File Services Configuration and Administration Guide*
- *PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*
- *PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function for Virtual NIC Mode*
- *PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function*
- *PRIMECLUSTER Web-Based Admin View Operation Guide*
- *PRIMECLUSTER Scalable Internet Services (SIS) Configuration and Administration Guide*
- *PRIMECLUSTER DR/PCI Hot Plug User's Guide*
- *PRIMECLUSTER Messages*

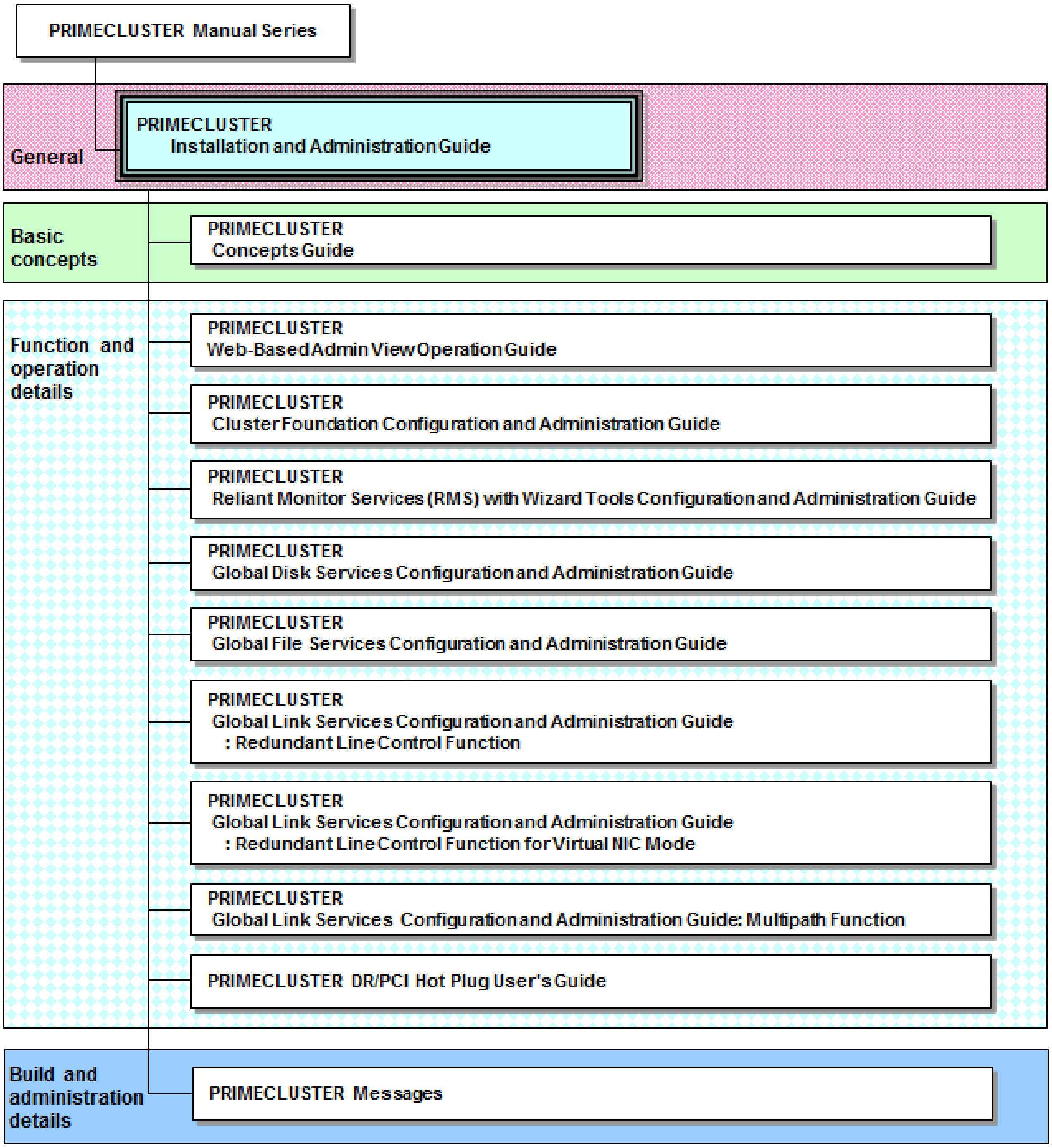
Note

The PRIMECLUSTER documentation includes the following documentation in addition to those listed above:

- PRIMECLUSTER Software Release Guide and Installation Guide

  The data is stored on "DVD" of each package. For details on the file names, see the documentation.

Manual Series



Manual Printing

If you want to print a manual, use the PDF file found on the DVD for the PRIMECLUSTER product. The correspondences between the PDF file names and manuals are described in the PRIMECLUSTER Installation Guide that comes with the product.

You will need Adobe Reader to read and print the PDF file. To get Adobe Reader, see Adobe Systems Incorporated's website.

Online Manuals

To allow users to view the online manuals, use the Cluster management server to register each user name to one of the user groups (wvroot, clroot, cladmin, or clmon).

For information on user group registration procedures and user group definitions, see "4.2.1 Assigning Users to Manage the Cluster"

Conventions

Notation

Prompts

Command line examples that require system administrator (or root) rights to execute are preceded by the system administrator prompt, the hash sign (#). Entries that do not require system administrator rights are preceded by a dollar sign ($).

Manual page section numbers

References to the UNIX operating system commands are followed by their manual page section numbers in parentheses - for example, cp(1)

The keyboard

Keystrokes that represent nonprintable characters are displayed as key icons such as [Enter] or [F1]. For example, [Enter] means press the key labeled Enter; [Ctrl-b] means hold down the key labeled Ctrl or Control and then press the [B] key.

Typefaces

The following typefaces highlight specific elements in this manual.

Typeface	Usage
`Constant Width`	Computer output and program listings; commands, file names, manual page names and other literal programming elements in the main body of text.
Italic	Variables that you must replace with an actual value.
Bold	Items in a command line that you must type exactly as shown.

Example 1

Several entries from an /etc/passwd file are shown below:

```
root:x:0:1:0000-Admin(0000):/:
sysadm:x:0:0:System Admin.:/usr/admin:/usr/sbin/sysadm
setup:x:0:0:System Setup:/usr/admin:/usr/sbin/setup
daemon:x:1:1:0000-Admin(0000):/:
```

Example 2

To use the cat(1) command to display the contents of a file, enter the following command line:

```
$ cat file
```

Notation symbols

Material of particular interest is preceded by the following symbols in this manual:

Point

Contains important information about the subject at hand.

Note

Describes an item to be noted.

Example

Describes operation using an example.

Information

Describes reference information.

See

Provides the names of manuals to be referenced.

Abbreviations

- Oracle Solaris might be described as Solaris, Solaris Operating System, or Solaris OS.
- If "Solaris X" is indicated in the reference manual name of the Oracle Solaris manual, replace "Solaris X" with "Oracle Solaris 10 (Solaris 10)" or the "Oracle Solaris 11 (Solaris 11)."

Export Controls

Exportation/release of this document may require necessary procedures in accordance with the regulations of your resident country and/or US export control laws.

Trademarks

UNIX is a registered trademark of The Open Group in the United States and other countries.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Microsoft, Windows, and Internet Explorer are registered trademarks of Microsoft Corporation in the United States and other countries.

NetWorker is a registered trademark or trademark of EMC Corporation in the United States and other countries.

lsof is used for this software.

Other product names are product names, trademarks, or registered trademarks of these companies.

Fujitsu M10 is sold as SPARC M10 by Fujitsu in Japan.

Fujitsu M10 and SPARC M10 are identical products.

Requests

- No part of this documentation may be reproduced or copied without permission of FUJITSU LIMITED.
- The contents of this documentation may be revised without prior notice.

Date of publication and edition

June 2015, Fifth edition
November 2015, 5.1 edition
November 2016, 5.2 edition
September 2017, 5.3 edition
July 2018, 5.4 edition

Copyright notice

Revision History

Revision	Location	Edition
Added the notes for the following environments: - Cluster system between guest domains among the different physical partitions (only between guest domains) - Cluster system between Kernel Zones among the different physical partitions (guest domain)	2.2.1.1.1 Cluster System Configuration in Oracle VM Server for SPARC Environment 2.2.1.2 Cluster System Operating in Oracle Solaris Kernel Zones Environment	5.1
Added the description for among the different physical partitions (only between guest domains).	2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	
Changed the descriptions of the number of the cluster applications.	2.3 Determining the Cluster System Operation Mode	
Added the description to Point in "Installation and Setup of Related Software."	3.2.2 Setting Up Disk Unit	
Changed the descriptions of Live Migration.	5.1.1 Setting Up CF and CIP 14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions 17.1 Design 17.2.1 Performing Live Migration of the Cluster on the Control Domains 17.4 Using with ServerView Resource Orchestrator Cloud Edition	
Changed the description, procedure, and note for cloning a cluster system.	5.1.1 Setting Up CF and CIP Appendix I Cloning the Cluster System Environment	
Added the check items for using the host name to set up the shutdown facility when using the SSH connection.	5.1.2.1.1 Checking XSCF Information 5.1.2.2.1 Checking Console Configuration 5.1.2.3.1 Checking Console Configuration 5.1.2.5.1 Checking XSCF Information 5.1.2.5.2 Checking ILOM Information	

Revision	Location	Edition
	5.1.2.5.3 Logging in to Global Zone Host 9.2.1 Changing an IP Address on the Public LAN	
Added the descriptions of SPARC M10-4S environment.	5.1.2.1.1 Checking XSCF Information 5.1.2.1.3 Using the Shutdown Configuration Wizard 5.1.2.5.1 Checking XSCF Information 5.1.2.5.4 Using the Shutdown Configuration Wizard 17.2.3 Performing Live Migration of the Cluster on the Guest Domains	
Changed the descriptions of the procedure to set the SNMP agent in XSCF.	5.1.2.1.2 Setting SNMP	
Added the notes for the survival priority.	5.1.2.1.3 Using the Shutdown Configuration Wizard 5.1.2.2.2 Using the Shutdown Configuration Wizard 5.1.2.3.2 Using the Shutdown Configuration Wizard 5.1.2.4.2 Using the Shutdown Configuration Wizard	
Changed the operation procedure when changing the IP address of the public LAN.	9.2.1 Changing an IP Address on the Public LAN	
Changed the descriptions of the OS installation, the control domain settings, and the sample configuration of guest domains.	14.1.1.1 Cluster System Between Guest Domains Within a Same Physical Partitions 14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions 14.1.1.3 Cluster System Between Control Domains	
Changed the description of Single-Node Cluster Operations.	16.1.2.1 Operation Mode	
Changed the description of the procedure when setting the IP address of CIP in the exclusive IP (Solaris 10).	16.2.4.6 Global Zone Environment Setup	
Changed the descriptions of the procedure of [Configuration change 2].	17.2.2 Changing Configuration after Setting Prerequisites	
Added the descriptions of the following items: - primary_PPAR-ID - XSCF-name1 - XSCF-name2	17.2.3 Performing Live Migration of the Cluster on the Guest Domains	
Changed the procedure when removing the shared class forcibly.	18.4 Changing the Setting in the Single-User Mode	
Added the procedure for registering Gds resources in userApplication.	18.6 Changing the Setting in the Multi-User Mode and Restoration of the Configuration Information of GFS/GDS	
Added the descriptions of port numbers and service names to be supported to Remarks of the following services: - /milestone/fjsvcldbm - /milestone/smawsf	H.2 SMF Service Lists	
Added the descriptions of the dependency relationship with other services to Remarks of the following service. - /network/fjsvhanet	H.2 SMF Service Lists	

Revision	Location	Edition
Added the description for the information that is necessary for setting the shutdown facility.	5.1.2.1.1 Checking XSCF Information	5.2
Changed the descriptions of the settings for XSCF.	5.1.2.1.3 Using the Shutdown Configuration Wizard	
Added the notes for the resource names of patrol diagnosis facility and procedure resources when checking the registration information of a cluster application.	6.7.2.1 Creating Standby Cluster Applications	
SPARC T5/T7/S7 series were added to the supported hardware.	1.7 Notes When Building a System 5.1 Initial Cluster Setup 9.1.2 Changing a CF Node Name 9.2.1 Changing an IP Address on the Public LAN 9.4.1 Changing the User Name and Password to Control the Console I.3.3 Changing the Settings in Multi-User Mode	5.3
Changed the description of "Synchronize time in the slew mode."	1.7 Notes When Building a System	
Added the notes.	Chapter 6 Building Cluster Applications Chapter 13 Backing Up and Restoring a PRIMECLUSTER System	
Added the note about the environment variables used with the scripts.	6.6 Setting Up Online/Offline Scripts	
Changed the description in "Outline" of MONITORONLY.	6.7.1.1 Creating Cmdline Resources	
Added the note to "Prerequisites for NFS Lock Failover."	6.7.1.2.1 Prerequisites	
Changed the note about the NFS entry of PRIMECLUSTER.	6.7.1.2.1 Prerequisites	
Changed the description about the setting contents of the cluster application.	6.11.1 Setting Contents of a Cluster Application	
Added the note about the maintenance mode.	6.12.2.1.4 Notes When Setting the NULLDETECTOR Flag 7.1.3.1 RMS Tree 7.2.2.6 Entering maintenance mode for Cluster Application	
Changed the note about the environment variables used with the scripts.	6.12.3 Notes on Scripts	
Added the procedure to clear the fault trace of the failure resource.	7.4.1.2 Corrective Action for Failed Resources	
Changed the note.	16.2.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone) 16.3.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone)	
Added the description and changed the note about kernel parameter.	16.2.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone) 16.3.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone) A.5.2 RMS Configuration	

Revision	Location	Edition
Added the description about local class.	Chapter 18 When Using Oracle VM Server for SPARC P2V Tool to Migrating a Cluster System	
Changed the description about the procedure to change the cluster interconnect in the migration destination.	18.6 Changing the Setting in the Multi-User Mode and Restoration of the Configuration Information of GFS/GDS	
Changed the utilized ports and remarks.	H.2 SMF Service Lists	
Changed the description of "Changes."	K.1.1 hvshut command	
Added the note when using the legacy file systems.	6.4.1.1 Design for using ZFS with a Cluster System	5.4
Changed the setting example of HV_NODENAME.	6.6 Setting Up Online/Offline Scripts 6.12.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts	
Added the information when the cluster application transits to Standby state.	6.7.3.1 Setting Exclusive Relationships Between Cluster Applications	
Changed the description in the table of relationship between AutoStartUp and StandbyTransitions.	6.11.1 Setting Contents of a Cluster Application	
Deleted the description when registering multiple Cmdlines in userApplication.	6.12.3 Notes on Scripts	
Changed the descriptions in Note.	10.4.1 Adding Fsystem Resource Dynamically	
Added the description when registering the Cmdline resources for monitoring the guest domain status.	14.2.1.1 During Installation	
Added the procedure necessary if the guest domain name was changed.	17.2.2 Changing Configuration after Setting Prerequisites	
Changed the setup procedure when building the cluster application.	Appendix F SafeCLUSTER Compatibility Function Settings	
Changed the descriptions in "Remarks" of the following SMF services: - /milestone/fjsvcldev - /milestone/smawsf	H.2 SMF Service Lists	

Contents

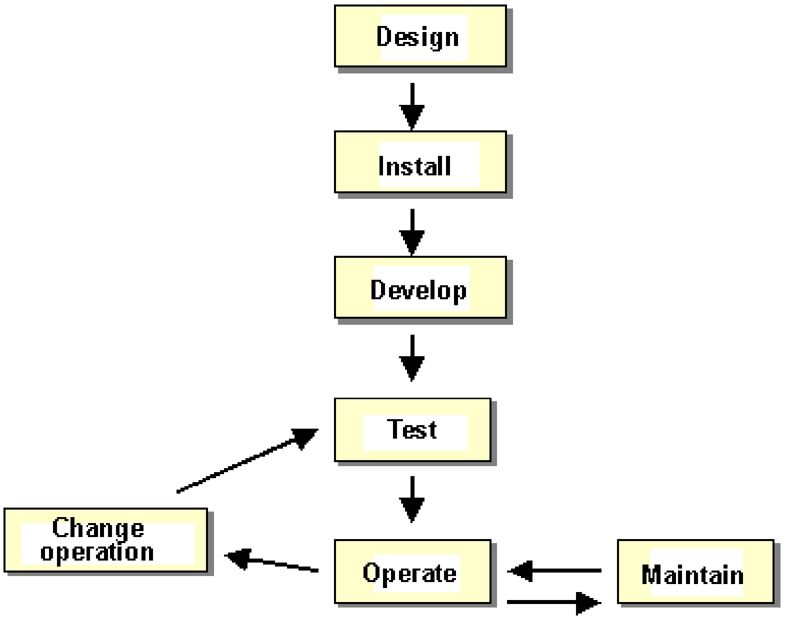
Part 1 Planning

Part 1 describes the workflow from PRIMECLUSTER design to installation and operation management. Users who are installing a PRIMECLUSTER system for the first time need to read this part.

Chapter 1 Build Flow

This chapter describes the workflow for building a PRIMECLUSTER system. To build a PRIMECLUSTER system, follow the procedure described below.

Figure 1.1 Flow of building a PRIMECLUSTER system



1.1 Planning

Before building a PRIMECLUSTER system, you must first design the system.

Designing a PRIMECLUSTER system

1. Select the PRIMECLUSTER products.

   Select the PRIMECLUSTER products required for the system you want to build.

   For details, see "2.1 PRIMECLUSTER Product Selection".

2. Design the system.

   Determine the operation environment for building the PRIMECLUSTER system. This includes selecting the applications to be used and determining the required hardware resources, such as the number of hosts, networks, and disk size.

   For details, see "2.2 System Design".

3. Determine the cluster system operation mode.

   Determine the number of nodes and the operation mode of the cluster system.

   For details, see "2.3 Determining the Cluster System Operation Mode".

4. Determine the operation mode for using Web-Based Admin View.

   Determine the operation mode for running Web-Based Admin View. Web-Based Admin View can manage up to 16 nodes.

   For details, see "2.4 Determining the Web-Based Admin View Operation Mode".

Note

In multi-node (3 or more nodes) operation, the 3-tier model, in which the cluster management server is installed separately, is recommended.

5. **Determine the cluster applications.**

   Determine the number of cluster applications. Also determine which nodes are to be used for each application.

6. **Determine the resources required for each cluster application.**

   Determine the resources required for each cluster application.

   Determine the switchover network type (IP address takeover, MAC address takeover, or node name takeover) and the takeover address.

   - Determine whether a user-defined RMS configuration script is to be used. Determine whether there are other items (for example, line switching units) to be used as resources.
   - Determine the cluster interconnect paths and quantity. Two or more cluster interconnects are recommended.
   - For a disk device, determine which nodes will be sharing the device, whether the device is to be used as a RAW device (database system), whether the device is to be used as a file system (general files), and whether the device is to be grouped.

7. **Determine the failover range of the cluster application.**

   Determine the trigger for cluster application failover.

   For details, see "2.5 Setting the Failover Timing of a Cluster Application."

See

For details on designing the system, see "Chapter 2 Site Preparation".

1.2 Installation

After completing the design of the PRIMECLUSTER system and determining the configuration of the PRIMECLUSTER system to be built, install the PRIMECLUSTER system.

Since the work will be performed based on the PRIMECLUSTER system design worksheet that was created, check that all items on the design worksheet have been entered.

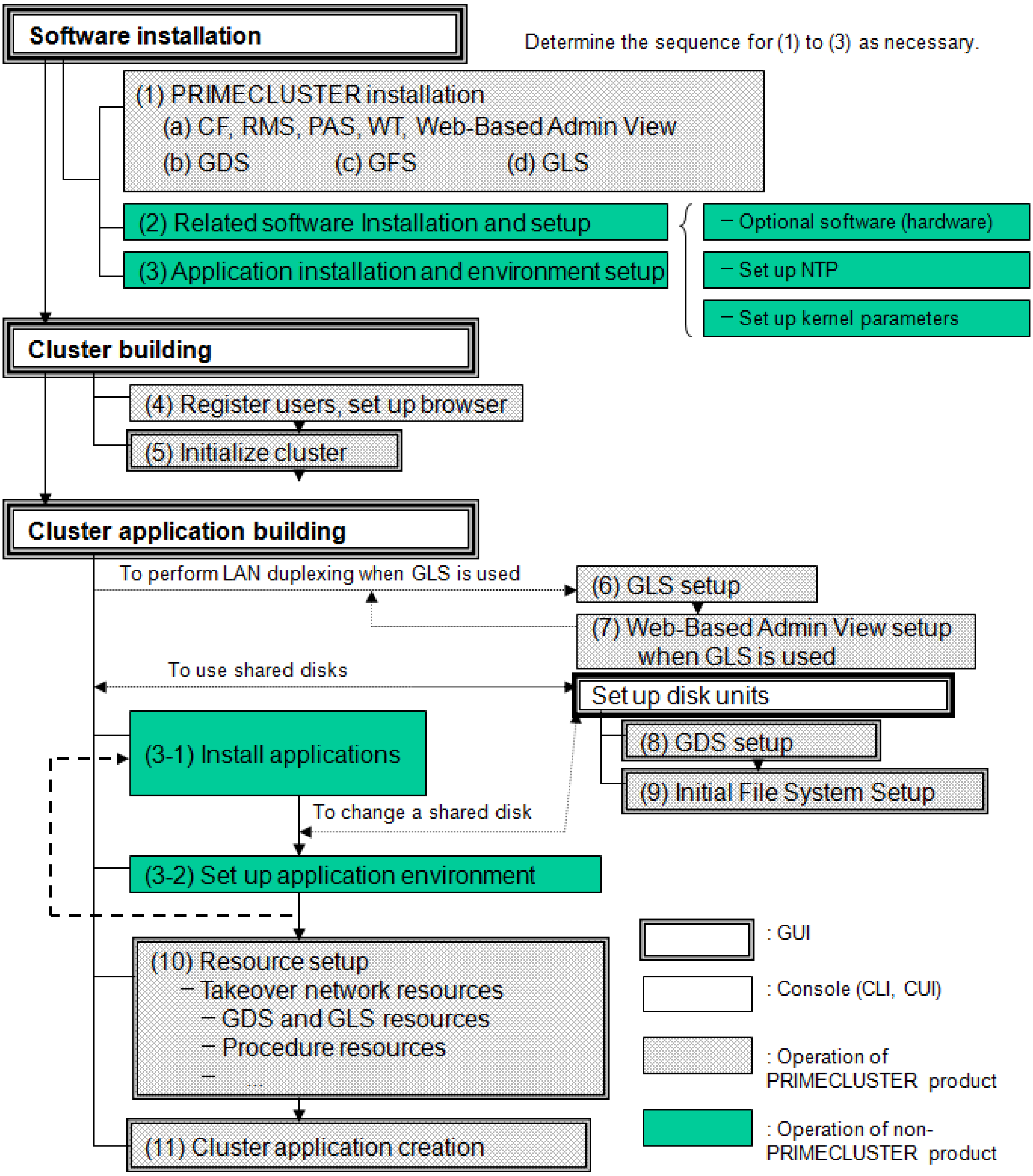
Install the PRIMECLUSTER system by performing the following procedure in sequence from (1).

Perform the operations described in the dotted line sections if the system design matches the described conditions.

If you are installing applications after you install the PRIMECLUSTER system, go back to the operations from the Application environment setup to the Application installation.

The screens to be used differ according to the operation. The work procedures to be performed with GUI from Web-Based Admin View and the work procedures to be performed with CLI and CUI from console screens are shown in separate boxes.

Figure 1.2 Flow of PRIMECLUSTER system installation



The abbreviations in the flowchart for PRIMECLUSTER system installation are explained below.

CF: Cluster Foundation

RMS: Reliant Monitor Services

PAS: Parallel Application Services

WT: Wizard Tools

GDS: Global Disk Services

GFS: Global File Services

GLS: Global Link Services

For detailed information on each item, refer as necessary to the corresponding manual reference section mentioned in the table below.

Table 1.1 Installation procedure and manual reference sections

Work items	Manual reference section	Required/Optional
System design	Chapter 2 Site Preparation	Required
Software installation		
(1) PRIMECLUSTER installation	3.1 PRIMECLUSTER Installation	Required
(2) Related software Installation and setup	3.2 Installation and Setup of Related Software	Optional
(3) Application installation and environment setup	3.3 Installation and Environment Setup of Applications	Optional
Cluster configuration		
(4) User registration/browser setup	Chapter 4 Preparation Prior to Building a Cluster	Required
(5) Initial cluster setup	5.1 Initial Cluster Setup	Required
Cluster application building		
(6) GLS setup	6.2 Initial GLS Setup	Optional (required to use GLS)
(7) Web-Based Admin View setup when GLS is used		
(8) GDS setup	6.3 Initial GDS Setup	Optional (required to use GDS)
(9) File system setup (file system creation)	6.4 Initial File System Setup	Optional (required to use ZFS)
(10) Resource setup	6.7.1 Setting Up Resources	Optional
(11) Cluster application creation	6.7 Setting Up Cluster Applications	Required
Cluster configuration in Oracle VM Server for SPARC Environment		
(12) Cluster systems in Oracle VM Server for SPARC environment	Chapter 14 Using PRIMECLUSTER in Oracle VM Server for SPARC Environment	Optional
Cluster configuration in Oracle Solaris Kernel Zones Environment		
(13) Cluster system in Oracle Solaris Kernel Zones environment	Chapter 15 Using PRIMECLUSTER in Oracle Solaris Kernel Zones Environment	Optional
Cluster configuration in Oracle Solaris Zones Environment		
(14) Cluster systems in Oracle Solaris Zones environment	Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment	Optional

GLS: Global Link Services

GDS: Global Disk Services

1.3 Development

To monitor a user application using PRIMECLUSTER, you need to create an RMS configuration script.

- Online script

  This script executes a process that sets the resources to Online or Standby.

- Offline script

  This script executes a process that sets the resources to Offline.

To check the state of a user application, the following RMS configuration script must be developed.

- Check script

  This script checks the state of the resource.

See

For details on the Online/Offline script and the Check script settings, see "6.6 Setting Up Online/Offline Scripts."

1.4 Test

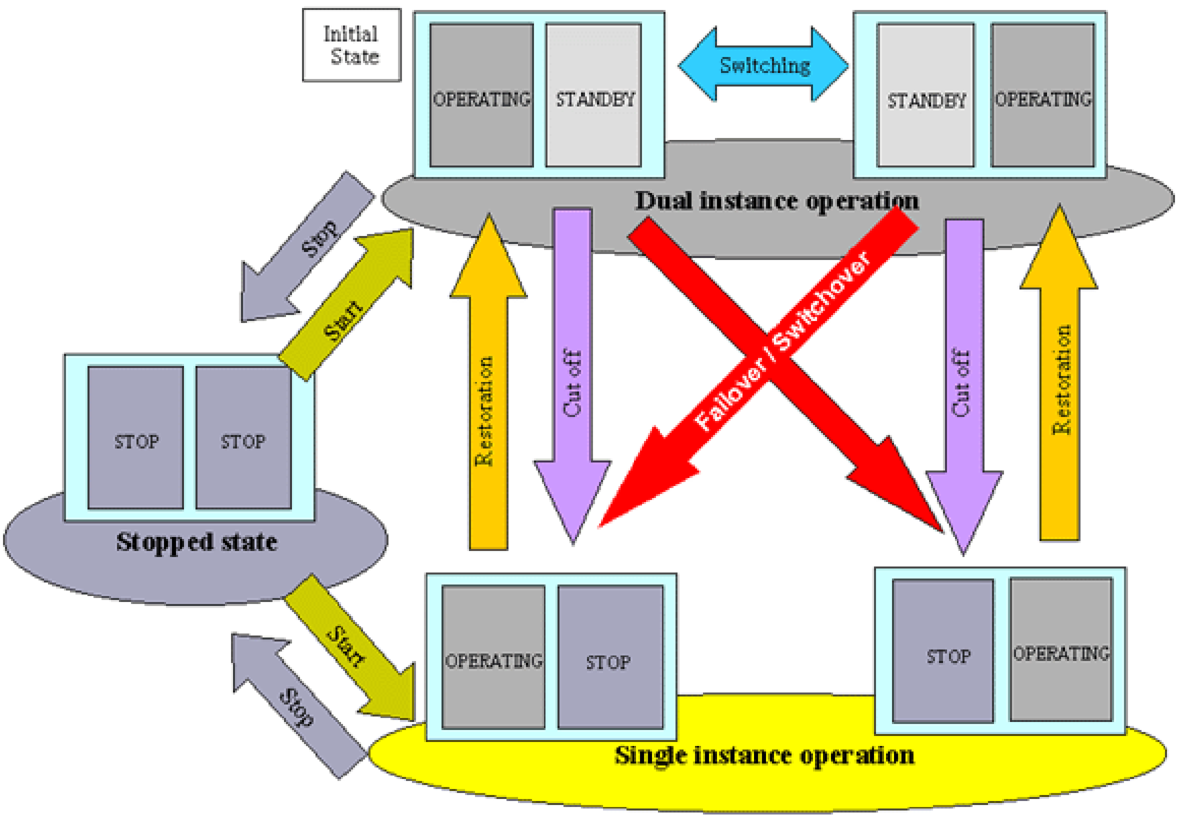
Purpose

When you build a cluster system using PRIMECLUSTER, you need to confirm before starting production operations that the entire system will operate normally and cluster applications will continue to run in the event of failures.

For 1:1 standby operation, the PRIMECLUSTER system takes an operation mode like the one shown in the figure below.

The PRIMECLUSTER system switches to different operation modes according to the state transitions shown in the figure below. To check that the system operates normally, you must test all operation modes and each state transition that switches to an operation mode.

Figure 1.3 State transitions of the PRIMECLUSTER system



PRIMECLUSTER System State

State	Description
Dual instance operation	A cluster application is running, and it can switch to the other instance in the event of a failure (failover). Two types of the dual instance operation are OPERATING and STANDBY.

State	Description
	Even if an error occurs while the system is operating, the standby system takes over ongoing operations as an operating system. This operation ensures the availability of the cluster application even after failover.
Single instance operation	A cluster application is running, but failover is disabled. Two types of the single instance operation are OPERATING and STOP. Since the standby system is not supported in this operation, a cluster application cannot switch to other instance in the event of a failure. So, ongoing operations are disrupted.
Stopped state	A cluster application is stopped.

The above-mentioned "OPERATING", "STANDBY", and "STOP" are defined by the state of RMS and cluster application as follows;

State	RMS state	Cluster application state	Remark
OPERATING	Operating	Online	-
STANDBY	Operating	Offline or Standby	-
STOP	Stopped	Unknown *	SysNode is Offline

* This state is shown when referring to the cluster application is stopped (STOP) in the state display icon of the rms tab on the GUI (Cluster Admin).

Main tests for PRIMECLUSTER system operation

Startup test

Conduct a startup test and confirm the following:

- View the Cluster Admin screen of Web-Based Admin View, and check that the cluster system starts as designed when the startup operation is executed.
- If an RMS configuration script was created, check that the commands written in the script are executed properly as follows.
    - For a command that outputs a message when it is executed, check that a message indicating that the command was executed properly is displayed on the console.
    - Check that the command has been executed properly by executing the "ps(1)" command.
- A new cluster application is not started automatically during the PRIMECLUSTER system startup. To start the cluster application automatically, you must set "AutoStartUp" for that cluster application. The AutoStartUp setting must be specified as a userApplication attribute when the application is created. For details, see "6.7.2 Creating Cluster Applications."

Clear fault

If a failure occurs in a cluster application, the state of that application changes to Faulted.

To build and run this application in a cluster system again, you need to execute "Clear Fault" and clear the Faulted state.

Conduct a clear-fault test and confirm the following:

- Check that the Faulted state of a failed application can be cleared without disrupting ongoing operations.
- If an RMS configuration script was created, check that the commands written in the script are executed properly as follows.
    - For a command that outputs a message when it is executed, check that a message indicating that the command was executed properly is displayed on the console.
    - Check that the command has been executed properly by executing the "ps(1)" command.

Switchover

Conduct a failover or switchover test and confirm the following:

- Check that failover is triggered by the following event:
    - When an application failure occurs
- Check that switchover is triggered by the following events:
    - When the OPERATING node is shut down
    - When an application is terminated by the exit operation
    - When an OPERATING cluster application is stopped
- Check that failover or switchover is normally done for the following:
    - Disk switchover

      Check that the disk can be accessed from the OPERATING node.

      For a switchover disk, you need to check whether a file system is mounted on the disk by executing the "df(1M)" command.
    - If the Cmdline resources are to be used, check that the commands written in the Start and Stop scripts for the Cmdline resources are executed properly.
        - For a command that outputs a message when it is executed, check that a message indicating that the command was executed properly is displayed on the console.
        - Check that the command has been executed properly by executing the "ps(1)" command.
- If IP address takeover is set, check that the process takes place normally by executing the "ifconfig(1M)" command.
- If MAC address takeover is set, check that the process takes place normally by executing the "ifconfig(1M)" command.
- If node name takeover is set, check that both the OPERATING and STANDBY nodes have the same node name after network takeover.
- If a line switching unit is set up, check that line switching takes place correctly.
- Check that an application is switched to other node.

You need to know the operation downtime in the event of a failure, so measure the switching time for each failure detection cause and check the recovery time.

Replacement test

Conduct a replacement and confirm the following:

- Check that the OPERATING and STANDBY instances of the OPERATING business application occur normally when the cluster application replacement is executed. Check the following:
    - If disk switchover is to be used, check that the disk can be accessed from the OPERATING node but not from the STANDBY node.

      For a switchover disk, you need to check whether a file system is mounted on the disk by executing the "df(1M)" command.
    - If Cmdline resources are to be used, check that the commands written in the Start and Stop scripts for the Cmdline resources are executed properly.
        - For a command that outputs a message when it is executed, check that a message indicating that the command was executed properly is displayed on the console.
        - Check that the command has been executed properly by executing the "ps(1)" command.
    - If IP address takeover is to be used, check that IP address takeover takes place normally.
    - Check that an application is switched to other node.

Stop

Conduct a stop test and confirm the following:

- Check that an OPERATING work process can be stopped normally by the stop operation.
- Check that work processes can be started by restarting all nodes simultaneously.

- If Cmdline resources are to be used, check that the commands written in the Start and Stop scripts for the Cmdline resources are executed properly.

  - For a command that outputs a message when it is executed, check that a message indicating that the command was executed properly is displayed on the console.

  - Check that the command has been executed properly by executing the "ps(1)" command.

Work process continuity

Conduct work process continuity and confirm the following:

- Generating some state transitions in a cluster system, check that the application operates normally without triggering inconsistencies in the application data in the event of a failure.

- For systems in which work processes are built as server/client systems, check that while a state transition is generated in the cluster system, work process services can continue to be used by clients, according to the specifications.

Cluster Node Forced Stop Test

Please check that the shutdown facility's settings are properly functioning.

With a view to the following, please conduct a test of whether or not there has once been a stop to the cluster nodes of which the cluster is comprised

- Check that the cluster nodes which are undergoing an error or are causing an OS error have undergone a forced stop.

- Check that the cluster interconnect has been disconnected and that the low-priority cluster nodes have undergone a forced stop.

Note

So as to detect an NIC linkdown event on both paths, please disconnect the cluster interconnect.

For example, if the both nodes are connected through a switch instead of being connected directly, please disconnect the two cluster interconnects from the same node side. If you perform a method of disconnection that does not allow for the detection of an NIC linkdown event on both paths, there will be time differences in detecting an error for each route and the node that detected the error first will have priority and stop peer node forcibly.

Also, please check that the crash dump for the cluster nodes that underwent forced stop has been collected.

See

- For information on the operation procedures for start, clear fault, failover, switchover, and stop, see "7.2 Operating the PRIMECLUSTER System".

- For descriptions of IP address takeover and node name takeover, see "6.7.1.5 Creating Takeover Network Resources".

1.5 Operation and Maintenance

After confirming that work processes can be continued no matter what state the cluster system lapses into, you can begin actual operations.

Note

The cluster system can continue work processes even if a failure occurs. However, work processes cannot be continued if another failure occurs during single node operation before the first failure is corrected. To enhance reliability, you need to eliminate the cause of the failure immediately and recover the dual node operation.

See

For details for collecting information required for an investigation, see "Appendix C Troubleshooting".

1.6 Operation Mode Change

Change the operation of the PRIMECLUSTER system when it becomes necessary to change the system configuration while the PRIMECLUSTER system is operating. The system configuration must be changed, for example, when a cluster application is added.

See

For details on changing the operation mode, see "Part 4 System Configuration Modification".

1.7 Notes When Building a System

This chapter describes notes you should be well aware of when building a PRIMECLUSTER system. Be sure to read through this before you start operation.

Synchronize time on all nodes to configure a cluster system

Connect to the NTP server and synchronize time on all nodes.
If the time is not synchronized on all nodes, a cluster may not operate properly.

For example, the following message is output or if the OnlinePriority attribute of the cluster application is set, the cluster application may not become Online on the desired node because the node, which was the last operating node at RMS startup, cannot be judged correctly.

(WRP, 34) Cluster host host is no longer in time sync with local node. Sane operation of RMS can no longer be guaranteed. Further out-of-sync messages will appear in the syslog.

(WRP, 35) Cluster host host is no longer in time sync with local node. Sane operation of RMS can no longer be guaranteed.

Synchronize time in the slew mode

To synchronize time on each node with NTP, use the slew mode to always adjust the time slowly. Do not choose the step mode, which is used for adjust the time rapidly.
For details, see the manual of OS and so on. Rapid time adjustment using NTP or time adjustment using running date command causes time inconsistency between nodes, which leads to the incorrect operation of cluster system.

Do no set Spanning Tree Protocol to cluster interconnects

If you set Spanning Tree Protocol to cluster interconnects, the access between them is suspended. Thus, a heartbeat communication may fail. Set Disable to the Status of Parameter Setting for STP (Spanning Tree Protocol) through a switching hub.

Do not set a filtering function in routes of cluster interconnects

The cluster interconnects in PRIMECLUSTER bundle multiple lines to perform communication with PRIMECLUSTER's own protocol (ICF protocol). Therefore, they cannot communicate with devices other than cluster nodes connected to the cluster interconnects. Thus, do not set the filtering function in routes of the cluster interconnects.

Set up kernel parameters necessary in a cluster

PRIMECLUSTER is operated by using a system resource. If this resource is insufficient, PRIMECLUSTER may not operate properly.

The volume of resources used in a system is set as a kernel parameter.
It varies depending on an environment on which your system is running. Estimate the volume of applicable resources based on the operation environment.

Moreover, change kernel parameters before building PRIMECLUSTER.
In addition to that, when you change kernel parameters, be sure to restart OS.

See

For details on a parameter value, see "A.5 Kernel Parameter Worksheet."

Enable system to collect a system dump or a clash dump

If either a system dump or a clash dump cannot be collected, it may take time to investigate the cause when a problem occurs. Moreover, it may not be able to identify its root cause.

Check that you can collect a system dump and a clash dump before building PRIMECLUSTER.

Configure the required Shutdown Facility depending on a server to be used

The required Shutdown Facility varies depending on a server to be used. See "5.1.2 Configuring the Shutdown Facility" to check the required Shutdown Facility according to a server that is to be used. After that, configure it.

Set the time to detect CF heartbeat timeout as necessary

For the time to detect CF heartbeat timeout, you should consider operational volumes at a peak hour, and then set it based on your customer's environment. The value should be about 10 seconds to 1 minute. The default value is 10 seconds.

See

For the method of setting the time to detect CF heartbeat timeout, see "2.1.2 cfset" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

Make sure to set the environment variable: RELIANT_SHUT_MIN_WAIT specifying the RMS shutdown wait time

The required time to stop RMS and cluster applications varies depending on an environment. Be sure to estimate its value corresponding to the configuration setup, and then set it.

See

For details on RELIANT_SHUT_MIN_WAIT, see "13.2 Global environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

For the method of referring to and changing RMS environment variables, see "13.1 Setting environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Set a communicable address to the IP address of the administrative LAN for Shutdown Facility regardless of the operation status of the cluster

If the IP address to which the availability of communication is dynamically changed according to the operation status of the cluster is set, the Shutdown Facility does not operate properly.
For example, if the IP address that was set to the NIC switching mode (physical IP address takeover) of Global Link Services (hereinafter GLS) is set to the administrative LAN for the Shutdown Facility, the availability of communication is changed depending on the startup or stop of GLS. If GLS is stopped and the communication is disabled, the Shutdown Facility does not operate properly.

Check the information of XSCF, ILOM, or ALOM used by the Shutdown Facility

If the settings are incorrect, the Shutdown Facility does not operate properly. The method of checking the information varies depending on the server type. Check the server type and see "5.1.2.1.1 Checking XSCF Information" for SPARC M10, "5.1.2.2.1 Checking Console Configuration" for SPARC Enterprise M3000, M4000, M5000, M8000, or M9000, "5.1.2.3.1 Checking Console Configuration" for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series, and "5.1.2.4.1 Checking Console Configuration" for SPARC Enterprise T1000 or T2000. Moreover, see "5.1.2.5.1 Checking XSCF Information" and "5.1.2.5.2 Checking ILOM Information" for Oracle Solaris Kernel Zones.

Set Locked to the mode switch for SPARC Enterprise Mx000 during operation

If the mode switch is not set properly, the forcible stop by SA_pprcir fails.

When specifying a shared disk unit as the hardware for the patrol diagnosis, set up the physical disk name of a shared disk unit to be the same in all nodes.

When the physical disk name of a shared disk unit varies depending on a node, you cannot set a shared disk unit to the hardware for the patrol diagnosis.

Chapter 2 Site Preparation

You must plan the items listed below before building the PRIMECLUSTER system.

Planning items

- PRIMECLUSTER product selection
- System design
- Cluster system operation mode
- Operation mode for using Web-Based Admin View
- Cluster applications and resources to be used by the cluster applications

Point

An overview of each PRIMECLUSTER product is described in the *"PRIMECLUSTER Concepts Guide."* Be sure to read the guide before designing the PRIMECLUSTER system.

2.1 PRIMECLUSTER Product Selection

The sequence for selecting PRIMECLUSTER products is as follows:

1. Select the products to be used.

   Select necessary PRIMECLUSTER products according to your environment.

   For details, see "2.1.1 Product Selection".

2. Select the functions to be used.

   Check if the products provide the functions you need.

   For details, see "2.1.2 Function Selection".

2.1.1 Product Selection

The product sets described below have been prepared for PRIMECLUSTER. Select the necessary products according to how the system will be used.

- **PRIMECLUSTER Enterprise Edition (EE)**

  All-in-one cluster providing the switching (HA) cluster and parallel database.

  This product is used for scalable operations, such as Oracle RAC and Symfoware.

- **PRIMECLUSTER HA Server (HA)**

  Switchover-type cluster system that features HA (switchover) cluster functions, volume management functions, system functions, and network multiplexing functions.

The following table shows the components (modules) that are included in each product.

Components		Products	
Names	**Features**	**EE**	**HA**
Cluster Foundation (CF)	Refers to the basic function that is required for user applications or other PRIMECLUSTER services to manage or communicate within the cluster.	Y	Y
Reliant Monitor Services (RMS)	Refers to the software monitoring function that is used to realize high-availability (HA)	Y	Y

Components		Products	
Names	Features	EE	HA
	of the application that is to be executed within the cluster.		
Wizard Tools	Refers to the function that is used to create an application that is to be controlled with RMS.	Y	Y
Web-Based Admin View	Refers to the function for realizing PRIMECLUSTER operations and monitoring with the GUI (management view).	Y	Y
System Information Output Tool	Collects system information required for failure investigation	Y	Y
Cluster Configuration Backup/Restore	Backup and restoration for cluster configuration	Y	Y
Global Link Services (GLS)	Provides highly reliable transmission routes by setting up redundant network.	Y	Y
Global File Services (GFS)	Refers to the function that is used to realize simultaneous access to the shared file system from multiple nodes to which the shared disk device is connected (only in Oracle Solaris 10 environment).	Y	Y
Global Disk Services (GDS)	Refers to the volume management function that is used to improve the availability and manageability of the data stored on the disk device.	Y	Y
Netcompo BASE	Online trace function (for GLS)	Y	Y
Scalable Internet Services (SIS)	Network load balancing	Y	-
Parallel Application Services (PAS)	Refers to the function that enables the high-performance and high-speed communication with the parallel databases.	Y	-

2.1.2 Function Selection

Check if the products provide the necessary functions, using the following documents:

- PRIMECLUSTER basic functions

  For information on the basic functions, see "2.3 PRIMECLUSTER modules" in *"PRIMECLUSTER Concepts Guide."*

- Other feature

  In addition to the PRIMECLUSTER basic functions, the following function is also provided:

  - Process monitoring function

    This function monitors the live state of applications and other processes.
    For details, see "6.7.1.7.1 What Is the Process Monitoring Function?"

  - Patrol diagnosis

    This function periodically diagnoses the following hardware which are connected to standby nodes:

    - Shared disk device

    - Network interface card

    For details, see "6.9 Setting Up Patrol Diagnosis."

2.2 System Design

You can use the following configuration of the cluster system. Use the "PRIMECLUSTER System Design Worksheet" to design the system in either case.

- Virtual Machine function not used
- Virtual Machine function used

The installation of the PRIMECLUSTER system is based on the completed "Appendix A PRIMECLUSTER System Design Worksheets." Make sure to create the worksheet.

Moreover, in a cluster system in SPARC M10, the asynchronous node monitoring can be performed and also the failed node can be forcibly stopped by making the communication route redundant. This is because if a LAN error occurs, you can use another LAN. See "2.2.2 XSCF Configuration in SPARC M10" to determine the communication route configuration for XSCF.

See

For details on the operation environment, see "2. Operating environment" in "PRIMECLUSTER Installation Guide."

Information

When designing the system on the native machine, it may be helpful to see "A.10 System Design Examples."

If using PRIMECLUSTER with Oracle VM Server for SPARC environment, see "Chapter 14 Using PRIMECLUSTER in Oracle VM Server for SPARC Environment."

If using PRIMECLUSTER with Oracle Solaris Kernel Zones environment, see "Chapter 15 Using PRIMECLUSTER in Oracle Solaris Kernel Zones Environment ."

If using PRIMECLUSTER with Oracle Solaris Zones environment, see "Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment."

2.2.1 Virtual Machine Function

The virtual machine function is to operate the PRIMECLUSTER systems in virtualized environments for Oracle Solaris.

There are the following virtualized environments:

- Oracle VM Server for SPARC
- Oracle Solaris Zones

Note

When installing PRIMECLUSTER in a virtual machine environment, do not perform the following procedures:

- Stopping the guest domain and the I/O root domain temporarily (suspend)
- Restarting the guest domain and the I/O root domain from the temporary stopped state(resume)

2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment

2.2.1.1.1 Cluster System Configuration in Oracle VM Server for SPARC Environment

The following cluster system configurations are supported in an Oracle VM Server for SPARC Environment:

- Cluster system between guest domains within a same physical partition (Supported only for SPARC M10)
- Cluster system between guest domains among different physical partitions (Supported only for SPARC M10)
- Cluster system between control domains

Monitoring and notes of each cluster system are as follows.

Cluster system configuration		Availability of building a cluster		Monitoring	Notes
		Guest domain	Control domain		
Cluster system between guest domains	Within the same physical partition	Y	N	- The cluster application error on the guest domain or the I/O root domain -The OS error on the guest domain or the I/O root domain	Since this environment only comprises one physical partition, all of the cluster nodes will be stopped when the physical partition failure occurs. Therefore, this mode is not suitable for the practical business.
	Among the different physical partitions (only between guest domains)	Y	-	- The cluster application error - The OS error on the guest domain or the I/O root domain - The hardware (network, shared disk and the route) faults	- You must build the cluster system between the cabinets. - When the physical partition error occurs, the node (guest domain) becomes LEFTCLUSTER state because the guest domain cannot be stopped forcibly. By building PRIMECLUSTER on the control domains, the cluster application will be switched over automatically.
	Among the different physical partitions (between control domains, also between guest domains)	Y	Y	- The cluster application error - The OS error on the control domain, the guest domain, or the I/O root domain - The hardware (network, shared disk and the route) faults - The physical partition error	You must build the cluster system between the cabinets.
Cluster system between control domains		N	Y	- The cluster application error on the control domain - The control domain OS error - The control domain hardware (network, shared disk and the route) faults - The error of the guest domain status (which is displayed by the ldm list-domain command)	PRIMECLUSTER does not monitor the status of guest domains and applications.

Availability of building a cluster Y: Required N: Not available -: Not required

Note

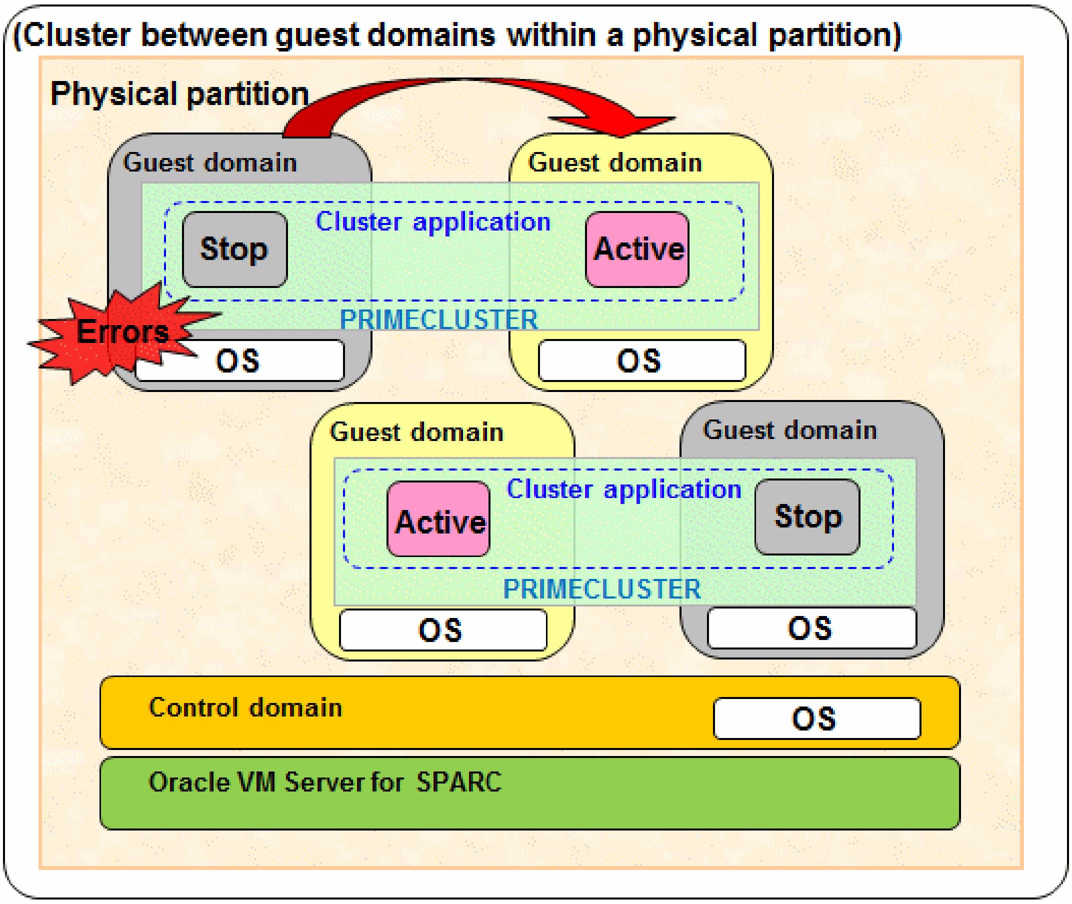
- A tagged VLAN interface cannot be used for the cluster interconnect.
- In the environment where a cluster application is built in the cluster system on the control domain, if an error occurs in the cluster application that is built on the control domain, the error may affect the guest domain to which the virtual device is provided by the control domain.

Cluster system between guest domains within a same physical partition

This configuration enables the cluster system to operate on guest domains or on I/O root domains within a single physical partition. This is effective when verifying operation of cluster applications operated on PRIMECLUSTER. The following types of error monitoring are performed in this configuration. This configuration supports only for SPARC M10.

- The cluster application error on the guest domain or the I/O root domain
- The OS error on the guest domain or the I/O root domain

Figure 2.1 Cluster system between guest domains within a same physical partition



Note

- Since this environment comprises a single physical partition, when the physical partition failure occurred, all cluster nodes are stopped. In consequence, the transaction comes to a stop. Therefore, this mode is not suitable for business operation.
- Specify the same type for domains in the cluster. A cluster cannot be configured with different types of domains, for example, between the guest domain and I/O root domain, or between the control domain and I/O root domain.
- When using a virtual disk as a shared disk of a cluster between guest domains in PRIMECLUSTER, you need to specify a timeout option of the virtual disk.

  [Specifying a timeout option]

  When omitting the timeout option, or specifying 0 to the option, an I/O error does not occur even the control domain or I/O root domain stops. You should wait for the recovery of the control domain or I/O root domain.
  When specifying a value greater than 0 to the timeout option, an I/O error will occur after the specified number of seconds has passed.

  The following explains how to specify a timeout option:

  Example 1: Specifying 15 (seconds) to the timeout when assigning a virtual disk.

```
# ldm add-vdisk timeout=15 vdisk0 disk0@primary-vds0 guest0
```

Example 2: Specifying 15 (seconds) to the timeout for the assigned virtual disk.

```
# ldm set-vdisk timeout=15 vdisk0 guest0
```

For details on the timeout option, see Oracle VM administration guide.

Cluster system between guest domains among the different physical partitions

This enables the cluster system to operate between guest domains or I/O root domains (including an I/O domain) among different physical partitions. In a cluster system that consist of only guest domains and I/O root domains, when the physical partition failure occurred, the nodes that construct the cluster may come into the LEFTCLUSTER state. For dealing with this, installing PRIMECLUSTER switches cluster applications on the guest domain or I/O root domain automatically even when the physical partition failure occurred. The following types of error monitoring are performed in this configuration. This configuration supports only for SPARC M10.

- Cluster application errors on a control domain(*), a guest domain, or an I/O root domain
- OS errors on a control domain(*), a guest domain, or an I/O root domain
- Hardware (network, shared disk and the route) faults
- Physical partition errors(*)

*) Only when PRIMECLUSTER is built on the control domain

However, use this configuration with careful consideration of system design because this function limits other functions.

Note

- When building the cluster system on multiple physical partitions within a single cabinet, the transaction comes to a stop if the cabinet failed. Therefore, you must build the cluster system between the cabinets.
- Specify the same type for domains in the cluster. A cluster cannot be configured with different types of domains, for example, between the guest domain and I/O root domain, or between the control domain and I/O root domain.
- When using a virtual disk as a shared disk of a cluster between guest domains in PRIMECLUSTER, you need to specify a timeout option of the virtual disk.

  [Specifying a timeout option]

  When omitting the timeout option, or specifying 0 to the option, an I/O error does not occur even the control domain or I/O root domain stops. You should wait for the recovery of the control domain or I/O root domain.
  When specifying a value greater than 0 to the timeout option, an I/O error will occur after the specified number of seconds has passed.

  The following explains how to specify a timeout option:

  Example 1: Specifying 15 (seconds) to the timeout when assigning a virtual disk.

  ```
  # ldm add-vdisk timeout=15 vdisk0 disk0@primary-vds0 guest0
  ```

  Example 2: Specifying 15 (seconds) to the timeout for the assigned virtual disk.

  ```
  # ldm set-vdisk timeout=15 vdisk0 guest0
  ```

  For details on the timeout option, see Oracle VM administration guide.

In addition, when PRIMECLUSTER is built on the control domain, note the following points as well:

- When creating the cluster application on the control domain, the guest domain, or the I/O root domain, do not specify the RMS priority (ShutdownPriority) attribute.
- Set the survival priority of the guest domains or I/O root domains so as to be the same order relation as that of the control domain.

- When a failure of the control domain (including the cluster application error) is detected and the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

- When a virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.

Figure 2.2 Cluster system between guest domains among the different physical partitions (only between guest domains)

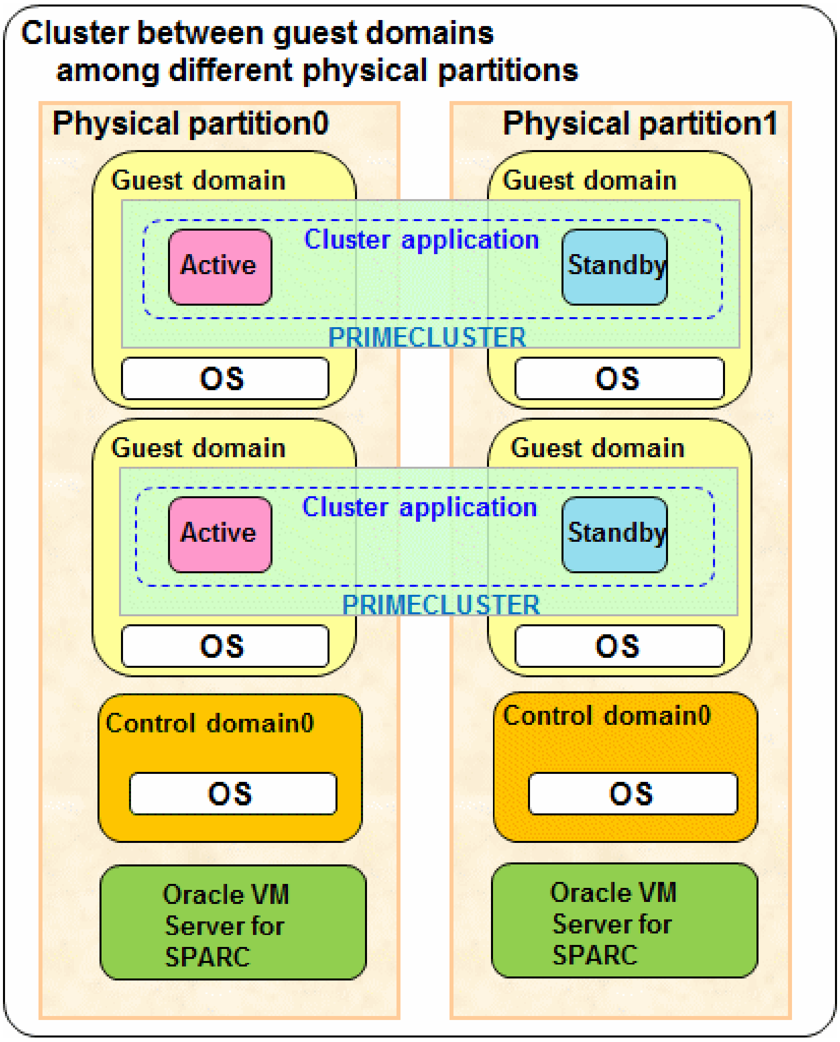


Figure 2.3 Cluster system between guest domains among the different physical partitions (between control domains, also between guest domains)

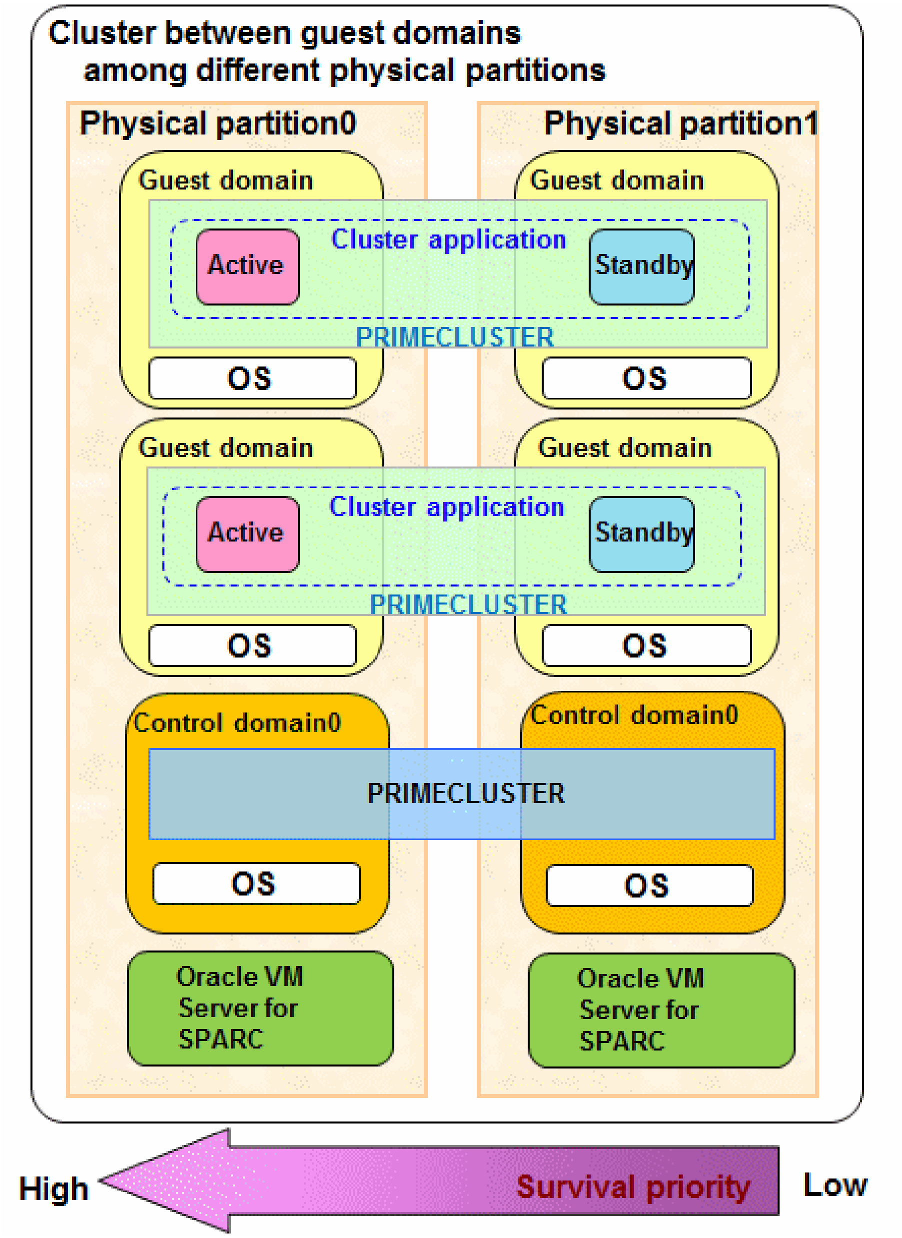
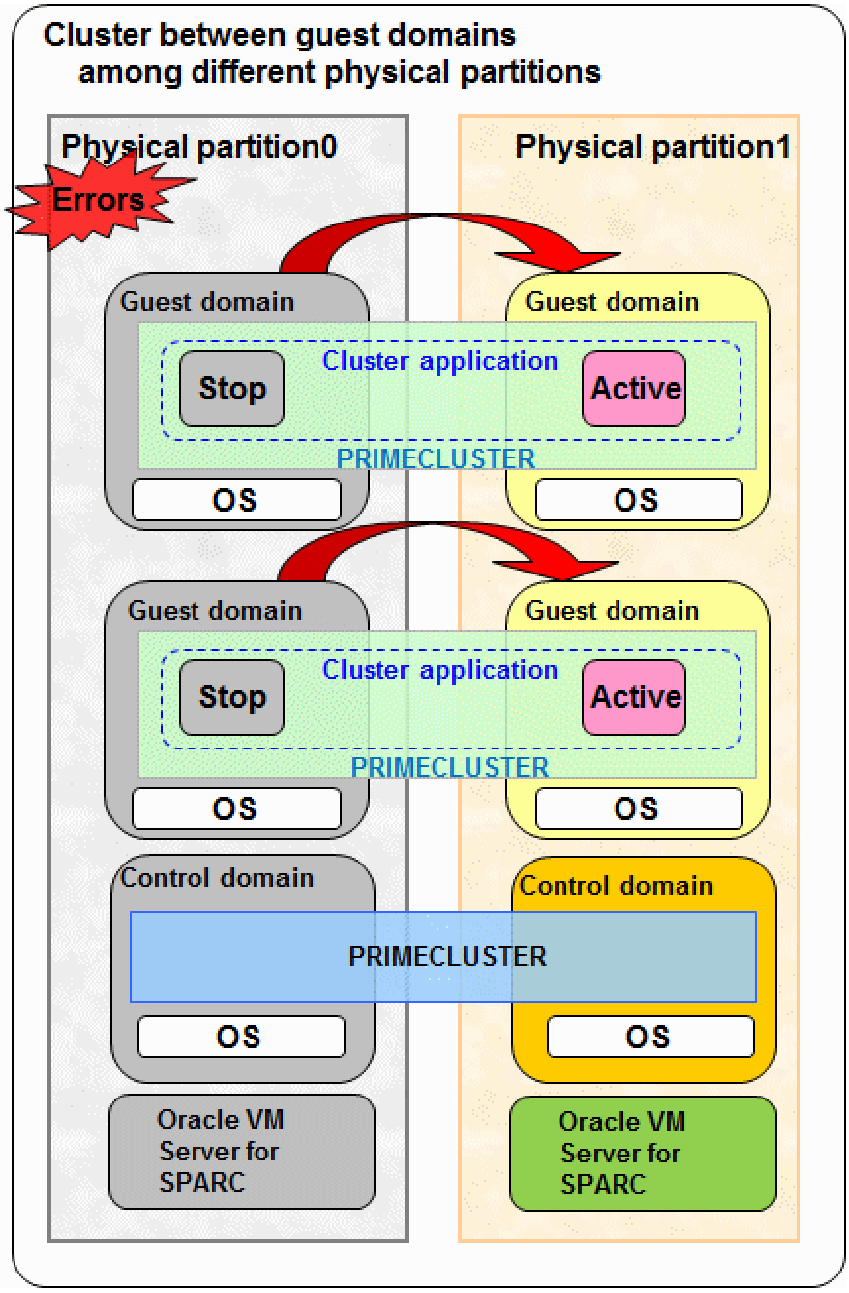


Figure 2.4 Switching image when the physical partition failure occurred



Cluster system between the control domain

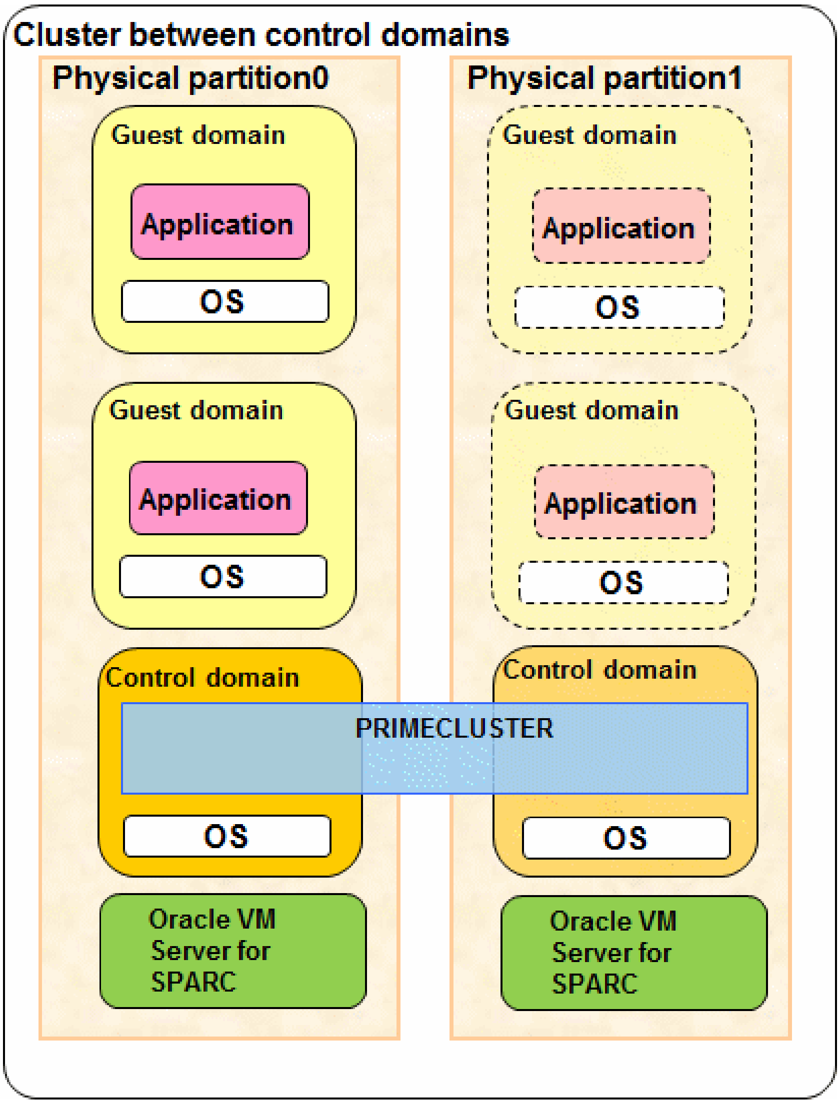
This configuration applies PRIMECLUSTER on the control domain in an environment where the guest domain is configured, so that the cluster on the control domain can monitor the state of the guest domain.

In this configuration, the operation can be continued even when the hardware (networks and disks) failed for starting the guest domain on the other control domain to continue the operation by executing failover of the control domain. Applying PRIMECLUSTER to the control domain monitors the following failures that disable applications on guest domains:

- The cluster application error on the control domain
- The control domain OS error
- The control domain hardware (network, shared disk and the route) fault
- The guest domain status (the state displayed on the ldm list-domain) error

When a failure occurred, the guest domain is switched to the standby system to realize the guest domain environment with high reliability.

Figure 2.5 Cluster system between control domains



Note

- PRIMECLUSTER does not monitor the status of guest domains and applications.

When using the cluster system between the control domains, the redundant line control method supported by GLS is only the NIC switching mode.

Note

- GLS must be installed in both control and guest domains.
- The I/O used in a guest domain must only be assigned to a virtual disk provided in a control domain.
- Multiple guest domains on the same control domain cannot share a GDS shared class. When configuring multiple guest domains, please create shared classes separately for each guest domain.
- When a failure of the control domain (including the cluster application error) is detected and the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

- When the virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.

2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment

Following two types of the Migration function can be used for a cluster system in Oracle VM Server for SPARC Environment:

- Live Migration
  Transferring an active guest domain.
- Cold Migration
  Transferring an inactive guest domain.

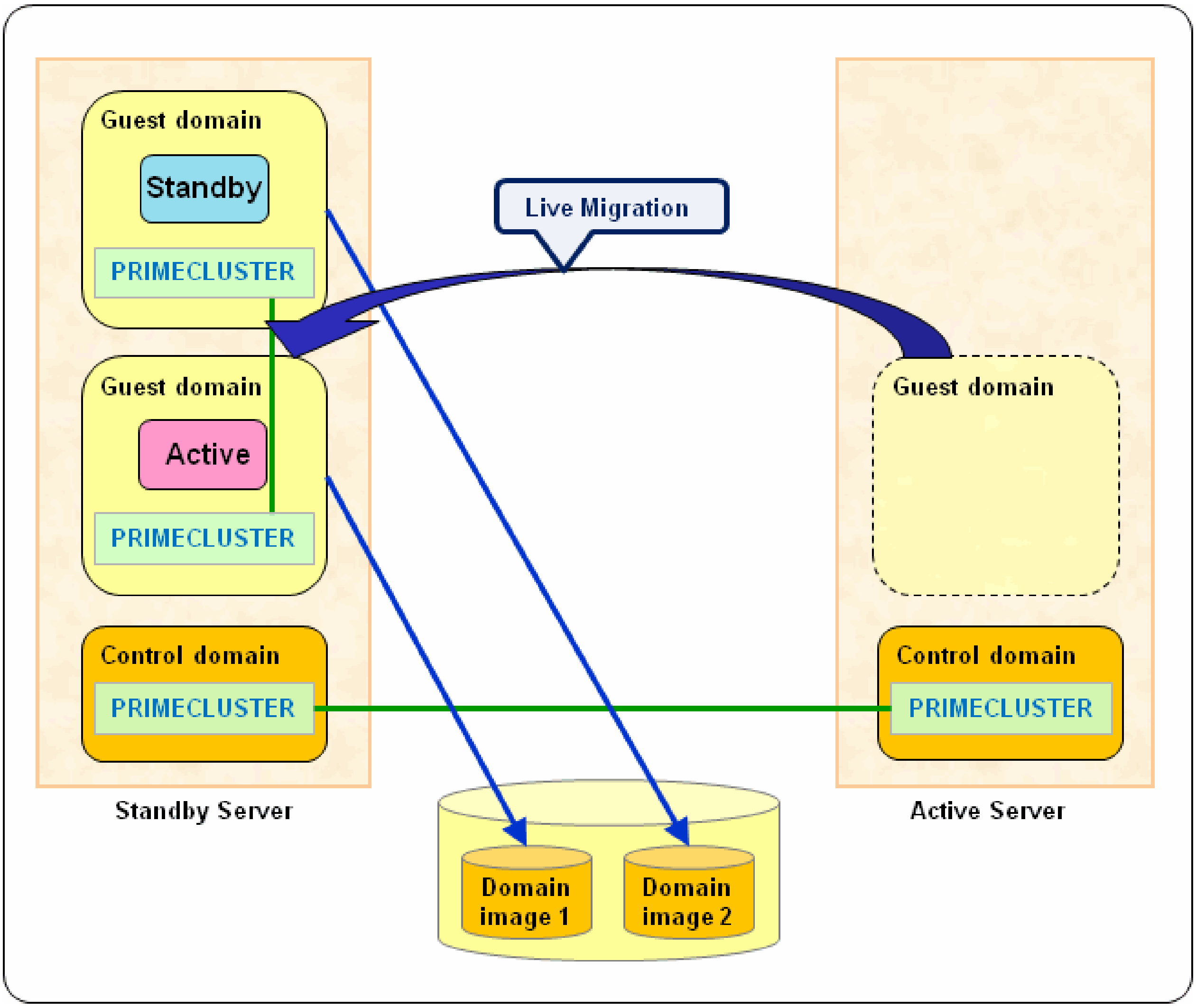
These functions can be used in combination with ServerView Resource Orchestrator Cloud Edition.

The Migration function of Oracle VM Server for SPARC can be used in the following cluster system configuration:

- Among the different physical partitions (only between guest domains)
- Among the different physical partitions (between control domains, also between guest domains)

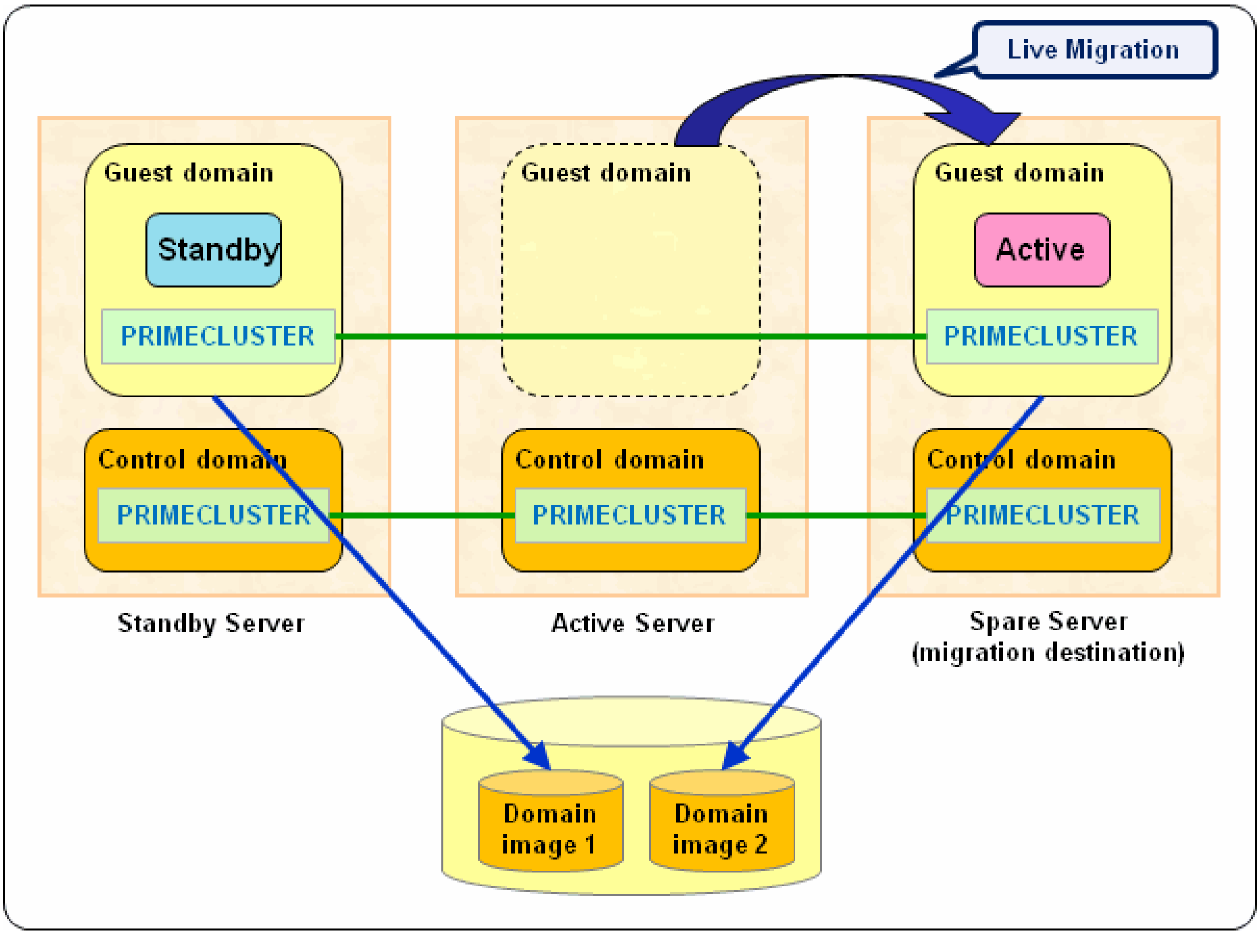
  To use the Migration function of Oracle VM Server for SPARC in a cluster system, you can do server maintenance while keeping a redundant configuration for active and standby servers.

Figure 2.6 Live Migration for a cluster system



You can also do server maintenance while keeping a redundant configuration for active and standby servers between physical partitions by configuring a cluster system using not only active and standby servers but also a spare server in a control domain.

Figure 2.7 Live Migration to a spare server



By the Cold Migration to an inactive guest domain, the guest domain can be started in a spare server.

A redundant configuration for active and standby servers can be maintained even during the maintenance of a standby server.

Figure 2.8 Cold Migration to a spare server (before performing)

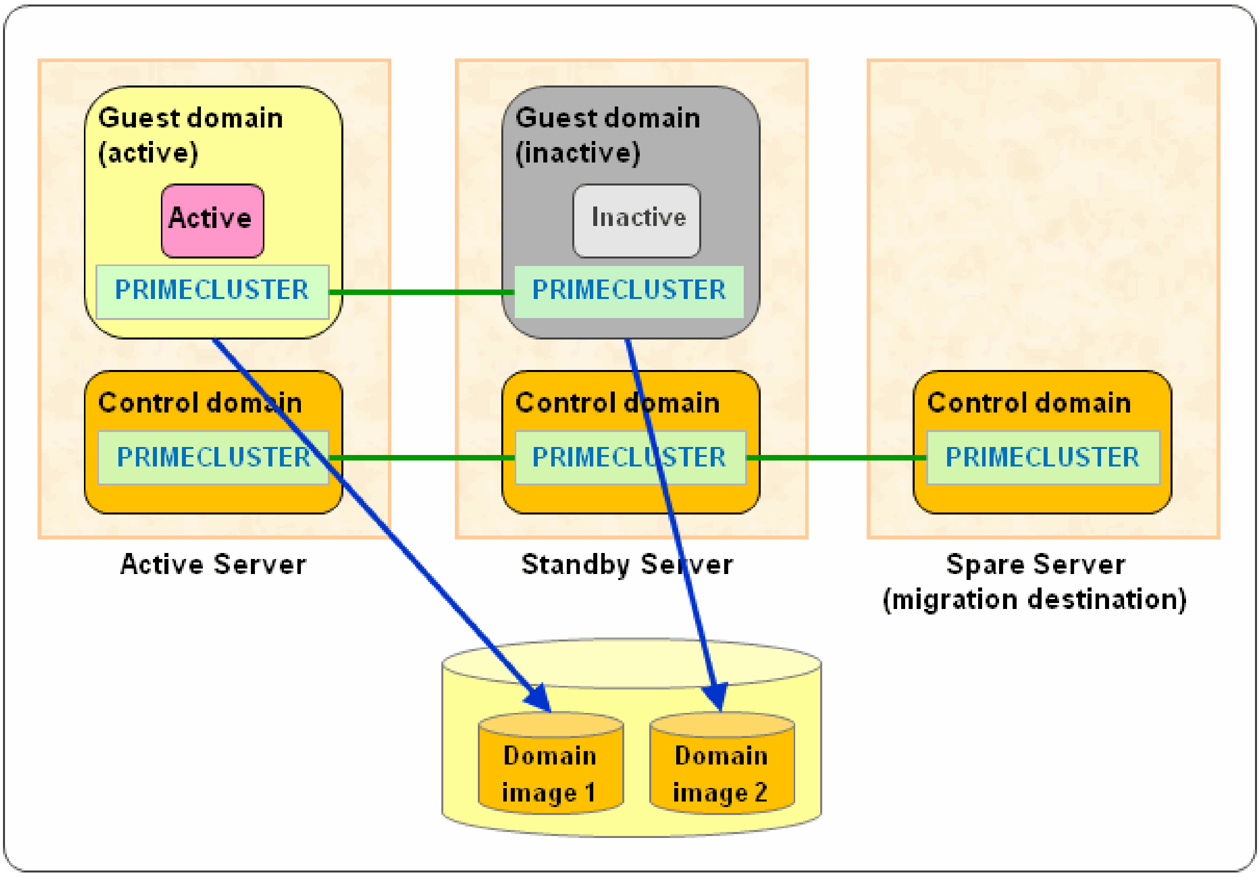


Figure 2.9 Cold Migration to a spare server (in performing)

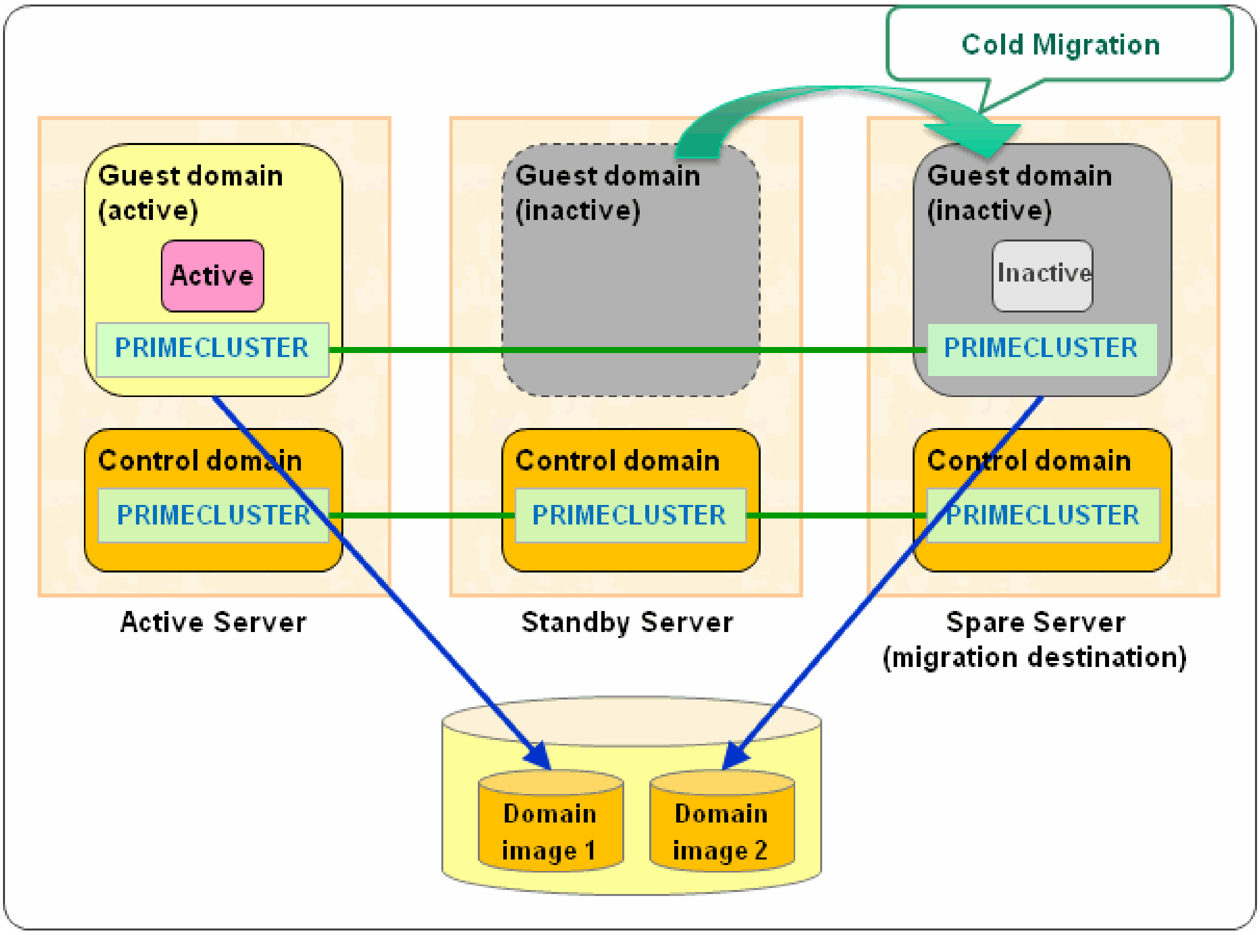
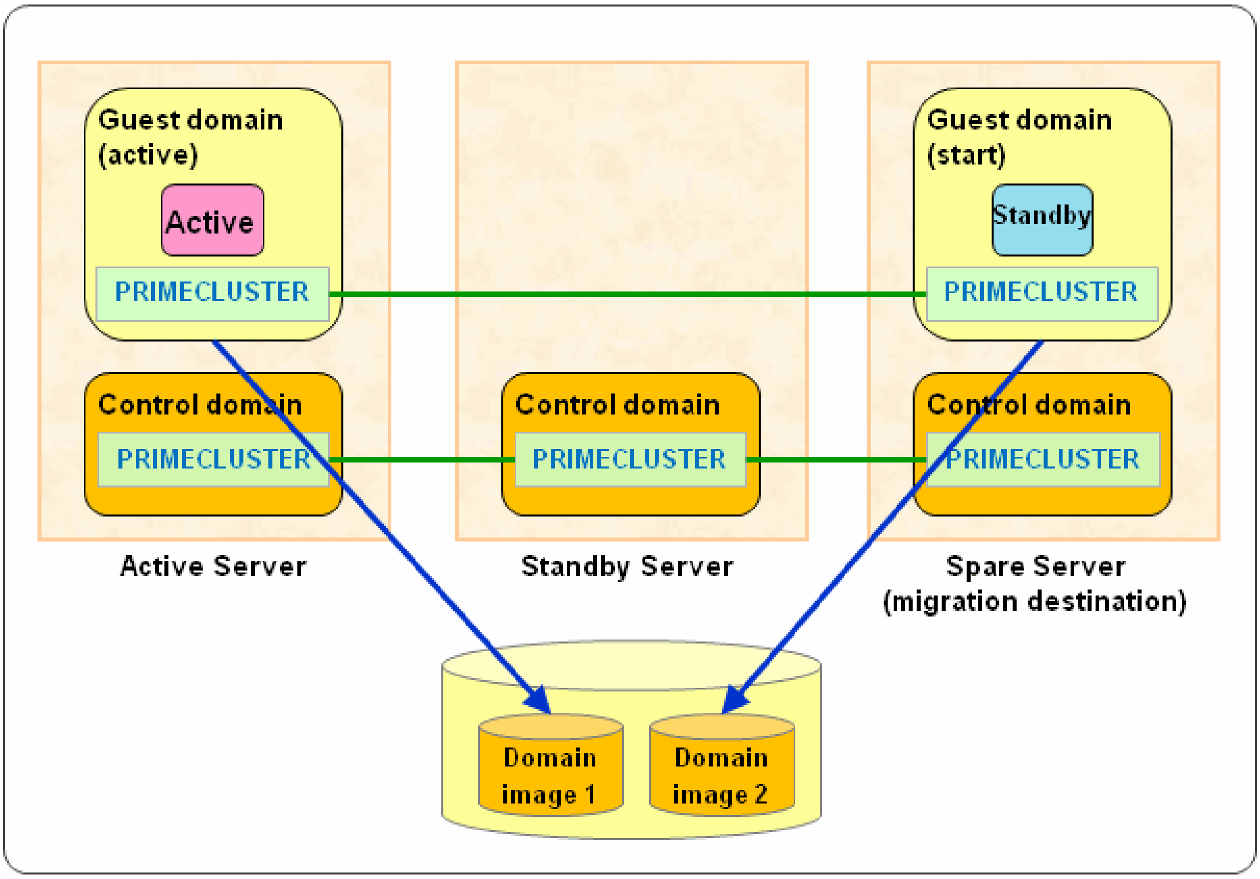


Figure 2.10 Cold Migration to a spare server (after performing)



Prerequisites are needed for using the Migration function of Oracle VM Server for SPARC in a cluster system. For details, see "Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment."

Note

- A cluster system is not switched during the Migration.
- Do not perform the Migration during a cluster system switchover.
- When configuring Oracle Solaris Kernel Zones environment on a guest domain, the Migration cannot be performed.

2.2.1.1.3 When Migrating a Cluster System in the Physical Environment to a Guest Domain in Oracle VM Server for SPARC Environment (Physical to Virtual)

In the physical environment, you can migrate a cluster system that uses PRIMECLUSTER 4.2A00 or later to a guest domain (or I/O root domain) in Oracle VM Server for SPARC Environment (Physical to Virtual: hereafter referred to as P2V). (Only SPARC M10 is supported)

See

- For how to migrate it with P2V, see "Chapter 18 When Using Oracle VM Server for SPARC P2V Tool to Migrating a Cluster System."
- For specification changes of PRIMECLUSTER after migration, see the following:
  - "Appendix K Changes in Each Version"
  - "*PRIMECLUSTER Global Link Services Configuration and Administration Guide 4.3: Redundant Line Control Function*"
  - "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide 4.3*"
- For system requirements and notes on migration, see "*Oracle VM Server for SPARC Administration Guide*."

Figure 2.11 Cluster system before migration

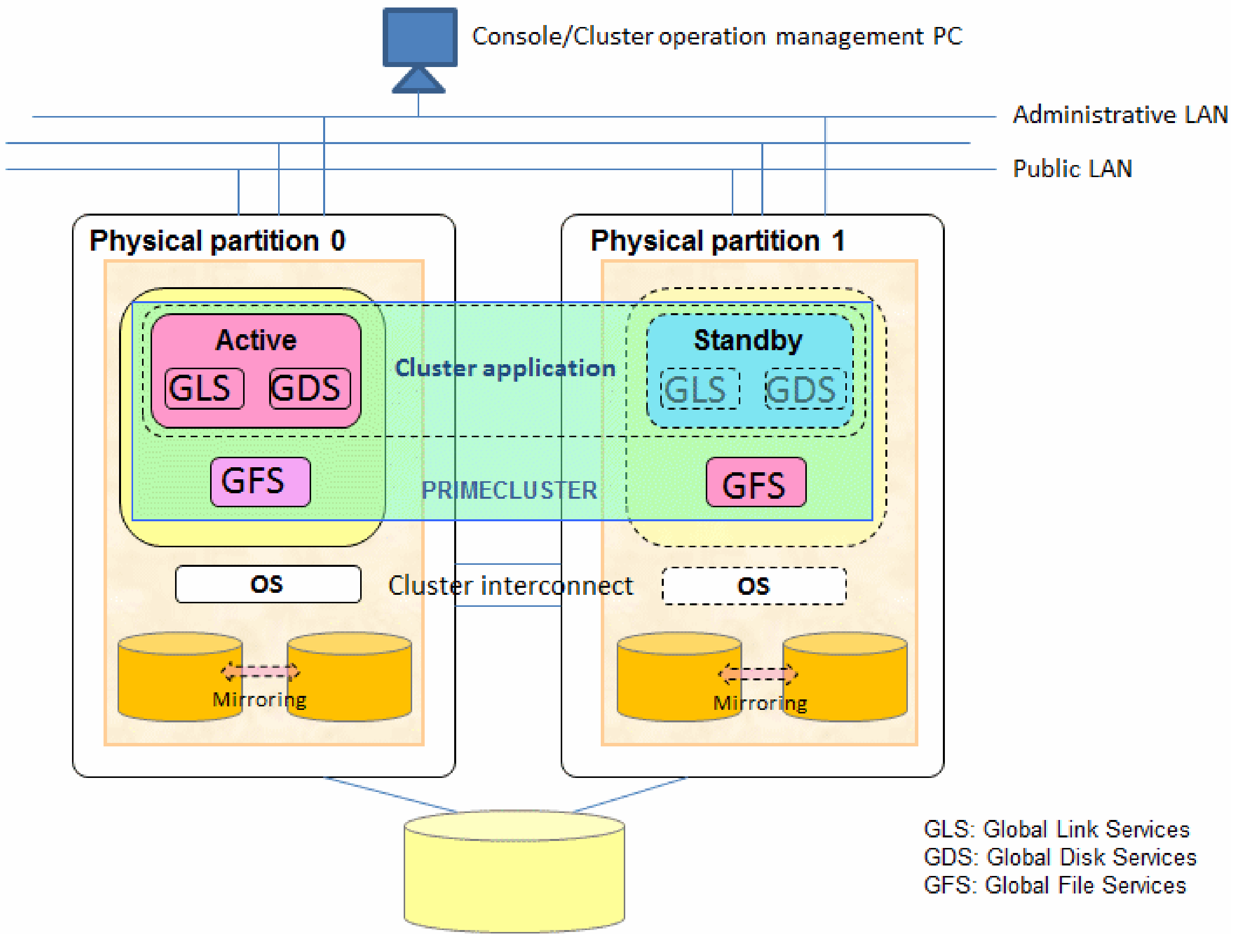
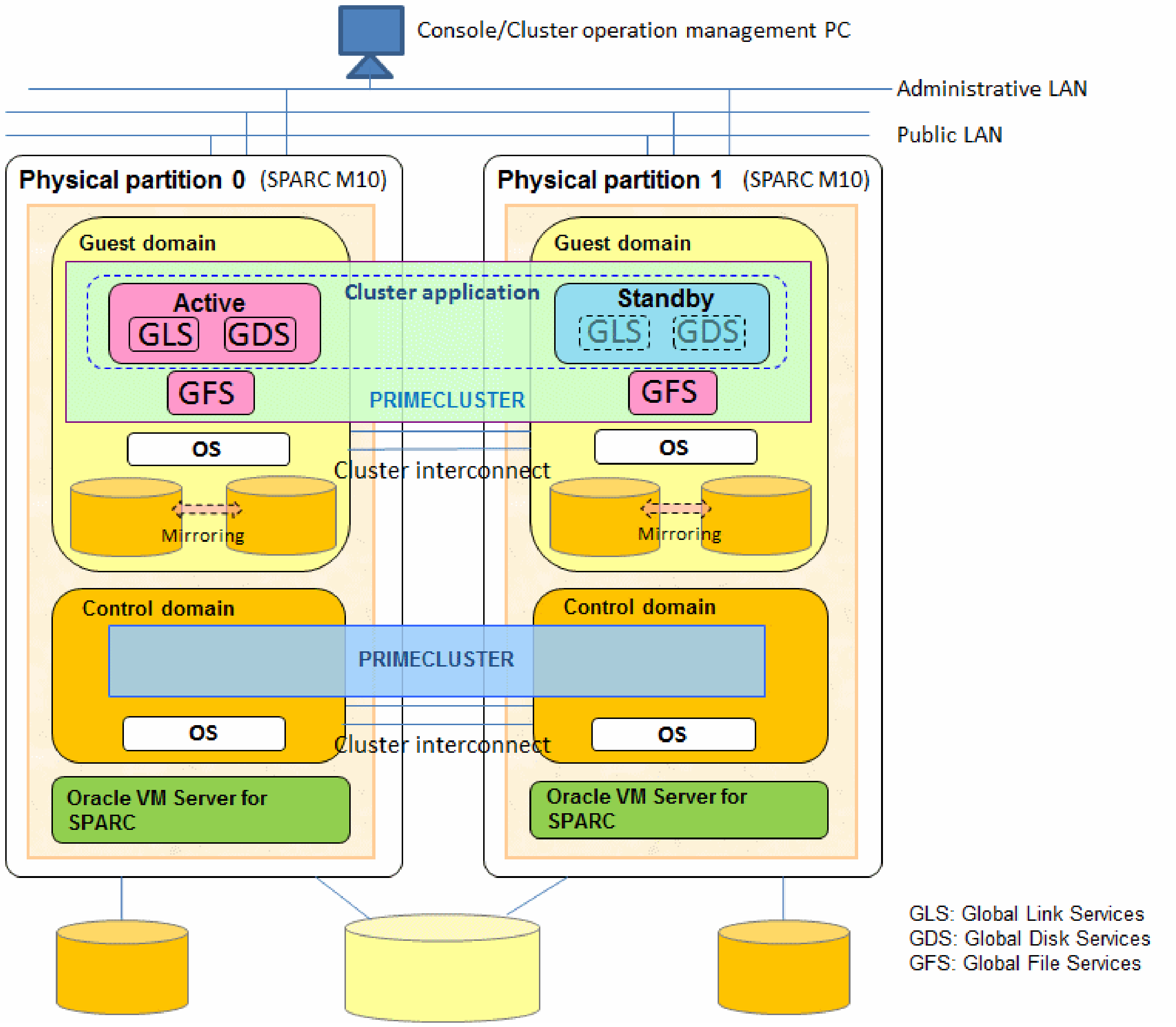


Figure 2.12 Cluster system after migration



System requirement for migration

- PRIMECLUSTER version

  PRIMECLUSTER 4.2A00 or later

- Supported OS

  Solaris 10

- GLS redundant line switching method

  NIC switching mode and GS/SURE linkage mode

- File system in a shared disk

  UFS, ZFS, and Global File Services (hereinafter GFS: only for Solaris 10)

Note

- Use Global Disk Services (hereinafter GDS) and GLS functions in the cluster system on a guest domain after migration.
- The disk size of GDS volumes in a shared disk must be the same in the migration source and migration destination.
- You must migrate user data with ETERNUS storage migration or LUN to LUN such as REC beforehand.

Use the same configuration as the migration source after the migration.

See

If a setting has to be changed after completing the cluster system migration, see the following to change the setting:

- "Part 4 System Configuration Modification"
- "PRIMECLUSTER Global Link Services Configuration and Administration Guide 4.3: Redundant Line Control Function"
- "PRIMECLUSTER Global Disk Services Configuration and Administration Guide 4.3"
- "PRIMECLUSTER Global File Services Configuration and Administration Guide 4.3"

Note

There are some important points when using a cluster system in an Oracle VM Server for SPARC Environment. For details, see "14.2 Precautions on Using Cluster Systems in Oracle VM Server for SPARC Environments."

2.2.1.2 Cluster System Operating in Oracle Solaris Kernel Zones Environment

The following cluster system configurations are supported in Oracle Solaris Kernel Zones environment. You can build an Oracle Solaris Kernel Zones environment in the physical environment, or on the control domain and guest domain in Oracle VM for SPARC environment. The guest domain includes the I/O root domain in this section.

- Cluster system between Kernel Zones within a same physical partition (physical environment/control domain)
- Cluster system between Kernel Zones within a same physical partition (guest domain)
- Cluster system between Kernel Zones among different physical partitions (physical environment/control domain)
- Cluster system between Kernel Zones among different physical partitions (guest domain)

Monitoring and notes of each cluster system are as follows.

Cluster system configuration		Availability of building a cluster			Monitoring	Notes
		Kernel Zones	Guest domain	Control domain		
Within the same physical partition	Cluster system between Kernel Zones (physical environment/ control domain)	Y	-	N	- The cluster application error on the Kernel Zone - The OS error on the Kernel Zone	Since this environment only comprises one physical partition, all of the cluster nodes will be stopped when the physical partition failure occurs. Therefore, this mode is not suitable for the practical business.
	Cluster system between Kernel Zones (guest domain)	Y	Y	N	- The cluster application error on the Kernel Zone or guest domain - The OS error on the Kernel Zone or guest domain	Since this environment only comprises one physical partition, all of the cluster nodes will be stopped when the physical partition failure occurs. Therefore, this mode is not suitable for the practical business.

Cluster system configuration		Availability of building a cluster			Monitoring	Notes
		Kernel Zones	Guest domain	Control domain		
Among the different physical partitions	Cluster system between Kernel Zones (physical environment/ control domain)	Y	-	Y	- The cluster application error - The OS error in the physical environment, on the control domain, or Kernel Zone - The hardware (network, shared disk and the route) failures - The physical partition error	You must build the cluster system between the cabinets.
	Cluster system between Kernel Zones (guest domain)	Y	Y	-	- The cluster application error - The OS error in the physical environment, on the control domain(*), guest domain, or Kernel Zone - The hardware (network, shared disk and the route) failures - The physical partition error(*)	- You must build the cluster system between the cabinets. - If you do not build PRIMECLUSTER in the physical environment or on the control domain, when a physical partition error occurs, the node (guest domain) becomes LEFTCLUSTER state because the guest domain cannot be stopped forcibly. By building PRIMECLUSTER in the physical environment or on the control domains, the cluster application will be switched over automatically.

Availability of building a cluster Y: Required N: Not available -: Not required

*) Only when PRIMECLUSTER is built in the physical environment or on the control domain

Note

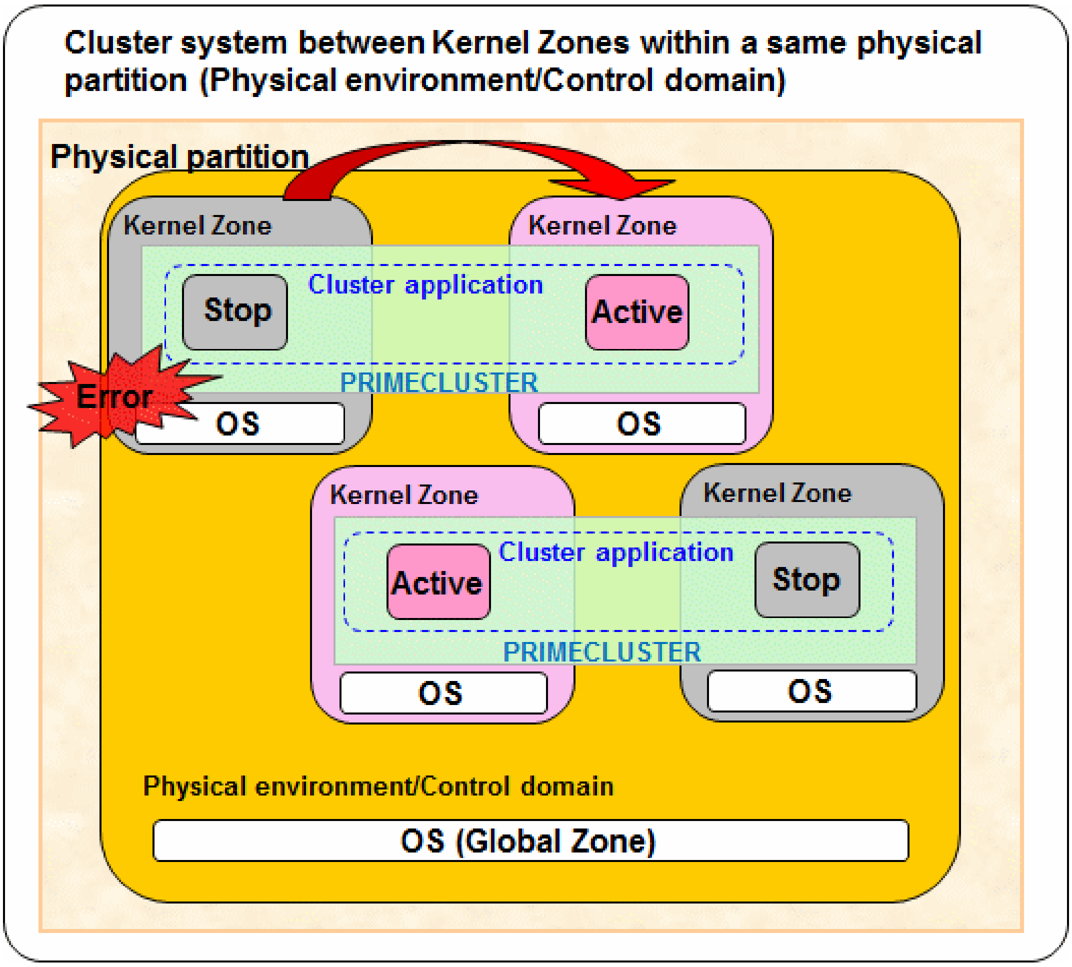
- A tagged VLAN interface cannot be used for the cluster interconnect.
- In the environment where a cluster application is built in the cluster system on the control domain or guest domain, if a cluster application error occurs on the control domain or guest domain, the error may affect Kernel Zones.

Cluster system between Kernel Zones within a same physical partition (physical environment/control domain)

This configuration enables the cluster system to operate between Kernel Zones that are built on the control domain within a single physical partition. This is effective when verifying operation of cluster applications operated on PRIMECLUSTER. The following types of error monitoring are performed in this configuration:

- The cluster application error on the Kernel Zone
- The OS error on the Kernel Zone

Figure 2.13 Cluster system between Kernel Zones within a same physical partition (physical environment/control domain)



Note

Since this environment only comprises one physical partition, all of the cluster nodes will be stopped when the physical partition failure occurs. Therefore, this mode is not suitable for the practical business.

Cluster system between Kernel Zones within a same physical partition (guest domain)

This configuration enables the cluster system to operate between Kernel Zones that are built on the guest domain within a single physical partition. This is effective when verifying operation of cluster applications operated on PRIMECLUSTER.

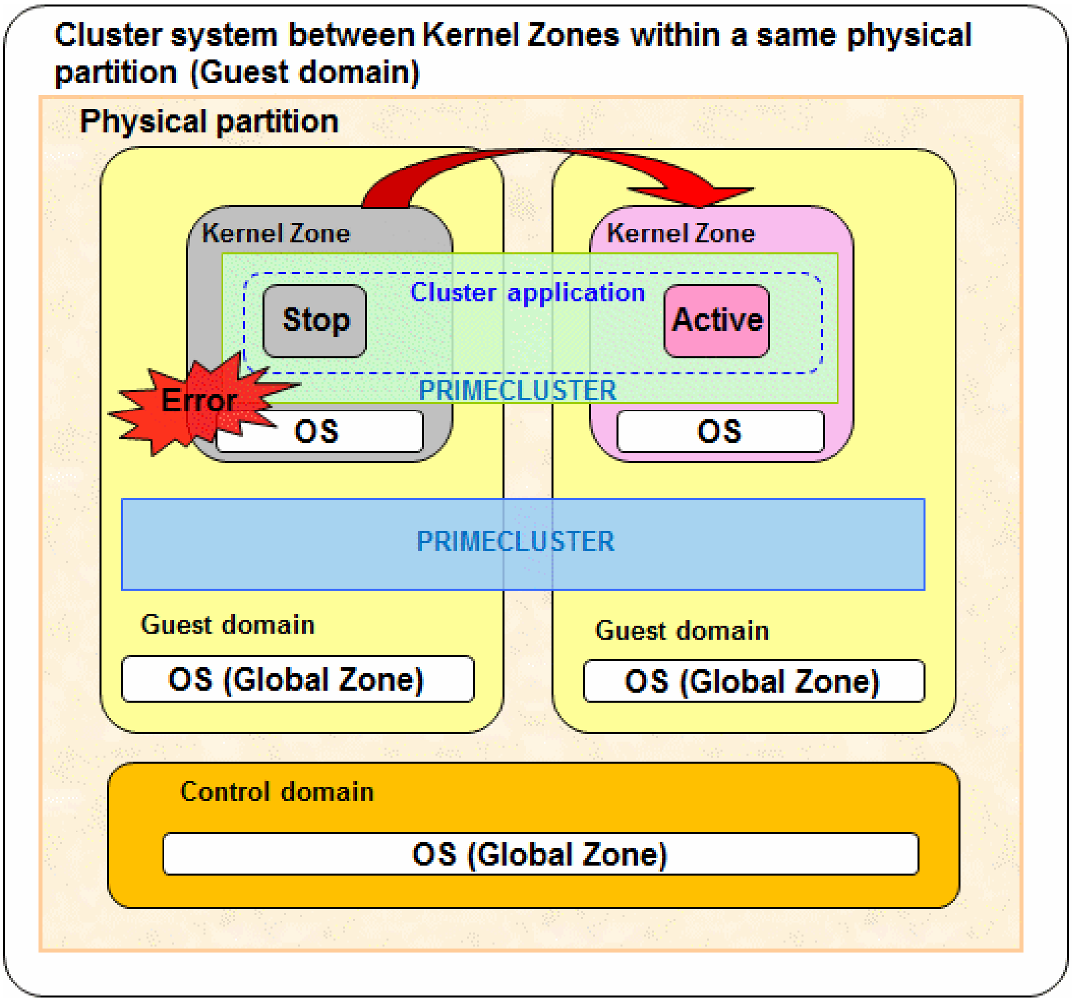
In a cluster system installed only on the Kernel Zones, when the guest domain error occurs, the nodes that construct the cluster may come into the LEFTCLUSTER state. For dealing with this, install PRIMECLUSTER also to the guest domain, and it is possible to switch over the cluster applications on the Kernel Zones automatically even when the guest domain error occurred.

The following types of error monitoring are performed in this configuration.

- The cluster application error on the Kernel Zone or guest domain
- The OS error on the Kernel Zone or guest domain

However, use this configuration with careful consideration of system design because this function limits other functions, such as disabling the RMS priority (ShutdownPriority) setting.

Figure 2.14 Cluster system between Kernel Zones within a same physical partition (guest domain)



Note

- Since this environment only comprises one physical partition, all of the cluster nodes will be stopped when the physical partition failure occurs. Therefore, this mode is not suitable for the practical business.
- When creating the cluster application on the guest domain, I/O root domain, or Kernel Zones, do not specify the RMS priority (ShutdownPriority) attribute.
- Set the survival priority of the Kernel Zones so as to be the same order relation as that of the guest domain.
- Specify the same type for domains in the cluster. A cluster cannot be configured with different types of domains, for example, between the guest domain and I/O root domain, or between the control domain and I/O root domain.

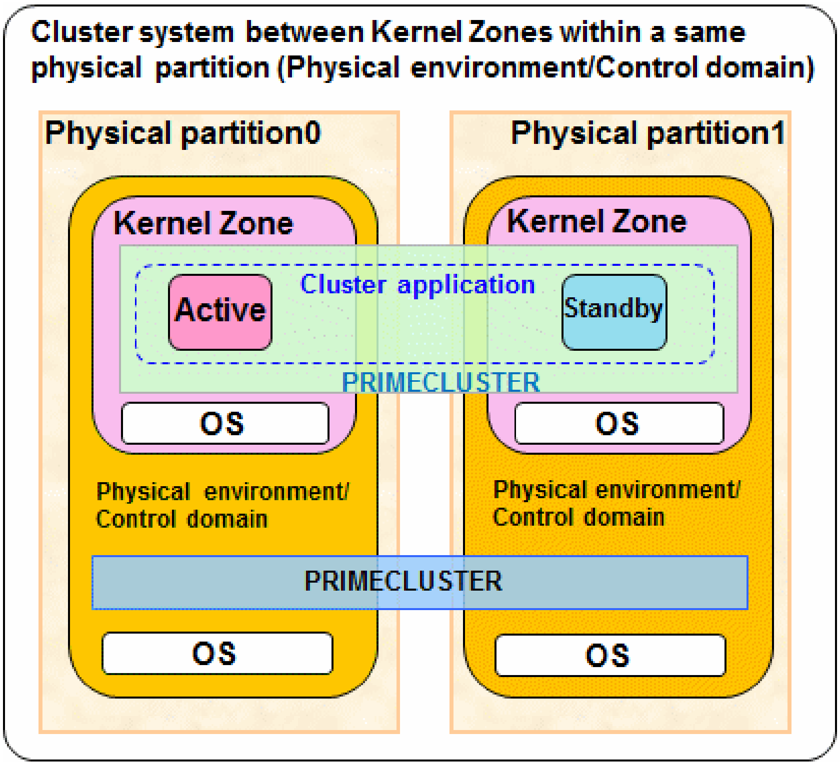
Cluster system between Kernel Zones among different physical partitions (control domain)

This enables the cluster system to operate between Kernel Zones on the control domain among different physical partitions. In a cluster system installed only on the Kernel Zones on the control domain, when the physical partition error occurs, the nodes that construct the cluster may come into the LEFTCLUSTER state. For dealing with this, install PRIMECLUSTER also to the control domain, and it is possible to switch over the cluster applications on the Kernel Zones automatically even when the physical partition error occurred.

- The cluster application error in the physical environment, on the control domain, or Kernel Zone
- The OS error in the physical environment, on the control domain, or Kernel Zone
- The hardware (network, shared disk and the route) failures
- The physical partition error

However, use this configuration with careful consideration of system design because this function limits other functions, such as disabling the RMS priority (ShutdownPriority) setting.

Figure 2.15 Cluster system between Kernel Zones among different physical partitions (control domain)



Note

- When creating the cluster application on the control domain, guest domain, or Kernel Zones, do not specify the RMS priority (ShutdownPriority) attribute.
- Set the survival priority of the Kernel Zones so as to be the same order relation as that of the physical environment/control domain.
- When building a cluster application on the control domain, if the control domain to be switched cannot be forcibly stopped due to a cluster application error on the control domain, the physical partition is forcibly stopped. Therefore, all the guest domains, I/O root domains, or Kernel Zones within the failed physical partition are stopped regardless of whether a cluster system exists.
- When a virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.
- Specify the same type for domains in the cluster. A cluster cannot be configured with different types of domains, for example, between the guest domain and I/O root domain, or between the control domain and I/O root domain.

Cluster system between Kernel Zones among different physical partitions (guest domain)

This enables the cluster system to operate between Kernel Zones built on the guest domain among different physical partitions. In a cluster system installed only on the Kernel Zones on the guest domain, when the physical partition error or the OS error on the guest domain occurs, the nodes that construct the cluster may come into the LEFTCLUSTER state. For dealing with this, install PRIMECLUSTER also to the control domain, and it is possible to switch over the cluster applications on the Kernel Zones automatically even when the physical partition error or the OS error on the guest domain occurred. The following types of error monitoring are performed in this configuration.

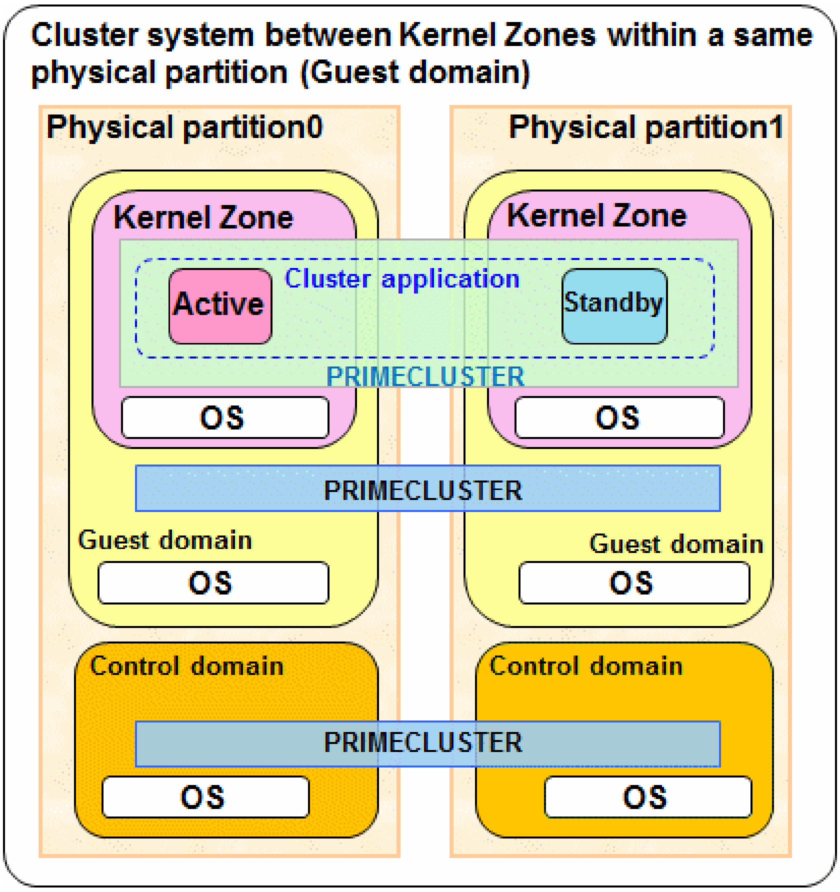
- The cluster application error on the control domain(*), guest domain, or Kernel Zone
- The OS error on the control domain(*), guest domain, or Kernel Zone
- The hardware (network, shared disk and the route) failures

- The physical partition error(*)

  *) Only when PRIMECLUSTER is built on the control domain

However, use this configuration with careful consideration of system design because this function limits other functions, such as disabling the RMS priority (ShutdownPriority) setting.

Figure 2.16 Cluster system between Kernel Zones among different physical partitions (guest domain)



Note

- When building a cluster application on the control domain or the guest domain, the control domain or the guest domain may be forcibly switched due to a switchover caused by a cluster application error on the control domain or the guest domain. Therefore, all the guest domains, I/O root domains, or Kernel Zones within the failed physical partition are stopped regardless of whether a cluster system exists. In addition, if the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

- A cluster cannot be configured with different types of domains, for example, between the guest domain and I/O root domain, or between the control domain and I/O root domain.

In addition, when PRIMECLUSTER is built on the domain, note the following points as well:

- When creating the cluster application on the control domain, guest domain, I/O root domain, or Kernel Zones, do not specify the RMS priority (ShutdownPriority) attribute.

- Set the survival priority of the Kernel Zones, guest domains, or I/O root domains so as to be the same order relation as that of the control domain.

- When a virtual I/O is set on the control domain, guest domain or Kernel Zone within the failed physical partition may be stopped regardless of whether a cluster system exists.

2.2.1.3 Cluster System Operating in Oracle Solaris Zones Environment

In an Oracle Solaris Zones environment, the applications on the non-global zone enter an inoperable status when an error occurs to the global zone or non-global zone.
Applying PRIMECLUSTER to the global zone and non-global zone provides status monitoring and a switchover function. Through these means, it becomes possible to switch over to a standby system in the event of an error occurring, and to achieve high reliability for the non-global zone.

You can build Oracle Solaris Zones environments on guest OS domains in Oracle VM Server for SPARC Environments (only for SPARC M10) as well as on physical server environments.

In addition, when the global zone is Solaris 10, the existing systems running on Solaris 8 or Solaris 9 can also be run on Solaris 10 by migrating them to the non-global zone with Oracle Solaris Legacy Containers (OSLC). (*1)

PRIMECLUSTER provides a status monitoring and switchover function for the non-global zone running on Solaris 8 or Solaris 9. Through these means, it becomes possible to switch over to a standby system in the event of an error occurring, and to achieve high reliability for the non-global zone running on Solaris 8 or Solaris 9.

(*1) To check whether the using middleware product is available in the non-global zone using Oracle Solaris Legacy Containers, see the respective middleware product manuals.

- Global zone status monitoring and switchover

  PRIMECLUSTER monitors the following statuses:

  - Global zone OS errors
  - Global zone hardware (network, shared disk, and the route) faults

  If PRIMECLUSTER detects an OS error, it stops all of the non-global zones operating on that global zone and switches them over to the standby system.

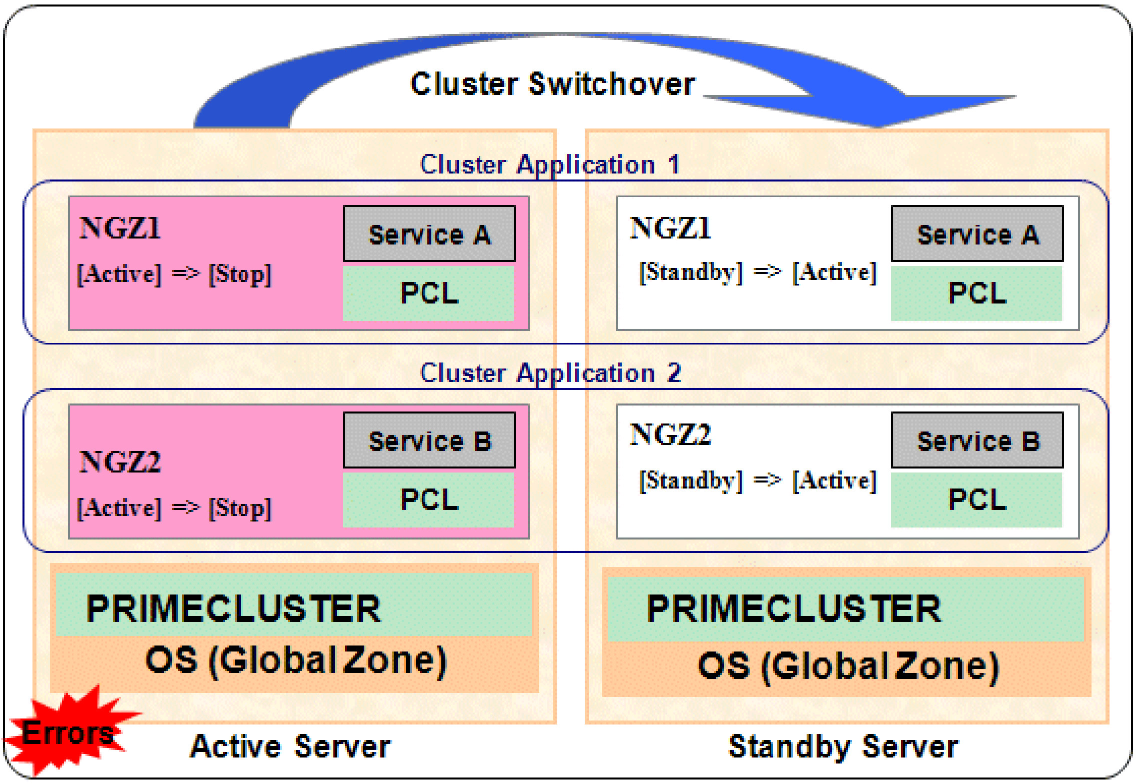
- Non-global zone status monitoring and switchover

  PRIMECLUSTER monitors the following statuses:

  - Non-global zone status
  - OS errors on the non-global zones
  - Status of applications operating on the non-global z

  If PRIMECLUSTER detects an error, it switches the affected non-global zones over to the standby system.

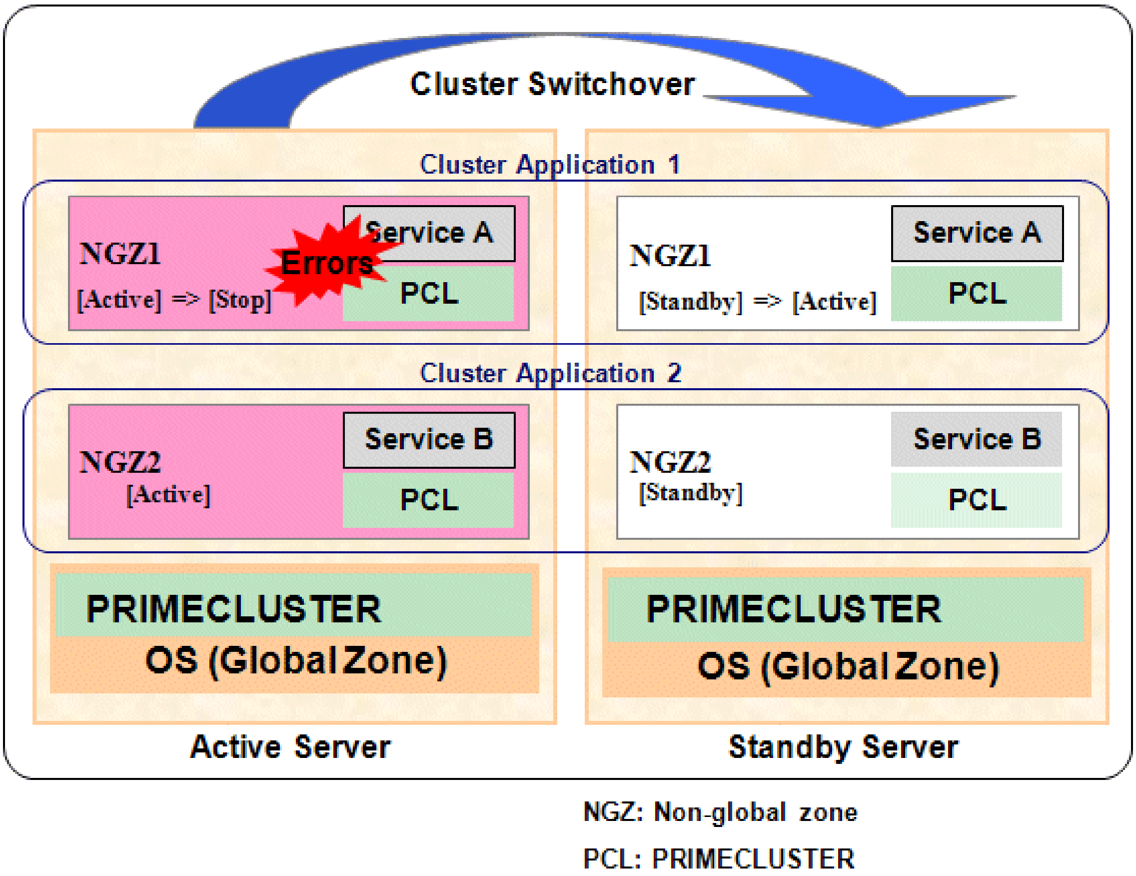
Figure 2.17 Switchover for When a Global Zone OS Error Occurs



NGZ: Non-global zone

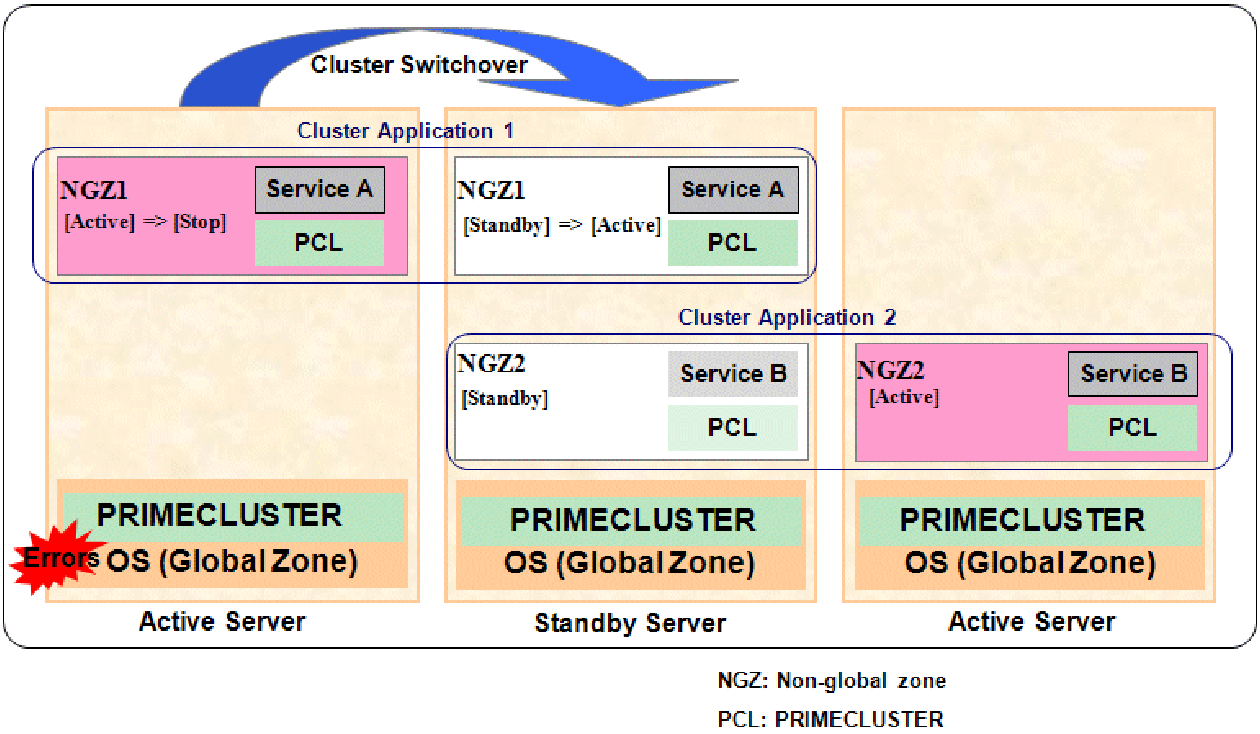
PCL: PRIMECLUSTER

Figure 2.18 Switchover for When some Application Error Occurs in a Non-Global Zone



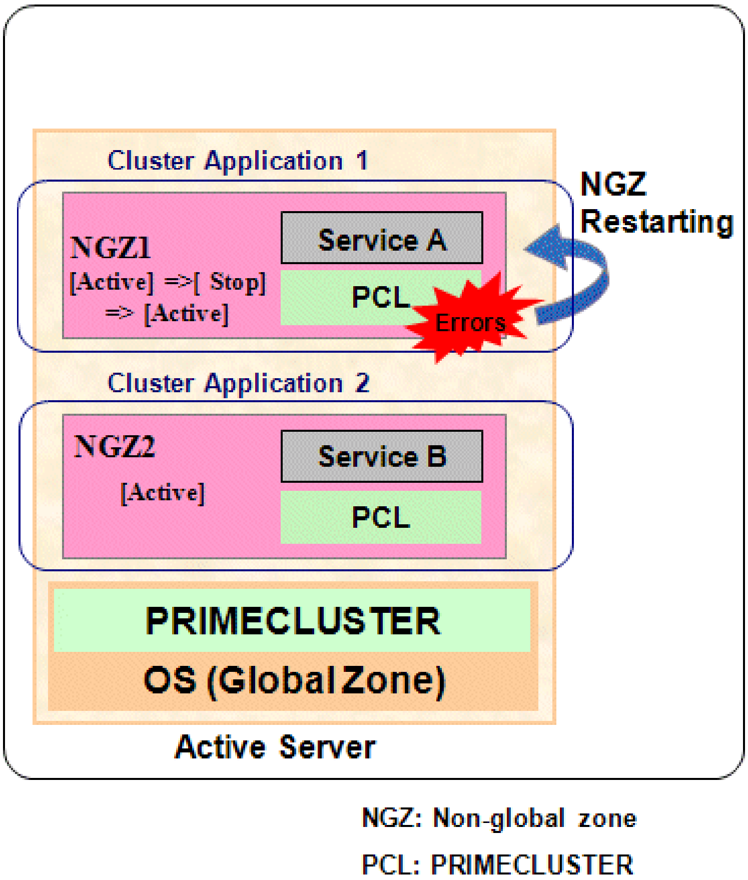
If using a cluster system comprised of three or more nodes, consolidation of the standby server becomes possible by preparing one standby server for the multiple operating servers. An example is shown below.

Figure 2.19 Switchover for When an OS Error for a Global Zone on a Three-Node Configuration Zones Environments Occurs



If using a single-node cluster comprised of one node, the status of OS and applications on the non-global zone are monitored. The availability is increased by restarting the non-global zone or an application on the non-global zone automatically to perform recovery when an error is detected. An example is shown in the following figure.

Figure 2.20 The Operations When an OS Error for a Non-Global Zone on a Single-Node Cluster Operation Zones Environments Occurs.



Note

- It is not possible to change the "cluster name" or "CF node name" in the non-global zone.
- The following functions or commands cannot be used in the non-global zone:
  - Automatic configure
  - Shared disk device connection confirmation
  - Operator intervention
  - Fault resource identification
  - Patrol diagnosis
  - clsyncfile (distributes a file between cluster nodes)
- The operations do not take over between the non-global zones operating on the same global zone.
- In the environment where a cluster application is built in the cluster system on the global zone, if an error occurs in the cluster application that is built on the global zone, the error may affect the non-global zone that is related to the global zone.

2.2.2 XSCF Configuration in SPARC M10

PRIMECLUSTER enables the following functions through connection to XSCF of SPARC M10.

- Asynchronous monitoring of a node

  Detects node errors immediately

- Forcible stop of a failed node

  Detects loss of heartbeat and stops the failed node forcibly

The asynchronous monitoring and forcible stop of a failed node make a communication route for XSCF redundant by using XSCF-LAN#0 and #1.

- XSCF-LAN#0: Administrative LAN
- XSCF-LAN#1: Asynchronous monitoring sub-LAN

The Asynchronous monitoring sub-LAN makes the path of the asynchronous monitoring and forcible stop of a failed node redundant. This LAN is connected by the other subnet differing from the administrative LAN.

By preparing the asynchronous monitoring sub-LAN, the failed node can be forcibly stopped even when an error occurs in the administrative LAN.

It is recommended that you use the asynchronous monitoring sub-LAN in a cluster configuration.

Note when the asynchronous monitoring sub-LAN is not used

When using XSCF-LAN#1 for any purposes other than the asynchronous monitoring sub-LAN due to the system requirements or other reasons, note the following points and build a configuration without the asynchronous monitoring sub-LAN.

- The communication for XSCF cannot be made redundant in this configuration. If an error occurs in the administrative LAN in this configuration, the connection to XSCF will not be available so that the asynchronous monitoring and forcible stop of a failed node cannot be done.
  If an error occurs in this situation (such as a node panic or loss of heartbeat), the service cannot be switched. When one of the following messages is displayed, check the cause. If XSCF has failed or the route to XSCF has failed, replace it immediately.

  ```
  SA SA_xscfsnmpg0p.so to test host <nodename> failed
  ```

  ```
  SA SA_xscfsnmpg1p.so to test host <nodename> failed
  ```

  ```
  SA SA_xscfsnmpg0r.so to test host <nodename> failed
  ```

  ```
  SA SA_xscfsnmpg1r.so to test host <nodename> failed
  ```

  ```
  SA SA_xscfsnmp0r.so to test host <nodename> failed
  ```

  ```
  SA SA_xscfsnmp1r.so to test host <nodename> failed
  ```

Sample configuration

Figure 2.21 Sample cluster configuration in a physical environment 1 (recommended configuration using the administrative LAN and asynchronous monitoring sub-LAN)

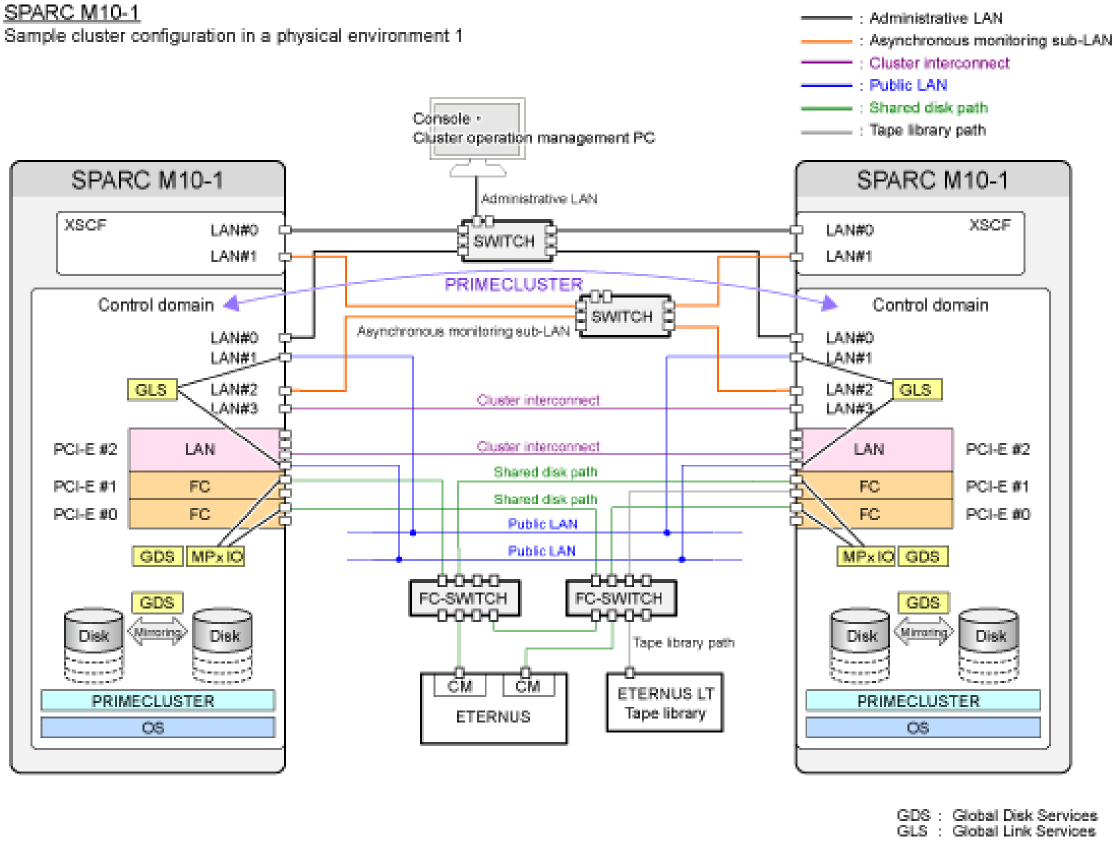


Figure 2.22 Sample cluster configuration in a physical environment 2 (configuration without using the asynchronous monitoring sub-LAN)

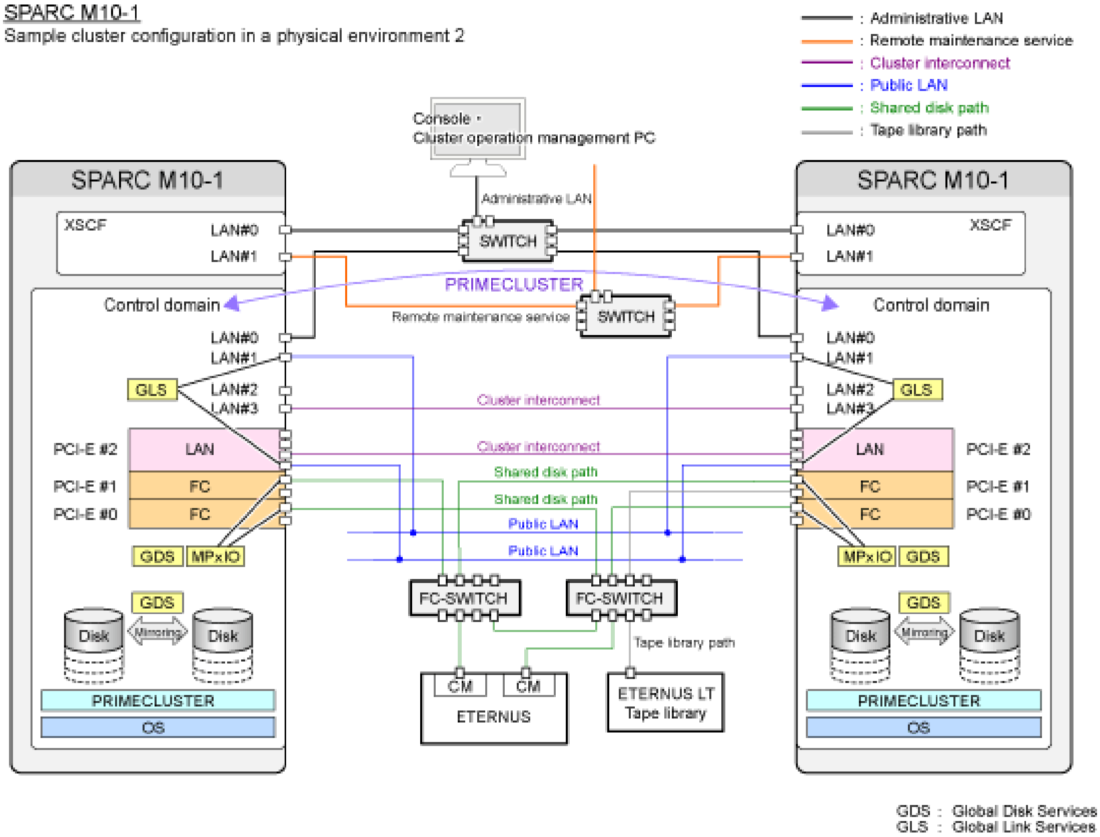


Figure 2.23 Sample cluster configuration in guest domains 1 (recommended configuration using the administrative LAN and asynchronous monitoring sub-LAN)

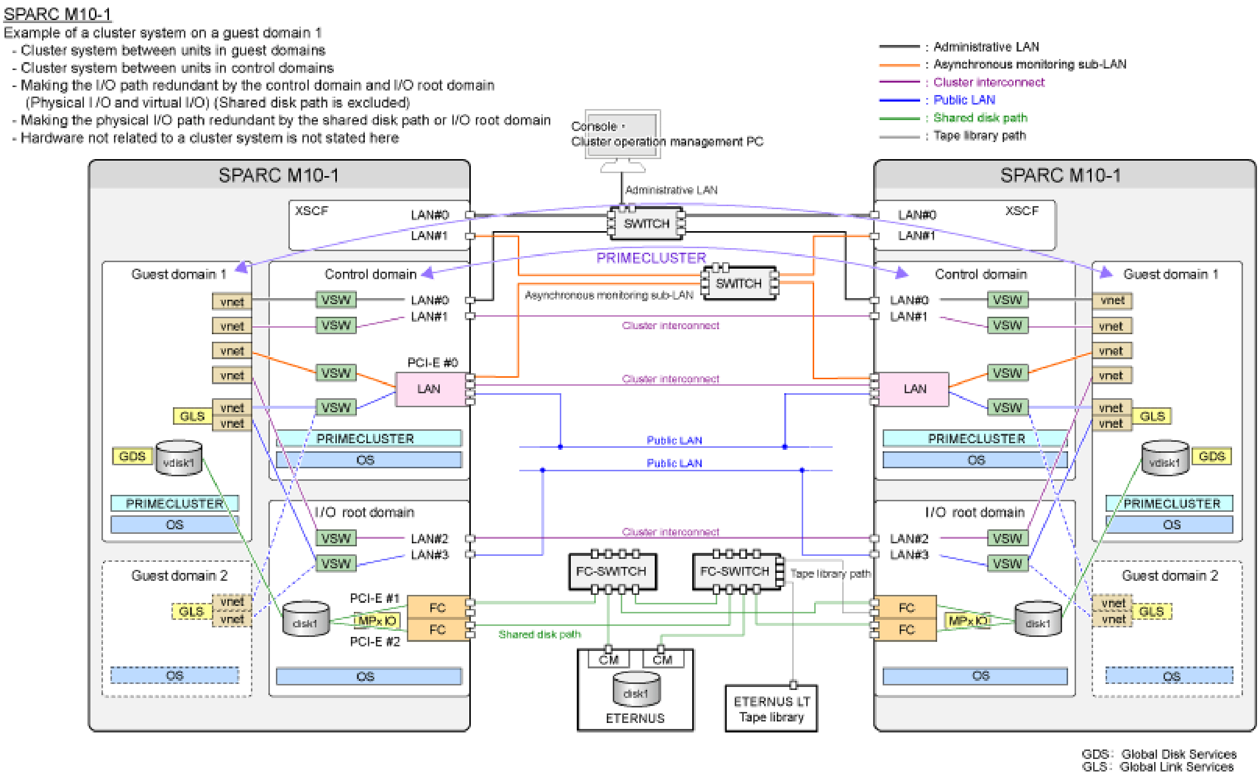
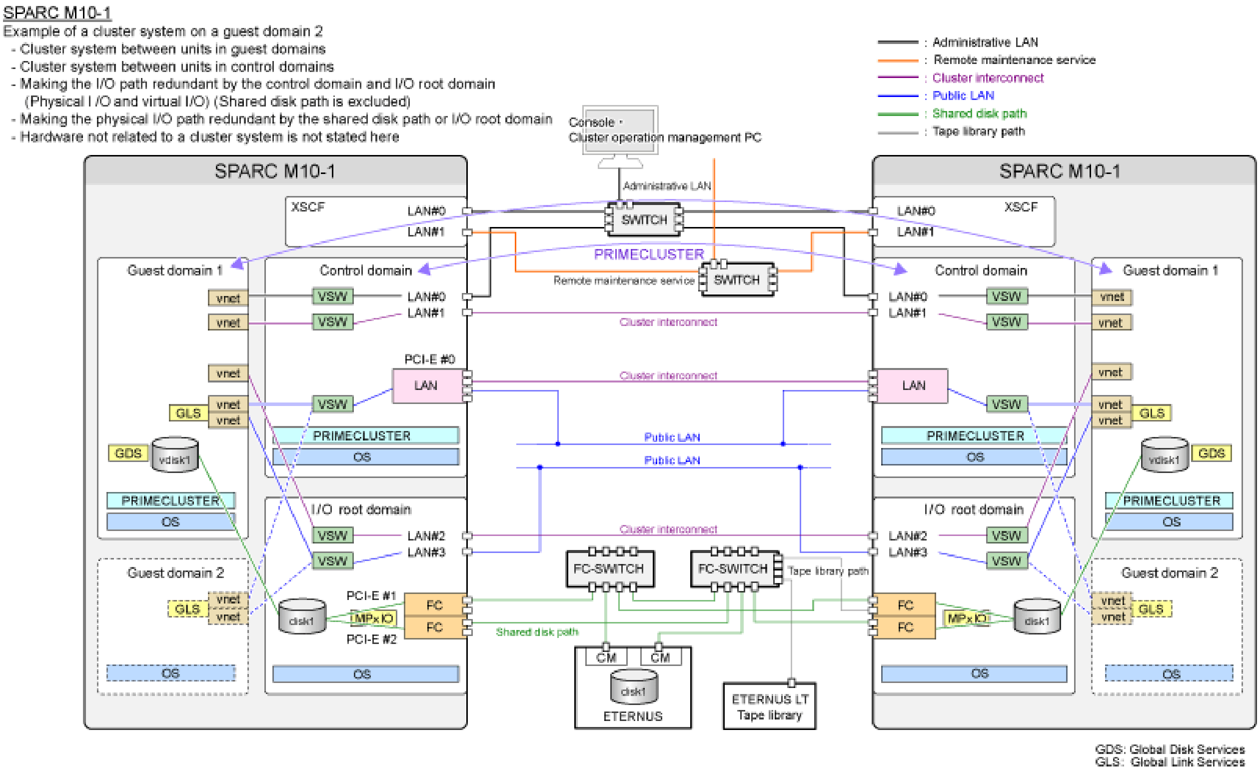


Figure 2.24 Sample cluster configuration in guest domains 2 (configuration without using the asynchronous monitoring sub-LAN)



2.3 Determining the Cluster System Operation Mode

To ensure that the cluster system to be built operates efficiently, you need to determine the number of nodes and an appropriate operation mode.

PRIMECLUSTER allows you to configure multiple cluster applications. The operation mode is determined depending on how you set up the cluster applications in the cluster system.

The main operation modes are listed below:

Classification	Operation mode	Number of cluster applications	Number of nodes
Standby operation	1:1 standby	1	2
	Mutual standby	2 to	2 to (number of supported nodes)
	N:1 standby	2 to	3 to (number of supported nodes)
	Cascade	1 to	3 to (number of supported nodes)
	Priority transfer	2 to	3 to (number of supported nodes)
	N:M standby	2 to	4 to (number of supported nodes)
Scalable operation	Scalable	2 to (*)	2 to (number of supported nodes)
	High-availability scalable operation	2 to (*)	2 to (number of supported nodes)
Single-node cluster operation		1 to	1

(*) For the scalable operation, a scalable cluster application is additionally required.

Note

Note that failover does not occur when a node is switched off. When designing the system, consider using an uninterruptible power supply (USP).

See

For supported configurations where PRIMECLUSTER is used in Oracle Solaris Zones environments, see "Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment."

2.3.1 Standby Operation

Standby operation has the following operation modes.

Information

The topologies for standby operation include hot-standby and cold-standby operation.

Hot-standby operation enables preliminary operation so that the operating state can be established immediately on the standby node. In hot-standby operation, the state of the cluster application running on the operating node will be Online, while that of the cluster application on the standby node will be Standby. To perform hot-standby operation, hot-standby must be supported by the PRIMECLUSTER product to be used, the ISV application, and the user applications.

Cold-standby operation does not allow the preliminary operation needed to establish the operating state immediately on the standby node. In cold-standby operation, the state of the cluster application on the operating node will be Online, while that of the standby node will be Offline.

1:1 standby

Definition

- It is an operation mode in which a cluster system consists of 2 nodes. One is operating, and the other is standby. When a failure occurs in the operating node, a cluster application switches to the standby node. This does not disrupt ongoing operation.

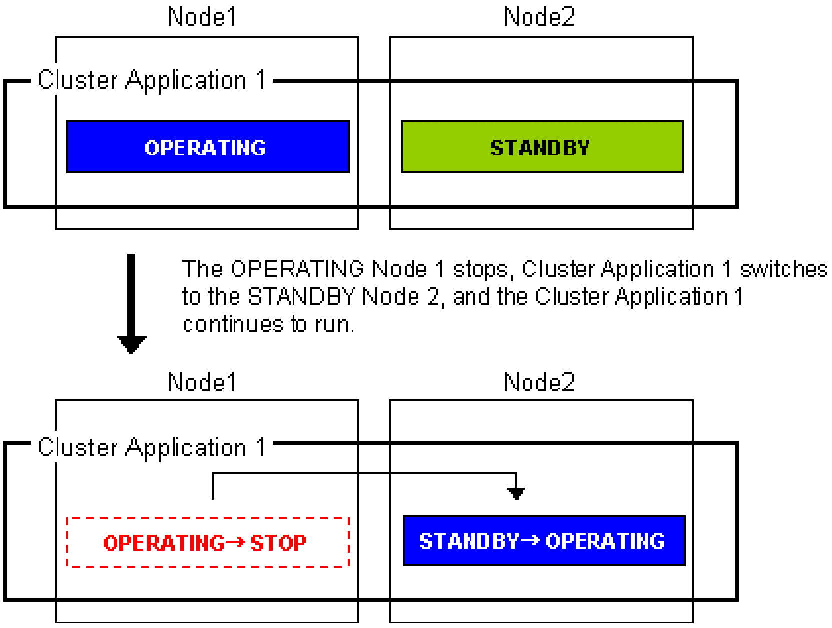
Advantage

- This operation mode ensures the availability of the cluster application even after failover.

Disadvantage

- Nodes cannot be used efficiently because of a redundant configuration.

Failover image



Mutual standby

Definition

- It is an operation mode in which a cluster system consists of 2 or more nodes. Normally, 2 nodes are used in this operation mode. Each node has one operating and one standby application. The operating cluster application has its own standby in each other's node.

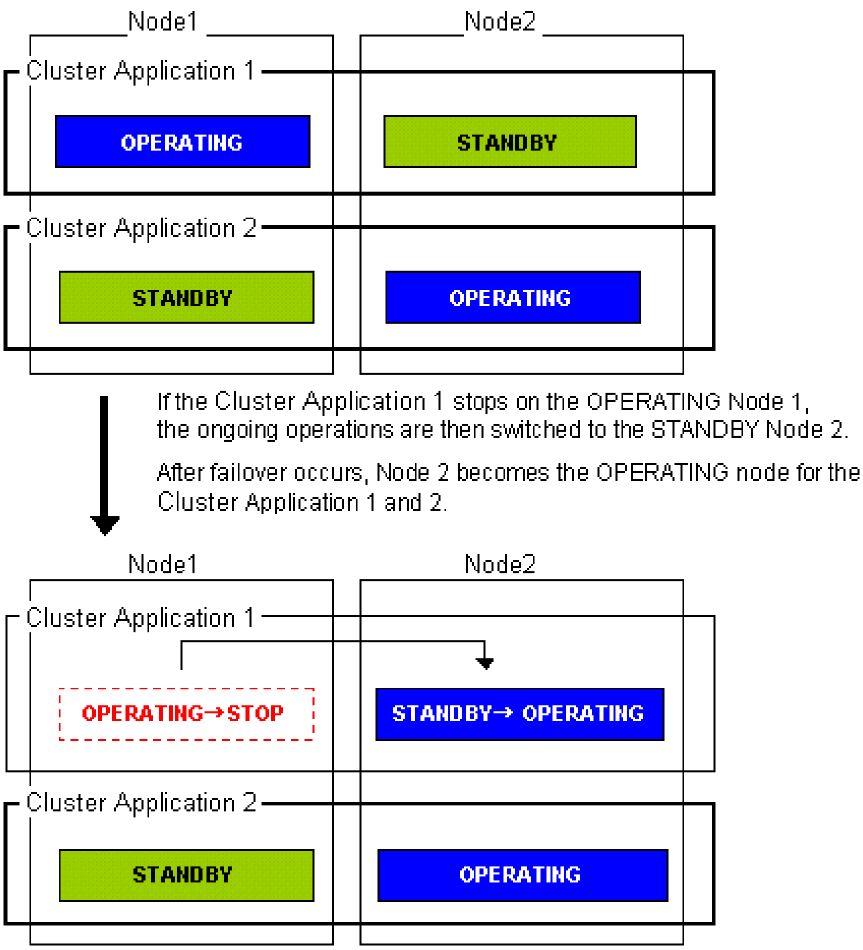
Advantages

- Since all nodes are used for cluster application operation, the nodes can be used efficiently.

Disadvantages

- If failover occurs for any of the cluster applications, the performance of the cluster applications may drop because two or more cluster applications will be operating in the failover node. For this operation mode, you need to estimate adequate resources.

Failover image



N:1 standby

Definition

- It is an operation mode in which a cluster system consists of 3 or more nodes. One is standby, and the others are operating. When a failure occurs in one of the operating nodes, a cluster application switches to the standby node. If a failure occurs in two or more operating nodes at the same time, the cluster applications switch to the standby node.

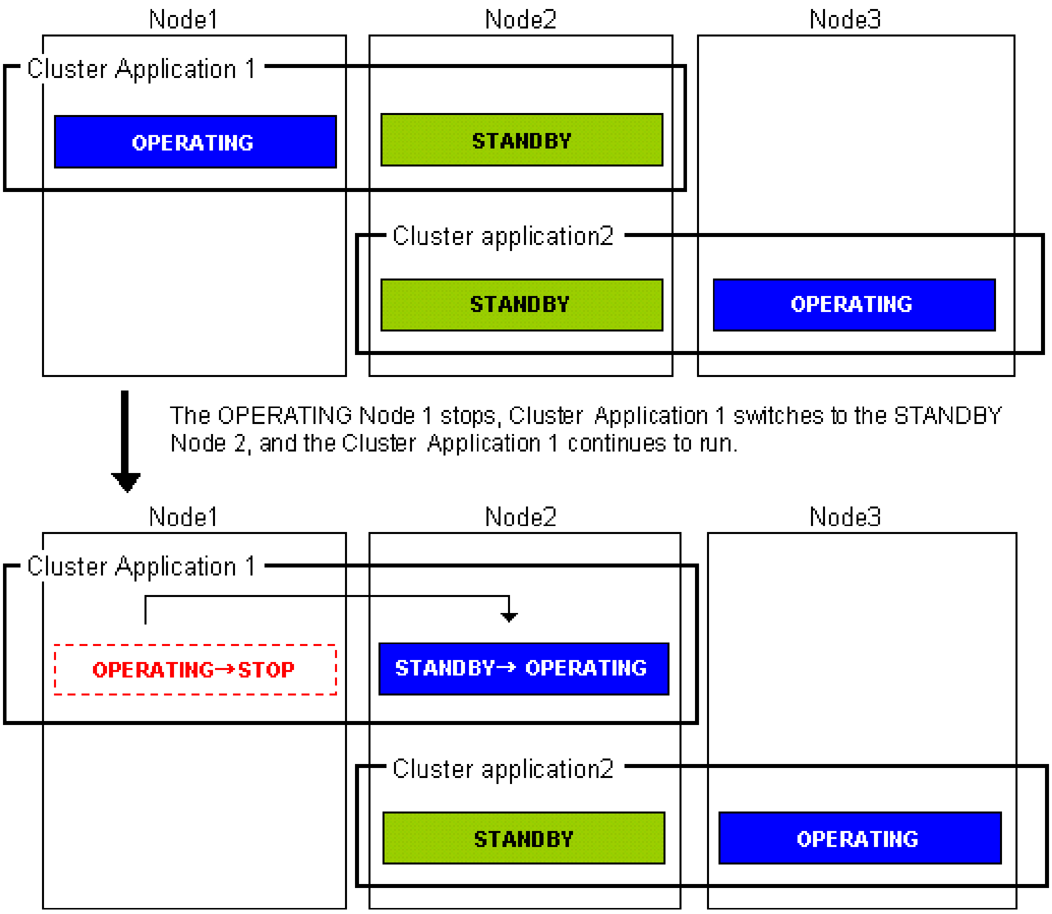
Advantages

- This operation mode ensures the availability of the cluster application even after failover.
- Since one node serves as the STANDBY node for multiple applications, the STANDBY cost can be reduced when the number of applications is large.

Disadvantages

- If failover occurs for multiple applications, the performance of the applications is reduced because multiple applications will be operating in one node.

Failover image



Cascade (using one cluster application)

Definition

- It is an operation mode in which a cluster system consists of 3 or more nodes: one is operating, and the others are standby. When a failure occurs in the operating node, a cluster application switches to one of the standby nodes. When a failover is even failed, this application switches to other standby node.

Advantages

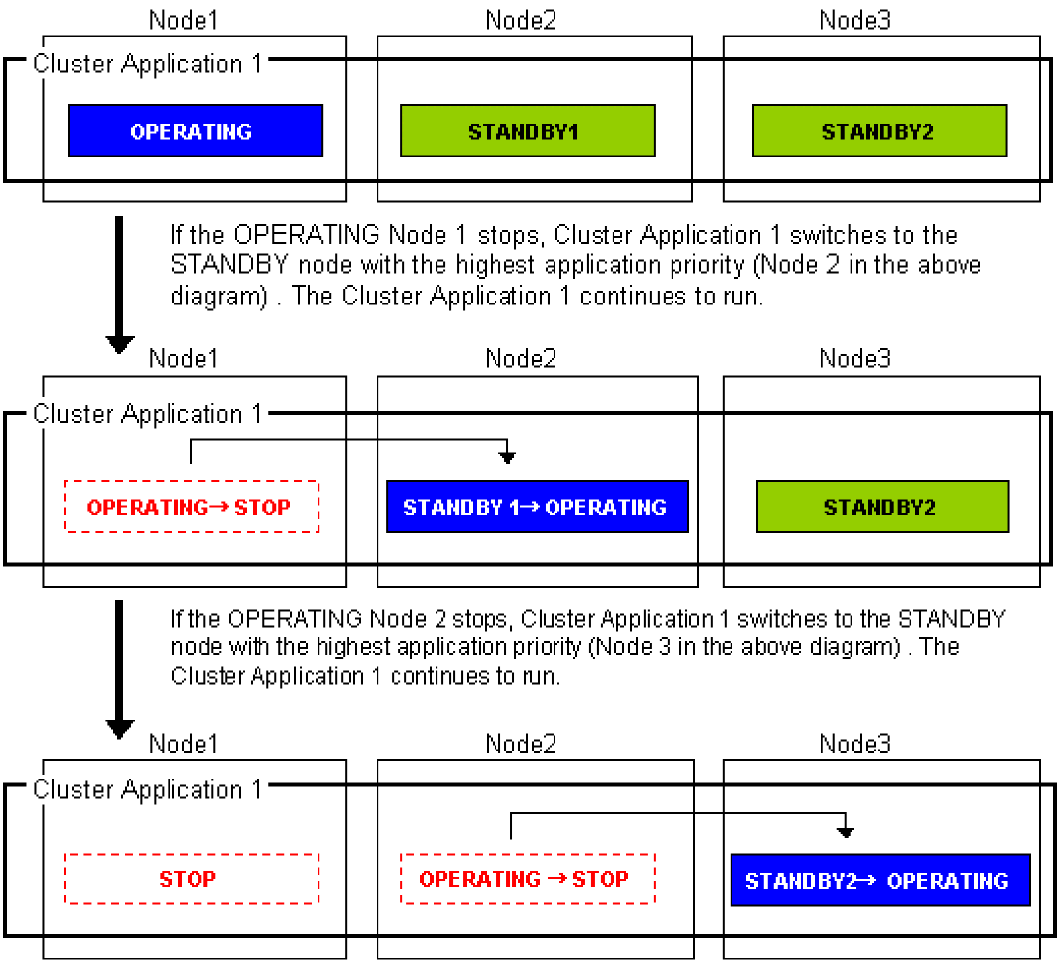
- Even after one node is stopped, the redundant configuration of the cluster application can be maintained by using other nodes. The availability is guaranteed during system maintenance.
- This operation mode ensures the availability of cluster applications even after failover.

Disadvantages

- As the system has a redundant configuration, nodes cannot normally be used efficiently.

Failover image

In this example, the nodes are defined in the sequence Node 1, Node 2, and Node 3 starting from the node with the highest cluster application priority. These nodes are defined when the cluster application is set up.



Priority transfer (application of N:1 standby)

Definition

- A single node functions as STANDBY for multiple cluster applications. For the other nodes, a single cluster application functions as OPERATING for every node of the other nodes while the other multiple cluster applications function as STOP.
- This topology uses the exclusivity function between cascade and cluster applications.

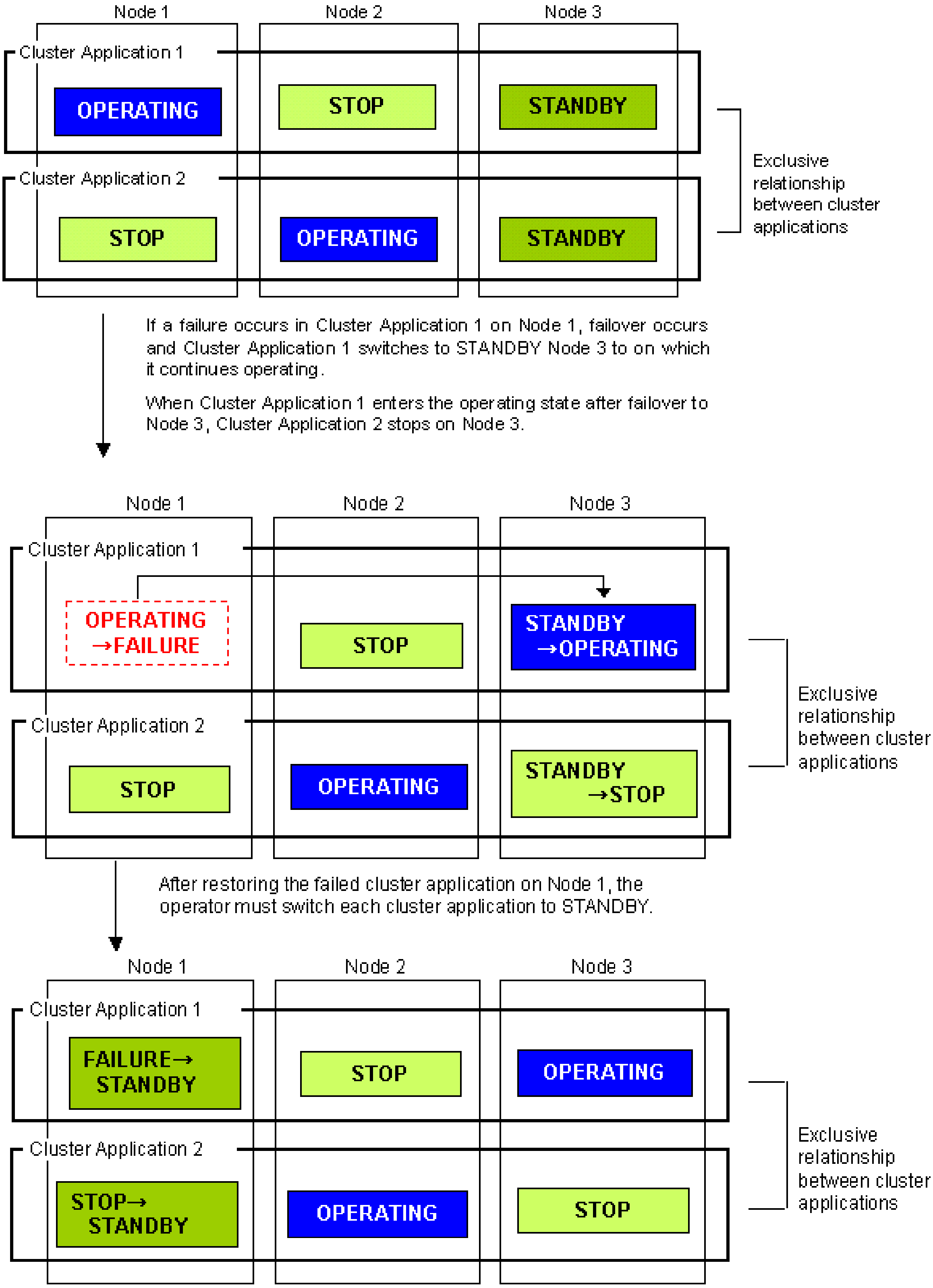
Advantages

- On that node on which a single cluster application is OPERATING, the other cluster applications do not become either OPERATING or STANDBY. Therefore, the throughput of that cluster application is guaranteed even after failover occurs.
- Because failback of the cluster application is not necessary during the restoration of a cluster application, a job can also be continued during the restoration.
- Since a single node is used as STANDBY exclusively for multiple cluster applications, the cost incurred for standby can be saved when there are many cluster applications.

Disadvantage

- Since a single node is used as STANDBY of multiple cluster applications, availability decreases when there are many cluster applications.
- If a failover occurs due to the occurrence of an error on a single node, the availability decreases because no standby node is available until the completion of the maintenance work.

Failover image



N:M standby (application of N:1 standby)

Definition

- Multiple nodes exclusively for standby server act as STANDBY for multiple cluster applications.

- For the nodes other than those that are used exclusively for standby, each of the nodes is used for a single OPERATING cluster application and multiple STOP cluster applications.
- This topology uses the exclusivity function between cascade and cluster applications.

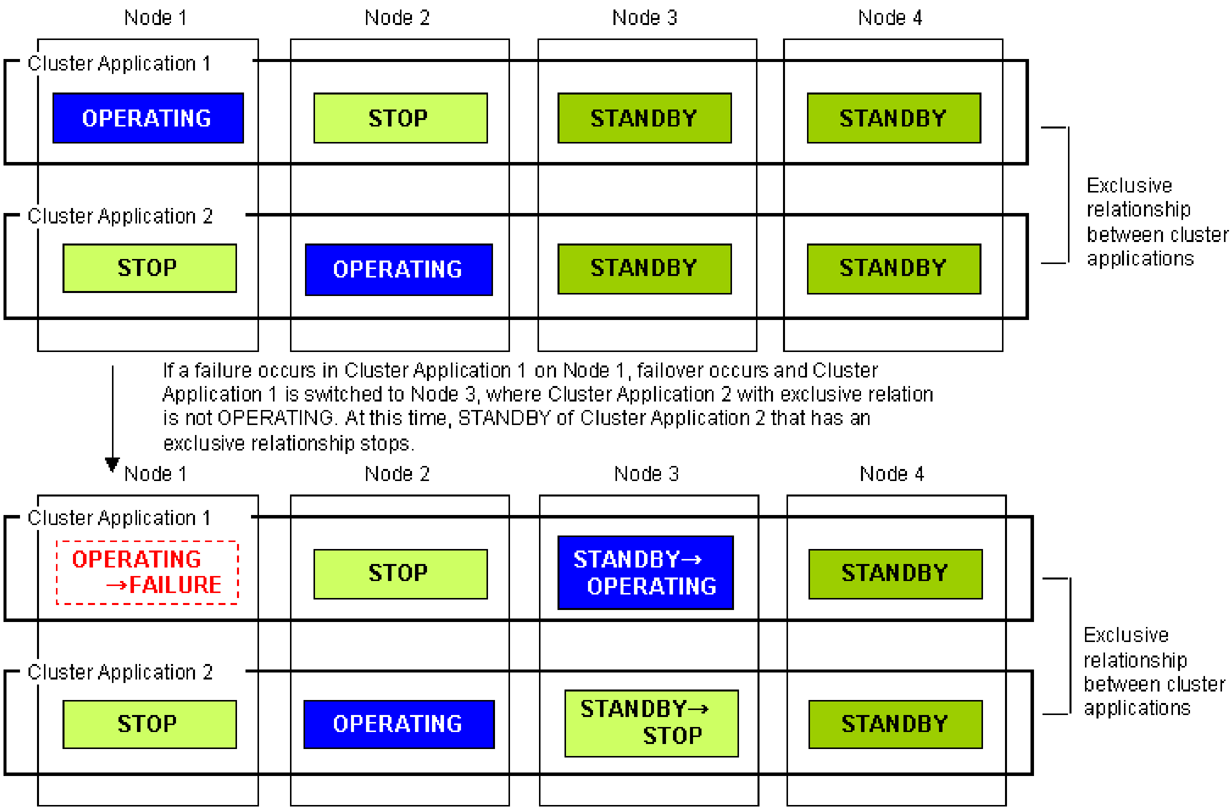
Advantages

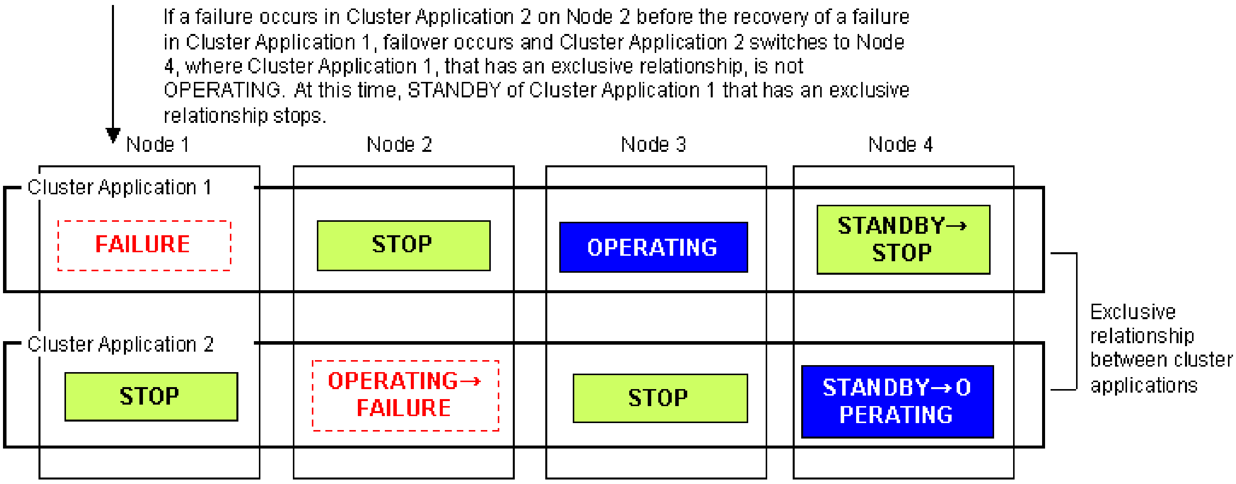
- On that node on which a single cluster application is in the OPERATING state, other cluster applications do not either become OPERATING or STANDBY. Therefore, the throughput of that cluster application is guaranteed even after failover occurs.
- Because failback of the cluster application is unnecessary during the restoration of a cluster application, a job can be also continued during the restoration.
- Since multiple nodes serve as STANDBY for multiple cluster applications, availability can be maintained even when there are many cluster applications is many.

Disadvantage

- Because a redundant configuration is used, the node usage efficiency is poor during normal operation.

Failover image





2.3.2 Scalable Operation

This section explains the topologies used for scalable operation:

Scalable

Definition

- A cluster system consists of two or more operating nodes, and all the nodes are used for online cluster applications. This operation mode is suitable for parallel jobs that use the I/O load balancing and load sharing on a parallel database.

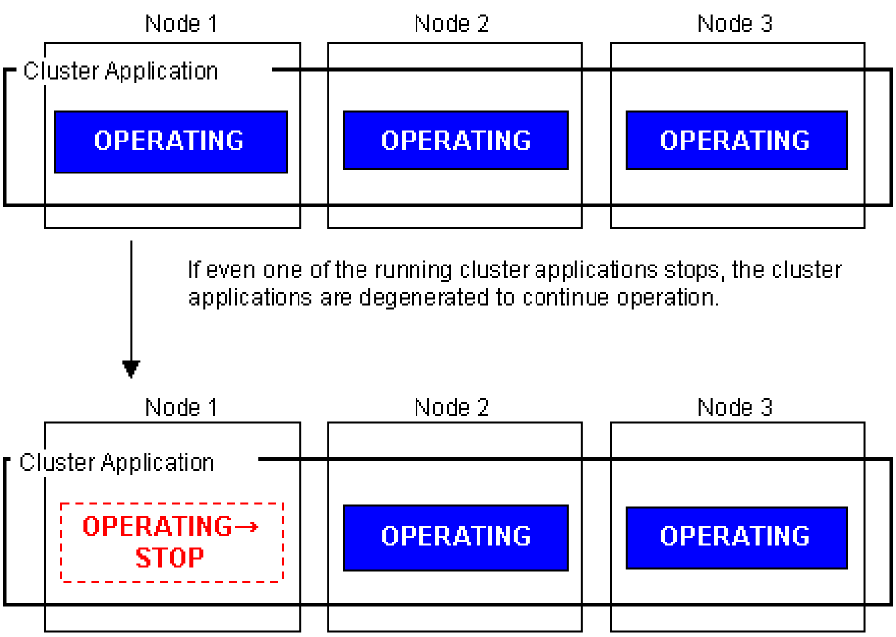
Advantages

- Throughput can be improved for a cluster application by expanding nodes.
- Degenerated operation is possible even if part of the cluster applications stops.

Disadvantage

- If part of the cluster applications stops, throughput of the cluster applications cannot be guaranteed because degenerated operation is assumed.

Failover image



Note

Scalable operation can be used in combination with some PRIMECLUSTER-related products. For information on the related products, see the "Table 19.1 PRIMECLUSTER product list."

High-availability scalable operation

Definition

- Refers to the topology in which standby operation is configured for each cluster application that constitutes scalable operation. Suitable for a parallel database for which scalability and availability are required, as well as parallel job execution for which load share/load balance is used.
- Standby operation that constitutes scalable operation can be combined with 1:1 standby and N:1 standby, with priority transfer.

Advantages

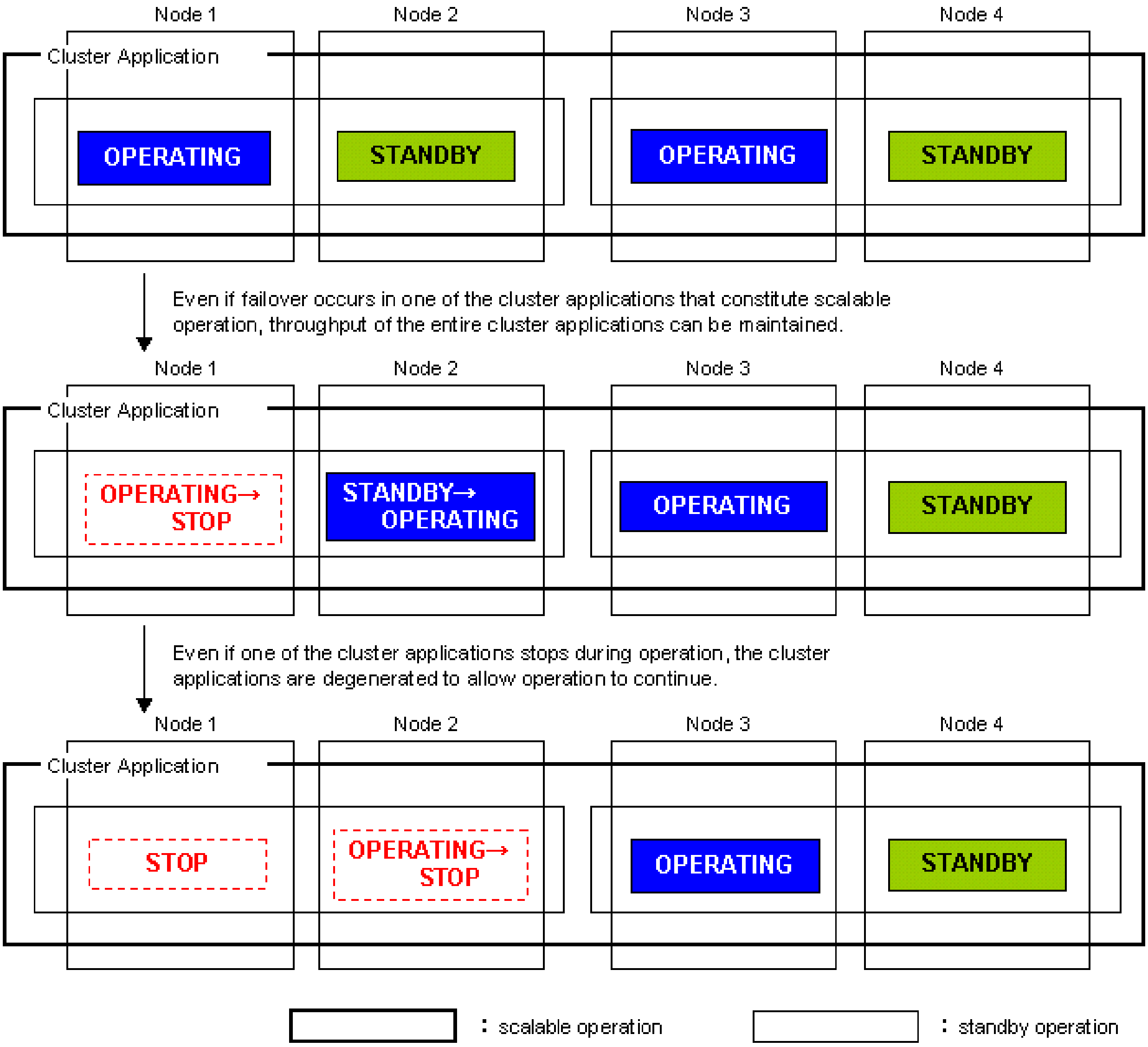
- Even if failover occurs in one of the cluster applications that constitute scalable operation, the throughput of all the cluster applications can be maintained by using a redundant configuration.
- Throughput of the cluster applications can be improved with node expansion.
- Degenerated operation is possible even if part of the cluster applications stops.

Disadvantage

- Since a redundant configuration is used, the node usage efficiency is poor during normal operation.

Failover image

The following illustrates failover when two 1:1 standby operations are combined to enable scalable operation.



Note

High-availability scalable operation can be used in combination with some PRIMECLUSTER-related products. For information on the related products, see the "Table 19.1 PRIMECLUSTER product list."

2.3.3 Single-Node Cluster Operation

This section explains the topologies used for single-node cluster operation:

Definition

- It is an operation mode in which a cluster system consists of one node.

Advantages

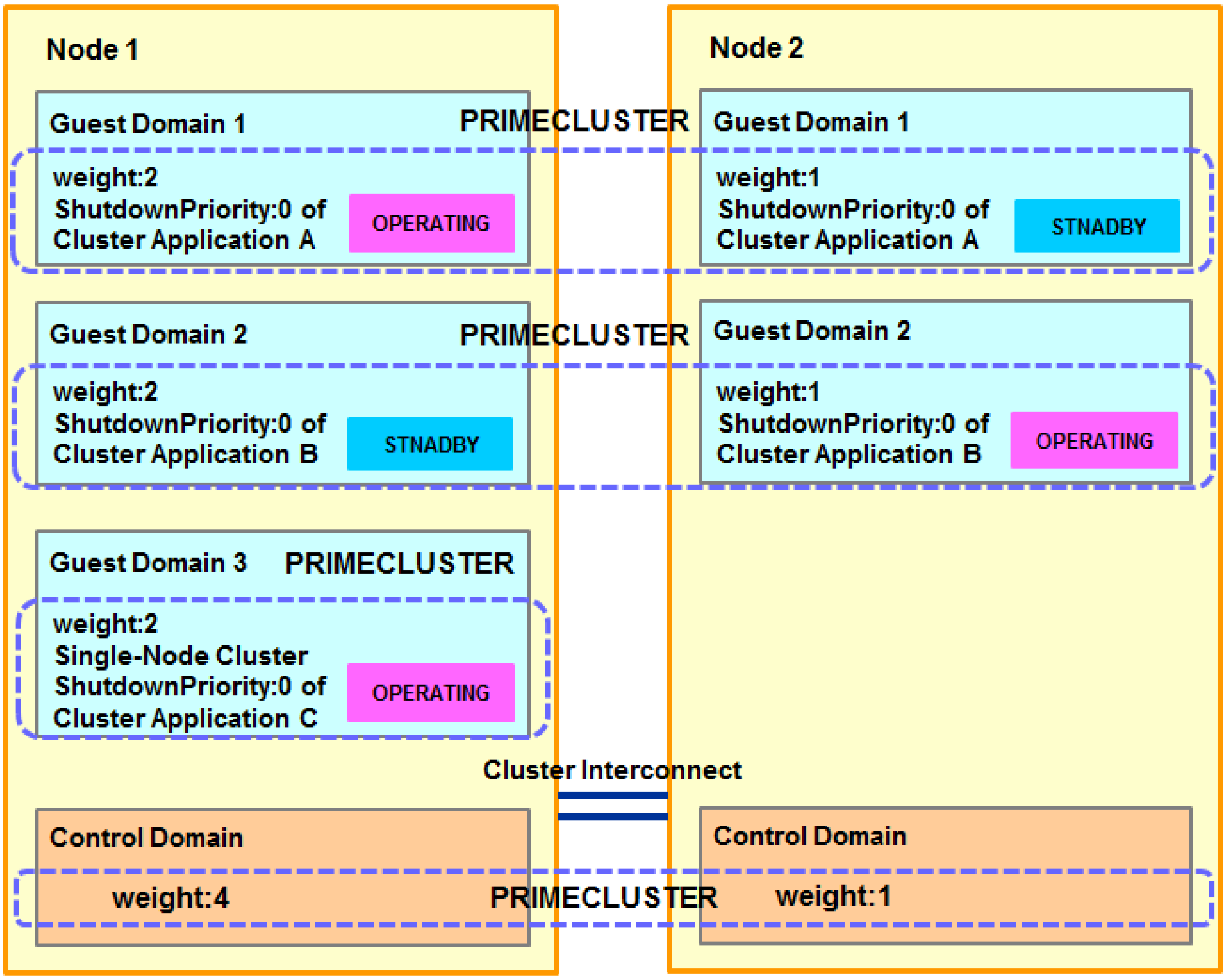
- This operation mode enables monitoring and control jobs on the node in a single node configuration.
- If an error occurs in the resource to which the AUTORECOVER attribute is set, the availability can be improved by automatically restarting the system for restoration.
- You can also use this mode as a development environment for creating and testing cluster applications.

Notes

- Jobs will be suspended in the case of a hardware failure because a single-node cluster has no hardware to switch to. Build a cluster with multiple nodes if you need to switch hardware when a hardware failure occurs.

- If multiple cluster systems exist in an environment in which the virtual machine function is used, build a single-node cluster on the highest priority node as the figure shown below.

Figure 2.25 Example: Single-node cluster configuration



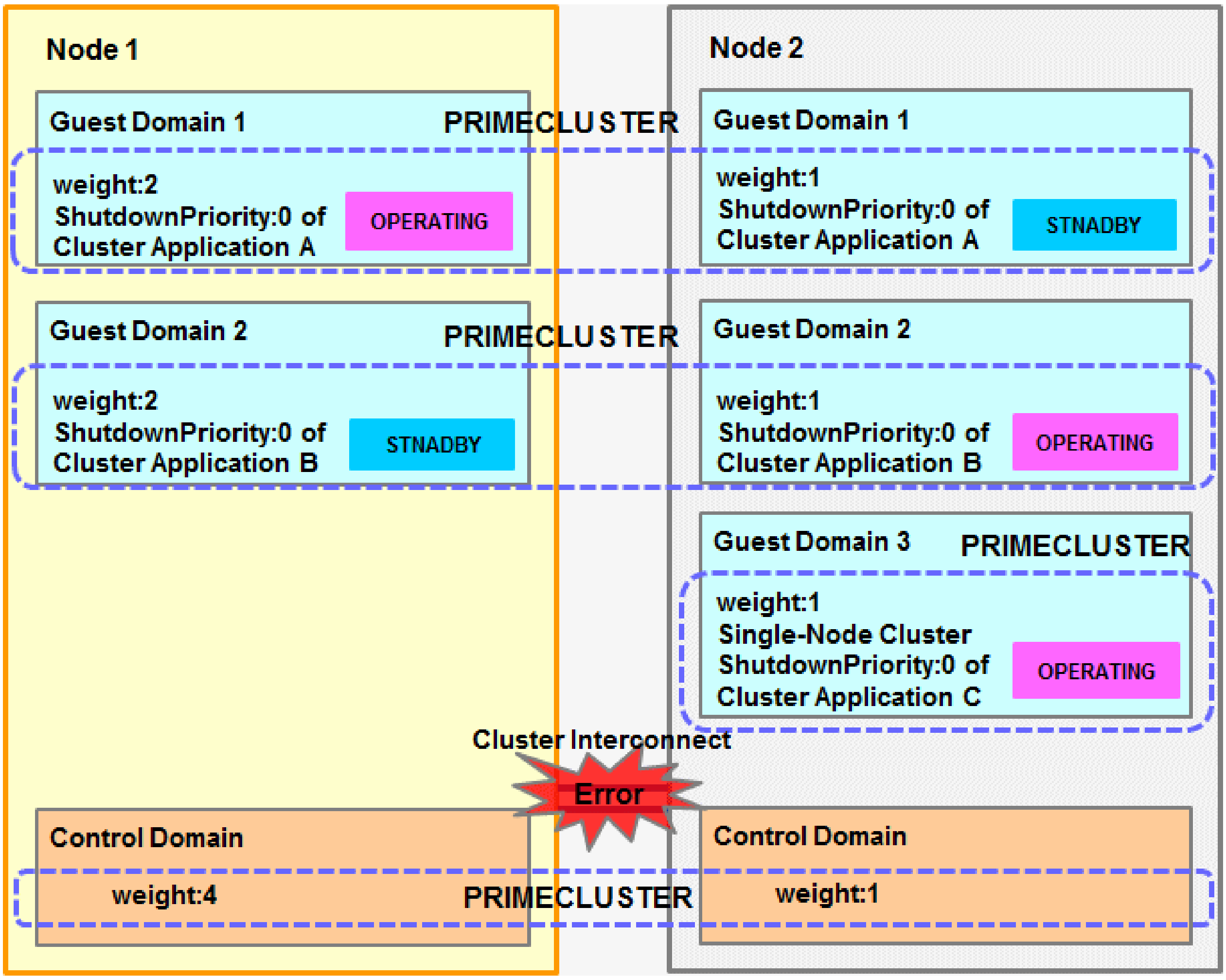
- In an environment in which the virtual machine environment is used, a guest domain on the single-node cluster is shut down under the following conditions (see the figure below):

  - Multiple cluster systems exist;

  - Priority is low for the node that includes the single-node cluster; and

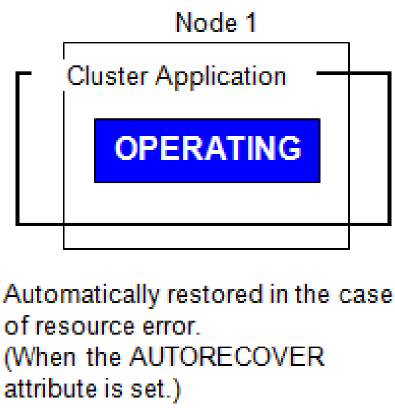
- The node is forcibly shutdown (due to an inter-node communication failure or other causes).

Figure 2.26 Example: Operation behavior in case of cluster interconnect error



Failover image

No failover occurs in the single-node cluster operation.



Note

You need at least one network interface card for cluster interconnect that is used in PRIMECLUSTER also in the single-node cluster operation.

2.4 Determining the Web-Based Admin View Operation Mode

Determine the operation mode of Web-Based Admin View according to your PRIMECLUSTER configuration through consideration of system performance.

This section describes operation modes and typical models of PRIMECLUSTER systems that run Web-Based Admin View, and provides a guideline for adopting models.

See

For information on the operation modes of Web-Based Admin View, see "1.2 Web-Based Admin View topology" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

Note

When a cluster contains 3 or more nodes, the 3-tier model, in which the cluster management server is not in the cluster, is recommended.

Roles of individual nodes

Web-Based Admin View adopts a logical 3-tier architecture, which consists of clients, a cluster management server, and cluster nodes. It has the following functions.

Clients

A client is a computer with which a user manages operations. Basically, the computer is a personal computer that uses a Web browser.

Management server

The cluster management server manages cluster operation and features web server functions. The server can be as a cluster node. The cluster management server can be duplexed. In this case the system will have a two-server configuration, consisting of a primary management server and a secondary management server.

Set up both primary and secondary management servers for redundancy.

You can dynamically move the secondary management server depending on the operation mode. The cluster management servers run Solaris.

Cluster nodes

Cluster nodes construct the PRIMECLUSTER system. The cluster nodes run Solaris.

Logical 3-tier architecture and operation models

Web-Based Admin View adopts a logical 3-tier architecture consisting of clients, management servers, and monitored nodes. Physically, the system can adopt a 2-tier architecture.

Typical operation modes that run Web-Based Admin View are introduced below.

2-tier model

In the 2-tier model, the cluster management server and the cluster node are used together on the same machine, and the client is on a machine other than the nodes and the management severs.

This model is used for a configuration where the number of nodes is relatively small.

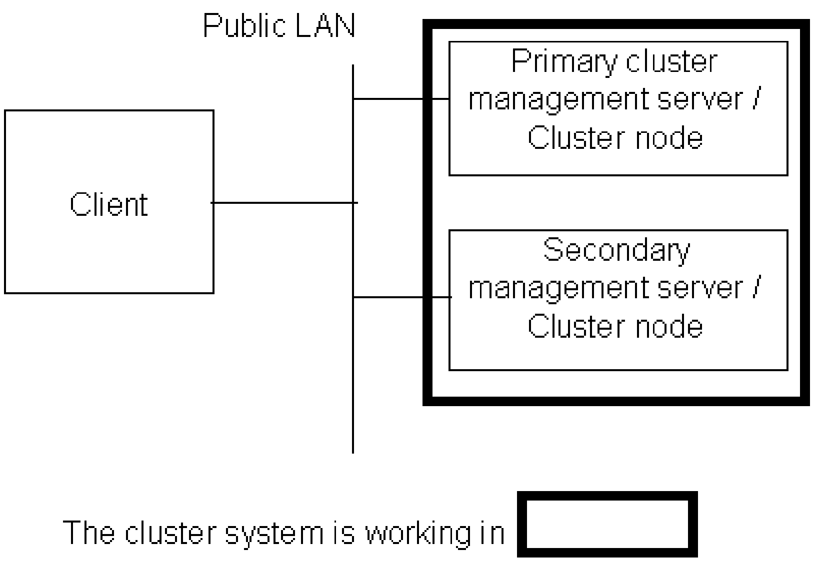
Normally, this model is used for an operation mode such as 1:1 standby or mutual standby in a 2-node configuration.

In the local operation, this model supports integrated management of the OS consoles of multiple nodes without linking to the cluster management server.

This model supports 2 types of topology, which are described below.

Topology where a network is shared

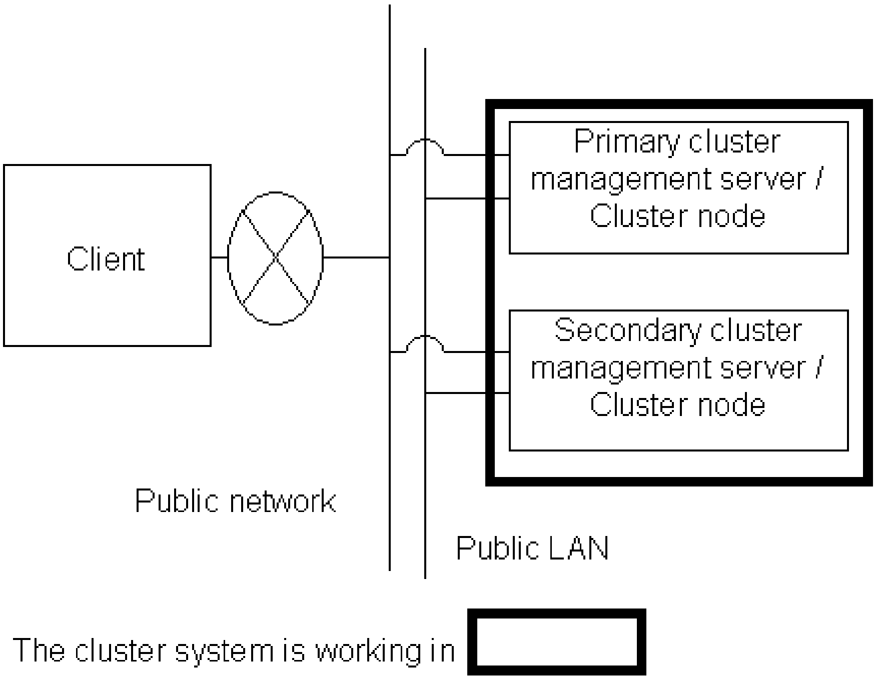
In this topology, the public LAN and the LAN that is connected to the management client are used together. You can adopt this topology if the network users and network range are limited for security. This is the default Web-Based Admin View configuration after PRIMECLUSTER installation.



Topology where separate LANs are used

In this topology, the public LAN and the LAN that is connected to the management client are separate. When using a management client from a public network, this topology is recommended for security. After the PRIMECLUSTER installation is done, you will need to modify the Web-Based Admin View configuration.

Specify the IP addresses used for a cluster node and a client respectively. For details, see "5.1.1 Setting Up CF and CIP."



3-tier model

In the 3-tier model, clients, cluster management servers, and cluster nodes are set up separately.

This model is adopted for configurations where the number of nodes is relatively large.

Normally, this model is used for integrated management of the PRIMECLUSTER system. You can also use this mode when you do not want to impose the load of running the management server on the cluster node or when you want to perform the integrated management of the PRIMECLUSTER system.

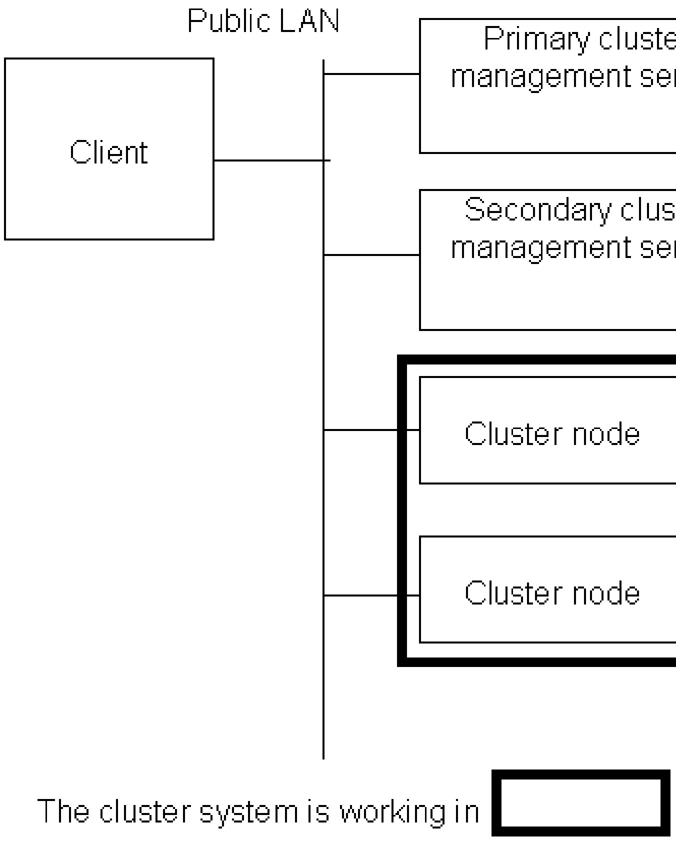
To maintain operation continuity and availability, Fujitsu recommends this 3-tire model for a configuration with 3 or more nodes.

In the server/client operation, this model supports integrated management of the OS consoles of multiple nodes, and can be operated on the cluster management server as well as on the client.

This model supports 2 types of topology, which are described below.

Topology where a network is shared

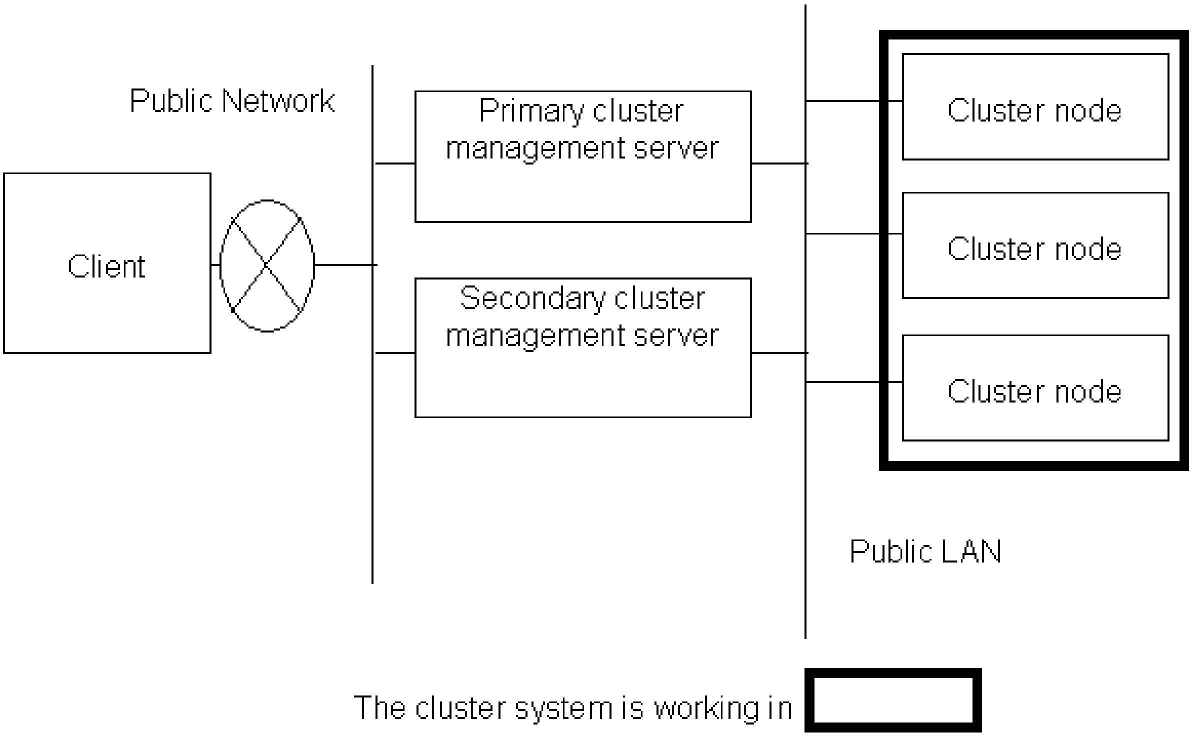
In this topology, the public LAN and the LAN that is connected to the management client are the same. You can adopt this topology if the network users and network range are limited for security. This is the default Web-Based Admin View configuration after PRIMECLUSTER installation.



Topology where separate LANs are used

In this topology, the public LAN and the LAN that is connected to the management client are separate. When using a management client from a public network, this topology is recommended for security. After PRIMECLUSTER installation is done, you will need to modify the Web-Based Admin View configuration.

Specify the IP addresses used for a cluster node and a client respectively. For details, see "5.1.1 Setting Up CF and CIP."



2.5 Setting the Failover Timing of a Cluster Application

Determine the trigger for cluster application failover. You can choose from the following:

Select the trigger for cluster application failover from the following:

Multiple selections are available for No. 2 to No. 4:

1. The cluster application does not automatically switch to other host.
2. The cluster application switches to other host in the event of a node failure.
3. The cluster application switches to other host in the event of a resource failure.
4. The cluster application switches to other host in the event of RMS shutdown.

See

The failover timing is set in "6.7.2 Creating Cluster Applications."

Part 2 Installation

This part describes procedures for installing the PRIMECLUSTER system and running Web-Based Admin View.

The operations include the procedures up to installing a new PRIMECLUSTER system.

For procedures on changing the PRIMECLUSTER system configuration after the system is installed, see "Part 4 System Configuration Modification."

Chapter 3 Software Installation

This chapter describes how to install required software for Solaris software and PRIMECLUSTER on native machines.

The following procedures are explained here:

- PRIMECLUSTER installation (overview)
- Installation and setup of related software
- Application installation and environment setup

See

If using PRIMECLUSTER with an Oracle VM Server for SPARC environment, refer to "Chapter 14 Using PRIMECLUSTER in Oracle VM Server for SPARC Environment."

If using PRIMECLUSTER with Oracle Solaris Kernel Zones environment, refer to "Chapter 15 Using PRIMECLUSTER in Oracle Solaris Kernel Zones Environment ."

If using PRIMECLUSTER with an Oracle Solaris Zones environment, refer to "Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment."

Note

- When mirroring the system disk using GDS in a ZFS boot environment, there are some notes on the installation of the operating system. For details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide".
- PRIMECLUSTER is guaranteed to work with a maximum umask value of "022." Do not modify the umask value.
- Change the system locale after setting up the shutdown facility.
  When changing the system locale before setting up the shutdown facility (for example, changing the system locale to EUC in the Oracle Solaris 11 environment), the FJSVcluster format message output in the /var/adm/messages file at the time of setting up the shutdown facility may become garbled. If this phenomenon occurs, restart the system to correct the garbled message.

3.1 PRIMECLUSTER Installation

There is a following method to install PRIMECLUSTER.

- Installation script

  The installation script is also called the CLI Installer. It is used to install PRIMECLUSTER on a system in which Solaris software and related Fujitsu software have been installed. This method is also used for the installation of the cluster management server.

See

For details on the installation procedures, see the *"PRIMECLUSTER Installation Guide."*

Note

When using Oracle Solaris Zones environments

When Oracle Solaris Zones of an exclusive IP zone exist on the system, perform the settings until "5.1.1 Setting Up CF and CIP" before booting a non-global zone.

If you execute a boot command of Oracle Solaris Zones while the settings of CF and CIP have not been done, the command will terminate abnormally.

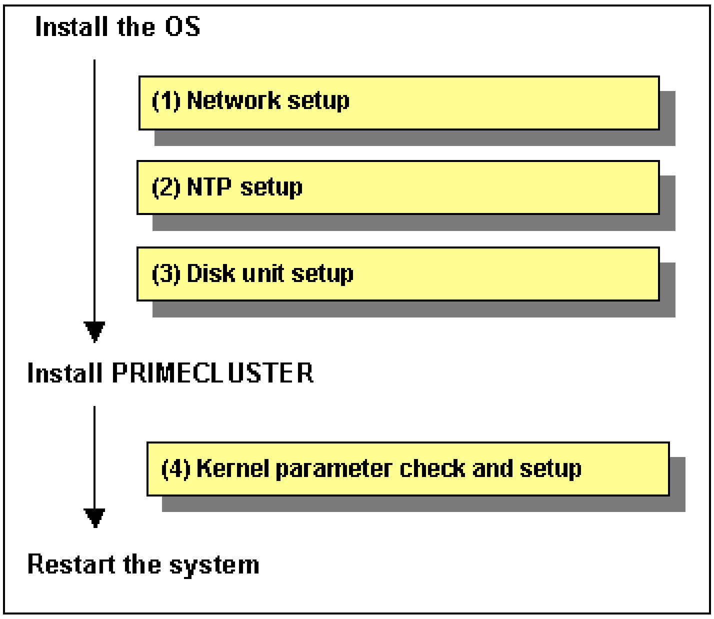
[Output example]

```
# zoneadm -z <zone_name> boot
zone '<zone_name>': WARNING: unable to add network interface '<NIC_name>': link busy
zone '<zone_name>': failed to add network device
zoneadm: zone '<zone_name>': call to zoneadmd failed
```

3.2 Installation and Setup of Related Software

After installation of the software products related to PRIMECLUSTER is done, you need to set up the OS and hardware for installing and operating PRIMECLUSTER.

Set up the following if necessary.



- Network setup

  If a network adapter for a public LAN has been installed, you need to configure the IP addresses after installing the OS.

  Perform this setup after installing the operating system.

  For details, see "3.2.1 Setting Up the Network".

- NTP setup

  This setup synchronizes the time in all of the nodes of the cluster system. This NTP setup needs to be done before installing PRIMECLUSTER.

- Disk unit setup

  When using a shared disk unit, you need to install and set up the related software product. This disk unit setup needs to be done before installing PRIMECLUSTER.

  For details, see "3.2.2 Setting Up Disk Units".

- Kernel parameter check and setup

  When operating the software products related to PRIMECLUSTER, you need to adjust kernel parameters according to the hardware configuration and applications. This kernel parameter setup should be done before operating PRIMECLUSTER.

  For details, see "3.2.3 Checking the Kernel Parameters".

3.2.1 Setting Up the Network

If a network adapter for a public LAN has been installed, the IP address setup is required.

See

- For details, see Oracle Solaris documents.
- For information on changing the public LAN that the PRIMECLUSTER system uses, see "9.2.1 Changing an IP Address on the Public LAN".

Information

Web-Based Admin View automatically sets up an interface that was assigned the IP address of the host name corresponding to the node on which PRIMECLUSTER was installed. This interface will be used as a transmission path between cluster nodes and cluster management server, and between cluster management servers and clients.

3.2.2 Setting Up Disk Units

Installation and Setup of Related Software

Install and set up the software products required for using shared disk units. See the installation guides and manuals of each software product.

Point

- If the disk unit uses a Fibre Channel interface, a Fibre Channel driver for the card needs to be installed and set up. For installation details, see "Installation Procedures" in *"FUJITSU PCI Fibre Channel x.x Guide for Solaris(TM) Environment."*
- If the disk array unit is used with a multipath configuration, refer to the following manuals for instructions on installing and setting up the software.
    - When the ETERNUS multipath driver (GRMPD) is used
        - "*ETERNUS Multipath Driver User's Guide*"
    - When the multipath disk control load balance option (MPLB) is used
        - *"INSTALLATION GUIDE - Multipath Disk Control load balance option x.x for Solaris(TM) Environment"*
        - *"Multipath Disk Control Load Balance option x.x Guide"*
    - When multipath disk control (MPHD) is used
        - "INSTALLATION GUIDE - Multipath Disk Control x.x for Solaris(TM) Environment "
        - *"Multipath Disk Control x,x Guide"*
- If an ETERNUS2000, ETERNUS3000, ETERNUS4000, ETERNUS6000, or ETERNUS8000 disk array is used with a single-path configuration, you need to install and setup a hard disk driver (HDDV).
  For details on installing HDDV, see the *"Hard Disk Driver x.x Guide."*
- One of the following settings is necessary when the ETERNUS multipath driver is used and each node has local disks only used by the node itself:
    - Allocate an instance number that is not duplicate to the multipath device of the local disk on each node.
    - If the instance number is duplicate in the multipath device of the local disk on each node, register the multipath device of the local disk to the exclusive device list before automatic resource registration.
      For details on the exclusive device list, see "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

  If the setting above is not performed even if the instance number is duplicate in the multipath device of the local disk on each node, the automatic resource registration fails.

Preparation for automatic shared disk configuration

Multipath disk

Before executing automatic shared disk configuration, format the logical units and assign labels to the units.

See

- For information on the procedures, refer to the manual ("Multipath Disk Control x,x Guide," "Multipath Disk Control load balance option," or "ETERNUS Multipath Driver User's Guide") of the software product to be used.
- For information on automatic configuration of shared disks, see "5.1.3 Initial Setup of the Cluster Resource Management Facility" and "6.3.1 Automatic Configuration of Shared Disks."

Note

At this stage, do not create logical paths for MPHD, MPLB, and GRMPD.

If a shared disk is automatically configured in a cluster system where MPLB or ETERNUS multipath driver is installed, and the disk supports the load balancing, the MPLB logical path will automatically be generated, then the resource will also be registered.

If you want to make the MPHD logical path automatically generate and the MPHD resource register by performing automatic configuration of a shared disk, it is necessary to set the parameter beforehand using the clsetacfparam(1M) command.

See

For details on the parameters to be set with the clsetacfparam(1M) command, see "5.3 Setting Up Shared Disk Connection Confirmation."

If the disk is not a multipath disk

Before executing automatic shared disk configuration, you need to assign a label to the disk by executing the "format(1M)" command.

See

For information on automatic configuration of shared disks, see "5.1.3 Initial Setup of the Cluster Resource Management Facility" and "6.3.1 Automatic Configuration of Shared Disks."

3.2.3 Checking the Kernel Parameters

You need to edit the values for the kernel parameters according to the environment.

Target node:

All nodes in which PRIMECLUSTER is to be installed

The kernel parameters differ according to the products and components to be used.

Check the Kernel Parameter Worksheet, and edit the value if necessary.

See

For information on the kernel parameters, see the "A.5 Kernel Parameter Worksheet".

Note

To enable the kernel parameter, it is necessary to reboot the system after the parameter setting.

3.3 Installation and Environment Setup of Applications

Install software products to be operated on the PRIMECLUSTER system and configure the environment as necessary.

To bring about application switchover in the event of a failure, you need to register the resources of software application to RMS. RMS will monitor these resources. For details, see "Chapter 6 Building Cluster Applications".

See

- For information on products supported by the PRIMECLUSTER system, see "Table 19.1 PRIMECLUSTER product list".
- For details on installing applications, see the manuals and installation guides for the individual applications.

Chapter 4 Preparation Prior to Building a Cluster

This chapter explains the preparation work that is required prior to building a cluster, such as starting up the Web-Based Admin View screen.

See

As preparation for building the cluster, check the operation environment. See "2 Operating environment" in the "PRIMECLUSTER Installation Guide."

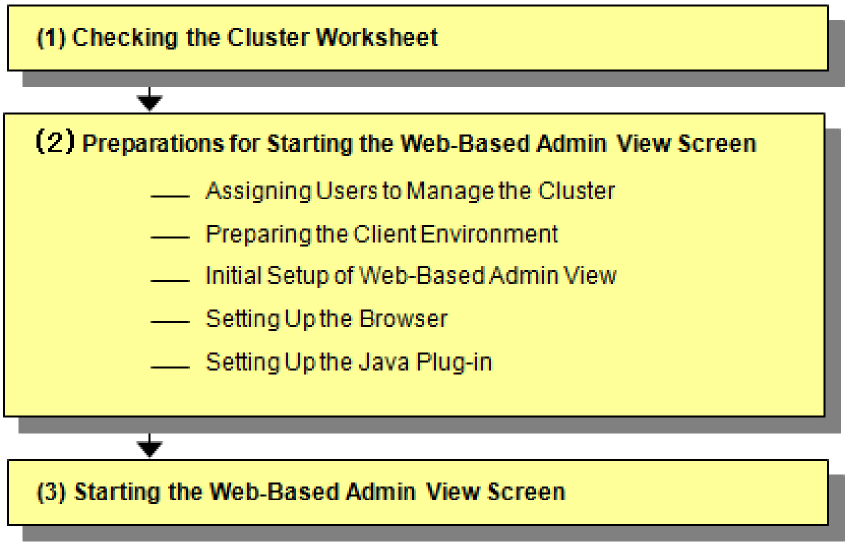


Table 4.1 Operation procedure and manual reference location for starting the Web-Based Admin View screen

	Work item	Execution Node	Required/ Optional	Manual reference location*
(1)	4.1 Checking the Cluster Worksheet	-	Required	This manual: Appendix A
(2)	4.2.1 Assigning Users to Manage the Cluster	Cluster node	Required	
	4.2.2 Preparing the Client Environment	Client	Required	**WEB** "3.1.2 Prerequisite client environment"
	4.2.3 Initial Setup of Web-Based Admin View	Cluster node	Required	
	4.2.4 Setting Up the Browser	Client	Required	**WEB** "3.1.3.1 Preparing the Web browser"
	4.2.5 Setting Up the Java Plug-in	Client	Required	**WEB** "3.1.3.2 Required for the Web Browser Environment"
(3)	4.3 Starting the Web-Based Admin View Screen	Client	Required	**WEB** "3.2 Screen startup"

***** The PRIMECLUSTER manual name is abbreviated as follows:

- WEB: *PRIMECLUSTER Web-Based Admin View Operation Guide*

4.1 Checking the Cluster Worksheet

Make certain of filling the worksheet. If there is missing information, you should specify values and complete the sheet.

See

For information on the worksheet, see "Appendix A PRIMECLUSTER System Design Worksheets".

4.2 Preparations for Starting the Web-Based Admin View Screen

Take the following steps for starting the GUI screen of Web-Based Admin View:

1. Environment setup

   You can set up the following in any order:

   - 4.2.1 Assigning Users to Manage the Cluster
   - 4.2.2 Preparing the Client Environment
   - 4.2.3 Initial Setup of Web-Based Admin View

2. Web environment preparation

   You need to set up the browser first.

   - 4.2.4 Setting Up the Browser
   - 4.2.5 Setting Up the Java Plug-in

4.2.1 Assigning Users to Manage the Cluster

Web-Based Admin View restricts access to specific operation management GUIs by using user groups in the management server.

The table below shows the groups used for operation management GUIs of PRIMECLUSTER.

Table 4.2 Operation management GUIs of Web-Based Admin View and authorized user groups

GUI name	UNIX user group name	Privileges
All GUIs	wvroot	Root authority. This group can execute all operations.
Cluster Admin	clroot	Root authority. This group can specify settings, execute management commands, and display information.
	cladmin	Administrator authority. This group cannot specify settings. It can execute management commands and display information.
	clmon	User authority. This group cannot specify settings and cannot execute management commands. It can only display information.
GDS (Global Disk Services)	sdxroot	Root authority. This group can use the GDS management view.

The groups for the operation management GUIs are defined as shown in the above table.

wvroot is a special user group, and is used for Web-Based Admin View and GUIs. Users belonging to this group are granted the highest access privileges for Web-Based Admin View and all kinds of operation management GUIs.

The system administrator can grant different access privileges to users according to the products that the users need to use.

For example, a user who belongs to the "clroot" group but not to "sdxroot" is granted all access privileges when opening the Cluster Admin screen but no access privileges when opening the Global Disk Services (GDS) GUIs.

You need to create the above UNIX groups for both the primary and secondary management servers as necessary, and assign users to these groups. The Web-Based Admin View group membership should maintain consistency among all management servers associated with a specific cluster system. To set a user group, execute the "useradd(1M)" command as follows:

```
# useradd -g wvroot usrname
```

The root user is granted the highest access privilege regardless of which group the root user belongs to.

For details about user groups, see "3.1.1 User group determination" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

4.2.2 Preparing the Client Environment

Prepare hardware, operating systems, and Web browsers of the clients supported by Web-Based Admin View.

See

For details, see "3.1.2 Prerequisite client environment" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

4.2.3 Initial Setup of Web-Based Admin View

4.2.3.1 Initial setup of the operation management server

When using Web-Based Admin View for the first time, you need to initialize the management server on each node. Take the following steps in the order listed below.

Operation Procedure:

1. Stop Web-Based Admin View on all the management servers and nodes.

```
# /etc/init.d/fjsvwvcnf stop
# /etc/init.d/fjsvwvbs stop
```

2. Set the IP addresses or host names of the primary management server and secondary management server.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server <primary-management-server-IP-address or primary-management-server-host-name>
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server <secondary-management-server-IP-address or secondary-management-server-host-name>
```

- In the case of multiple-node cluster operation
  Execute the commands on all the management servers and nodes, referring to the example.
  Example: when the primary management server IP is 10.20.30.40 and the secondary management server IP is 10.20.30.41

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server 10.20.30.41
```

- In the case of the single-node cluster operation
  For each IP address or host name of the primary management server and secondary management server, specify the node's own IP address or host name.
  Example: when the node's own IP address is 10.20.30.40

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server 10.20.30.40
```

  In this case, no value is displayed in secondary management server information field on the Web-Based Admin View screen.

3. Restart Web-Based Admin View on all the management servers and nodes.

```
# /etc/opt/FJSVwvbs/etc/bin/wvCntl restart
# /etc/init.d/fjsvwvcnf start
```

See

Web-Based Admin View has some different operation management modes. For further details, see "1.2.2 System topology" and "Chapter 7 Web-Based Admin View setup modification" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

Note

Only the IP addresses can be set to the primary management server and secondary management server.

4.2.3.2 Confirming Web-Based Admin View Startup

This section describes the procedure for confirming whether Web-Based Admin View has been started.

Confirmation Procedure

Check that all node information is output by executing the "wvstat(1M)" command on the connected management server.

Example:

In a two-node configuration consisting of node1(10.20.30.40) and node2(10.20.30.41), node1 is the primary management server and node2 is the secondary management server.

```
# /etc/opt/FJSVwvbs/etc/bin/wvstat
primaryServer  10.20.30.40 node1  http=10.20.30.40   Run  1d+7h36m
primaryServer   Sessions: 0
primaryServer   Nodes: 2
        10.20.30.40 node1   SunOS-5.11       1d+7h36m
        10.20.30.41 node2   SunOS-5.11       1d+7h36m
secondaryServer 10.20.30.41 node2  http=10.20.30.41   Run  1d+7h36m
secondaryServer Sessions: 0
secondaryServer Nodes: 2
        10.20.30.41 node2   SunOS-5.11       1d+7h36m
        10.20.30.40 node1   SunOS-5.11       1d+7h36m
#
```

If the information is not properly displayed, Web-Based Admin View has not been started or there may be an error in the Web-Based Admin View settings. Restart Web-Based Admin View and execute the operation again. If node information is still not displayed, refer to the *"PRIMECLUSTER Web-Based Admin View Operation Guide"* and check the parameter settings.

For details on the wvstat(1M) command, see the manual page.

4.2.3.3 Setting the Web-Based Admin View Language

The language environment in which Web-Based Admin View operates is set to English as default. Even though the client has a Japanese environment, the text of cluster resource management facility messages that are sent from the cluster node is displayed in English.

If you want to display the messages in Japanese, take the following steps to set up environment variables of Web-Based Admin View. You need to set up the variables using a system administrator access privilege on all nodes and the cluster management servers that construct a cluster system.

This operation must be executed with the system administrator authority for all cluster nodes and the cluster management server that make up the cluster system.

Table 4.3 Environment variable for the operation language of Web-Based Admin View

Attribute	Variable	Possible values	Meaning
sys	Lang	C, ja	Language environment in which Web-Based Admin View operates. C: Operates in English. ja: Operates in Japanese.

Attribute	Variable	Possible values	Meaning
			If this variable is not set, Web-Based Admin View operates in the English environment.

Operation Procedure:

1. Stop Web-Based Admin View on all the management servers and nodes.

```
# /etc/init.d/fjsvwvcnf stop
# /etc/init.d/fjsvwvbs stop
```

2. Add the environment variable to the definition file (/etc/opt/FJSVwvbs/etc/webview.cnf) of Web-Based Admin View, and set the language.
Execute the following command on all the management servers and nodes, referring to the example.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam -add <attribute> <environment-variable> <set-value>
```

Example: Add the environment variable and set the operation language to Japanese.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam -add sys lang ja
```

3. Restart Web-Based Admin View on all the management servers and nodes.

```
# /etc/opt/FJSVwvbs/etc/bin/wvCntl restart
# /etc/init.d/fjsvwvcnf start
```

Note

- For Web-Based Admin View to display messages in Japanese, the language environment of the personal computers that are being used as clients must be set to Japanese. If a client has an English environment, the message contents turn into garbled characters by the above setting change.
- To change the environment variable again after it is added by the above procedure, execute the following command:

  ```
  # /etc/opt/FJSVwvbs/etc/bin/wvSetparam lang <setting_value>
  ```

  For details on the command, see "4.5.3 Environment variable modification" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

4.2.4 Setting Up the Browser

Set up a Web browser on the clients.

See

See "3.1.3.1 Preparing the Web browser" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

4.2.5 Setting Up the Java Plug-in

Install the Java Plug-in on the clients.

See

For details on the supported Java Plug-in versions, see "4.2.2 Preparing the Client Environment." For instructions on setting up the Java Plug-in, see "3.1.3.2 Required for the Web Browser Environment" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

4.3 Starting the Web-Based Admin View Screen

After completing all the preparations, start the Web-Based Admin View GUI screen.

Operation Procedure:

1. Start the Web browser in the client.

2. Specify the URL in the following format, and access the cluster management server:

   http://<*host-name*>:<*port-number*>/Plugin.cgi

   <host-name>

   The IP address or the host name (**httpip**) that clients use to access the primary or secondary management server.

   The default value of **httpip** is the IP address that is assigned to the node name that is output when "uname -n" is executed.

   <port-number>

   Specify "8081."

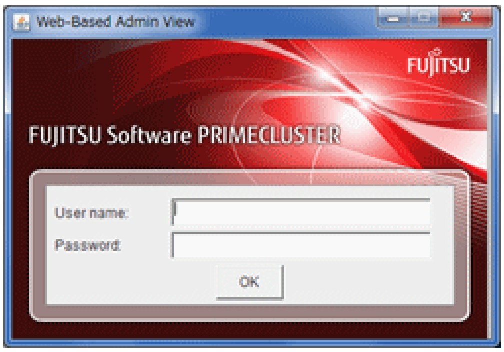
   If the port number has been changed, specify the up-to-date number.
   For instructions on changing the http port number, see "7.2.1 http port number" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

   Note

   - If the Web-Based Admin View screen cannot be started when the host name is specified in **<*host-name*>**, specify the IP address directly that corresponds to the host name.
   - When specifying the IPv6 address for <*host-name*>, enclose it in blankets "[ ]".
     (Example: http://[1080:2090:30a0:40b0:50c0:60d0:70e0:80f0]:8081/Plugin.cgi)
   - Note that the access method may be different depending on the operation management product. To use operation management products that have different access methods at the same time, see "3.3.4 Concurrent use of operation management products with different access methods" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*
   - For information on the IP address or host name (httpip) used by clients, see the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

3. When the Web-Based Admin View is started, the following window appears.

   Enter a user name and password that have been registered to the management server, and click *OK*.



Note

You cannot close this window by clicking "x" at the top right corner.

4. When authentication ends, you can use Web-Based Admin View.

Note

- After starting the Web-Based Admin View screen, do not change the page by pressing the *Forward/Next*, *Back*, or *Reread/Refresh* buttons.
- The screen of some browsers may hang.
  - If the browser is terminated, restart the browser and display Web-Based Admin View.
  - Reread the URL or restart the browser if the browser does not work (no response).
  - The operation can be continued after the Web-Based Admin View is started by moving the mouse on the browser if the page switches to a black screen.
- If "Welcome to Web-Based Admin View" does not appear after you read the URL of the Java Plug-in with Internet Explorer, an appropriate Java Plug-in may not be installed. Confirm that an appropriate Java Plug-in is installed by using "Add/Remove Programs" in the control panel. If the Java Plug-in is not installed or if an older Java Plug-in version that is not supported is installed, see the *"PRIMECLUSTER Web-Based Admin View Operation Guide"* and install the Java Plug-in. Also, if the "security warning" dialog box appears, and prompts you to specify whether the "Java Plug-in" is to be installed and executed, select *No*.
- If the secondary cluster management server is set to operate dynamically, there is a function that connects automatically to the primary or secondary management server that is operating at that time even if the URL of a specific monitoring node is specified. For details, see "7.4 Secondary management server automatic migration" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*
- If repeated errors occur during the authentication of Step 3, the message 0016 may be displayed and you may not be able to log in. For the action to take if this happens, see "Symptom 17" in "Appendix B Troubleshooting" of the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*
- If some problems occur while you are using Web-Based Admin View, see "Appendix A Message" and "Appendix B Troubleshooting" of the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

4.4 Web-Based Admin View Screen

When you start Web-Based Admin View, the Web-Based Admin View screen is displayed.

The left area of the Web-Based Admin View screen displays the currently supported functions as **operation menus**.

4.4.1 Operation Menu Functions

Web-Based Admin View screen supports the functions shown below.

See "Menu Outline."

Figure 4.1 Web-Based Admin View screen

Menu Outline

The operation menus are divided into the following two types:

a. Management screens and manuals of operation management products that are provided by PRIMECLUSTER

b. Management screens and manuals of operation management products that are presented by a software product other than PRIMECLUSTER, such as the Enhanced Support Facility (ESF)

- Operation menu of a.
    - Operation management product name (PRIMECLUSTER)

      You can operate the screen of the operation management product.

      For details, see the manual provided with each operation management product.

        - Global Cluster Services (CF, CRM, RMS)
        - Global Disk Services
        - Global File Services (only in Oracle Solaris 10 environment)
    - Web-Based Admin View tools

      These tools display the Web-Based Admin View log and allow you to set the operation environment. For details, see "Part 3 Web-Based Admin View tools menu" in the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*
    - Manual

      The PRIMECLUSTER online manual is displayed.
- Operation menu of b.

  Buttons for starting the management screens of installed operation management products other than the PRIMECLUSTER products are displayed. In this system, this menu category is displayed because Enhanced Support Facility (ESF) is installed.

  For details, see the manual provided with each operation management product.

    - Operation management product name (other than PRIMECLUSTER)

      You can operate the management screens of installed operation management products other than the PRIMECLUSTER products.

- Machine Administration

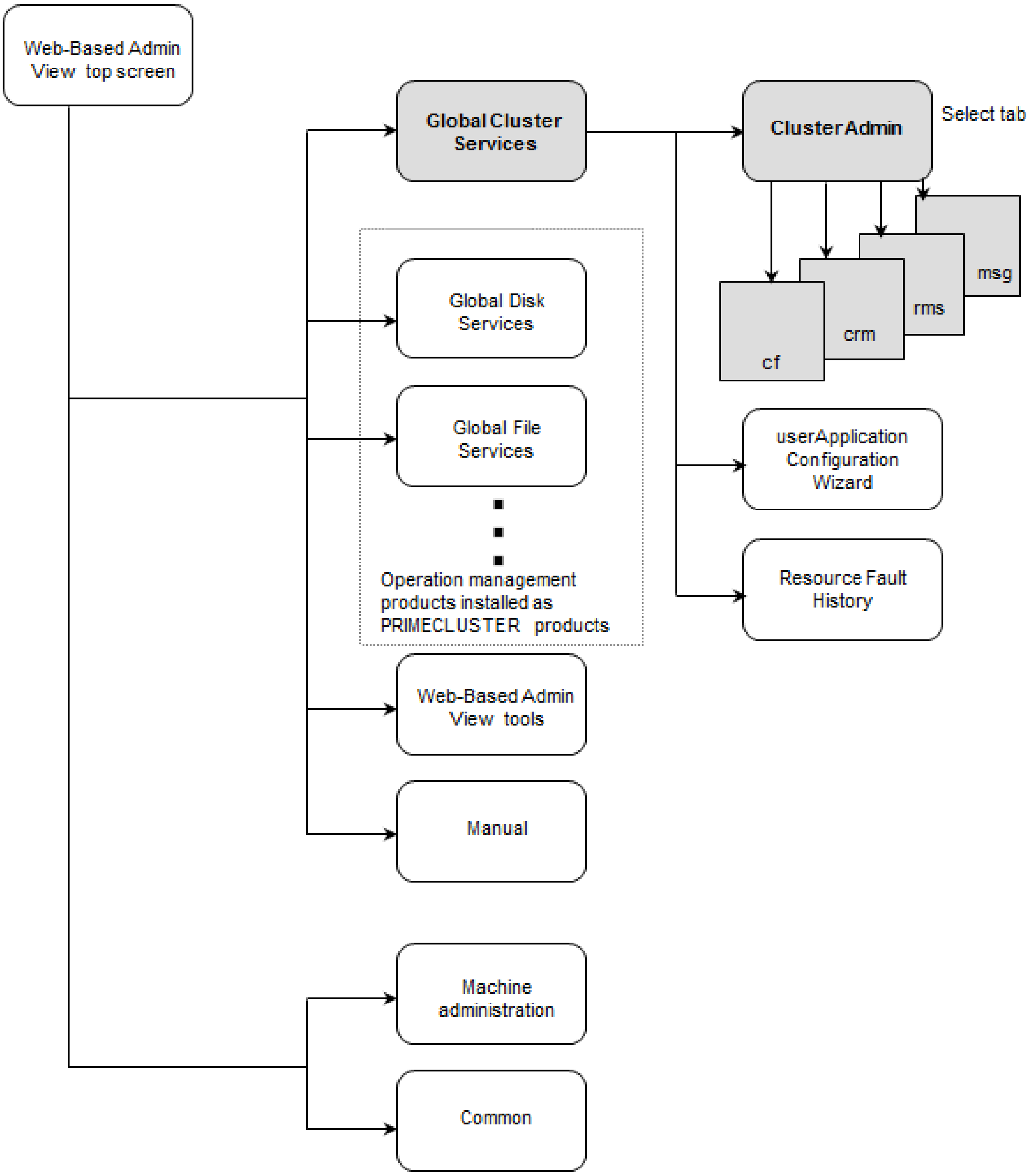
- MultiPathDisk view

- Common

  You can refer to manuals that are available as online manuals. For details, see the *"PRIMECLUSTER Web-Based Admin View Operation Guide."*

Operation menu transition diagram

Shown below are the transitions from the top screen of Web-Based Admin View to the other screens.



On the Cluster Admin screen, you can switch the window by clicking the following tab:

- cf: Cluster Foundation

- crm: Cluster Resource Management

- rms: Reliant Monitor Services

- msg: Message

The following sections describe the screens found after the Global Cluster Services menu.

4.4.2 Global Cluster Services Menu Functions

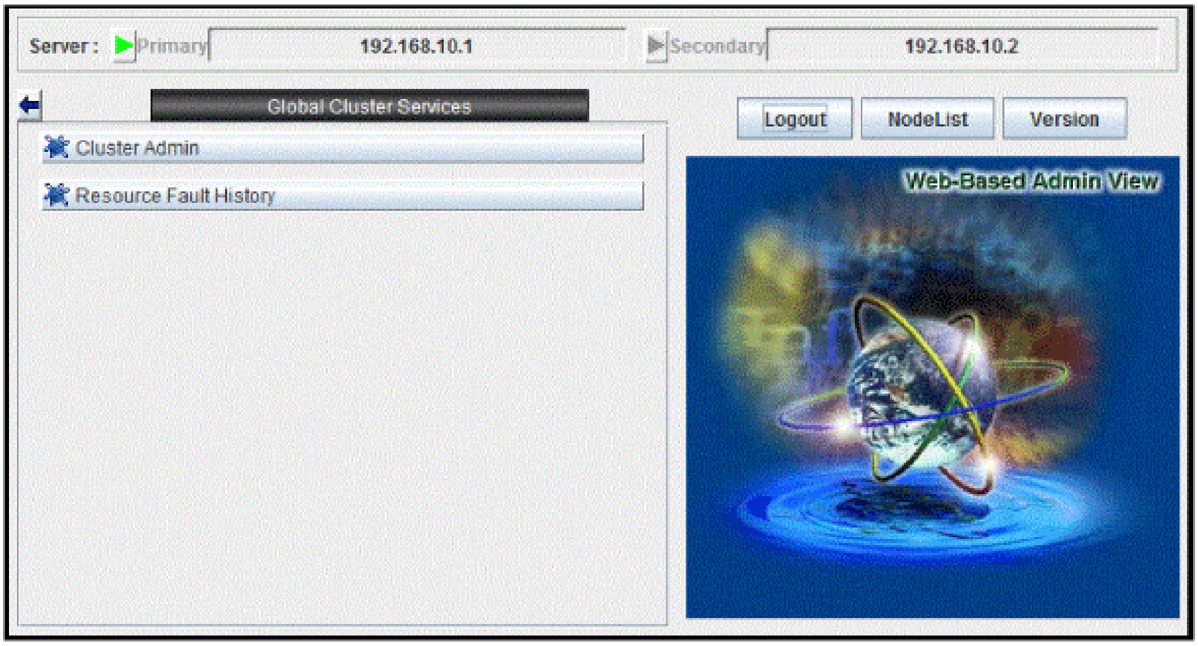
Display procedure

Web-Based Admin View screen -> *Global Cluster Services*

Exit procedure

To return to the Web-Based Admin View screen, click the arrow next to *the Global Cluster Services*.

Figure 4.2 Web-Based Admin View screen (Global Cluster Services menu)



Overview of the Global Cluster Services menu

- Cluster Admin

  This function allows you to monitor the status of the PRIMECLUSTER system and operate the system.

- userApplication Configuration Wizard

  This function allows you to create cluster applications.

- Resource Fault History

  This function allows you to display the resource fault history. For details, see "C.3.2 Resource Fault History."

4.4.3 Cluster Admin Functions

Display procedure

Web-Based Admin View screen -> Select *Global Cluster Services*. -> Select *Cluster Admin*. -> Node selection screen -> Select the node.

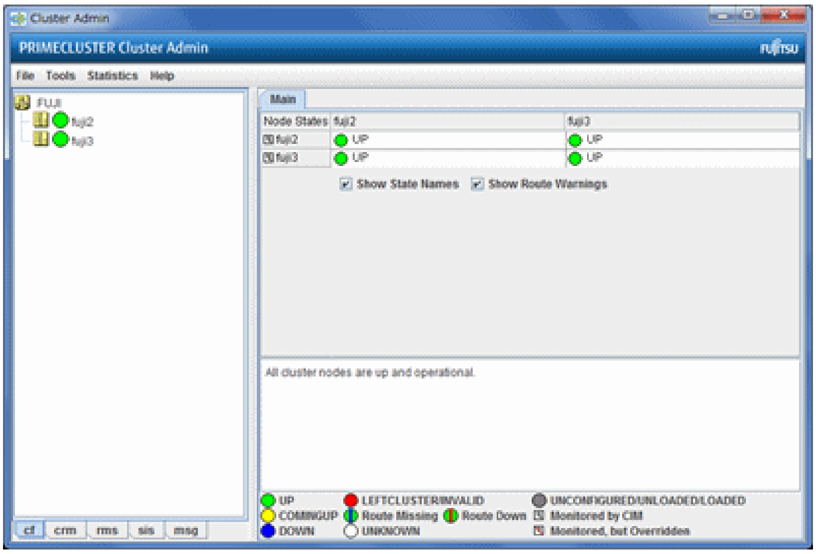
Exit procedure

Select the *Exit* in the *File* menu. -> Confirmation screen -> Select the *Yes*. -> *Global Cluster Services* menu

Cluster Admin supports the functions described below.

The manual reference locations are indicated in "Overview of Cluster Admin."

Figure 4.3 Web-Based Admin View screen (Cluster Admin)



Overview of Cluster Admin

- cf (CF: Cluster Foundation)

  This function allows you to manage, build, monitor, and diagnose the cluster.

  Reference location: "5.1 Initial Cluster Setup", "Chapter 7 Operations"

- crm (CRM: Cluster resource management facility)

  This function manages the resource database, which contains information about the hardware devices (including shared disks, line switching units, and network interface cards).

  Reference location: "5.1.3 Initial Setup of the Cluster Resource Management Facility", "Chapter 7 Operations"

- rms (RMS: Reliant Monitor Services)

  This function monitors the status of the cluster system and manages applications and resources.

  Reference location: "Chapter 7 Operations"

- msg (Message)

  Cluster control messages are displayed.

  Reference location: "Chapter 7 Operations"

4.4.4 userApplication Configuration Wizard Functions

Display procedure

Web-Based Admin View screen -> *Global Cluster Services* -> *userApplication Configuration Wizard*

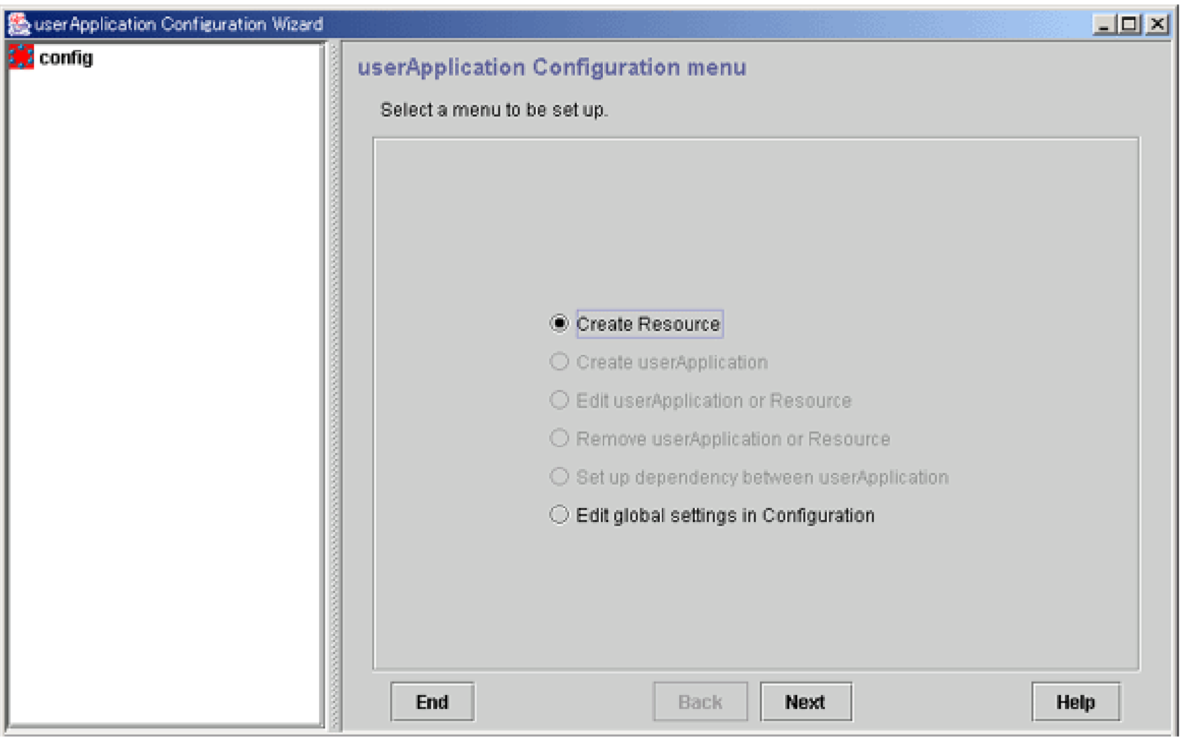
Exit procedure

Click *the End* button. -> Global Cluster Services menu

userApplication Configuration Wizard supports the functions described below.

The manual reference locations are indicated in "Overview of userApplication Configuration Wizard."

Figure 4.4 Web-Based Admin View screen (userApplication Configuration Wizard)



Overview of userApplication Configuration Wizard

- Create Resource

  This function creates a new resource.

  Reference location: "6.7.1 Setting Up Resources"

- Create userApplication

  This function creates cluster applications.

  Reference location: "6.7.2 Creating Cluster Applications"

- Edit userApplication or Resource

  This function edits the attributes of cluster applications or resources.

  Reference locations: "10.6.2 Changing the Attributes Used by a Resource or a Resource Interface" and "11.1 Changing the Operation Attributes of a Cluster Application"

- Remove userApplication or Resource

  This function removes cluster applications or resources.

  Reference location: "10.2 Deleting a Cluster Application"

- Set up dependency between userApplication

  This function sets up dependency between cluster applications.

  Reference location: "6.7.3 Setting Up Dependency Relationships Between Cluster Applications"

- Edit global settings in Configuration

  This function edits global settings in configuration.

  Reference location: "6.7.4 Editing global settings in Configuration"

4.5 Exiting the Web-Based Admin View Screen

To exit the Web-Based Admin View screen, follow the procedure below.

Logging out of the screen

To log out of the Web-Based Admin View screen, follow the procedure below.

1. Close all screens if the management screen of the operation management product is displayed.
2. When only the Web-Based Admin View screen is displayed, select the *Logout*.

Exiting the screen

To exit the Web-Based Admin View screen, follow the procedure below.

1. Log out from the Web-Based Admin View screen according to "Logging out of the screen" described above.
2. The login screen will be displayed. To exit the Web-Based Admin View screen, execute one of the following operations while the login screen is still displayed:
    - Terminate the Web browser.
    - Specify another URL in the Web browser to switch the screen. (Enter a new URL or specify a bookmark.)
    - Select the *Back* button of the browser.

Note

- To terminate the Web browser, select the *Close* in the *File* menu, or click the "x" at the top right corner of the screen.
- At the login screen, clicking the "x" at the top right corner of the screen will not terminate the screen.
- The login screen will remain temporarily after exiting the browser.

Chapter 5 Building a Cluster

The procedure for building a PRIMECLUSTER cluster is shown below.

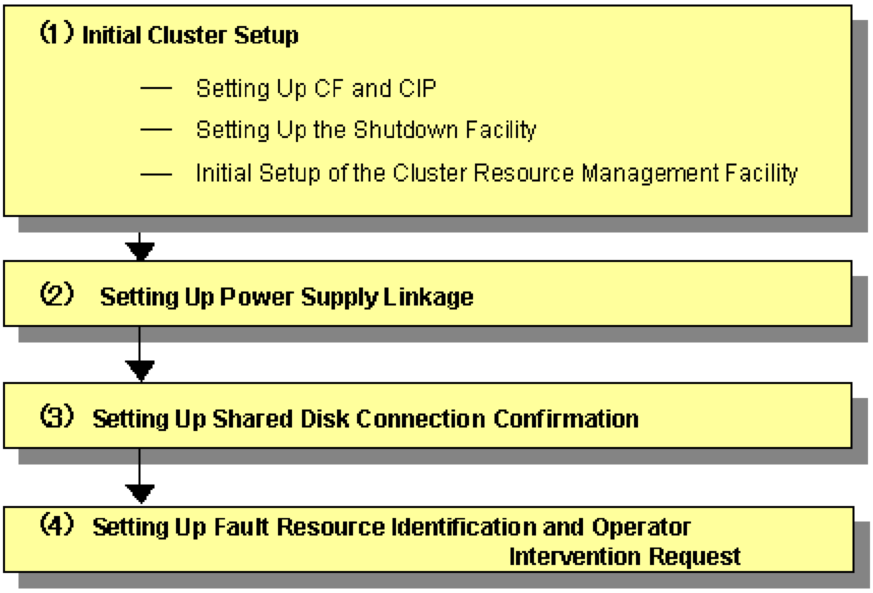


Table 5.1 Cluster configuration procedure and manual reference locations

	Work item	Execution Node	Required/ Optional	Manual reference location*1
(1)	5.1 Initial Cluster Setup			
	5.1.1 Setting Up CF and CIP	All nodes	Required	**CF** "2.1 CF, CIP and CIM configuration"
	5.1.2 Configuring the Shutdown Facility	All nodes	Required*2	**CF** "8 Shutdown Facility"
	5.1.3 Initial Setup of the Cluster Resource Management Facility	All nodes	Required	**CF** "4.3 Resource database configuration"
(2)	5.2 Setting Up Power Supply Linkage	All nodes	Optional	"Machine Administration Guide"
(3)	5.3 Setting Up Shared Disk Connection Confirmation	All nodes	Required	
(4)	5.4 Setting Up Fault Resource Identification and Operator Intervention Request	1 node	Required	

***1** The PRIMECLUSTER manual names are abbreviated as follows:

- **CF**: *PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*

***2** When configuring a single node cluster, it is not necessary to setup the shutdown facility.

5.1 Initial Cluster Setup

This section describes the initial setup of the PRIMECLUSTER cluster.

For details on the setup methods, see the reference locations indicated in the table below.

	Contents	Manual reference location*1
1	Checking Console Configuration*2	- For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000: 5.1.2.2.1 Checking Console Configuration - For SPARC Enterprise T1000, T2000: 5.1.2.4.1 Checking Console Configuration - For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series: 5.1.2.3.1 Checking Console Configuration
2	5.1.1 Setting Up CF and CIP (setting up the cluster configuration information and the IP addresses)	**CF** "2.1 CF, CIP and CIM configuration"
3	5.1.2 Configuring the Shutdown Facility	**CF** "8 Shutdown Facility"
4	5.1.3 Initial Setup of the Cluster Resource Management Facility	**CF** "4.3 Resource database configuration"

*1 The PRIMECLUSTER manual names are abbreviated as follows:

- **CF**: *PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*

*2 "1. Check console information" is the prerequisite of "3. Configure Shutdown Facility". Be aware that you need to check the console information before "2. Configure CF and CIP".

5.1.1 Setting Up CF and CIP

Set up Cluster Foundation (CF) and CIP using the CF Wizard of Cluster Admin. Enter the information described in the table below. The worksheet for the CF and CIP setup is the "A.6 Cluster Configuration Worksheet". For details on the setting, see "2.1.4 Example of creating a cluster" in the *"PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide"*.

Setup item	Description	Described in worksheet
Cluster name	Define the name of the cluster systems. Use up to 31 printable ASCII characters (except space, line feed, and tab characters) for each name. Cluster names are always processed as uppercase characters.	"Cluster name" for "Cluster Foundation (CF)"
Cluster nodes	Select the nodes that will construct a cluster system.	"Node name (uname-n)" for "Node 1" and "Node 2"
CF node names	Set the names of the nodes that construct the cluster. Use up to 11 lower-case characters and symbols ("-" and "_") for each node name. For the first letter of the CF node names, set a lower-case character.	"CF node name" for "Node 1" and "Node 2"
Cluster interconnects	In each node of the cluster, determine the network interface to be used in CF inter-node communication. A representative network interface is the Ethernet device. Set the network interfaces to be used for CF inter-node communication so that they are activated when the system is started. However, IP addresses must not be assigned to the interfaces.	"Interconnect" - "Device name" for "Node 1" and "Node 2"
IP interconnects	Optional. This setup allows you to operate CF over IP. In a Solaris 11 environment, this function is not available.	"IP interconnect" for "Cluster Foundation (CF)"

Setup item	Description	Described in worksheet
CIP subnets	Set the following items, and set the IP address used by CF: - CIP subnet count - Host suffix - Subnet number - Subnet mask	"Subnets," "Subnet number," and "Subnet mask" for "Cluster Foundation (CF)", and "Interconnect" - "Subnet mask" for "Node 1" and "Node 2"
Usage confirmation of CF remote services	Check whether the following functions are to be enabled: - Remote file copy (cfcp) - Remote command execution (cfsh) With the default settings, these services are "Not selected". Make sure to select them for the following cases: - To use RMS - To use the Migration function in Oracle VM Server for SPARC Environment	"Usage of CF remote services" for "Cluster Foundation (CF)"
Cluster Integrity Monitor (CIM) configuration	Set the nodes to be monitored by CIM.	"Node in CF quorum set" for "Node 1" and "Node 2"

Note

- If you enable any one of the CF remote services, do not connect the following systems in the same cluster interconnect:
    - Systems that have a security problem
    - Systems in which the cluster interconnects are not secured
- If you used the userApplication Configuration Wizard (GUI), the two remote services, "remote file copy" and "remote command execution," will be enabled automatically. If the cluster interconnects are not secured, comment out the following two lines in the "/etc/default/cluster.config" file for all cluster nodes after you build userApplication:

```
CFCP    "cfcp"
CFSH    "cfsh"
```

  The results are as follows:

```
#CFCP    "cfcp"
#CFSH    "cfsh"
```

- A CF or CIP setup error can be due to one of the following causes:
    - A cluster interconnect is disconnected or incorrectly connected.
    - A network interface for a cluster interconnect is not activated.
    - An IP address is assigned to a network interface for a cluster interconnect.
- The cluster node name will be automatically used for the CF node name. The name string must consist of eleven characters or less. Change it if necessary.
- Do not add CIP node names manually to the /etc/inet/hosts file because the CF Wizard automatically updates the /etc/inet/hosts files on each node in the cluster.
- Check that CIP node names are not set to the /etc/inet/hosts file for each node in a cluster before executing the CF wizard. When the CIP node names are set, delete the entry of CIP node names.

- In Oracle Solaris VM for SPARC environment, the physical NIC of the cluster interconnects can be shared between the multiple cluster systems. In this case, separating network segments is unnecessary. In the same network segment, the physical NIC to be used as the cluster interconnects can be shared. However, note the following points for designing:
  - The IP address to be assigned for the cluster interconnects should not be duplicated between each cluster system.
  - The different cluster name needs to be created for each cluster system.
- A tagged VLAN interface cannot be used for the cluster interconnect.
- In Solaris 11 or later environment, when configuring the cluster system of the copy destination that is cloned by using Unified Archives, the symbolic links to some device files cannot be cloned in the cluster system environment of the copy destination.
  Apply the PRIMECLUSTER patch T012012SP-01 or later, and execute the following command to restore the file in the cluster system environment of the copy destination.

```
# /opt/SMAW/SMAWcf/bin/cfrecoverdev
```

Note

In the case of the single-node cluster operation

- When constructing multiple single-node clusters, specify a different cluster name per each node, such as including the node' own name in the name.
- For the network interface specified to the cluster interconnect, select a dedicated network interface device described in Note of "2.3.3 Single-Node Cluster Operation."
- Select [No] to the message of "Do you wish to invoke the Shutdown Facility Wizard to configure this cluster?" output after configuring CF in ClusterAdmin since the setting of the shutdown facility is unnecessary.

See

For information on the corrective action to be applied when the setting of the cluster interconnect fails, see "10 Diagnostics and troubleshooting" in the *"PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."*

5.1.2 Configuring the Shutdown Facility

This section explains the procedure for configuring the shutdown facility with the shutdown configuration wizard.

The configuration procedure for the shutdown facility varies depending on the machine type. Check the machine type of hardware and set an appropriate shutdown agent.

The following table shows the shutdown agent necessary by machine type.

<table>
<tr><th colspan="4" rowspan="3">Server machine type name</th><th colspan="5">Shutdown agent</th></tr>
<tr><th>XSCF SNMP</th><th>RCI</th><th>XSCF</th><th>ALOM</th><th>ILOM</th></tr>
<tr><th>Panic/Reset/
PPAR Reset</th><th>Panic/
Reset</th><th>Panic/
Reset/
Break</th><th>Break</th><th>Panic/Reset</th></tr>
<tr><td>SPARC Servers</td><td colspan="3">SPARC M10</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>N</td></tr>
<tr><td rowspan="3">SPARC Enterprise</td><td rowspan="3">M3000
M4000
M5000
M8000
M9000</td><td rowspan="2">Japan</td><td>Fujitsu</td><td>N</td><td>Y</td><td>Y</td><td>N</td><td>N</td></tr>
<tr><td>Other than Fujitsu</td><td>N</td><td>N</td><td>Y</td><td>N</td><td>N</td></tr>
<tr><td colspan="2">Other than Japan</td><td>N</td><td>N</td><td>Y</td><td>N</td><td>N</td></tr>
</table>

Server machine type name		Shutdown agent				
		XSCF SNMP	RCI	XSCF	ALOM	ILOM
		Panic/Reset/ PPAR Reset	Panic/ Reset	Panic/ Reset/ Break	Break	Panic/Reset
	T1000 T2000	N	N	N	Y	N
	T5120 T5220 T5140 T5240 T5440	N	N	N	N	Y (*1)
SPARC	T3 series	N	N	N	N	Y
	T4 series	N	N	N	N	Y
	T5 series	N	N	N	N	Y
	T7 series	N	N	N	N	Y
	S7 series	N	N	N	N	Y

Y: Required N: Not required
(*1) When using ILOM Reset, you need firmware for SPARC Enterprise server (System Firmware 7.1.6.d or later).

The following table shows the shutdown agent necessary for virtualized environments.

Table 5.2 Shutdown agents necessary for virtualized environments (Oracle VM Server for SPARC environment)

Server machine type name		Shutdown agent						
		XSCF SNMP					ILOM	
		Control domain			Guest domain		Control domain	
		Panic	Reset	PPAR Reset	Panic	Reset	Panic	Reset
SPARC Servers	SPARC M10	Y (*1)	Y (*1)	Y (*1)	Y (*2)	Y (*2)	N	N
SPARC	T3 series T4 series T5 series T7 series S7 series	N	N	N	N	N	Y	Y

Y: Required N: Not required
(*1) Required if used in the cluster between control domains.
(*2) Required if used in the cluster between guest domains.

Table 5.3 Shutdown agents necessary for virtualized environments (Oracle Solaris Kernel Zones environment)

Server machine type name		Shutdown agent									
		KZONE			XSCF SNMP					ILOM	
		Kernel Zone			Control domain			Guest domain		Control domain	
		Panic	Reset	Check	Panic	Reset	PPAR Reset	Panic	Reset	Panic	Reset
SPARC Servers	SPARC M10	Y	Y	Y	Y (*1)	Y (*1)	Y (*1)	Y	Y	N	N

Server machine type name		Shutdown agent									
		KZONE			XSCF SNMP					ILOM	
		Kernel Zone			Control domain			Guest domain		Control domain	
		Panic	Reset	Check	Panic	Reset	PPAR Reset	Panic	Reset	Panic	Reset
SPARC	T4 series T5 Series T7 Series S7 Series	Y	Y	Y	N	N	N	N	N	Y (*1)	Y (*1)

Y: Required N: Not required

(*1) Not required to configure the shutdown facility if you configure cluster system between Kernel Zones within a same physical partition.

Note

- When you are operating the shutdown facility by using one of the following shutdown agents, do not use the console.

  - XSCF Panic
  - XSCF Reset
  - XSCF Break
  - ILOM Panic
  - ILOM Reset

  If you cannot avoid using the console, stop the shutdown facility of all nodes beforehand. After using the console, cut the connection with the console, start the shutdown facility of all nodes, and then check that the status is normal. For details on stop, start, and the state confirmation of the shutdown facility, see the manual page describing sdtool(1M).

- In the /etc/inet/hosts file, you must describe the IP addresses and the host names of the administrative LAN used by the shutdown facility for all nodes. Check that the IP addresses and host names of all nodes are described.

- When you set up asynchronous RCI monitoring, you must specify the timeout interval (kernel parameter) in /etc/system for monitoring via SCF/RCI. For kernel parameter settings, see the section "A.5.1 CF Configuration."

- If a node's AC power supply is suddenly disconnected during operation of the cluster system, the PRIMECLUSTER, after putting the node for which the power supply was cut into LEFTCLUSTER status, may disconnect the console. In this instance, after confirming that the node's power supply is in fact disconnected, cancel the LEFTCLUSTER status using the cftool -k command. Afterwards, reconnect the console and switch on the power supply to the node.

- If the SCF/RCI is malfunctioning or if there is the detection of a hardware error such as the RCI cable being disconnected or detection of redundant RCI address settings, it will take a maximum of 10 minutes (from the time that the error is detected or the shutdown facility is started up) until those statuses are reflected to the sdtool -s display or shutdown facility status display screen.

- After setting the shutdown agent, conduct the cluster node forced stop test to check that the cluster nodes have undergone a forced stop correctly. For details on the cluster node forced stop test, see "1.4 Test."

- For using the Migration function of Oracle VM Server for SPARC, see "Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment."

- To make the administrative LAN, used in the shutdown facility, redundant by GLS, use the logical IP address takeover function of NIC switching mode, and configure the physical IP address for the administrative LAN of the shutdown facility.

- When using SPARC T5/T7/S7 series, make sure to read "5.1.2.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series" or "5.1.2.5 For Oracle Solaris Kernel Zones" before the settings.

See

For details on the shutdown facility and the asynchronous monitoring function, refer to the following manuals:

- "2.3.1.7 PRIMECLUSTER SF" in "PRIMECLUSTER Concepts Guide."
- "8. Shutdown Facility" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide".

5.1.2.1 For SPARC M10

5.1.2.1.1 Checking XSCF Information

The SNMP asynchronous monitoring function of the shutdown facility uses XSCF.

The connection method to XSCF can be selected from SSH or the telnet. Default connection is SSH.

Confirm the following settings concerning XSCF before setting the shutdown facility.

- Commonness
    - The log in user account must be made excluding root for the shutdown facility, and the platadm authority must be given.
    - The configuration information of the logical domains should be saved by the control domain before the showdomainstatus command is executed in XSCF. The state of the logical domains that configures the cluster should be displayed.
    - In the configuration where the asynchronous monitoring sub-LAN is not used, the host name corresponding to the IP address of XSCF-LAN#0 must be defined in /etc/inet/hosts.
- At the SSH connection
    - In XSCF, SSH must be effective in connected permission protocol type from the outside.
    - User inquiries of the first SSH connection (such as generation of the RSA key) must be completed by connecting to XSCF from all the cluster nodes via SSH using the log in user account for the shutdown facility.

      When using the host name for setting XSCF name, the first SSH connection with the host name must be completed.
- At the telnet connection
    - In XSCF, telnet must be effective in connected permission protocol type from the outside.

Note

When the connection to XSCF is a serial port connection alone, it is not supported in the shutdown facility. Connect to XSCF via SSH or telnet by using XSCF-LAN.

Record the following information that is necessary for setting the shutdown facility.

	Information	Description
(1)	PPAR-ID	Identification ID of the physical partition (PPAR) in which the logical domain of the cluster node belongs to. In SPARC M10-1 and M10-4, the value is "0". In SPARC M10-4S, the value is an integer ranged from 0 to 15. Execute the showpparstatus -a command on XSCF to display all PPAR-ID status. In the results of the showpparstatus -a command, PPAR-ID is displayed as two-digit by prefixing zero when it is one-digit. In this case, excluding the prefixing zero, take one-digit as a memo. Example: If zero is prefixed to the PPAR-ID status in the results of the showpparstatus -a command below, take zero as a memo if PPAR-ID is "00". Take one as a memo if PPAR-ID is "01".

	Information	Description
		`XSCF> showpparstatus -a` `PPAR-ID        PPAR Status` `00             Running` `01             Running` `XSCF>`
(2)	Domain-name	Logical domain name of the cluster node. Execute the virtinfo -a command on each node, and take the logical domain name as a memo. `# virtinfo -a` `Domain role: LDoms control I/O service root` `Domain name: primary` `             ^^^^^^^logical domain name` `Domain UUID: xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx` `Control domain: xxxxx` `Chassis serial#: xxxxxxxxxx` `#`
(3)	XSCF-name1	Host name or IP address for XSCF-LAN#0 of the unit in which the logical domain of the cluster node exists. (*1, *2, *3)
(4)	XSCF-name2	Host name or IP address for XSCF-LAN#1 of the unit in which the logical domain of the cluster node exists. (*1, *2, *3)
(5)	User-Name	User name to log in to the unit in which the logical domain of the cluster node exists. (*4, *5)
(6)	Password	Password to log in to the unit in which the logical domain of the cluster node exists. (*4, *5)
(7)	Administrative LAN	Administrative LAN of the cluster node used for the shutdown facility.
(8)	Asynchronous monitoring sub-LAN	Asynchronous monitoring sub-LAN of the cluster node used for the shutdown facility. (*3)

*1) When the network routing is set, the IP address of XSCF need not be the same to management LAN segment of the cluster node.

*2) For SPARC M10-4S environment, specify the XSCF takeover IP address.

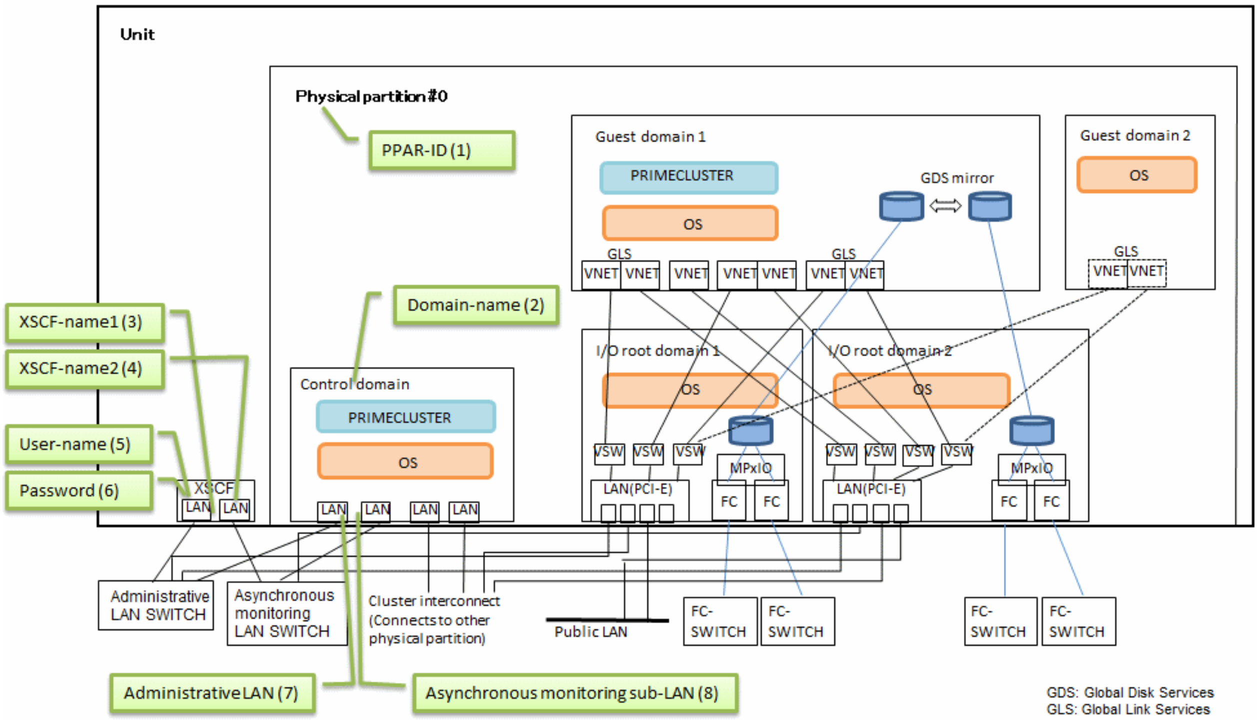
*3) In the configuration where the asynchronous monitoring sub-LAN is not used, record the IP address of XSCF-LAN#0 to "XSCFname1" and the host name corresponding to the IP address of XSCF- name1 to "XSCF-name2." For the configuration where the asynchronous monitoring sub-LAN is not used, see "2.2.2 XSCF Configuration in SPARC M10."

*4) In the environment where XSCF is duplexed, a combination of a user name and a password for 2 of the XSCF must be the same.

*5) To use the Migration function, set a combination of a user name and password for the XSCF and the connection method to the XSCF to be consistent on all nodes.

See the figure below to check the information used to set the shutdown facility.

Figure 5.1 Information used to set the shutdown facility when configuring the cluster on the control domain (reference)



See

For information on how to configure and confirm XSCF, see the "SPARC M10 Systems System Operation and Administration Guide".

5.1.2.1.2 Setting SNMP

Make settings for SNMP to use the SNMP asynchronous monitoring function.

Note

Port numbers for SNMP need to be changed under the following condition. For details, see "9.2.4 Changing Port Numbers for SNMP".

- When the port number 9385 used for the shutdown facility overlaps with a port number of the other products.

Setting up information related to the SNMP agent of XSCF

Set up the SNMP agent on all XSCF in the cluster.

1. Execute the showsnmp command to display SNMP settings.

```
XSCF> showsnmp
```

2. Execute the setsnmp command to set up the trap transmission for all the nodes configuring the cluster.

```
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous monitoring
sub-LAN on node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node2]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous monitoring
sub-LAN on node2]
```

Example

- XSCF on node1

```
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
monitoring sub-LAN on node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node2]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
monitoring sub-LAN on node2]
```

- XSCF on node2

```
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
monitoring sub-LAN on node1]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN on
node2]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
monitoring sub-LAN on node2]
```

3. Execute the setsnmp command to enable the SNMP agent.

```
XSCF> setsnmp enable
```

4. Execute the showsnmp command to check that the settings are enabled.

```
XSCF> showsnmp
```

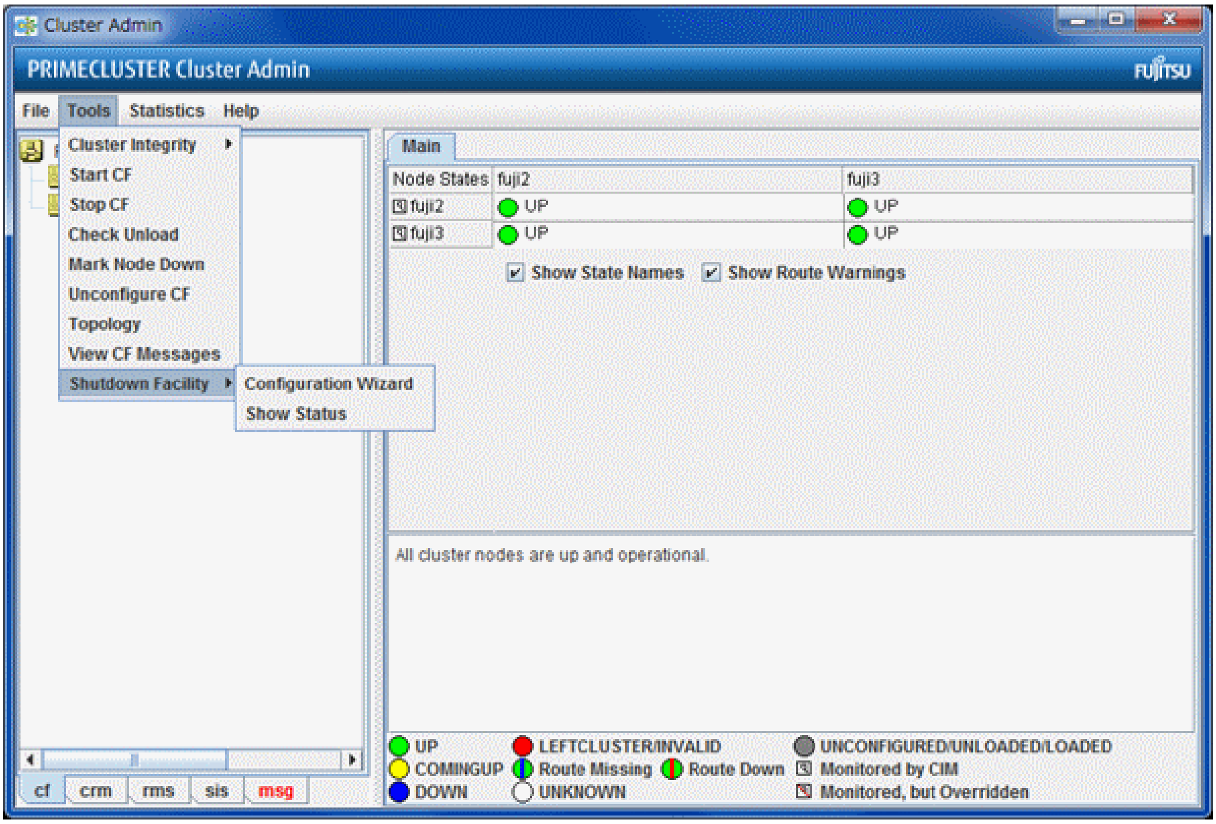
See

For information on how to configure and confirm XSCF related to SNMP agents, see the "SPARC M10 Systems System Operation and Administration Guide".

5.1.2.1.3 Using the Shutdown Configuration Wizard

Starting up the shutdown configuration wizard

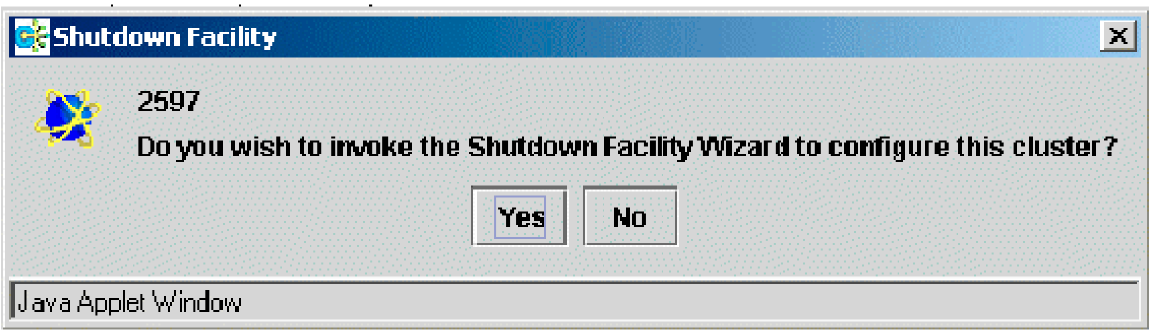
From the CF main window of the Cluster Admin screen, select the *Tool* menu and then *Shutdown Facility -> Configuration Wizard*. The shutdown configuration wizard will start.



Note

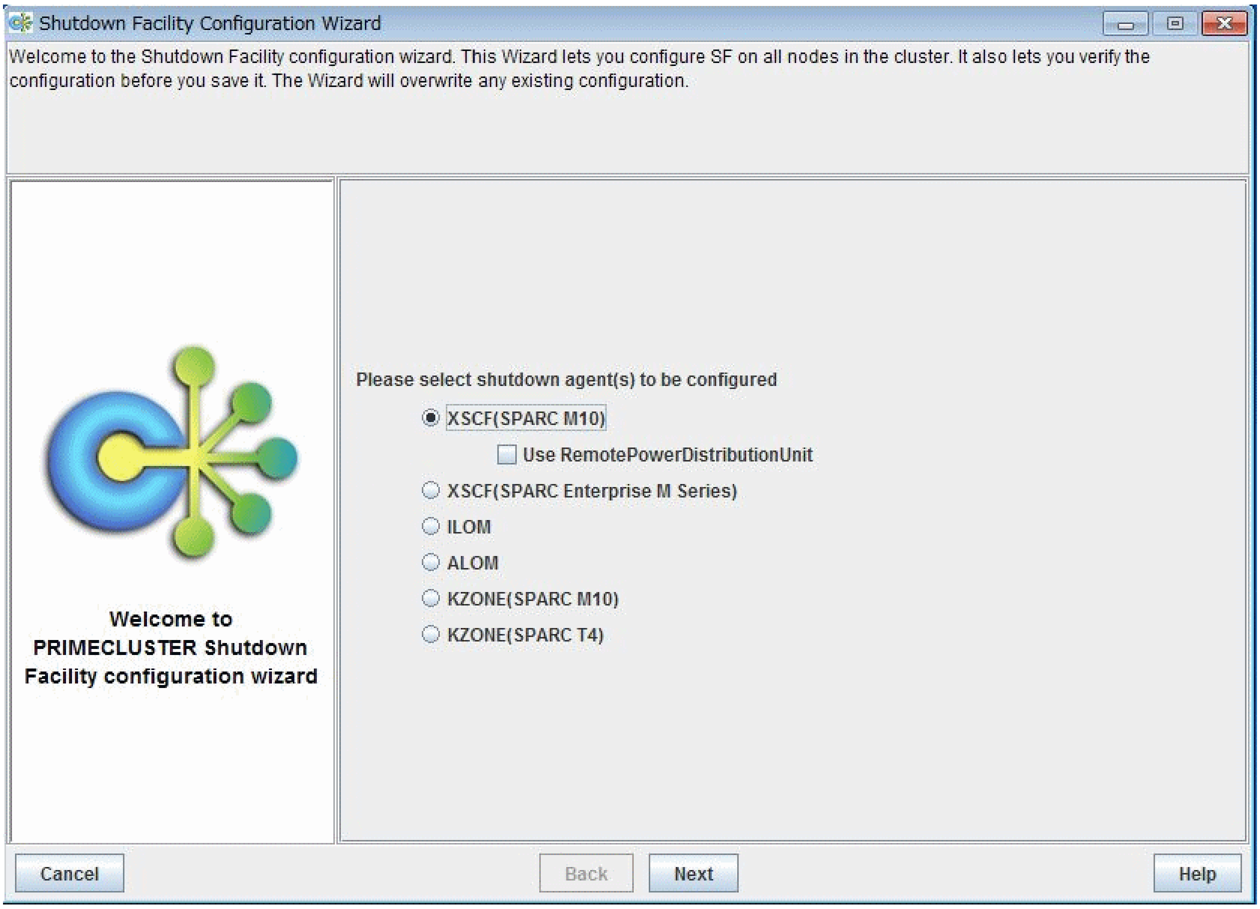
You can also configure the shutdown facility immediately after you complete the CF configuration with the CF wizard.

The following confirmation popup screen will appear. Click *Yes* to start the shutdown configuration wizard.



Selecting a shutdown agent

The selection screen for the shutdown agent will appear.



Confirm the hardware machine type and select the appropriate shutdown agent.

- For SPARC M10

  Select *XSCF*(*SPARC M10*).

  The following shutdown agents are automatically set.

  - Cluster systems between control domains:
    - XSCF SNMP Panic XSCF-LAN#0(Domain)
    - XSCF SNMP Panic XSCF-LAN#1(Domain)
    - XSCF SNMP Reset XSCF-LAN#0(Domain)
    - XSCF SNMP Reset XSCF-LAN#1(Domain)
    - XSCF SNMP Reset XSCF-LAN#0(PPAR)
    - XSCF SNMP Reset XSCF-LAN#1(PPAR)
  - Cluster systems between guest domains:
    - XSCF SNMP Panic XSCF-LAN#0(Domain)
    - XSCF SNMP Panic XSCF-LAN#1(Domain)
    - XSCF SNMP Reset XSCF-LAN#0(Domain)
    - XSCF SNMP Reset XSCF-LAN#1(Domain)

  When using the Remote Power Distribution Unit, check the "Use RemotePowerDistributionUnit" box.

  The following shutdown agent is automatically set:

  - RPDU Reset

After selection, click *Next*.

Information

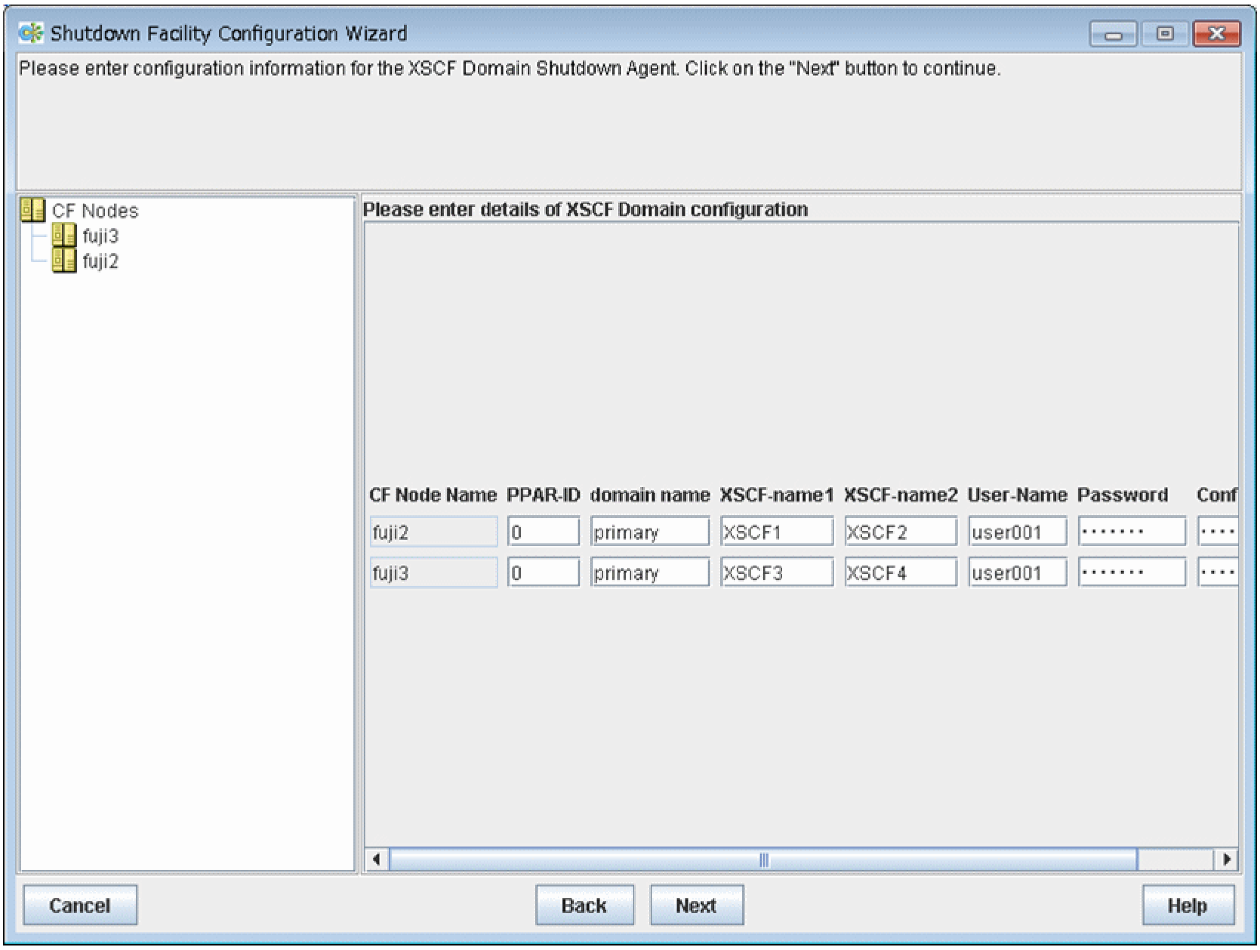
If you select a shutdown agent, the timeout value is automatically set. For details on the timeout value of the shutdown agent, see "8.2.2 Configuration file of SF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

- For XSCF Domain Panic/XSCF Domain Reset/XSCF PPAR Reset/RPDU Reset

  Timeout value = 20 (seconds)

Configuring XSCF

The screen for entering the information of XSCF will appear.



Enter the settings for XSCF that you recorded in "5.1.2.1.1 Checking XSCF Information".

PPAR-ID

Specify the identification ID of the physical partition (PPAR) in which the logical domain of the cluster node belongs to.

Make sure to enter "0" for SPARC M10-1 and M10-4.

For SPARC M10-4S, specify the integer ranged from 0 to 15.

Domain-name

Specify the logical domain name of the cluster node.

During initial setup, the logical domain name acquired from each node will be displayed as the initial value.

When changing the settings, the previous set value will be displayed on the screen.

Check that the logical domain name displayed is correct.

Change the logical domain name if it is wrong.

Execute the virtinfo -a command on each node and enter the logical domain name displayed.

Specify the input character string using up to 255 characters starting with an alphabetic letter and consisting of only alphanumeric characters and "-" (hyphens) and "." (period).

XSCF-name1

Specify the host name or IP address for XSCF-LAN#0 of the cabinet in which the logical domain of the cluster node exists.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

XSCF-name2

Specify the host name or IP address for XSCF-LAN#1 of the cabinet in which the logical domain of the cluster node exists.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

User-Name

Enter the user name to log in to the XSCF of the cabinet where the logical address of the cluster node exists.

Password

Enter the password to log in to the XSCF of the cabinet where the logical address of the cluster node exists.

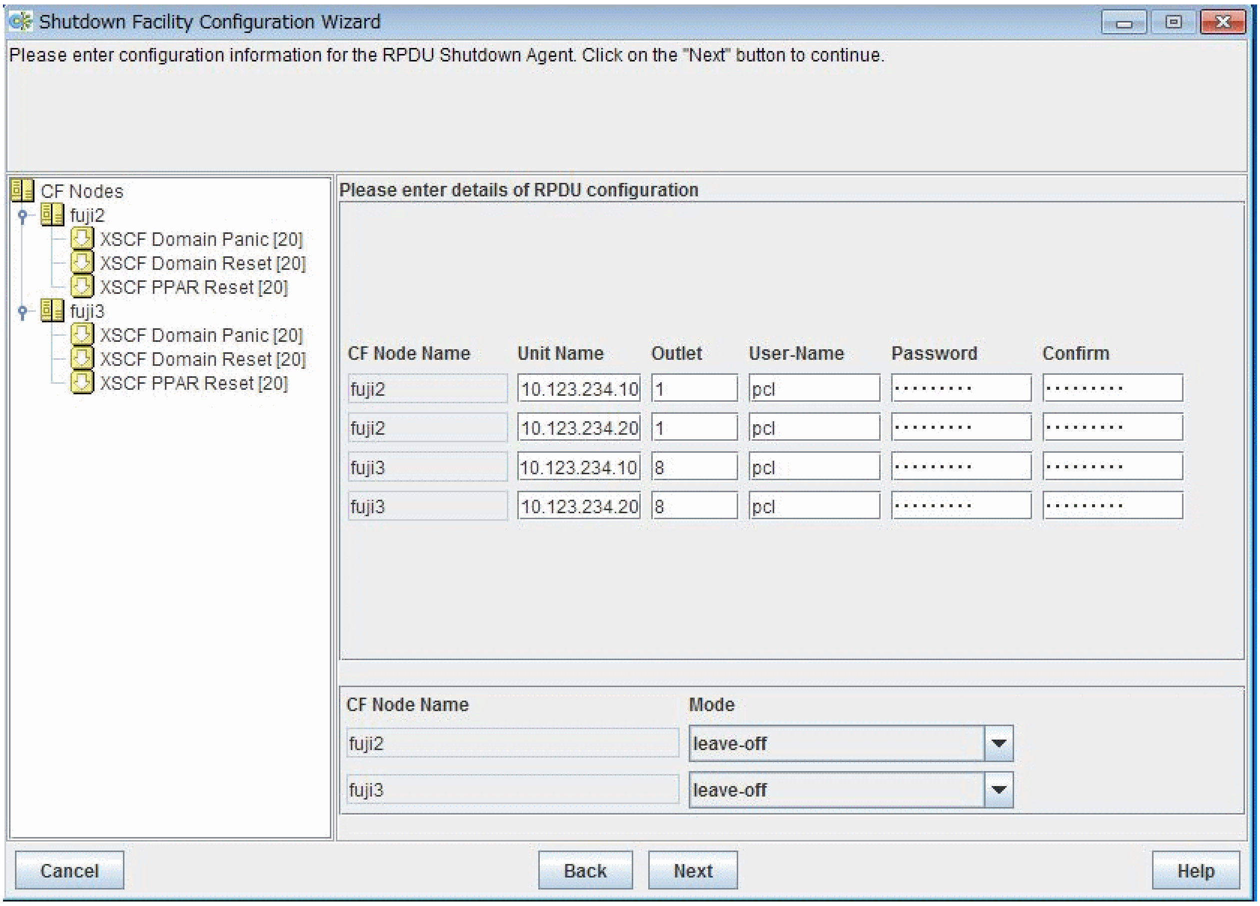
Note

- In the environment where XSCF is duplexed, a combination of a user name and a password for 2 of the XSCF must be the same.
- To use the Migration function, set a combination of a user name and password for the XSCF and the connection method to the XSCF to be consistent on all nodes.
- In the configuration where the asynchronous monitoring sub-LAN is not used, specify the IP address of XSCF-LAN#0 to "XSCF-name1" and the host name corresponding to the IP address specified in XSCF- name1 to "XSCF-name2." For the configuration where the asynchronous monitoring sub-LAN is not used, see "2.2.2 XSCF Configuration in SPARC M10."

Upon the completion of configuration, click *Next*.

Setting the Remote Power Distribution Unit

If you check the "Use RemotePowerDistributionUnit" box for selecting a shutdown agent, the settings for the Remote Power Distribution Unit is displayed.



Unit Name

Specify the IP address for the Remote Power Distribution Unit to which the main power and redundant power of a cluster node are connected.

Available IP addresses are IPv4 addresses.

Outlet

Specify the outlet number (1 to 8) of the Remote Power Distribution Unit to which the main power and redundant power of a cluster node are connected.

User-Name

Specify the user for the shutdown facility to log into the Remote Power Distribution Unit that was created in "J.2.3 Creating a User for Shutdown Facility."

Enter "pcl" here.

Password

Enter a password to log in to the Remote Power Distribution Unit.

Mode

Specifies a behavior after stopping a node forcibly.

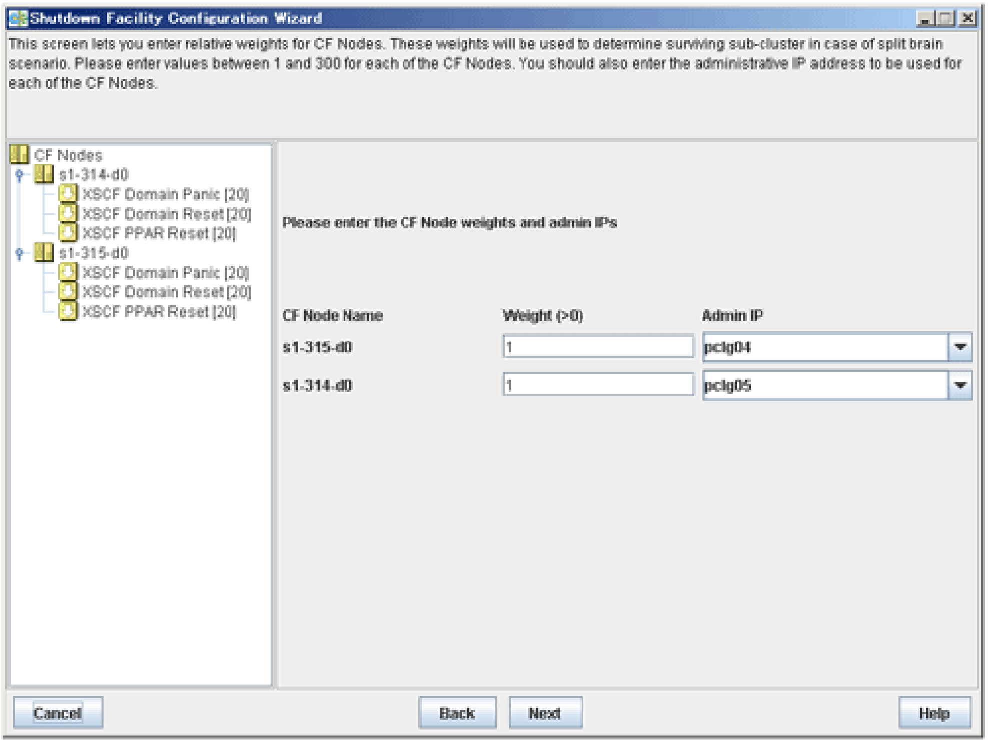
Select either "cycle" or "leave-off."

When "cycle" was selected, stop the node forcibly, and then reboot it.

When "leave-off" was selected, stop the node forcibly, and then cut its power.

Entering node weights and administrative IP addresses

The screen for entering the weights of the nodes and the IP addresses for the administrative LAN will appear.



Enter the weights of the nodes and the IP addresses for the administrative LAN.

Weight

Enter the weight of the node that constitutes the cluster. Weight is used to identify the survival priority of the node group that constitutes the cluster. Possible values for each node range from 1 to 300.
For details on survival priority and weight, refer to the explanations below.

Admin IP

Enter an IP address directly or click the tab to select the host name that is assigned to the administrative IP address.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

Upon the completion of configuration, click *Next*.

Survival priority

Even if a cluster partition occurs due to a failure in the cluster interconnect, all the nodes will still be able to access the user resources. For details on the cluster partition, see "1.2.2.1 Protecting data integrity" in "PRIMECLUSTER Concepts Guide".
To guarantee the consistency of the data constituting user resources, you have to determine the node groups to survive and those that are to be forcibly stopped.
The weight assigned to each node group is referred to as a "Survival priority" under PRIMECLUSTER.
The greater the weight of the node, the higher the survival priority. Conversely, the less the weight of the node, the lower the survival priority. If multiple node groups have the same survival priority, the node group that includes a node with the name that is first in alphabetical order will survive.

Survival priority can be calculated based on the following formula:

Survival priority = SF node weight + ShutdownPriority of userApplication

Note

When SF calculates the survival priority, each node will send its survival priority to the remote node via the administrative LAN. If any communication problem of the administrative LAN occurs, the survival priority will not be able to reach. In this case, the survival priority will be calculated only by the SF node weight.

SF node weight (Weight):

Weight of node. Default value = 1. Set this value while configuring the shutdown facility.

userApplication ShutdownPriority:

Set this attribute when userApplication is created. For details on how to change the settings, see "11.1 Changing the Operation Attributes of a Cluster Application".

See

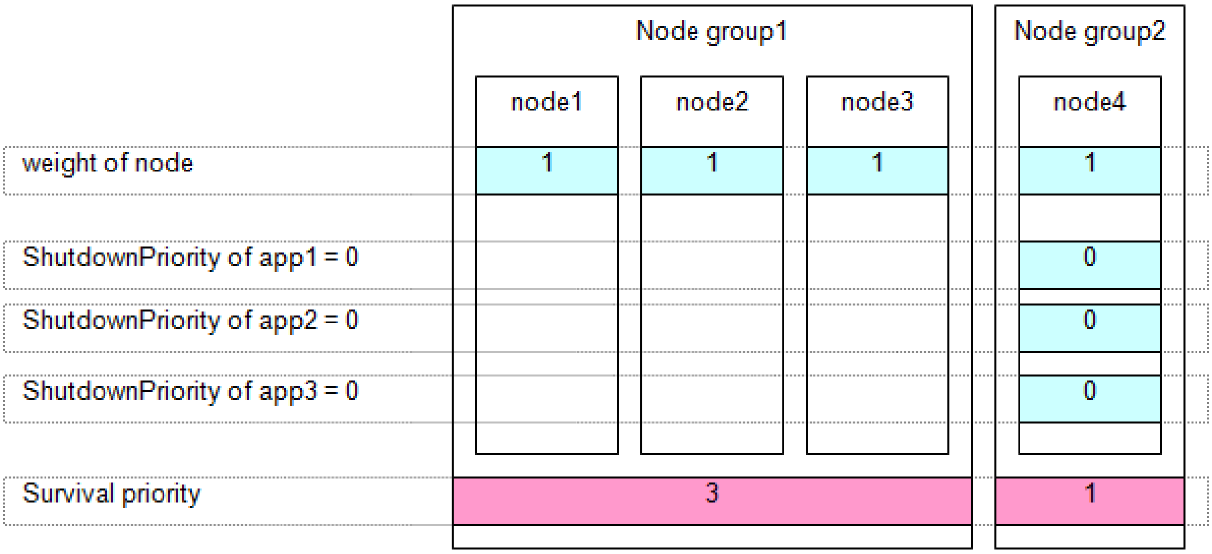
For details on the ShutdownPriority attribute of userApplication, see "6.7.5 Attributes".

Survival scenarios

The typical scenarios that are implemented are shown below:

[Largest node group survival]

- Set the weight of all nodes to 1 (default).
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

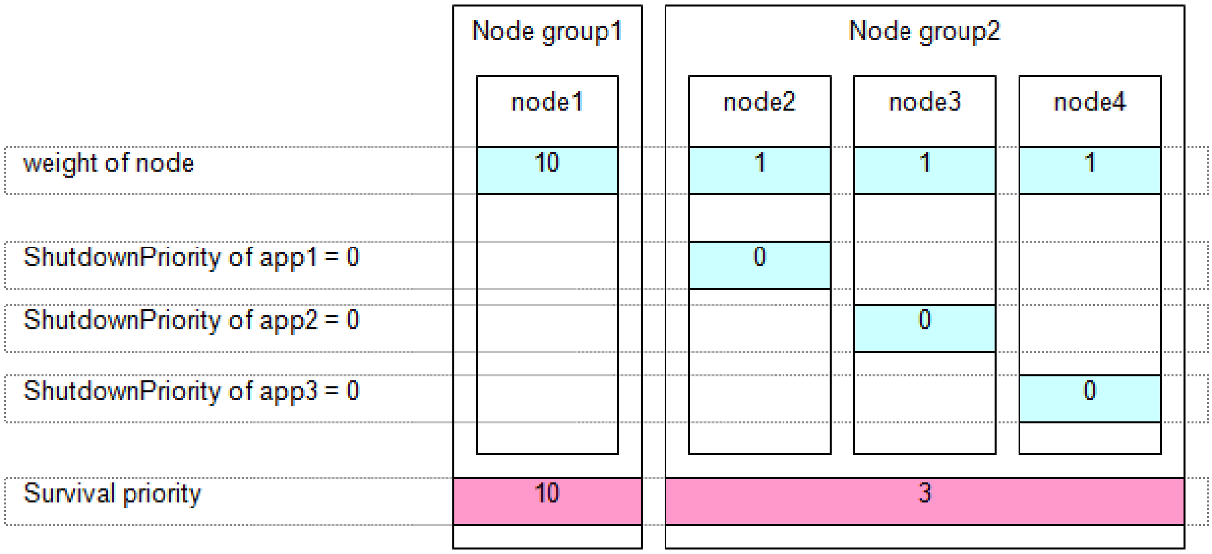


[Specific node survival]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes.

- Set the ShutdownPriority attribute of all user applications to 0 (default).

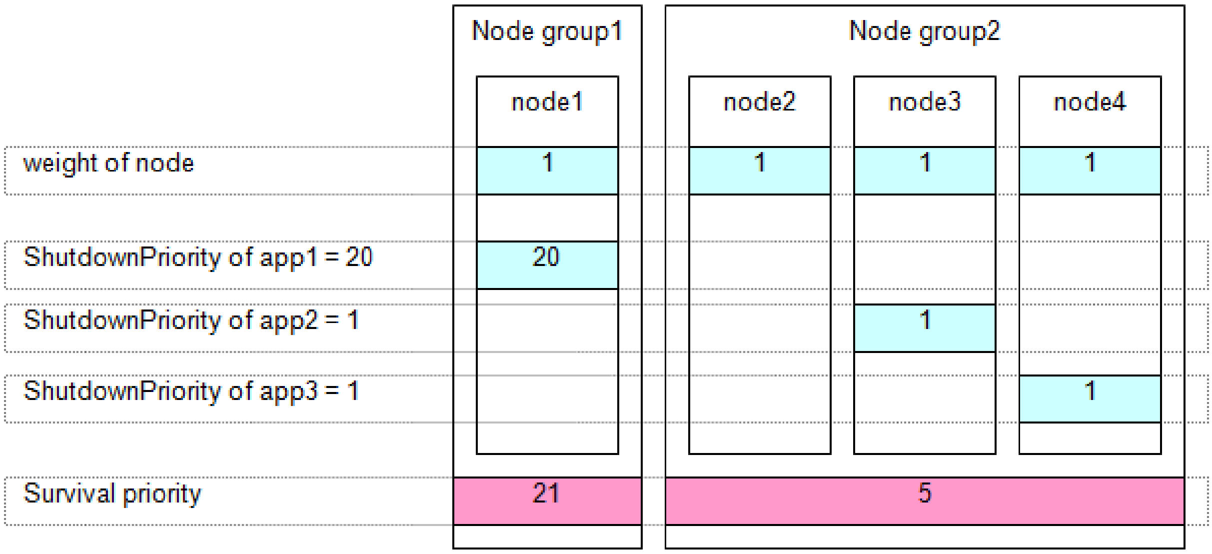
  In the following example, node1 is to survive:



[Specific application survival]

- Set the "weight" of all nodes to 1 (default).
- Set the ShutdownPriority attribute of the user application whose operation is to continue to a value more than double the total of the ShutdownPriority attributes of the other user applications and the weights of all nodes.

  In the following example, the node for which app1 is operating is to survive:



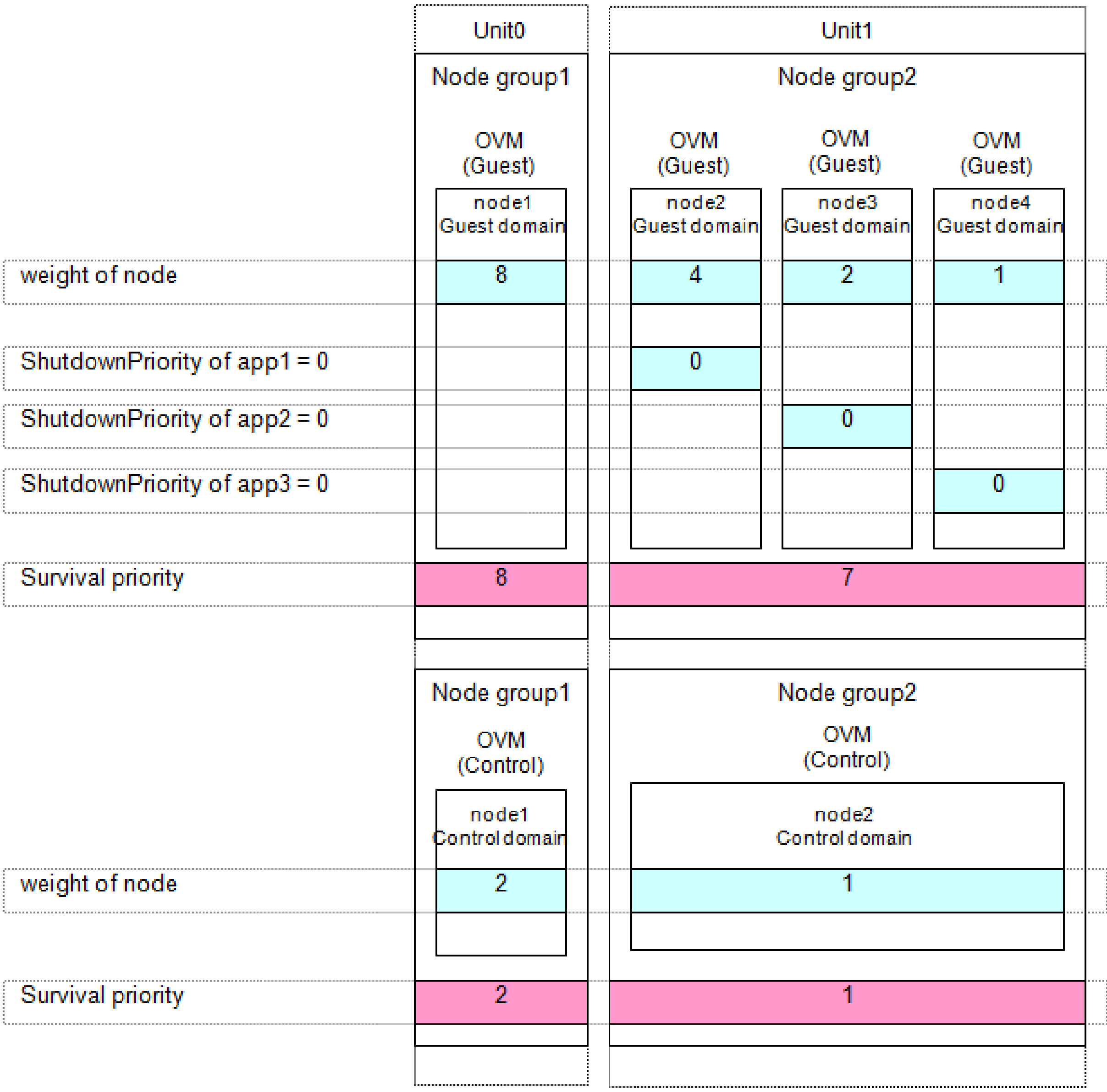
[Combination of the cluster system between control domains and the cluster system between guest domains for specific control domain survival (recommended)]

- Set the "weight" the nodes to a power of 2 (1,2,4,8,16,...) in ascending order of the survival priority on each cluster system..
- The order relation of "weight" set for guest domains must be the same as the corresponding control domains.

  For example, if the survival priority of host1 is higher than that of host2 between control domains, the survival priority of node1 (corresponding to host1) must be higher than those of node2 to 4 (corresponding to host2) between guest domains.

- Set the ShutdownPriority attribute of all user applications to 0 (default).

  In the following example, nodes are to survive in the order of node1, node2, node3, and node4.



[Combination of the cluster system between control domains and the cluster system between guest domains for the largest control domain survival]

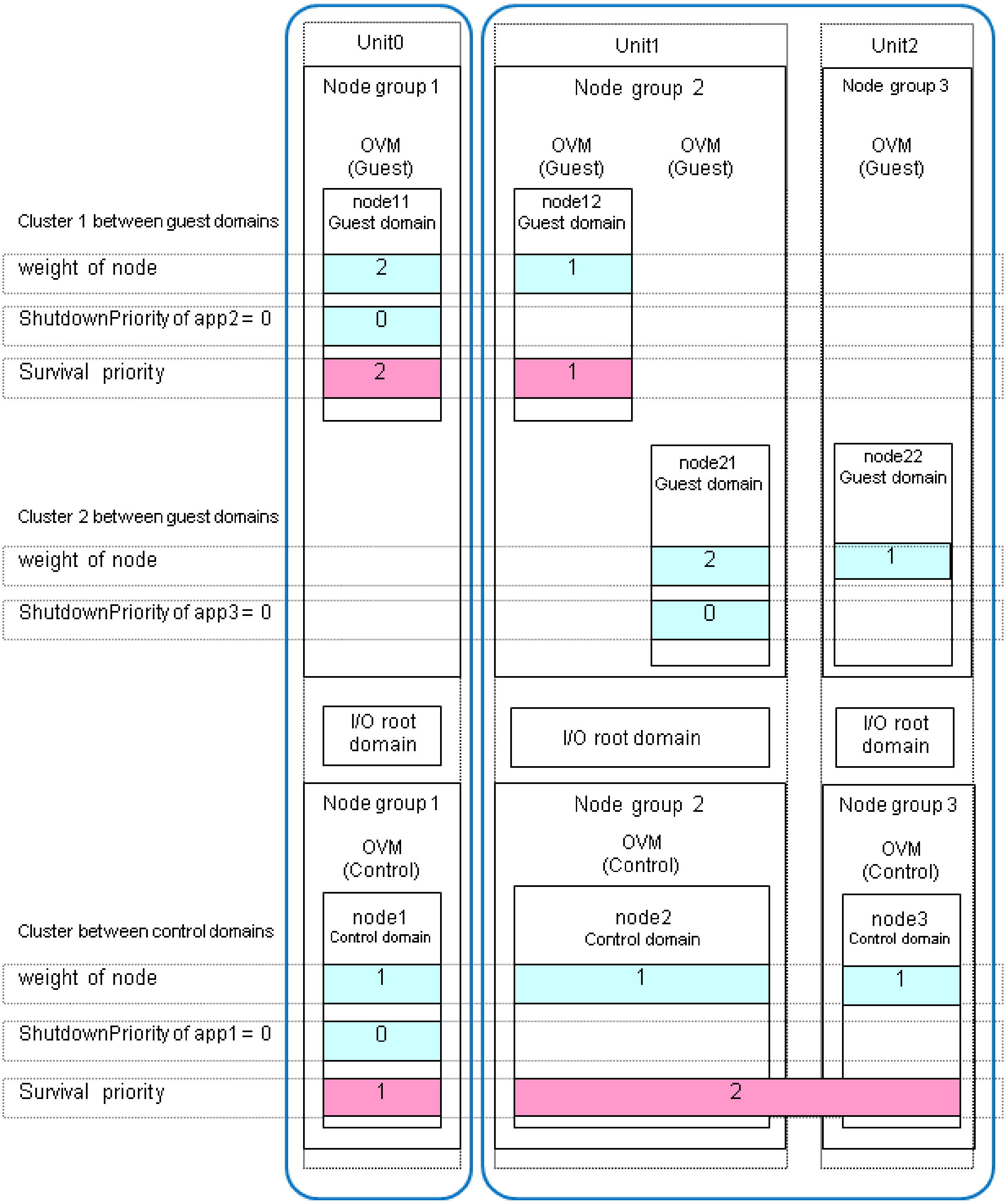
Note

- If the physical partition is reset, note that operations in the cluster system between guest domains may stop.
- Create the I/O root domain for this setting.

- Set the node weight of the control domain to 1 (default).
- Set the ShutdownPriority attribute of the user application of the control domain to 0 (default).
- Set either "Specific node survival" or "Specific application survival" for the node of the guest domain.

In the following example, "Specific node survival" is set for the guest domain.

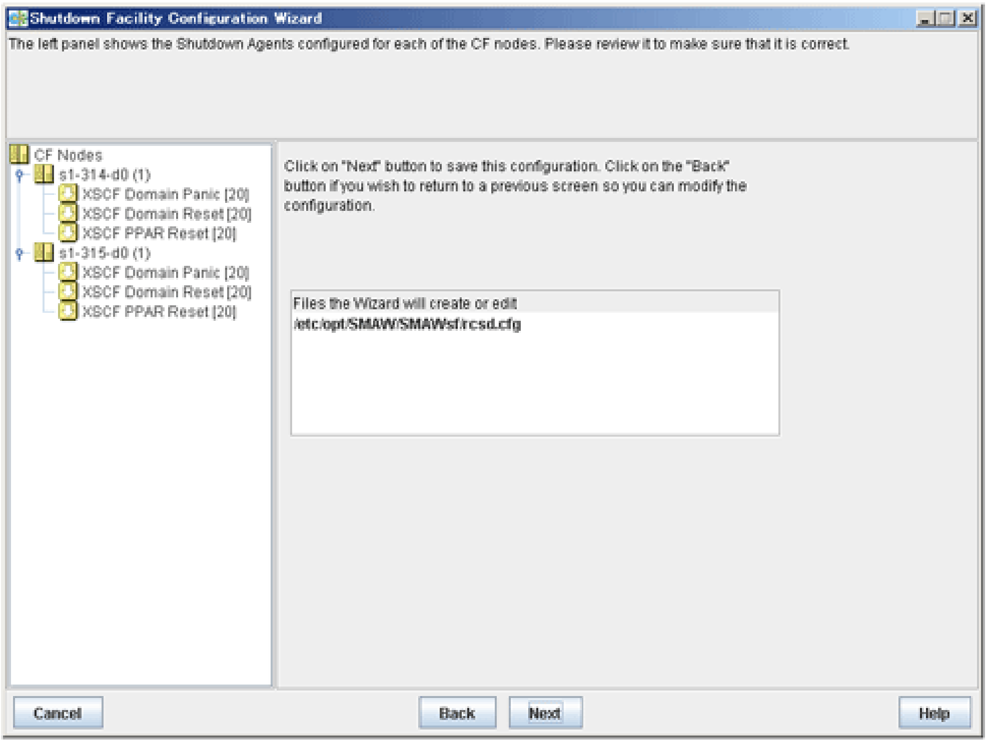
In this case, in the cluster 1 between guest domains, node 11 is saved as a survival node and node 12 is forcibly stopped while node 2 and node 3 are saved as survival nodes and node 1 is forcibly stopped in the cluster between control domains. If the physical partition of unit 0 is reset, note that operations in the cluster 1 between guest domains will stop.



Saving the configuration

Confirm and then save the configuration.

In the left-hand panel of the window, those nodes that constitute the cluster are displayed (SF node weight is displayed in brackets after those nodes), as are the shutdown agents that are configured for each node.



Click *Next*. A popup screen will appear for confirmation.

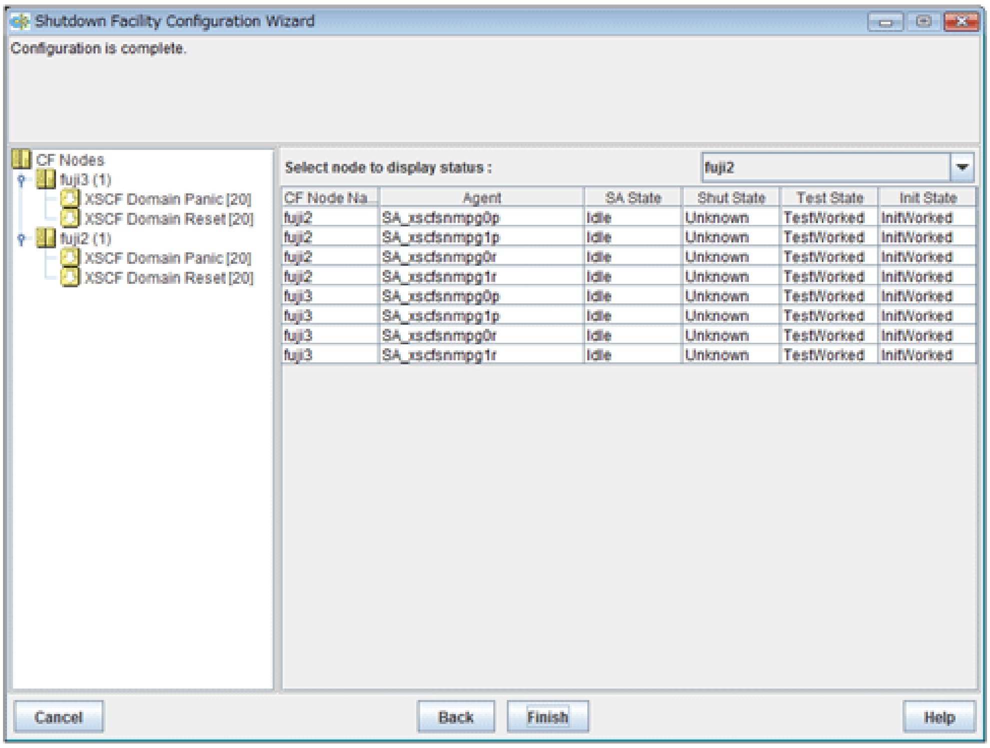
Select *Yes* to save the setting.

Displaying the configuration of the shutdown facility

If you save the setting, a screen displaying the configuration of the shutdown facility will appear. On this screen, you can confirm the configuration of the shutdown facility on each node by selecting each node in turn.

Information

You can also view the configuration of the shutdown facility by selecting *Shutdown Facility -> Show Status* from the *Tool* menu.



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

- If "TestFailed" is displayed in the test state, the configuration information of the logical domains may not be saved. Use the ldm add-spconfig command to save the information when it is not saved.

- If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the agent or hardware configuration may contain an error. Check the /var/adm/messages file and the console for an error message. Then, apply appropriate countermeasures as instructed the message that is output.

- If the connection to XSCF from the shutdown facility is telnet and also the setting of the connection to XSCF is SSH, the test state becomes TestFailed at this point in time. Confirm that the shutdown facility is operating normally, after performing the "5.1.2.1.4 Setting of the connection method to the XSCF".

See

For details on how to respond to the error messages that may be output, see "PRIMECLUSTER Messages."

Checking if the SNMP trap is received

Check if the SNMP trap is received in all XSCFs that constitute a cluster system.

1. Execute the following command in XSCF.

   The pseudo error notification trap is issued.

```
XSCF> rastest -c test
```

2. For two IP addresses for the management LAN and the asynchronous monitoring sub-LAN of XSCF, check that the pseudo error notification trap is output to the /var/adm/messages file on all nodes. For the environment where the asynchronous monitoring sub-LAN is not used, the pseudo error notification trap is output only for the IP address for the management LAN. If the output message contains the following words "FF020001" and "M10-Testalert," it means that the SNMP trap has been received successfully.

   Example: When the IP address of XSCF is "192.168.10.10"

```
snmptrapd[Process ID]: [ID 702911 daemon.warning] 192.168.10.10 [192.168.10.10]: Trap
DISMAN-EVENT-MIB::sysUpTimeInstance = Timeticks: (3557) 0:00:35.57, SNMPv2-MIB::snmpTrapOID.0 =
OID: SNMPv2-SMI::enterprises.211.1.15.4.1.2.0.1,
SNMPv2-SMI::enterprises.211.1.15.4.1.1.12.2.1.13.100.0.254.0.254.0 = INTEGER: 3,
SNMPv2-SMI::enterprises.211.1.15.4.1.2.1.2.0 = INTEGER: 1,
SNMPv2-SMI::enterprises.211.1.15.4.1.1.4.3.0 = STRING: "PZ31426053",
SNMPv2-SMI::enterprises.211.1.15.4.1.1.4.2.0 = STRING: "SPARC M10-1",
SNMPv2-SMI::enterprises.211.1.15.4.1.1.4.1.0 = "", SNMPv2-SMI::enterprises.
211.1.15.4.1.2.1.14.0
= STRING: "FF020001", SNMPv2-SMI::enterprises.211.1.15.4.1.2.1.15.0 = STRING: "Oct 27
10:54:34.288
JST 2014", SNMPv2-SMI::enterprises.211.1.15.4.1.2.1.16.0 =
STRING: "https://support.oracle.com/msg/M10-Testalert
<https://support.oracle.com/msg/M10-Testalert>", SNMPv2-SMI::enterprises.211.1.15.4.1.2.1.17.0
=
STRING: "TZ1422A010  ", SNMPv2-SMI::enterprises.211.1.15.4.1.2.1.18.0 = STRING: "CA07363-D011
```

   If the pseudo error notification trap is not output, there may be an error in the SNMP settings. See the "5.1.2.1.2 Setting SNMP" to correct the settings.

   See

   For the content of the pseudo error notification trap, see "Fujitsu M10/SPARC M10 Systems XSCF MIB and Trap Lists."

5.1.2.1.4 Setting of the connection method to the XSCF

The default of setting of the connection method to the XSCF is SSH connection for SPARC M10.

The procedure when changing to the telnet connection is the following.

Change of the connection method

Execute the following command in all nodes to change a connection method.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -m -t telnet
```

After changing the connection method, execute the clsnmpsetup -l command to check that "telnet" is displayed in the "connection-type" field.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -l
device-name cluster-host-name PPAR-ID domain-name IP-address1 IP-address2 user-name connection-type
-------------------------------------------------------------------------------------------------
xscf        node1             1       primary     xscf11      xscf12      xuser     telnet
xscf        node2             2       primary     xscf21      xscf22      xuser     telnet
```

Note

To use the Migration function, set a combination of a user name and password for the XSCF and the connection method to the XSCF to be consistent on all nodes.

Starting up the shutdown facility

Execute the following command in each node, and confirm the shutdown facility has started.

```
# /opt/SMAW/bin/sdtool -s
```

If the state of configuration of shutdown facility is displayed, shutdown facility is started.

If "The RCSD is not running" is displayed, shutdown facility is not started.

If shutdown facility is started, execute the following command, and restart the shutdown facility.

```
# /opt/SMAW/bin/sdtool -r
```

If shutdown facility is not started, execute the following command, and start the shutdown facility.

```
# /opt/SMAW/bin/sdtool -b
```

5.1.2.2 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000

5.1.2.2.1 Checking Console Configuration

In SPARC Enterprise M3000, M4000, M5000, M8000, and M9000, XSCF is used. The connection method to XSCF as the shutdown facility can be selected from SSH or the telnet.

Default connection is SSH.

Please confirm the following settings concerning XSCF before setting the shutdown facility.

- Commonness
    - The log in user account must be made excluding root for the shutdown facility, and the platadm authority must be given.
- At the SSH connection
    - In XSCF, SSH must be effective in connected permission protocol type from the outside.
    - User inquiries of the first SSH connection (such as generation of the RSA key) must be completed by connecting to XSCF from all the cluster nodes via SSH using the log in user account for the shutdown facility.

      When using the host name for specifying XSCF name, the first SSH connection with the specified host name must be completed.
- At the telnet connection
    - In XSCF, telnet must be effective in connected permission protocol type from the outside.

Note

When the connection to XSCF is a serial port connection alone, it is not supported in the shutdown facility. Please use XSCF-LAN.

Moreover, record the following information on XSCF.

- XSCF IP address or an XSCF host name registered in the "/etc/inet/hosts" file of the node
- Log in user account and password for shutdown facility in XSCF

  *1) When the network routing is set, IP address of XSCF need not be the same to management LAN segment of the cluster node.

See

For information on how to configure and confirm XSCF, see the "*XSCF User's Guide*".

5.1.2.2.2 Using the Shutdown Configuration Wizard

The required shutdown agent varies depending on the hardware machine type.

Check the following combinations of the hardware machine types and shutdown agents.

- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by companies other than Fujitsu in Japan, or
  SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 with logos of both Fujitsu and Oracle provided in other than Japan
  - XSCF Panic
  - XSCF Break
  - XSCF Reset
- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 (other than above) provided by Fujitsu in Japan
  - RCI Panic
  - XSCF Panic
  - XSCF Break
  - RCI Reset
  - XSCF Reset

Setting up the operation environment for the asynchronous RCI monitoring

This setting is required only for the following cases:

- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by Fujitsu in Japan

When you set up asynchronous RCI monitoring, you must specify the timeout interval (kernel parameter) in /etc/system for monitoring via SCF/RCI.

See

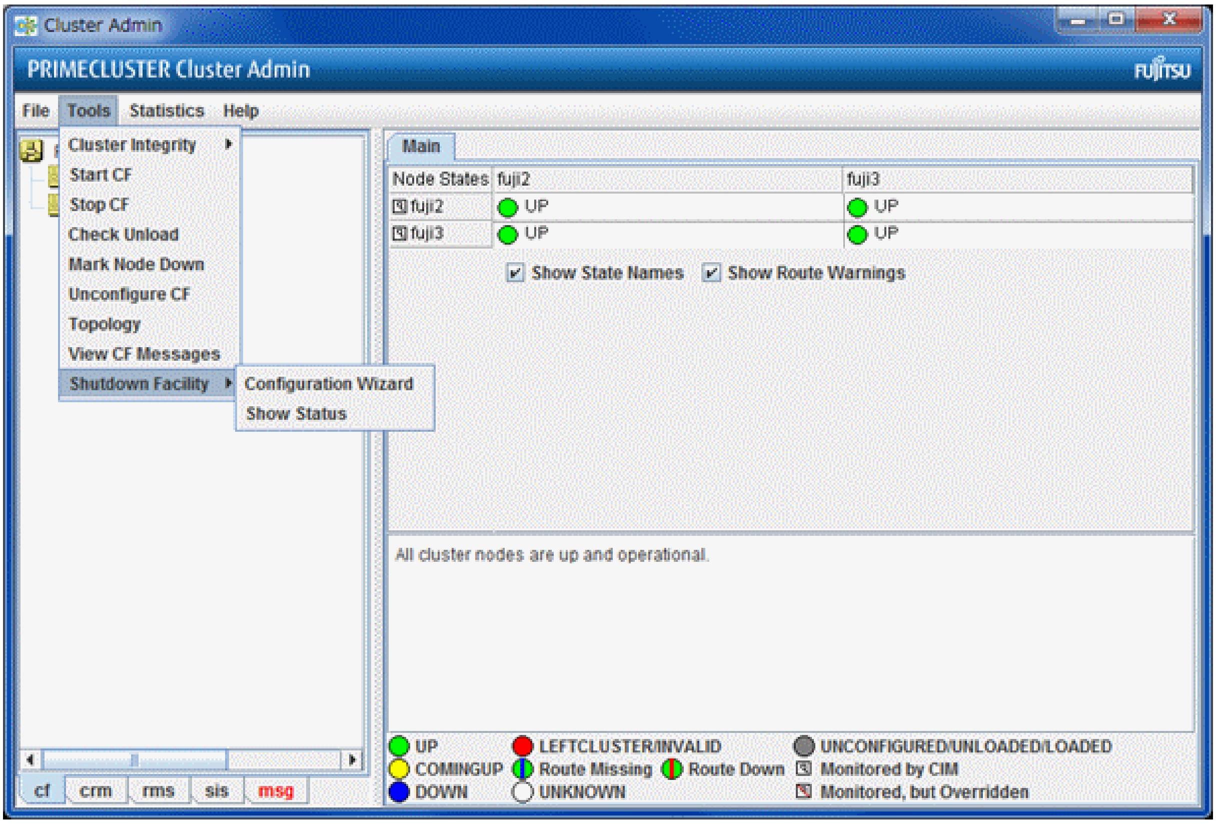
For kernel parameter settings, see "A.5.1 CF Configuration."

Note

You need to reboot the system to enable the changed value.

Starting up the shutdown configuration wizard

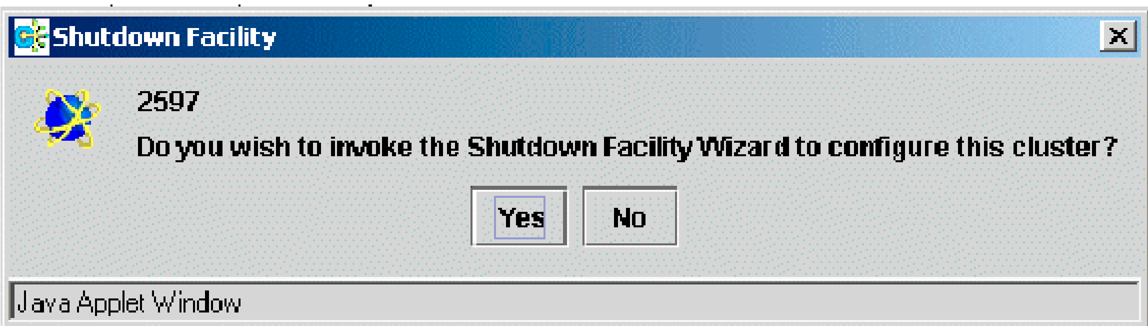
From the CF main window of the Cluster Admin screen, select the *Tool* menu and then *Shutdown Facility -> Configuration Wizard*. The shutdown configuration wizard will start.



Note

You can also configure the shutdown facility immediately after you complete the CF configuration with the CF wizard.

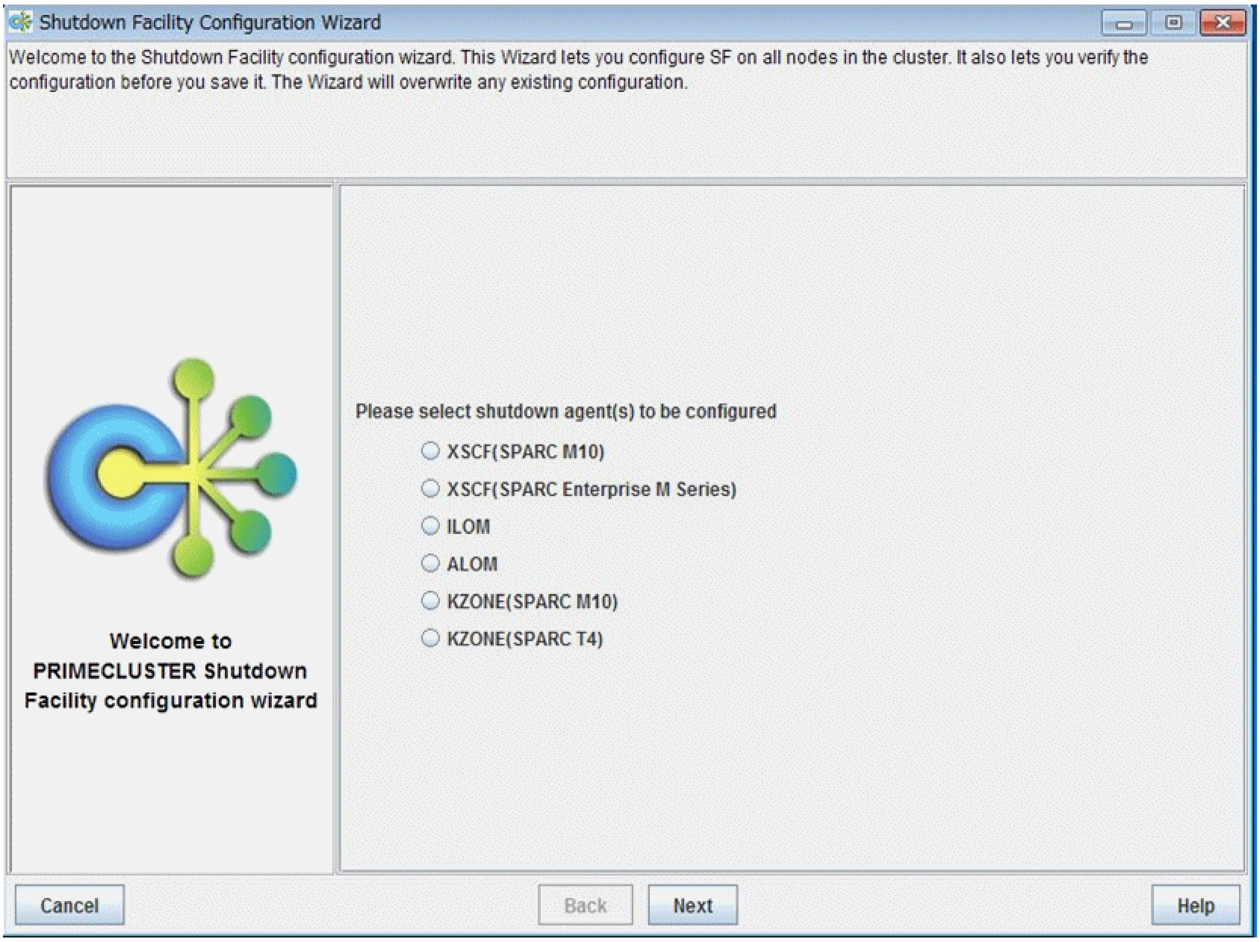
The following confirmation popup screen will appear. Click *Yes* to start the shutdown configuration wizard.



Selecting a shutdown agent

The selection screen for the shutdown agent will appear.

Figure 5.2 Selecting a shutdown agent

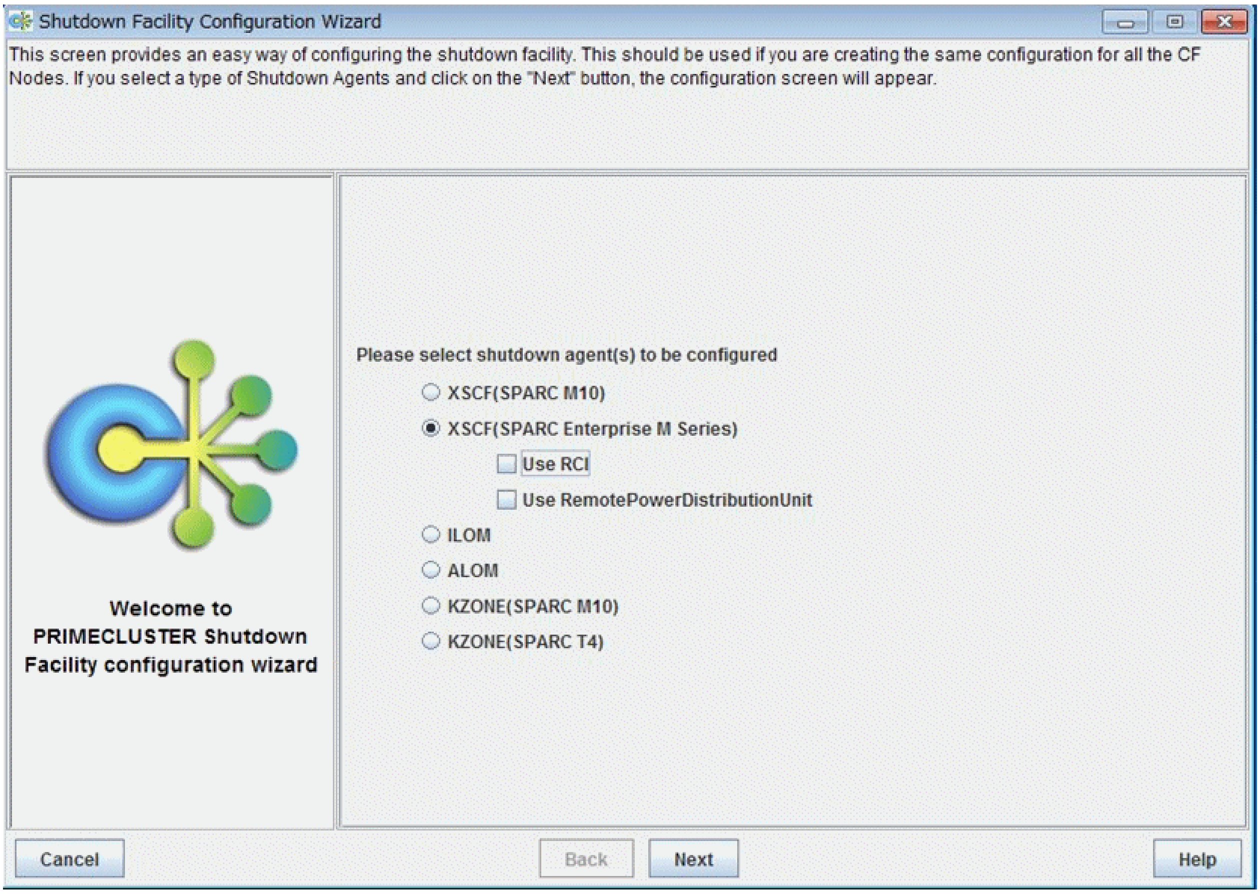


Confirm the hardware machine type and select the appropriate shutdown agent.

a. SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by companies other than Fujitsu in Japan or SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 with logos of both Fujitsu and Oracle provided in other than Japan

b. SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 (other than above) provided by Fujitsu in Japan

a) SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by companies other than Fujitsu in Japan or SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 with logos of both Fujitsu and Oracle provided in other than Japan



Select *XSCF* (*SPARC Enterprise M-series*).

If you select *XSCF* (*SPARC Enterprise M-series*), *Use RCI* is displayed. Clear the checkbox of *Use RCI*.

The following shutdown agents are automatically set:

- XSCF Panic
- XSCF Break
- XSCF Reset

When using the Remote Power Distribution Unit is used in the SPARC Enterprise M3000, M4000, or M5000 environment, check the "User RemotePowerDistributionUnit" box.

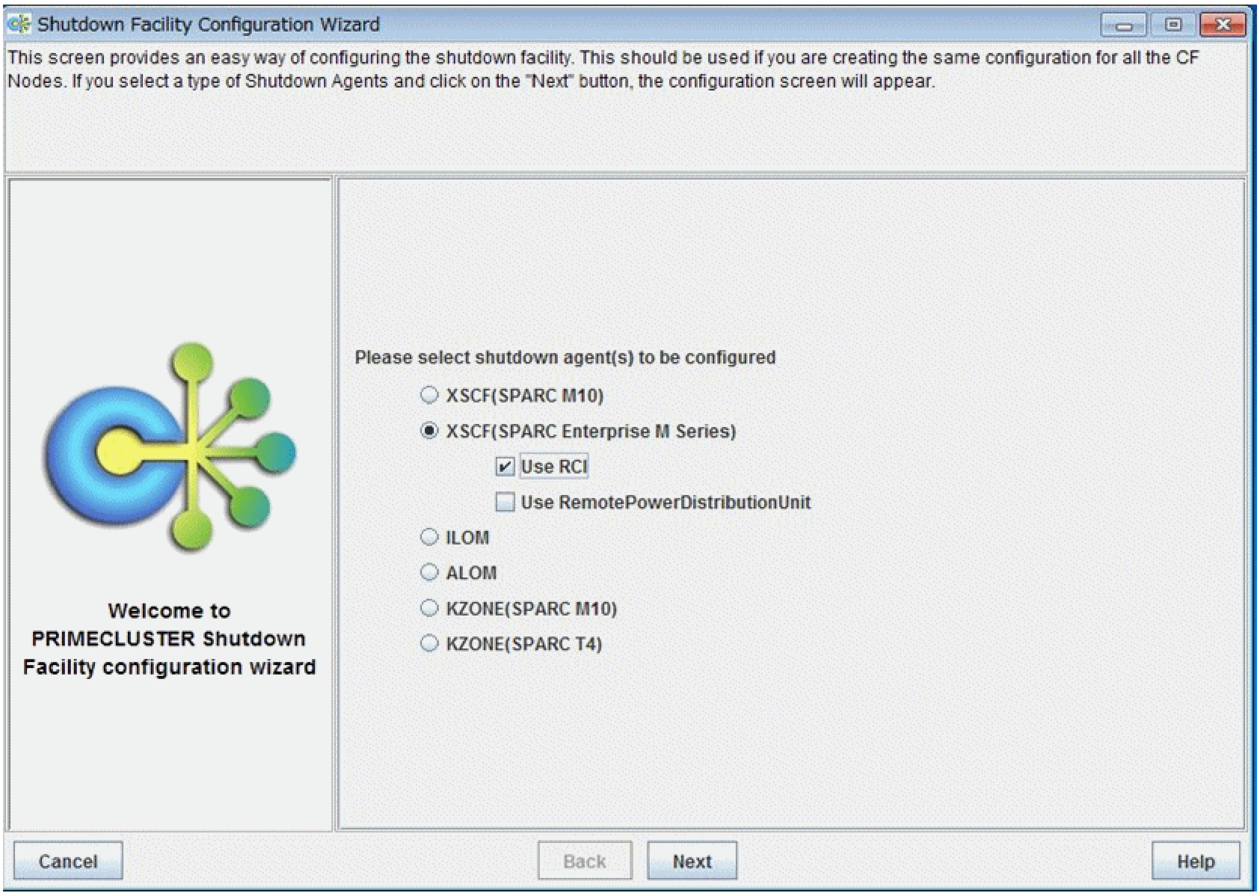
In addition, select "Use Model: M3000, 4000" or "Use Model: M5000" depending on the hardware machine type.

The following shutdown agent is automatically set:

- RPDU Reset

After selection, click *Next*.

b) SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 (other than above) provided by Fujitsu in Japan



Select *XSCF* (*SPARC Enterprise M-series*).

If you select *XSCF* (*SPARC Enterprise M-series*), *Use RCI* is displayed, however do not clear the checkbox of *Use RCI*.

The following shutdown agents are automatically set:

- RCI Panic
- XSCF Panic
- XSCF Break
- RCI Reset
- XSCF Reset

When using the Remote Power Distribution Unit is used in the SPARC Enterprise M3000, M4000, or M5000 environment, check the "User RemotePowerDistributionUnit" box.

In addition, select "Use Model: M3000, 4000" or "Use Model: M5000" depending on the hardware machine type.

The following shutdown agent is automatically set:

- RPDU Reset

After selection, click *Next*.

Information

If you select a shutdown agent, the timeout value is automatically set. For details on the timeout value of the shutdown agent, see "8.2.2 Configuration file of SF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

- For XSCF Panic/XSCF Break
    - 4 or fewer nodes

      Timeout value = 20 (seconds)
    - 5 or more nodes

      Timeout value = 6 x number of cluster nodes + 2 (seconds)

      Example for 5 nodes: 6 x 5 + 2 = 32 (seconds)
- For XSCF Reset
    - 4 or fewer nodes

      Timeout value = 40 (seconds)
    - 5 or more nodes

      Timeout value = 6 x number of cluster nodes + 22 (seconds)

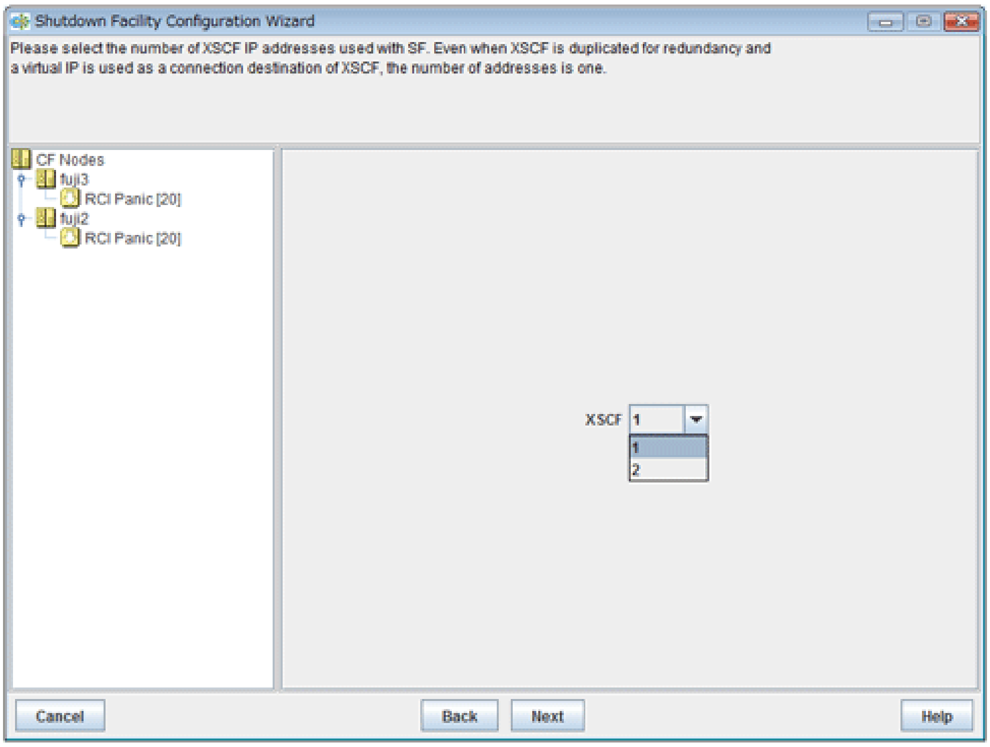
      Example for 5 nodes: 6 x 5 + 22 = 52 (seconds)
- For RCI Panic/RCI Reset/RPDU Reset

  Timeout value = 20 (seconds)

Configuring XSCF

The screen for entering the information of XSCF will appear.

Figure 5.3 Selecting the number of XSCF IP addresses



Select the number of XSCF IP addresses to use in the shutdown facility.

Note

If XSCF unit is duplexed but XSCF-LAN is not duplexed, the number of XSCF IP addresses is *1*.

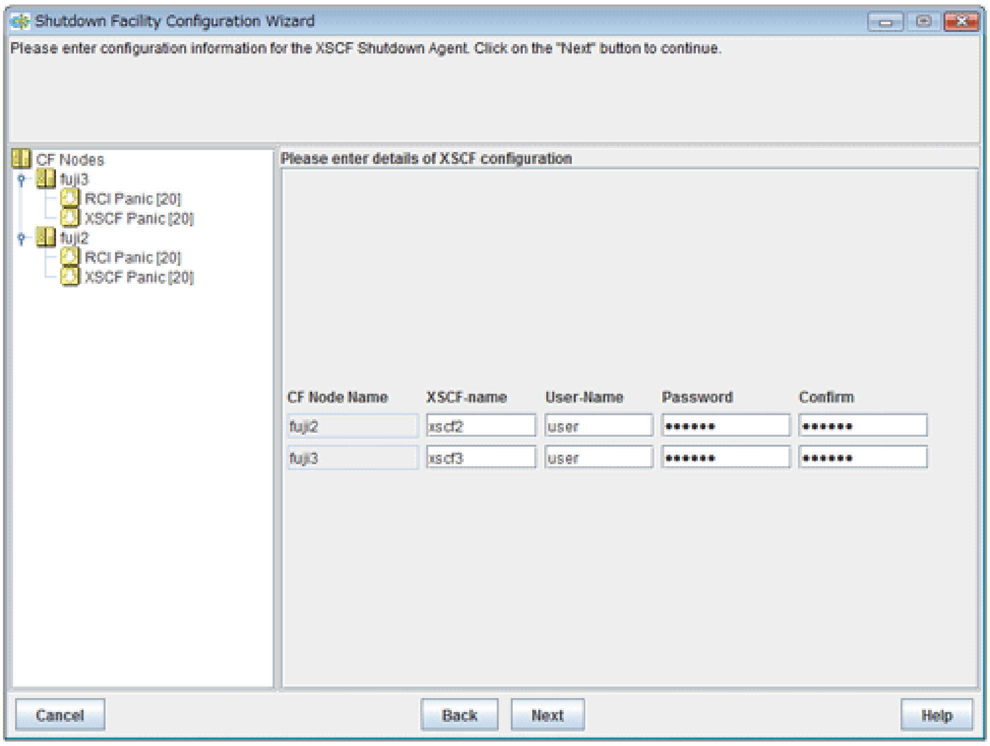
In this case, specify the virtual IP (takeover IP address) for the XSCF IP addresses.

Select the number of XSCF IP addresses, and click *Next*.

The screen to set the information of XSCF will appear.

"a) For selecting [1] for the number of XSCF IP addresses" and "b) For selecting [2] for the number of XSCF IP addresses" are respectively explained below.

a) For selecting [1] for the number of XSCF IP addresses



Enter the settings for XSCF that you recorded in "5.1.2.2.1 Checking Console Configuration".

XSCF-name

Enter the IP address of XSCF or the host name of XSCF that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

User-Name

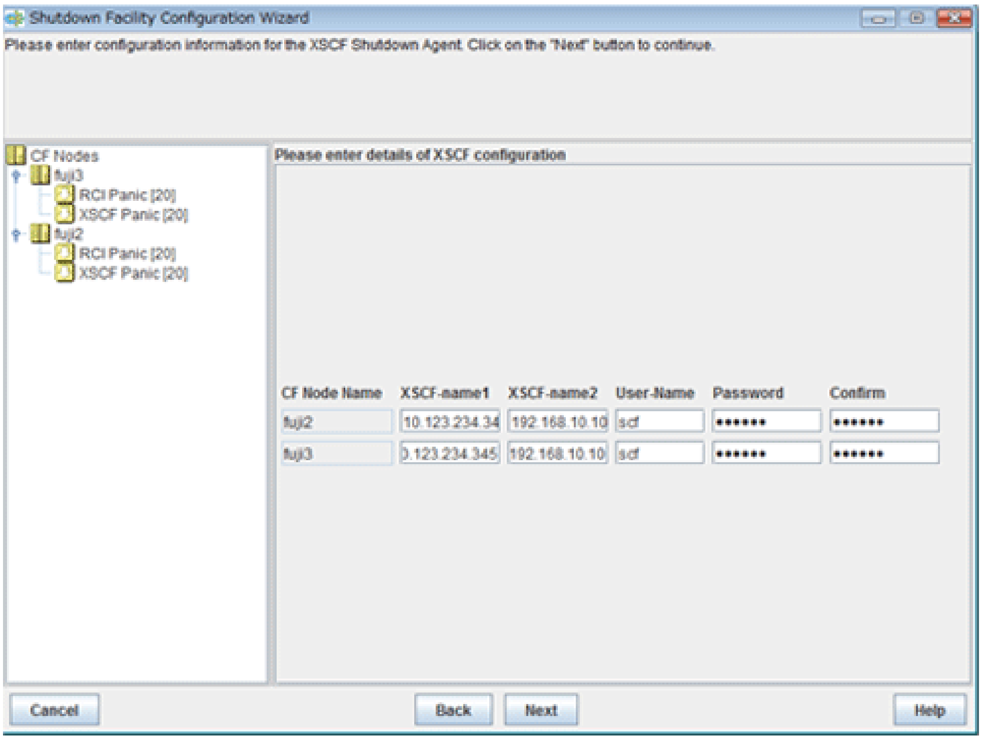
Enter a user name to log in to XSCF.

Password

Enter a password to log in to XSCF.

Upon the completion of configuration, click *Next*.

b) For selecting [2] for the number of XSCF IP addresses



Enter the settings for XSCF that you recorded in "5.1.2.2.1 Checking Console Configuration".

XSCF-name1

Enter the IP address of XSCF-LAN#0 or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

XSCF-name2

Enter the IP address of XSCF-LAN#1 or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

User-Name

Enter a user name to log in to XSCF.

Password

Enter a password to log in to XSCF.

Note

A combination of a user name and a password for 2 of the XSCF must be the same.

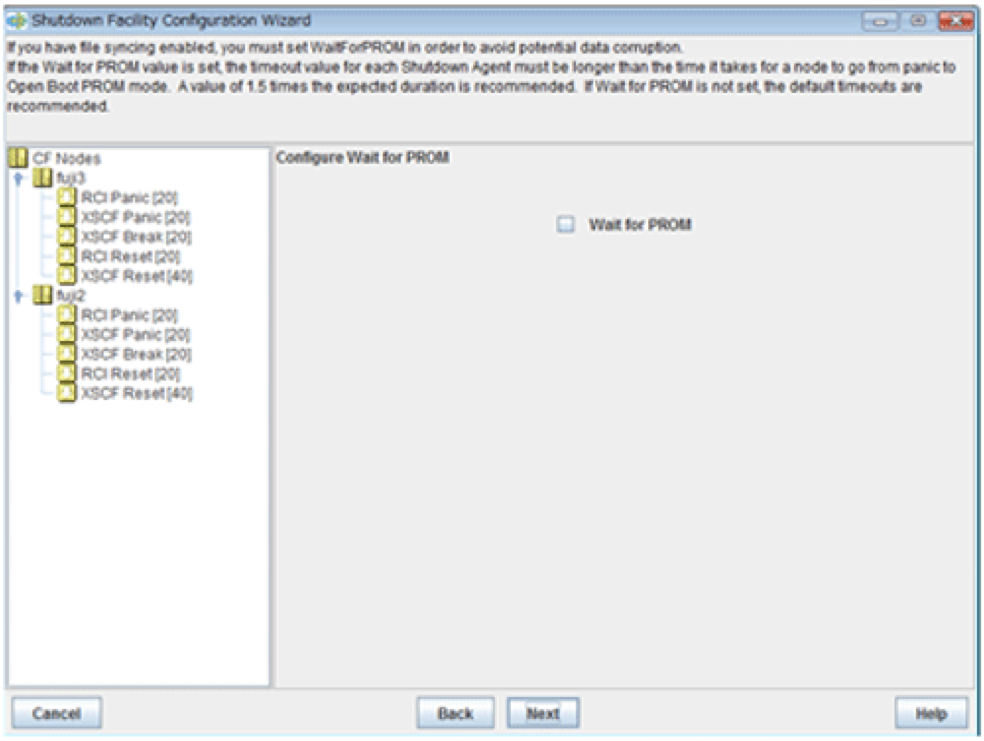
Upon the completion of configuration, click *Next*.

Configuring Wait for PROM

Note

Wait for PROM is currently not supported.
You do not have to select the checkbox, and then click *Next*.

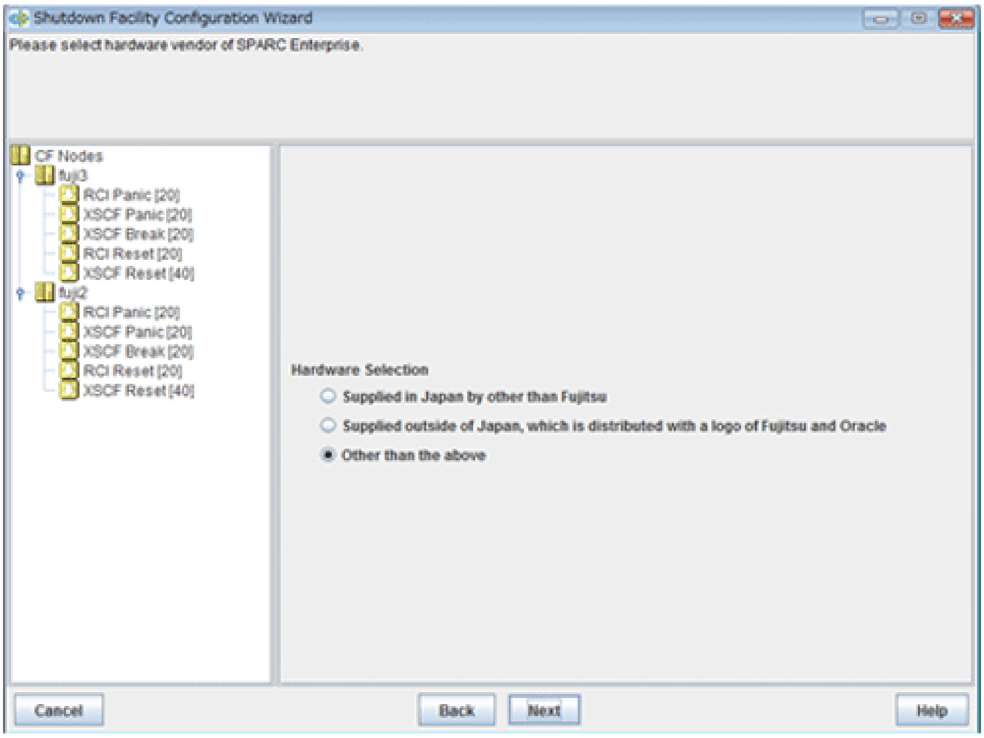
Figure 5.4 Configure Wait for PROM



Configuring hardware selection

If you select *XSCF* (*SPARC Enterprise M-series*) as the shutdown agent, the screen for selecting hardware will appear.

Figure 5.5 Configuring hardware selection



- For SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by companies other than Fujitsu in Japan

  Select "*Supplied in Japan by other than Fujitsu*".

- For SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 with logos of both Fujitsu and Oracle provided in other than Japan

  Select "*Supplied outside of Japan which is distributed with a logo of Fujitsu and Oracle*".

- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 (other than above) provided by Fujitsu in Japan

  Select "*Other than the above*".

Upon the completion of configuration, click *Next*.

Setting the Remote Power Distribution Unit

If you check the "Use RemotePowerDistributionUnit" box for selecting a shutdown agent, the settings for the Remote Power Distribution Unit is displayed.

Figure 5.6 For SPARC Enterprise M3000 and M4000

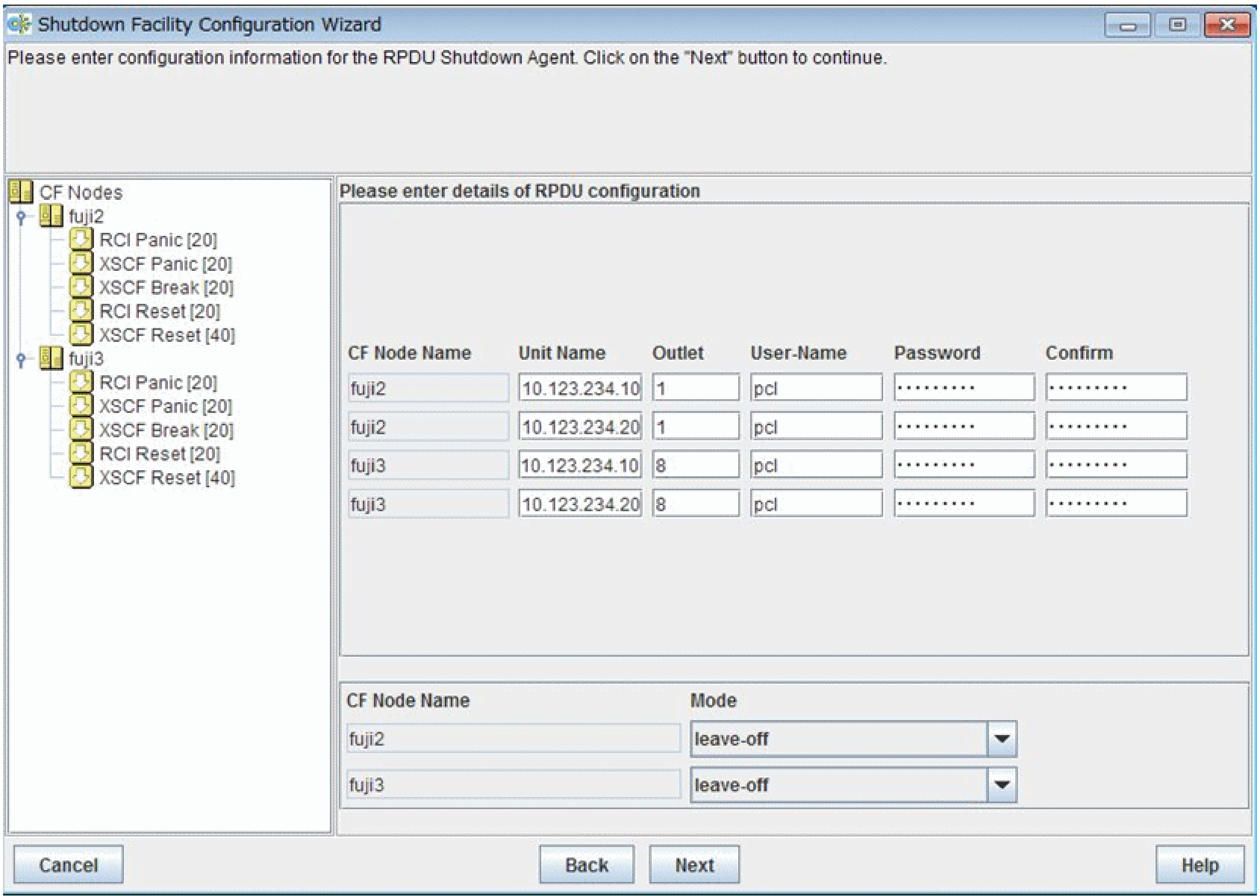
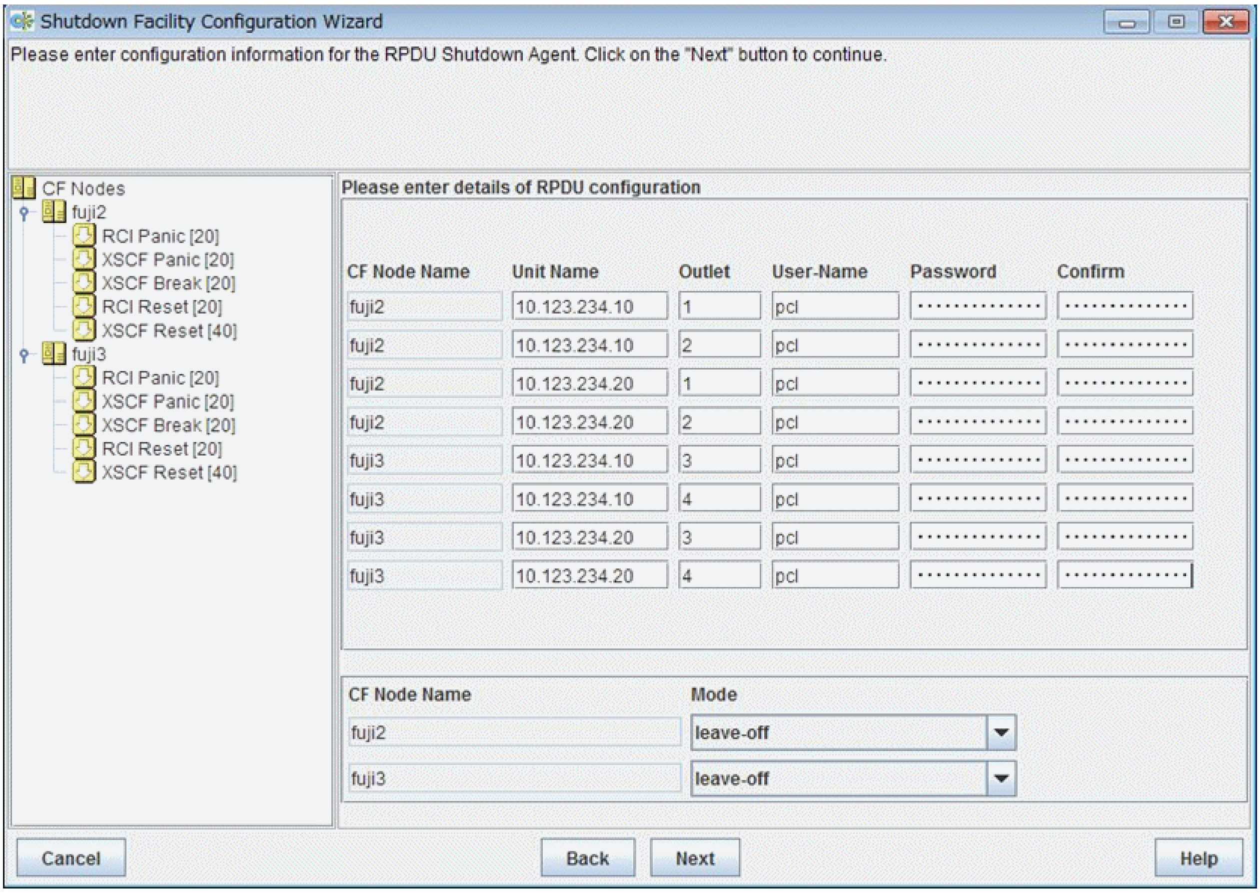


Figure 5.7 For SPARC Enterprise M5000



Unit-Name

Specify the IP address for the Remote Power Distribution Unit to which the main power and redundant power of a cluster node are connected.

Available IP addresses are IPv4 addresses.

Outlet

Specify the outlet number (1 to 8) of the Remote Power Distribution Unit to which the main power and redundant power of a cluster node are connected.

User-Name

Specify the user for the shutdown facility to log into the Remote Power Distribution Unit that was created in "J.2.3 Creating a User for Shutdown Facility."

Enter "pcl" here.

Password

Enter a password to log in to the Remote Power Distribution Unit.

Mode

Specifies a behavior after stopping a node forcibly.

Select either "cycle" or "leave-off."

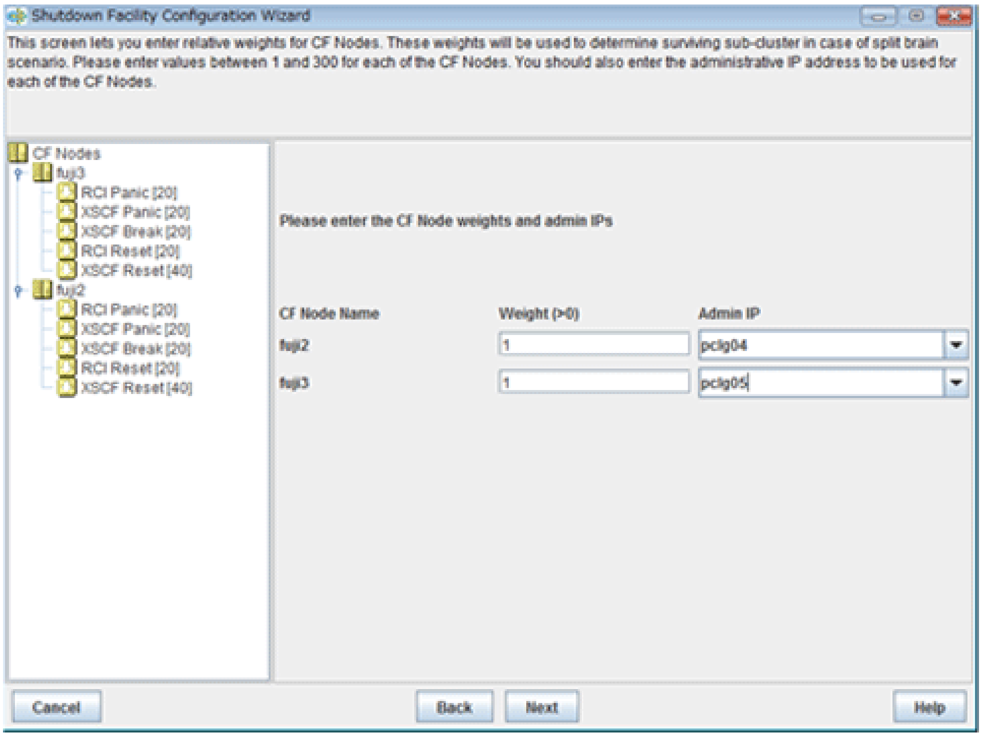
When "cycle" was selected, stop the node forcibly, and then reboot it.

When "leave-off" was selected, stop the node forcibly, and then cut its power.

Entering node weights and administrative IP addresses

The screen for entering the weights of the nodes and the IP addresses for the administrative LAN will appear.

Figure 5.8 Entering node weights and administrative IP addresses



Enter the weights of the nodes and the IP addresses for the administrative LAN.

Weight

Enter the weight of the node that constitutes the cluster. Weight is used to identify the survival priority of the node group that constitutes the cluster. Possible values for each node range from 1 to 300.
For details on survival priority and weight, refer to the explanations below.

Admin IP

Enter an IP address directly or click the tab to select the host name that is assigned to the administrative IP address.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

Upon the completion of configuration, click *Next*.

Survival priority

Even if a cluster partition occurs due to a failure in the cluster interconnect, all the nodes will still be able to access the user resources. For details on the cluster partition, see "1.2.2.1 Protecting data integrity" in "PRIMECLUSTER Concepts Guide".
To guarantee the consistency of the data constituting user resources, you have to determine the node groups to survive and those that are to be forcibly stopped.
The weight assigned to each node group is referred to as a "Survival priority" under PRIMECLUSTER.
The greater the weight of the node, the higher the survival priority. Conversely, the less the weight of the node, the lower the survival priority. If multiple node groups have the same survival priority, the node group that includes a node with the name that is first in alphabetical order will survive.

Survival priority can be calculated based on the following formula:

Survival priority = SF node weight + ShutdownPriority of userApplication

Note

When SF calculates the survival priority, each node will send its survival priority to the remote node via the administrative LAN. If any communication problem of the administrative LAN occurs, the survival priority will not be able to reach. In this case, the survival priority will be calculated only by the SF node weight.

SF node weight (Weight):

Weight of node. Default value = 1. Set this value while configuring the shutdown facility.

userApplication ShutdownPriority:

Set this attribute when userApplication is created. For details on how to change the settings, see "11.1 Changing the Operation Attributes of a Cluster Application".

See

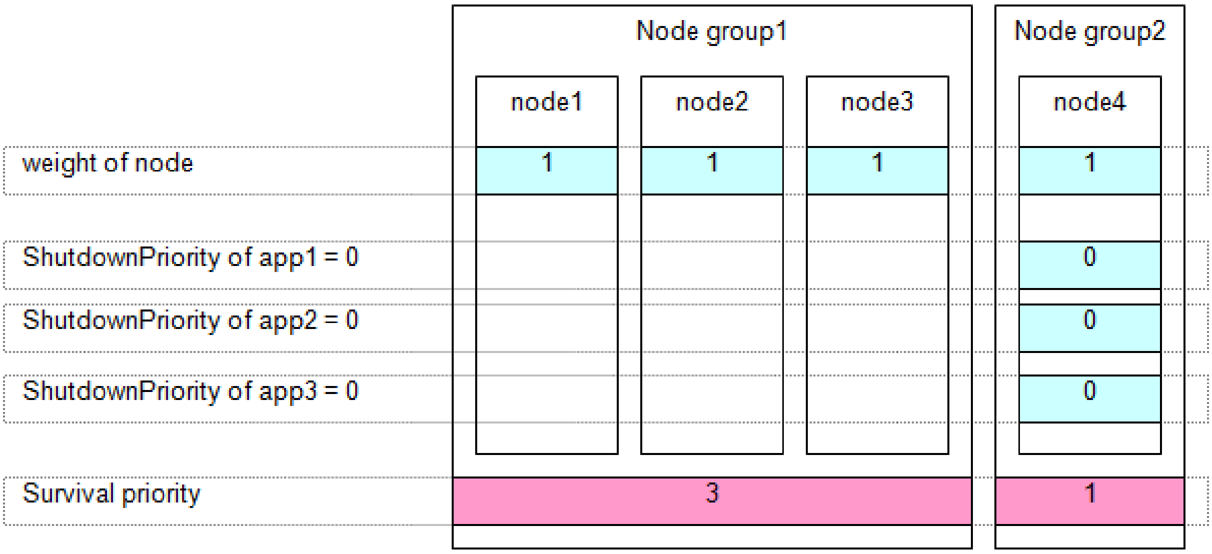
For details on the ShutdownPriority attribute of userApplication, see "6.7.5 Attributes".

Survival scenarios

The typical scenarios that are implemented are shown below:

[Largest node group survival]

- Set the weight of all nodes to 1 (default).
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

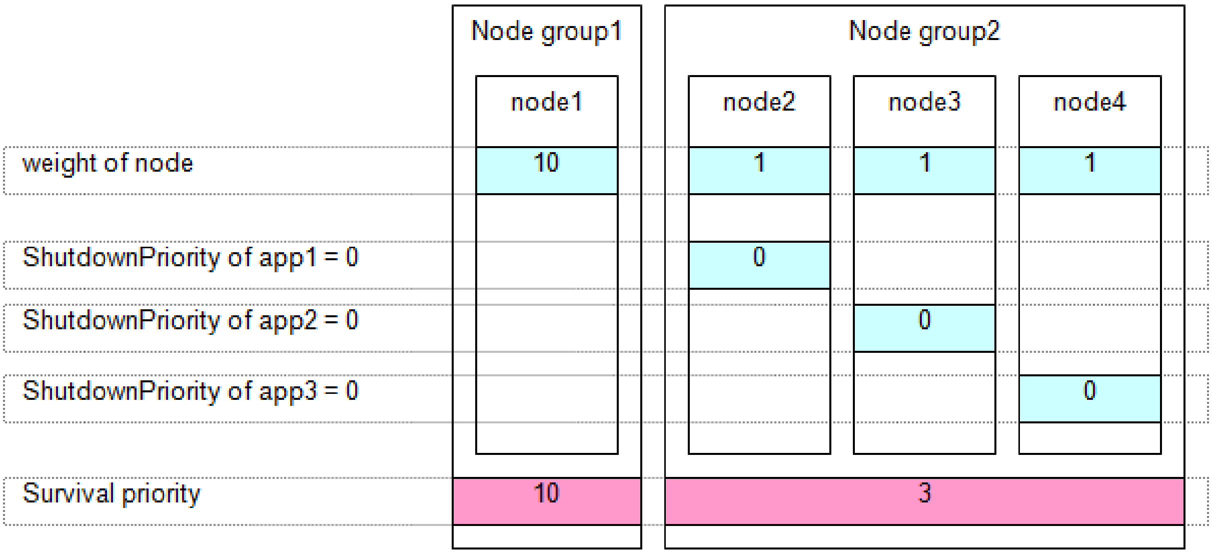


[Specific node survival]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes.

- Set the ShutdownPriority attribute of all user applications to 0 (default).

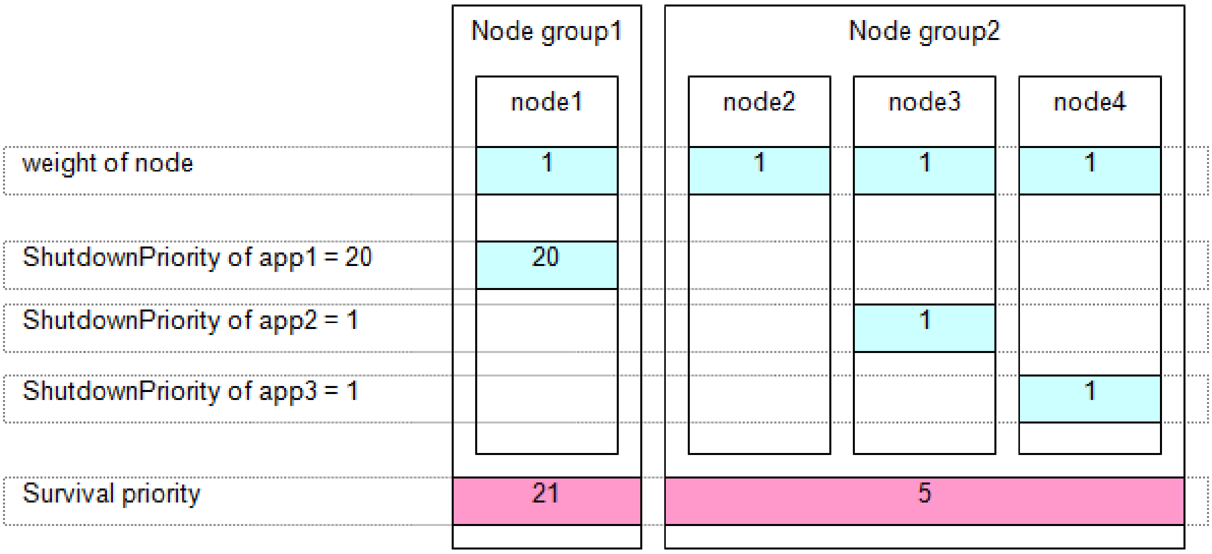
  In the following example, node1 is to survive:



[Specific application survival]

- Set the "weight" of all nodes to 1 (default).
- Set the ShutdownPriority attribute of the user application whose operation is to continue to a value more than double the total of the ShutdownPriority attributes of the other user applications and the weights of all nodes.

  In the following example, the node for which app1 is operating is to survive:

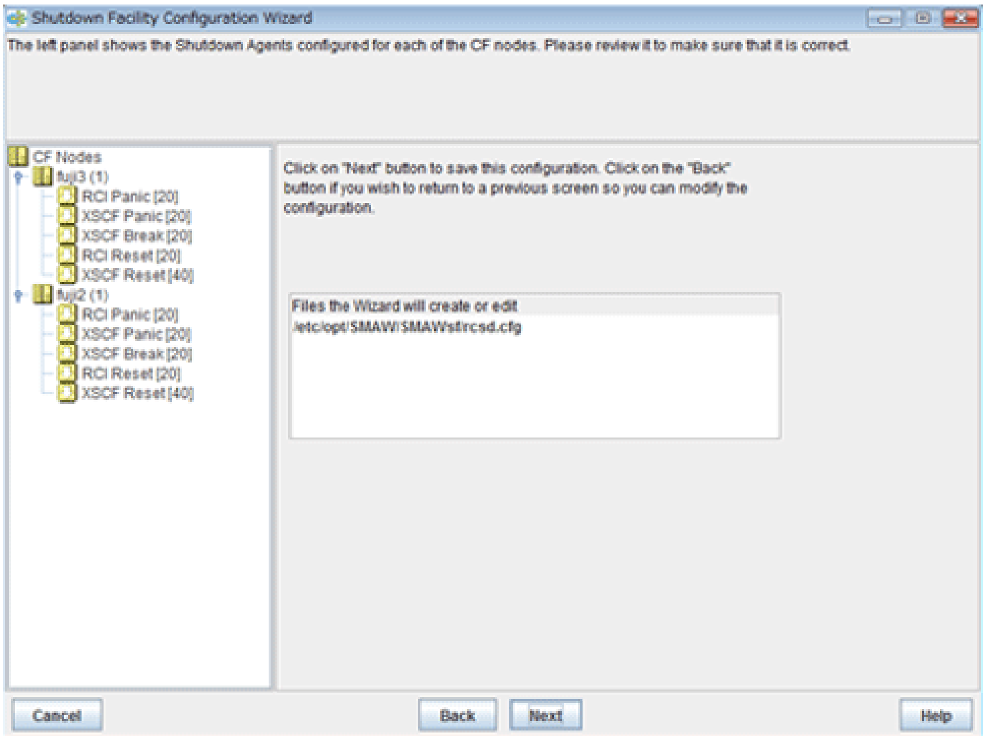


Saving the configuration

Confirm and then save the configuration.

In the left-hand panel of the window, those nodes that constitute the cluster are displayed (SF node weight is displayed in brackets after those nodes), as are the shutdown agents that are configured for each node.

Figure 5.9 Saving the configuration



Click *Next*. A popup screen will appear for confirmation.

Select *Yes* to save the setting.

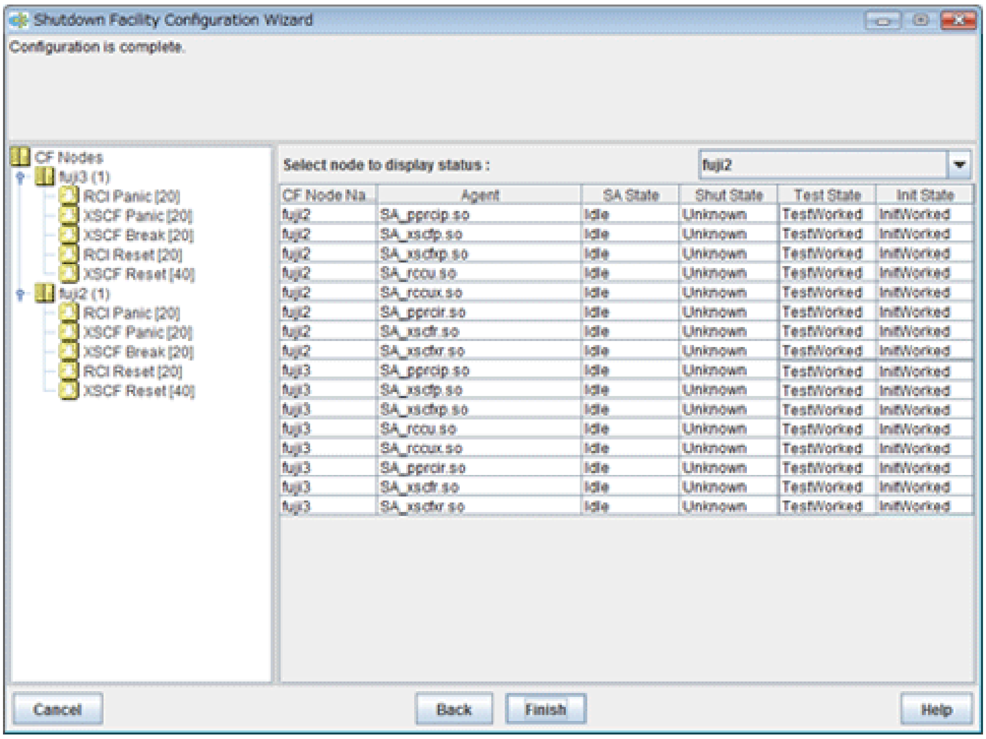
Displaying the configuration of the shutdown facility

If you save the setting, a screen displaying the configuration of the shutdown facility will appear. On this screen, you can confirm the configuration of the shutdown facility on each node by selecting each node in turn.

Information

You can also view the configuration of the shutdown facility by selecting *Shutdown Facility* -> *Show Status* from the *Tool* menu.

Figure 5.10 Show Status



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

- If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the agent or hardware configuration may contain an error. Check the /var/adm/messages file and the console for an error message. Then, apply appropriate countermeasures as instructed the message that is output.

- If the connection to XSCF from the shutdown facility is telnet and also the setting of the connection to XSCF is SSH, the test state becomes TestFailed at this point in time. Confirm that the shutdown facility is operating normally, after the "5.1.2.2.3 Setting of the connection method to the XSCF".

See

For details on how to respond to the error messages that may be output, see "PRIMECLUSTER Messages."

5.1.2.2.3 Setting of the connection method to the XSCF

The default of setting of the connection method to the XSCF is SSH connection, in the SPARC Enterprise M3000, M4000, M5000, M8000, or M9000. The procedure when changing to the telnet connection is the following.

Change of the connection method

Execute the following command in all nodes to change a connection method.

```
# /etc/opt/FJSVcluster/bin/clrccusetup -m -t telnet
```

After changing the connection method, execute the clrccusetup -l command to check that "telnet" is displayed in the "connection-type" field.

```
# /etc/opt/FJSVcluster/bin/clrccusetup -l
Device-name cluster-host-name IP-address host-name user-name connection-type
-------------------------------------------------------------------------------
xscf        fuji2             xscf2      1         xuser     telnet
xscf        fuji3             xscf3      1         xuser     telnet
```

Starting up the shutdown facility

Execute the following command in each node, and confirm the shutdown facility has started.

```
# /opt/SMAW/bin/sdtool -s
```

If the state of configuration of shutdown facility is displayed, shutdown facility is started.

If "The RCSD is not running" is displayed, shutdown facility is not started.

If shutdown facility is started, execute the following command, and restart the shutdown facility.

```
# /opt/SMAW/bin/sdtool -r
```

If shutdown facility is not started, execute the following command, and start the shutdown facility.

```
# /opt/SMAW/bin/sdtool -b
```

5.1.2.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series

5.1.2.3.1 Checking Console Configuration

In SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series, ILOM is used.

Check the following settings concerning ILOM before setting the shutdown facility.

- The log in user account is made for the shutdown facility, and CLI mode of that is set to the default mode (*1).
- User inquiries of the first SSH connection (such as generation of the RSA key) are completed by connecting to ILOM from all the cluster nodes via SSH using the log in user account for the shutdown facility.

  When using the host name for specifying ILOM name, the first SSH connection with the specified host name must be completed.

If you are using ILOM 3.0, please check the following settings as well.

- The log in user account for the shutdown facility must be set to one of the following privileges:
  - If the keyswitch_state parameter is set to normal
    - Console, Reset and Host Control, Read Only (cro)
    - Operator(*2)

        - If the keyswitch_state parameter is set to locked
            - Admin, Console, Reset and Host Control, Read Only (acro)
            - Administrator(*2)

    If a necessary privilege is not set, TestFailed or KillFailed of shutdown agent would be occurred.

- The log in user account for the shutdown facility must not be using SSH host-based key authentication.

Moreover, record the following information on ILOM.

- ILOM IP address(*3)
- Log in user account and password for shutdown facility in ILOM

*1) You can check if CLI mode of the log in user account is set to the default mode by the following procedure.

1. Log in CLI of ILOM.
2. Check prompt status.
   Prompt status that is set to the default mode.
   ->
   Prompt status that is set to alom mode.
   sc>

*2) Due to compatibility of ILOM 3.0 with ILOM 2.x, this operation is also available for users with administrator or operator privileges from ILOM 2.x.

*3) When the network routing is set, the IP address of ILOM need not be the same to management LAN segment of the cluster node.

See

For details on how to make and check ILOM settings, please refer to the following documentation.

- For ILOM 2.x:
    - "Integrated Lights Out Manager User's Guide"
- For ILOM 3.0:
    - "Integrated Lights Out Manager (ILOM) 3.0 Concepts Guide"
    - "Integrated Lights Out Manager (ILOM) 3.0 Web Interface Procedures Guide"
    - "Integrated Lights Out Manager (ILOM) 3.0 CLI Procedures Guide"
    - "Integrated Lights Out Manager (ILOM) 3.0 Getting Started Guide"

5.1.2.3.2 Using the Shutdown Configuration Wizard

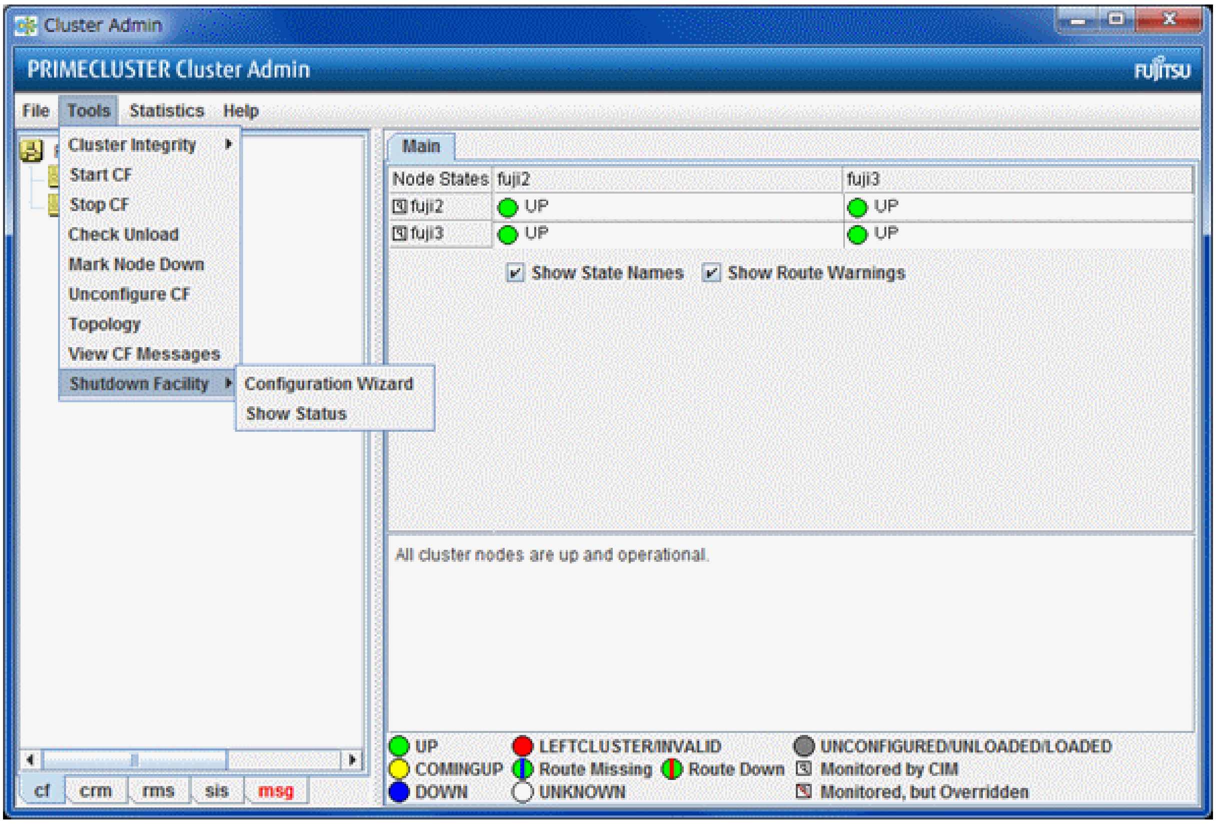
The required shutdown agent varies depending on the hardware machine type.

Check the following combinations of the hardware machine types and shutdown agents.

- For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, and SPARC T3, T4, T5, T7, S7 series
    - ILOM Panic
    - ILOM Reset

Starting up the shutdown configuration wizard

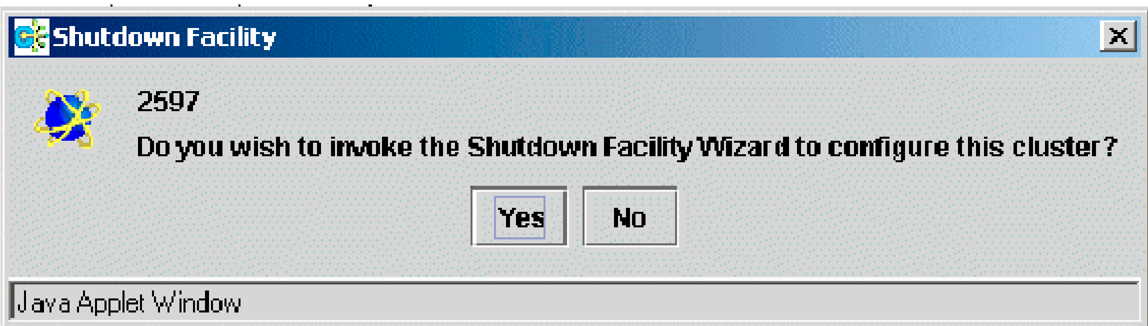
From the CF main window of the Cluster Admin screen, select the *Tool* menu and then *Shutdown Facility -> Configuration Wizard*. The shutdown configuration wizard will start.



Note

You can also configure the shutdown facility immediately after you complete the CF configuration with the CF wizard.

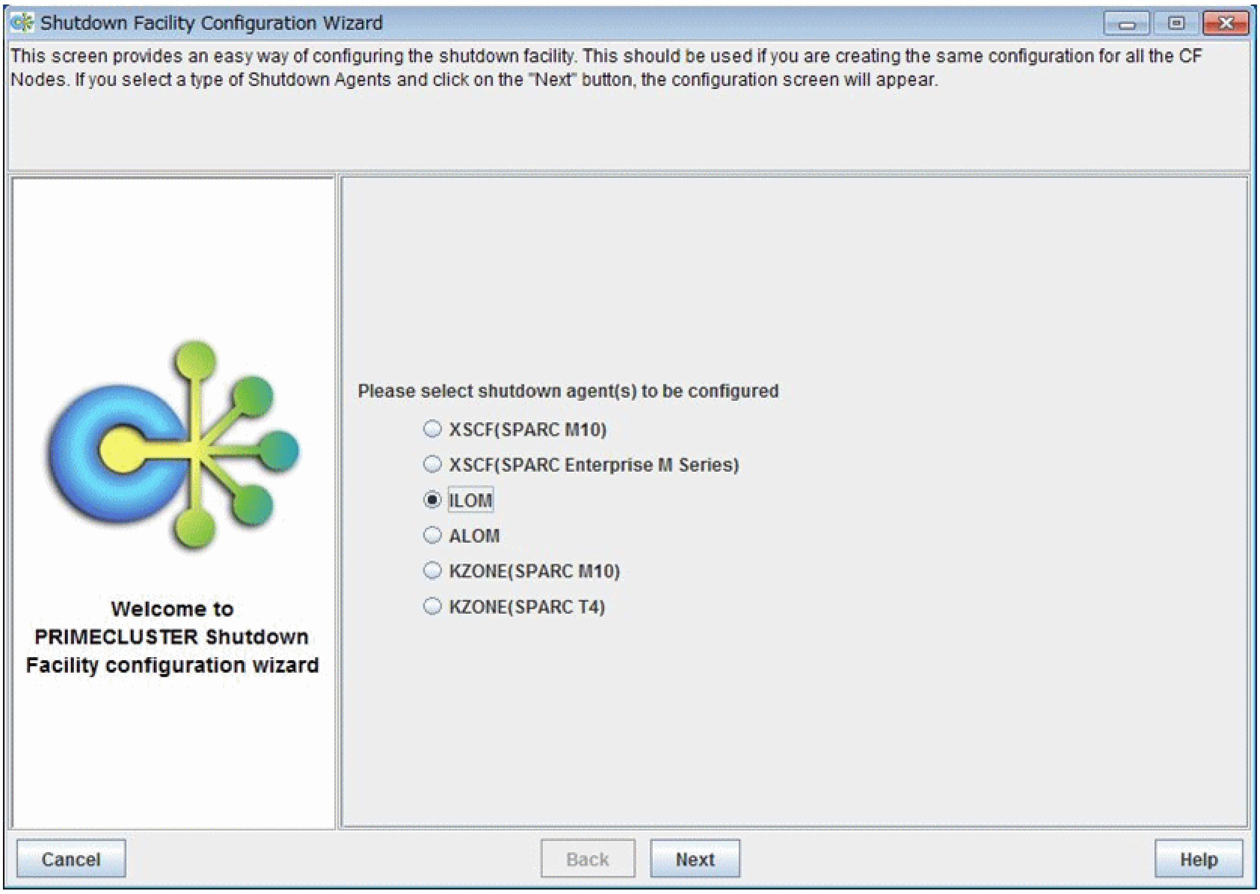
The following confirmation popup screen will appear. Click *Yes* to start the shutdown configuration wizard.



Selecting a shutdown agent

The selection screen for the shutdown agent will appear.

Figure 5.11 Selecting a shutdown agent



Confirm the hardware machine type and select the appropriate shutdown agent.

- For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, and SPARC T3, T4, T5, T7, S7 series

  Select *ILOM*.

  The following shutdown agents are automatically set:

  - ILOM Panic
  - ILOM Reset

Select *ILOM*, and then click *Next*.

Information

If you select a shutdown agent, the timeout value is automatically set. For details on the timeout value of the shutdown agent, see "8.2.2 Configuration file of SF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

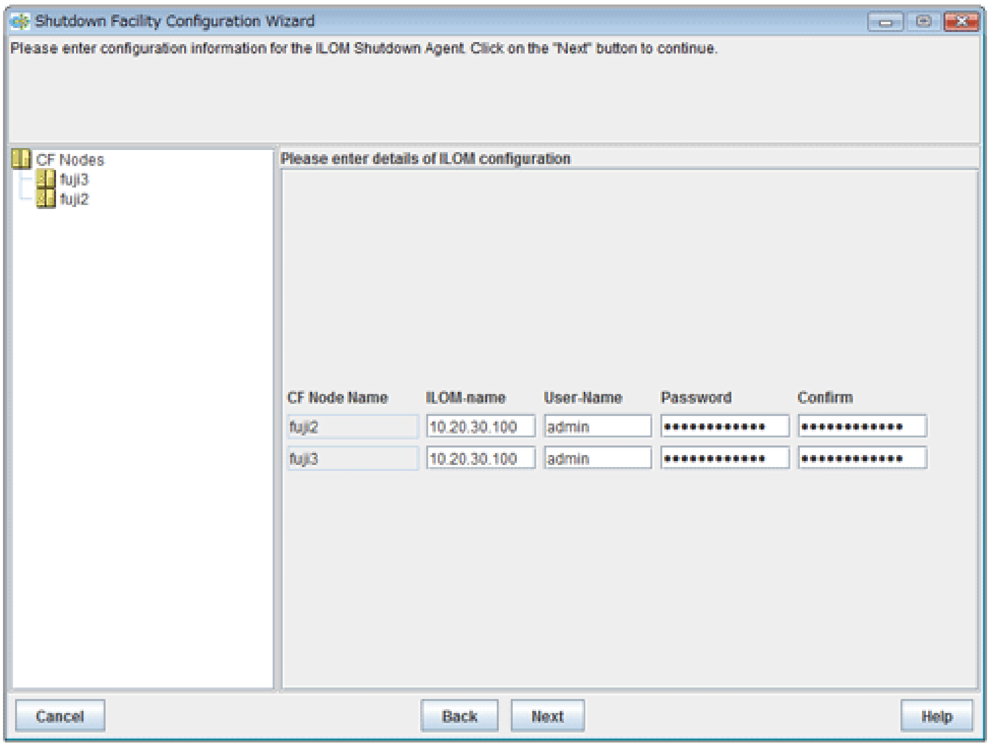
- For ILOM Panic/ILOM Reset

  Timeout value = 70 (seconds)

Configuring ILOM

The screen for entering the information of ILOM will appear.

Figure 5.12 Configuring ILOM



Enter the settings for ILOM that you recorded in "5.1.2.3.1 Checking Console Configuration".

ILOM-Name

Enter the IP address of ILOM or the host name of ILOM that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

User-Name

Enter a user name to log in to ILOM.

Password

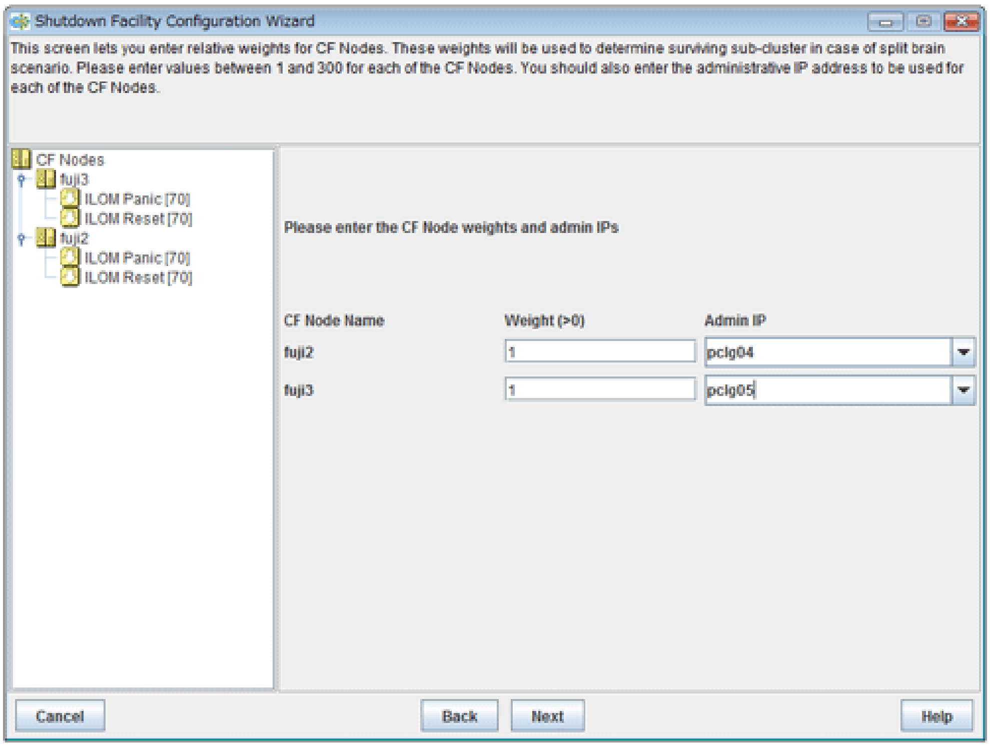
Enter a password to log in to ILOM.

Upon the completion of configuration, click *Next*.

Entering node weights and administrative IP addresses

The screen for entering the weights of the nodes and the IP addresses for the administrative LAN will appear.

Figure 5.13 Entering node weights and administrative IP addresses



Enter the weights of the nodes and the IP addresses for the administrative LAN.

Weight

Enter the weight of the node that constitutes the cluster. Weight is used to identify the survival priority of the node group that constitutes the cluster. Possible values for each node range from 1 to 300.
For details on survival priority and weight, refer to the explanations below.

Admin IP

Enter an IP address directly or click the tab to select the host name that is assigned to the administrative IP address.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

Upon the completion of configuration, click *Next*.

Survival priority

Even if a cluster partition occurs due to a failure in the cluster interconnect, all the nodes will still be able to access the user resources. For details on the cluster partition, see "1.2.2.1 Protecting data integrity" in "PRIMECLUSTER Concepts Guide".
To guarantee the consistency of the data constituting user resources, you have to determine the node groups to survive and those that are to be forcibly stopped.
The weight assigned to each node group is referred to as a "Survival priority" under PRIMECLUSTER.
The greater the weight of the node, the higher the survival priority. Conversely, the less the weight of the node, the lower the survival priority. If multiple node groups have the same survival priority, the node group that includes a node with the name that is first in alphabetical order will survive.

Survival priority can be calculated based on the following formula:

Survival priority = SF node weight + ShutdownPriority of userApplication

Note

When SF calculates the survival priority, each node will send its survival priority to the remote node via the administrative LAN. If any communication problem of the administrative LAN occurs, the survival priority will not be able to reach. In this case, the survival priority will be calculated only by the SF node weight.

SF node weight (Weight):

Weight of node. Default value = 1. Set this value while configuring the shutdown facility.

userApplication ShutdownPriority:

Set this attribute when userApplication is created. For details on how to change the settings, see "11.1 Changing the Operation Attributes of a Cluster Application".

See

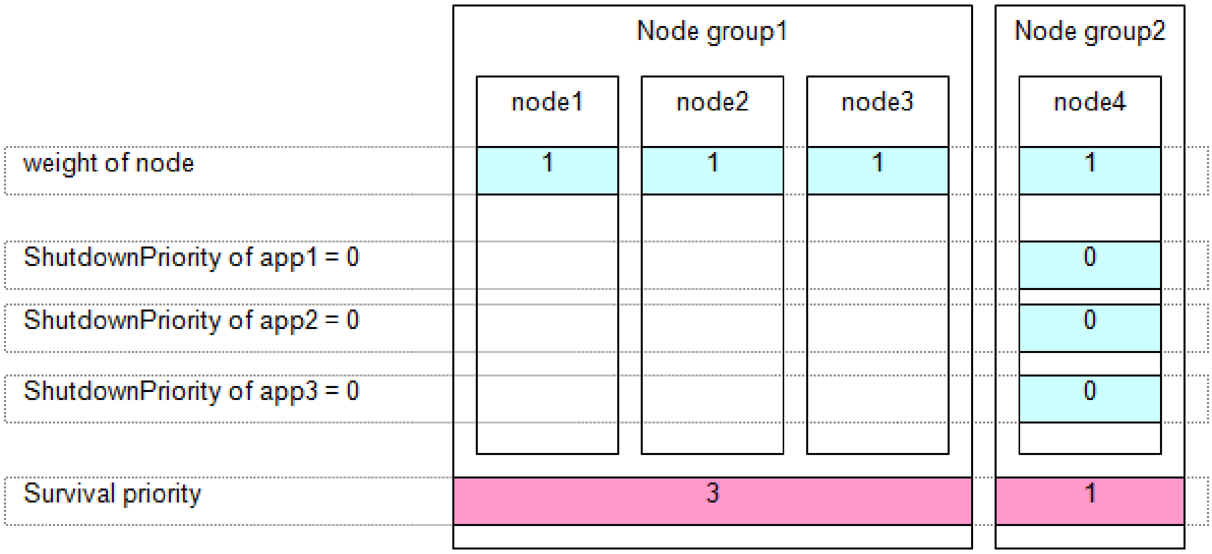
For details on the ShutdownPriority attribute of userApplication, see "6.7.5 Attributes".

Survival scenarios

The typical scenarios that are implemented are shown below:

[Largest node group survival]

- Set the weight of all nodes to 1 (default).
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

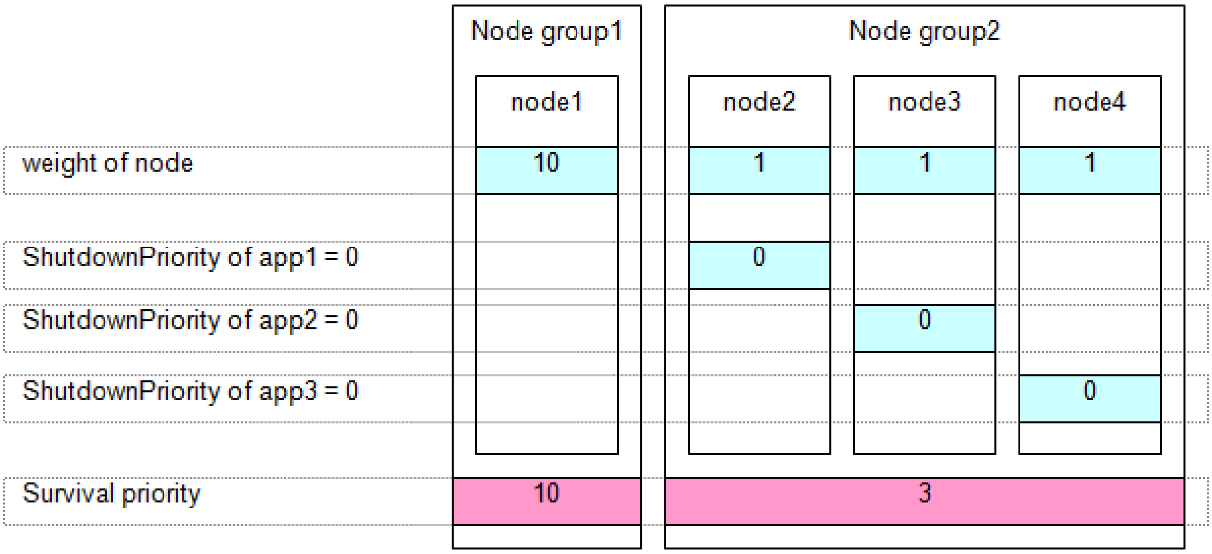


[Specific node survival]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes.

- Set the ShutdownPriority attribute of all user applications to 0 (default).

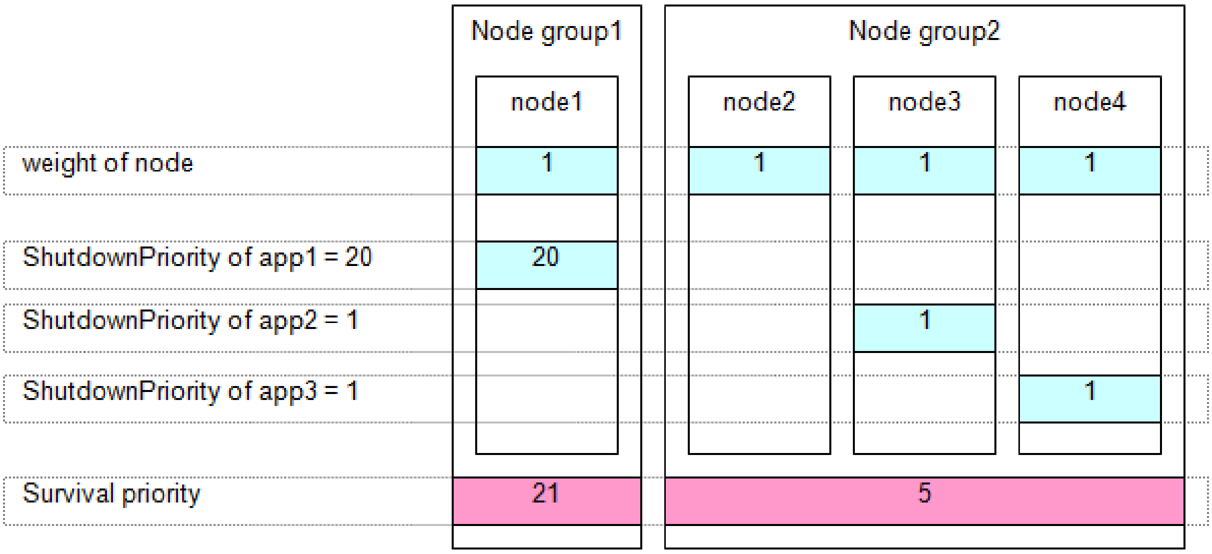
  In the following example, node1 is to survive:



[Specific application survival]

- Set the "weight" of all nodes to 1 (default).
- Set the ShutdownPriority attribute of the user application whose operation is to continue to a value more than double the total of the ShutdownPriority attributes of the other user applications and the weights of all nodes.

  In the following example, the node for which app1 is operating is to survive:

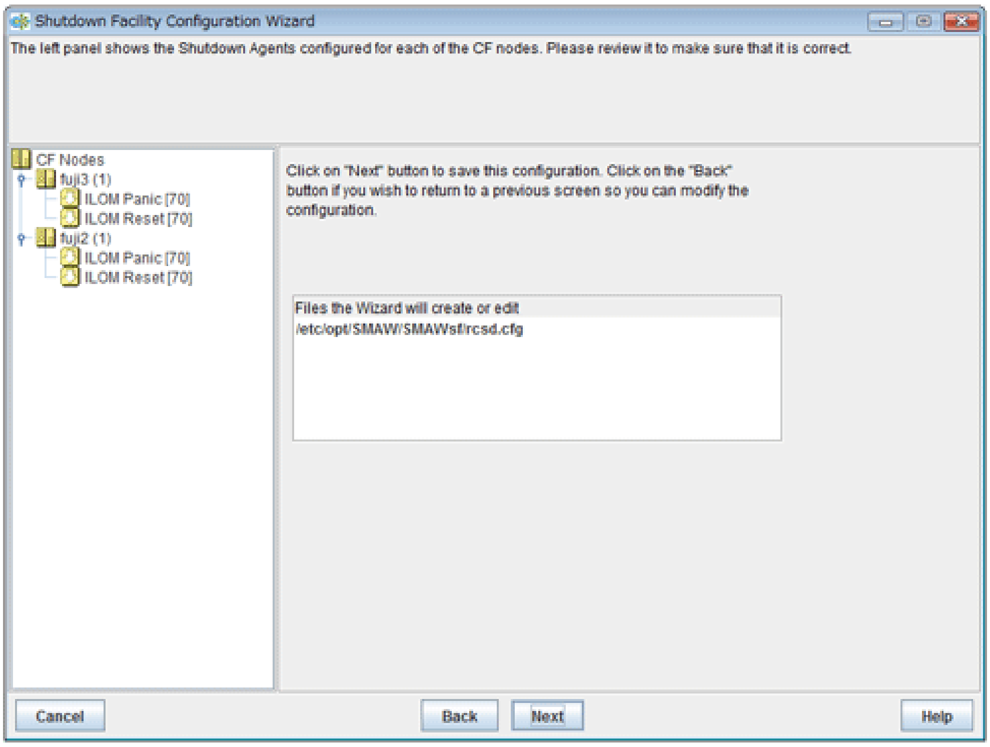


Saving the configuration

Confirm and then save the configuration.

In the left-hand panel of the window, those nodes that constitute the cluster are displayed (SF node weight is displayed in brackets after those nodes), as are the shutdown agents that are configured for each node.

Figure 5.14 Saving the configuration



Click *Next*. A popup screen will appear for confirmation.

Select *Yes* to save the setting.

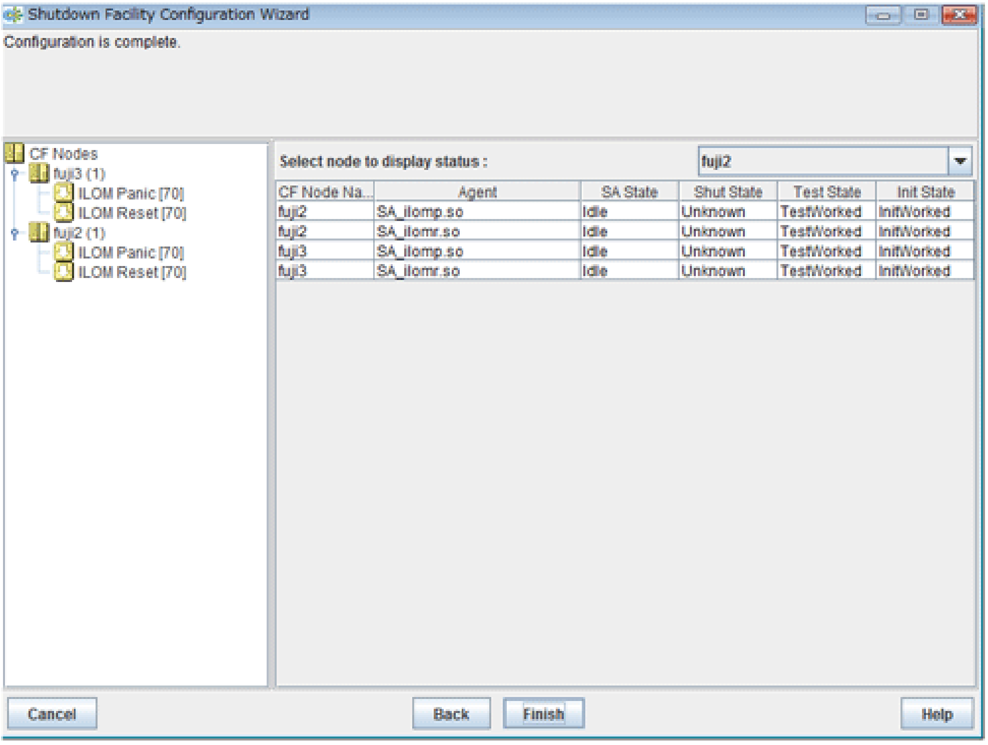
Displaying the configuration of the shutdown facility

If you save the setting, a screen displaying the configuration of the shutdown facility will appear. On this screen, you can confirm the configuration of the shutdown facility on each node by selecting each node in turn.

Information

You can also view the configuration of the shutdown facility by selecting *Shutdown Facility -> Show Status* from the *Tool* menu.

Figure 5.15 Show Status



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the agent or hardware configuration may contain an error. Check the /var/adm/messages file and the console for an error message. Then, apply appropriate countermeasures as instructed the message that is output.

5.1.2.4 For SPARC Enterprise T1000, T2000

5.1.2.4.1 Checking Console Configuration

ALOM in console can be used by SPARC Enterprise T1000 or T2000.

Confirm the following settings concerning ALOM before setting the shutdown facility.

- The log in user account must be made for the shutdown facility, and c level (console access) authority must be given.
- In connected permission protocol type from the outside, the telnet must be effective.

- The following ALOM composition variables must not have been changed from default.
  - if_emailalerts : false(default)
  - sc_clieventlevel : 2(default)
  - sc_cliprompt : sc(default)

Note

- As a default, the connection permission from the outside to ALOM is SSH. In that case, it is not supported in the shutdown facility.
- When the connection to ALOM is a serial port connection alone, it is not supported in the shutdown facility.

Moreover, record the following information on ALOM.

- ALOM IP address(*1) or an ALOM host name registered in the "/etc/inet/hosts" file.
- User name used to log in the ALOM.
- Password used to log in the ALOM.

*1) When the network routing is set, Internet Protocol address of ALOM need not be the same to management LAN segment of the cluster node.

See

For information on how to configure and confirm ALOM, see the "*Advanced Lights out Management (ALOM) CMT Guide*".

5.1.2.4.2 Using the Shutdown Configuration Wizard

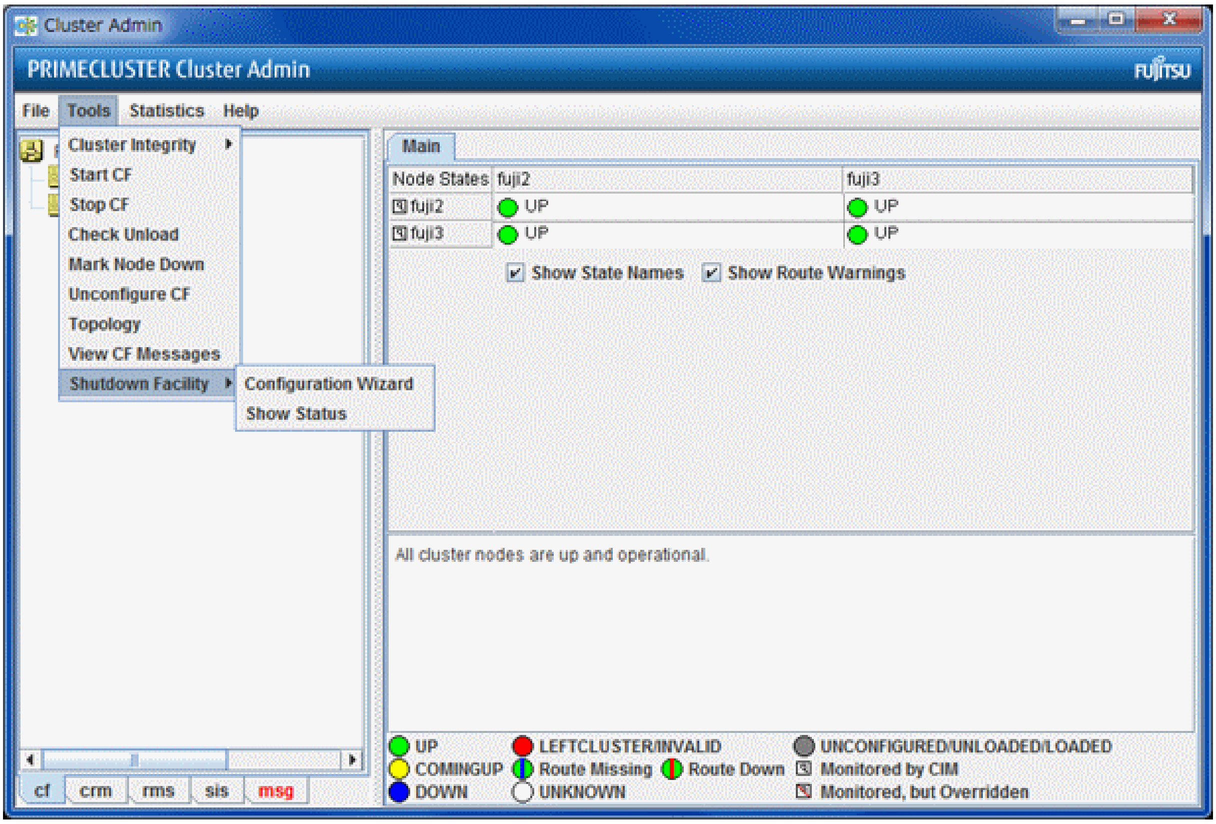
The required shutdown agent varies depending on the hardware machine type.

Check the following combinations of the hardware machine types and shutdown agents.

- For SPARC Enterprise T1000, T2000
  - ALOM Break

Starting up the shutdown configuration wizard

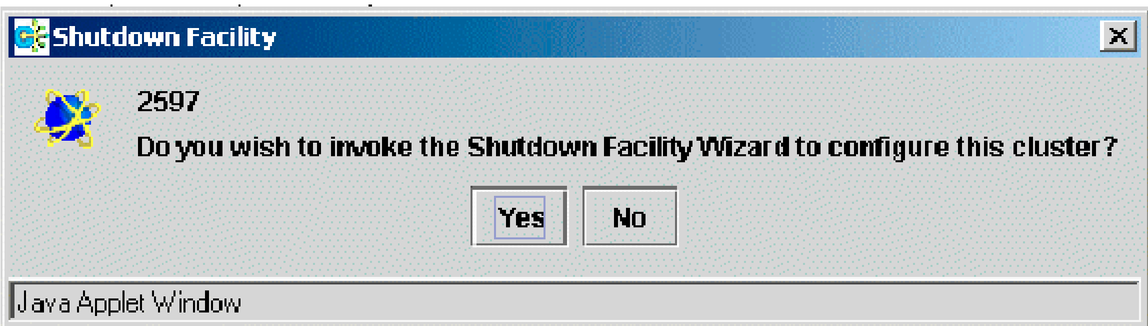
From the CF main window of the Cluster Admin screen, select the *Tool* menu and then *Shutdown Facility -> Configuration Wizard*. The shutdown configuration wizard will start.



Note

You can also configure the shutdown facility immediately after you complete the CF configuration with the CF wizard.

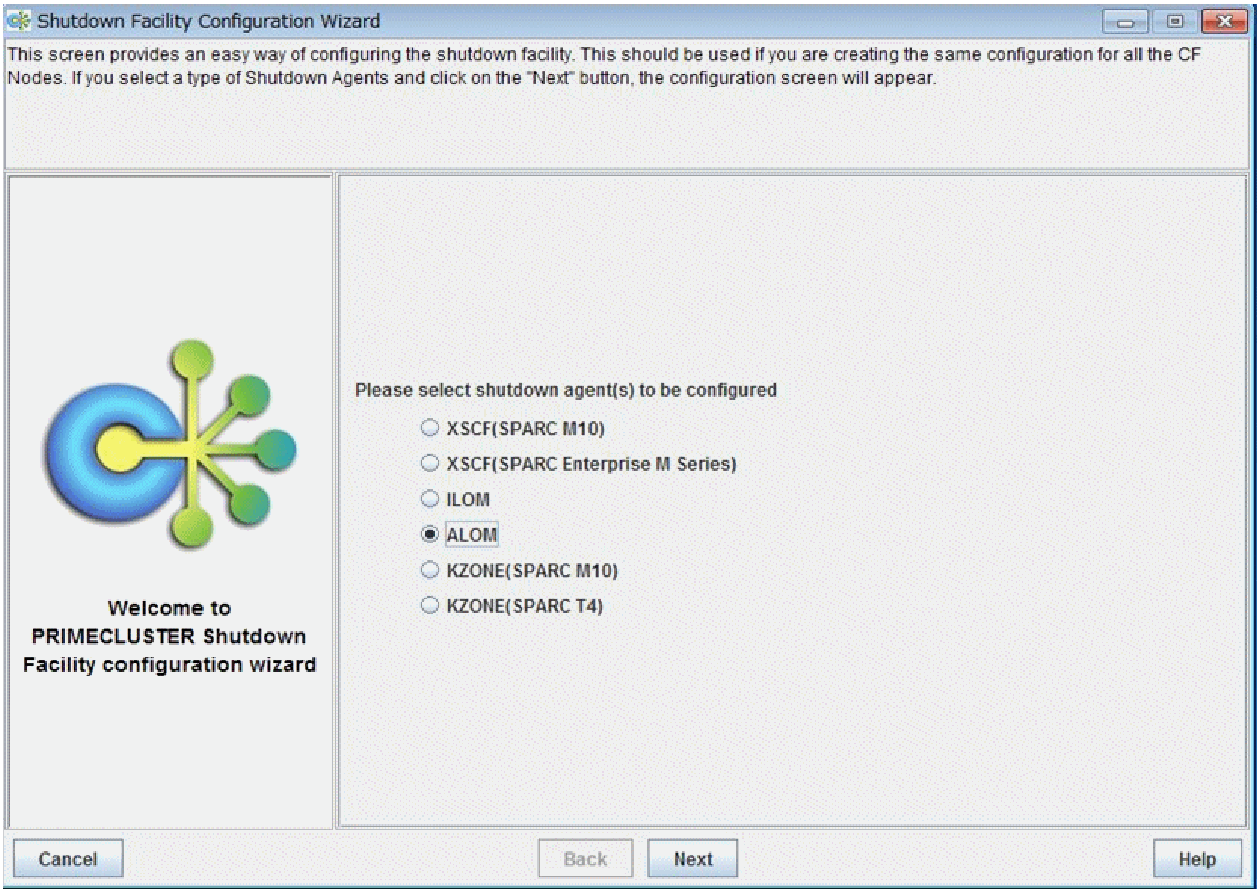
The following confirmation popup screen will appear. Click *Yes* to start the shutdown configuration wizard.



Selecting a shutdown agent

The selection screen for the shutdown agent will appear.

Figure 5.16 Selecting a shutdown agent



Confirm the hardware machine type and select the appropriate shutdown agent.

- For SPARC Enterprise T1000, T2000

  Select *ALOM*.

  The following shutdown agent is automatically set:

  - ALOM Break

Select *ALOM*, and then click *Next*.

Information

If you select a shutdown agent, the timeout value is automatically set. For details on the timeout value of the shutdown agent, see "8.2.2 Configuration file of SF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

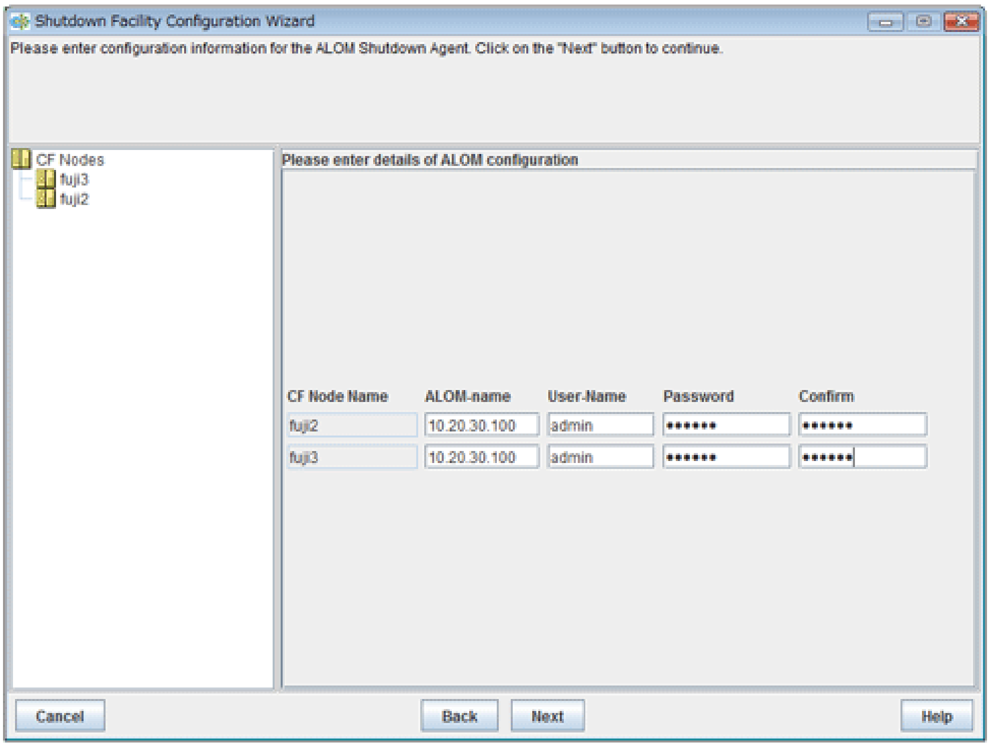
- For ALOM Break

  Timeout value = 40 (seconds)

Configuring ALOM

The screen for entering the information of ALOM will appear.

Figure 5.17 Configuring ALOM



Enter the settings for ALOM that you recorded in "5.1.2.4.1 Checking Console Configuration".

ALOM-Name

Enter the IP address of ALOM.

Available IP addresses are IPv4 addresses.

User-Name

Enter a user name to log in to ALOM.

Password

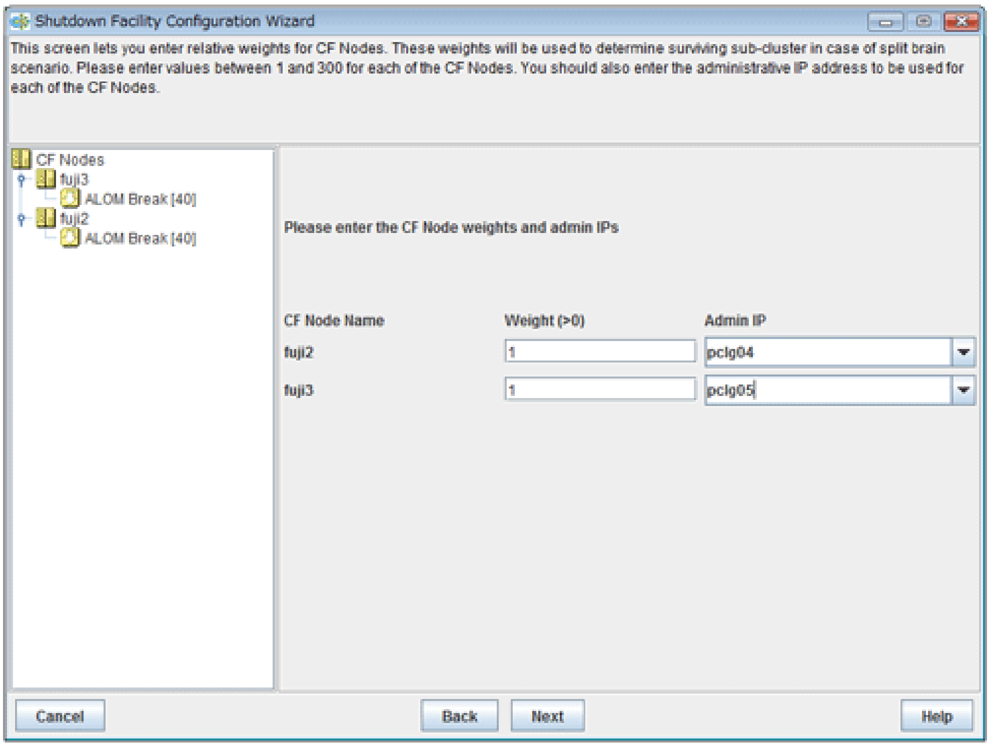
Enter a password to log in to ALOM.

Upon the completion of configuration, click *Next*.

Entering node weights and administrative IP addresses

The screen for entering the weights of the nodes and the IP addresses for the administrative LAN will appear.

Figure 5.18 Entering node weights and administrative IP addresses



Enter the weights of the nodes and the IP addresses for the administrative LAN.

Weight

Enter the weight of the node that constitutes the cluster. Weight is used to identify the survival priority of the node group that constitutes the cluster. Possible values for each node range from 1 to 300.
For details on survival priority and weight, refer to the explanations below.

Admin IP

Enter an IP address directly or click the tab to select the host name that is assigned to the administrative IP address.

Available IP addresses are IPv4 addresses.

Upon the completion of configuration, click *Next*.

Survival priority

Even if a cluster partition occurs due to a failure in the cluster interconnect, all the nodes will still be able to access the user resources. For details on the cluster partition, see "1.2.2.1 Protecting data integrity" in "PRIMECLUSTER Concepts Guide".
To guarantee the consistency of the data constituting user resources, you have to determine the node groups to survive and those that are to be forcibly stopped.
The weight assigned to each node group is referred to as a "Survival priority" under PRIMECLUSTER.
The greater the weight of the node, the higher the survival priority. Conversely, the less the weight of the node, the lower the survival priority. If multiple node groups have the same survival priority, the node group that includes a node with the name that is first in alphabetical order will survive.

Survival priority can be calculated based on the following formula:

Survival priority = SF node weight + ShutdownPriority of userApplication

Note

When SF calculates the survival priority, each node will send its survival priority to the remote node via the administrative LAN. If any communication problem of the administrative LAN occurs, the survival priority will not be able to reach. In this case, the survival priority will be calculated only by the SF node weight.

SF node weight (Weight):

Weight of node. Default value = 1. Set this value while configuring the shutdown facility.

userApplication ShutdownPriority:

Set this attribute when userApplication is created. For details on how to change the settings, see "11.1 Changing the Operation Attributes of a Cluster Application".

See

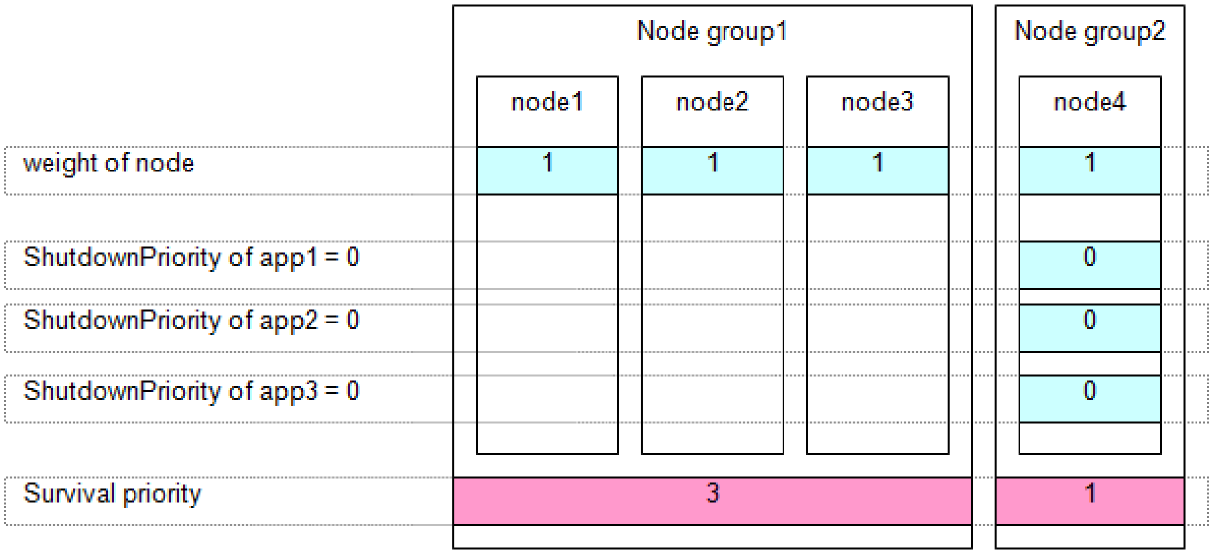
For details on the ShutdownPriority attribute of userApplication, see "6.7.5 Attributes".

Survival scenarios

The typical scenarios that are implemented are shown below:

[Largest node group survival]

- Set the weight of all nodes to 1 (default).
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

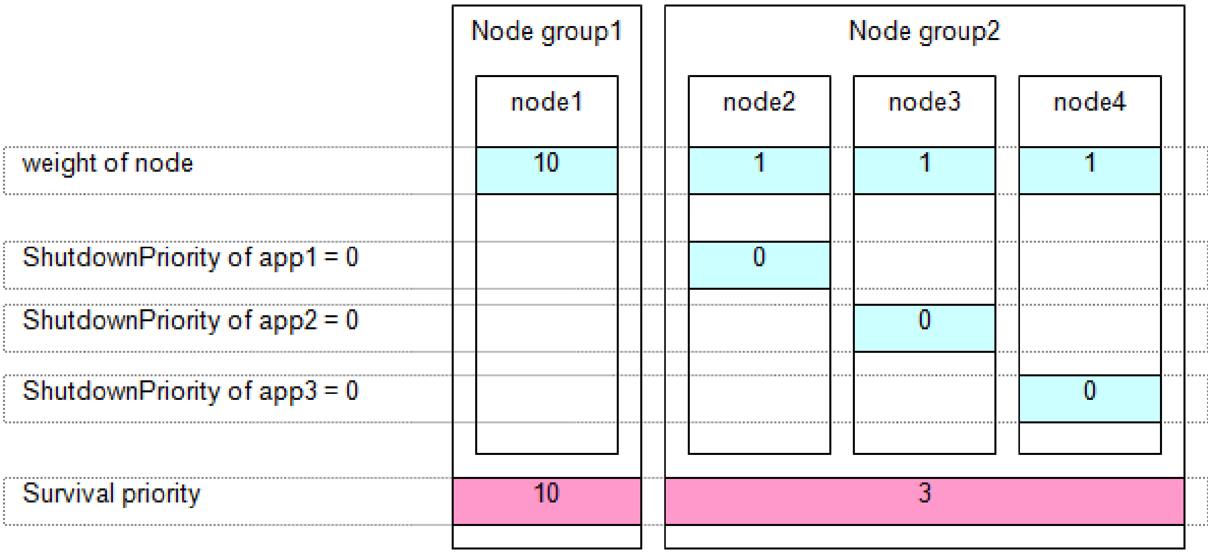


[Specific node survival]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes.

- Set the ShutdownPriority attribute of all user applications to 0 (default).

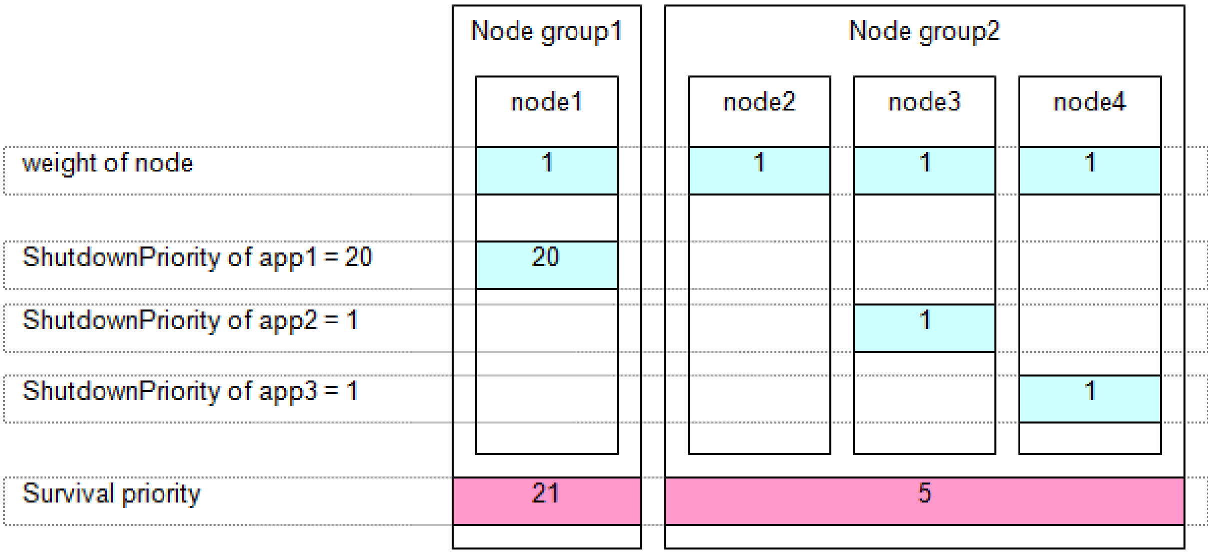
  In the following example, node1 is to survive:



[Specific application survival]

- Set the "weight" of all nodes to 1 (default).
- Set the ShutdownPriority attribute of the user application whose operation is to continue to a value more than double the total of the ShutdownPriority attributes of the other user applications and the weights of all nodes.

  In the following example, the node for which app1 is operating is to survive:

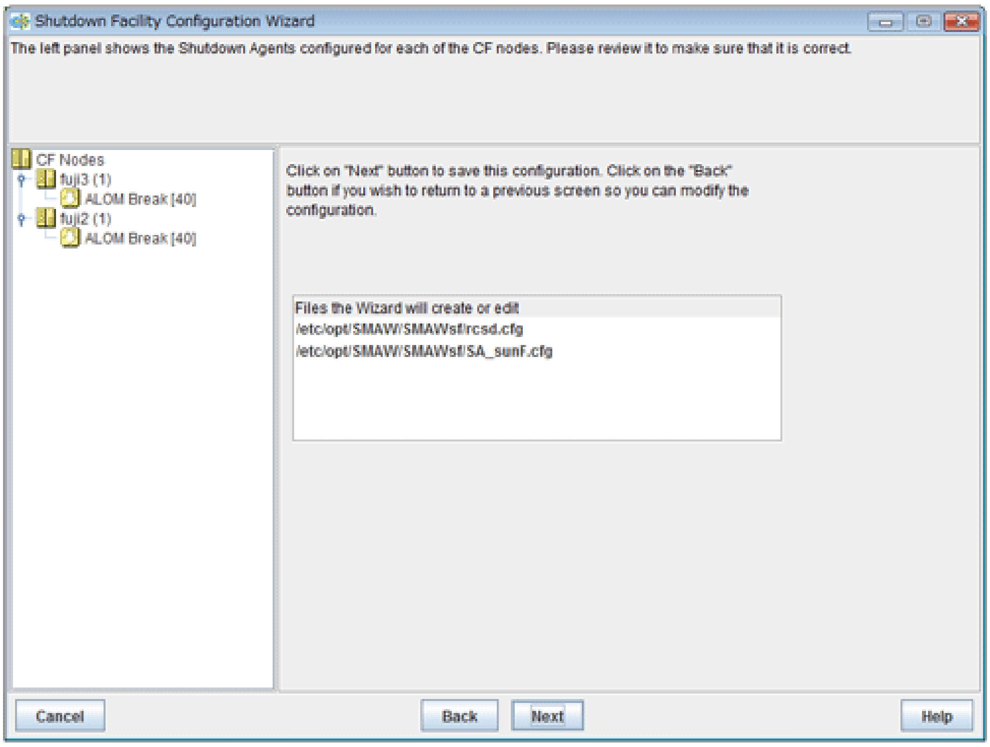


Saving the configuration

Confirm and then save the configuration.

In the left-hand panel of the window, those nodes that constitute the cluster are displayed (SF node weight is displayed in brackets after those nodes), as are the shutdown agents that are configured for each node.

Figure 5.19 Saving the configuration



Click *Next*. A popup screen will appear for confirmation.

Select *Yes* to save the setting.

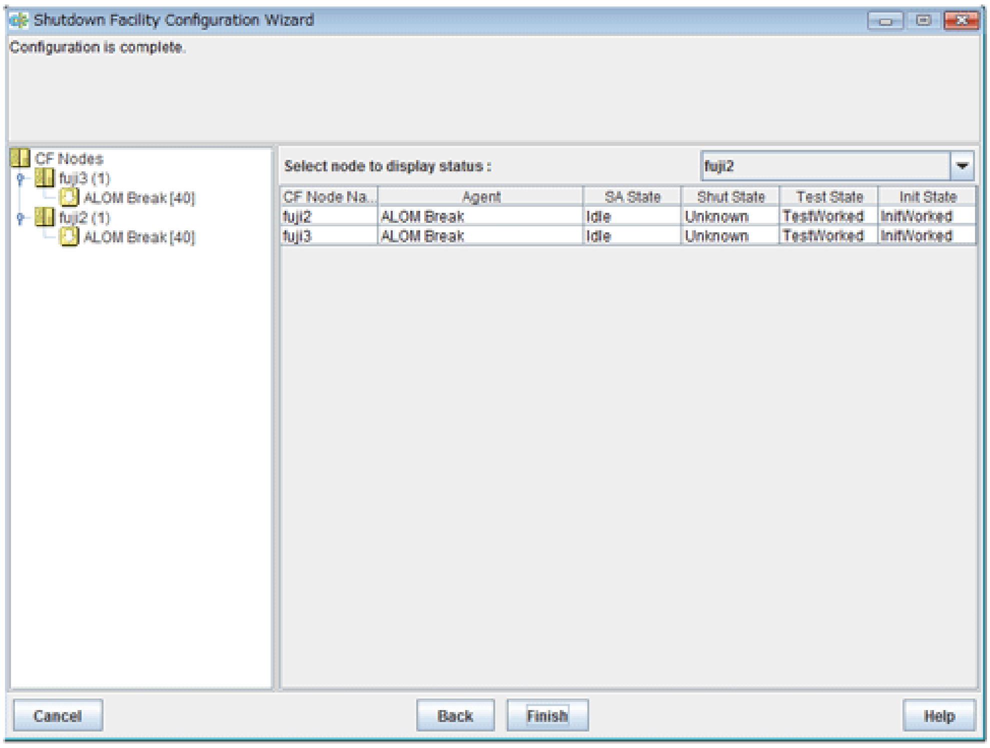
Displaying the configuration of the shutdown facility

If you save the setting, a screen displaying the configuration of the shutdown facility will appear. On this screen, you can confirm the configuration of the shutdown facility on each node by selecting each node in turn.

Information

You can also view the configuration of the shutdown facility by selecting *Shutdown Facility -> Show Status* from the *Tool* menu.

Figure 5.20 Show Status



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the agent or hardware configuration may contain an error. Check the /var/adm/messages file and the console for an error message. Then, apply appropriate countermeasures as instructed the message that is output.

5.1.2.5 For Oracle Solaris Kernel Zones

5.1.2.5.1 Checking XSCF Information

When building Kernel Zones in SPARC M10, the KZONE shutdown agent uses XSCF.

The connection method to XSCF can be selected from SSH or the telnet. Confirm the following settings concerning XSCF before setting the shutdown facility.

- Commonness
    - The log in user account must be made excluding root for the shutdown facility, and the pparop authority must be given.

      When building a cluster system on the control domain, you can also use the login user account set in "5.1.2.1.1 Checking XSCF Information."

    - In the configuration where the asynchronous monitoring sub-LAN is not used, the host name corresponding to the IP address of XSCF-LAN#0 must be defined in /etc/inet/hosts.

- At the SSH connection
    - In XSCF, SSH must be effective in connected permission protocol type from the outside.
    - User inquiries of the first SSH connection (such as generation of the RSA key) must be completed by connecting to XSCF from all the cluster nodes via SSH using the log in user account for the shutdown facility.

      When using the host name for specifying XSCF name, the first SSH connection with the specified host name must be completed.

- At the telnet connection
    - In XSCF, telnet must be effective in connected permission protocol type from the outside.

Note

When the connection to XSCF is a serial port connection alone, it is not supported in the shutdown facility. Use the SSH connection or the telnet connection that uses XSCF-LAN.

Moreover, record the following information on XSCF.

- XSCF IP address (*1 and *2) or an XSCF host name registered in the "/etc/inet/hosts" file of the node
- Log in user account and password for shutdown facility in XSCF

  *1) When the network routing is set, the IP address of XSCF need not be the same to management LAN segment of the cluster node.

  *2) For SPARC M10-4S environment, specify the XSCF takeover IP address.

See

For information on how to configure and confirm XSCF, see the "SPARC M10 Systems System Operation and Administration Guide".

5.1.2.5.2 Checking ILOM Information

When building Kernel Zones to SPARC T4, T5, T7, S7 series, use ILOM for the KZONE shutdown agent.

Check the following settings concerning ILOM before setting the shutdown facility.

- The log in user account is made for the shutdown facility, and CLI mode of that is set to the default mode (*1).
- User inquiries of the first SSH connection (such as generation of the RSA key) are completed by connecting to ILOM from all the cluster nodes via SSH using the log in user account for the shutdown facility.

  When using the host name for specifying ILOM name, the first SSH connection with the specified host name must be completed.

- The log in user account for the shutdown facility must be set to one of the following privileges:
    - Console, Read Only (cro)

  If a necessary privilege is not set, TestFailed or KillFailed of shutdown agent would be occurred.

- The log in user account for the shutdown facility must not be using SSH host-based key authentication.

Moreover, record the following information on ILOM.

- ILOM IP address(*2)
- Log in user account and password for shutdown facility in ILOM

*1) You can check if CLI mode of the log in user account is set to the default mode by the following procedure.

1. Log in CLI of ILOM.
2. Check prompt status.

   Prompt status that is set to the default mode.

   ->D53

   Prompt status that is set to alom mode.

   sc>

*2) When the network routing is set, the IP address of ILOM need not be the same to management LAN segment of the cluster node.

See

For details on how to make and check ILOM settings, please refer to the following documentation.

- "Integrated Lights Out Manager (ILOM) 3.0 Concepts Guide"
- "Integrated Lights Out Manager (ILOM) 3.0 Web Interface Procedures Guide"
- "Integrated Lights Out Manager (ILOM) 3.0 CLI Procedures Guide"
- "Integrated Lights Out Manager (ILOM) 3.0 Getting Started Guide"

5.1.2.5.3 Logging in to Global Zone Host

To access a host (control domain or guest domain) where Kernel Zones from Kernel Zones via SSH, you need to complete the user inquiry of the first SSH connection (RSA key generation).

On all Kernel Zones (cluster nodes), log in to the hosts where all Kernel Zones operate with the user for the shutdown facility set in "Creating a user for the shutdown facility" of "15.1.1 Software Installation and Configuration of Cluster Environment."

When using the host name for specifying the globalzone hostname, log in with the specified host name.

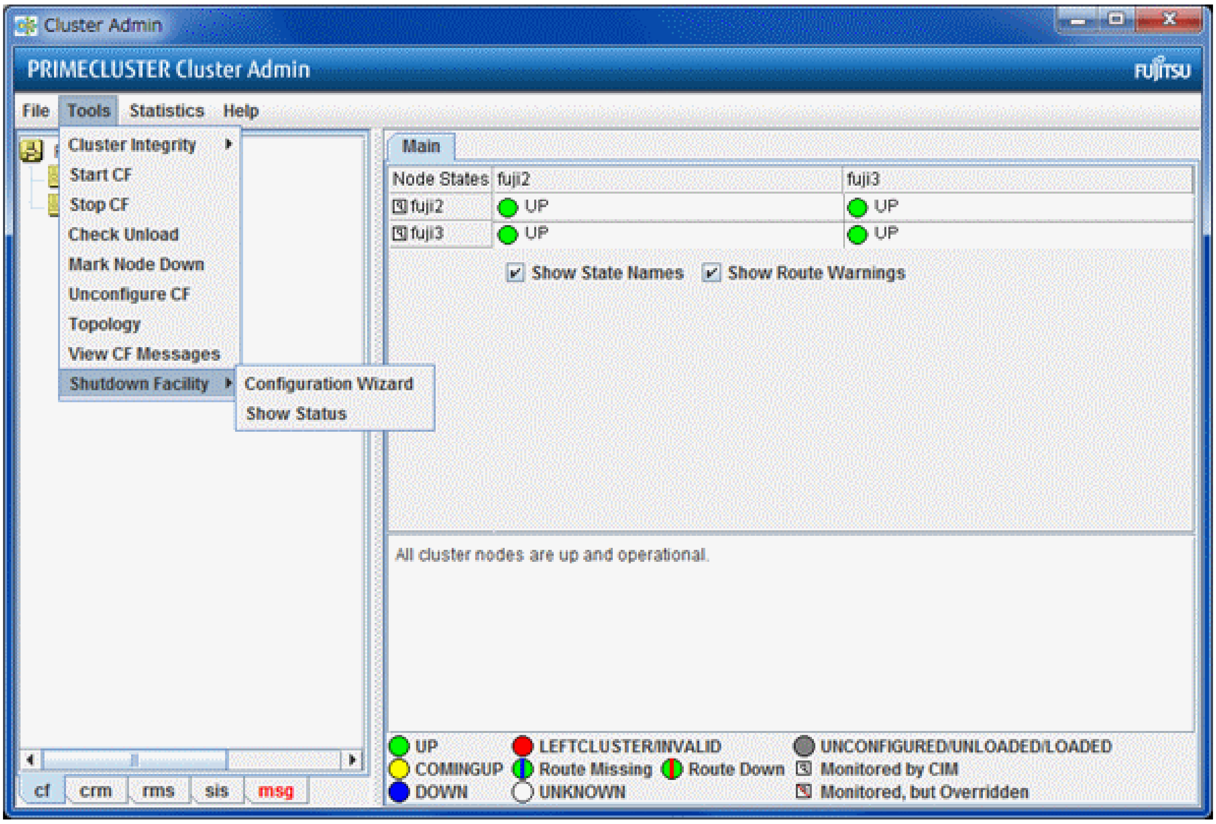
Example: when the user for the shutdown facility is "user1"

```
# ssh -l user1 XXX.XXX.XXX.XXX
The authenticity of host 'XXX.XXX.XXX.XXX (XXX.XXX.XXX.XXX)' can't be established.
RSA key fingerprint is xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx.
Are you sure you want to continue connecting (yes/no)? Yes    <- Input yes.
```

5.1.2.5.4 Using the Shutdown Configuration Wizard

Starting up the shutdown configuration wizard

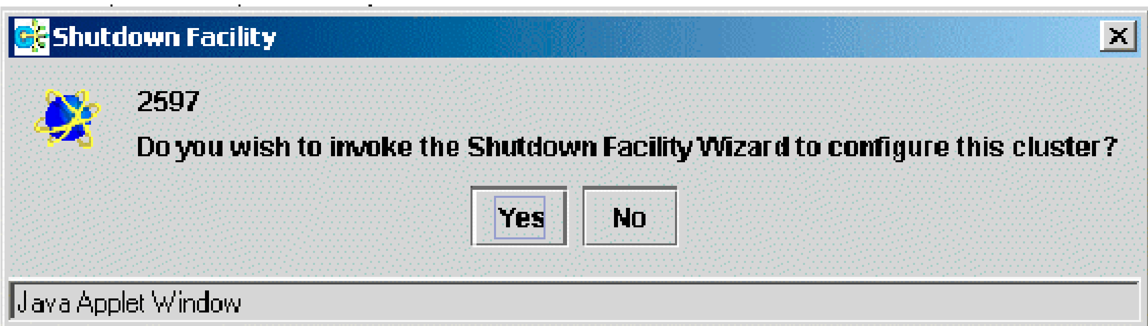
From the CF main window of the Cluster Admin screen, select the Tool menu and then Shutdown Facility -> Configuration Wizard. The shutdown configuration wizard will start.



Note

You can also configure the shutdown facility immediately after you complete the CF configuration with the CF wizard.

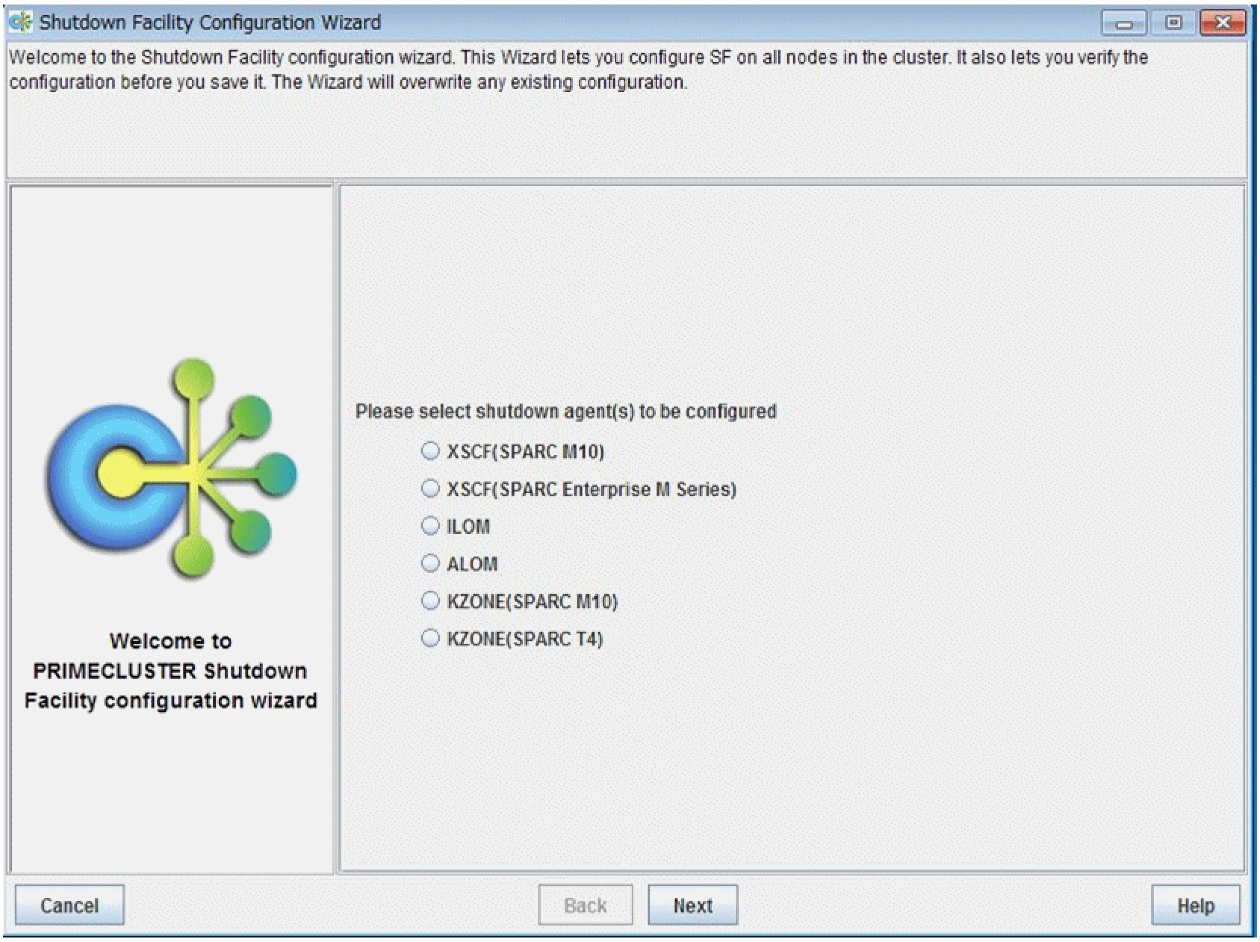
The following confirmation popup screen will appear. Click Yes to start the shutdown configuration wizard.



Selecting a shutdown agent

The selection screen for the shutdown agent will appear.

Figure 5.21 Selecting a shutdown agent



Confirm the hardware machine type and select the appropriate shutdown agent.

- For Oracle Solaris Kernel Zones

  When building Kernel Zones in SPARC M10, select [KZONE(SPARC M10)].

  When building Kernel Zones in SPARC T4, T5, T7, S7 series, select [KZONE(SPARC T4)].

  The following shutdown agents are automatically set.

  - KZONE Panic
  - KZONE Reset
  - KZONE Check

Information

If you select a shutdown agent, the timeout value is automatically set. For details on the timeout value of the shutdown agent, see "8.2.2 Configuration file of SF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

- For KZONE Panic

  Timeout value = 45 (seconds)

- For KZONE Reset

  Timeout value = 70 (seconds)

- For KZONE Check
  - For SPARC M10

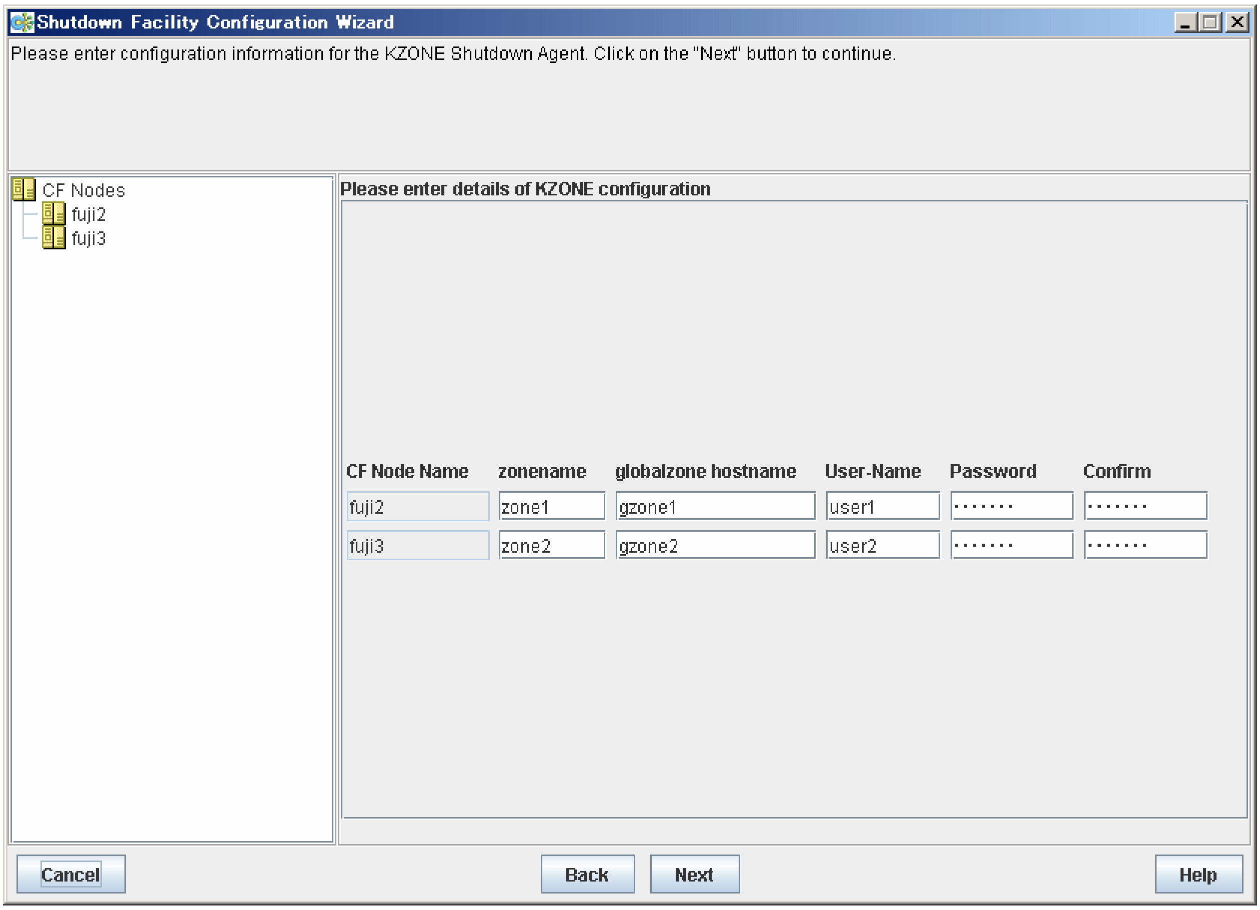
    Timeout value = 20 (seconds)
  - For SPARC T4, T5, T7, S7 series

    Timeout value = 30 (seconds)

Configuring Kernel Zones

The screen for entering the information of Kernel Zones will appear.

Figure 5.22 Configuring Kernel Zones



zone name

Enter a Kernel Zone name.

Prior to entering a Kernel Zone name, execute the zoneadm list -cv command on the global zone host to check that "solaris-kz" is displayed in BRAND of the zone name to be entered.
If a brand other than "solaris-kz" is displayed, it cannot be used because it is not a Kernel Zone.

globalzone hostname

Enter the IP address of the global zone host (control domain or guest domain) where a Kernel Zone operates, or the host name that is registered in the /etc/inet/hosts file.

If the Kernel Zone was built on the control domain, enter the IP address or host name of the control domain.

If the Kernel Zone was built on the guest domain, enter the IP address or host name of the guest domain.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

User-Name

Enter a user name to log in to the global zone host.

Specify the user for the shutdown facility that was created in "Creating a user for the shutdown facility" of "15.1.1 Software Installation and Configuration of Cluster Environment."

Password

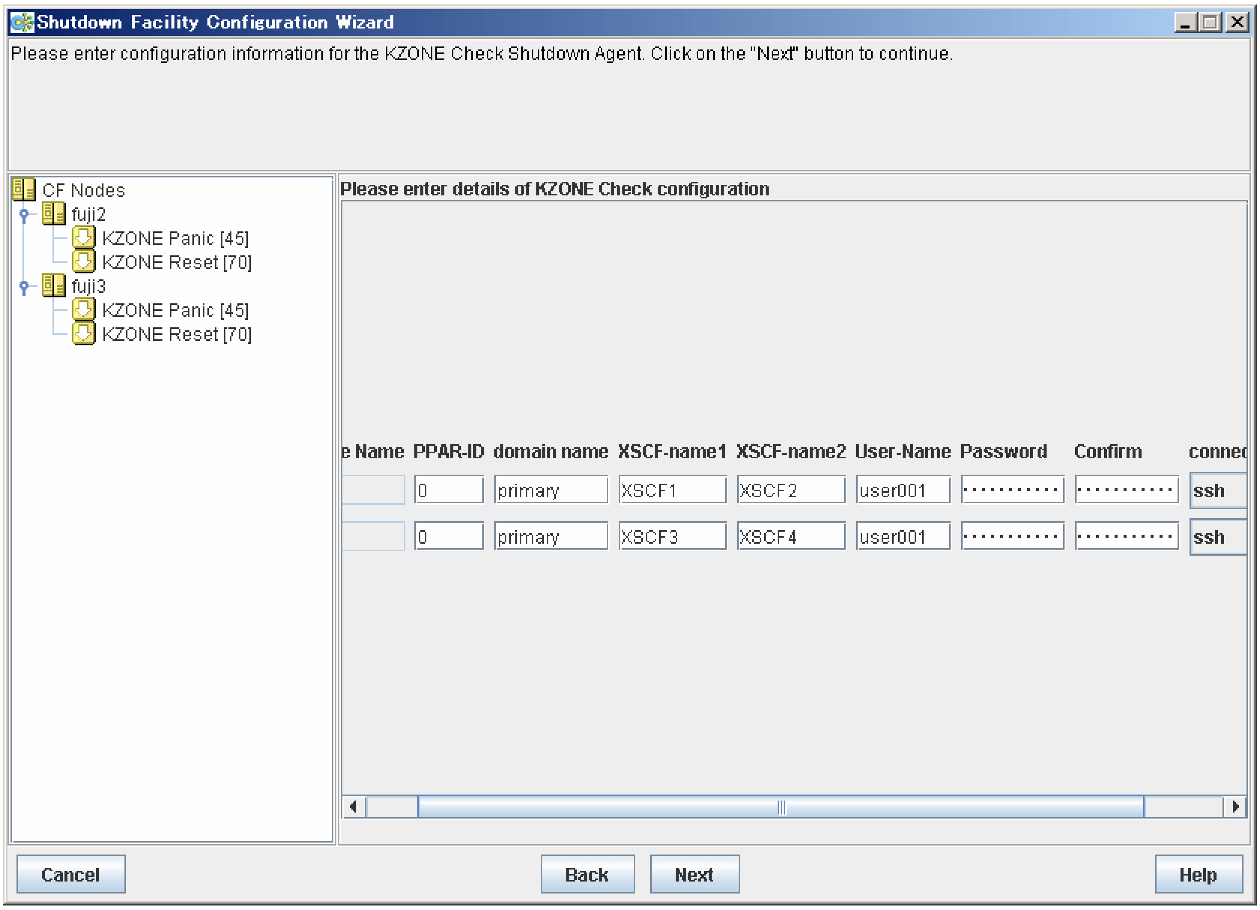
Enter a password to log in to the global zone host.

Upon the completion of configuration, click Next.

Configuring XSCF

When building a Kernel Zone in SPARC M10, the screen for entering the information of the XSCF will appear.

Figure 5.23 Configuring XSCF



Enter the settings for XSCF that you recorded in "5.1.2.1.1 Checking XSCF Information."

PPAR-ID

Specify the identification ID of the physical partition (PPAR) in which the logical domain of the cluster node belongs to.

Make sure to enter "0" for SPARC M10-1 and M10-4.

For SPARC M10-4S, specify the integer ranged from 0 to 15.

domain-name

Enter the domain name of the host (control domain or guest domain) where Kernel Zones operate.

For the control domain, enter "primary".

For the guest domain, prior to entering a domain name, execute the virtinfo -a command on the guest domain, and then enter the domain name displayed in the Domain name.

Specify the input character string using up to 255 characters starting with an alphabetic letter and consisting of only alphanumeric characters and "-" (hyphens) and "." (period).

XSCF-name1

Enter the IP address of XSCF-LAN#0 of the host where Kernel Zones operate or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

XSCF-name2

Enter the IP address of XSCF-LAN#1 of the host where Kernel Zones operate or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

User-Name

Enter a user name to log in to XSCF.

Password

Enter a password to log in to XSCF.

connection method

Select a connection method for XSCF

Select "ssh" or "telnet."

Note

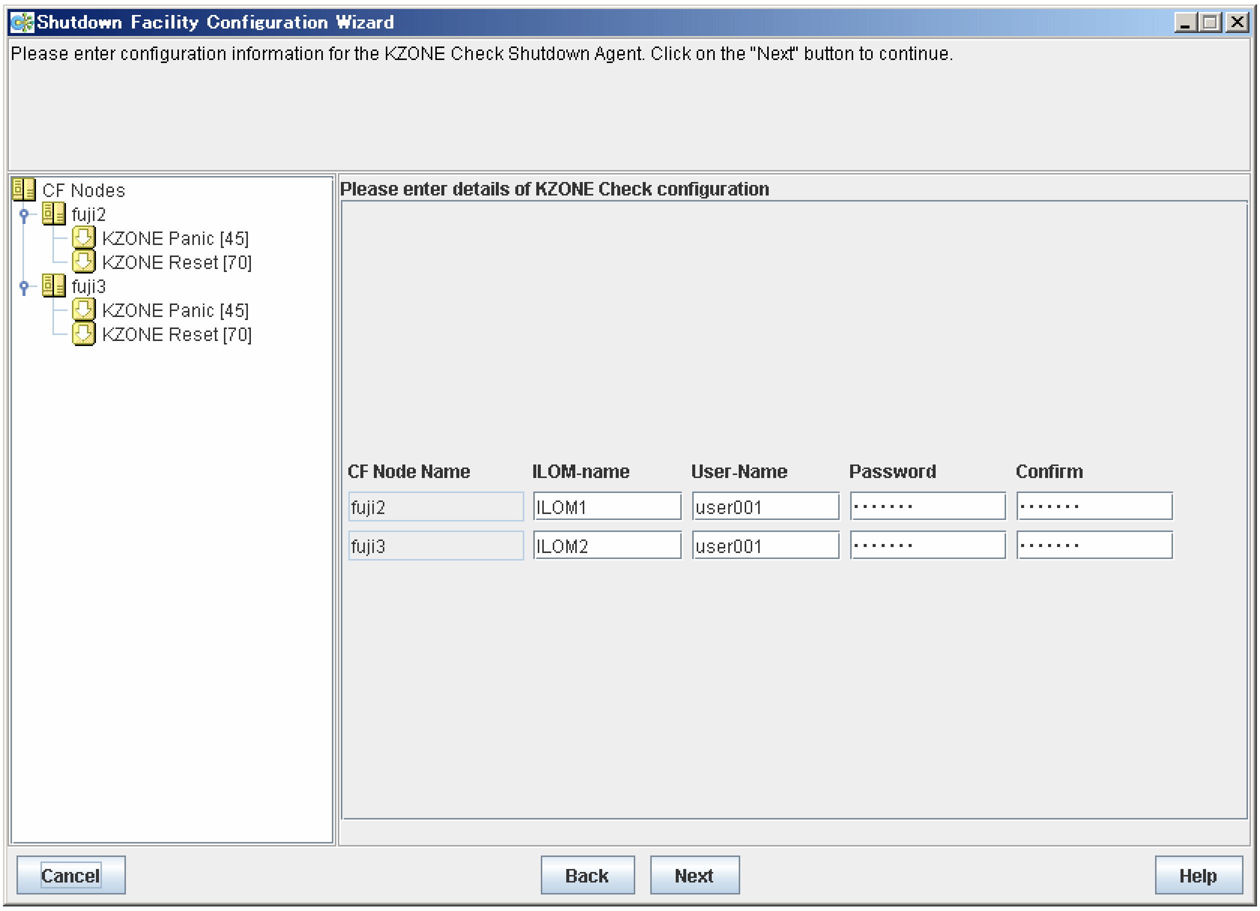
- In the environment where XSCF is duplexed, a combination of a user name and a password for 2 of the XSCF must be the same.
- To change the connection type after setting the shutdown facility, perform the setting using the shutdown facility wizard from be beginning again.
- In the configuration where the asynchronous monitoring sub-LAN is not used, specify the IP address of XSCF-LAN#0 to "XSCF-name1" and the host name corresponding to the IP address specified in XSCF- name1 to "XSCF-name2." For the configuration where the asynchronous monitoring sub-LAN is not used, see "2.2.2 XSCF Configuration in SPARC M10."
- If an incorrect value has been entered, the cluster system in a Kernel Zone cannot be switched even if an error occurs in the host (control domain or guest domain) where the Kernel Zone operates.

Upon the completion of configuration, click Next.

Configuring ILOM

When building Kernel Zones in SPARC T4, T5, T7, S7 series, the screen for entering the information of ILOM will appear.

Figure 5.24 Configuring ILOM



Enter the settings for ILOM that you recorded in "5.1.2.5.2 Checking ILOM Information."

ILOM-Name

Enter the IP address for ILOM of the host where Kernel Zones operate or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

User-Name

Enter a user name to log in to ILOM.

Password

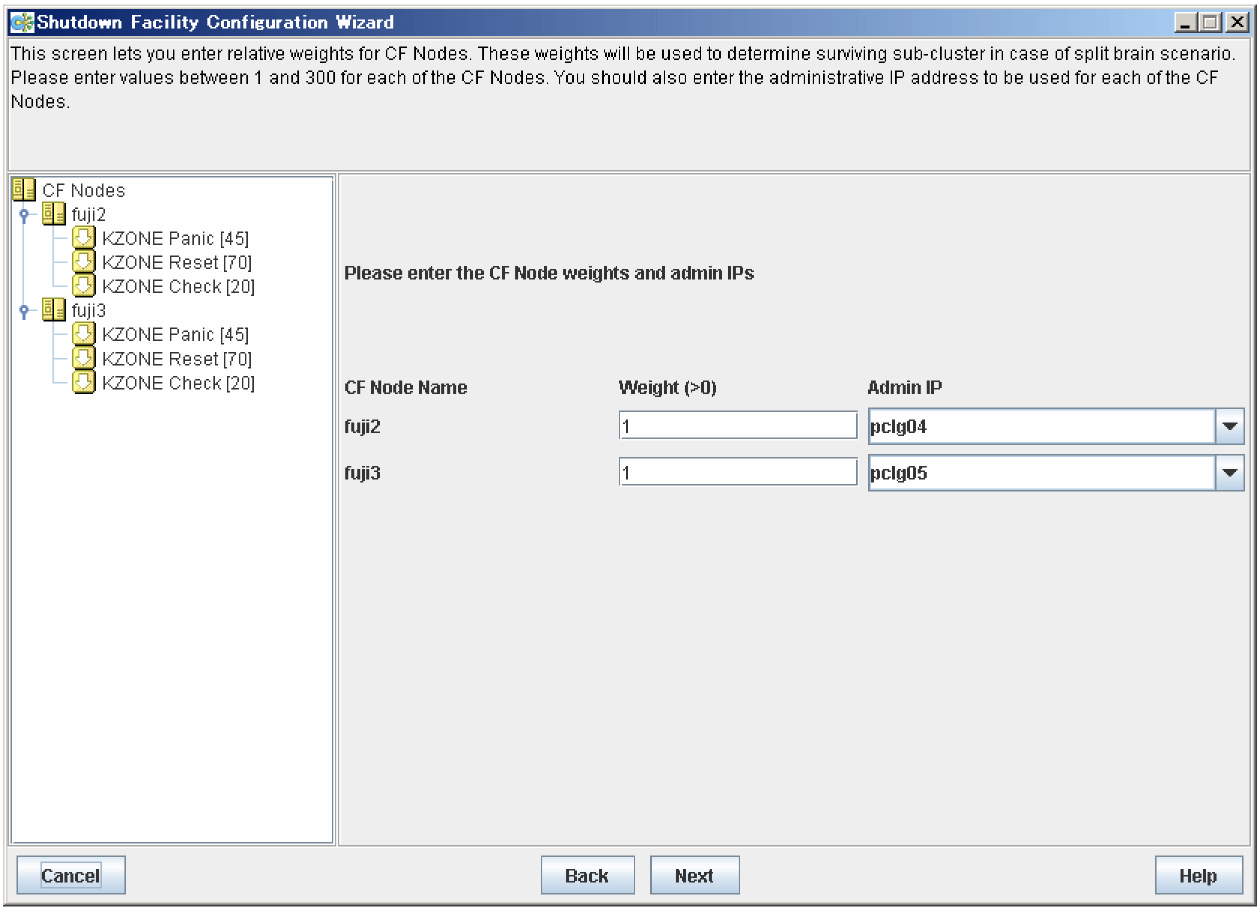
Enter a password to log in to ILOM.

Upon the completion of configuration, click Next.

Entering node weights and administrative IP addresses

The screen for entering the weights of the nodes and the IP addresses for the administrative LAN will appear.

Figure 5.25 Entering node weights and administrative IP addresses



Enter the weights of the nodes and the IP addresses for the administrative LAN.

Weight

Enter the weight of the node that constitutes the cluster. Weight is used to identify the survival priority of the node group that constitutes the cluster. Possible values for each node range from 1 to 300.

For details on survival priority and weight, refer to the explanations below.

Admin IP

Enter an IP address directly or click the tab to select the host name that is assigned to the administrative IP address.

Available IP addresses are IPv4 and IPv6 addresses.

IPv6 link local addresses are not available.

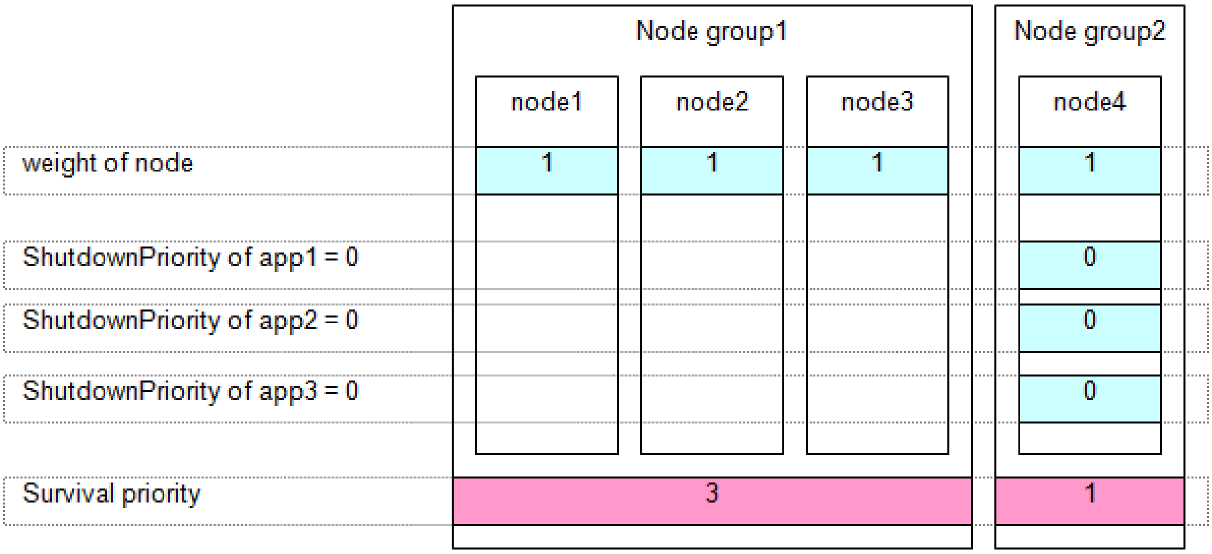
Upon the completion of configuration, click Next.

Survival priority

The typical scenarios that are implemented are shown below:

[Largest node group survival]

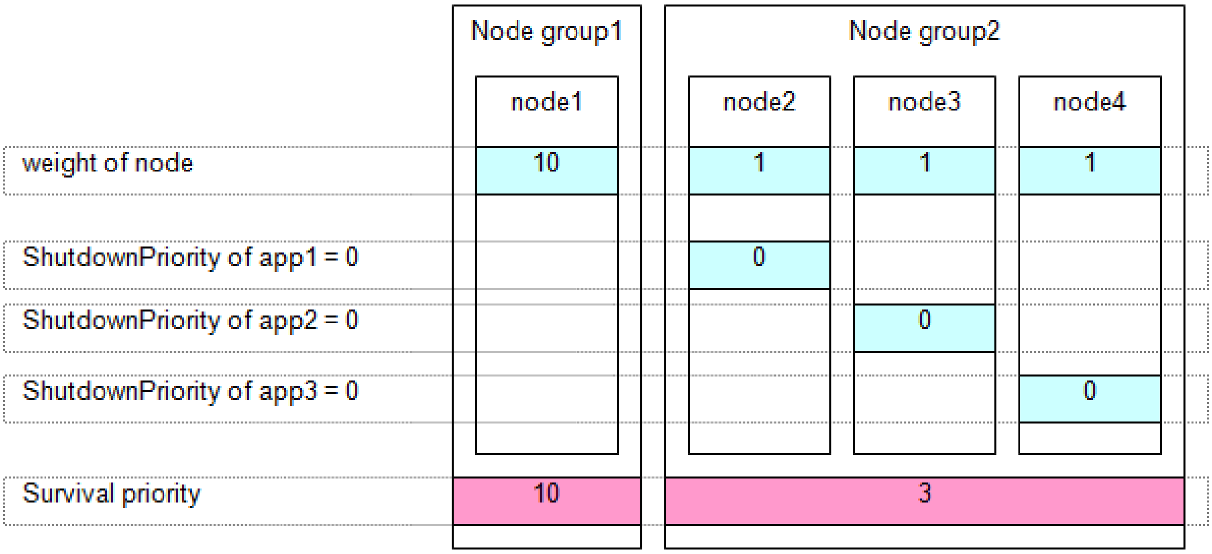
- Set the weight of all nodes to 1 (default).
- Set the attribute of ShutdownPriority of all user applications to 0 (default).



[Specific node survival]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes.
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

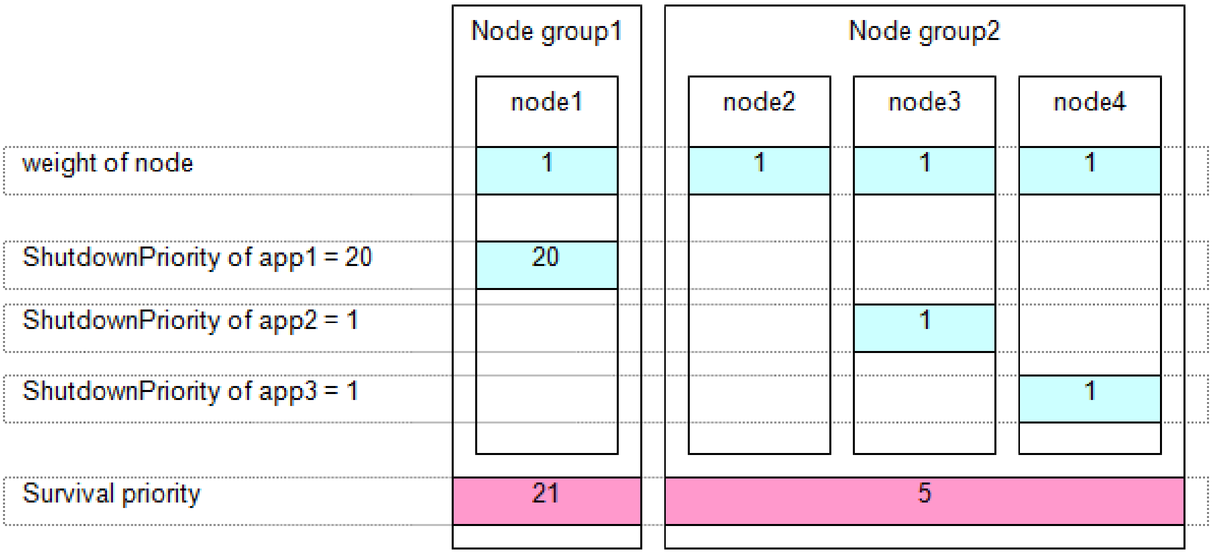
In the following example, node1 is to survive:



[Specific application survival]

- Set the weight of all nodes to 1 (default).
- Set the ShutdownPriority attribute of the user application whose operation is to continue to a value more than double the total of the ShutdownPriority attributes of the other user applications and the weights of all nodes.

In the following example, the node for which app1 is operating is to survive:



[Combination of the cluster system between control domains or guest domains and the cluster system among Kernel Zones for specific control domain survival]

- Set the "weight" the nodes to a power of 2 (1,2,4,8,16,...) in ascending order of the survival priority on each cluster system.
- The order relation of "weight" set for Kernel Zones must be the same as the corresponding control domains.

  For example, if the survival priority of host1 is higher than that of host2 between control domains, the survival priority of node1 (corresponding to host1) must be higher than those of node2 to 4 (corresponding to host2) between Kernel Zones.
- Set the attribute of ShutdownPriority of all user applications to 0 (default).

In the following example, nodes are to survive in the order of node1, node2, node3, and node4.

Figure 5.26 Building Kernel Zones on the control domain

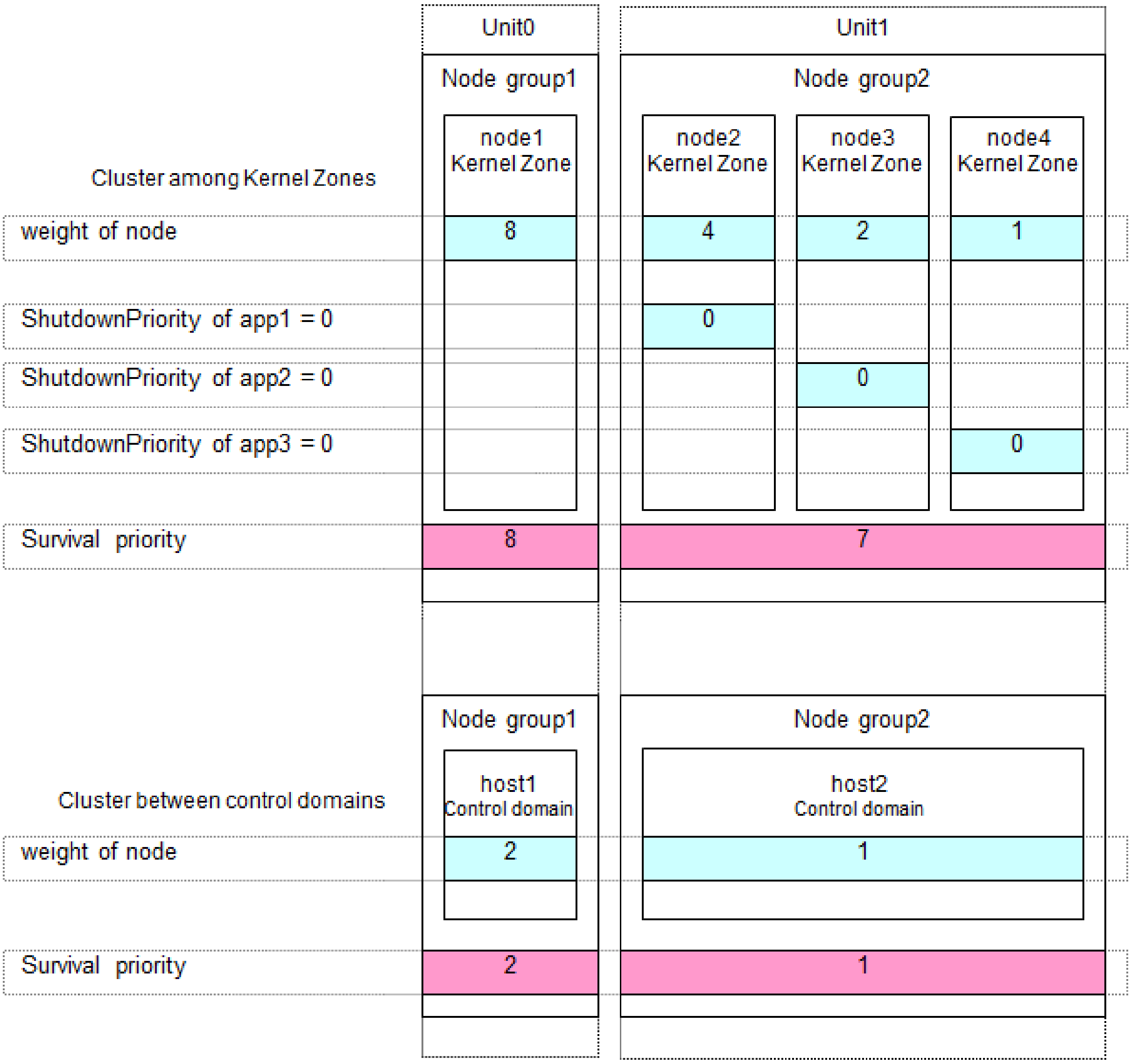
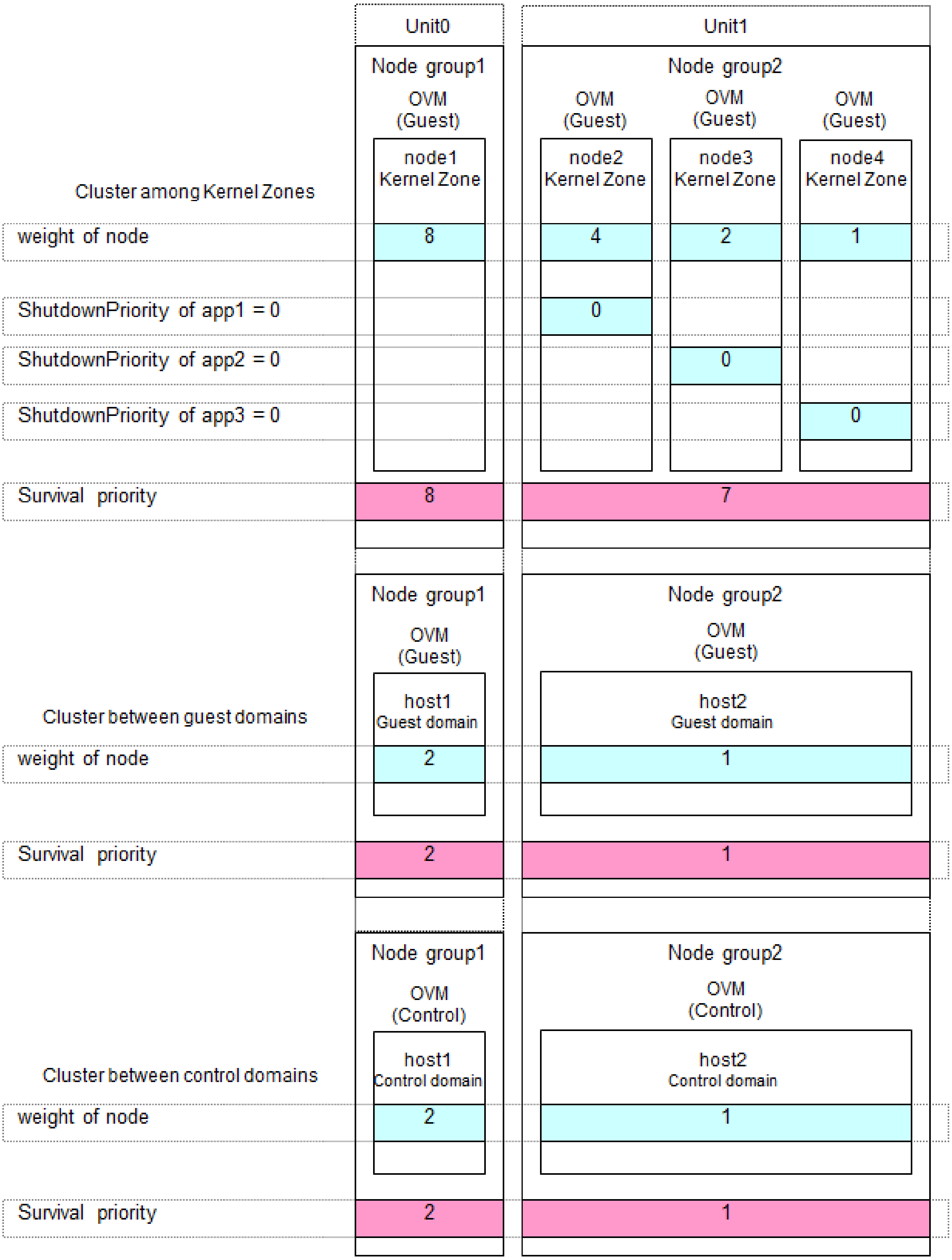


Figure 5.27 Building Kernel Zones on the guest domains

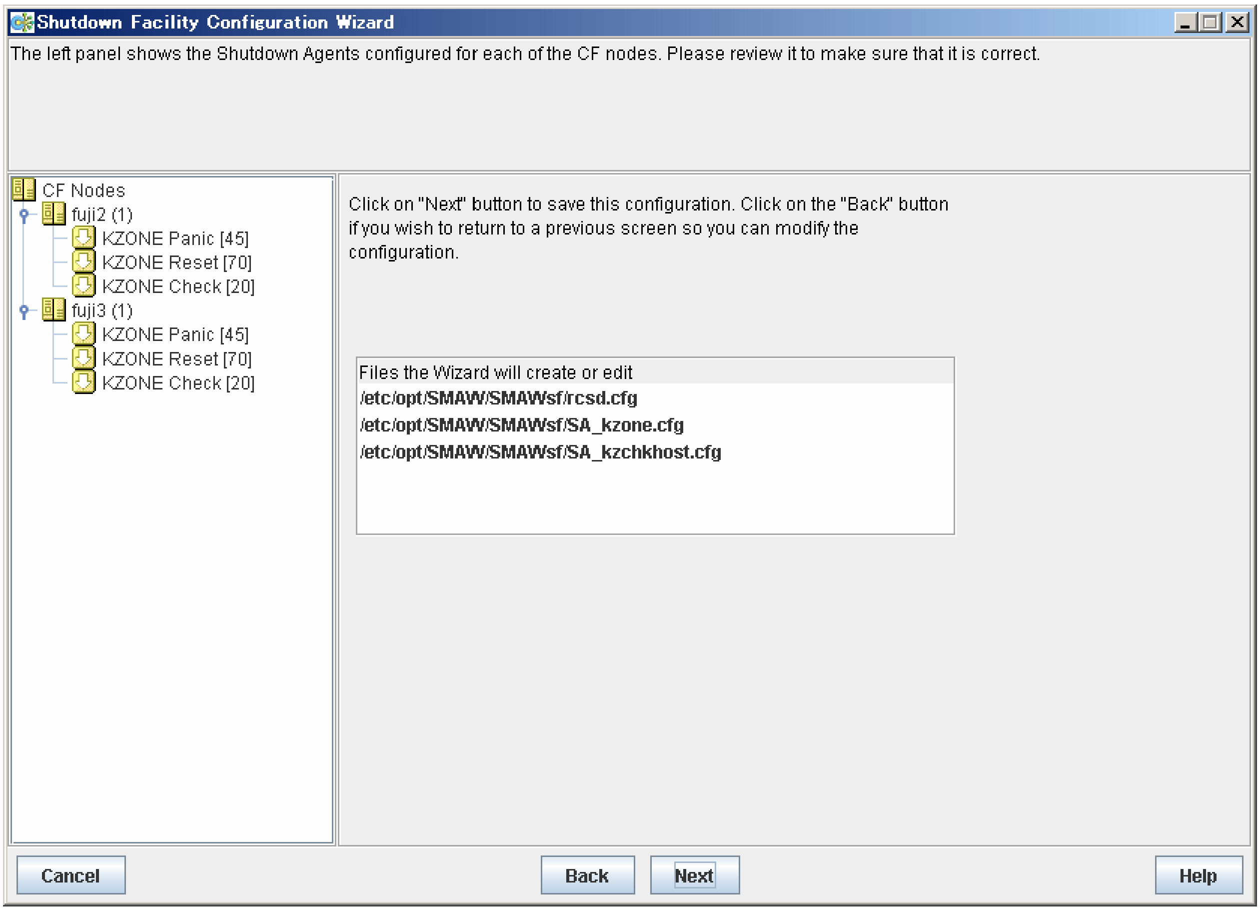


Saving the configuration

Confirm and then save the configuration.

In the left-hand panel of the window, those nodes that constitute the cluster are displayed (SF node weight is displayed in brackets after those nodes), as are the shutdown agents that are configured for each node.

Figure 5.28 Saving the configuration



Click Next. A popup screen will appear for confirmation.

Select Yes to save the setting.

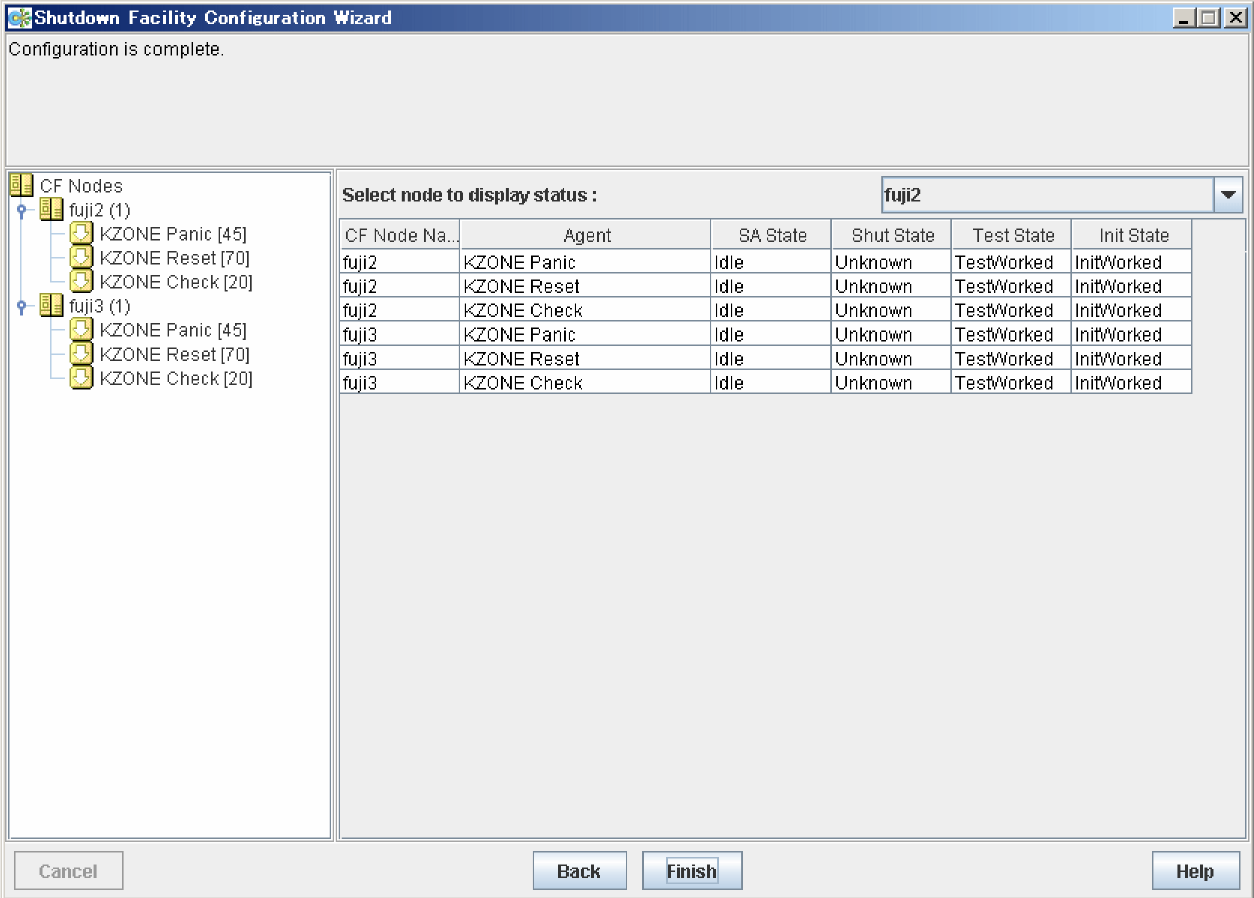
Displaying the configuration of the shutdown facility

If you save the setting, a screen displaying the configuration of the shutdown facility will appear. On this screen, you can confirm the configuration of the shutdown facility on each node by selecting each node in turn.

Information

You can also view the configuration of the shutdown facility by selecting Shutdown Facility -> Show Status from the Tool menu.

Figure 5.29 Displaying the state of configuration of shutdown facility



Shut state

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click Finish. Click Yes in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

- If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the agent or hardware configuration may contain an error. Check the /var/adm/messages file and the console for an error message. Then, apply appropriate countermeasures as instructed the message that is output.

5.1.3 Initial Setup of the Cluster Resource Management Facility

This section explains how to set up the resource database that the cluster resource management facility (CRM) manages.

Set up the CRM resource database according to the following procedure:

1. Initial setup

   Set up the resource database that CRM manages.

2. Automatic configure

   Register the connected hardware devices (shared disks, line switching units, and network interface cards) to the resource database that CRM manages.

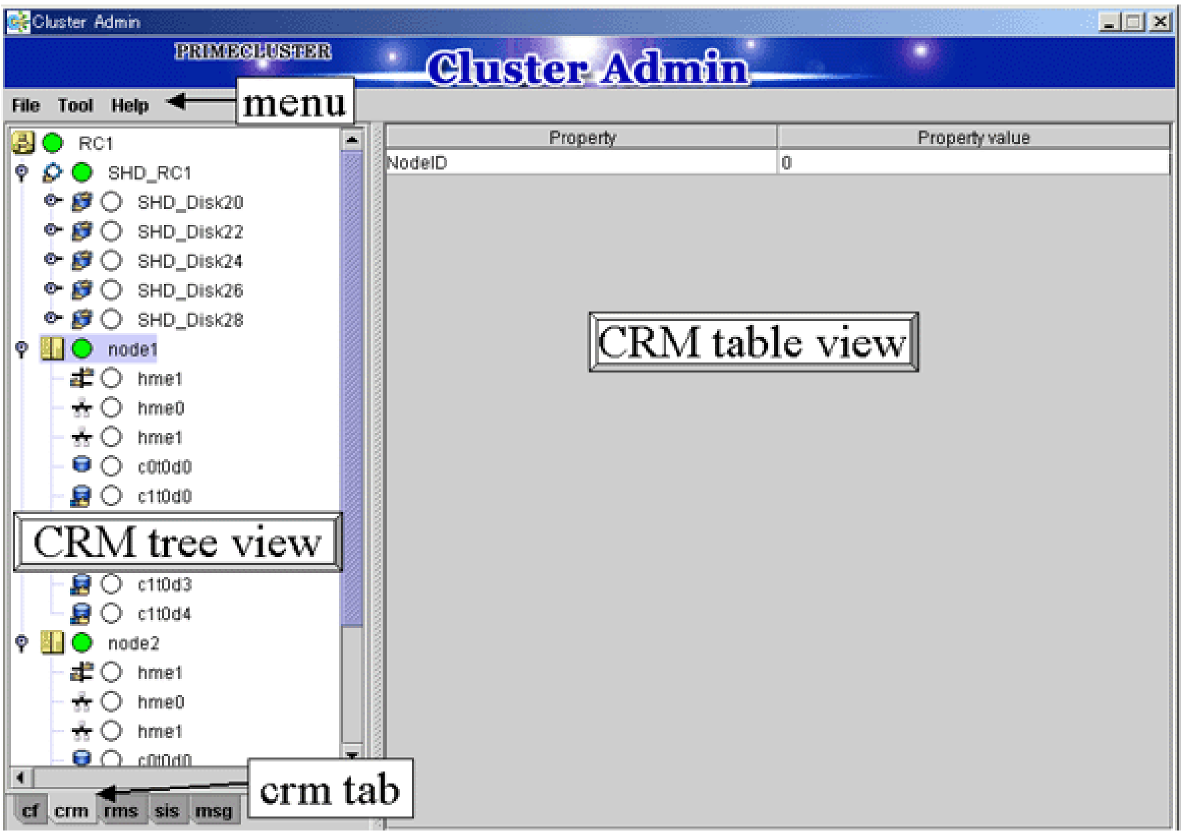
Note

You can check the NIC (network interface card) to be registered by executing the "ifconfig(1M)" command in the plumb up state.

Set up the CRM resource database from the CRM main window. Use the CRM main window as follows:

Operation procedure

1. Select *PRIMECLUSTER -> Global Cluster Services -> Cluster Admin* in the *Web-Based Admin View operation* menu.
2. When the "Cluster Admin" screen is displayed, select the *crm* tab.



The areas shown in the screen are described below.

Menu bar

   This area displays the menu. See "7.1.2.1.3 Operations".

CRM tree view

   This area displays the resources registered to CRM. The resources are displayed in a tree structure.
   For details on the colors and status of the icons displayed in the tree, see "7.1.2.1 Displayed Resource Types".

CRM table view

   This area displays attribute information for the resource selected in the CRM tree view. For information on the displayed information, see "7.1.2.2 Detailed Resource Information".

5.1.3.1 Initial Configuration Setup

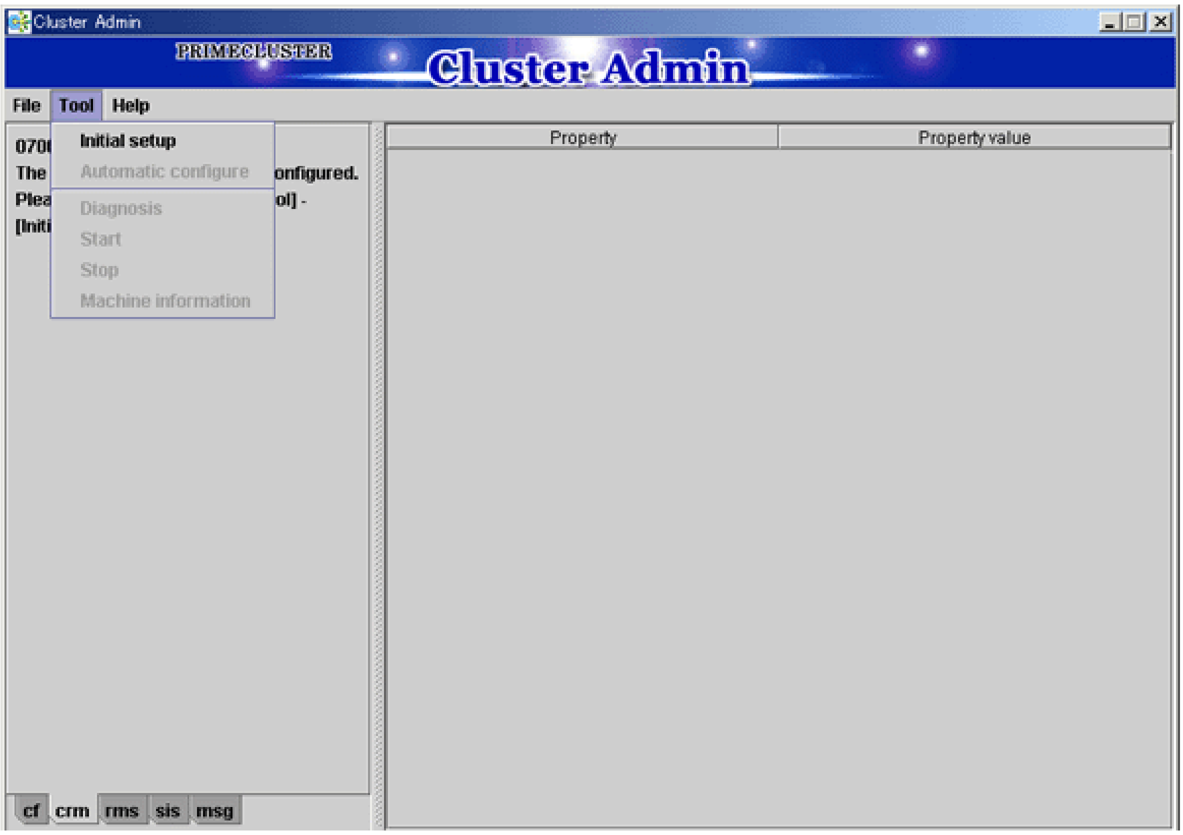
Set up the resource database that CRM manages.

When setting up the initial configuration, make sure that all nodes in the cluster have been started and that CF configuration is completed.

Operation procedure

1. Select the *Initial setup* in the *Tool* menu.

Figure 5.30 Screen for cluster resource management facility

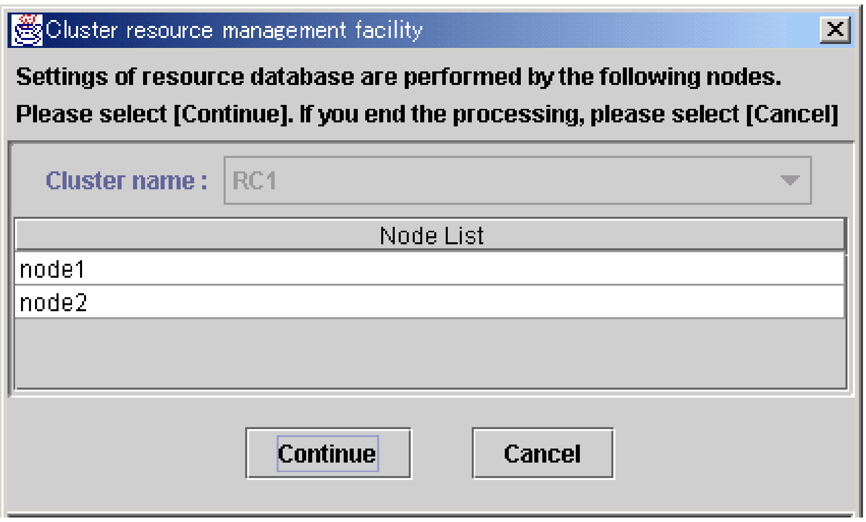


Note

The *Initial setup* can be selected only if the resource database has not been set.

2. The screen for initial setup is displayed.

Figure 5.31 Screen for initial setup



Cluster name

This area displays the names of the clusters that make up the resource database. The cluster names displayed here were defined during CF configuration.

Node List

This area displays the list of the nodes that make up the resource database.

Note

Check that the nodes that were configured in the cluster built with CF and the nodes displayed here are the same.

If the nodes do not match, check the following:

- Whether all nodes displayed by selecting the *cf* tab in the Cluster Admin screen are Up.
- Whether Web-Based Admin View is operating in all nodes.

For instructions on checking this, see "4.2.3.2 Confirming Web-Based Admin View Startup".

Continue button

Click this button to set up the resource database for the displayed cluster.

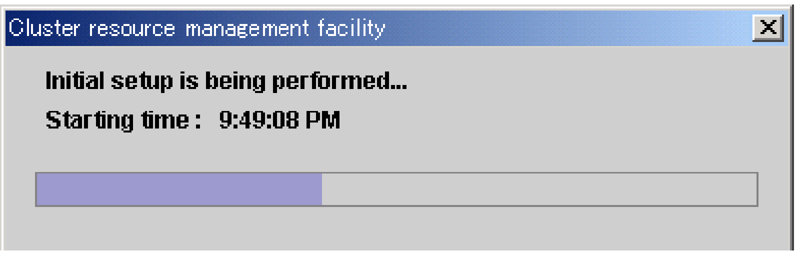
Initial setup is executed on all nodes displayed in the Node list.

Cancel button

Click this button to cancel processing and exit the screen.

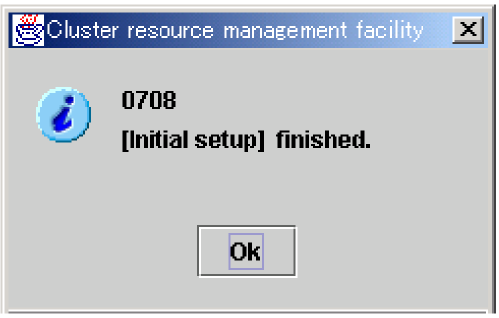
3. Check the displayed contents, and click the *Continue* to start initial setup.

4. The screen below is displayed during execution of initial setup.



5. When initial setup ends, the following message is displayed.

   If *Ok* is selected, the automatic configure screen is displayed. For details on automatic configure, see "5.1.3.2 Automatic Configure".



Note

- If a message appears during operation at the CRM main window, or if a message dialog box entitled "Cluster resource management facility" appears, see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in the "*PRIMECLUSTER Messages*."
- If you want to add, delete, or rename a disk class from the *Global Disk Services* screen after executing Initial Setup from the CRM main window, close the *Cluster Admin* screen.

5.1.3.2 Automatic Configure

The automatic configure process registers the connected hardware devices (shared disks, line switching units, and network interface cards) to the resource database.

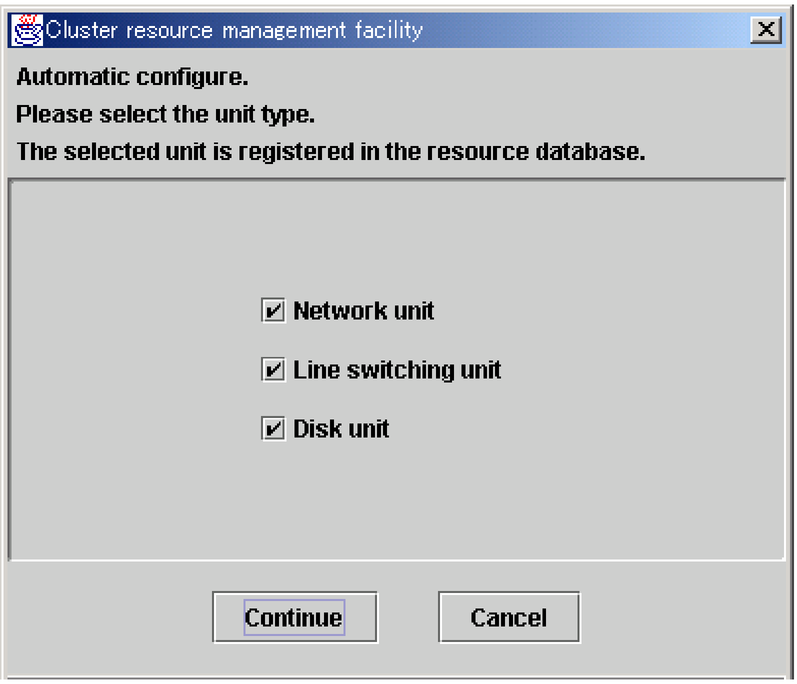
Operation Procedure

1. Display the automatic configure screen by one of the following methods:

   - Respond to the "0708 [*Initial setup*] finished." message, which is displayed after initial setup is completed.

- Select *Tool* -> *Automatic configure* from the CRM tree view on the CRM main window of Cluster Admin while the cluster name displayed at the highest level is selected.

Figure 5.32 Screen for automatic configure



Network unit

Select this item to detect network interfaces automatically and register them to the resource database.

You must select this item to use public LAN takeover.

Line switching unit

Select this item to detect connected line switching units automatically and register them to the resource database.

You must select this item to use the line takeover function.

Do not select this item in an Oracle Solaris 11 environment.

Disk unit

Select this item to detect connected shared disk units automatically and register them to the resource database. You must select this item to use system disk mirroring or to use Global Disk Services (GDS) and Global File Services (GFS) in shared disk unit takeover.

Note

- When you use the multipath disk control function to set shared disk units, the following messages may be displayed on the console, depending on the configuration:

```
WARNING: mphd2048: paths setup information of command is reverse against device one.
WARNING: mphd2049: paths setup information of command is reverse against device one.
```

  These messages indicate that connection of multipath disk that is stored in multipath disk control and actual connection of multipath disk are not identical. But multipath disk control automatically reconfigures the multipath according to the actual connection. You can ignore these messages.

- When specifying a shared disk unit as the hardware for the patrol diagnosis, set up the physical disk name (such as c1t4d0) of a shared disk unit to be the same in all nodes, and then execute the automatic configuration. When the physical disk name of a shared disk unit varies depending on a node, you cannot set a shared disk unit to the hardware for the patrol diagnosis.

Continue button

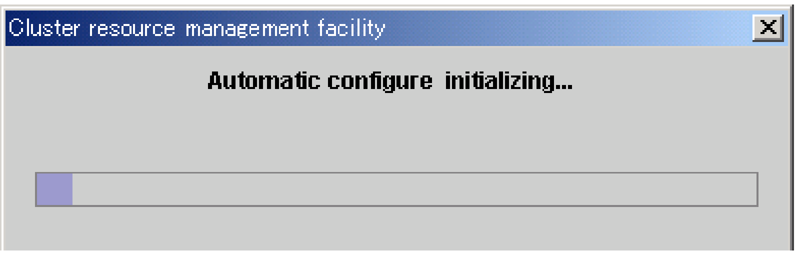
Click this button to detect the selected device type automatically and register it to the resource database.

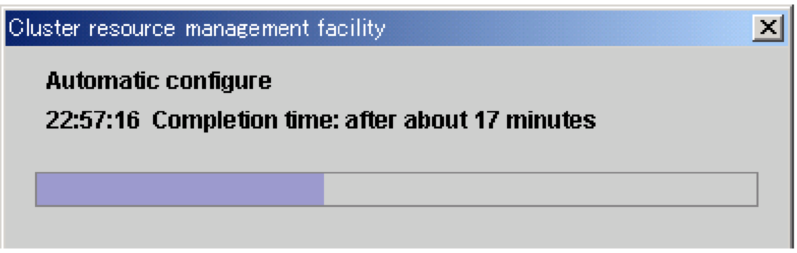
Cancel button

Click this button to cancel processing and exit the screen.

2. Select the device type, and click the *Continue*.

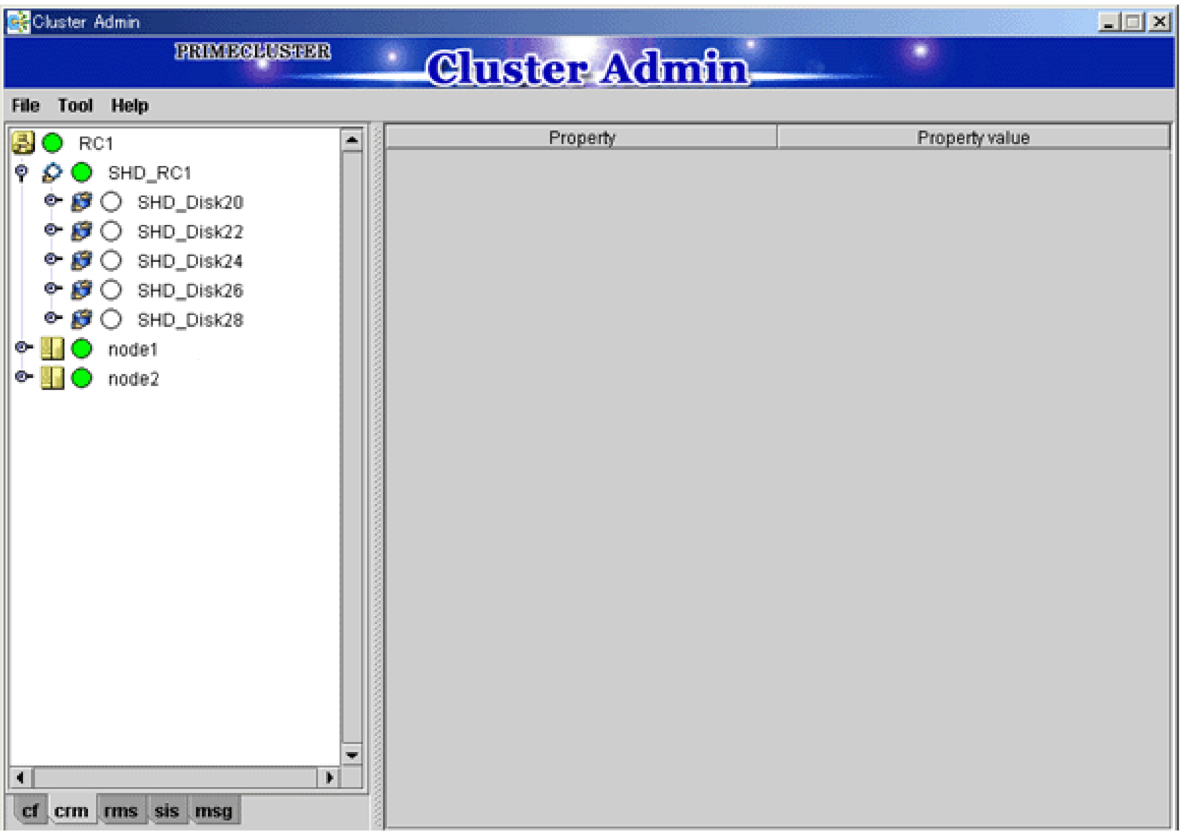
Processing is executed according to the following flow:





3. When automatic configure is completed, the following message is displayed:

4. When the initial setup and automatic configure are completed, the following screen appears.



5. When automatic configuration is completed, go to the CRM main window and confirm that the resource registration is completed by checking the following.

   - Whether the number of shared disk unit resources is less than the actual device configuration.
   - Whether any disks other than shared disk unit are registered as shared disk unit.
   - Whether the number of public LAN resources is less than the actual device configuration.
   - Whether all connected line switching units are displayed.

   If the actual device configuration and the resources do not match each other as described above, automatic detection may have been disabled because of one of the following causes:

   - There is a connection path failure between a host device and a disk array unit.
   - A disk array unit is not ready.
   - A network adapter failed.
   - A network adapter driver failed.
   - A line switching unit is not connected properly.
   - A line switching unit is not turned on.

   If the resources are not registered correctly, first review the above causes. Then in the CRM main window, select the *Automatic configure* in the *Tool* to re-register the resources.

Note

- If a message appears during operation at the CRM main window, or if a message dialog box entitled "Cluster resource management facility" appears, see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in the "*PRIMECLUSTER Messages*."

- If you want to add, delete, or rename a disk class from the *Global Disk Services* screen after executing automatic configuration from the CRM main window, close the *Cluster Admin* screen.

- The following message may be displayed on the console during automatic configuration:

```
FJSVcluster: Error: dcmevmd: 6000: An internal error occurred.
(function:_evm_res_all_callback detail:0x9060b00-0x4-0-0)
```

If this message is displayed, all the hardware devices registered to the resource database may not be displayed on the CRM tree view on the CRM main window. Restart the Cluster Admin screen and open the CRM main window again.

5.2 Setting Up Power Supply Linkage

If you want to set up power supply linkage, see the "*System Operation and Administration Guide*".

5.3 Setting Up Shared Disk Connection Confirmation

Shared disk connection confirmation is a function that performs the following:

- Check that the shared disk is powered.
- Check that the cable is properly connected.
- Detect new shared disks.

Note

If the paths of a disk array unit are multiplexed, and the paths are controlled by a multipath disk control function, shared disk connection confirmation will not detect an error even if an error occurs in part of the multiplexed paths.

If an error is detected in a shared disk unit or if a new shared disk unit is detected, shared disk connection confirmation outputs a message to syslogd(1M) and the CRM main window.

This function allows you to specify whether or not RMS activation is to be suppressed in the node in which the failure was detected. If a failure is detected in a shared disk unit, the userApplication might not operate normally and user resources in that shared disk unit might become damaged. It is therefore recommended that you specify that RMS activation should be suppressed if a failure is detected. After PRIMECLUSTER is installed, shared disk connection confirmation is initially disabled by default. To change this, follow the procedures described below.

Note

- Set up shared disk connection confirmation after the initial setup of the Cluster Resource Management facility is completed.
- To view the manual pages of each command, add "/etc/opt/FJSVcluster/man" to the MANPATH variable.

Enabling shared disk connection confirmation

Take the following steps on each node. Shared disk connection confirmation will be executed from the next node startup.

RMS activation is to be suppressed if a failure is detected

1. Add the following to the "RELIANT_INITSCRIPT" environment variable of RMS:

   ```
   /etc/opt/FJSVcluster/bin/clinitscript
   ```

   Example: Add the following to /opt/SMAW/SMAWRrms/bin/hvenv.local.

   ```
   export RELIANT_INITSCRIPT=/etc/opt/FJSVcluster/bin/clinitscript
   ```

2. Execute the "clsetacfparam(1M)" command.

   Example:

   ```
   # /etc/opt/FJSVcluster/bin/clsetacfparam -p auto
   ```

   Note

   If there is an error in the shared disk unit, it is detected by the registered script (/etc/opt/FJSVcluster/bin/clinitscript) and RMS stops. In this case, the following RMS message will be sent to syslogd(1M):

   "(INI, 12): FATAL ERROR: InitScript returned non-zero exit code 1."

RMS activation is not to be suppressed if a failure is detected

1. Execute the "clsetacfparam(1M)" command.

   Example:

   ```
   # /etc/opt/FJSVcluster/bin/clsetacfparam -p auto
   ```

Disabling shared disk connection confirmation

To cancel the setting previously made for "Enabling shared disk connection confirmation," perform the following procedure on all the nodes. After you next boot the node, shared disk connection will not be confirmed.

"Suppress RMS activation if an error is detected" was previously configured

1. Delete the following from the RELIANT_INITSCRIPT environment variable of RMS.

   ```
   /etc/opt/FJSVcluster/bin/clinitscript
   ```

   Example: Delete the following from /opt/SMAW/SMAWRrms/bin/hvenv.local.

   ```
   export RELIANT_INITSCRIPT=/etc/opt/FJSVcluster/bin/clinitscript
   ```

2. Execute the "clsetacfparam(1M)" command.

   Example:

   ```
   # /etc/opt/FJSVcluster/bin/clsetacfparam -p none
   ```

"Do not suppress RMS activation if an error is detected" was specified

1. Execute the "clsetacfparam(1M)" command.

   Example:

   ```
   # /etc/opt/FJSVcluster/bin/clsetacfparam -p none
   ```

See

- For details on the RMS environment variables, see "13 Appendix - Environment Variables" in the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide".
- For details on the "clsetacfparam(1M)" command, see the manual page for clsetacfparam(1M).

5.4 Setting Up Fault Resource Identification and Operator Intervention Request

The fault resource identification is a function that outputs a message to syslogd(1M) and Cluster Admin and a history of failed resources to Resource Fault History if a failure occurs in a resource or node that is registered to a cluster application.

After setting the initial configuration of the resource database, this function enables fault resource identification and operator intervention request. An example of a message displayed by fault resource identification is shown below.

```
6750 A resource failure occurred.  SysNode:node1RMS userApplication:app0 Resource:apl1
```

The operator intervention request function displays a query-format message to the operator if a failed resource or a node in which RMS has not been started is found when a cluster application is started. The messages for operator intervention requests are displayed to syslogd(1M) and Cluster Admin.

```
1421 The userApplication "userApplication" did not start automatically because not all of the nodes
where it can run are online.
Forcing the userApplication online on the SysNode "SysNode" is possible.
Warning: When performing a forced online, confirm that RMS is started on all nodes in the cluster,
manually shutdown any nodes where it is not started and then perform it. For a forced online, there is
a risk of data corruption due to simultaneous access from several nodes. In order to reduce the risk,
nodes where RMS is not started maybe forcibly stopped.
Are you sure wish to force online? (no/yes) Message No: 1001
```

See

For details on the messages displayed by the fault resource identification function and the messages displayed by the operator intervention request function, see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in the "*PRIMECLUSTER Messages*."

Note

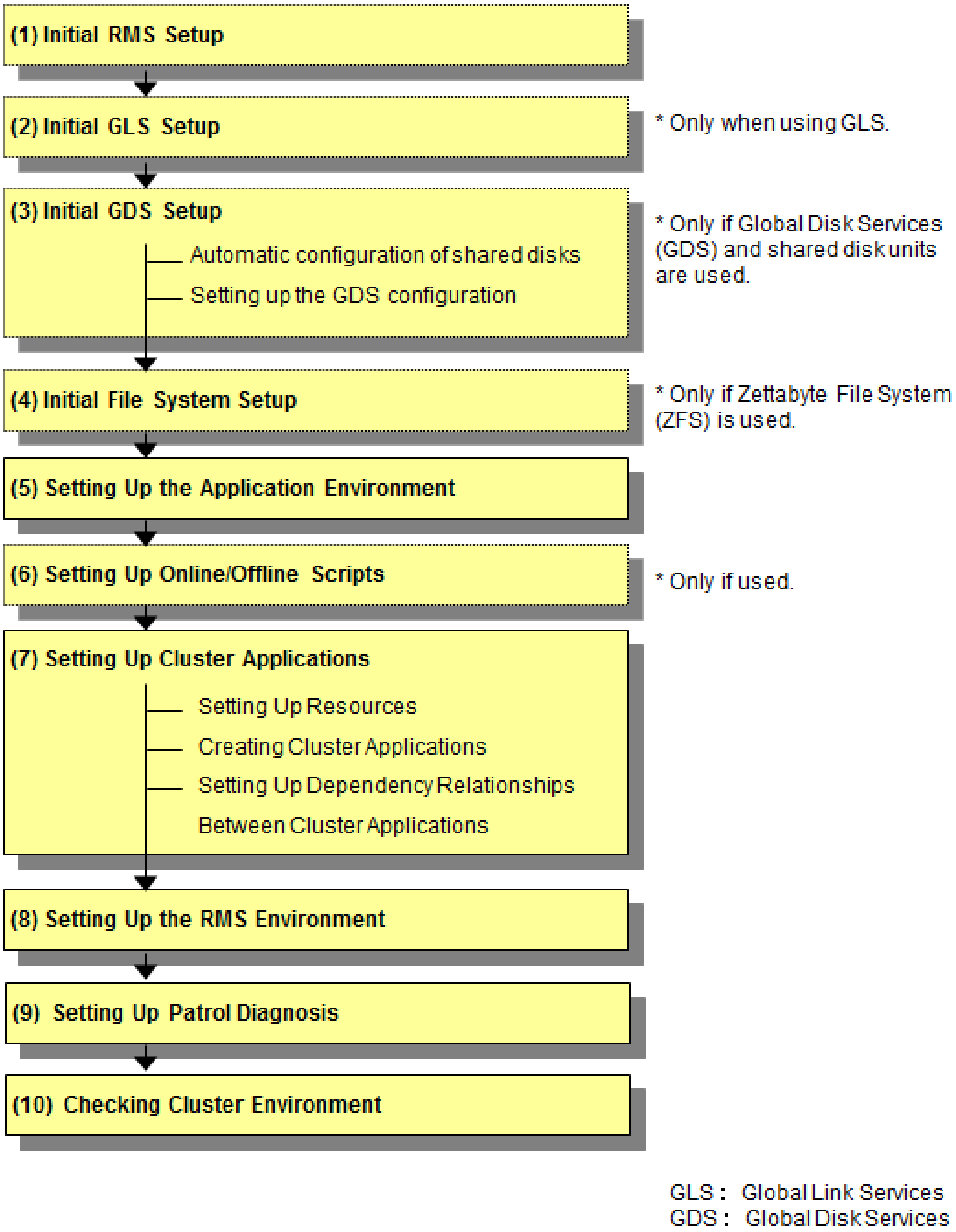
To view the manual pages of each command, add "/etc/opt/FJSVcluster/man" to the MANPATH variable.

Changing the operation settings of fault resource identification and operator intervention request

Change the settings with the clsetparam command. For details, see the manual page for clsetparam.

Chapter 6 Building Cluster Applications

The procedure for building a cluster application is shown below.



Note

When using RMS, make sure to configure the cluster application.

Table 6.1 Application building procedure and manual reference locations

Work item		Execution Nodes	Required/ optional	Manual reference location*
(1)	6.1 Initial RMS Setup	All nodes	Required	**RMS** "8.1.2 Starting RMS automatically at boot time"
(2)	6.2 Initial GLS Setup	All nodes	Optional (required when GLS is used)	GLSR GLSM
(3)	6.3.1 Automatic Configuration of Shared Disks	All nodes	Optional	**CF** "4.4.3 Automatic resource registration"
	6.3.2 GDS Configuration Setup	All nodes	Optional	**GDSG** "Chapter 5 Operations"
(4)	6.4 Initial File System Setup	All nodes	Optional (required when ZFS is used)	
(5)	6.5 Setting Up the Application Environment	All nodes	Required	Manuals for each application
(6)	6.6 Setting Up Online/Offline Scripts	All nodes	Optional	**RMS** "2.9 Environment variables", "13 Appendix - Environment variables"
(7)	6.7 Setting Up Cluster Applications	All nodes	Required	"Solaris X Reference Manual Collection," "Solaris Answer Book"
	6.7.1 Setting Up Resources			
	6.7.2 Creating Cluster Applications			
	6.7.3 Setting Up Dependency Relationships Between Cluster Applications			
(8)	6.8 Setting Up the RMS Environment	All nodes	Required	**RMS** "2.9 Environment variables", "13 Appendix - Environment variables"
(9)	6.9 Setting Up Patrol Diagnosis	All nodes	Required	
(10)	6.10 Checking the Cluster Environment	All nodes	Required	

***** The names of the reference PRIMECLUSTER manuals and function name are abbreviated as follows:

- **RMS:** *PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*
- **CF:** *PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*
- **GDSG:** *PRIMECLUSTER Global Disk Services Configuration and Administration Guide*
- **GLSR:** *PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*
- **GLSM:** *PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function*
- **GDS:** Global Disk Services
- **GLS:** Global Link Services

6.1 Initial RMS Setup

When using RMS, you need to check "A.7.1 RMS Setup Worksheet" and change the following environment variable as required.

- Automatic startup of RMS (HV_RCSTART)

The default value is "Start up automatically" in this version.

If you want to set RMS not to be started automatically along with node startup, select "Does not start up automatically."

We recommend that you set "Start up automatically" when using RMS.

See

For information on how to check and change the environment variables of RMS automatic startup, see "8.1.2 Starting RMS automatically at boot time" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

6.2 Initial GLS Setup

This section outlines the steps for configuring Global Link Services (GLS).

6.2.1 GLS Setup

For information on the initial GLS setup, see "Chapter 5 Operation on Cluster System" in the "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*" and "Chapter 7 Administration on a Cluster System" in the "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function*."

This section describes how to set up "the Single system without NIC sharing of the NIC switching mode (IPv4)" that GLS (redundant line control function) provides. This procedure is described in the example below.

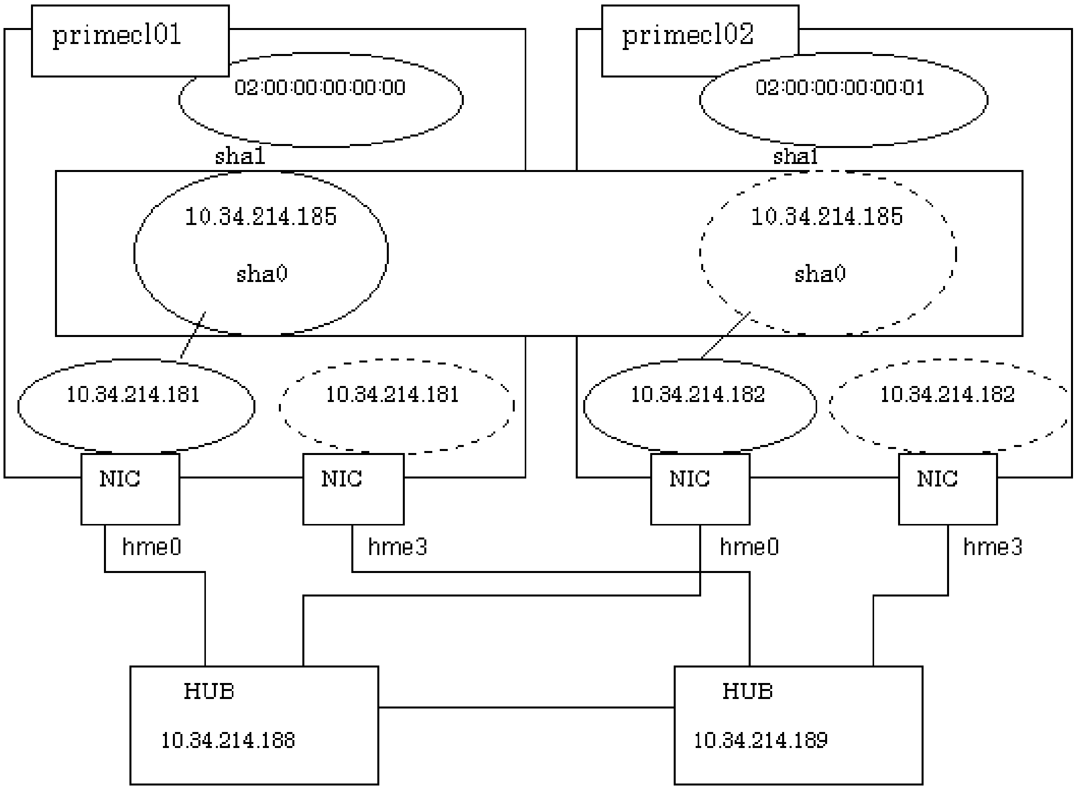
Interface names listed in examples of setting up vary depending on the environment. Replace interface names according to the environment. For Solaris 11 OS, the default interface name is netX (X means the instance number).

Note

You can use PRIMECLUSTER "takeover network" and GLS "IP address takeover" together in a same cluster system; however, you cannot configure them on the same interface. If this occurs, communication using takeover IP addresses is disenabled.

For example, if you select hme1 as the interface when you set up the PRIMECLUSTER "takeover network," do not set hme1 in the GLS environment settings (do not specify hme1 using "-t" option of the "hanetconfig create" command).

The setup values correspond to those on the "A.7.2 GLS Setup Worksheet."



Operation Procedure:

If the operating node is [HOST-primecl01]

1. Setting up the system

    1. Define the IP address and Host name in /etc/inet/hosts file.

```
10.34.214.185   takeoverIP   # Virtual IP
10.34.214.181   primecl01    # primecl01 Physical IP
10.34.214.182   primecl02    # primecl02 Physical IP
10.34.214.188   swhub1       # primary HUB IP
10.34.214.189   swhub2       # secondary HUB IP
```

    2. Setting up interface to use.

        - For Solaris 10

          Add the host names from above definitions to the /etc/hostname.hme0 file.

          Contents of /etc/hostname.hme0

```
primecl01
```

        - For Solaris 11

          Set up with the interface and the host name defined above using ipadm(1M) command.

          Interface net0 (corresponding to hme0 in the figure)

```
# /usr/sbin/ipadm create-ip net0
# /usr/sbin/ipadm create-addr -T static -a primecl01/24 net0/v4
```

    3. Define the subnet masks in the /etc/inet/netmasks file.

```
10.0.0.0    255.255.255.0
```

2. Rebooting

Run the following command and reboot the system. After rebooting the system, verify the interface set using ifconfig command is enabled.

```
# /usr/sbin/shutdown -y -i6 -g0
```

3. Creating of virtual interface

For the underlined parameter, specify the physical IP address of the node.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.34.214.185 -e 10.34.214.181 -t hme0,hme3
```

Check that the virtual interface has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.181     hme0,hme3

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+------------------------+
```

4. Setting up the HUB monitoring function

For the underlined parameter, specify the IP addresses of the hubs to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+---------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
         hub-hub poll       =    OFF
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES
```

5. Setting up the Standby patrol monitoring function

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]
```

```
 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.181     hme0,hme3
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+-------------------------+
```

6. Creating the takeover virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

Check that the registration has been done correctly.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+--------------------------------------------------+
 sha0:65    10.34.214.185    -
```

If the standby node is [HOST-primecl02]

1. Setting up the system

    1. Define the IP address and Host name in /etc/inet/hosts file. Defined content is same as HOST-primecl01.

    2. Setting up interface to use.

        - For Solaris 10

          Add the host names from above definitions to the /etc/hostname.hme0 file.

          Contents of /etc/hostname.hme0

          ```
          primecl02
          ```

        - For Solaris 11

          Set up with the interface and the host name defined above using ipadm(1M) command.

          Interface net0 (corresponding to hme0 in the figure)

          ```
          # /usr/sbin/ipadm create-ip net0
          # /usr/sbin/ipadm create-addr -T static -a primecl02/24 net0/v4
          ```

    3. Define the subnet masks in the /etc/inet/netmasks file.

       ```
       10.0.0.0    255.255.255.0
       ```

2. Rebooting

   Run the following command and reboot the system. After rebooting the system, verify the interface set using ifconfig command is enabled.

   ```
   # /usr/sbin/shutdown -y -i6 -g0
   ```

3. Creating of virtual interface

   For the underlined parameter, specify the physical IP address of the node.

   ```
   # /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.34.214.185 -e 10.34.214.182 -t hme0,hme3
   ```

   Check that the virtual interface has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.182     hme0,hme3

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+---------------------------+
```

4. Setting up the HUB monitoring function

   For the underlined parameter, specify the IP addresses of the hubs to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

   Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+---------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
         hub-hub poll       =    OFF
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES
```

5. Setting up the Standby patrol monitoring function

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

   Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.182     hme0,hme3
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+---------------------------+
```

6. Creating the takeover virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

Check that the resources have been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+---------------+--------------------------------------------------+
 sha0:65    10.34.214.185    -
```

Post-setup processing

After the OPERATING and STANDBY node setup is done, create the Gls resources, and register them to the cluster application.

For details, see "6.7.1.4 Creating Gls Resources" and "6.7.2 Creating Cluster Applications."

See

For information on GLS (redundant line control function) and other operation modes, see "Appendix B Examples of Setting Up" in the "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*."

6.2.2 Setting Up Web-Based Admin View When GLS is Used

To use Web-Based Admin View in a network that was made redundant with GLS, you must set up Web-Based Admin View after setting up the NIC switching mode or the fast switching mode.

See

For setup details, see "2.3 Setup with GLS" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

6.3 Initial GDS Setup

Take the following steps to configure Global Disk Services (GDS). If shared disk units are to be used, you are recommended to use GDS to manage the shared disk units.

Note

If you plan to add, delete, or rename a disk cluster from the Global Disk Services screen, close the Cluster Admin screen before starting the operation.

6.3.1 Automatic Configuration of Shared Disks

Before setting up GDS, you must have configured the shared disks using automatic configuration. If the disk units were already configured with automatic configuration in "5.1.3.2 Automatic Configure," skip this operation.

Note

If the following procedures are taken, execute automatic configuration of disk units before setting up GDS.

- Disk units are not set up for automatic configuration during the initial setup
- Shared disk unit was added after the initial configuration
- Disk labels are changed

6.3.1.1 Executing Automatic Configuration

After setting up the disk units according to the procedure described in "3.2.2 Setting Up Disk Units," either execute *Automatic configure* from the Tools menu of the CRM main window in Cluster Admin, or execute the following command in any one of the nodes in the cluster system:

```
# /etc/opt/FJSVcluster/bin/clautoconfig -r
```

Note

If disk units were specified during automatic resource registration in "5.1.3.1 Initial Configuration Setup," the resources have already been created. If you want to check whether the disk units were registered as resources, you can reexecute automatic resource registration.

See

For details, see the manual page for the "clautoconfig(1M)" command or "4.4.3 Automatic resource registration" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

6.3.2 GDS Configuration Setup

The GDS setup operations are described below.

These setup operations can be performed in any sequence.

- Setting Up System Disk Mirroring

  Set up system disk mirroring by one of the following methods corresponding to the boot environment of the system.

  - 6.3.2.1 Setting Up System Disk Mirroring in the ZFS Boot Environment

  - 6.3.2.2 Setting Up System Disk Mirroring in the UFS Boot Environment

- Setting Up Shared Disks

  Set up the shared disk volumes according to 6.3.2.3 Setting Up Shared Disks.

Note

When one or more single nodes for which the class has already been created are converted to the cluster system, the setting of GDS is needed. For details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

6.3.2.1 Setting Up System Disk Mirroring in the ZFS Boot Environment

If mirroring a system disk under a ZFS boot environment, perform one of the following setups.

- If mirroring under GDS

- If registering to GDS and mirroring under ZFS

  Refer to "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

The values to be set for each item are to correspond with those provided in "A.8.1 System Disk Mirror Setup Worksheet."

6.3.2.2 Setting Up System Disk Mirroring in the UFS Boot Environment

Take the following setup procedures to enable system disk mirroring in the UFS boot environment.

The setup values correspond to the values that were specified in the "A.8.1 System Disk Mirror Setup Worksheet." In the operation procedure, the "System Disk Mirror Setup Worksheet" is abbreviated as "worksheet."

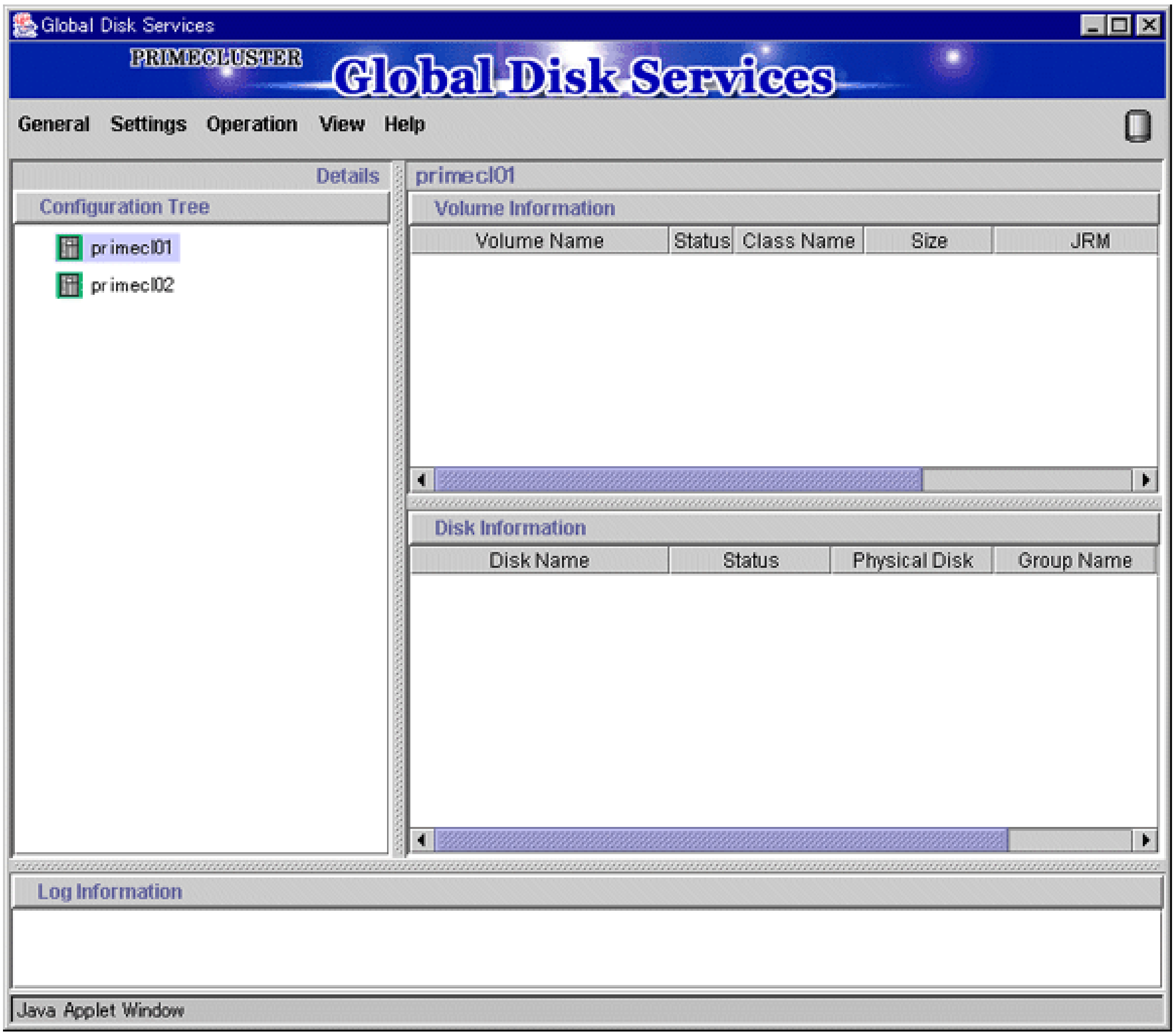
See

For setup details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Operation Procedure:

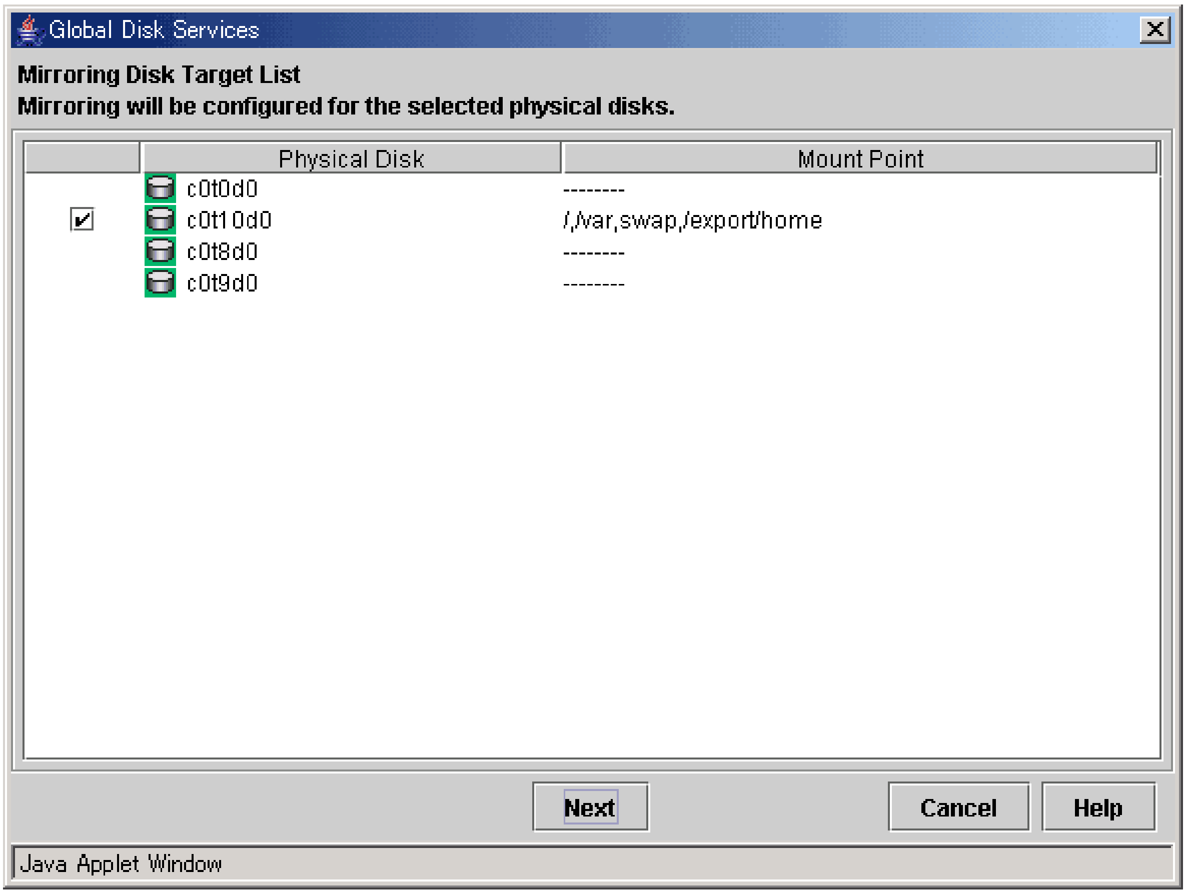
1. Select *Global Disk Services* at the Web-Based Admin View top screen.

   The GDS Management screen (hereinafter main screen) is displayed.



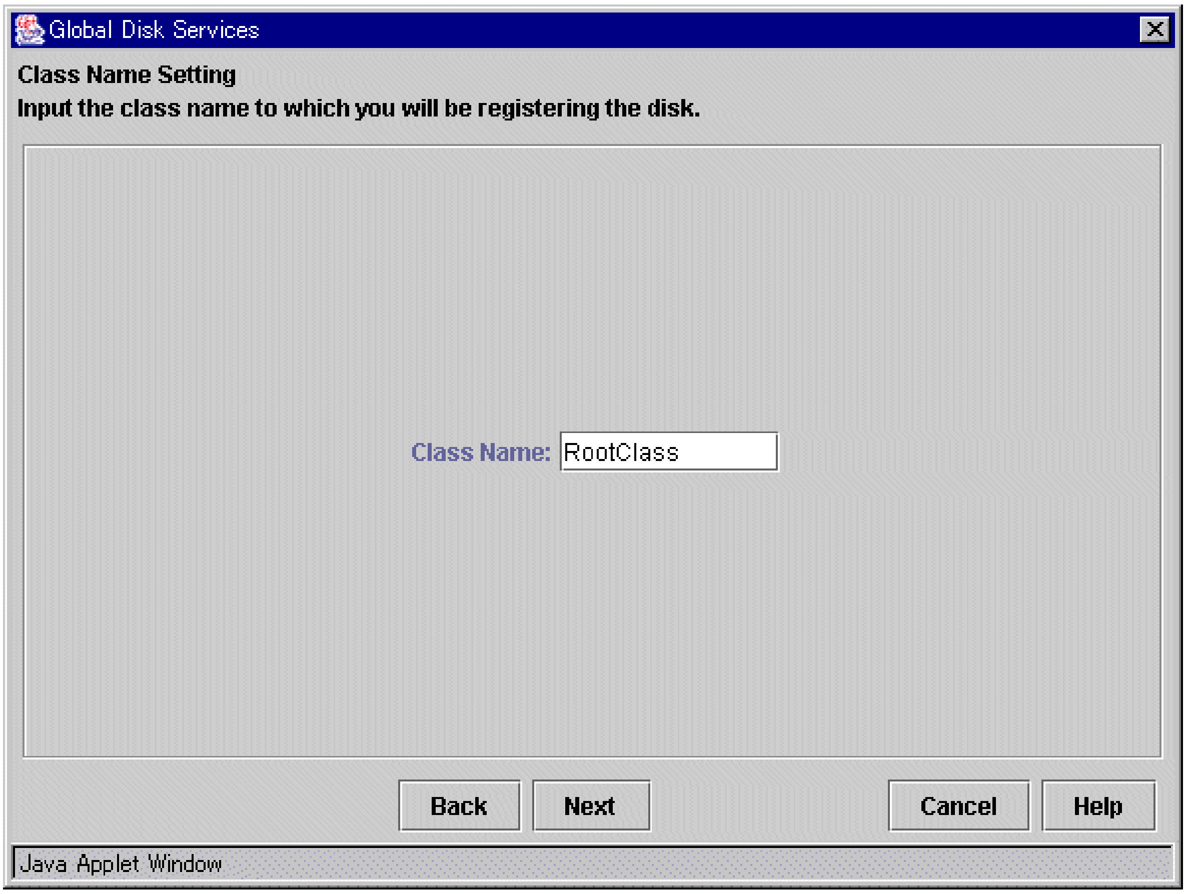
2. From the GDS configuration tree, select the node in which the system disk mirror is to be set, click the *Settings* menu, and select *System Disk Settings*.
A list of disks that can be used for mirrored disks for the selected node is displayed.

Select the system disk ("Physical disk name" on the worksheet), and click *Next*.



3. Specify class name of the root class.

   Enter the class name ("Class name" on the worksheet) of the root class, and click *Next*.

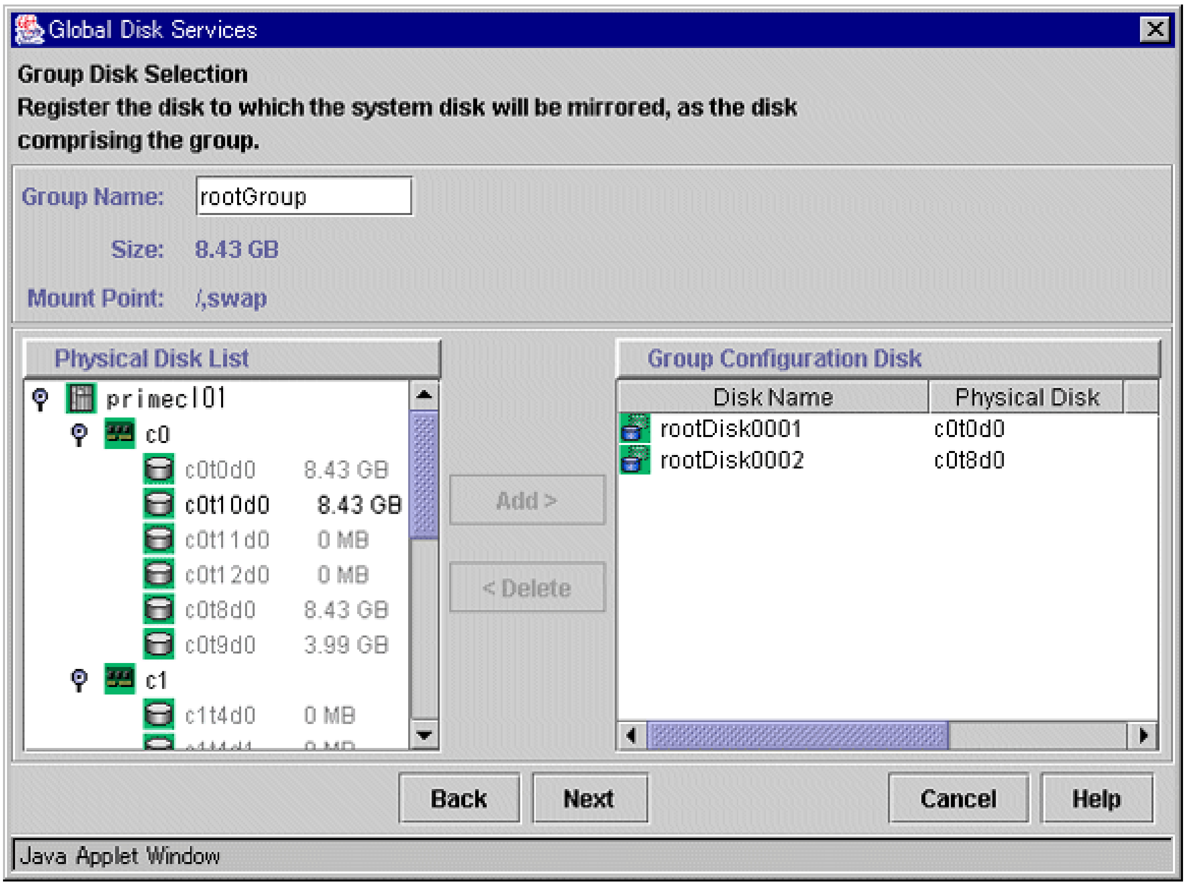


Note

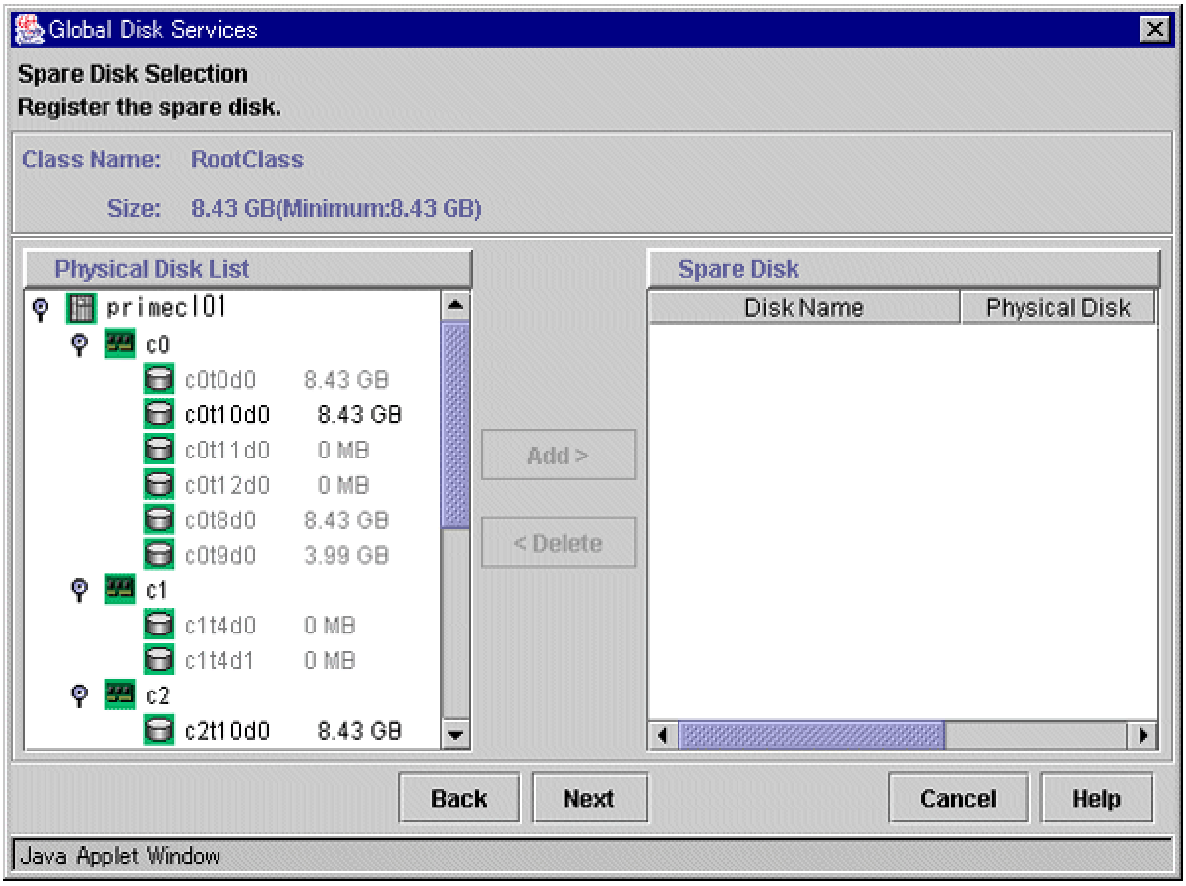
Specify the class name so that the class names of the root class are not duplicated among cluster nodes.

4. Add a mirror disk to the group.

Enter the "Group Name" ("Group name" on the worksheet), then from the "Physical Disk List," select the "mirror disk" ("Mirror disk name" on the worksheet) for the system disk, and click *Add*.
Check that the "mirror disk" that was selected is displayed in "Group Configuration Disk," and then click *Next*.

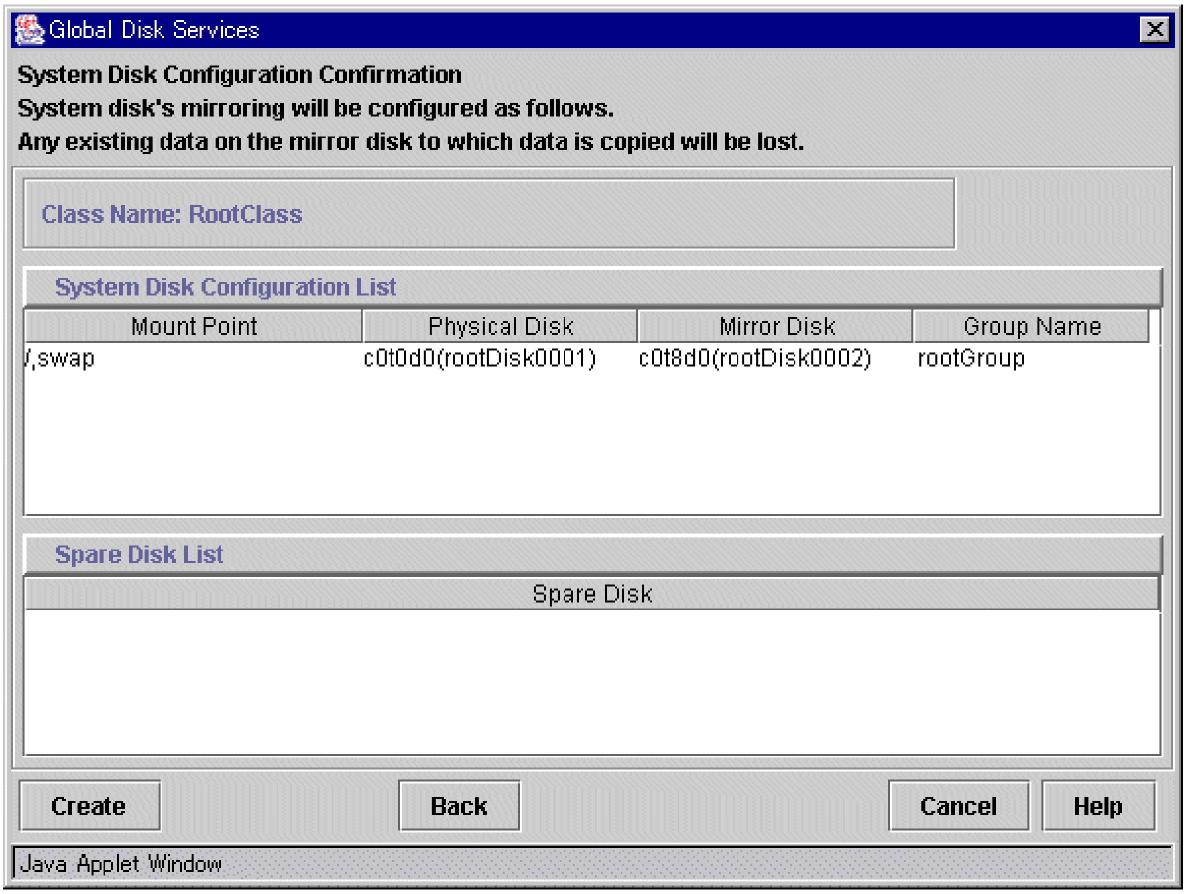


5. Select a spare disk ("Spare disk name" on the worksheet) from the "Physical Disk List," and click *Add*.
Check that the spare disk that was selected is registered to "Spare Disk," and then click *Next*.
If a spare disk is unnecessary, go to Step 6.



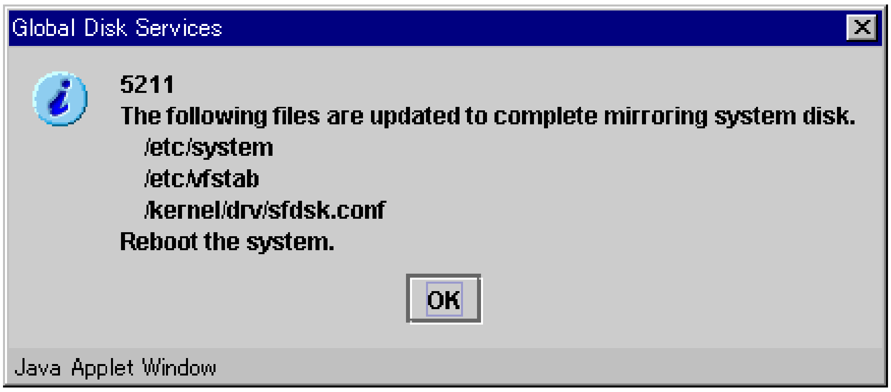
6. Check the system disk configuration.

Check the physical disk name and the mirror disk name, and then click *Create*.



After creation of the system disk is completed, the following screen is displayed.
Check the screen contents, and then click *OK*.

Set up mirroring for the system disk of primecl02 on each node, and then, restart all the nodes.



6.3.2.3 Setting Up Shared Disks

Set up the shared disks according to the following procedures:

- Volume setup
- File system setup

Volume setup

There are four types of volumes:

a. Single volume

b. Mirror volume

c. Striped volume

d. Volume created in a concatenation group

This section separately describes the volume setup procedures for a single volume (a) and for other volumes (b, c, d). For details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

The values to be set for the individual items correspond to the values in the "GDS Configuration Worksheet."

Note

If you plan to add, delete, or rename a disk class from the GDS Management screen (hereinafter main screen), close the Cluster Admin screen before starting the operation.

Single volume setup

If you are not using a single volume, this setup is unnecessary.

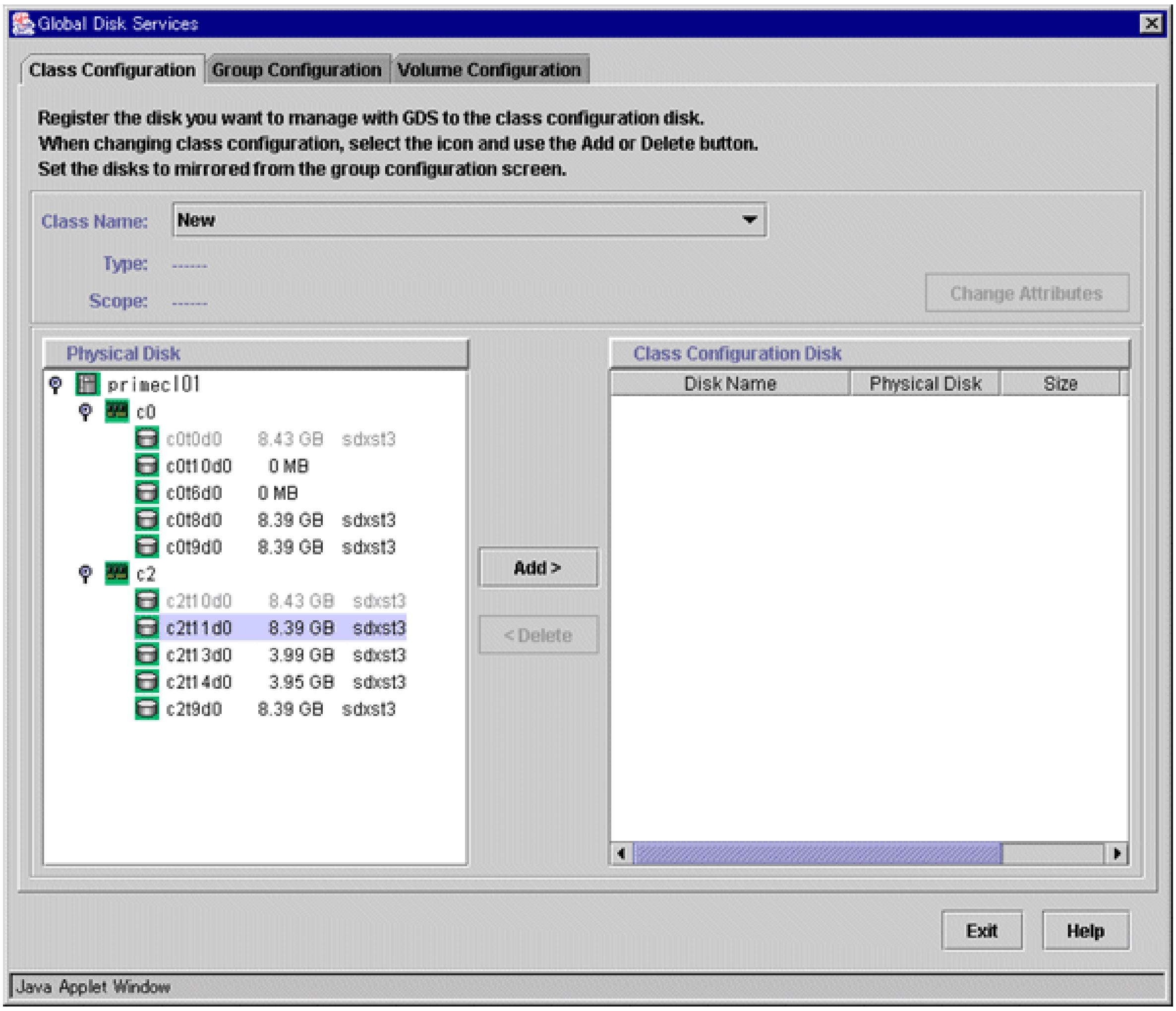
Operation Procedure:

1. Start the main screen

   Choose *Global Disk Services* on the Web-Based Admin screen.

2. Disk class creation and physical disk registration

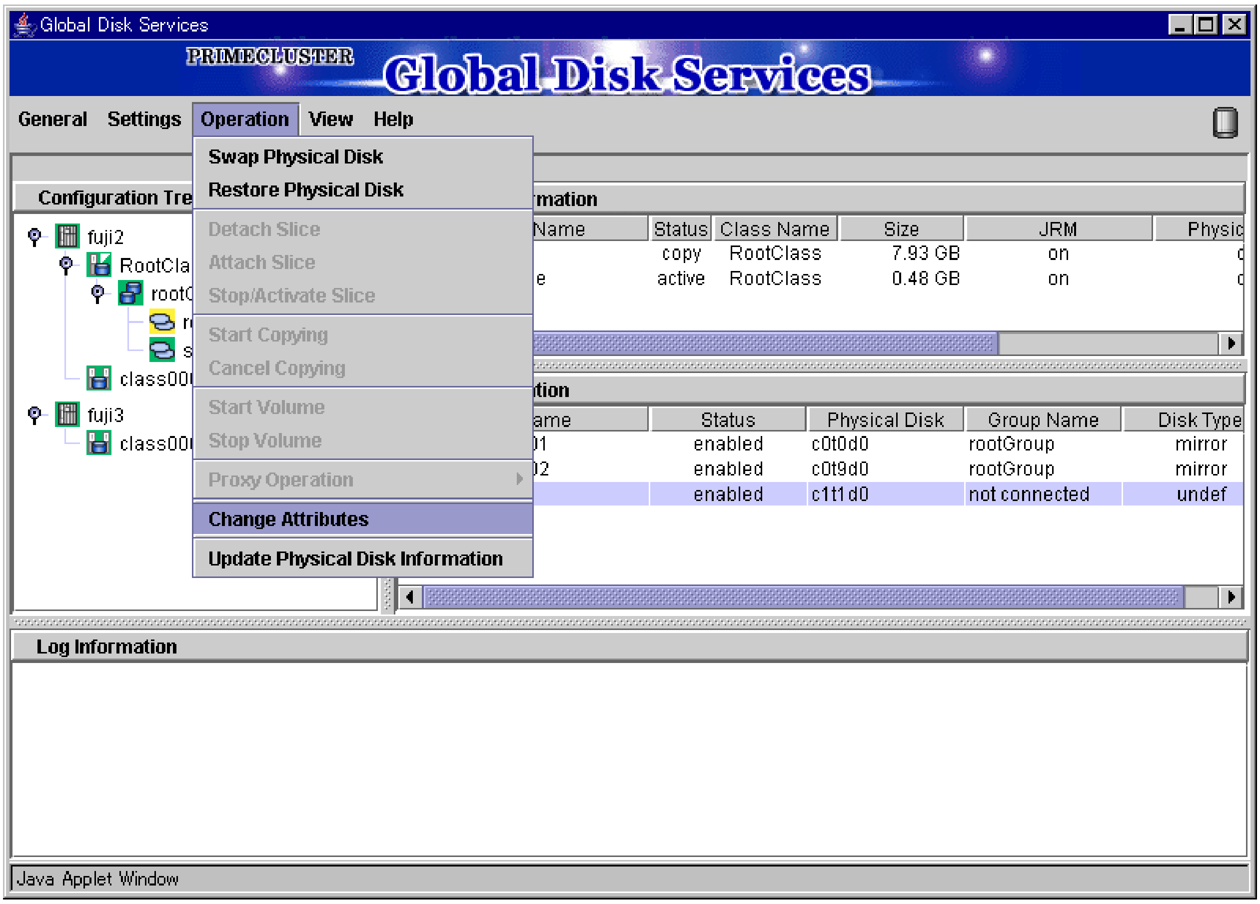
At the main screen, select *Class Configuration* from the *Settings* menu.



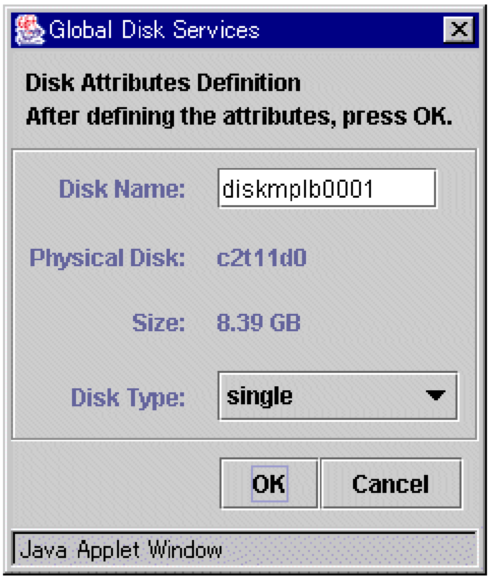
At the above screen, select the physical disk to be registered from the "*Physical Disk*" list, and then click *Add*. When *Add* is clicked, the class attribute definition screen opens. Enter the *Class name* but do not change the *Type* value (leave the value as "shared"). Then click *Exit*.

3. Disk type attribute setup

   At the main screen, select the disk that was registered in Step 1 from the disk information field, and select *Operation* -> *Change Attributes* from the menu bar.



   Set *Disk Type* to "single," and then click *OK*.

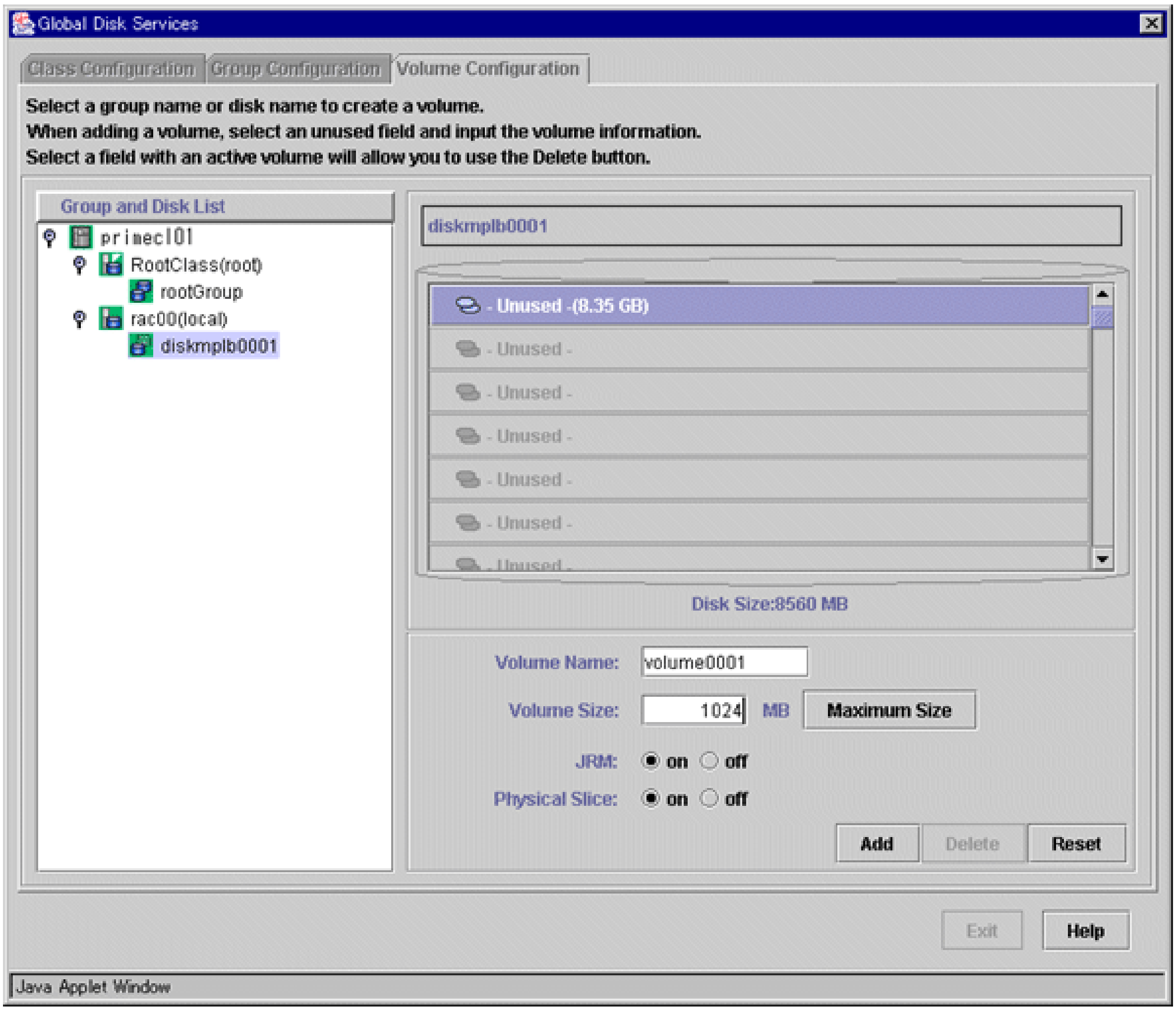


4. Volume creation

   Select *Settings* -> *Volume Configuration*, and then select the disk that was registered in Step 1 from the *Group and Disk List*. Select "*Unused*" in the volume diagram, and enter the "Volume Name," the "Volume Size," and the volume attributes.

Click *Add* to enable the settings.

Check the settings, and then click *Exit*.



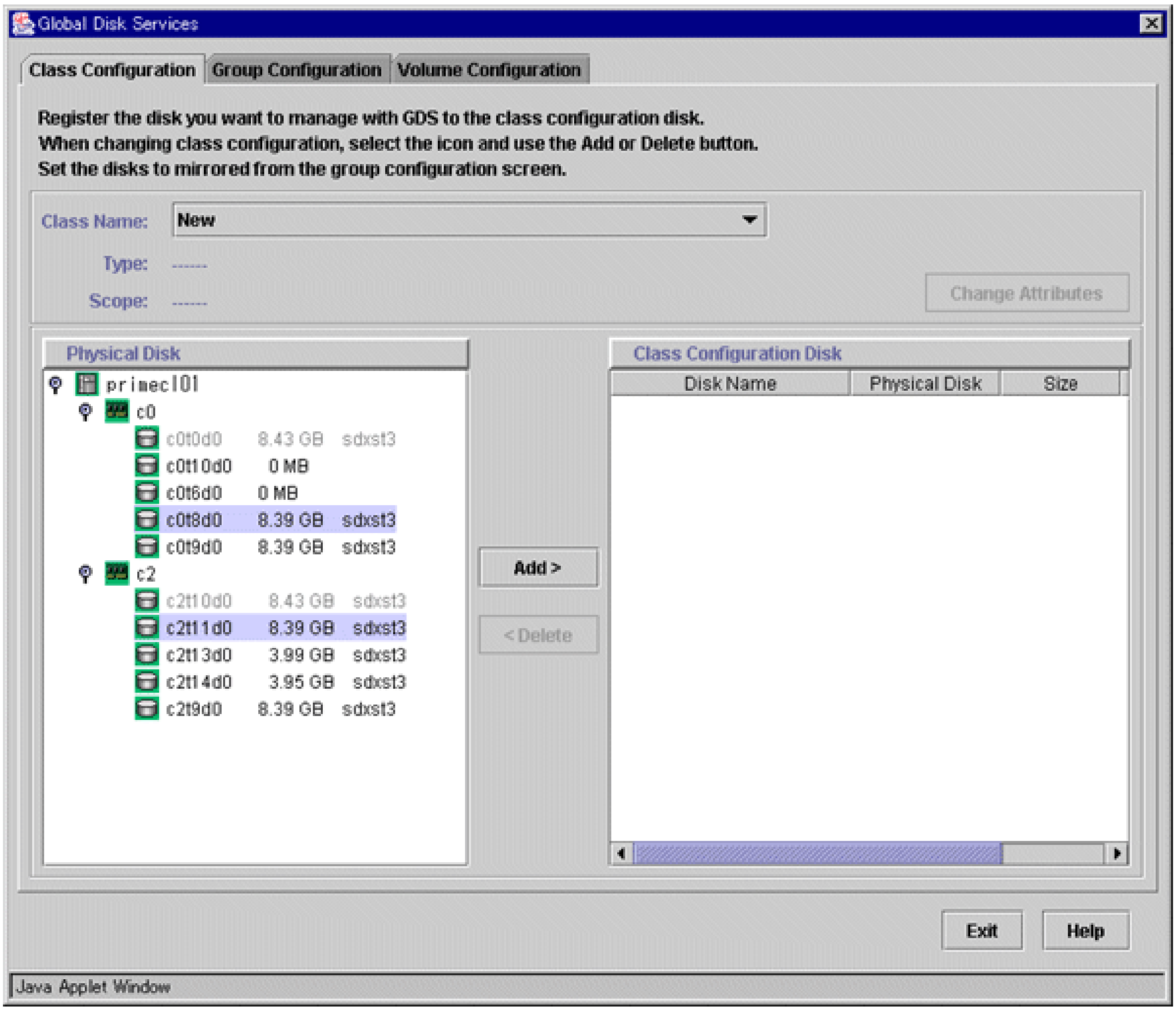
Setup for other volumes

If you are using only a single volume, this setup is unnecessary.

Operation Procedure:

1. Creating disk classes and registering physical disks

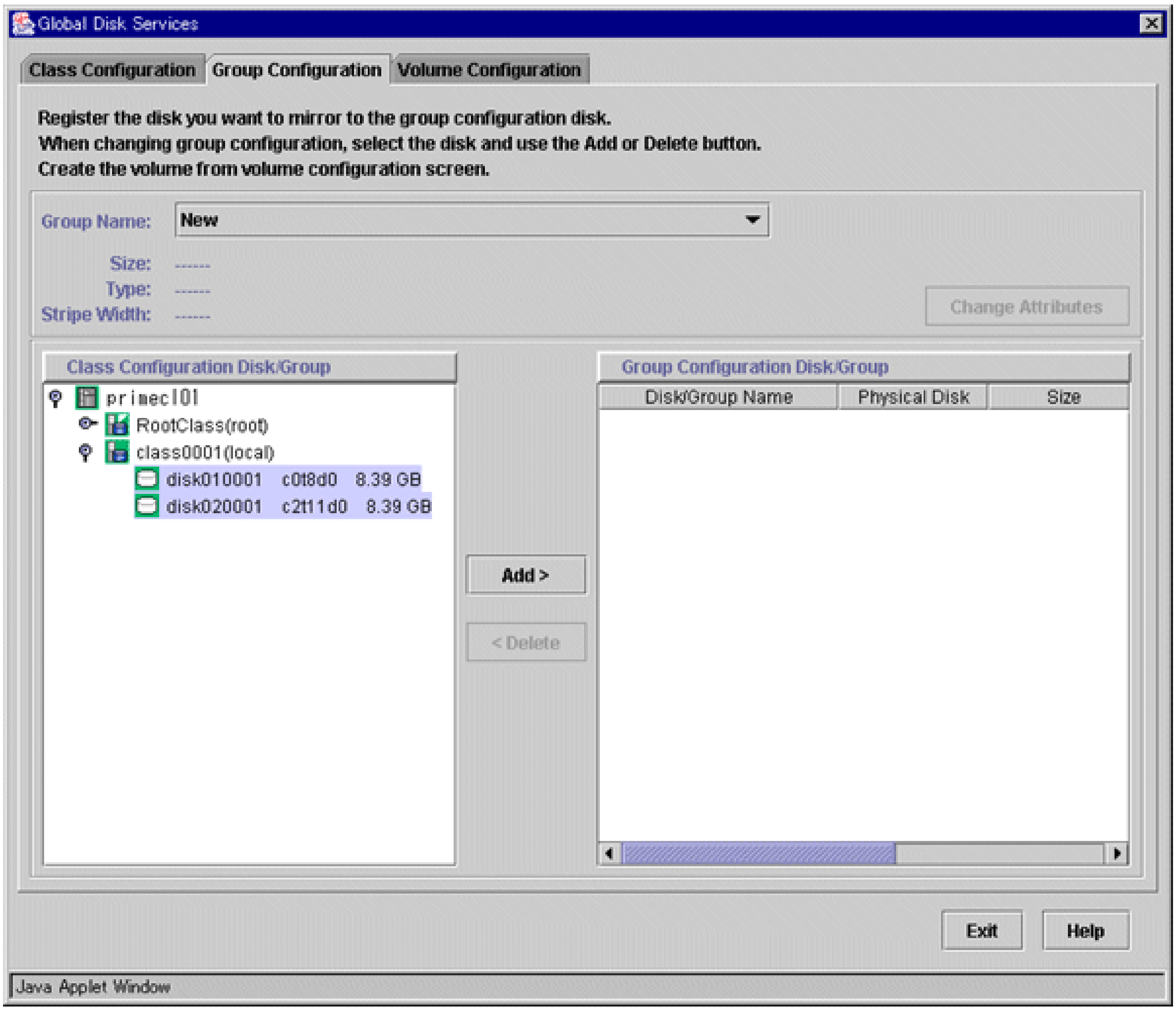
   At the main screen, select *Class Configuration* from the *Settings* menu.



   At the above screen, select the physical disk to be registered from the *Physical Disk* list, and then click *Add*. When *Add* is clicked, the class attribute definition screen opens. Enter the "Class name" but do not change the "Type" value (leave the value as "shared"). Then click *Exit*.

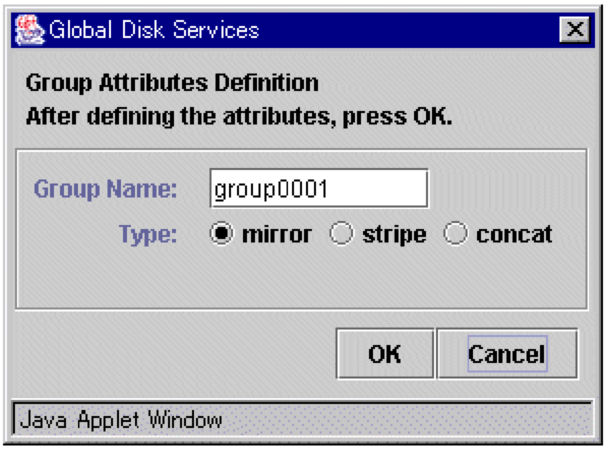
2. Setting up the disk group configuration

   Click the Group Configuration tab.



   At the above screen, select the disk to be added to the group from the *Class Configuration Disk/Group* list, and then click *Add*.

   Enter the "Group Name", "Type" and "Stripe Width" in the group attribute definition screen, and then click *OK*.

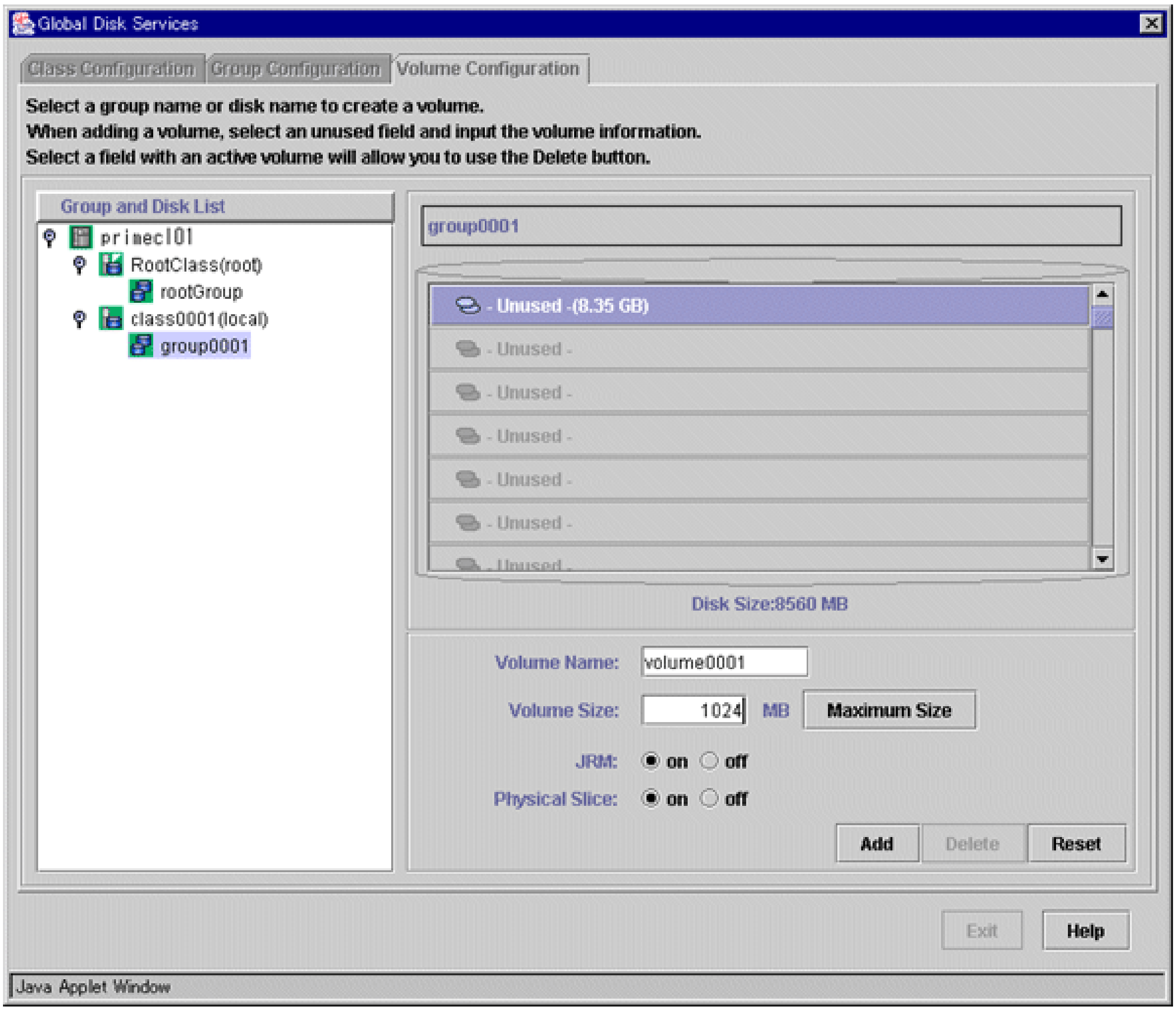


3. Creating a volume

   Click the *Volume Configuration* tab, and select the group that was created in Step 2 from the *Group and Disk List*. Select *Unused* in the volume diagram, and enter the "Volume Name," the "Volume Size," and the volume attributes.

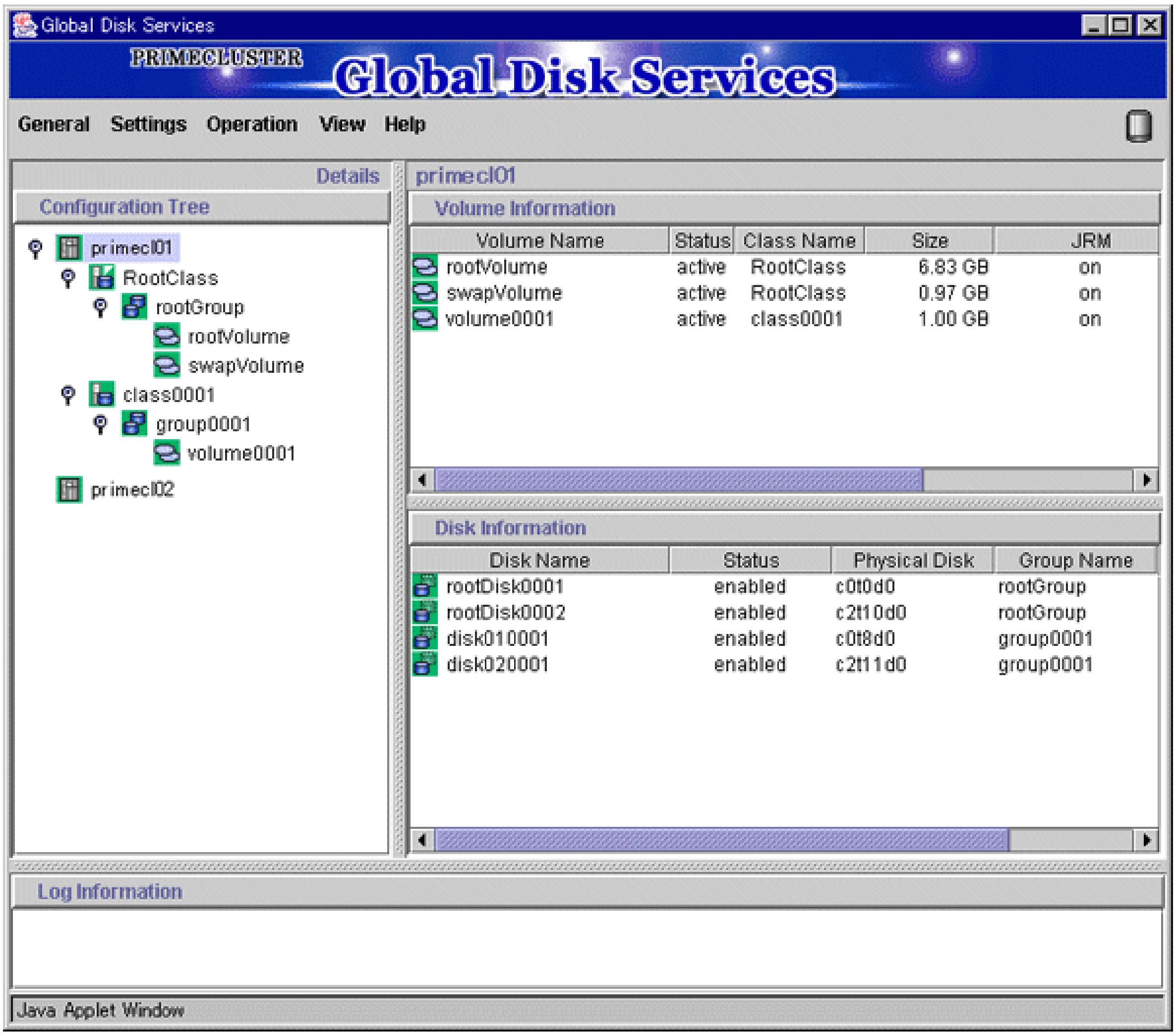
Click *Add* to enable the settings.

Check the setup information, and then click *Exit*.



4. Checking the configuration

The disk configuration is displayed as shown below.



File system setup

Create a file system for the created volume.

Example

When the class name is Class1, volume name is Volume1, and file system type is ufs.

```
# newfs /dev/sfdsk/Class1/dsk/Volume1
```

See

- For how to create a file system, see the manual of the file system.
- To use ZFS, follow the procedures described in "6.4 Initial File System Setup" to set up the file system.

6.4 Initial File System Setup

This section explains the Initial File System Setup to register Fsystem resources.

6.4.1 If using ZFS

This section explains monitoring targets, support configurations and required setup procedure before registering to Fsystem resources when using ZFS with PRIMECLUSTER.

6.4.1.1 Design for using ZFS with a Cluster System

Monitoring Facility

PRIMECLUSTER provides the following monitoring functions for ZFS file system which is configured on ZFS storage pool.

- Monitoring of the ZFS storage pool status (the status displayed by the zpool list command)
- Monitoring of the mount status of the ZFS file system created on the ZFS storage pool
- Monitoring of the NFS share status for the ZFS file system created on the ZFS storage pool

Supported Configurations

The ZFS configurations supported with PRIMECLUSTER are as follows.

- ZFS storage pool device

  GDS physical special files (example: /dev/sfdsk/class/dsk/volume0001) only

- ZFS file system type

  Both non-legacy file system (*1) and legacy file system (*2) are supported.

  Also, both file systems can exist together on one ZFS storage pool.

  *1: This is the default file system for ZFS. It is mounted or unmounted when the ZFS storage pool is imported or exported.

  *2: This is the file system which has the mountpoint property of "legacy." As with the UFS file system, it is managed using the mount/umount commands and /etc/vfstab.pcl.

- Highest level ZFS file system

  Make the highest level file system (the file system automatically created during the creation of the ZFS storage pool) a non-legacy file system.

- Dataset

  There are no restrictions on the type of dataset that can be created on the ZFS storage pool, but only file systems can be used for monitoring.

Notes on Using Legacy File System

Since there are the following disadvantages when using the legacy file system in PRIMECLUSTER, it is recommended that you use the non-legacy file system.

- Since mounting and unmounting are not performed as a part of the ZFS storage pool control, as with the UFS file systems, resources are created, also mounting and unmounting are performed for each file system. As such, the Online/Offline process takes more time for legacy file systems than for non-legacy file systems.

Note

- Create the ZFS storage pool and the legacy ZFS file systems created upon it as one Fsystem resource.
- The directory, which is the same name as the ZFS storage pool, cannot be used as a mountpoint of the legacy ZFS file system (For example, setting "legacy" to the mountpoint property of the ZFS storage pool "app1" and setting /etc/vfstab.pcl to mount it to "/app1").
- To hierarchically mount the legacy ZFS file systems on the ZFS storage pool, hierarchically mount them on the same ZFS storage pool.

Notes on Using Multiple File Systems in Combination

If you want to mount the legacy ZFS file system or UFS file system to the directory on the ZFS file system controlled by the Fsystem resource, use a legacy mount point.

Specifically, create a new dataset on the ZFS storage pool, and then make the dataset to be mounted as the legacy mount point. In addition, mount the legacy ZFS file system and UFS file system to the directory under the legacy mount point.

Example

The procedure for mounting the UFS file system on the ZFS storage pool "app1" is as follows:

1. Create the dataset "app1/zfsmnt" as the legacy ZFS file system, and then set the /etc/vfstab.pcl file to mount the dataset to /zfsmnt.
2. Set the /etc/vfstab.pcl file to mount the UFS file system to /zfsmnt/ufsmnt.

Fsystem resource cannot be created with the following configurations (combination):

- Configuration to mount the dataset on the storage pool to the non-legacy ZFS file system on another storage pool.
- Configuration to mount the UFS file system to non-legacy ZFS file system.

Example

The UFS file system cannot be mounted to the /app/data directory under /app where the ZFS storage pool "app1" was imported.

- Configuration to import the ZFS storage pool to the mountpoint of the legacy ZFS file system which controlled as the Fsystem resource.
- Configuration to import the ZFS storage pool to the mountpoint other than ZFS file system which controlled as the Fsystem resource.

Example

The mountpoint /mnt/data of the ZFS storage pool cannot be created on the UFS mountpoint /mnt controlled as Fsystem resource.

6.4.1.2 Setup Procedure

1. GDS Configuration Setup

   See "6.3.2 GDS Configuration Setup" and "PRIMECLUSTER Global Disk Services Configuration and Administration Guide", create a shared disk.

   The GDS physical special file which is the target on the node where the following operation is performed needed to be accessed.

2. Creating the ZFS storage pool

   Perform the procedure up to "5. export of the ZFS storage pool" by one of the nodes which configure a cluster.

   Create the ZFS storage pool by the zpool create command. The following example shows that the storage pool name is the app and the GDS physical special file (/dev/sfdsk/class/dsk/volume0001) is used.

   ```
   # zpool create app /dev/sfdsk/class/dsk/volume0001
   # zfs list -r app
   NAME USED AVAIL REFER MOUNTPOINT
   app 178K 129G 28.5K /app
   ```

   For details on the command to use, see the Solaris ZFS management guide.

   The highest level of the ZFS file system is automatically created if the ZFS storage pool is created as above.

3. Creating the ZFS file system

   Create the ZFS file system by the zfs command. Three non-legacy files of app/home, app/config and app/data are created in the example below.

```
# zfs create app/home
# zfs create app/config
# zfs create app/data
# zfs list -r app
NAME USED AVAIL REFER MOUNTPOINT
```

When creating a non-legacy file system, set the mountpoint property to legacy. The following is the example of setting the file system app/data to legacy.

```
# zfs set mountpoint=legacy app/data
```

Information

It is not a problem even if the "-o mountpoint=legacy" is specified and the mountpoint property is set when creating the file system by the zfs create.

4. Prerequisites for Fsystem resources

   See "Note" in "6.7.1.2 Creating Fsystem Resources" and "6.7.1.2.1 Prerequisites" to perform the prerequisites for registering the Fsystem resources.

   For settings to share in NFS, see the procedure in "6.7.1.2.1 Prerequisites." Also, when sharing the non-legacy file systems in NFS, the sharenfs property of ZFS must be setup.

   For information on how to set up the sharenfs property, see the manual ZFS (1M) of ZFS. The following is an example how to set the sarenfs (specifying on) to the file system app/home.

```
# zfs set sharenfs=on app/home
```

5. export of the ZFS storage pool

   Export the ZFS storage pool created above by the zpool export command.

```
# zpool export app
```

6.4.1.3 Notes on the Operation

If using ZFS in PRIMECLUSTER, there are the following notes on the operation.

- Do not allocate a file to the import destination and the mount destination or mount the other file system. For the Online process of the ZFS resource, the ZFS storage pool is imported and the ZFS file system is mounted. Therefore, if a file is allocated to the directory of the import destination or the mountpoint of the file system and the other file system is mounted, there may be a case where the startup of the userApplication and Failover are failed.
- When starting the userApplication, the ZFS storage pool need to be exported. After creating the ZFS storage pool and complete the settings, follow the step 5) "6.4.1.2 Setup Procedure" and export the ZFS storage pool. Moreover, when importing it automatically with the purpose of a backup, export the userApplication before startup. If it is imported, the startup of the userApplication is failed.
- After creating the ZFS storage pool, do not access via the physical special file of GDS (/dev/sfdsk/class/dsk/volume and /dev/sfdsk/class/rdsk/volum)
- In the non-legacy zfs file system, do not execute unshare command for the mountpoint of the dataset where "on" was set to the sharenfs property or the share.nfs property.

  In addition, do not delete the ZFS sharing with zfs set -c. If those commands are executed, a resource error is detected and userApplication cannot be started. This is because the mountpoint is not published as NFS when the pool and the dataset start up.
- If unshare command is executed mistakenly, stop RMS on all nodes, import the non-legacy ZFS file system manually, and then execute share command to the directory where unshared command was executed.

  After that, delete the ZFS sharing, and then export the non-legacy ZFS file system.

- If the ZFS sharing was deleted with zfs set -c mistakenly, stop RMS on all nodes, import the non-legacy ZFS file system manually, and then set "off" to the sharenfs property or the share.nfs property for the dataset where the ZFS sharing is to be deleted. After that, export the non-legacy ZFS file system.

6.5 Setting Up the Application Environment

Configure an environment for the applications to be used in the PRIMECLUSTER system.

The environment configuration for the individual applications may involve registering resources to the PRIMECLUSTER system.

There are also products that require you to set up an environment that uses the shared disk units and takeover networks that were set in this chapter.

See

See the manuals for the individual applications.

6.6 Setting Up Online/Offline Scripts

Create Online and Offline scripts to start and stop ISV applications and user applications in line with the userApplication state transition.

Set the created scripts as Cmdline resources and set those resources in userApplication.

For details, see "6.7.1.1 Creating Cmdline Resources."

- An Online script is started when userApplication is switched to Online.
- An Offline script is started when userApplication is switched to Offline.
- A Check script is used to monitor the state of the resource that is started or stopped with an Online or Offline script.

This section presents script examples and describes notes on script creation.

Note

Environment variables set in each server (such as "/etc/profile") are not guaranteed to be inherited by Online, Offline, and Check scripts. Therefore, make sure to define the environment variables used with these scripts in each script.

Sample scripts

This section shows samples of the Online and Offline scripts, which are set as Cmdline resources.

Figure 6.1 Start script/Stop script

```
#!/bin/ksh
#
# Script.sample
#    Sample of Online/Offline Script
#
# Copyright(c) 2002 FUJITSU LIMITED.
# All rights reserved.
#
# $1 -c : OnlineScript
#    -u : OfflineScript

if [[ $1 = "-c" ]]; then
      # Start your application
elif [[ $1 = "-u" ]]; then
      # Stop your application
else
      # Default operation
      exit 1  # Error
fi
exit 0
```

The above script sample is both for the Start script and Stop script.

An example of Check script is shown below:

Figure 6.2 Check script

```
#!/bin/sh
#
# Script.sample.check
#    Sample of Check script
#
# Copyright(c) 2003 FUJITSU LIMITED.
# All rights reserved.
#
# Check the current state of target resource.

# If status is Online:
      exit 0
# If status is not Online:
      exit 1
```

Set up the above scripts in the Cmdline resource as shown below:

- Start script $FULL_PATH/Script.sample -c
- Stop script $FULL_PATH/Script.sample -u

- Check script $FULL_PATH/Script.sample.check

For information on how to set up these scripts, see "6.7.1.1 Creating Cmdline Resources."

Notes on script creation

Hot-standby operation

To enable hot-standby operation of the Cmdline resources, the following must be prepared:

- Online/Offline/Check scripts that support hot-standby operation.
- The setting of attributes for the Cmdline resources

1. Create the Online, Offline, and Check scripts to support hot-standby operation. The sample scripts are shown below.

Figure 6.3 Start script/Stop script (hot-standby operation)

```
#!/bin/sh
#
# Script.sample
#    Sample of Online/Offline Script
#
# Copyright(c) 2003 FUJITSU LIMITED.
# All rights reserved.
#
# $1 -c : OnlineScript
#    -u : OfflineScript

if [[ $1 = "-c" ]]; then
      if [ ${HV_LAST_DET_REPORT} = "Offline" ]; then
            if [ ${HV_INTENDED_STATE} = "Standby" ]; then
                    # commands for Offline -> Standby
            else
                    # commands for Offline -> Online
            fi
      else
            # commands for Standby -> Online
      fi
elif [[ $1 = "-u" ]]; then
      if [ ${HV_LAST_DET_REPORT} = "Standby" ]; then
            # commands for Standby -> Offline
      else
            # commands for Online -> Offline
      fi
else
      # Default operation
      exit 1 # Error
fi
exit 0
```

The following example shows Check script that supports hot-standby operation.

Figure 6.4 Check script (hot-standby operation)

```
#!/bin/sh
#
# Script.sample.check
#    Sample of Check script
#
# Copyright(c) 2003 FUJITSU LIMITED.
# All rights reserved.
#

# Check the current state of target resource.
# If status is Online:
   exit 0
# If status is Standby:
   exit 4
# If status is Faulted:
  exit 2
# If status is Offline:
   exit 1
```

2. Setting attributes for the Cmdline resources

   Enable the STANDBYCAPABLE and the ALLEXITCODES attributes.

   For details, see "6.7.1.1 Creating Cmdline Resources."

Online/Offline script exit code

The state transition process of userApplication changes according to the exit code of the Online/Offline script:

0: Normal exit

The system assumes that the state transition of the Cmdline resource was processed normally, and state transition processing of the userApplication continues. If all the resources of the userApplication are processed normally, the state transition of the userApplication is also processed normally.

Other than 0: Abnormal exit

The system assumes that an error occurred during the state transition of the Cmdline resources and interrupts state transition processing of the userApplication.

Check script exit code

The state of the Cmdline resource is determined by the exit code of Check script. The exit code and the Cmdline resource are associated each other as follows:

0: Indicates the Online state.

Other than 0: Indicates the Offline state.

When ALLEXITCODES variables of the Cmdline resources are enabled, Check script will provide more detailed state of the resource. The exit code and Cmdline resource are associated each other as follows:

0: Indicates the Online state.

1: Indicates the Offline state.

2: Indicates the Faulted state.

3: Indicates the Unknown state.

4: Indicates the Standby state.

5: Indicates the Onlinewarning state.

6: Indicates the Offlinefaulted state.

For details, see "HELP."

Timeout

If script processing is not completed within the specified time, a timeout occurs, script processing is interrupted by the SIGTERM signal, and state transition ends with an error.

Default: 300 seconds

The timeout value can be specified with the TIMEOUT flag value of the Cmdline resources.

When creating the Cmdline resource, you need to set up a timeout value in "Setting up Cmdline flags" of "6.7.1.1 Creating Cmdline Resources." If a timeout occurs when a Cmdline resource is used, change the timeout value to an appropriate value according to the instructions in "10.6.2 Changing the Attributes Used by a Resource or a Resource Interface."

Environment variables

When the script is executed, the environment variables shown in the table below are set.

Environment variable	Outline
HV_APPLICATION	This variable sets the userApplication name that the resource belongs to. Example) app1
HV_AUTORECOVER	The value of this variable indicates whether the script is triggered by AutoRecover or not (1 or 0). For details on AutoRecover, see "12 Appendix - Attributes" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*." 0: Not triggered by AutoRecover 1: Triggered by AutoRecover
HV_FORCED_REQUEST	This variable sets a value that indicates whether or not forced failover was requested by operator intervention. 0: Forced failover was not requested. 1: Forced failover was requested.
HV_NODENAME	This variable contains the resource name. Example) ManageProgram000_Cmd_APP1, RunScriptsAlways000_Cmd_APP1
HV_OFFLINE_REASON	This variable sets the trigger for bringing the resource Offline. SWITCH: The resource was set to Offline because of a userApplication switchover request (hvswitch). STOP: The resource was set to Offline because of a userApplication stop request (hvutil -f, hvutil -c) FAULT: The resource was set to Offline because of a resource fault. DEACT: The resource was set to Offline because of a userApplication deactivate request (hvutil -d) SHUT: The resource was set to Offline because of an RMS stop request (hvshut)
HV_SCRIPT_TYPE	This variable sets the type of script that was executed. Online: Online script Offline: Offline script
HV_LAST_DET_REPORT	This variable sets the state of the current resources. Online: Online state Offline: Offline state

Environment variable	Outline
	Standby: Standby state Faulted: Faulted state Warning: Warning state
HV_INTENDED_STATE	This variable sets the resource state that is expected after state transition is completed. Online: Online state Offline: Offline state Standby: Standby state Faulted: Faulted state Warning: Warning state
NODE_SCRIPTS_TIME_OUT	This variable sets the timeout duration (seconds) of the script. Example) 300

When executing the Check script, the following environment variables in the above table are set:

- HV_APPLICATION
- HV_NODENAME

See

- For details on hvenv.local, see "2.9 Environment Variables" in the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."
- For details on the RMS environment variables, see "13 Appendix - Environment Variables" in the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

6.7 Setting Up Cluster Applications

This section explains how to set up a cluster application by using the cluster application GUI.

The terms "userApplication" and "Resource" that appear on the setup screen refer to "cluster application" and "resource" respectively.

Information

- After you finish setting up the cluster application, start the cluster applications. For instructions on starting the application, see "7.2.2.1 Starting a Cluster Application."
- For instructions on changing a cluster application, see "10.3 Changing a Cluster Application." For instructions on deleting a cluster application, see "10.2 Deleting a Cluster Application."
- For the setting contents and notes on a cluster application that depends on the operation, see "6.11 Setting Contents and Notes on Cluster Application."

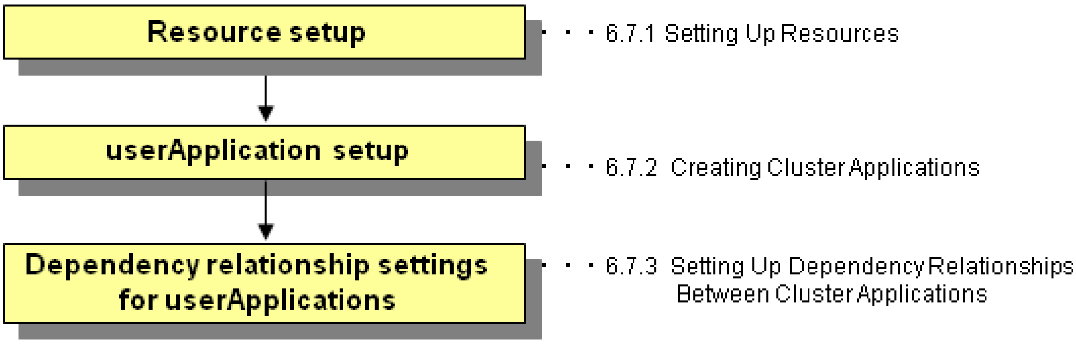
Note

- Set up the cluster application and resources based on the cluster application and resource information in the "A.9 Cluster Application Worksheets," which was created in the planning stage. If you need to change the cluster application after it is created, the worksheet is helpful.
- You cannot share one resource with multiple userApplication.

- If there are files with the same name as shown below on setup of a cluster application, you cannot set up a cluster application properly. If files with the same name exist, delete them before setting up a cluster application.
    - /tmp/hostcheck
    - /tmp/sdxinfo
    - /tmp/work
    - /tmp/workc
    - /tmp/workn
    - /tmp/macaddr
    - /tmp/wgcnf000.tmp
    - /tmp/wgcnf001.tmp
    - /tmp/wgcnf002.tmp
    - /tmp/wgcnf003.tmp
- If you use userApplication Configuration Wizard, the CF remote services (remote file copy and remove command execution) are enabled automatically. For instructions on disabling these definitions, see Notes in "5.1.1 Setting Up CF and CIP."
- Do not make any settings that cause multiple cluster system nodes to be monitored with a single primary management server of Web-Based Admin View.
  Use a single primary management server of Web-Based Admin View to monitor a single cluster system.
- If you click <Registration> on the confirm registration screen for userApplication or Resource, the dialog boxes will appear (0805 followed by 0813) to show the status of the configuration process. While these dialog boxes are displayed, do not operate the userApplication Configuration Wizard screen, which is on the background.
- The RMS configuration created in the userApplication Configuration Wizard cannot be used in the RMS Wizard. Moreover, the RMS configuration created in the RMS Wizard cannot be used in the userApplication Configuration Wizard.

Setting up cluster applications by using GUIs

Operation Procedure:



1. Resource setup

   Create all resources that make up the userApplication.

2. userApplication setup

   Use all the resources and create the userApplication.

   Note

   - RMS will not start if there is any remaining resource that is not part of userApplication. Delete unnecessary resources.

- To create multiple userApplication, repeat steps 1) and 2).

3. Dependency relationship settings for userApplication

   Dependency relationship settings are enabled when two or more userApplication are created.

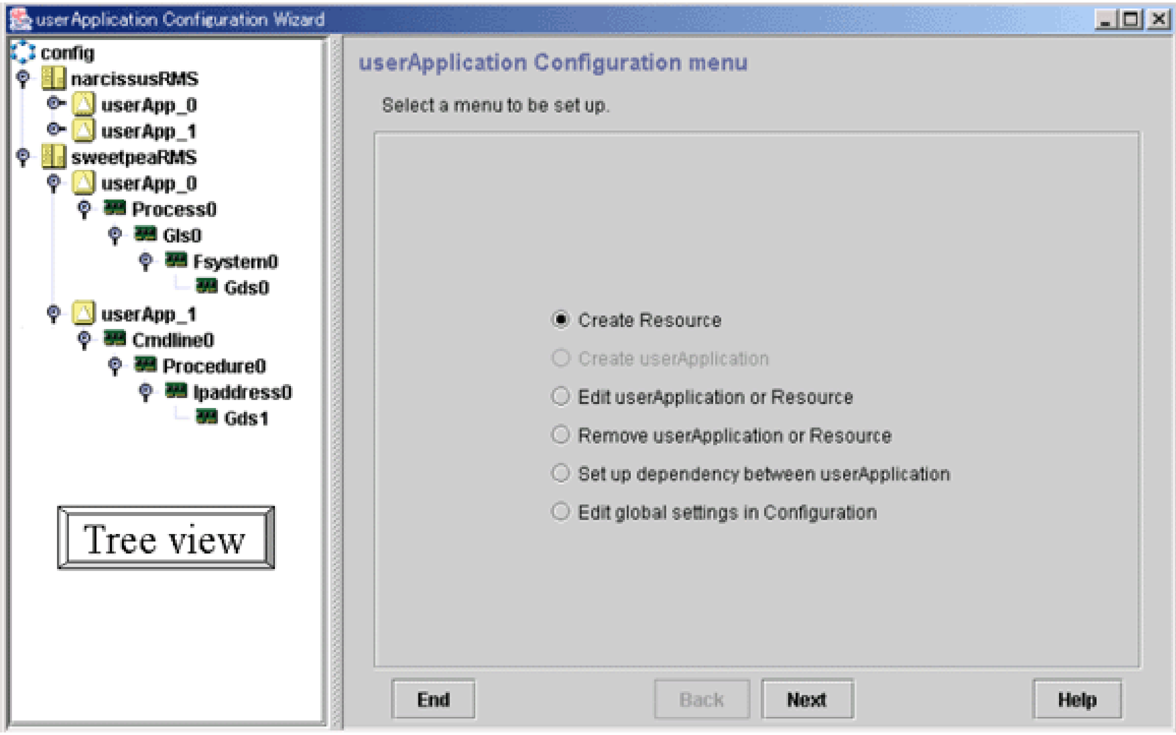
   These settings define how each userApplication is controlled on the same cluster node.

Starting the userApplication Configuration Wizard

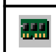
1. At the Web-Based Admin View top screen, select *Global Cluster Services*.
2. At the Global Cluster Services screen, select *userApplication Configuration Wizard*.

Explanation of GUI screen and icons

The userApplication Configuration Wizard screen has the configuration shown below.



The following table shows the icons that are displayed in the tree view of the userApplication Configuration Wizard.

Icon	Description
	Indicates the configuration.
	Indicates that a discrepancy was found in the configuration information. A resource that does not belong to any userApplication or userApplication with no resources was found.
	Indicates SysNode.
	Indicates userApplication.
	Indicates userApplication that was created with the TURNKEY Wizard. The userApplication Configuration Wizard does not allow change and deletion operations for this userApplication and its Resources.
	Indicates a controller.
	Indicates a resource.

6.7.1 Setting Up Resources

This section explains how to register resources to RMS using the "userApplication Configuration Wizard" screen.

The resources include the following:

Resource type	Outline
Cmdline resources	Use this type to set a user-created script file or command as a resource. The resource is required to start or stop the ISV application or application in line with the userApplication state transition, and in the same way, to cause the state transition with the application stop.
Fsystem resources	Mounts a file system at userApplication startup.
Gds resources	Sets up a disk class in userApplication. The disk class is defined in Global Disk Services. In the case of the single-node cluster operation, create GDS volumes in a local class and do not create Gds resources.
Gls resources	Sets up a takeover IP address in userApplication. The Ipaddress is defined in Global Link Services (network redundancy).
Takeover network resources	Sets up a takeover network in userApplication. The two types of takeover addresses are takeover IP address and takeover node name.
Procedure resources	Sets up a state transition procedure in userApplication.
Process monitoring resources	Sets up a user-created program to be monitored by the process monitoring function in userApplication.
Line switching resources	Sets up a line switching unit in userApplication. (used simultaneously with Netcompo WAN control) Line switching resources are not available in Solaris 11 environments.
ISV resources	Sets up an ISV (Independent Software Vendor) application in userApplication.

Note

CUI (hvw) cannot display or operate any configuration that was created by using resources other than those listed above.

Note

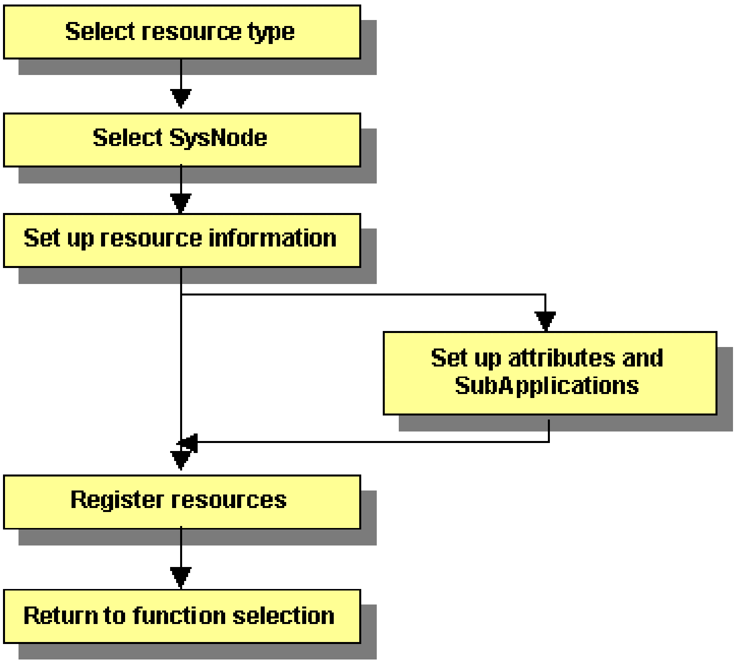
Resource name

When you create a resource, first select a resource type and then enter a resource name. The resource name must be a string of **18 or fewer characters** consisting of alphanumeric characters and "_" (underscore), **starting with an uppercase alphabetic character**. For this name, do not use the same name ("config") as the Configuration name or the same name ("Cmdline" and "Fsystem" etc.) as resource type name.

Example: Cmdline0

The flow of resource creation is shown below.

Figure 6.5 Resource creation flow



Note

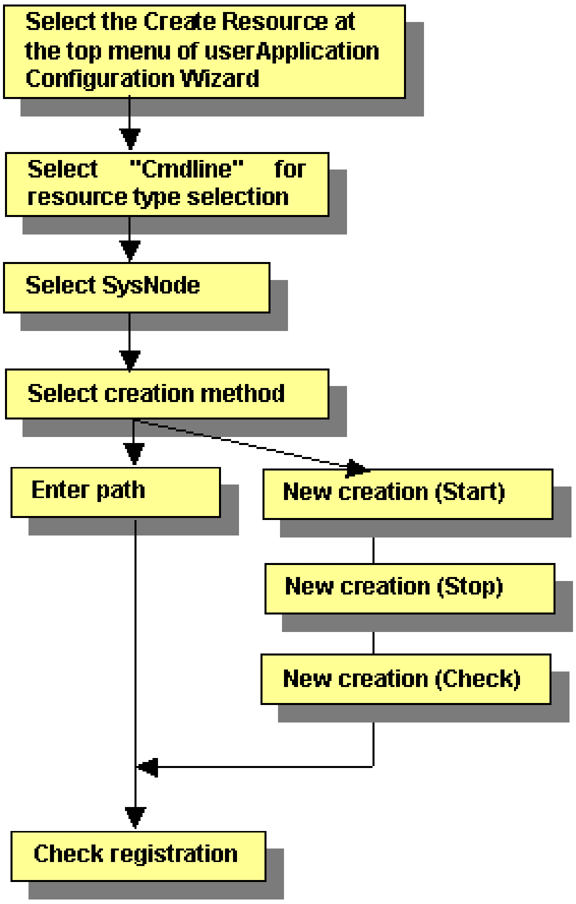
When specifying an association (priority setting) between resources of the same type, create the resource starting from the one with the higher priority.

6.7.1.1 Creating Cmdline Resources

This section describes the procedure for creating Cmdline resources.

For the notes when setting Cmdline resources, see "6.12 Notes When Setting Cmdline Resources."

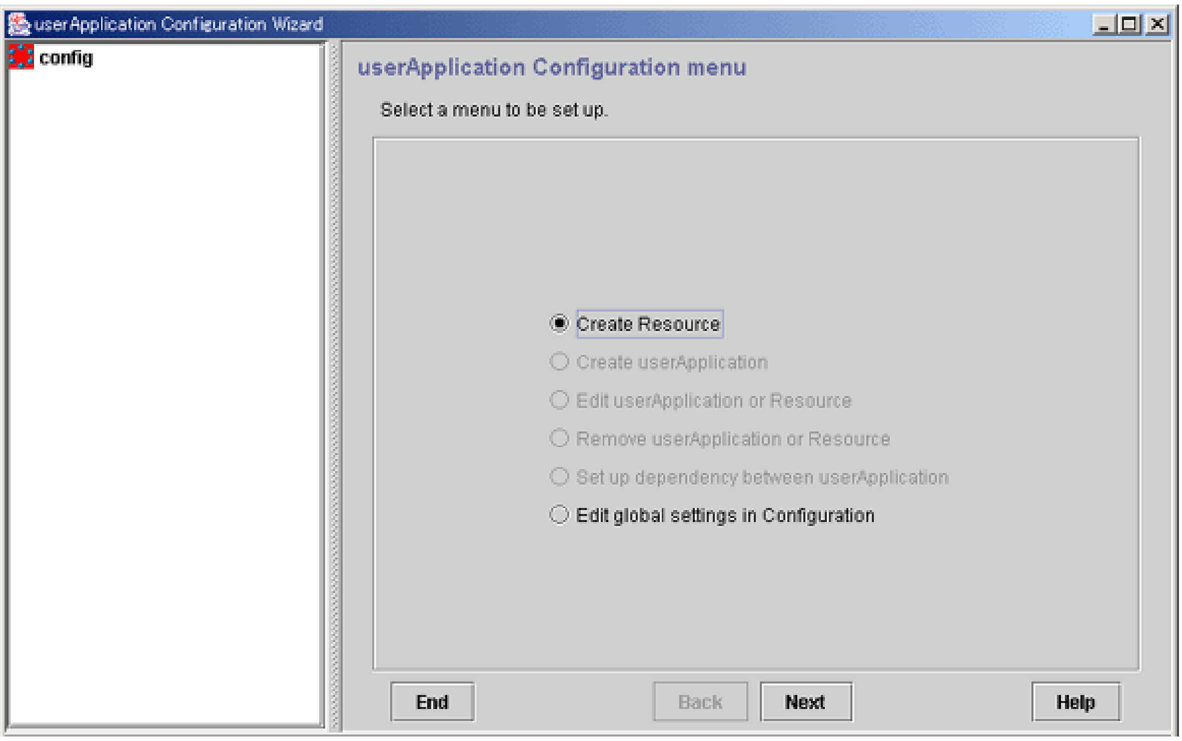
Figure 6.6 Flow of Cmdline resource creation



Creating resources

At the top menu of the "userApplication Configuration Wizard," select *Create Resource*.

Figure 6.7 Top menu of the userApplication Configuration Wizard

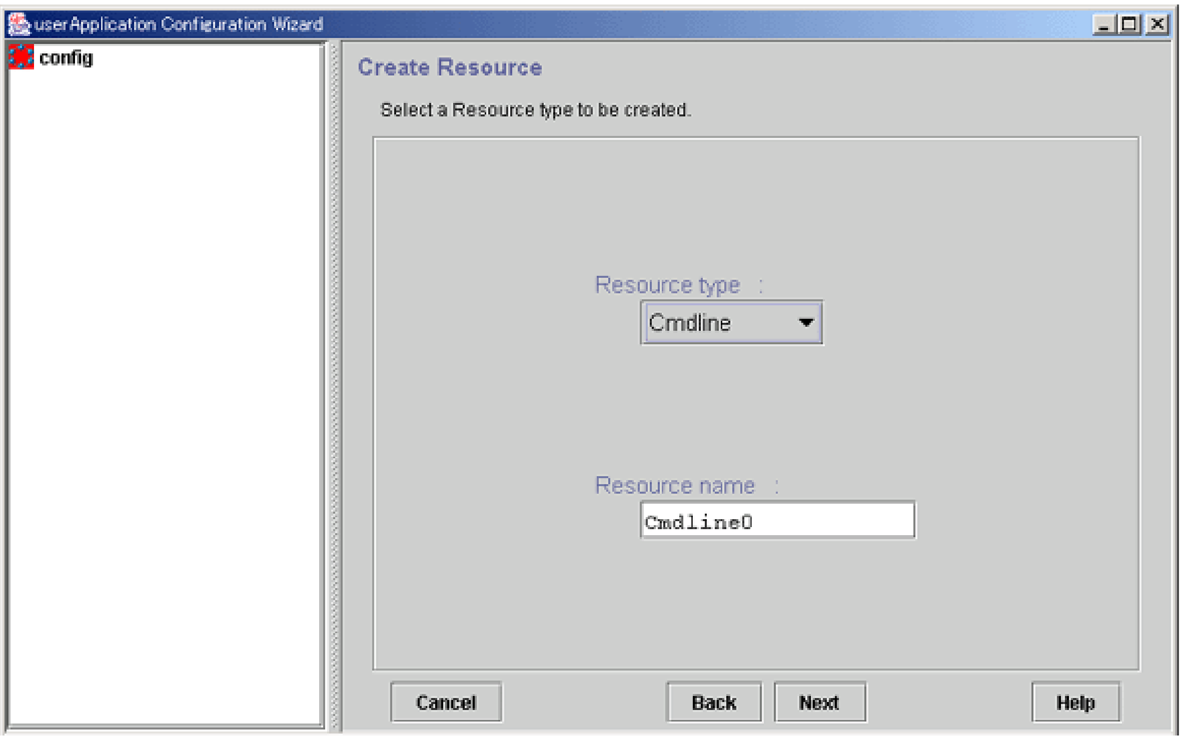


Click *Next* to go to the "Create Resource" screen.

Selecting the resource type

Select the type of resource that you want to create.

Figure 6.8 Create Resource



Resource Type

Select the type of resource to be created from the types described in the table below.

- **Cmdline** (Cmdline resources)
- **Fsystem** (Fsystem resources)
- **Gds** (Gds resources)
- **Gls** (Gls resources)
- **Ipaddress** (Takeover network resources)
- **Procedure** (Procedure resources)
- **Process** (Process monitoring resources)
- **SH_SWLine** (Line switching resources)
- **Oracle, Nwcl, Nwsv, Nwst** (ISV resources)

Resource Name

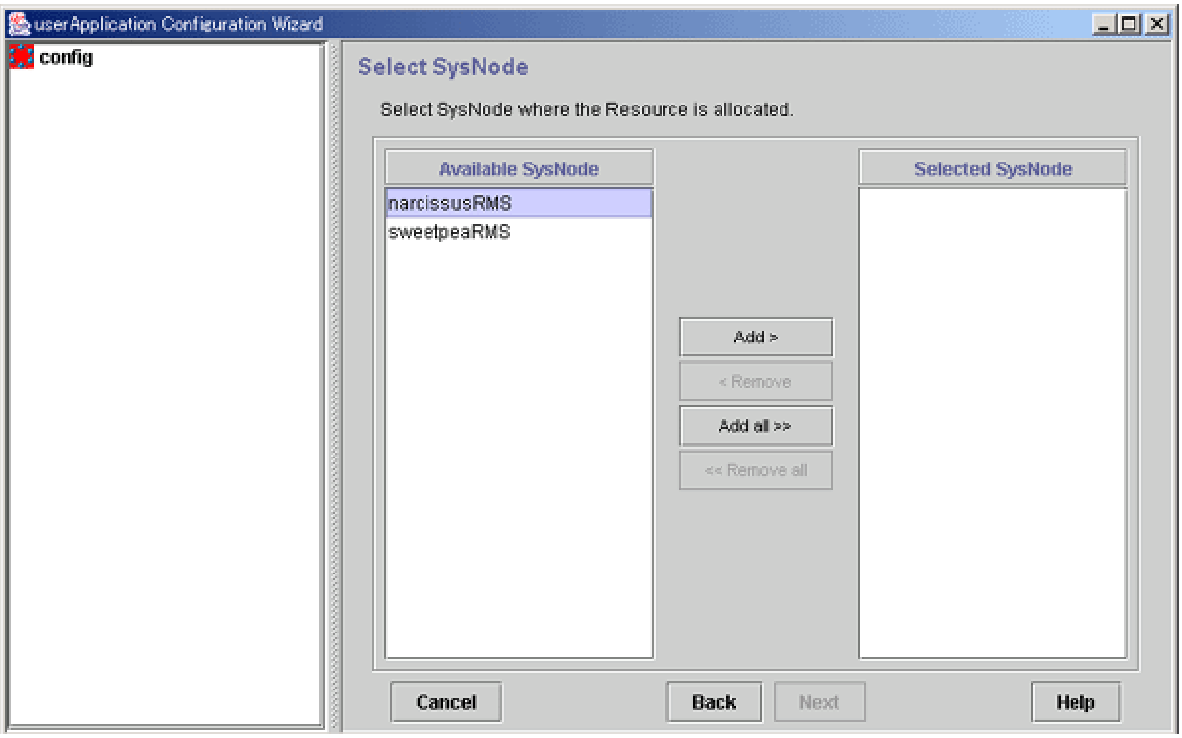
When you create a resource, first select a resource type and then enter a resource name. The resource name must be a string of **18 or fewer characters** consisting of alphanumeric characters and "_" (underscore), **starting with an uppercase alphabetic character**. For this name, do not use the same name ("config") as the Configuration name.

Click *Next* to go to the "Select SysNode" screen.

Selecting the SysNode

Select the SysNode in which the resource is to exist. SysNode refers to the CIP node name used by RMS.

Figure 6.9 Select SysNode



Available SysNode

Of the nodes in a cluster system, only the names accessible through Web-Based Admin View are displayed.

Selected SysNode

The SysNode name in which the resource is to exist is displayed. This information is used as filtering information for the interfaces used by each resource.

From the *Available SysNode*, select the SysNode, and then click *Add*. To add all the listed SysNodes, click *Add all*.

To delete a SysNode, select the SysNode to be deleted from *Selected SysNodes*, and then click *Remove*. To delete all listed SysNodes, click *Remove all*.

Click *Next* to go to the "Set up each resource" page.

Selecting the Cmdline creation method

Cmdline resources are general-purpose resources. By defining three scripts for starting, stopping, and monitoring the user application (program) to be monitored, you can create a resource that matches the purpose.
The scripts correspond to the following processes in RMS

Start script

It starts a program in Online processing.

Stop script

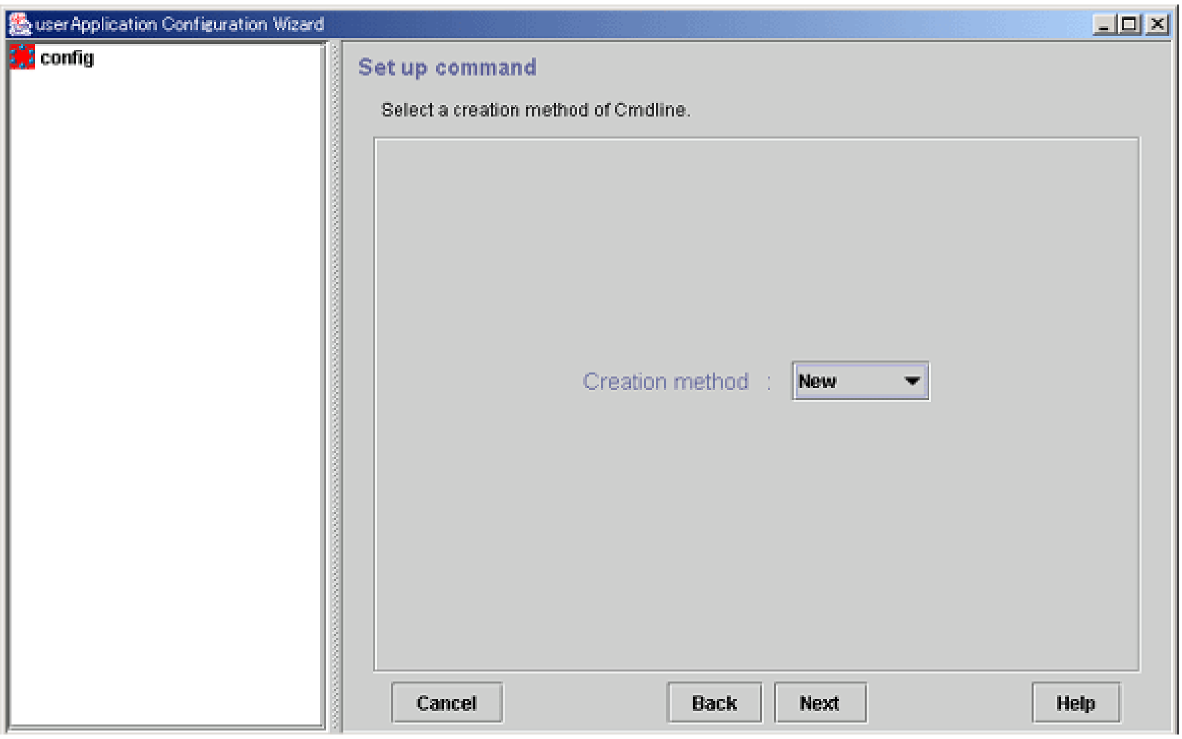
It stops a program in Offline processing.

Check script

This script is started periodically.
It notifies the user program state.

Select a method of configuring the script files and commands as the resource.

Figure 6.10 Set up command



Creation methods

Select the creation method from the following:

- New

  You can create a new script file.

- Path Input

  You can specify the existing script or command.
  Click *Next*.

After completing the setup, click *Next*. This will lead you to the following operation according to the selected item:

New: Creating new script files

Path input: Creating scripts by path input

Creating new script files

Note

When you create a script, you must carefully create a script end value, a timeout value, and environmental variable. For details, see "Notes on script creation" in "6.6 Setting Up Online/Offline Scripts."

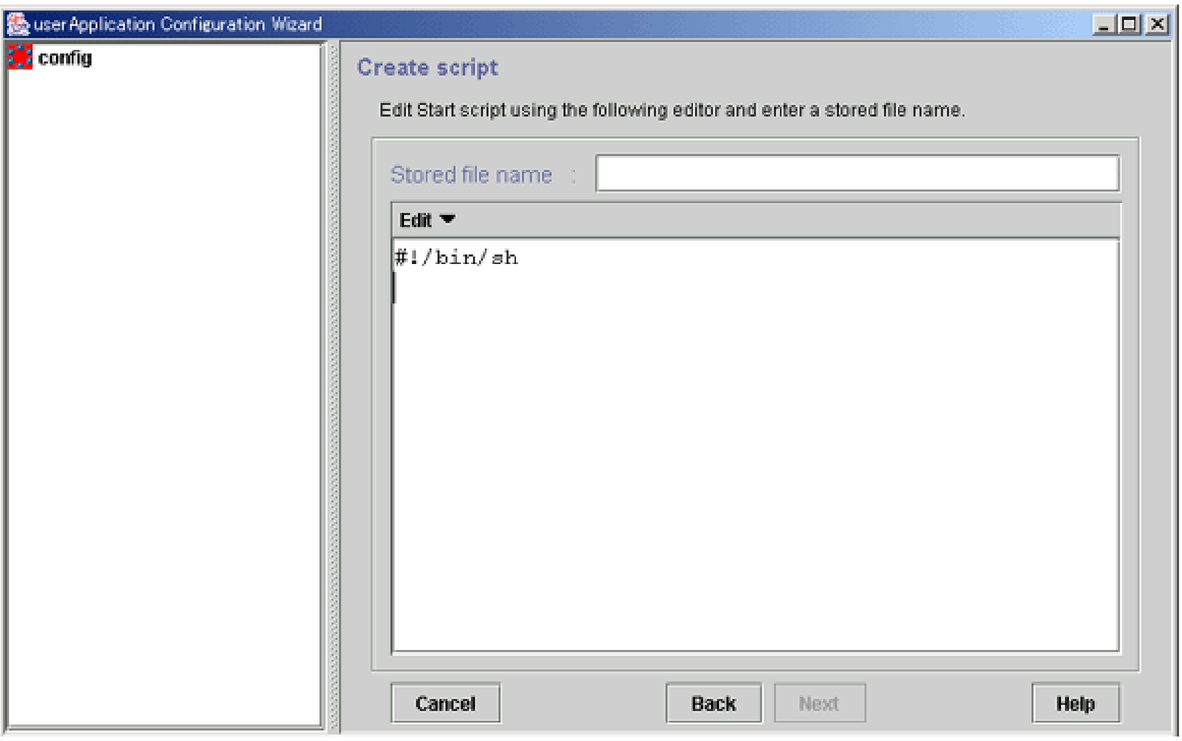
For details on the script sample, see "Sample scripts" in "6.6 Setting Up Online/Offline Scripts."

If you specify New for Creation method of Cmdline, the Start script creation screen appears.

Start script creation

The Cmdline contents can be input directly into the editing area. The text copy, cut, and paste operations are enabled only in the editing area.

Figure 6.11 Start script creation



Stored file name

Specify the file name to be saved.

Note

The following characters cannot be used in the file path and the arguments.

```
= \ ~ % @ &
```

If you need to use those characters, create the script specified with usable names and arguments, and then execute the script that starts, stops, and monitors the resource.

- Start script
  /opt/FJSVwvucw/scripts/start

- Stop script
  /opt/FJSVwvucw/scripts/stop

- Check script
  /opt/FJSVwvucw/scripts/check

Note

When specifying arguments, separate the arguments with "blank" characters. Since the scripts that can be created in this screen are Bourne shell by default, the following string is displayed at the beginning of the editing area:

```
#!/bin/sh
```

Edit

You can edit scripts from the simplified menu. The editing operation is described below:

Item	Operation
Move cursor	You can move the cursor to any position by placing the mouse pointer at the target position and single-clicking the mouse pointer or by operating the arrow keys on the keyboard in the top, bottom, left, or right directions.
Insert text	You can input characters at the cursor position.
Delete text	To delete 1 character Press the [*Delete*] key to delete the character at the cursor position. Press the [*Backspace*] key to delete the character positioned directly before the cursor position. To delete a string with one or more characters, use the following keys: [*Delete*] key [*Backspace*] key *Edit* -> *Delete* from the menu bar
Cut text	Select a string with one or more characters in the editing area and go to *Edit* -> *Cut*.
Copy text	Select a string with one or more characters in the editing area and go to *Edit* -> *Copy*.
Paste text	Position the cursor at the position where the character string that was previously cut or copied within the editing area is to be pasted, and then select *Edit* -> *Paste*.

Note

These script files are not removed even when processing is cancelled during Cmdline creation, or the Cmdline resource or GUI package (FJSVwvucw) is removed. If you want to remove the files, you need to remove the "/opt/FJSVwvcuw/scripts/start" directory. For example, execute the "rm(1)" command to remove one script file as follows:

```
# rm /opt/FJSVwvucw/scripts/start/XXXXXXXX
```

To remove the directory, specify "/opt/FJSVwvucw/scripts/start" or "/opt/FJSVwvucw."

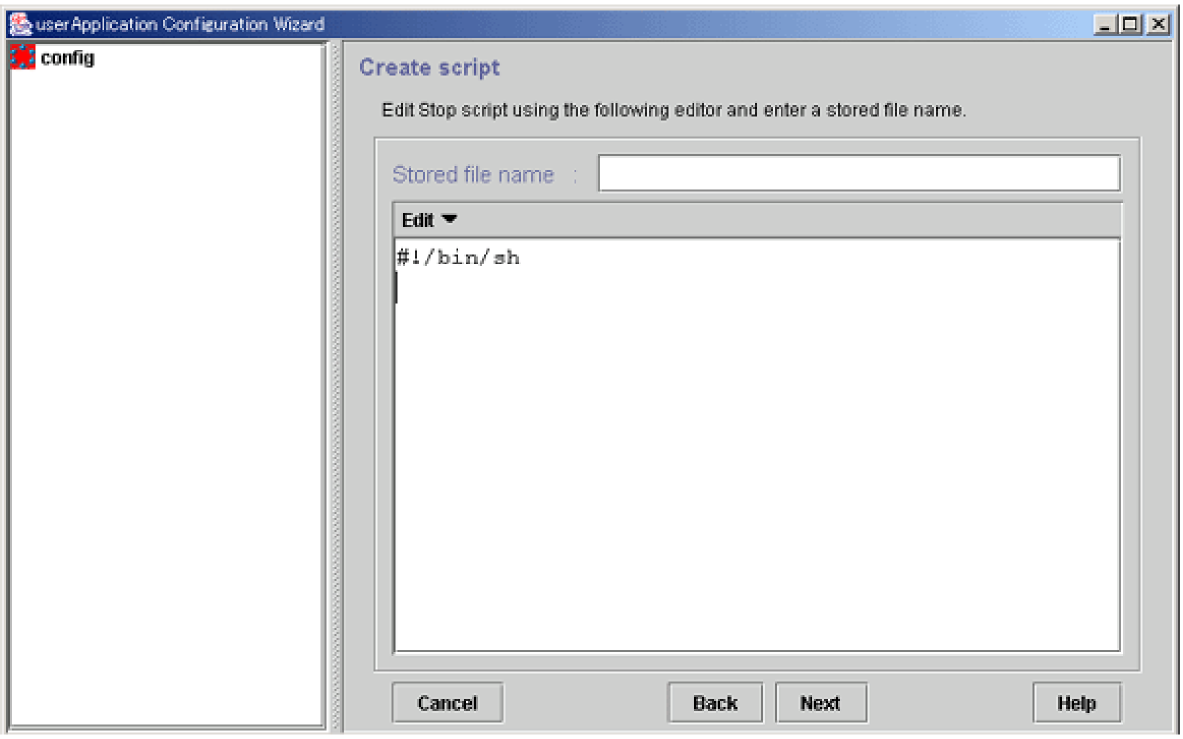
Note that, when you delete a Stop script or Check script, replace the "start" portion in the above directory with "stop" or "check."

Click *Next* to go to "Stop script creation."

Stop script creation

Create a Stop script by using the same procedure as that for the Start script.

Figure 6.12 Stop script creation



Click *Next* to go to "Check script creation."

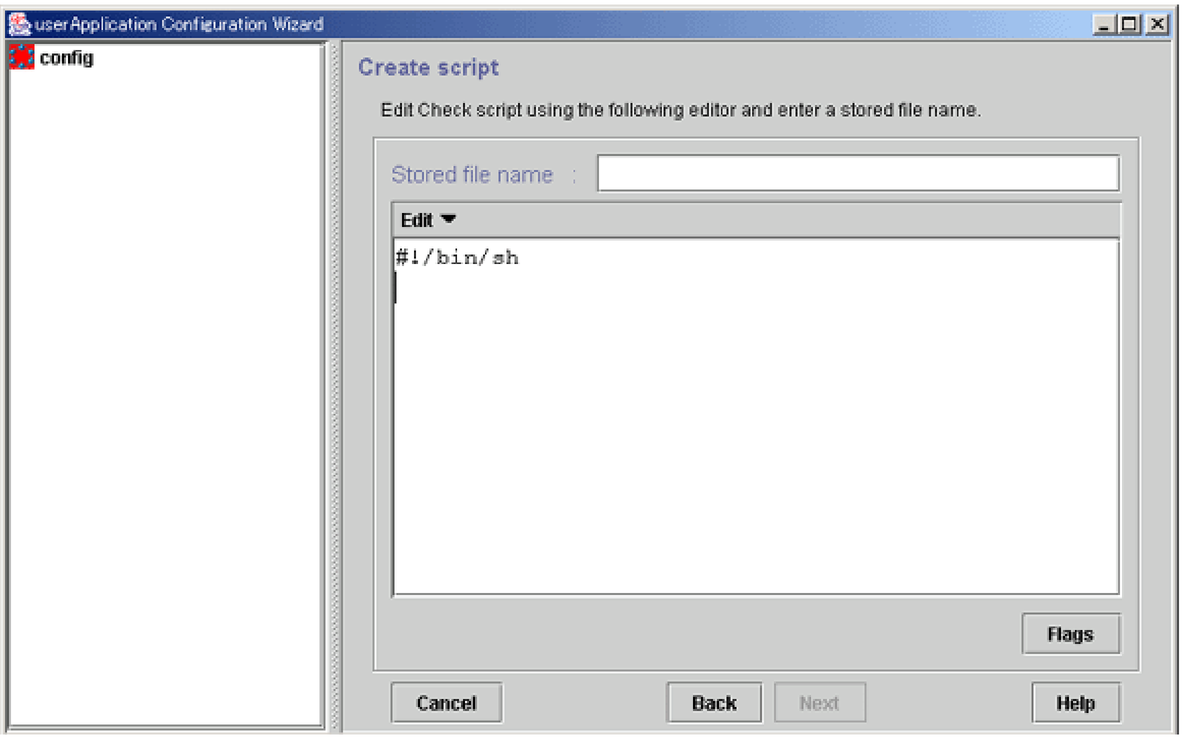
Note

You can omit setup of this script file. Just click *Next*.

Check script creation

Create a Check script by using the same procedure as that for the Start script.

Figure 6.13 Check script creation



Flags button

You can set script attributes from a dialog box. See "Setting up Cmdline flags," for details.

Note

If "NULLDETECTOR" was set to "Yes" at the "Set Cmdline Flag" screen, a Check script does not need to be configured. In this case, the Check script is not configured as resource information because the Check script is not started from RMS.

Creating scripts by path input

If a script already exists, you can enter a path to select the script.

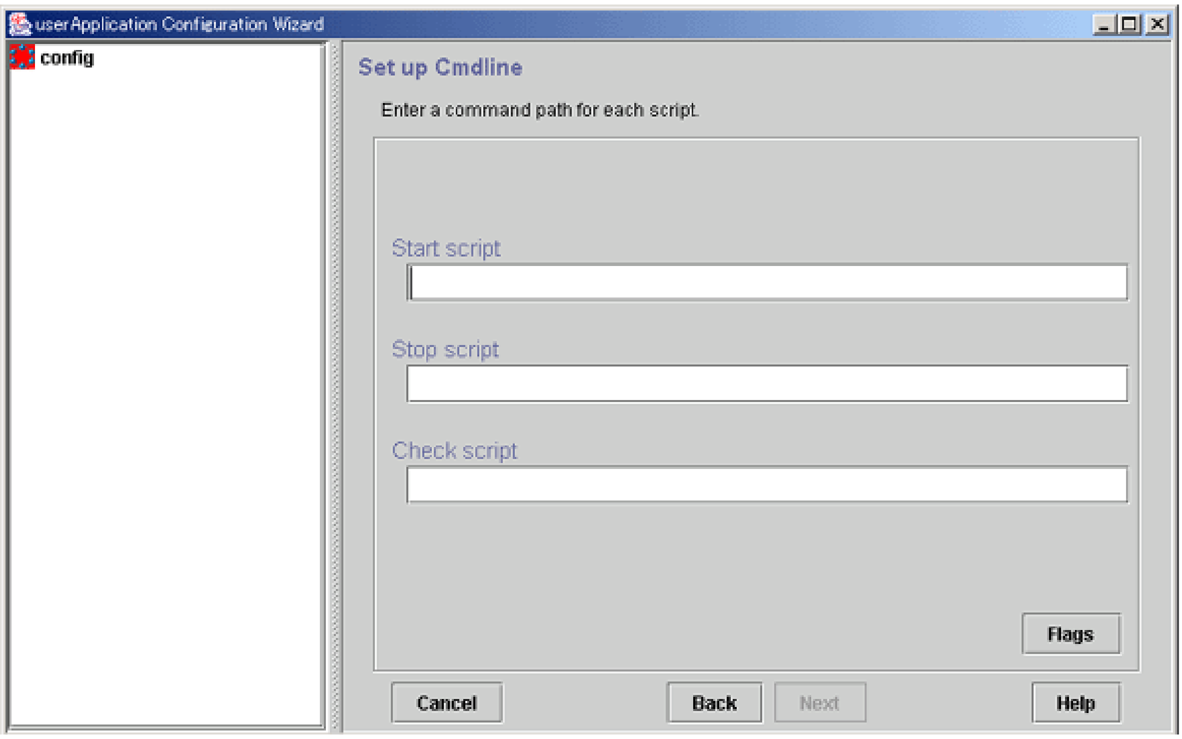
Note

The following characters cannot be used in the file path and the arguments.

```
= \ ~ % @ &
```

If you need to use those characters, create the script specified with usable names and arguments, and then execute the script that starts, stops, and monitors the resource.

Figure 6.14 Cmdline setup



Start script

Enter the path of the Start script. If you are specifying arguments, separate the arguments with "blanks."
The Start script is always a required script. The script must be entered with a full path name.

Stop script

Enter the path of the Stop script. If you are specifying arguments, separate the arguments with "blanks."
The Stop script is not required information. The script must be entered with a full path name.

Check script

Enter the path of the Check script. If you are specifying arguments, separate the arguments with "blanks." The script must be entered with a full path name.
Although the Check script is always a required script, it does not have to be specified if NULLDETECTOR was set to "Yes" in the "Set Cmdline Flag" screen. Even if the script is setup, the information is deleted during resource registration because the Check script is not started from RMS.

Flags button

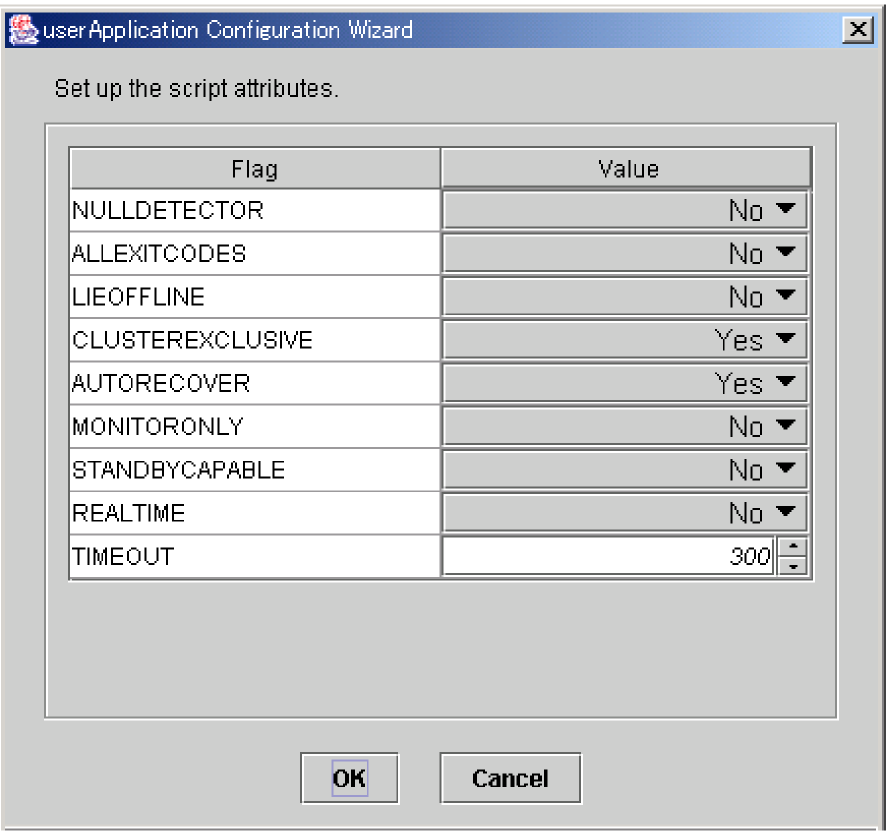
Set the script attributes. For information, see "Setting up Cmdline flags," which is described later.

Setting up Cmdline flags

The method of a controlling script defined to a Cmdline resource can be adjusted by changing the flag values.

The following default value is for Cmdline resource. Other resources have a default value individually.

For the attributes that can be set to the Cmdline resources, see the following.



After completing the setup, click *OK* to return to the "Set up Cmdline" screen.

Flag	Outline
NULLDETECTOR	If an attribute value is set to "Yes," Check script is disabled. The resource state is determined only depending on what Online or Offline script of the Cmdline resource shows when each script is executed in conjunction with Online or Offline processing of the userApplication. In this case, the resource state is unmonitored. Moreover, all values of other Flags are set to "No." Below indicates the relationship between the Check script and the default value: The Check script is set The default value is "No." The Check script is not set The default value is "Yes."
ALLEXITCODES	If the flag is set to "No", the exit code of the Check script will be interpreted as follows: 0: Online Other than 0: Offline If the flag is set to "Yes", the exit code of the Check script will be interpreted as follows: 0: Online 1: Offline 2: Faulted 3: Unknown 4: Standby 5: Online warning 6: Offline faulted The default value is "No." * Do not use them as return values within the script because the values displayed in 3, 5,

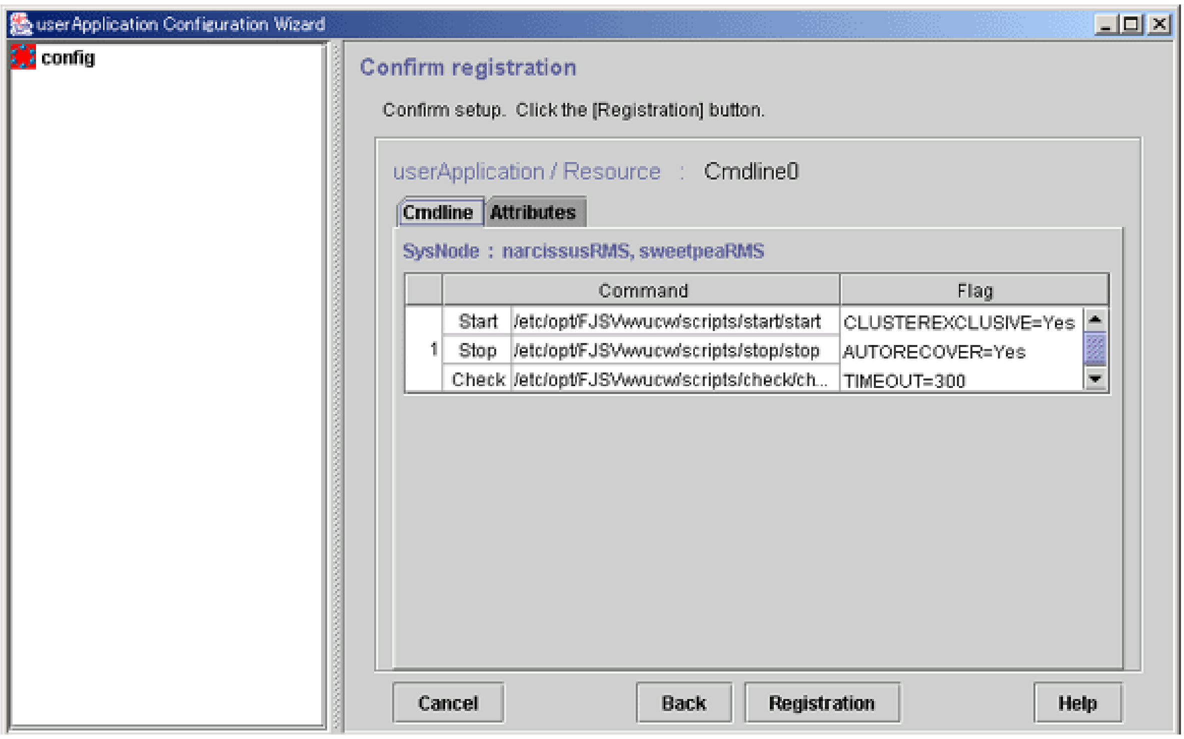
Flag	Outline
	6 and other than the values above indicate a special status. They are only allowed when PRIMECLUSTER products specified.
LIEOFFLINE	If the attribute is set to "Yes" and the Stop script is not set, the Offline processing of the resource is interpreted as it was processed successfully. However, for the resource status, the current status is displayed. If the Stop script is specified, the failure of the script triggers a fault processing. The default value is "No."
CLUSTEREXCLUSIVE	If the attribute is set to "Yes," the resource needs to be Online on a single node at the same time in a cluster system. If the resource becomes Online on two or more nodes at the same time because of a script problem and so on, the state of userApplication to which this resource belongs becomes Inconsistent. The default value is "Yes." It is recommended to set "Yes."
AUTORECOVER	If the attribute is set to "Yes," it tries to restart on the same node before userApplication is failed over in the even to a resource failure. This recovery is tried only once. The default value is "Yes."
MONITORONLY	This flag controls whether the "faulted" condition in the resource makes the userApplication Faulted. If this flag is set to yes, the resource becomes Faulted, but this does not make the userApplication Faulted. The system does not allow this value to be set to "Yes" for Cmdline resources. The default value is "No."
STANDBYCAPABLE	If this flag is set to "Yes," RMS sets the StandbyCapable attribute to 1 for this resource. For detailed information regarding this attribute, see "12 Appendix -Attributes" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*." The default value is "No."
REALTIME	If the flag is set to "No", the Check script will be started in the TS class (time-sharing scheduling) . If this flag is set to "Yes," the Check script is started with the RT class (real time mode). Note that since the operating system assigns the highest priority to processes that are started with the RT class, any bugs in the script or commands may have a large effect on system performance. The default value is "No."
TIMEOUT	This flag sets the timeout interval (seconds) for program start and stop processing. The default value is "300."

Check Cmdline registration information

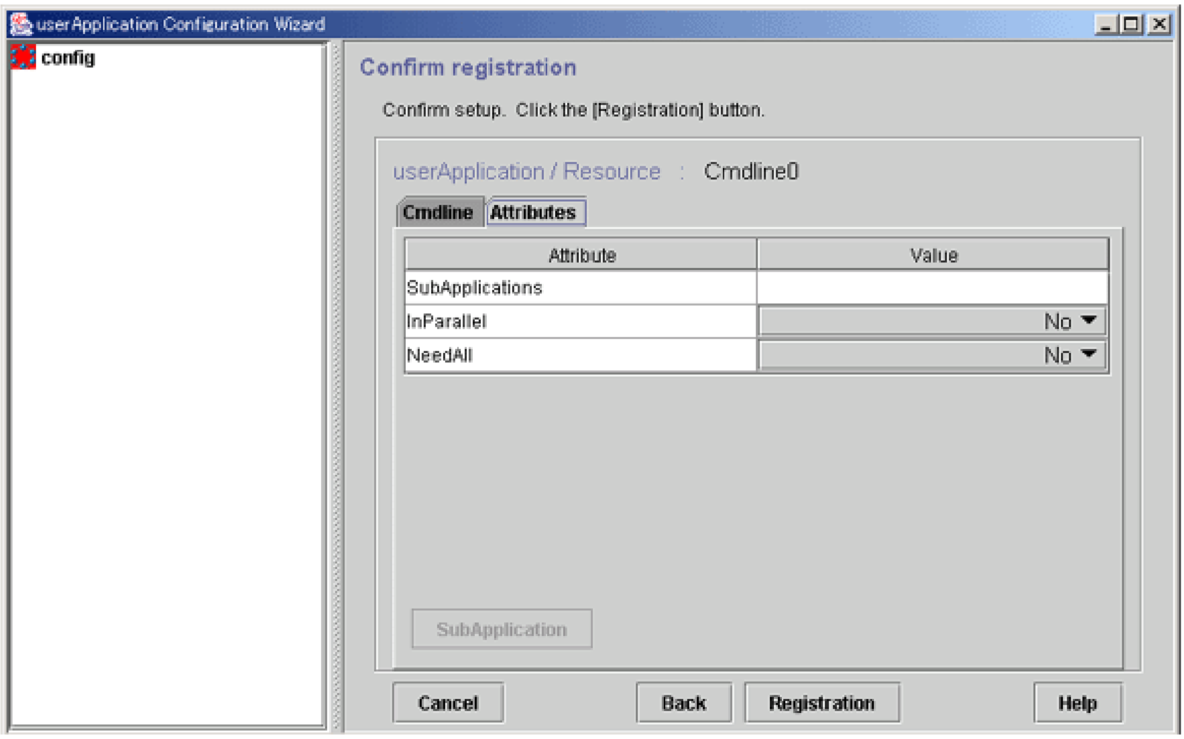
You can check a list of the scripts that has been set up previously. You can also configure resource attributes by selecting the *Attributes* tab and switching the screen.

For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.15 Confirmation of registration information



The attributes that are displayed in the "Flag" column are only those Resource attributes for which a value has been set. For example, if *No* is set for the AUTORECOVER attribute, then it means that AUTORECOVER attribute has not been set, and it is not displayed in the Flag column.



***SubApplication* button**

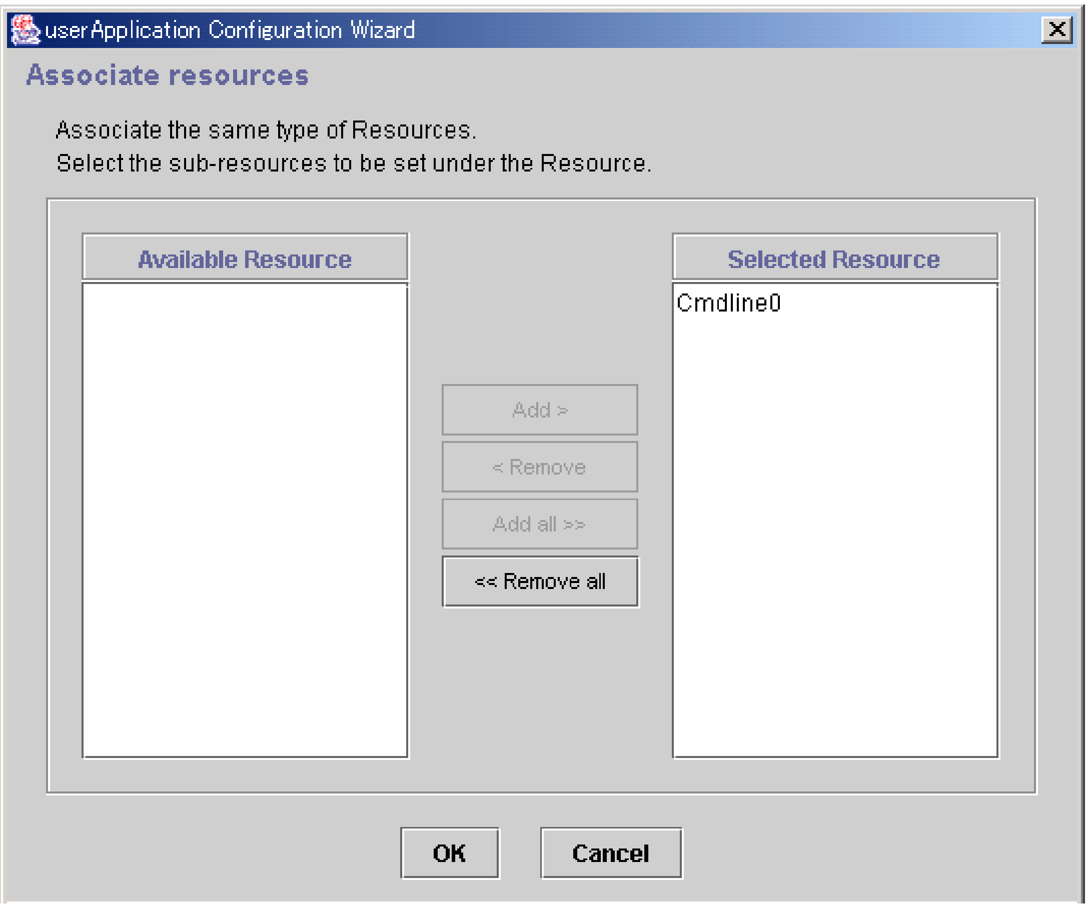
Associate the other Cmdline resources or process monitoring resources with the Cmdline resource that has been created above. See "Resource association."

After checking the registration information, click *Registration*.

Resource association

Use this function to configure a startup sequence among resources of the same type.
By assigning startup priorities to resources, you can clarify the order in which the resources are to be switched to Online or Offline.
In Online processing, resources that have a high startup priority are switched to Online before resources that have a low startup priority.
Conversely in Offline processing, resources that have a low startup priority are switched to Offline before resources that have a high startup priority.
If a startup priority is not assigned or if resources have the same priority, the sequence is undefined.

Figure 6.16 Associate resources



Available Resource

Available resource is referred to as the resource of the same type that can be associated, and satisfies the following conditions.

- The resource is of the same type as the resource that was called.
- The resource is not being used by another cluster application.

Note: Cmdline and process monitoring resources are considered the same type of resources.

Selected Resource

Resources to be set under the current resource being created.

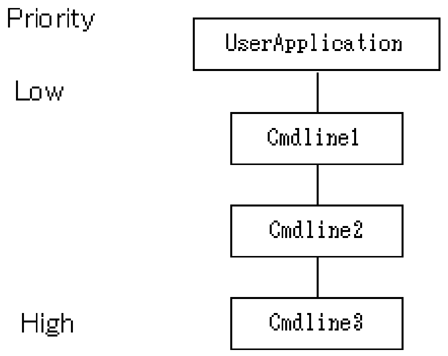
From *Available Resource* select the resource to be configured under the current resource, and then click *Add*. To add all listed resources, click *Add all*. To delete a resource from under the current resource, select the resource to be deleted from the *Selected Resource*, and then click *Remove*. To delete all listed resources, click *Remove all*.

After completing the setup, click *OK*, and return to the "Confirm Registration Information" screen.

Example

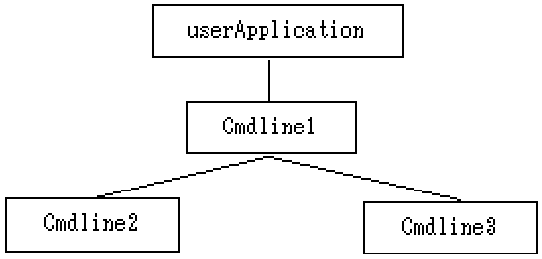
In the configuration shown below, the startup priority has been set to resources of the same type.

To build this configuration, build the resources according to the procedure shown below.



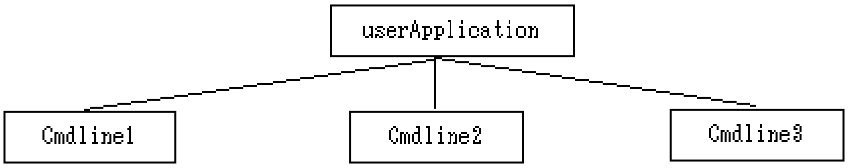
1. Create Cmdline3.
2. Create Cmdline2, and set Cmdline3 as subApplication.
3. Create Cmdline1, and set Cmdline2 as subApplication.
4. Create userApplication, and set Cmdline1 as subApplication.

The procedure for building the following configuration is shown below.



1. Create Cmdline3.
2. Create Cmdline2.
3. Create Cmdline1, and set Cmdline2 and Cmdline3 as subApplication.
4. Create userApplication, and set Cmdline1 as subApplication.

In the configuration shown below, the other resources were not set as subApplication during the creation of resource.



6.7.1.2 Creating Fsystem Resources

Set up an Fsystem resource if you want to mount a file system when userApplication is started.
To control multiple mount points in parent-child relationships, create the file system as one Fsystem resource.
For the notes when setting Fsystem Resource, see "6.13 Notes When Setting Fsystem Resource."
You need to work on the following prerequisites to create an Fsystem resource.

Mount point characteristics

- Mount points are used to mount UFS and other local file systems.
- The file system can be shared as an NFS file system in the network. The file system can also be set up so that it is not shared.

- When a file system is shared in the network, the NFS Lock Failover function becomes enabled.

Notes on using the file lock facility

If you use an NFS server function, set it so that the Lock information for NFS is succeeded.

If a transfer of the Lock information for NFS is not possible, you can get multiple Lock information for the same file. Therefore, the application that uses fcnlt(2) and lockf(3C) cannot execute the exclusive control. And it becomes the factor to cause file destruction.

But the following attention is necessary on using this function. Execute the next attention whether an application uses a lock unidentified.

- On the node operation with NFS servers, do not use the NFS client function. Do not implement NFS mount.
- A file system in which NFS is shared with the operation of NFS servers must be used only from the NFS client. Do not use a file directly from an application on a node on which NFS servers are operating.
- When failover occurs, the following processing is performed. The access from the client is reserved until the NFS lock recovery processing is completed though the state of the resource becomes Online when the start of statd and lockd is completed.
  - Stopping the statd(1M) and lockd(1M) daemon
  - Starting the statd(1M) and lockd(1M) daemon
  - NFS lock recovery processing
- When a mountpoint is shared using NFS, only one Fsystem resource including the mountpoint can be set to a single userApplication. In addition, the Fsystem resource must include the mountpoint where the NFS Lock Failover was set.

```
userApplication0
    Fsystem0
        mountpoint0  (NFS share, NFS Lock Failover)
        mountpoint1  (NFS share)
        mountpoint2  (NFS share)
        mountpoint3  (NFS share)
```

Note

It is not configured as about the following.

- Setting multiple Fsystem resources including the NFS lock Failover setting to a single userApplication

```
userApplication0
    Fsystem0
        mountpoint0  (NFS share, NFS Lock Failover)
        mountpoint1  (NFS share)
    Fsystem1
        mountpoint2  (NFS share, NFS Lock Failover)
        mountpoint3  (NFS share)
```

- Setting multiple Fsystem resources including the NFS share setting to a single userApplication

```
userApplication0
    Fsystem0
        mountpoint0  (NFS share, NFS Lock Failover)
        mountpoint1  (NFS share)
    Fsystem1
        mountpoint2  (NFS share)
        mountpoint3  (NFS share)
```

Note that the effect that NFS Lock cannot be temporarily get from NFS clients for the NFS file system controlled by userApplication1 may occur in the case of a failover on userApplication0 with the configuration below.

- Setting was multiple userApplication/Fsystem

```
userApplication0
    Fsystem0
```

```
        mountpoint0  (NFS share, NFS Lock Failover)
        mountpoing1  (NFS share)
userApplication1
    Fsystem1
        mountpoint2  (NFS share, NFS Lock Failover)
        mountpoing3  (NFS share)
```

Notes on using NFS

Protocols

NFS can use TCP and UDP protocols. After using TCP for failover and failback in sequence, recovery sometimes takes longer than with UDP. Therefore, please use the UDP protocol in PRIMECLUSTER.

Creating 31 or more mount points in Fsystem resources

The minimum Timeout value of the Fsystem resource is determined by the formula " No. of mount points in Fsystem resources x 6 seconds".

The default timeout is 180 seconds. If you set 31 or more mount points in Fsystem resources, it is necessary to change Timeout to the proper value based on "No. of mount points in Fsystem resources x 6 seconds" by selecting the *Attributes* tab in the screen shown in "Checking Fsystem registration information."

6.7.1.2.1 Prerequisites

This section describes operations that must be performed before the file system is set up with the GUI. Although not mentioned in the procedure, editor commands such as those featured by vi(1) are generally used in file editing.

Editing the /etc/vfstab.pcl file

If using ZFS

1. Register the ZFS Storage Pool to /etc/vfstab.pcl

   Perform the following procedure up to "2 Registering the legacy file system to /etc/vfstab.pcl" with all nodes which configure the cluster.

   Add the entry to /etc/vfstab.pcl for the ZFS storage pool created above. Set the following information for the respective items.

   - bdev : "#RMS#<ZFS storage pool name>"
   - cdev : "<ZFS storage pool name>"
   - mount point : "<mountpoint>"
   - fstype : "zfs"
   - runlevel, auto mount, flags : "-"

   If the mountpoint is not set to the ZFS storage pool (default), set the name of the mountpoint as the ZFS storage pool name with the prefix "/".

   Below is an example of registering two entries, the one of which is the ZFS storage pool named app1 and the mountpoint is set to /mnt, the other is the ZFS storage pool named app2 and the mountpoint is not set.

   ```
   # bdev cdev mountpoint fstype runlevel auto mount flags
   #RMS#app1 app1 /mnt zfs - - -
   #RMS#app2 app2 /app2 zfs - - -
   ```

2. Register the legacy file system to /etc/vfstab.pcl

   If one has created a legacy file system, in addition to the ZFS storage pool, add entries for each of the legacy file systems. Set the following information for the respective items.

   - bdev : "#RMS#<file system name>"
   - cdev : "<file system name>"

- mount point : "<mountpoint>"

- fstype : "zfs"

- runlevel, auto mount, flags : "-"

Below, is an example of settings for when creating two legacy file system app/mpl and app/mp2 to the ZFS storage pool app and with their respective mountpoints being /appdata1 and /appdata2.

```
# bdev cdev mountpoint fstype runlevel auto mount flags
#RMS#app app /app zfs - - -
#RMS#app/mp1 app/mp1 /appdata1 zfs - - -
#RMS#app/mp2 app/mp2 /appdata2 zfs - - -
```

If using UFS

Define the mount point in the "/etc/vfstab.pcl" file on all nodes where userApplication is configured to use Fsystem.

```
#RMS#/dev/sfdsk/class0001/dsk/volume0001 /dev/sfdsk/class0001/rdsk/volume0001 /disk1 ufs - no -
#RMS#/dev/sfdsk/class0001/dsk/volume0002 /dev/sfdsk/class0001/rdsk/volume0002 /disk2 ufs - no -
```

Note

- Lines must begin with "#RMS#".

- You cannot set different volumes on the same mount point.

Editing autofs

The file system and mount point to register into the Fsystem resources cannot be used in autofs.

When the device of a file system and the directory of a mount point added in "Editing the /etc/vfstab.pcl file" are used in autofs, delete the corresponding entry from the /etc/auto_master.

Preparations for sharing a file system in a network (NFS)

See

For settings when sharing the non-legacy ZFS file systems in NFS, see "6.4.1.2 Setup Procedure."

1) Create Ipaddress or Gls resources.

Since a takeover network is necessary when an NFS service is used as a cluster application, create an Ipaddress or a Gls resource.

For details on the setup procedure, see "6.7.1.5 Creating Takeover Network Resources" and "6.7.1.4 Creating Gls Resources."

Note

Set this resource to the same userApplication as the Fsystem resource.

2) Carry out preparations for enabling NFS Lock Failover.

Select the "Edit global settings in Configuration" command from the menu, and specify the settings for enabling NFS Lock Failover.

For details on the setup procedure, see "6.7.4 Editing global settings in Configuration."

Note that you must register one or more UFS or the legacy ZFS file systems with Fsystem resources as the location of directories. NFS file lock information to be registered in "6.7.4 Editing global settings in Configuration" is stored there.

Moreover, you need to change the version of NFS to use and the settings of SMF for enabling NFS Lock Failover. Perform the following procedure corresponding to the environment you are using.

- When you use NFS Lock Failover in Solaris 10, perform "Prerequisites for NFS Lock Failover (for Solaris 10)."

- When you use NFS Lock Failover in Solaris 11, perform "Prerequisites for NFS Lock Failover (for Solaris 11)."

Prerequisites for NFS Lock Failover (for Solaris 10)

Perform the following procedure for all cluster nodes to use NFS Lock Failover.

1) Log in to the corresponding cluster node as system administrator.

2) Change the setting file of NFs /etc/default/nfs as below.

```
# Sets the maximum version of the NFS protocol that will be registered
# and offered by the server. The default is 4.
#NFS_SERVER_VERSMAX=4
NFS_SERVER_VERSMAX=3
       .
       .
# Sets the maximum version of the NFS protocol that will be used by
# the NFS client. Can be overridden by the "vers=" NFS mount option.
# If "vers=" is not specified for an NFS mount, this is the version
# that will be attempted first. The default is 4.
#NFS_CLIENT_VERSMAX=4
NFS_CLIENT_VERSMAX=3
```

3) Execute the command below for each node to configure a cluster.

```
# /usr/bin/script -a /var/opt/reliant/log/smfchg_cnf.log
# /usr/sbin/svcadm -v disable -s network/nfs/client
# /usr/sbin/svcadm -v disable -s network/nfs/server
# /usr/sbin/svcadm -v disable -s network/nfs/nlockmgr
# /usr/sbin/svcadm -v disable -s network/nfs/status
# /usr/sbin/svccfg -v delete -f network/nfs/status
# /usr/sbin/svccfg -v delete -f network/nfs/nlockmgr
# /usr/sbin/svccfg -v delete -f network/nfs/server
# /usr/sbin/svccfg -v delete -f network/nfs/client
# /usr/sbin/svccfg -v import /var/svc/manifest/network/nfs/status.xml
# /usr/sbin/svccfg -v import /var/svc/manifest/network/nfs/nlockmgr.xml
# /usr/sbin/svccfg -v import /var/svc/manifest/network/nfs/server.xml
# /usr/sbin/svccfg -v import /var/svc/manifest/network/nfs/client.xml
# /usr/sbin/svccfg -v -s network/nfs/status setprop network/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/status setprop rpcbind/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/status setprop filesystem-local/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/status addpg startd framework
# /usr/sbin/svccfg -v -s network/nfs/status addpropvalue startd/duration astring: transient
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop network/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop rpcbind/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop status/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop filesystem-minimal/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr addpg startd framework
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr addpropvalue startd/duration astring: transient
# /usr/sbin/svccfg -v -s network/nfs/server setprop nlockmgr/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/client setprop nlockmgr/restart_on=astring: none
# /usr/sbin/svccfg -s svc:/network/nfs/status setprop stop/exec = astring: ":true"
# /usr/sbin/svccfg -s svc:/network/nfs/nlockmgr setprop stop/exec = astring: ":true"
# /usr/sbin/svcadm -v refresh network/nfs/status
# /usr/sbin/svcadm -v restart network/nfs/status
# /usr/sbin/svcadm -v enable -s network/nfs/status
# /usr/sbin/svcadm -v refresh network/nfs/nlockmgr
# /usr/sbin/svcadm -v enable -s network/nfs/nlockmgr
# /usr/sbin/svcadm -v refresh network/nfs/server
# /usr/sbin/svcadm -v enable -s network/nfs/server
# /usr/sbin/svcadm -v refresh network/nfs/client
# /usr/sbin/svcadm -v enable -s network/nfs/client
```

Note

It may fail to activate network/nfs/server service, but continue the process even if this would happen.

4) Confirm that three SMF services below are online.

```
# /usr/bin/svcs | grep nfs
online          0:51:53 svc:/network/nfs/client:default
online          0:51:54 svc:/network/nfs/status:default
online          0:51:54 svc:/network/nfs/nlockmgr:default
```

5) Exit the script command.

```
# exit
```

Prerequisites for NFS Lock Failover (for Solaris 11)

Execute the following procedures in all the cluster nodes which use the NFS Lock Failover.

1) Login to the target cluster node using a system administrator.

2) Execute the following commands for each node which configures the clusters.

```
# /usr/bin/script -a /var/opt/reliant/log/smfchg_cnf.log
# /usr/sbin/svcadm -v disable -s network/nfs/client
# /usr/sbin/svcadm -v disable -s network/nfs/server
# /usr/sbin/svcadm -v disable -s network/nfs/nlockmgr
# /usr/sbin/svcadm -v disable -s network/nfs/status
# /usr/sbin/svccfg -v delete -f network/nfs/status
# /usr/sbin/svccfg -v delete -f network/nfs/nlockmgr
# /usr/sbin/svccfg -v delete -f network/nfs/server
# /usr/sbin/svccfg -v delete -f network/nfs/client
```

3) Start svccfg, and update the settings for SMF.

```
# svccfg
  svc:> select network/nfs/status
  svc:/network/nfs/status> delcust
  svc:/network/nfs/status> select network/nfs/nlockmgr
  svc:/network/nfs/nlockmgr> delcust
  svc:/network/nfs/nlockmgr> select network/nfs/server
  svc:/network/nfs/server> delcust
  svc:/network/nfs/server> select network/nfs/client
  svc:/network/nfs/client> delcust
  svc:/network/nfs/client> quit
```

4) Execute the following commands for the each node which configure the clusters.

```
# /usr/sbin/svccfg -v import /lib/svc/manifest/network/nfs/status.xml
# /usr/sbin/svccfg -v import /lib/svc/manifest/network/nfs/nlockmgr.xml
# /usr/sbin/svccfg -v import /lib/svc/manifest/network/nfs/server.xml
# /usr/sbin/svccfg -v import /lib/svc/manifest/network/nfs/client.xml
```

5) Change the NFS settings as shown below.

```
# /usr/sbin/sharectl set -p server_versmax=3 nfs
# /usr/sbin/sharectl set -p client_versmax=3 nfs
```

6) Execute the following commands for the each node which configures the clusters in sequence.

```
# /usr/sbin/svccfg -v -s network/nfs/status setprop network/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/status setprop rpcbind/restart_on=astring: none
```

```
# /usr/sbin/svccfg -v -s network/nfs/status setprop filesystem-local/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/status addpg startd framework
# /usr/sbin/svccfg -v -s network/nfs/status addpropvalue startd/duration astring: transient
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop network/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop rpcbind/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop status/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr setprop filesystem-minimal/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr addpg startd framework
# /usr/sbin/svccfg -v -s network/nfs/nlockmgr addpropvalue startd/duration astring: transient
# /usr/sbin/svccfg -v -s network/nfs/server setprop nlockmgr/restart_on=astring: none
# /usr/sbin/svccfg -v -s network/nfs/client setprop nlockmgr/restart_on=astring: none
# /usr/sbin/svccfg -s svc:/network/nfs/status setprop stop/exec = astring: ":true"
# /usr/sbin/svccfg -s svc:/network/nfs/nlockmgr setprop stop/exec = astring: ":true"
# /usr/sbin/svcadm -v refresh network/nfs/status
# /usr/sbin/svcadm -v restart network/nfs/status
# /usr/sbin/svcadm -v enable -s network/nfs/status
# /usr/sbin/svcadm -v refresh network/nfs/nlockmgr
# /usr/sbin/svcadm -v enable -s network/nfs/nlockmgr
# /usr/sbin/svcadm -v refresh network/nfs/server
# /usr/sbin/svcadm -v enable -s network/nfs/server
# /usr/sbin/svcadm -v refresh network/nfs/client
# /usr/sbin/svcadm -v enable -s network/nfs/client
```

Note

It may fail to activate network/nfs/server service, but continue the process even if this would happen.

7) Confirm the following three SMF services are online.

```
# /usr/bin/svcs | grep nfs
online          0:51:53 svc:/network/nfs/client:default
online          0:51:54 svc:/network/nfs/status:default
online          0:51:54 svc:/network/nfs/nlockmgr:default
```

8) Exit the script command.

```
# exit
```

3) Set up the NFS entry of PRIMECLUSTER

When all file systems to be shared in NFS are the non-legacy ZFS file systems, make a directory with UFS or the legacy ZFS file systems separately, and then add the entry of the file system. The NFS file lock information registered in "6.7.4 Editing global settings in Configuration" is stored there.

When all file systems to be shared in NFS are UFS or the legacy ZFS file systems, add entries of the file systems.

In both cases, edit the /etc/dfs/dfstab.pcl file.add. Add the entries of the file systems to be shared in NFS on all nodes that make up userApplication.

```
#RMS# share -F nfs /disk1
#RMS# share -F nfs /disk2
```

Note

Be sure to start each entry with "#RMS#."

The definition of each entry follows when the file system is shared in NFS manually by the share (1M) command.

For UFS or the legacy ZFS file systems, a file system can be shared in NFS by the mount point defined in /etc/vfstab.pcl.

For the non-legacy ZFS file systems, the sharing process of NFS on Fsystem resources is not performed because ZFS performs the sharing process of NFS. NFS shared status is monitored when sharenfs property or share.nfs property is set to "on". At this time, the mountpoint of dataset is monitored to be described in /etc/dfs/sharetab.

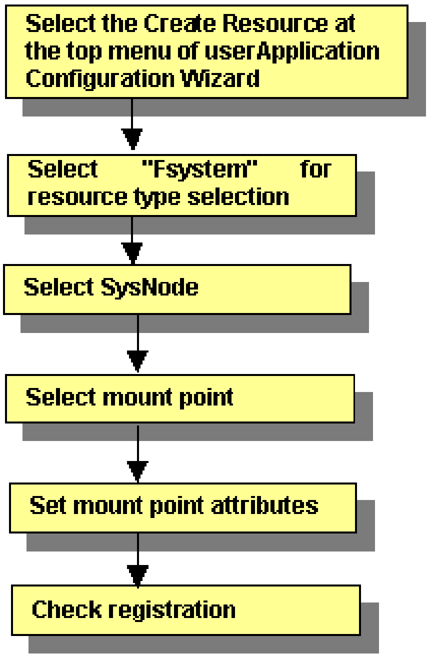
If another file system is used

The disk must be formatted and the file system must be created beforehand. For information on formatting and file system commands, see the "Solaris X Reference Manual Collection." If you plan to use ZFS as the file system, see "6.4.1 If using ZFS" when you build the file system.

6.7.1.2.2 Setup Method

This section describes how to create Fsystem resources.

Figure 6.17 Flow of Fsystem resource creation

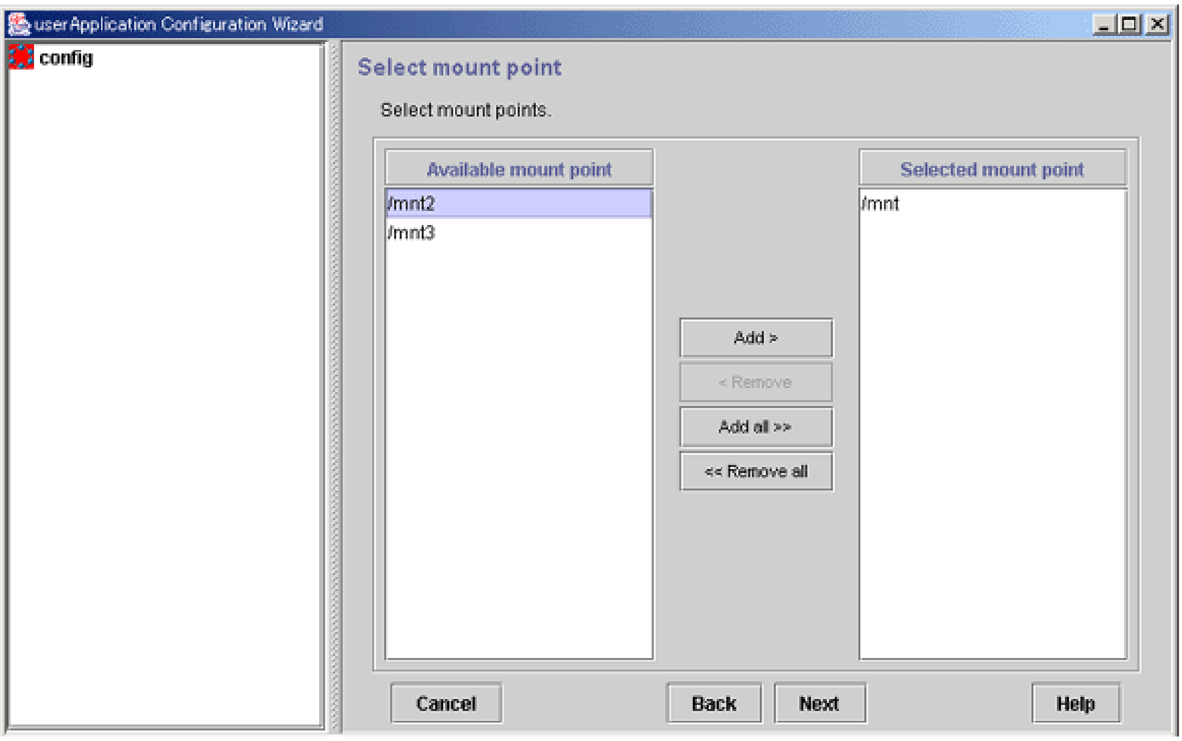


For information on the above operations up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resources." This section describes the operations starting from "Select mount point."

Selecting mount points

Select the mount points.

Figure 6.18 Select mount point



Available mount point

Names of mount points where the file system can be mounted.
The mount points that were configured for "/etc/vfstab.pcl" (described earlier) are displayed.

Selected mount point

Names of mount points to be used for file system mounting.

From *Available mount point*, select the mount point at which the file system is to be mounted, and then click *Add*. To add all listed mount points, click *Add all*. To delete a mount point on which the file system is not going to be mounted, select the mount point to be deleted from *Selected mount point*, and then click *Remove*. To delete all listed mount points, click *Remove all*.

After completing the setup, click *Next* to go to "Setting mount point attributes."

Note

Setup of the ZFS storage pool's resource is performed with the create Fsystem resource screen. Since the ZFS storage pool's highest level mount point that was defined to the /etc/vfstab.pcl file is displayed on the mount point selection screen, please select that and embed the resource in the userApplication.

Even if there is the ZFS dataset used as a ufs file system or legacy file system, the setup is performed with the create Fsystem resource screen.

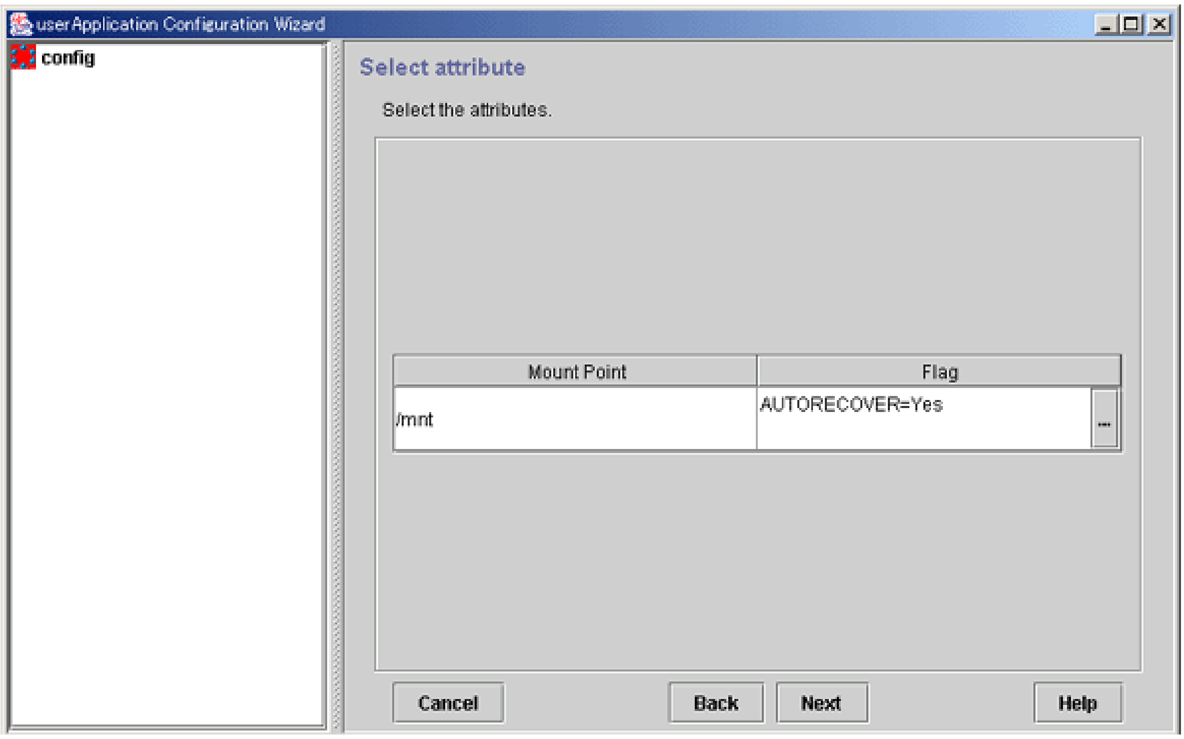
The setup of the legacy file system is also performed with the create Fsystem resource screen. Select the mount point with the mount point selection screen and embed the resource in the userApplication.

If the ZFS dataset used as a legacy file system exists on the ZFS storage pool of a shared disk, set a mountpoint corresponding to the ZFS storage pool's highest level mountpoint and the dataset of the legacy file system to one Fsystem resource.

Setting mount point attributes

Set attributes for the mount points.

Figure 6.19 Select attribute



Mount point

The mount name is displayed.

Flag

The attributes that are set for the individual mount points are displayed.

[...] button

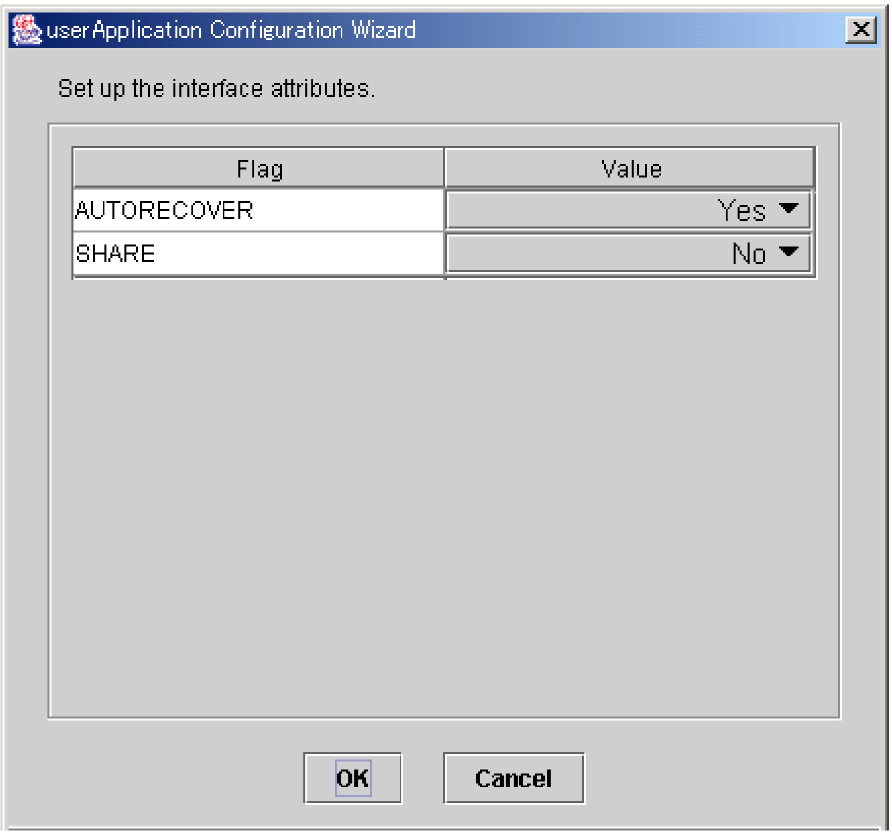
Press this button to set attributes for controlling mount point management. When this button is clicked, the Set Mount Point Attributes screen shown below is displayed.

After completing the setup, click *Next* to go to "Fsystem Registration Information Confirmation."

Attribute setup for mount points

Configure the attributes for controlling the mount points.

Figure 6.20 Attribute setup for mount points



After completing the setup, click *OK* to go to the "Select Attribute" screen.

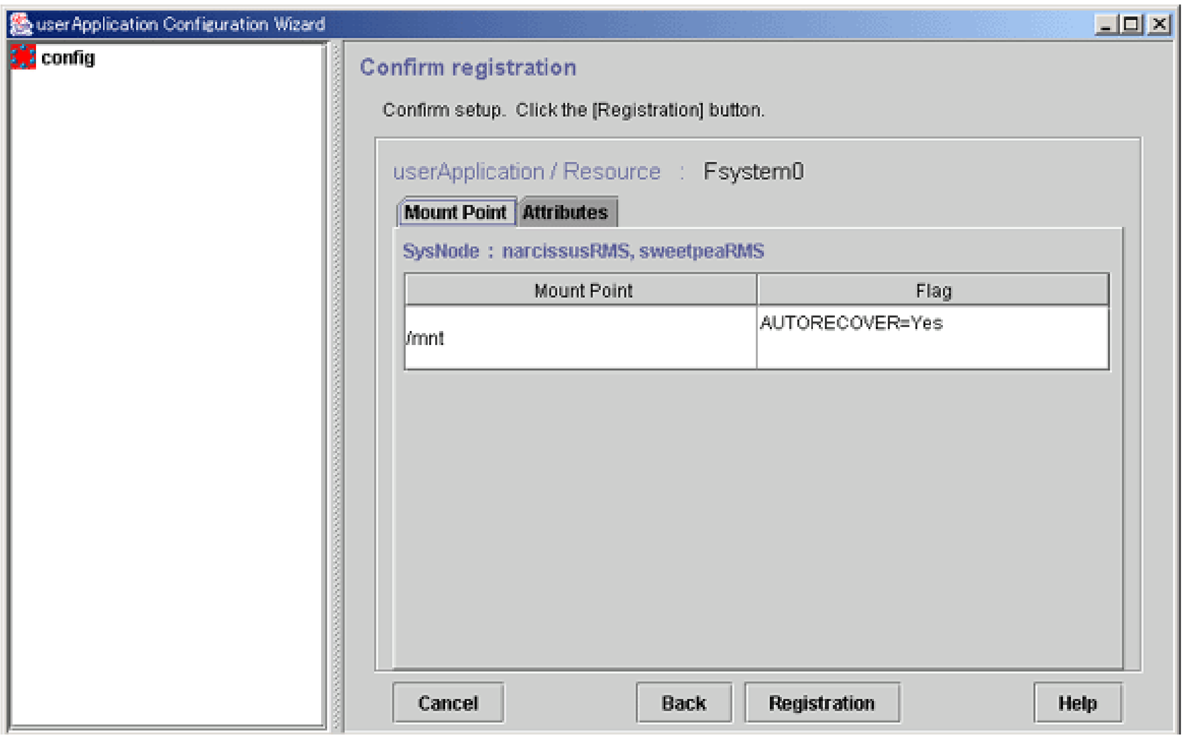
Flag	Outline
AUTORECOVER	If this flag is set to "Yes" and the specified file system is unmounted, RMS automatically attempts to remount the specified file system when it is unmounted. If this attempt fails, Fault processing is initiated. The default value is "Yes."
SHARE	If this flag is set to Yes, a mounted directory is shared (NFS). In such a case, make a specific entry to "/etc/dfs/dfstab.pcl". See "Preparations for sharing a file system in a network (NFS)" in "6.7.1.2.1 Prerequisites." The default value is "No."
NFSLOCKFAILOVER	If this flag is set to "Yes," Lock information for NFS Lock Failover is stored for this mount point. To use NFS Lock Failover, you must set this flag to "Yes" for one of the mount points. If you set this attribute to "Yes," you must also set SHARE to "Yes" at the same time. The default value is "No." For each Fsystem resource, this flag can be set for only one mount point. Even if userApplication has multiple Fsystem resources and NFS Lock Failover is to be used, set this attribute to "Yes" for only one mount point in userApplication.

Checking Fsystem registration information

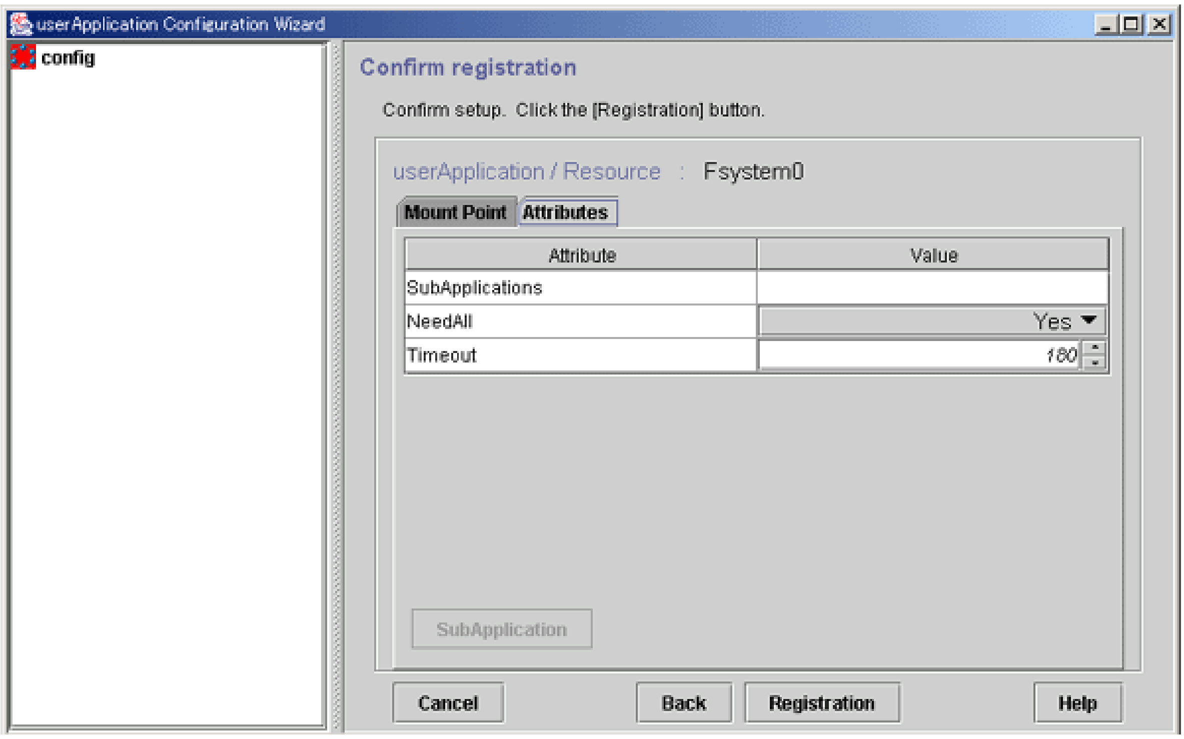
Check the Fsystem registration information. You can also select the *Attributes* tab, and set resource attributes by switching the screen.

For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.21 Confirmation of registration information



The attributes that are displayed in the "Flag" column are only those Resource attributes for which a value has been set. For example, if "No" is set for an attribute that takes a "Yes" or "No" setting, such as AUTORECOVER, that attribute is not displayed in the "Flag" column.



SubApplication button

This button is used for specifying other resources that the current resource depends on. For an Fsystem resource, this button is disabled. For multiple directories in parent-child relationships, configure the directories in one resource.

Check the registration contents, and then click *Registration*.

6.7.1.3 Creating Gds Resources

Set up resources for the disk classes that are defined in Global Disk Services (GDS).

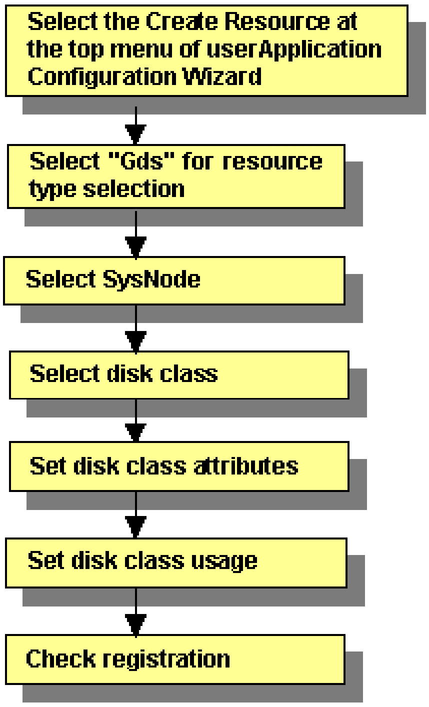
The disk classes must be created before this screen is used.

Note

- In the case of the single-node cluster operation, create GDS volumes in a local class and do not create Gds resources.
- If a disk class is registered to a Gds resource when equivalent volume copy is being executed, the equivalent copy operation is canceled. Therefore, the data becomes invalid in some of the slices. However, recovery is not necessary because equivalent copy is executed automatically when the volume is started. If you want to recover the slice state before the volume is started, complete Gds resource setup and then start equivalent copy. For instructions on starting equivalent copy, see "5.3.6 Copying Operation" or "D.10 sdxcopy - Synchronization copying operation" in the "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

For instructions on creating disk classes, see "6.3.2.3 Setting Up Shared Disks."

Figure 6.22 Flow of Gds resource creation

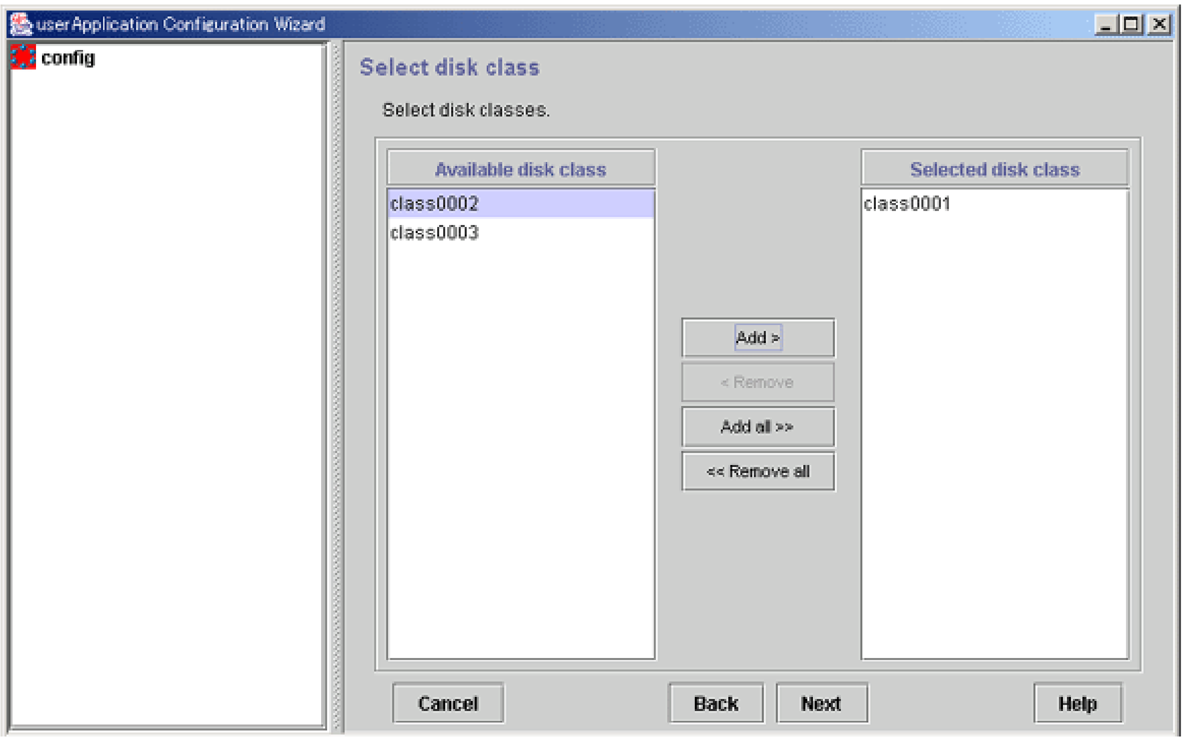


For information on the above operations up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resources." This section describes the operations starting from "Select disk class."

Selecting the disk class

Select the disk class.

Figure 6.23 Select disk class



Available Disk Class

Names of disk classes that can be set.
The disk classes (shared disks) that are defined in GDS are displayed.

Selected Disk Class

Names of disk classes to be set.

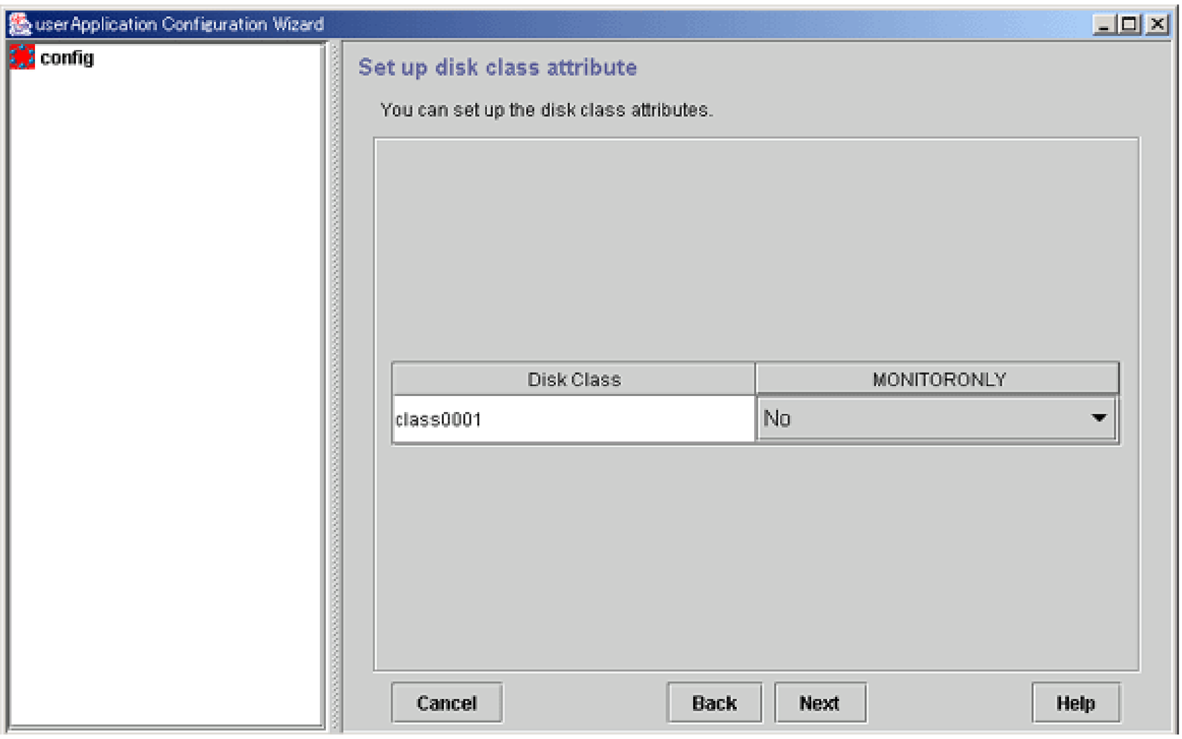
Select disk classes from *Available disk class*, and then click *Add*. To add all listed disk classes, click *Add all*. To delete a disk class to be set, select the disk class to be deleted from *Selected disk class*, and then click *Remove*. To delete all listed disk classes, click *Remove all*.

After completing the setup, click *Next* to go to the "Set Disk Class Attribute" screen.

Disk class attribute setup

Set the disk class attribute.

Figure 6.24 Set up disk class attribute



Disk Class

The disk classes (common disks) that were defined in GDS (described earlier) are displayed.

Flag	Outline
MONITORONLY	This flag determines whether disk class failures are to be reported to the userApplication. If "Yes" is set and a disk class failure occurs, the disk class is switched to faulted state but the Gds resources remain online, and userApplication failover does not occur. Carry out the preliminary design, including the higher applications" before determining the disk classes for which this flag is to be set to "Yes." The default value is "No."

The MONITORONLY attribute must be set to "No" for at least one disk class.

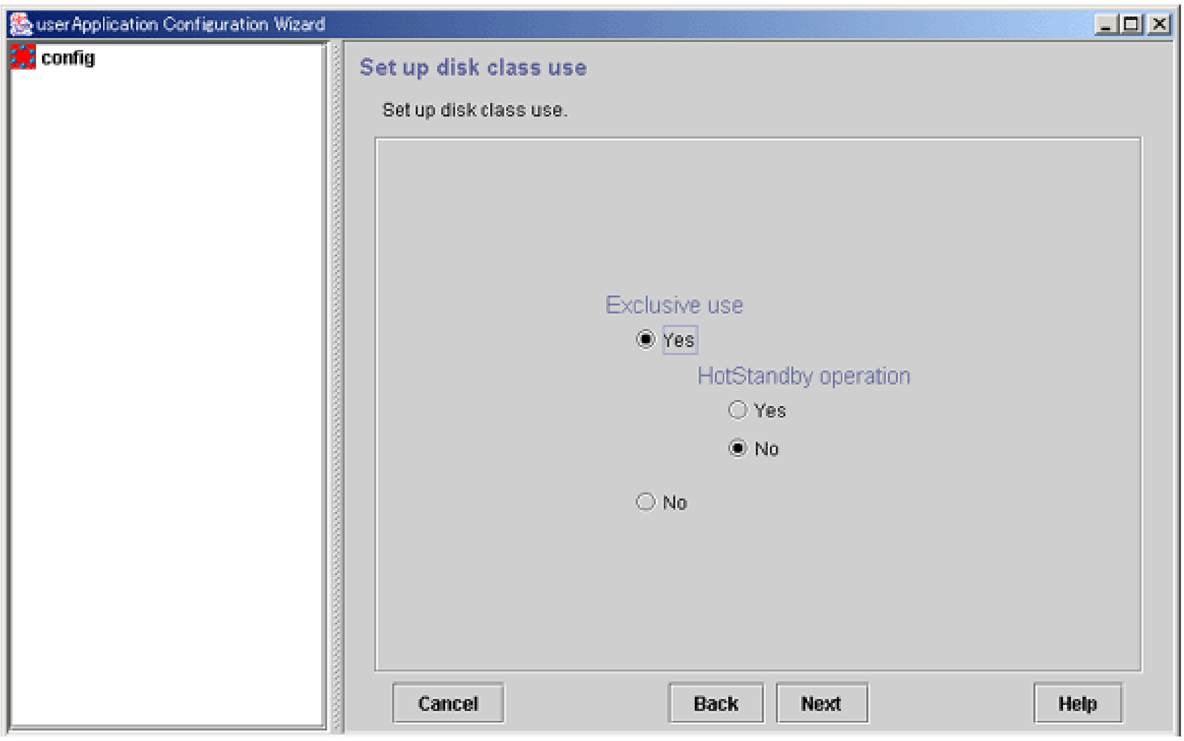
This will prevent userApplication failover in the event of all disk class failures and RMS' notifying the upper applications of the failures.

After completing the setup, click *Next* to go to the "Set up disk class use" screen.

Setting the disk class use

Configure the use of the disk class.

Figure 6.25 Set up disk class use



Exclusive use

Select "*Yes*" or "*No*" for *Exclusive use*. If you select "*Yes*", also select "*Yes*" or "*No*" for *HotStandby operation*.

Specify the options according to the use of the shared disk.
After completing the setup, click *Next* to go to the "Confirm Registration Information" screen.

Shared disk uses

Table 6.2 Shared disk uses and setting methods

Use	Exclusive use	HotStandby operation
Switched disk	Yes	No
Shared disk for simultaneous access	Yes	Yes
Simultaneous shared disk	No	-

Shared disks have the following features for each use:

- Switched disk

  Only the OPERATING node can use the shared disk.
  The non-OPERATING nodes cannot access the volume.

- Shared disk for simultaneous access

  All nodes can use the shared disk simultaneously.
  Select this use if applications that require disk access from a node other than the OPERATING node are set inside the same userApplication.
  The exclusive control for maintaining data integrity in the shared disk must be handled by the applications, and not by PRIMECLUSTER.

- Simultaneous shared disk

  Select this use if multiple userApplications share the disk classes. An example is when Oracle real application clusters are used.
  The exclusive control for maintaining data integrity in the shared disk must be handled by the applications, and not by PRIMECLUSTER.

Note

- During operation of a cluster application, try not to have applications or user processes that are not related to the cluster applications use the shared disks. If this occurs, the state transition might fail.

- Do not create a home directory for user accounts on a switchover disk. The reasons are as follows:

  - The file system on the switching disk is mounted only on the OPERATING node of the cluster application.

  - PRIMECLUSTER forcibly terminates process that uses the file system on the switching disk by executing the "kill(1M)" command to unmount the file system in the event of cluster application failover.

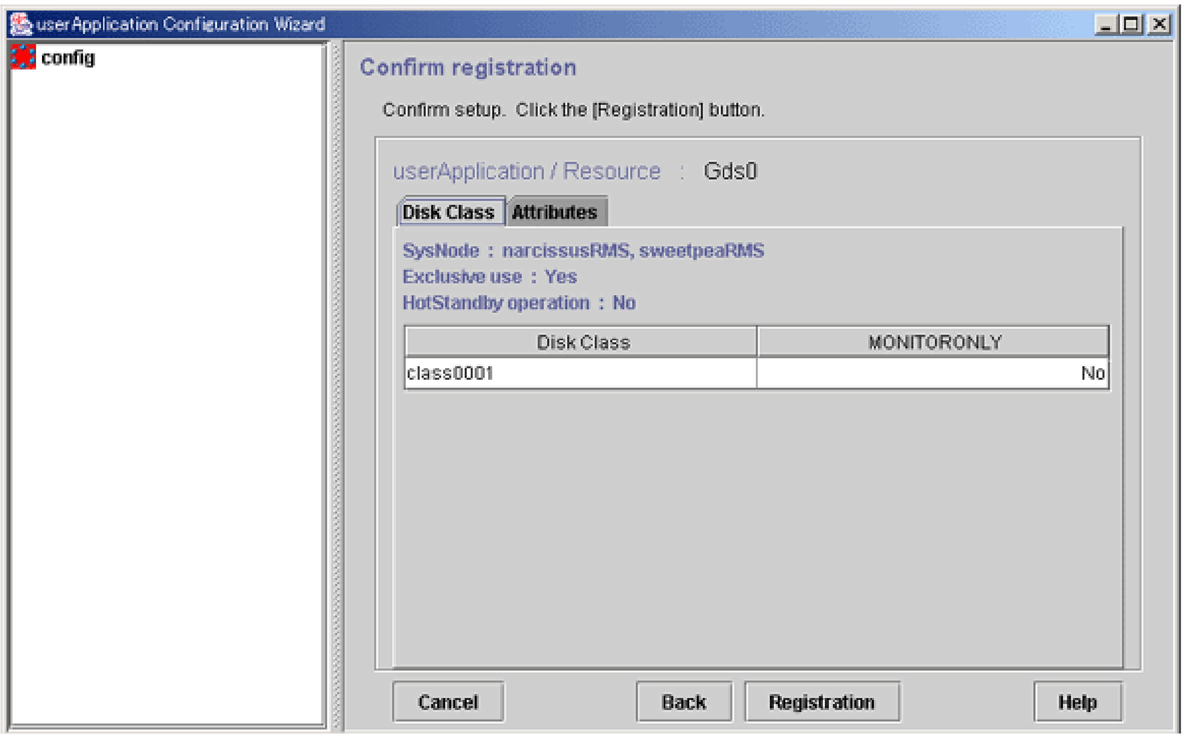
  - If the file system cannot be unmounted properly because of a failure of forced termination of process, PRIMECLUSTER might forcibly stop the node to prevent the file system data on the shared disk from being destroyed by double mounting.

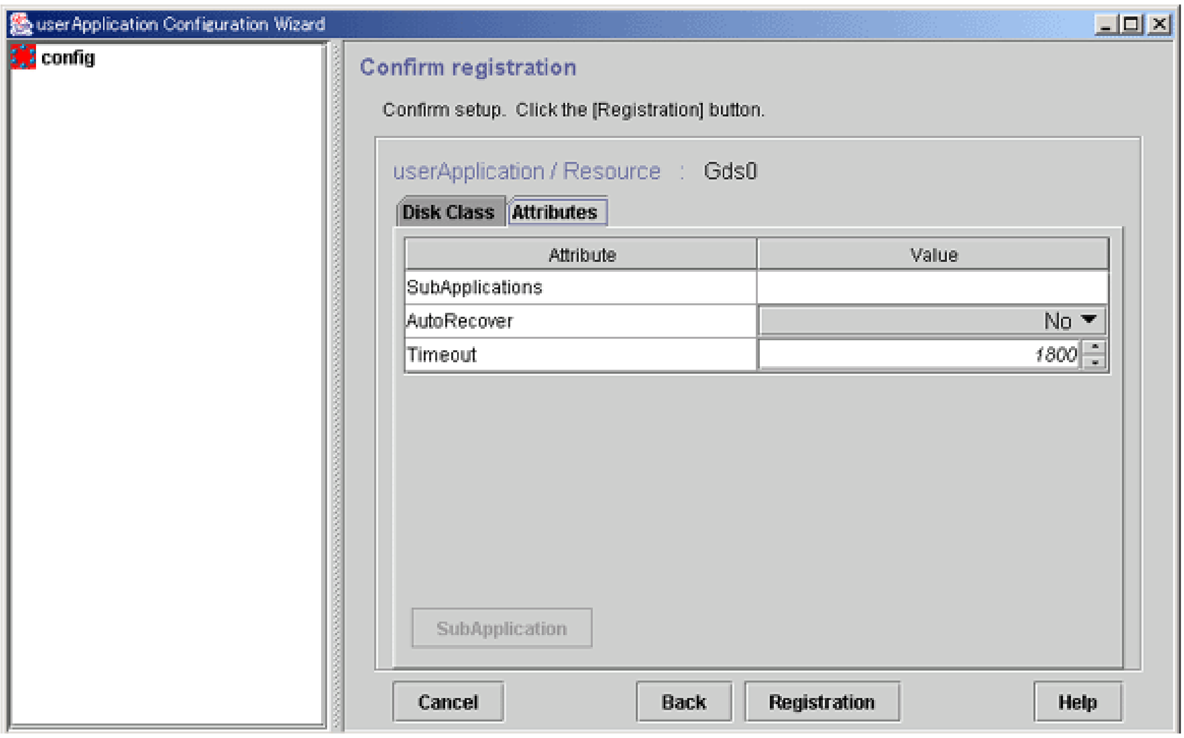
Checking registration information for Gds resources

Check the registration information for the Gds resources. You can also set resource attributes by selecting the *Attributes* tab and switching the screen.

For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.26 Confirm registration





***SubApplication* button**

Use this button to associate a previously created Gds resource under the current Gds resource. This button can be selected only if there are resources of the same type that can be associated. For setting instructions, see "Resource association" in "6.7.1.1 Creating Cmdline Resources."

Check the registration information, and then click *Registration*.

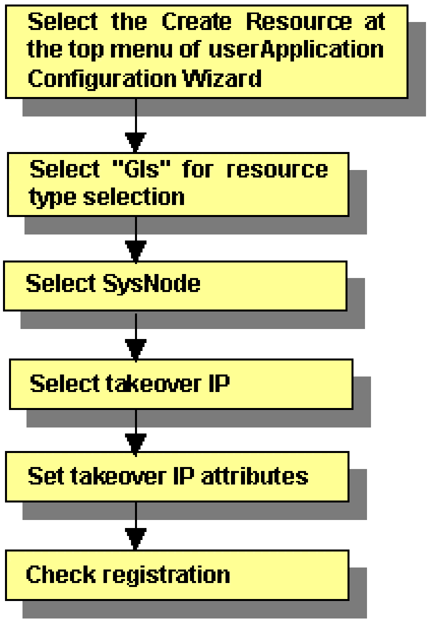
6.7.1.4 Creating Gls Resources

Configure the takeover IP addresses that are defined by the redundant line control function of Global Link Services (GLS).

Note

For details on using the multipath function of Global Link Services (GLS), see "Chapter 7 Administration on a Cluster System" in "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function*."

Figure 6.27 Flow of Gls (redundant line control function) resource creation

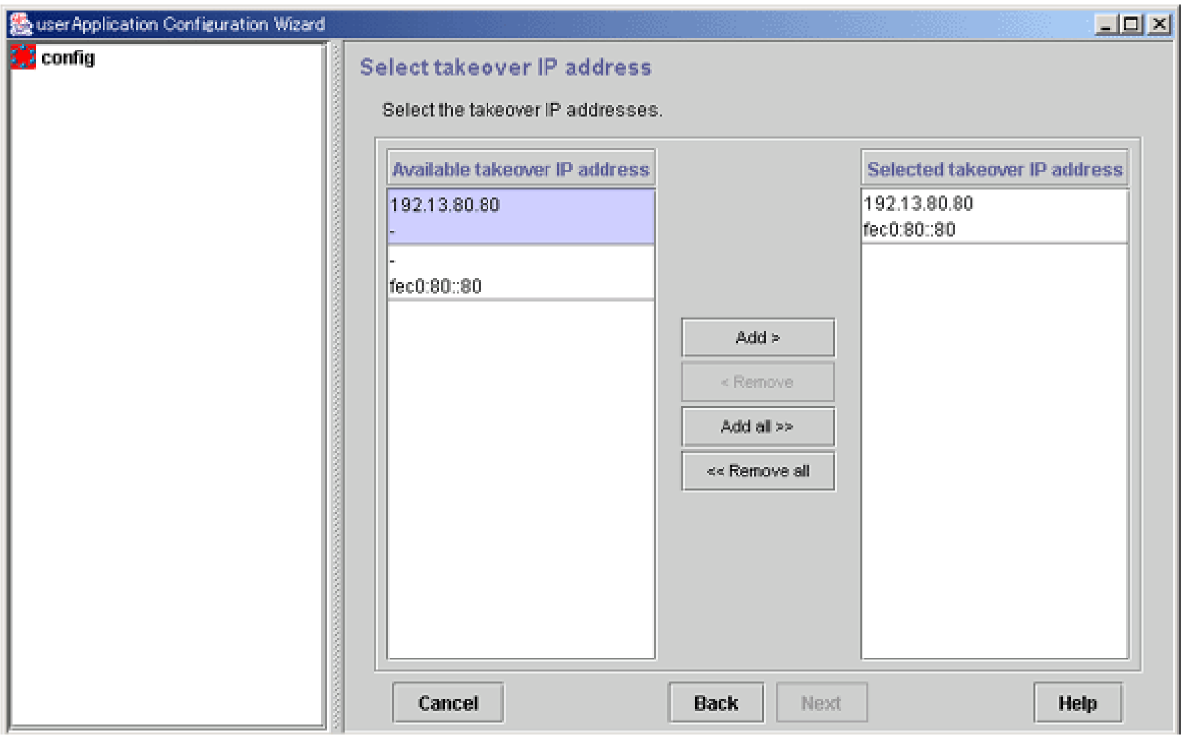


For information on the above operations up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resourcesmdline Resources." This section describes the operations from "Select takeover IP."

Selecting the takeover IP address

Select the takeover IP address.

Figure 6.28 Takeover IP address selection



Available takeover IP address

IP addresses that can be taken over.

Selected takeover IP address

Takeover IP addresses.

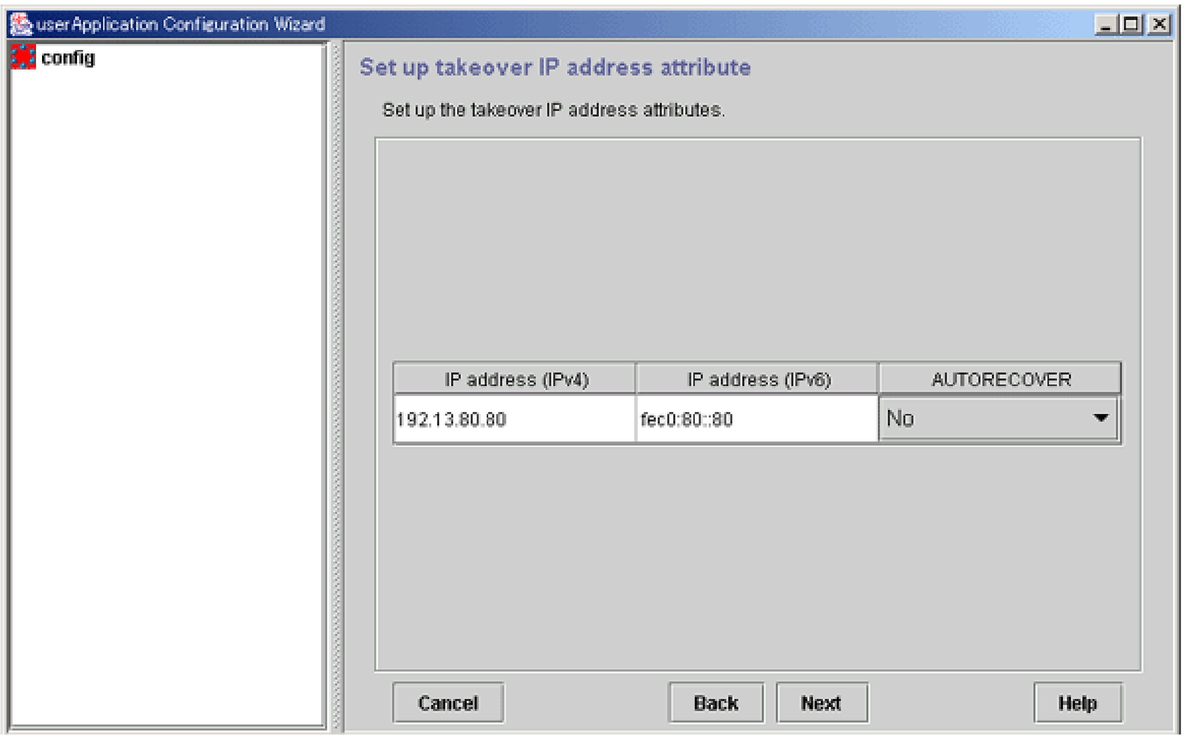
From *Available takeover IP address*, select takeover IP addresses, and then click *Add*. To add all listed takeover IP addresses, click *Add all*. To delete a takeover IP address, select the takeover IP address to be deleted from *Selected takeover IP address*, and then click *Remove*. To delete all listed takeover IP address, click *Remove all*.

After completing the setup, click *Next* to go to the "Set Takeover IP Address Attribute" screen.

Setting attributes of the takeover IP address

Set the attributes of the takeover IP address.

Figure 6.29 Attribute setup for the takeover IP address



After completing the setup, click *Next* to go to "Checking Gls Resource Registration Information."

Flag	Outline
AUTORECOVER	If the flag is set to "Yes," RMS attempts to restore the faulted resource for a given amount of time to prevent userApplication being switched to the other host. The default timeout is 60 seconds. You can change it by tuning the resource attribute Timeout. If the faulted resource does not recover within a specified time, userApplication is moved to the other host. If the flag is set to "No", RMS switches userApplication to the other host on detecting the faulted resource.

Checking registration information for Gls resources

Check the registration information for the Gls resources. You can also set resource attributes by selecting the *Attributes* tab and switching the screen.

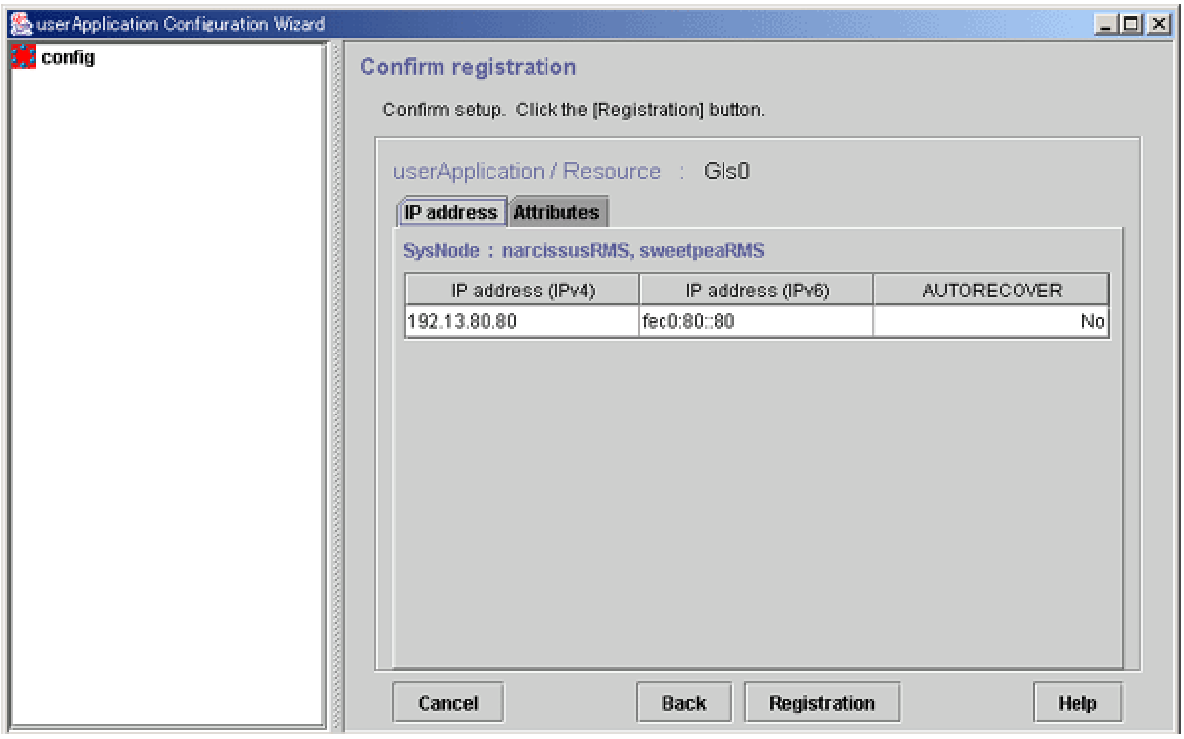
For information on the resource attributes, see "6.7.5 Attributes."

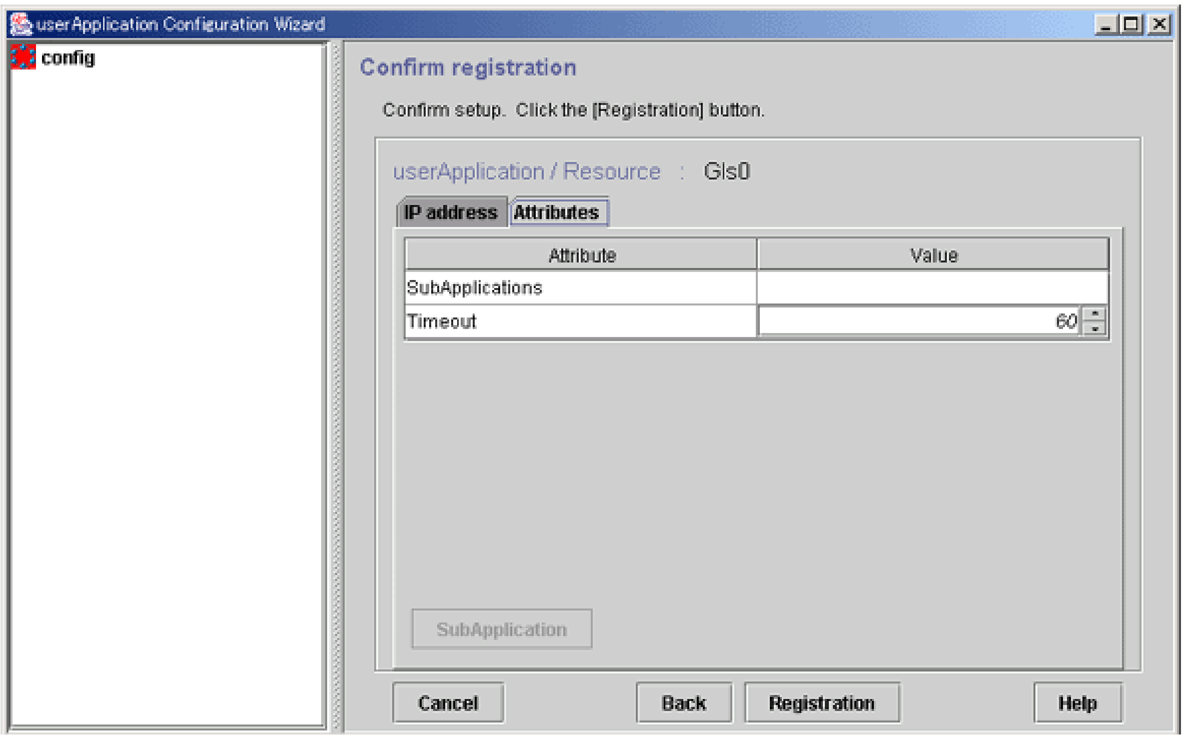
Gls resource names are displayed as GlsX (X is a number, such as 0 or 1).

To register a Gls resource to a cluster application, select this resource name from *Available Resources* on the Select Resource screen.

For details, see "6.7.2 Creating Cluster Applications."

Figure 6.30 Confirmation of registration information





***SubApplication* button**

Use this button to associate a previously created Gls resource under the current Gls resource. This button can be selected only if there are resources of the same type that can be associated. For setting instructions, see "Resource association" in "6.7.1.1 Creating Cmdline Resources."

After checking the registration information, click *Registration*.

6.7.1.5 Creating Takeover Network Resources

Normally to use a takeover network, use either "Gls" (Global Link Services) or "Ipaddress" as the resource type. If the availability of the takeover network is needed, use "Gls."

A takeover network must be set up if you are building a system like a client/server system that communicates with cluster services that operate in the cluster system.

Takeover networks allow communication to continue with the same network name from outside the cluster even if a cluster application that operates in the cluster system undergoes failover.

Takeover network types

The takeover network types are IP address takeover and node name takeover.

- IP address takeover

  When switchover takes place, the defined IP address is taken over to the OPERATING node.

  This is the basic function of a takeover network.

- Node name takeover

  The node name* is also taken over to the OPERATING node.

  Use this type if a program that operates as a cluster application in the cluster node recognizes node names.

  * This is the same value as the host name that is obtained when uname -n is executed.

  In a Solaris 11 environment, change the node name by using the svccfg(1M) command and so on.

Files that are edited automatically by the GUI

If the GUI is used to execute the setup, certain files are edited as shown below.

The files differ according to the takeover network type.

```
# Start of lines added by FJSVwvucw - DO NOT DELETE OR CHANGE THIS LINE
# Mon Aug 05 21:01:43 JST 2002
```

<Specify information that is dependent on that particular file. For example, for /etc/inet/hosts, the information becomes as follows:>

```
192.168.246.100 Ipaddress01
# End of lines added by FJSVwvucw - DO NOT DELETE OR CHANGE THIS LINE
```

- IP address takeover

  /etc/inet/hosts

  /usr/opt/reliant/etc/hvipalias

- Node name takeover

  /etc/inet/hosts

  /usr/opt/reliant/etc/hvipalias

  /etc/nodename

Note

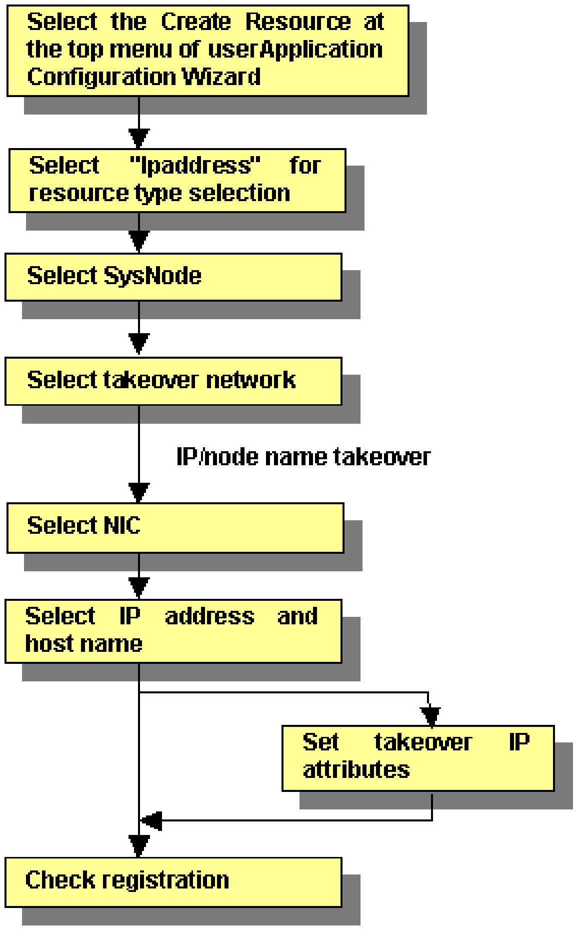
- In node name takeover, the same node name is set to both the OPERATING node and the STANDBY node.
- All cluster nodes must have at least one network interface card so the card can be used.
- Although a PRIMECLUSTER "takeover network" and the "IP address takeover" function of GLS can be set up in the same cluster system, do not use them on the same interface. If the two are used on the same interface, connection using takeover IP addresses will be disenabled.
  For example, if you select hme1 as the interface to be used when you set up the PRIMECLUSTER "takeover network," do not set hme1 in the GLS environment settings (do not specify hme1 in the -t option of the "hanetconfig create" command).

- You cannot assign takeover networks to network interfaces used in the shared or exclusive IP zones, which are delegated to non-global zones.

6.7.1.5.1 Setup Method

This section describes how to create takeover network resources.

Figure 6.31 Flow of takeover network resource creation

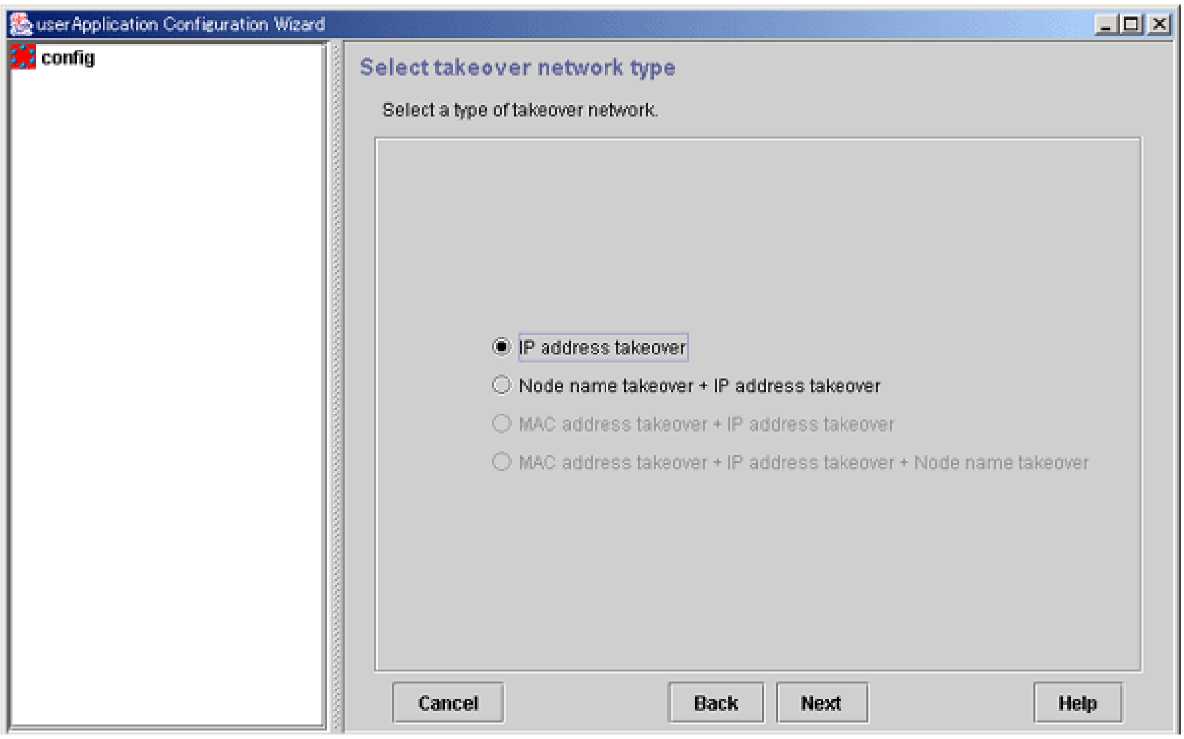


For information on the above operations up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resourcesmdline Resources." This section describes the operations starting from "Select takeover network."

Selecting a takeover network

Select the takeover network type.

Figure 6.32 Network type selection



IP address takeover

Select this item to enable IP address takeover.

Node name takeover + IP address takeover

Select this item to enable node name takeover and IP address takeover.
If node name takeover has already been set, you are not allowed to set twice.

"MAC address takeover + IP address takeover" and "MAC address takeover + node name takeover + IP address takeover" cannot be selected.

You can configure a network interface for each SysNode. The takeover network settings are enabled for one network interface:

- Multiple IP address takeover settings are enabled for each network interface.
- Only one node name takeover setting is enabled in a cluster system.

After completing the setup, click *Next* to go to the "Select Interface" screen.

Note

- To activate a network interface card (NIC) when the system is started, you must do the following in advance:
  - For Solaris 10

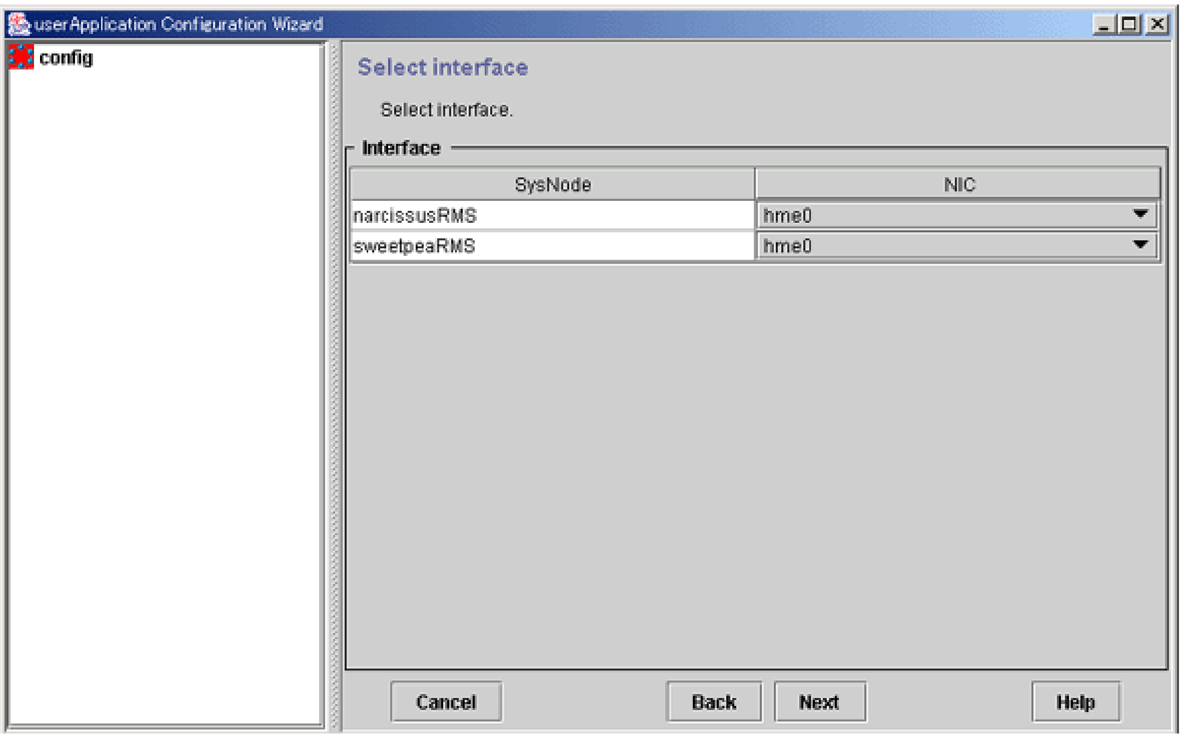
    Create the "/etc/hostname.*network_interface_name*" file, and then define the IP address (or the host name) in the file.
  - For Solaris 11

    Use the ipadm(1M) command for setting. Specify a different value from other IP addresses.
- Node name takeover changes the uname of the system. To set node name takeover for userApplication, you need to restart all nodes configuring userApplication.
  If a node name takeover resource has been deleted, check whether /etc/nodename has been properly changed in all nodes, and then restart all nodes. If /etc/nodename has not been properly changed, change the node name by executing the "setuname(1M)" command, and then restart the nodes.
  Also check whether /etc/nodename is set properly when system conversion or software upgrade is implemented. If /etc/nodename is not set properly, change the node name by executing the "setuname(1M)" command, and then restart the nodes.

Selecting an interface

Select a network interface card (NIC).
The NIC cards registered to Cluster Resource Manager are displayed.

Figure 6.33 Select interface



Interface

Select the network interface to be used in each SysNode.

After completing the setup, click *Next* to go to the "Select IP address and host name" screen.

Selecting or creating an IP address or host name

Select or create the takeover IP address or host name.

Figure 6.34 Selection screen for IPv4 address and host name

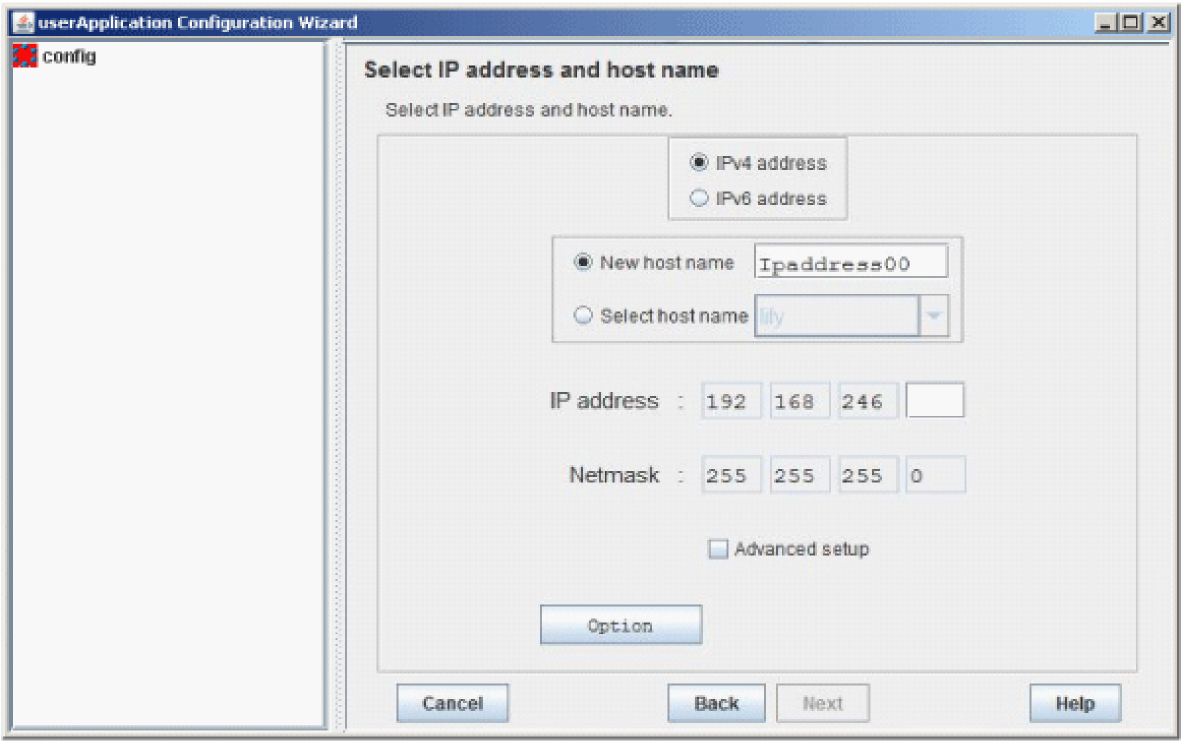
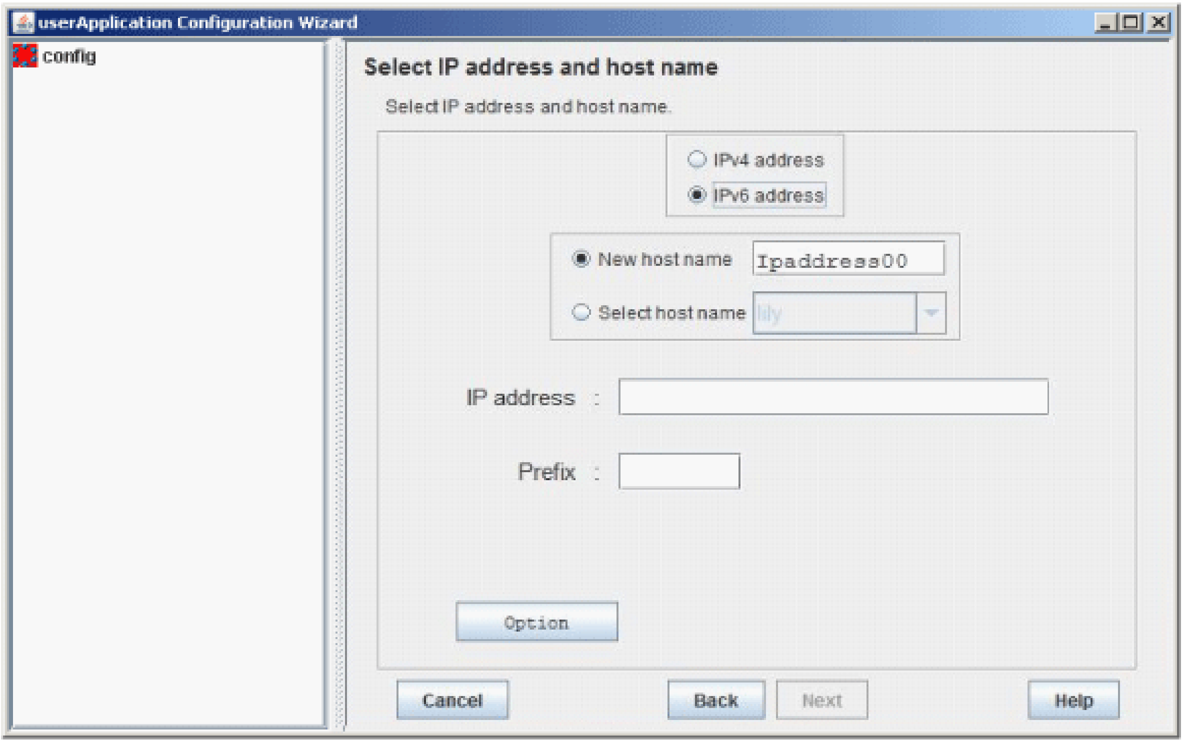


Figure 6.35 Selection screen for IPv6 address and host name



IPv4 address

Select "IPv4 address" when creating a takeover IP address of an IPv4 address.

IPv6 address

Select "IPv6 address" when creating a takeover IP address of an IPv6 address.

New host name

You can set a new takeover IP address and a takeover node name. The setup information is added to the "/etc/inet/hosts" or "/usr/opt/reliant/etc/hvipalias" file on all nodes configuring the cluster system.
Specify a character string of up to 14 characters that begins with an alphabet letter and consists of only alphanumeric characters.

Select host name

You can select the IP address or node name from information that has been set. If IP addresses or node names were previously configured to /etc/inet/hosts and /usr/opt/reliant/etc/hvipalias, select the IP address or node name from those settings.

IP address

Enter the takeover IP address.
When you select "IPv4 address," enter 0 to 255 numbers in the address input area.
When you select "IPv6 address," enter an IPv6 address.

Netmask (when you select "IPv4 address")

Enter the net mask value.
Enter 0 to 255 numbers in the address input area.

Prefix length (when you select "IPv6 address")

Set the prefix length.
Enter 0 to 128 numbers in the prefix length input area.

Advanced setup (when you select "IPv4 address")

Set up a masked net mask or change an IP address.

Option button

Press this button to configure attributes for the takeover IP address. For more information, see "Setting up Takeover IP Address Attributes" below.

Note

If /usr/opt/reliant/etc/hvipalias and /etc/inet/hosts files have been modified to accommodate the entries for the takeover IP address and node name takeover in advance (as described in the Prerequisites section), message 0840 appears when the Next button is clicked. This message confirms whether the existing values are to be used without modification. To use the values, select *Yes*.

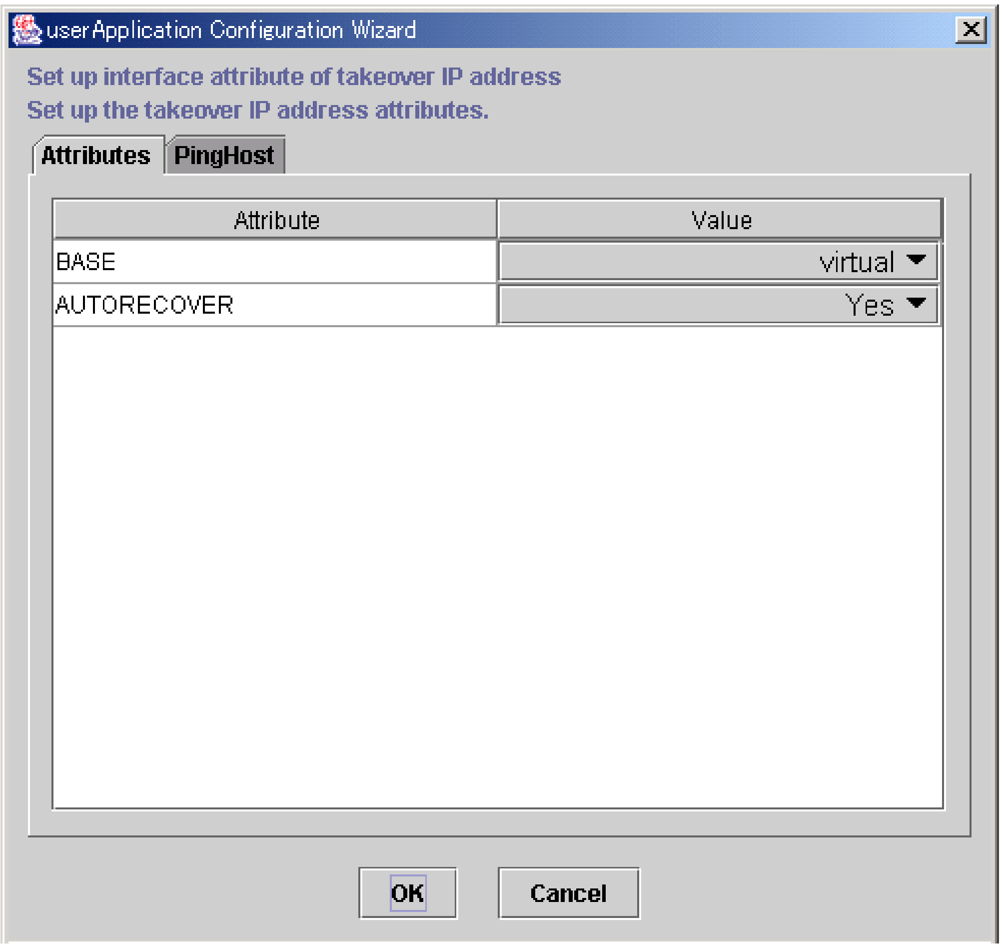
If you select *No*, the existing settings in the file are deleted and the GUI re-creates the information. In this case, takeover network information can be deleted automatically from /usr/opt/reliant/etc/hvipalias and /etc/inet/hosts when the Ipaddress resource is deleted.

IPv6 link local addresses are not available for IP addresses.

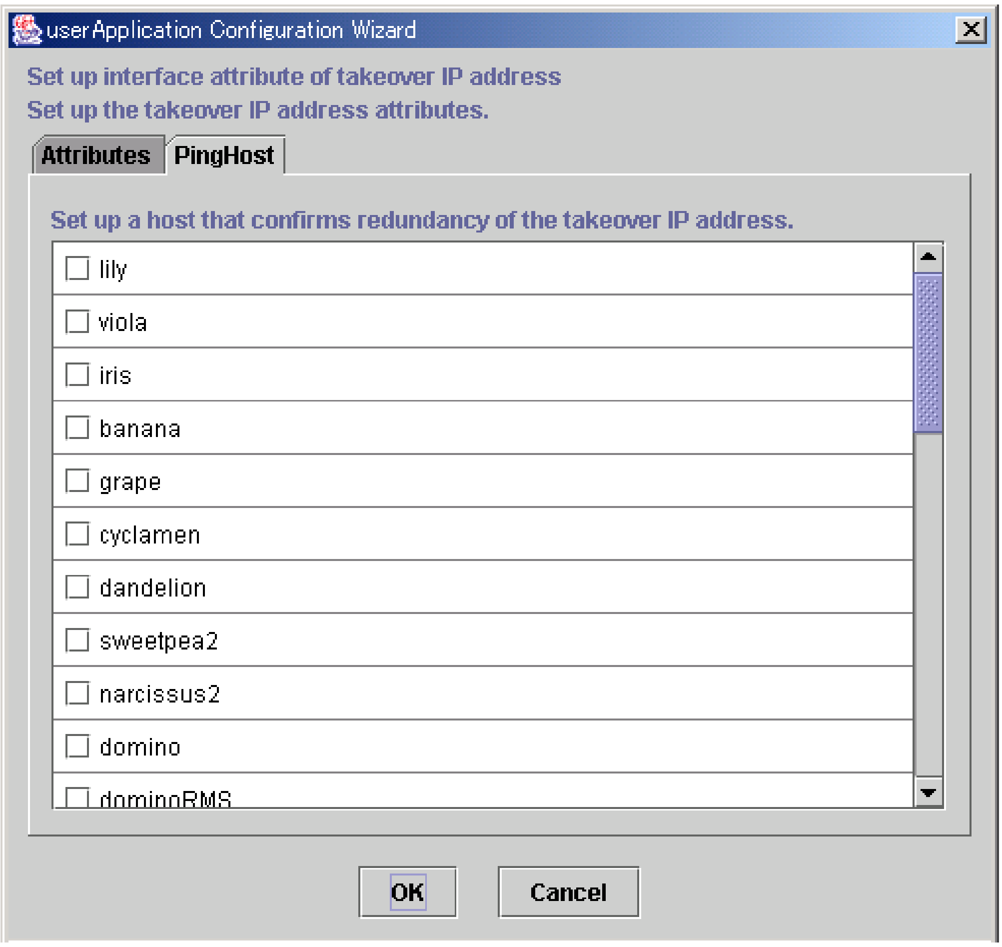
Setting up Takeover IP Address Attributes

Configure the interface attributes for the takeover IP address.

Figure 6.36 Attribute setup for takeover IP address



Flag	Outline
BASE	If this flag is set to "base," the specified address is assigned to the physical interface. If this flag is set to "virtual," the specified address is assigned to the virtual interface. The default value is "virtual." "base" can be selected only when using an IPv4 address for a takeover IP address.
AUTORECOVER	If this flag is set to "Yes," RMS automatically attempts to create an interface if the specified address becomes disabled. A failure of this attempt might trigger Fault processing. The default value is "Yes."



After the setup is completed, click *OK* to return to the "*Select IP address and host name*" screen.

The availability of the takeover IP address is verified by executing the "ping" command.

We recommend that you specify two or more hosts, which are not used for the cluster system, and are in the same network segment that does not use a hub or router. This prevents adverse effects from hub and router failures.

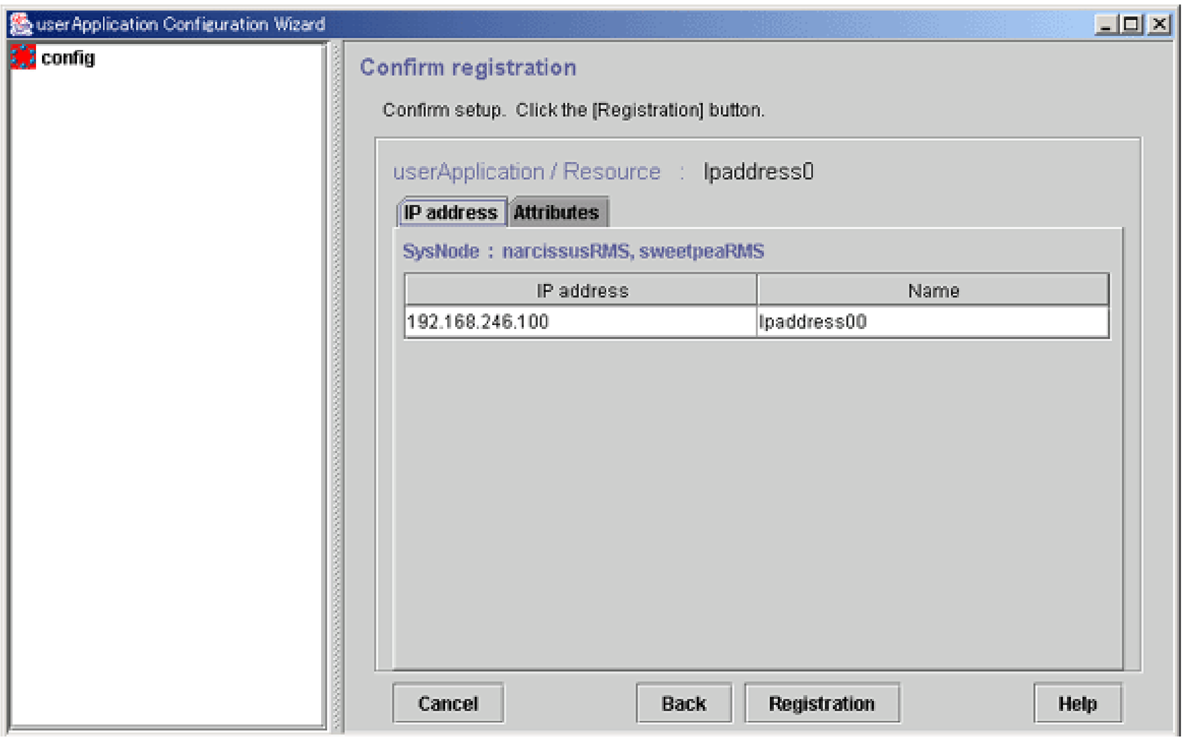
The host information to be used by PingHost must be in /etc/inet/hosts.

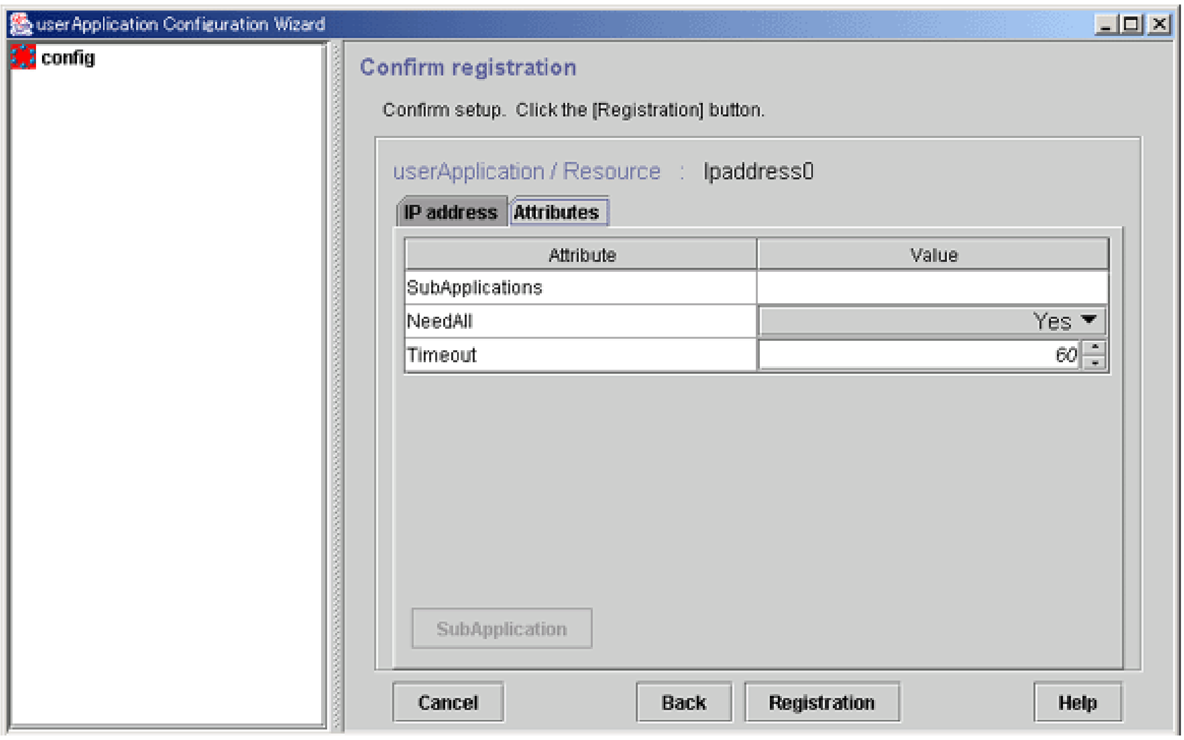
Checking registration information for the takeover network

Check the registration information for the takeover network. You can also set resource attributes by selecting the *Attributes* tab and switching the screen.

For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.37 Confirm registration





SubApplication button

Select this button to associate a previously created takeover network under the current takeover network. This button can be selected only if there are resources of the same type that can be associated. For setting instructions, see "Resource association" in "6.7.1.1 Creating Cmdline Resources."

Check the registration information, and then click *Registration*.

6.7.1.6 Creating Procedure Resources

A procedure resource must be created to migrate a SynfinityCluster product to PRIMECLUSTER. Procedure resources can be created only for those products described in "Part 7 PRIMECLUSTER Products."

6.7.1.6.1 Prerequisites

To create a procedure resource, you must first create a state transition procedure and register the procedure to the resource database on all nodes in the cluster system.

Note

If a process is started without process monitoring from a state transition procedure, the hard- and software limits of the file descriptor (fd) for the process may be changed.

If the hard- and software limits of the file descriptor (fd) for the system are less than 1024, the hard- and software limits of the file descriptor (fd) for the process started from a state transition procedure will be 1024.

In the process started from a state transition procedure, if you need to set the value less than 1024 for the hard- and software limits of the file descriptor (fd), add a "ulimit" definition at the beginning of processing for the state transition procedure, and specify the hard- and software limits of the file descriptor (fd).

Example

For setting both hard- and software limits of the file descriptor (fd) to 256:

```
#!/bin/sh
ulimit -n 256
...
```

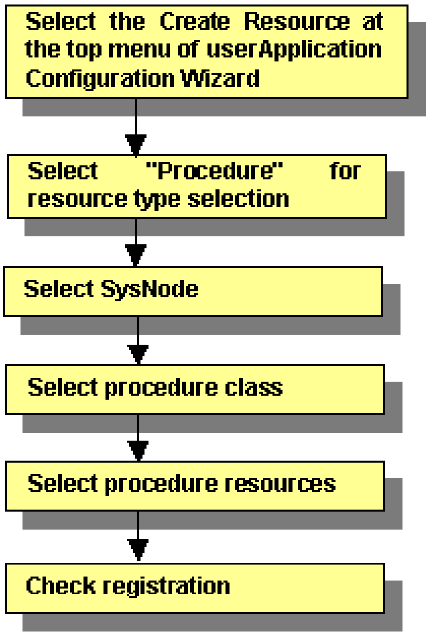
See

For details on how to register a state transition procedure, see "Appendix E Registering, Changing, and Deleting State Transition Procedure Resources for SynfinityCluster Compatibility."

6.7.1.6.2 Setup Instructions

Register the procedure resources that were registered to the cluster resource manager to RMS.

Figure 6.38 Flow of procedure resource creation

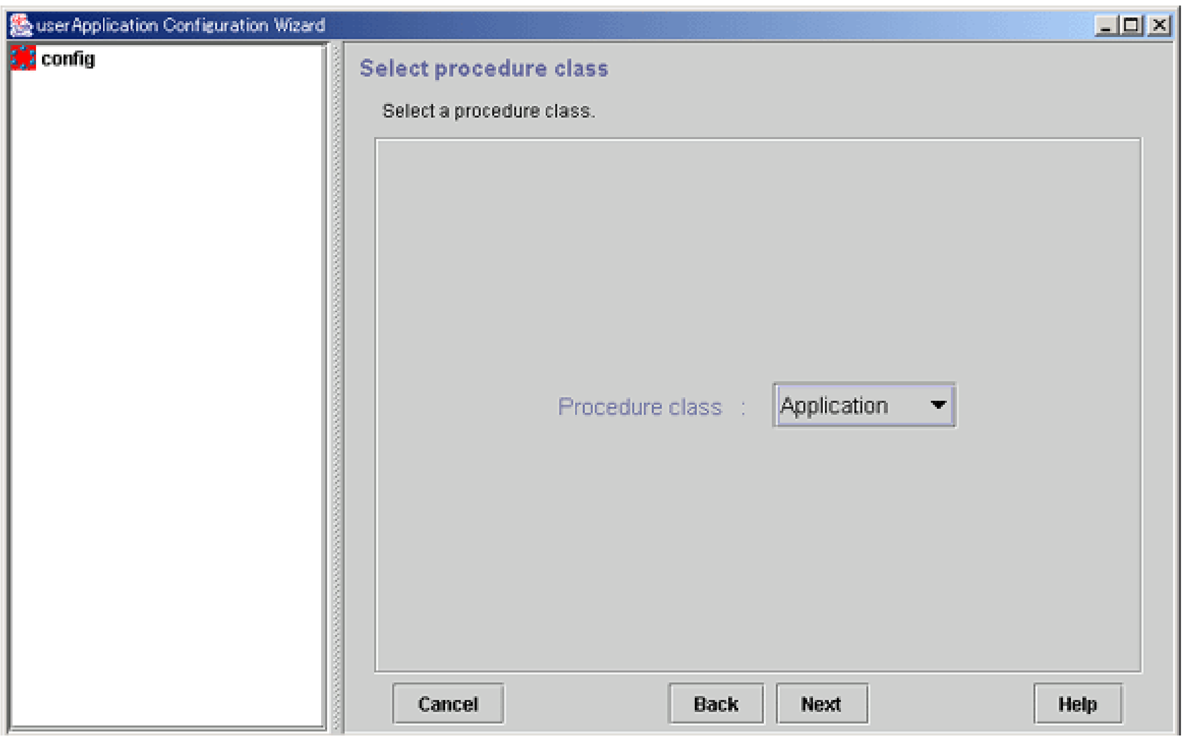


For information on the above operations up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resourcesmdline Resources." This section describes the operations from *Select procedure class*.

Selecting the procedure class

Select the procedure class to be created.

Figure 6.39 Select procedure class



Procedure class

Select the procedure class. Only the resource classes of resources that have been registered to the cluster resource manager are displayed. For example, if only procedure resources of the Application class have been registered, only *Application* is displayed.
The class resource manager provides four standard procedure classes:

- Application

  This class is used to make general applications to be cluster-aware.

- BasicApplication

  It is a class used for DBMS.

- SystemState2

  It is used to make a part of OS functionality cluster-aware. This OS function should be started at "/etc/rc2.d" in a non-cluster single node. SystemState2 does not automatically start during OS startup. It is used to start a cluster application only on the OPERATING node.

- SystemState3

  It is used to make a part of OS functionality cluster-aware. This OS function should be started at "/etc/rc3.d" in a non-cluster single node. SystemState3 does not automatically start during OS startup. It is used to start a cluster application only on the OPERATING node.

Note

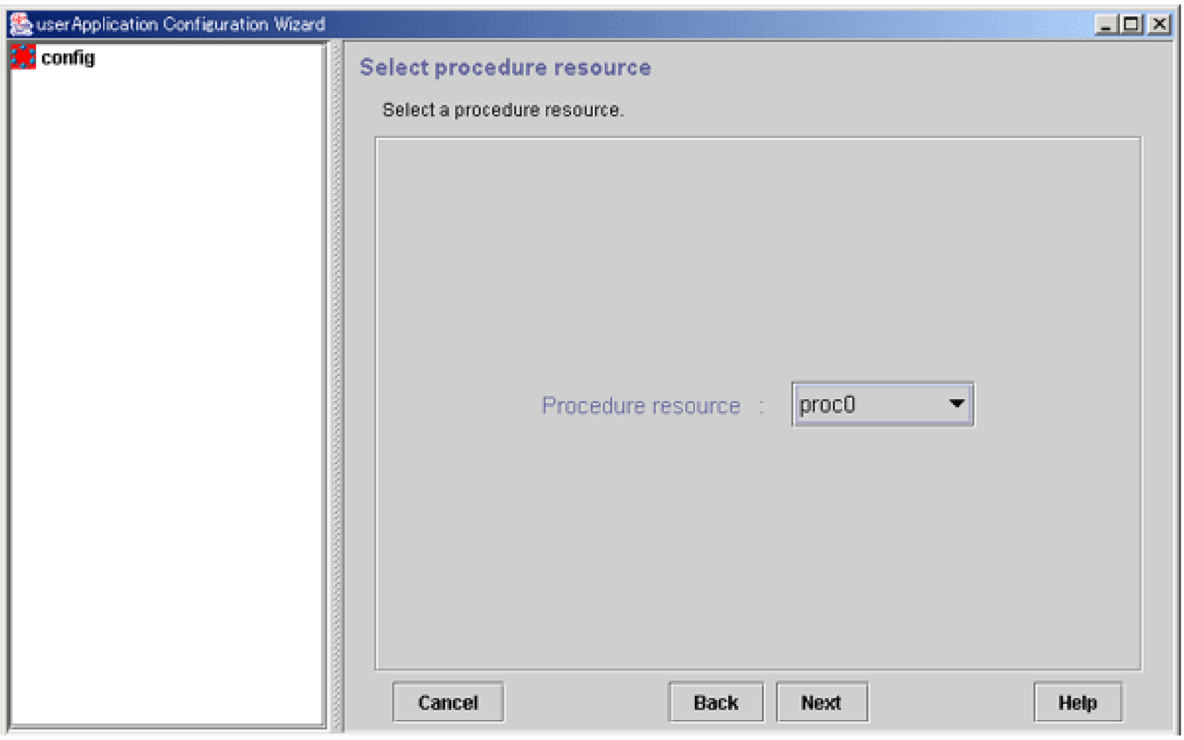
To create an application resource that is set in one cluster application on each node, the application resource name should be the same as the application type on each node.

After completing the setup, click *Next* to go to the "Select procedure resource" screen.

Selecting the procedure resource

Select the procedure resource to be created.

Figure 6.40 Select procedure resource



Procedure resource

Select the procedure resource to be created from the displayed list. The procedure resource name consists of up to 32 characters.

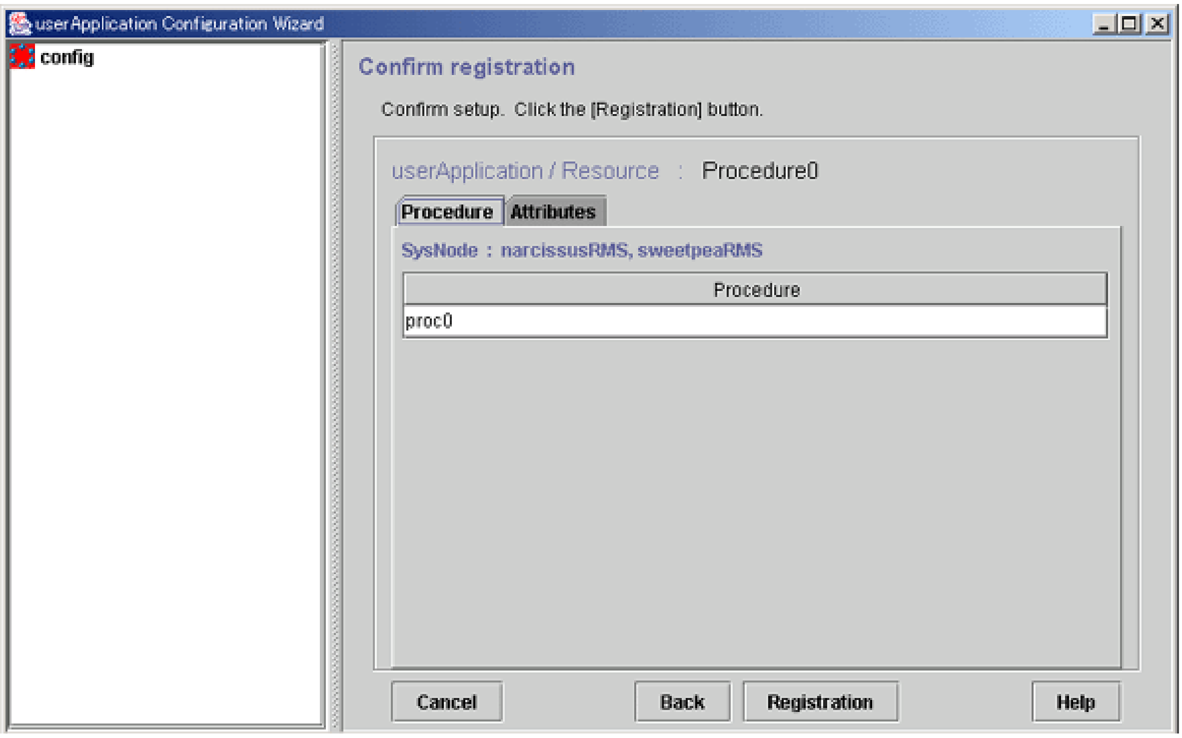
After completing the setup, click *Next* to go to the "Confirm registration" screen.

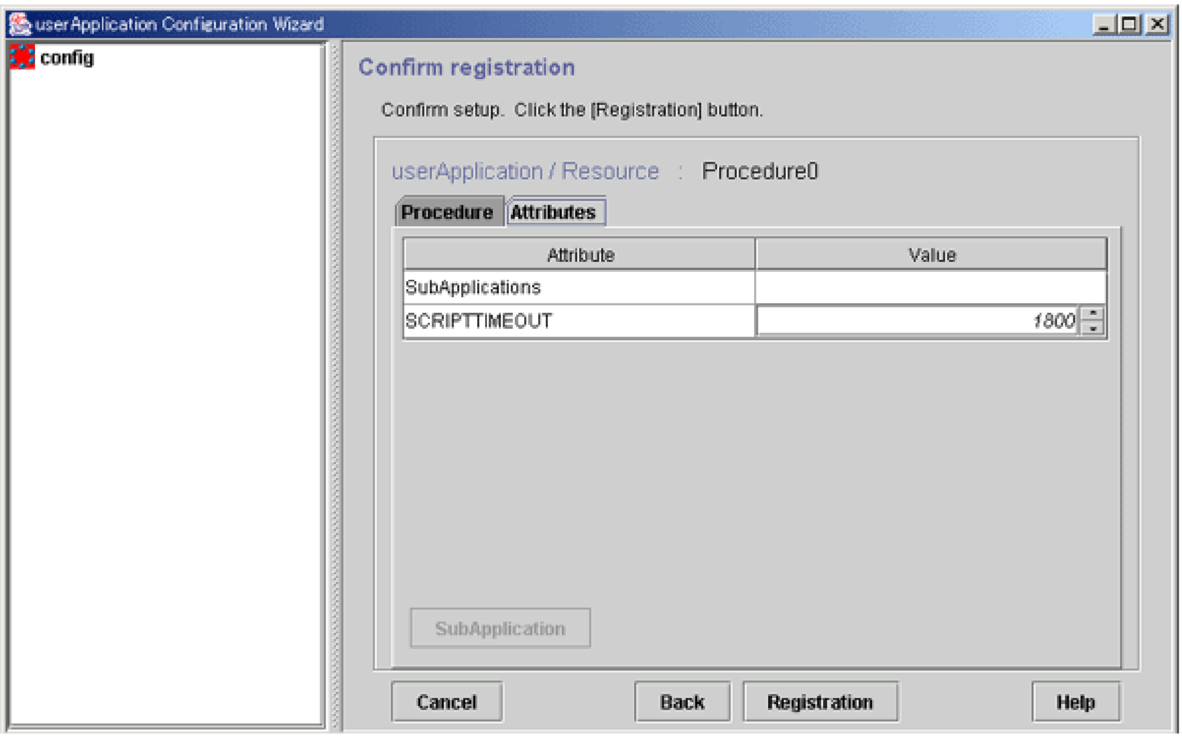
Checking the registration information for the procedure resource

Check the registration information for the procedure resource. You can also set resource attributes by selecting the *Attributes* tab and switching the screen.

For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.41 Confirm registration





***SubApplication* button**

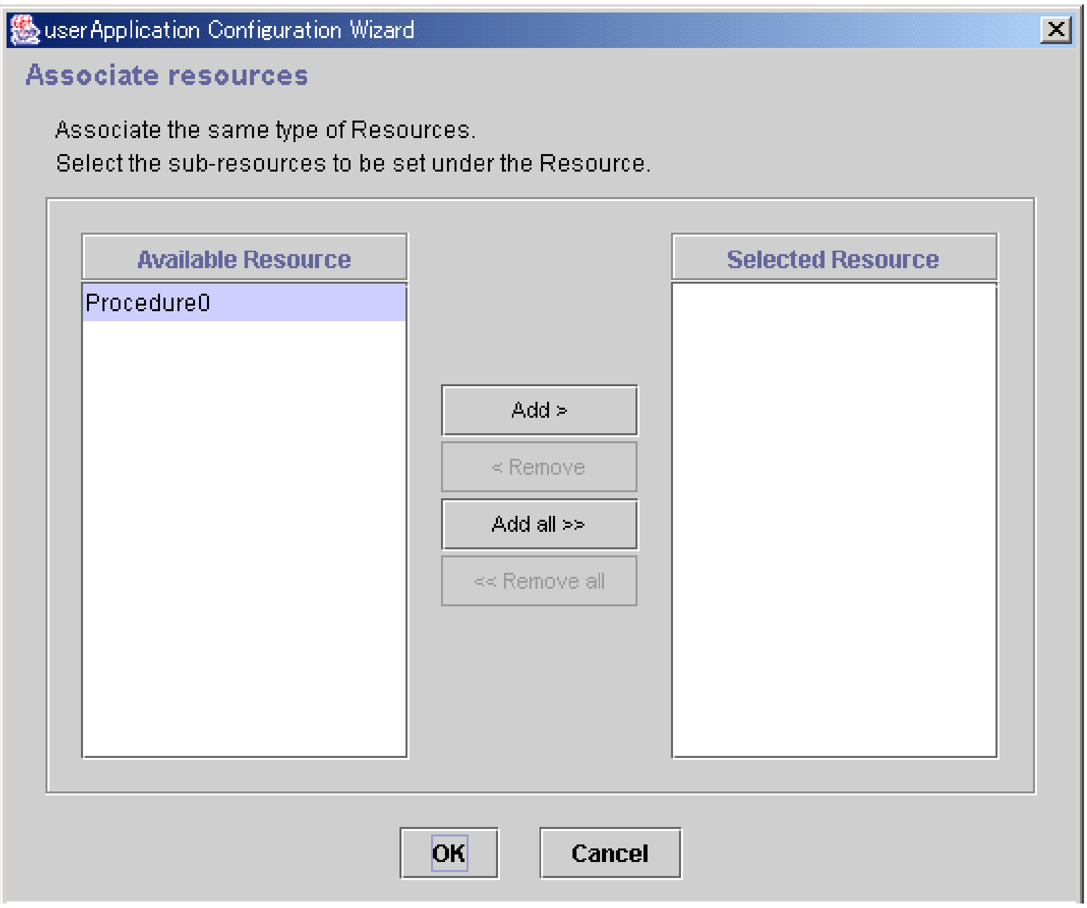
This button associates the other procedure resources to the procedure resource that has been created above. For more information, see "Associate resources," which is described later.

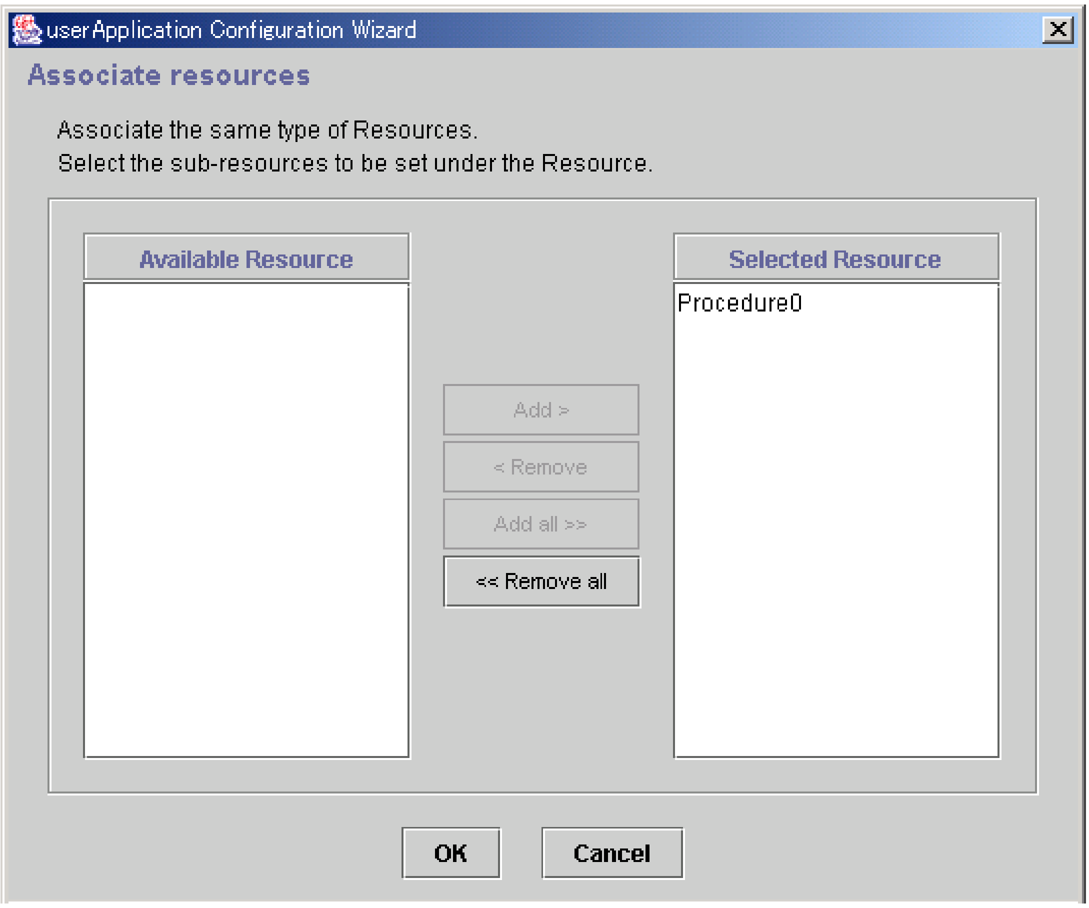
Check the registration information, and then click *Registration*.

Resource Association

Use resource association to assign a startup order to resources of the same type.

Figure 6.42 Associate resources





Available Resource

Available resource is referred to as the resource of the same type that can be associated, and satisfies the following conditions.

- The resource is of the same type as the resource that was called.
- The resource is not being used by another cluster application.

Selected Resource

Resources to be set under the current resource being created.

From *Available Resource*, select the resources to be configured under the current resource, and then click *Add*. To add all listed resources, click *Add all*. To remove a resource, select the resource to be removed from *Selected Resource*, and then click *Remove*. To delete all listed resources, click *Remove all*.

After completing the setup, select *OK* to return to the "*Confirm registration*" screen.

6.7.1.7 Creating Process Monitoring Resources

This section describes how to set up process monitoring resources.

Before describing the setup method, this section also outlines the process monitoring function and prerequisites for specific uses.

6.7.1.7.1 What Is the Process Monitoring Function?

Overview of the process monitoring function

The process monitoring function monitors the live state of processes. The main features are as follows:

- Changes in the live status of a process can be monitored.
  (This setup is quite easy, so the user does not need to prepare commands for monitoring the live status of a process.)
- Notifies RMS of the live state of any process immediately, and this provides high-speed switchover.

- If any process terminates abnormally because of an unexpected error, that process is automatically restarted.

A relationship diagram of the process monitoring function and RMS is shown below. The process monitoring function consists of three components: the "clmonproc" command, the Process Monitoring Daemon (prmd), and the Detector (hvdet_prmd).

- "clmonproc" command

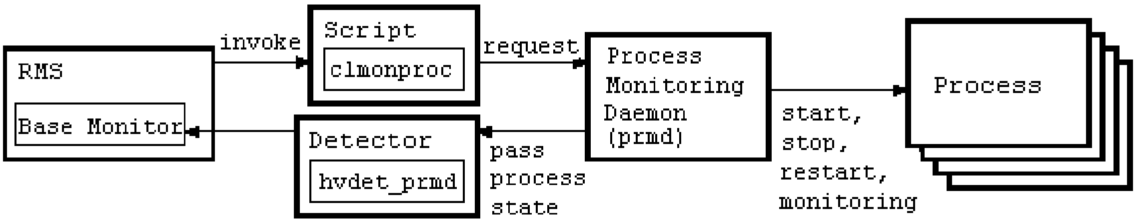
  The "clmonproc" command is executed from the Online or Offline script. The command requests prmd to start a specified process and to stop live monitoring.

- prmd daemon

  prmd is a daemon process that starts a process and stops live monitoring according to requests received from the "clmonproc" command. If the live state of a process being monitored changes, prmd notifies hvdet prmd immediately.

- hvdet_prmd daemon

  After receiving change information on the live state of a process from prmd, the "hvdet prmd" process notifies the RMS Base Monitor (BM) of the changes.



Benefits of using the process monitoring function

Described below are the benefits of using the process monitoring function.

Easy setup

Since prmd monitors whether there are any processes to be monitored, the user does not need to create a check command for each process to be monitored. The check command is used to determine whether the process to be monitored exists. Therefore with little work, the user can easily monitor the existence of processes.

High-speed detection of abnormal process termination

If the process monitoring function is not used, abnormal termination of a monitored process is detected by using a Cmdline resource to execute the "aforementioned check" command periodically. This delays detection of abnormal termination of a monitored process by execution time interval of the check command. However, if the process monitoring function is used, prmd uses signal processing to detect abnormal termination in monitored processes. This process monitoring function allows abnormal process termination to be detected at high speed compared to when check commands are executed periodically.

Automatic restart of any process that terminates abnormally

If any process terminates abnormally because of an unexpected error, the process monitoring function restarts that process automatically.

Reduction of CPU resource consumption

To shorten the time required to detect abnormal termination of a monitored process without using the process monitoring function, you must shorten the execution time interval of the check command. However, since this leads to frequent generation and execution of the check command, many CPU resources may be used up. Generally a command like the "ps" command is used as the check command. However, when a command that uses relatively more CPU resources, like the "ps" command, is used, the CPU resource consumption may become even more pronounced.

When the process monitoring function is used, prmd uses a signal process to monitor abnormal termination of the monitored process. A process that uses many CPU resources, such as one that issues a check command periodically, is not executed.

With the method that uses Cmdline resources, the number of check commands increases in proportion to the number of RMS objects because a check command is executed for each RMS object that is defined by the process to be monitored. Therefore if many check commands are executed periodically, many CPU resources may be used.

When the process monitoring function is used, it is always just one prmd that monitors the live stage of the process. Therefore, prmd does not use many CPU resources in proportion to the increase in the number of processes to be monitored.

6.7.1.7.2 Prerequisites

Normally, you do not need to work on this prerequisite for using the process monitoring function. This is required only when you want to use the processing monitoring function for the specific purpose described below.

The identification number of the detector (the state notification module for notifying RMS) used by the process monitoring function is the same as the identification number to be used by other functions.

Note

- The GUI uses "0" (default value) as the detector identification number of the process monitoring function. If a detector identification number other than "0" is already used, use that value.
- This setup is required only for changing the detector identification number from the default value "0."
- In this setup, the setting of the detector identification number must not exceed the allowed value "127."
- Configure the same detector identification number on all cluster nodes.
- Stop RMS and all running cluster applications on all nodes before changing the identification number.

Example

In the example described below, the detector identification number of the process monitoring function is changed from the initial value "0" to "2."

If the process monitoring function is being used with the initial value "0," the setting <*detector_name_of_process_monitoring_function*>.g<*identification_number*> will exist. Remove that setting.

```
# cd /usr/opt/reliant/bin
# ls -l  hvdet_prmd.g0
lrwxrwxrwx   1 root     other        31 Dec 20 12:21 hvdet_prmd.g0 -> /usr/opt/reliant/bin/hvdet_prmd
# rm hvdet_prmd.g0
```

Execute the "clmonsetdet" command as shown below. For the command argument, specify "2," which is the new detector identification number.

```
# /etc/opt/FJSVcluster/bin/clmonsetdet 2
```

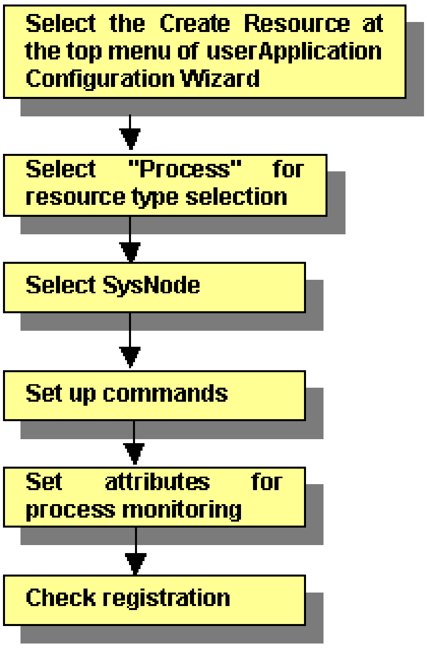
Check whether the settings were set correctly.

```
# ls -l  hvdet_prmd.g2
lrwxrwxrwx   1 root     other        31 Dec 27 12:21 hvdet_prmd.g2 -> /usr/opt/reliant/bin/hvdet_prmd
#
```

6.7.1.7.3 Setup Instructions

This section explains how to create process monitoring resources.

Figure 6.43 Flow of process monitoring resource creation

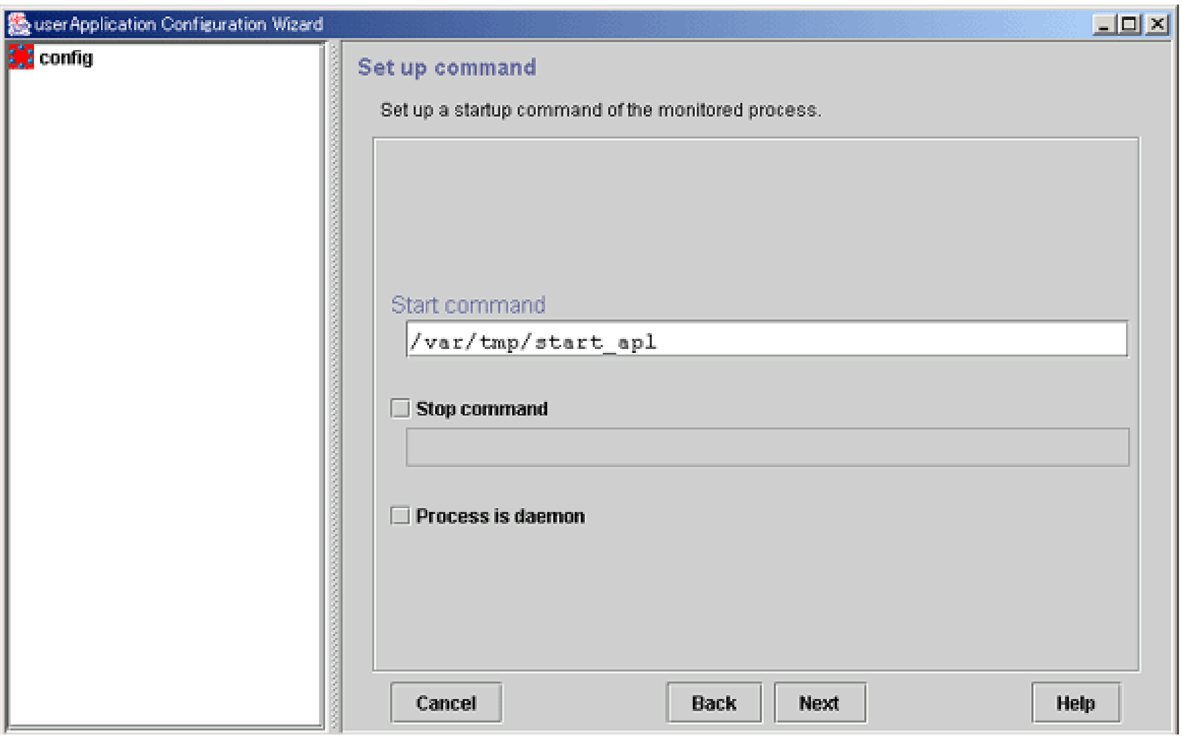


For information on the above operation up to "Select SysNode," see "6.7.1.1 Creating Cmdline Resourcesmdline Resources." This section describes the operations from "Set startup path."

Inputting the process startup command

Enter the startup path of the process to be monitored.

Figure 6.44 Set up command



Start command

Enter the program name for starting the process to be monitored during online processing using a full path. If spaces are included in the full pathname, the pathname must be enclosed in double quotation marks (").

For example, enter the command line as follows:

```
"/var/tmp 1/start_apl"
```

The process monitoring function cannot monitor the following processes. It is necessary to specify a start command for each process monitoring resource.

- Programs that exit out of own after initiating the other programs in the background.

  Example: Shell script that exits out of own after starting three programs of "prog0", "prog1", and "prog2".

```
#!/bin/sh

prog0 &
prog1 &
prog2 &

exit 0
```

Be aware that the process monitoring function cannot monitor child process that is generated with the start command.

Stop command

Enter the method for stopping the monitored process during Offline processing.
If there is a command for stopping the monitored process, set the checkbox to ON, and enter the command line using a full path. If spaces are included in the command line, the command line must be enclosed in double quotation marks (").

For example, enter the command line as follows:

```
"/var/tmp 1/stop_apl"
```

If the stop command is omitted, the software exit signal (hereafter referred as SIGTERM) is sent from the process monitoring facility to stop the monitored process. Some processes might not be stopped by SIGTERM. In that case, Offline processing will fail.

Process is daemon

Specify this item if the monitored process is to be operated as a daemon.
A daemon refers to a process that executes the following types of processes at startup:

- Execution in the background
- Promotion to process group leader

The former refers to child process generation when the fork system call is issued and termination of the parent process when the exit system call is issued. The latter refers to promotion to process group leader when the setpgrp system call is issued.

Note

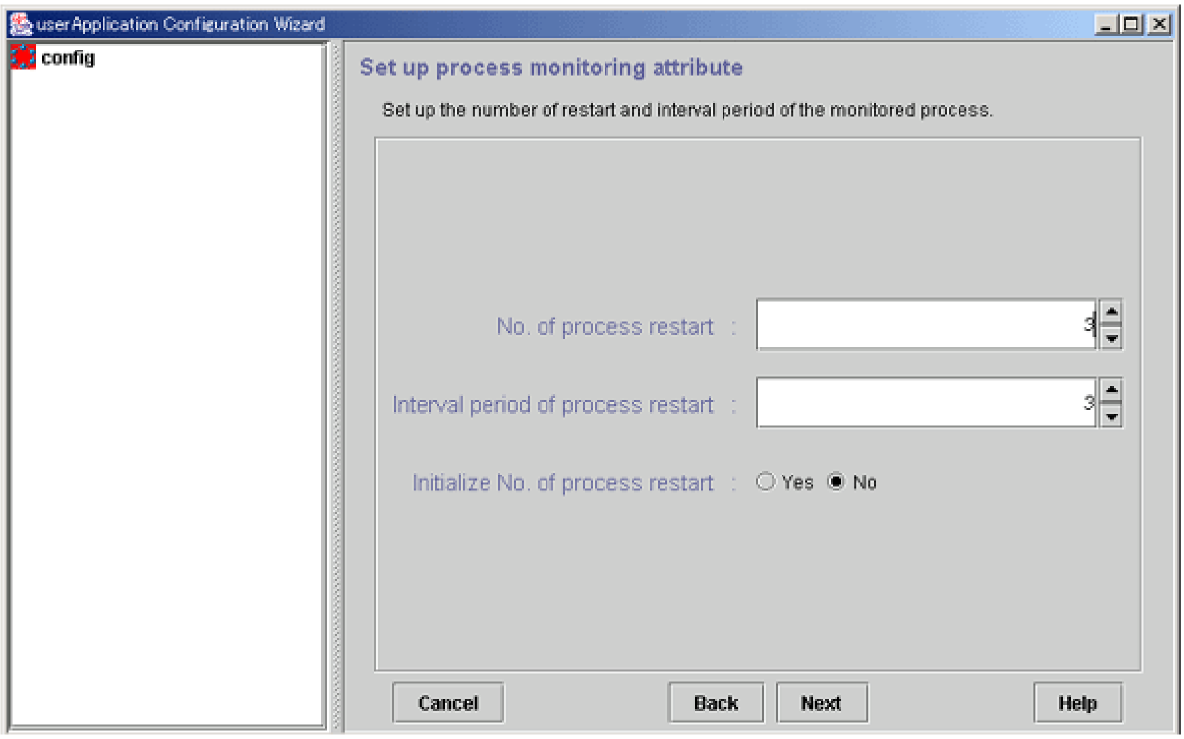
- The following characters cannot be used in the file name or the arguments: back slash ("\"), tilde ("~"), percent sign ("%"), ampersand ("&"), and at sign ("@").
- Single quotation marks ("'") and tabs cannot be entered.
- If a stop command is entered, a script for executing that command is created automatically. The script is stored in the following directory:

  `/opt/FJSVwvucw/scripts/process_stop`

Setting attributes for the process

Enter attributes for the process to be monitored

Figure 6.45 Set up process monitoring attribute



No. of process restart

Specify the number of times the monitored process is to be restarted between 0 and 99 (default is 3). If 0 is specified and the monitored process stops, the process will become Faulted.

Interval period of process restart

This is the interval from when the process monitoring facility determines that the process has stopped until the facility executes restart. The specification range is 0 to 3600 seconds (default: 3 seconds).

Initialize No. of process restart

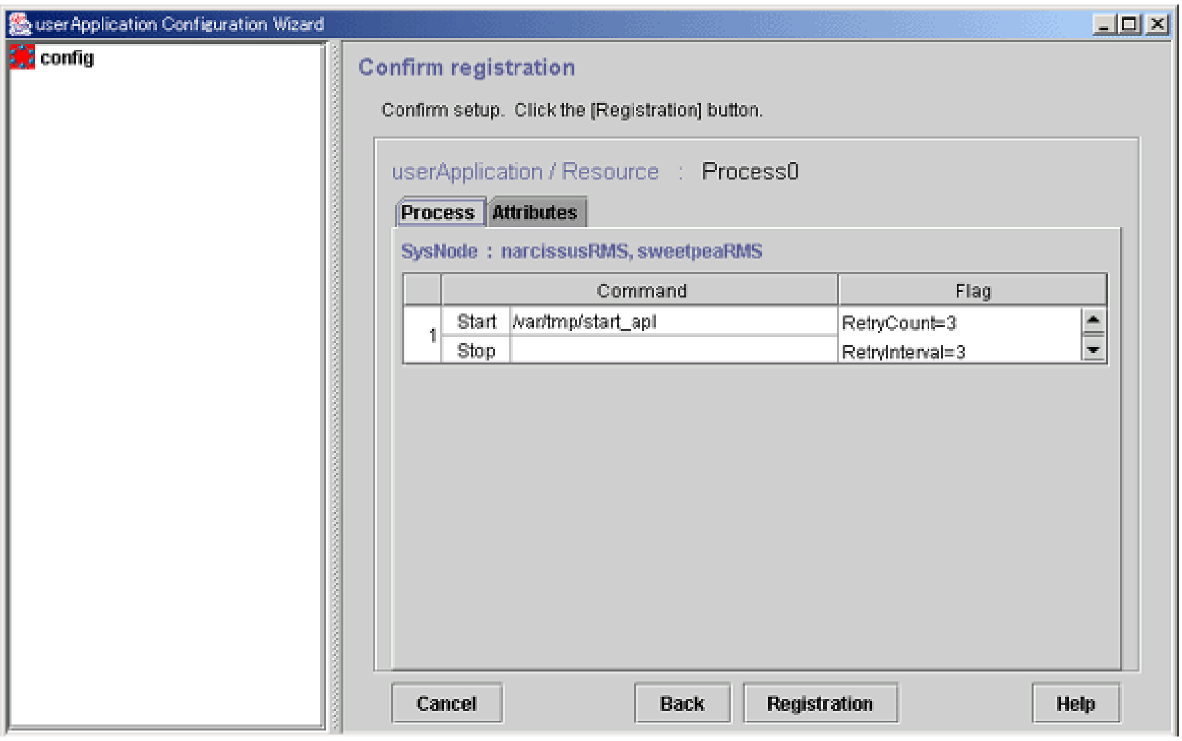
Specify whether or not the counter that has the specified *No. of process restart* value in the process monitoring facility as its maximum value is to be initialized periodically. If you select *Yes*, the counter is initialized every "*No. of process restart* value x 60 seconds." If you select *No*, the counter is not initialized periodically.

Checking the registration information of the process monitoring resource

Check the registration information for the process monitoring resource. You can also configure resource attributes by selecting the *Attributes* tab and switching the screen.

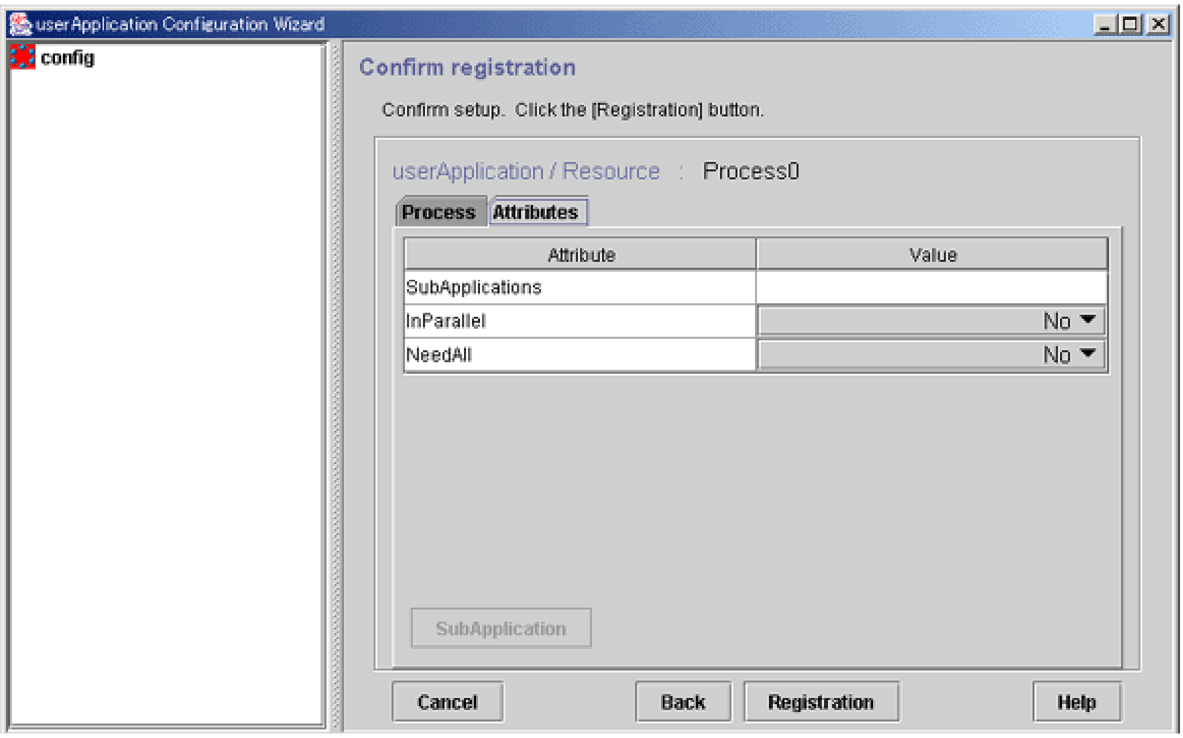
For information on the resource attributes, see "6.7.5 Attributes."

Figure 6.46 Confirm registration



Flag

RetryCount indicates the number of times the process is to be restarted.
RetryInterval indicates the interval before starting the process.
Initialize=Yes indicates that the retry count of the process is to be initialized periodically. If *No* was specified for *Initialize No. of process restart* in the "Set up process monitoring attribute" screen, this attribute is not displayed.
Daemon=Yes indicates that the process is to be started as a daemon. If *Process is daemon* was not checked at the command setup screen, this attribute is not displayed.



SubApplication button

This button associates the other processing monitoring resources to the Cmdline or process monitoring resource that has been created above. This button can be selected only if there are resources that can be associated. For setting instructions, see "Resource association" in "6.7.1.1 Creating Cmdline Resourcesmdline Resources."

After checking the registration information, click *Registration*.

6.7.1.8 Creating Line Switching Unit Resources

This section describes how to create line switching unit resources.

You should complete on the prerequisites before creating a line switching unit resource.

Line switching unit resources are available only in an Oracle Solaris 10 environment.

6.7.1.8.1 Prerequisites

You must always perform the setup described below before using a line switching unit resource.

Resource registration

To use a line switching unit, you must first register a switching line resource (SH_SWLine class) to the resource database.

This section describes the procedure for registering a switching line resource to the resource database.

Registration procedure flow

1. Check the resource name of the line switching unit.
2. Register the switching line resource.
3. Check the registration information.

Registration procedure

1. Confirmation of the resource name of the line switching unit resource

   Check the resource names of the line switching units that are registered to the resource database by using the "clgettree(1)" command.

```
# clgettree
Cluster 1 cluster
     Domain 2 CLUSTER
          Shared 7 SHD_CLUSTER
                  SHD_DISK 21 SHD_Disk21 UNKNOWN
                          DISK 22 c5t0d0 ON node1
                          DISK 27 c4t0d0 ON node2
                  SH_SWU 18 SWU2002 UNKNOWN
          Node 3 node1 ON
                  Ethernet 29 hme0 ON
                  DISK 19 c0t0d0 UNKNOWN
                  DISK 22 c5t0d0 ON
          Node 5 node2 ON
                  Ethernet 30 hme0 ON
                  DISK 25 c0t0d0 UNKNOWN
                  DISK 27 c4t0d0 ON
```

   In this example, the line indicating "SH_SWU" shows the resource name is "SWU2002". If you cannot confirm the resource name, see "5.1.3.2 Automatic Configure" and register the line switching units to the resource database.

   See

   - For details on the information that is output by executing the "clgettree(1)" command, see "Explanation of resources related to line switching units."

- For details on the "clgettree(1)" command, see the manual page.

1. Registration of switching line resources

   Execute the "claddswursc(1M)" command to add the switching line resources to the resource database.

```
# claddswursc -k sh_swl_1 -s SWU2002 -0 node1 -1 node2 -m 0x3
```

   In this example, the line switching unit called "SWU2002," which was checked in step 2, is used to register the switching line resource called "sh_swl_1."

   Port 0 of the switching unit is connected to "node1," and port 1 to "node2."

   Since the two switching units LSU01 and LSU00 are to be used, the mask value is set to 0x3.

   See

   For details on the "claddswursc(1M)" command, see the manual page.

2. Confirmation of registration information

   Execute the "clgettree(1)" command to check that the switching unit resources of the line switching unit have been registered to the resource database.

   Example) Resources of the "SWLine" class are the switching unit resources of the line switching unit.
   The resources of the "SH_SWLine" class show a sharing relationship of the switching units for the line switching unit.

```
     # /etc/opt/FJSVcluster/bin/clgettree
     Cluster 1 cluster
          Domain 2 CLUSTER
               Shared 7 SHD_CLUSTER
                       SHD_DISK 21 SHD_Disk21 UNKNOWN
                               DISK 22 c5t0d0 ON node1
                               DISK 27 c4t0d0 ON node2
                       SH_SWU 18 SWU2002 UNKNOWN
                               SH_SWLine 19 sh_swl_1 UNKNOWN
                                     SWLine 35 sh_swl_1P0 UNKNOWN node1
                                     SWLine 36 sh_swl_1P1 UNKNOWN node2
               Node 3 node1 ON
                       Ethernet 29 hme0 ON
                       DISK 19 c0t0d0 UNKNOWN
                       DISK 22 c5t0d0 ON
                       SWLine 35 sh_swl_1P0 UNKNOWN
               Node 5 node2 ON
                       Ethernet 30 hme0 ON
                       DISK 25 c0t0d0 UNKNOWN
                       DISK 27 c4t0d0 ON
                       SWLine 36 sh_swl_1P1 UNKNOWN
```

Explanation of resources related to line switching units

This section explains the information that is output by the "clgettree" command.

[Output example]

```
Cluster 1 cluster
     Domain 2 CLUSTER
          Shared 7 SHD_CLUSTER
               SHD_DISK 21 SHD_Disk21 UNKNOWN
                       DISK 22 c5t0d0 ON node1
                       DISK 27 c4t0d0 ON node2
               SH_SWU 18 SWU2002 UNKNOWN                         .....(a)
                       SH_SWLine 19 sh_swl_1 UNKNOWN              .....(b)
                             SWLine 35 sh_swl_1P0 UNKNOWN node1   .....(c)
```

```
                       SWLine 36 sh_swl_1P1 UNKNOWN node2   .....(d)
        Node 3 node1 ON
                Ethernet 29 hme0 ON
                DISK 19 c0t0d0 UNKNOWN
                DISK 22 c5t0d0 ON
                SWLine 35 sh_swl_1P0 UNKNOWN
        Node 5 node2 ON
                Ethernet 30 hme0 ON
                DISK 25 c0t0d0 UNKNOWN
                DISK 27 c4t0d0 ON
                SWLine 36 sh_swl_1P1 UNKNOWN
```

[Explanation]

(a) This line identifies a line switching unit resource.

In the example shown in the above figure, "SWU2002" is the resource name of the resource representing the line switching unit.
This resource is displayed when PRIMECLUSTER recognizes the line switching unit when automatic resource registration is executed.

(b) This line identifies a switching line shared resource of the line switching unit.

In the example shown in the above figure, "sh_swl_1" is the resource name of the switching line shared resource (switching line name).
This resource is displayed when the switching line resources of the line switching unit are registered to the resource database with the "claddswursc" command.
As shown in the example of the above figure, this resource belongs to the "SH_SWLine" class.

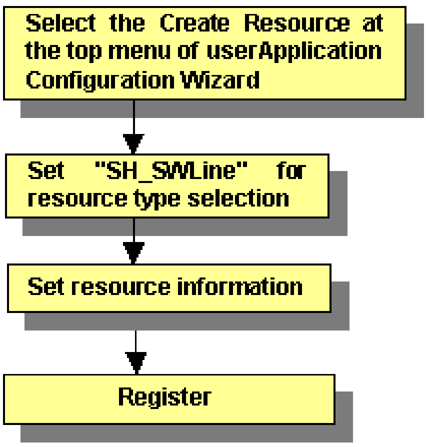
(c),(d)

These lines represent switching line resources of the line switching unit.
In the example shown in the above figure, "sh_swl_1P0" and "sh_swl_1P1" are resource names for switching line resources.
These resources are displayed when the switching line resources of the line switching unit are registered to the resource database with the "claddswursc" command.
As shown in the example of the above figure, these resources belong to the "SWLine" class.

6.7.1.8.2 Setup Procedure

This section describes the procedure for setting up a line switching unit resource.

Figure 6.47 Flow of creating a line switching unit resource



For information on the above operations up to "Selecting SH_SWLine for resource type selection," see "6.7.1.1 Creating Cmdline Resourcesmdline Resources." This section describes the operations from "Set resource information."

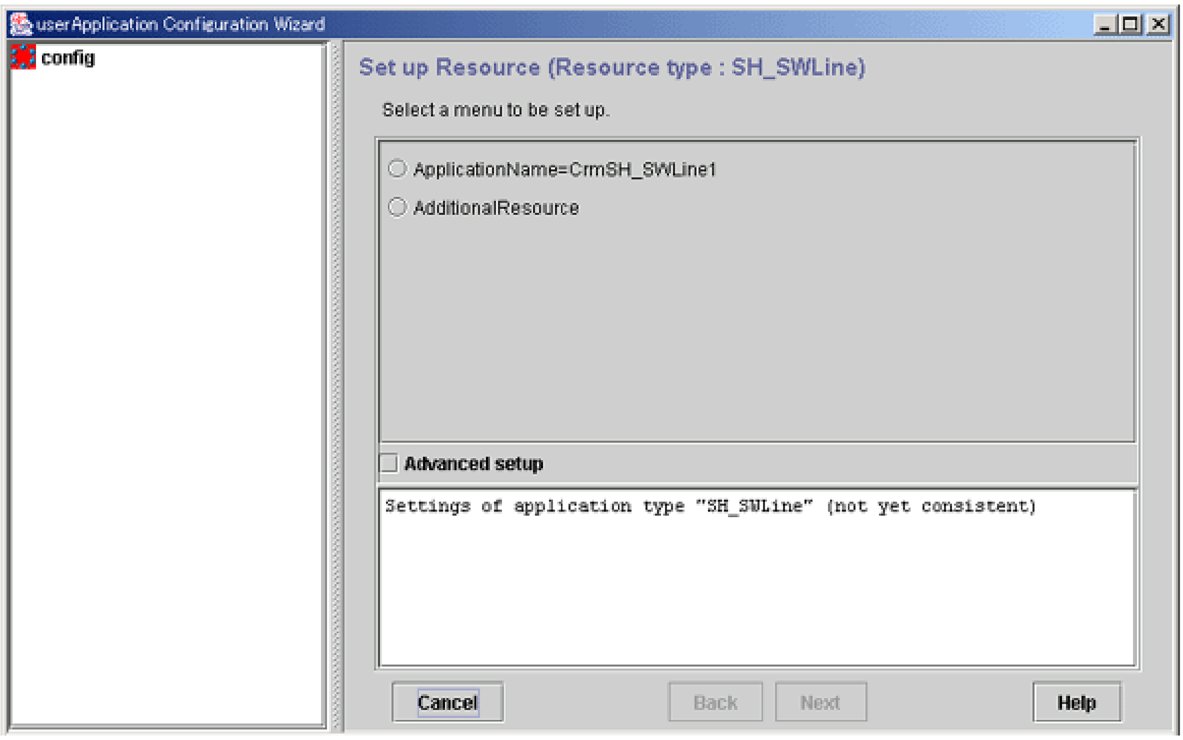
Note

The displayed Help information when you click the *Help* button includes the information for CUI only.

Setting resource information

1. Display the top screen.

   Select *SH_SWLine* for the resource type selection. The following screen is displayed:



   To terminate the selection process, click the *Cancel* button.

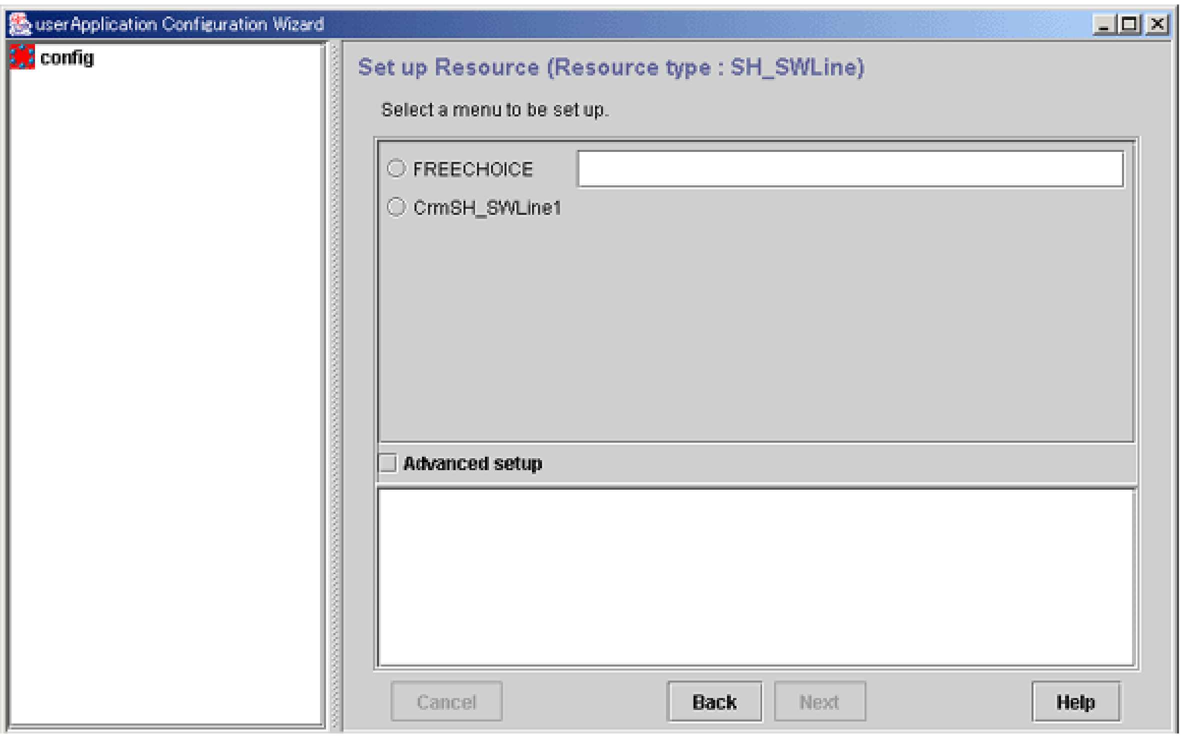
2. Check the name of the line switching unit resource.

   name in *ApplicationName=name* is the name of the line switching unit resource to be created.

   - If the name is correct, go to step 5.
   - To change the name, go to step 3.

3. Select *ApplicationName=name* and click the *Next* button.

   The screen for changing the resource name is displayed.

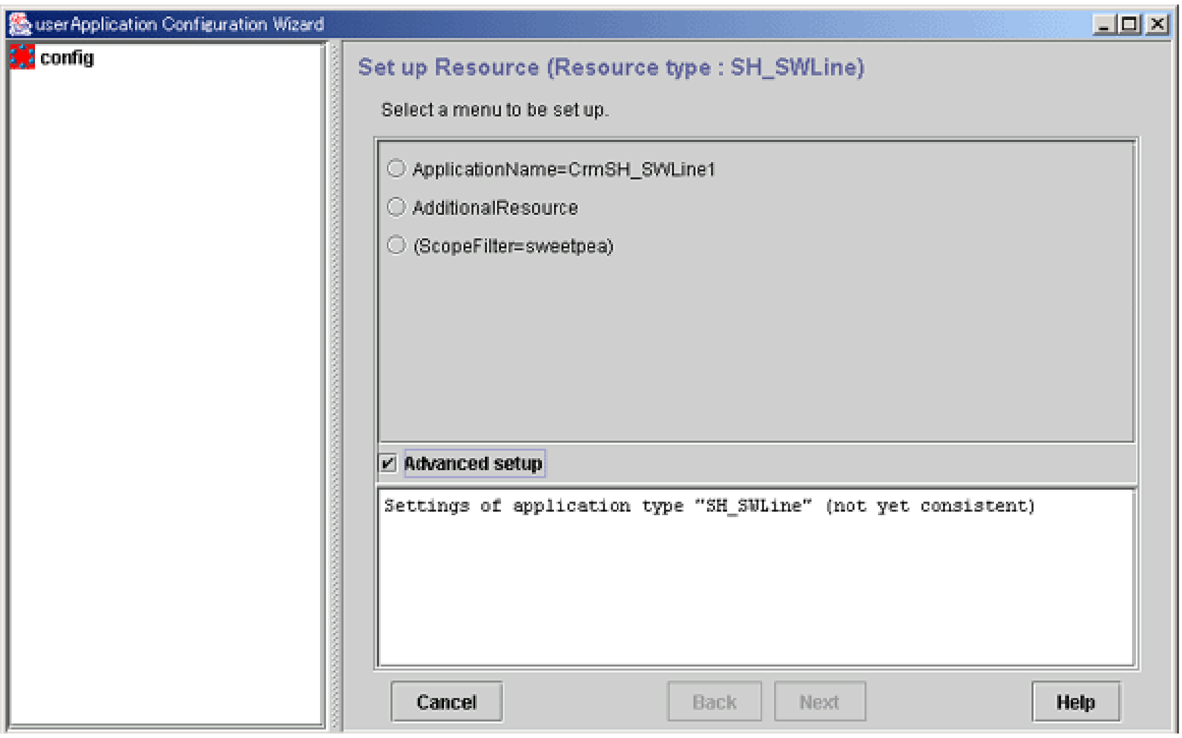


4. Click *FREECHOICE* and enter the resource name. After entering the name, click the *Next* button.

   An updated top screen containing the entered information is displayed. Go to step 2 to check the information.

5. Select the *Advanced setup* checkbox.

   A new menu will be added to the screen.



6. Check the nodes to which the line switching unit is connected.

   node-names in *ScopeFilter=node-name* shows the CF node names, which are separated with colons (":").

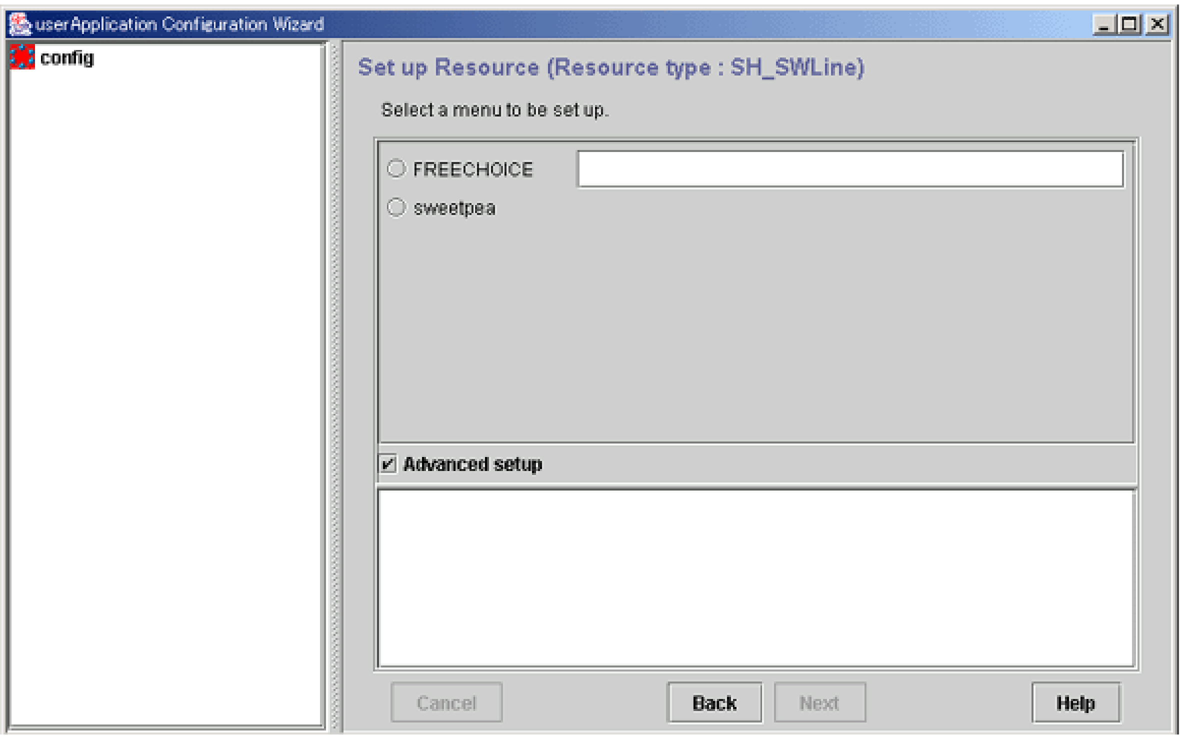
   - If all nodes to which the line switching unit is connected are displayed, go to step 9.
   - If there is an error, go to step 7.

Information

Of the switching line resources (SH_SWLine class) that are registered to the resource database, those resources that are common to these nodes can be added to the line switching unit resource according to the procedure described later.

7. Select *ScopeFilter=node-names*" and click *Next* button.

   The screen for entering the nodes to which the line switching unit is connected is displayed.

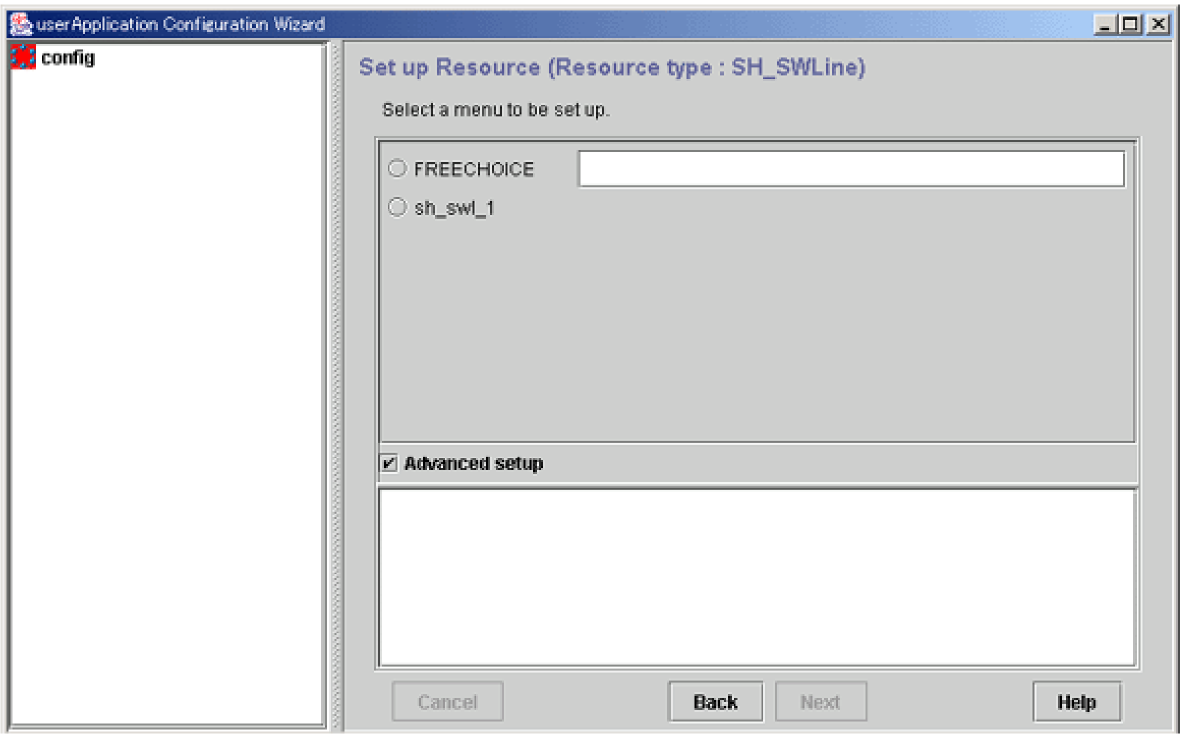


8. Click *FREECHOICE* and enter the names of the CF nodes to which the line switching unit is connected. Separate the names with colons (":"). After entering the CF node names, click the *Next* button.

   An updated top screen containing the entered information is displayed. Go to step 5 and check the information.

9. Select *AdditionalResource* and then click the *Next* button.

   The screen for selecting switching line resources that are registered to the resource database is displayed.

   In the example shown in this screen, you can select the switching line resource called *sh_swl_1*. If no other data is displayed besides *FREECHOICE*, the following causes can be considered:

   - There is an error in the nodes that were checked in step 5.
   - No switching line resource is registered to the resource database.

   For the first case, return to step 5 and check the nodes. To return to step 5, click the *Back* button.

   For the second case, see "6.7.1.8.1 Prerequisites" and register a switching line resource to the resource database. To do this, click the *Back* button, and then click the *Cancel* button in the displayed screen. This stops the setup process, and you can start again from the beginning.

10. Select switching line resource from the candidate list, and click *Next*.

    The top menu to which the switching line resource was added is displayed.

    The information is displayed as follows:

```
   cluster_resource[number]=rsc-name
         number     A number corresponding to the sequence in which the resource was added is
assigned.
         rsc-name   The resource name that was added is displayed.
```

Note

The attribute must not be changed. If you add a switching line resource, the Attribute menu is added. If you use this menu to change any attribute of the switching line resource, RMS will not operate correctly.

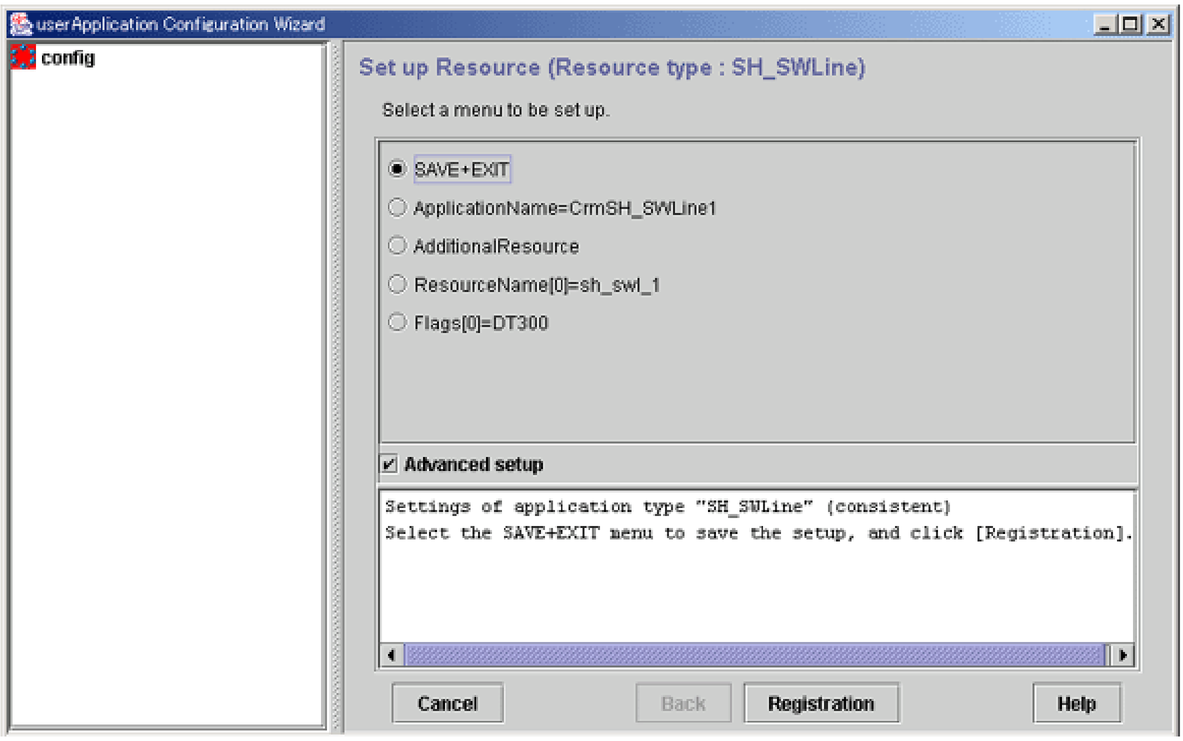
Go to step 1 of "Registering the information."

Registering the information

1. Select *SAVE+EXIT*.
   When you select *SAVE+EXIT*, the *Next* button changes to the *Registration* button.

   Note

   If there is information that can be registered, the *SAVE+EXIT* menu will be added to the top screen.

2. Click the *Registration* button.

   A line switching unit resource will be created according to the specified information, and the "userApplication Configuration Wizard Menu" screen will be displayed.
   This ends the procedure.

6.7.1.9 Creating ISV Resources

For information on ISV resource creation, refer to the manuals for the individual products.

6.7.2 Creating Cluster Applications

This section describes how to use the "userApplication Configuration Wizard" to register cluster applications (userApplication) to RMS.

Note that the operation procedure varies depending on the topology of the cluster application.

Note

userApplication name

When you create a cluster application, enter a name of the userApplication and then select a topology.
The userApplication name must be a string of 18 or fewer characters starting with an alphabetic letter and consisting of alphanumerics and "_" (underscore) only.

Moreover, for this name, do not use the same name ("generic") as the type of userApplication.

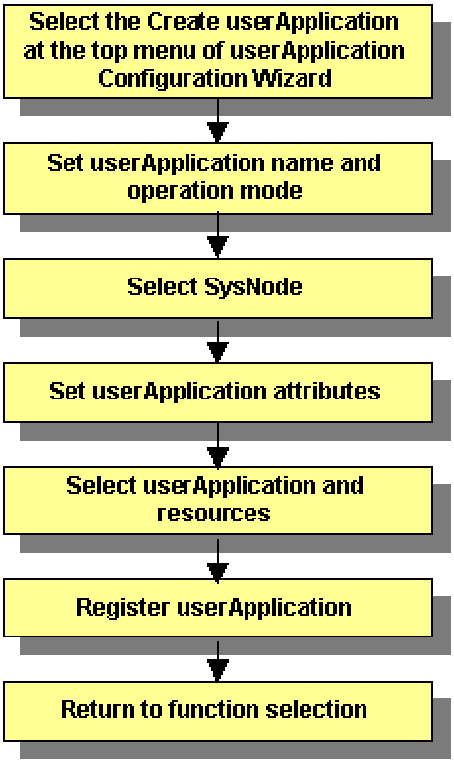
Example: userApp_0

6.7.2.1 Creating Standby Cluster Applications

This section explains how to register standby cluster applications to RMS by using the "userApplication Configuration Wizard."

The flow of the process for registering the Standby cluster applications is shown below.

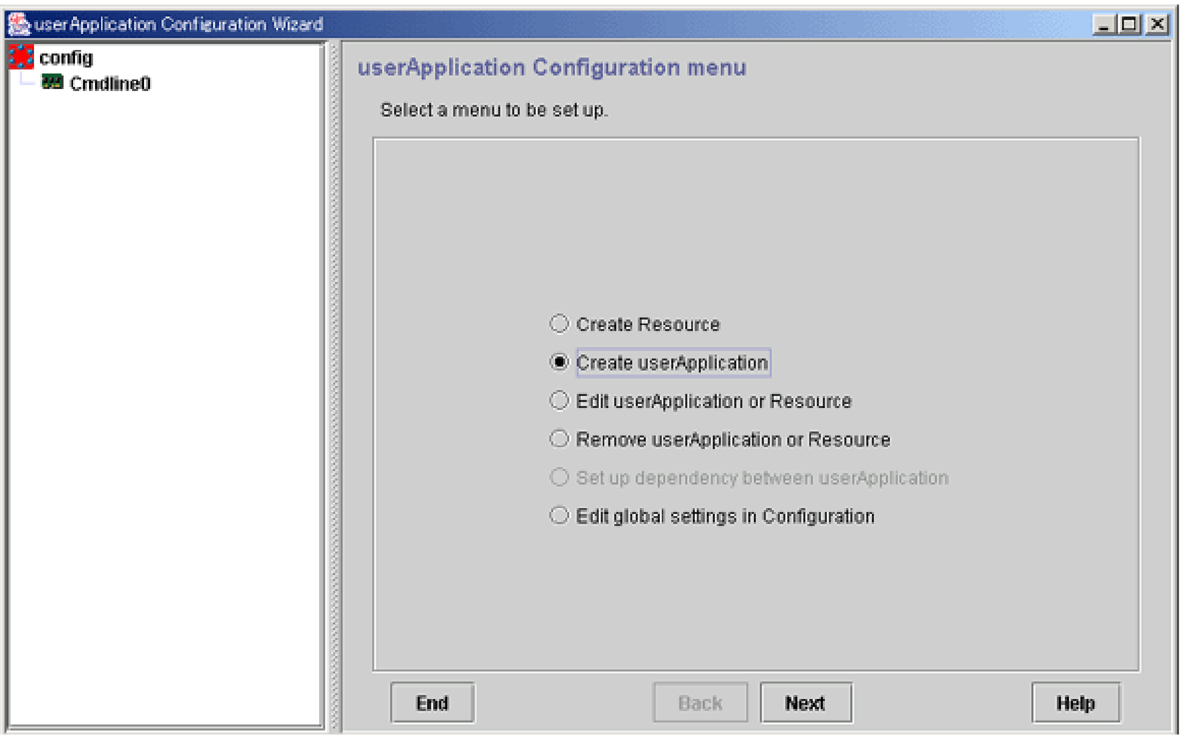
Figure 6.48 Flow of cluster application registration



Starting cluster application creation

At the top menu of the "userApplication Configuration Wizard," select *Create userApplication*.

Figure 6.49 Top menu of userApplication Configuration Wizard

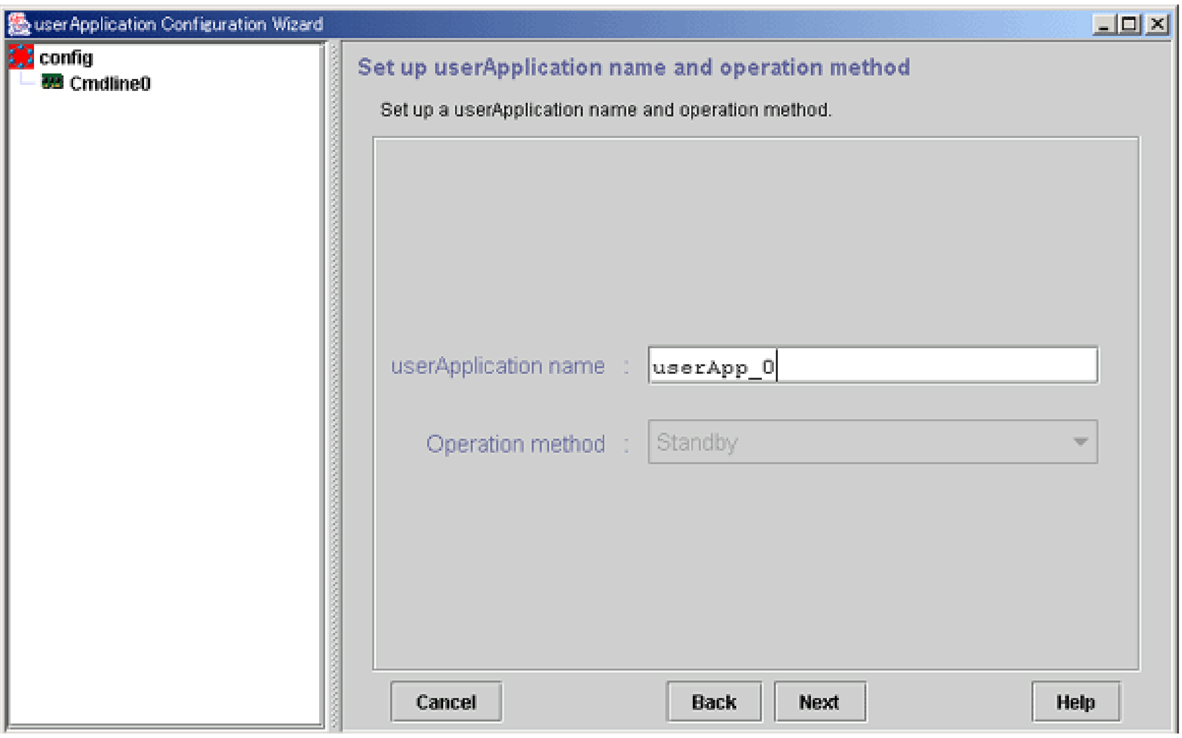


Click *Next* to go to the "Set up userApplication name and operation method" screen.

Setting the userApplication name and the operation method

Enter the userApplication name and the operation method.

Figure 6.50 Set up userApplication name and operation method



userApplication name

Specify the name of the cluster application.
Specify the input character string using up to 18 characters starting with an alphabetic letter and consisting of only alphanumeric characters and "_" (underscore).

Operation method

Specify Standby (standby operation) or Scalable (scalable operation) as the operation method of the cluster application.

- Standby

  This is the default operation method. When multiple SysNode are selected, "priority" values must be set.

- Scalable

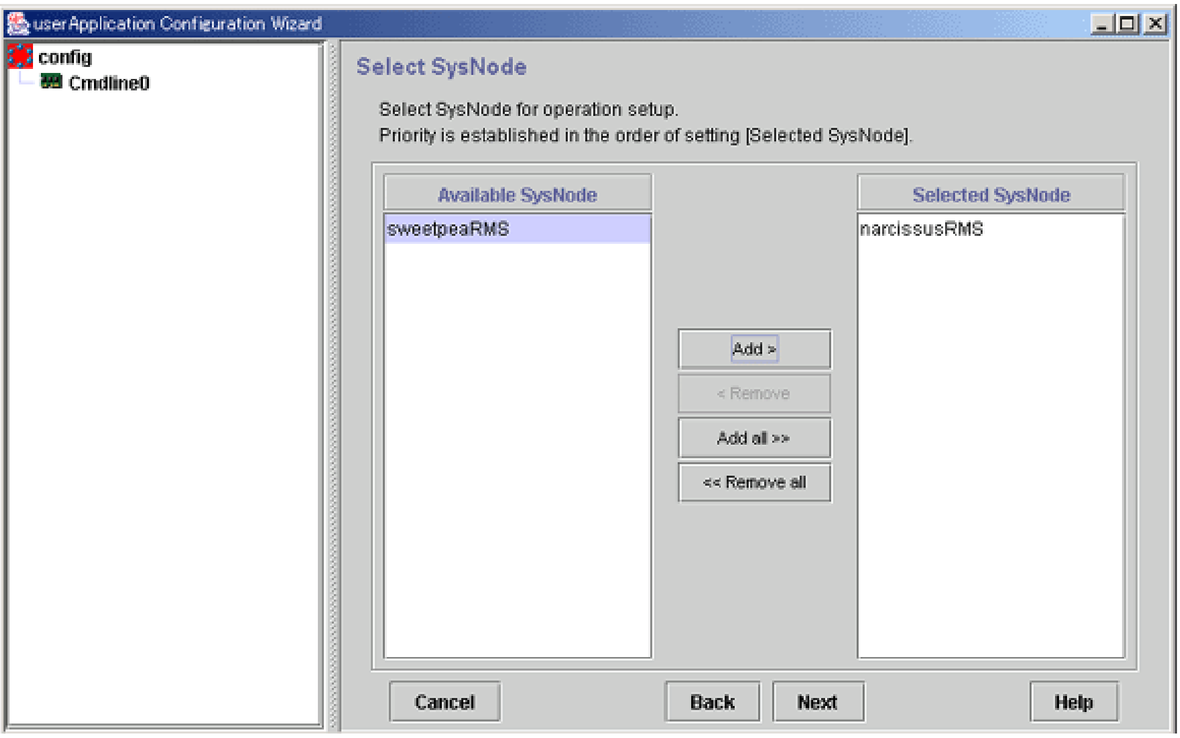
  This is only for the cluster application resources displayed in *Resource Selection*. The priority does not need to be set even if multiple SysNode are selected.

After completing the setup, click *Next* to go to the "Select SysNode" screen.

Selecting a SysNode

Select the SysNode in which the cluster application is to be configured.

Figure 6.51 Select SysNode



Available SysNode

Contains names of available SysNodes that belong to the specified RMS configuration.

Selected SysNode

The name of the SysNode in which the cluster application is to be operated. The priority of SysNode is determined by the sequence in which SysNode are listed in this box. SysNode displayed at the top of the list becomes the Online SysNode when userApplication is started.

From *Available SysNode*, select the SysNodes in which the cluster application is to be built, and then click *Add*. To add all listed SysNodes, click *Add all*. To remove a SysNode, select the SysNode to be removed from *Selected SysNode*, and then click *Remove*. To remove all listed SysNodes, click *Remove all*.

After completing the setup, click *Next* to go to the "Set up attribute" screen.

Setting cluster application attributes

Refer to the items below when setting up the cluster application attributes:

- To start the cluster application automatically when the cluster is started

  Set "AutoStartUp" to "Yes."

  Note

  If you use a standby cluster application for scalable operation, you must set "AutoStartUp" to "No." Scalable operation controls the startup of a standby cluster application.

- Action to take if another error occurs during failover (double failure)

  Set "HaltFlag" to "Yes."

  Note

  To ensure safe operation, be sure to set this item to "Yes."

- For Standby operation

  To monitor the resource state of the standby node, set "StandbyTransitions" to "ClearFaultRequest|StartUp|SwitchRequest."

  Note

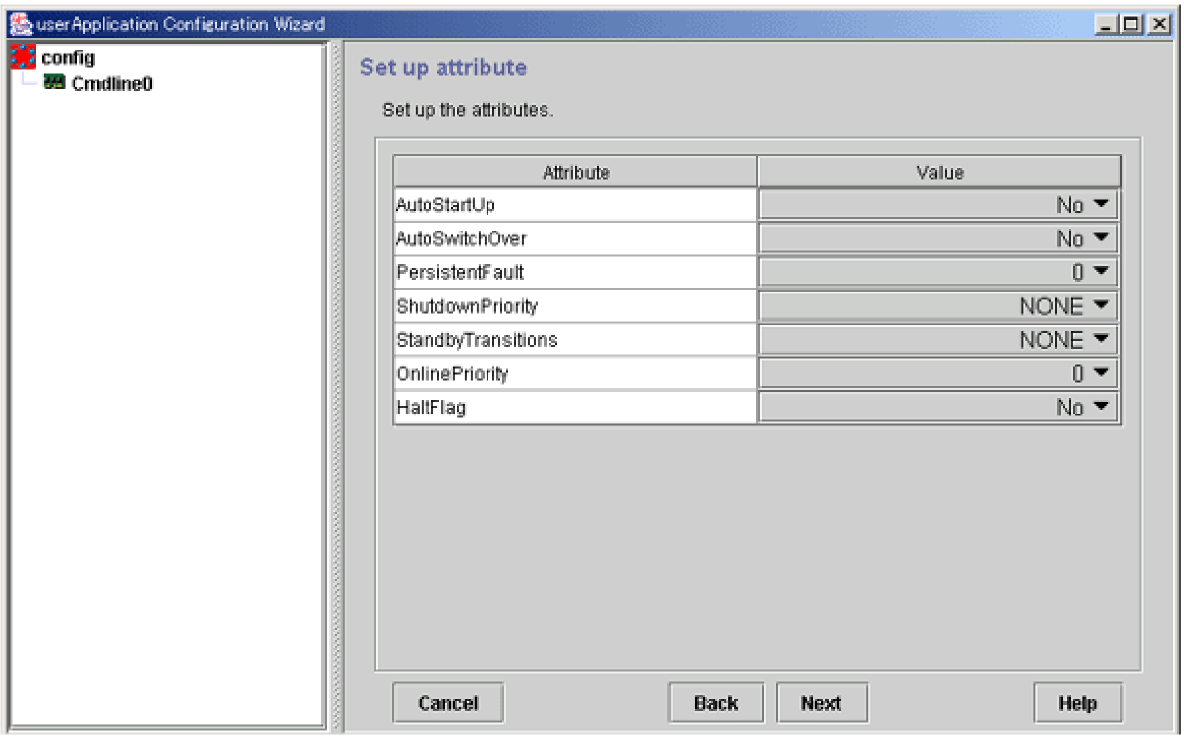
  If GLS (Redundant Line Control Function) is used, please make sure to set "ClearFaultRequest|StartUp|SwitchRequest" for "StandbyTransitions" so that the Gls resources on a Standby node can be monitored.

- In the case of the single-node cluster operation

  - Do not set a value to "AutoSwitchOver".
  - Set "No" to "HaltFlag".
  - Do not set a value to "ShutdownPriority".
  - Do not set a value to "OnlinePriority".

For details on attribute, see "6.7.5 Attributes."

Figure 6.52 Set up attribute

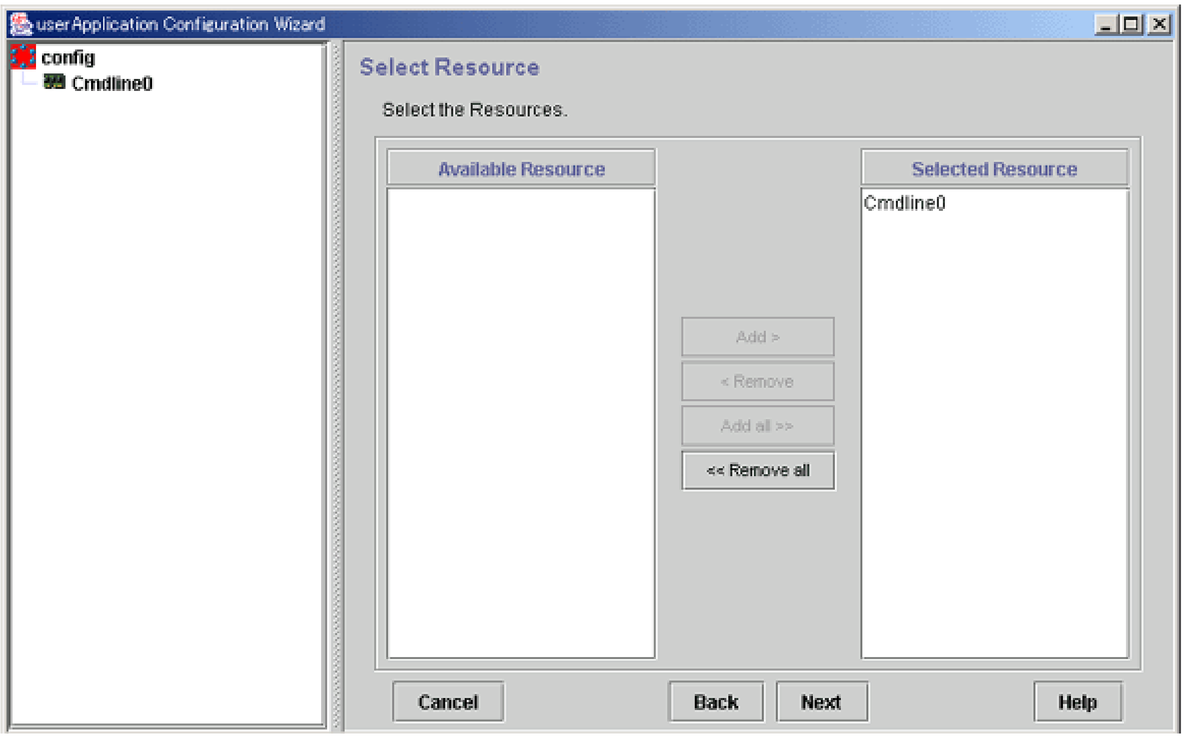


After completing the set up, click *Next* to go to the "Select Resource" screen.

Selecting resources

Enter the resources to be used by the cluster application.

Figure 6.53 Select Resource



Available Resource

The names of the resources that can be shared among the selected SysNodes are displayed. An available resource is one that was created with the same name in all the SysNodes that are specified by the cluster application.
You cannot share one resource with multiple userApplication. The resource that is used for other userApplication is not displayed.

Selected Resource

The resources to be registered to the cluster application are displayed.

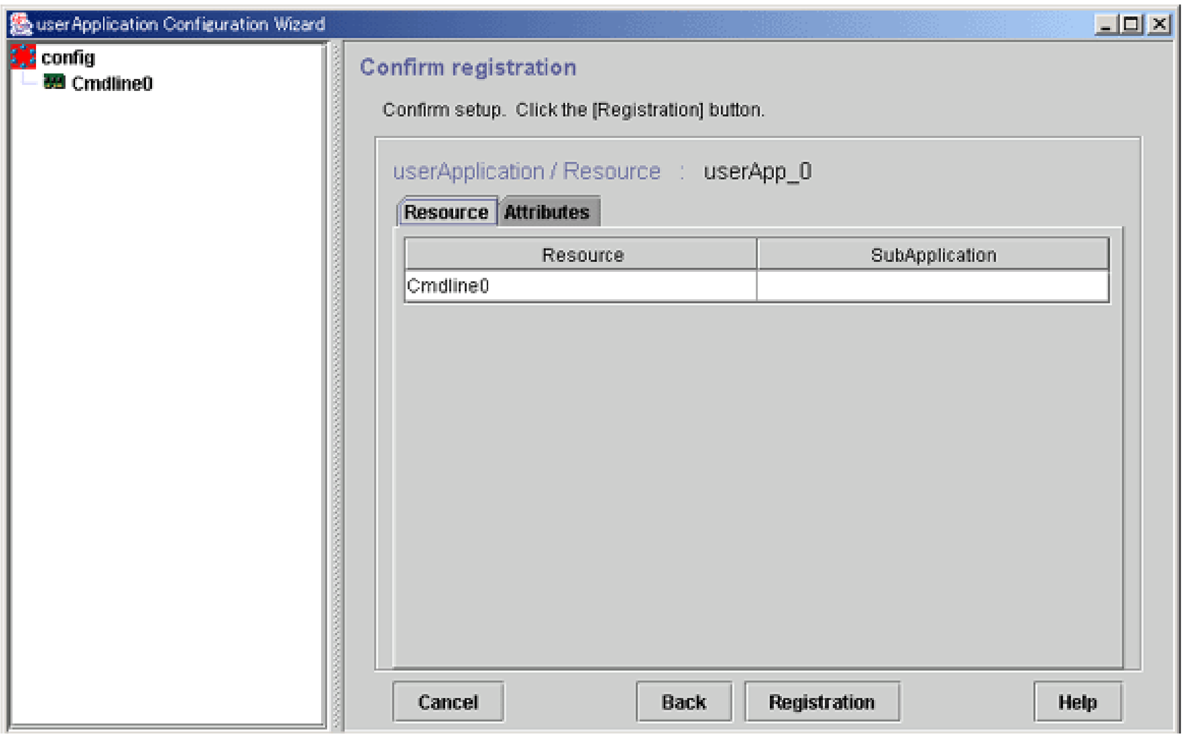
From *Available Resource*, select the resources to be used by the cluster application, and then click *Add*. To add all listed resources, click *Add all*. To remove a resource, select the resource to be removed from *Selected Resource*, and then click *Remove*. To remove all listed resources, click *Remove all*.

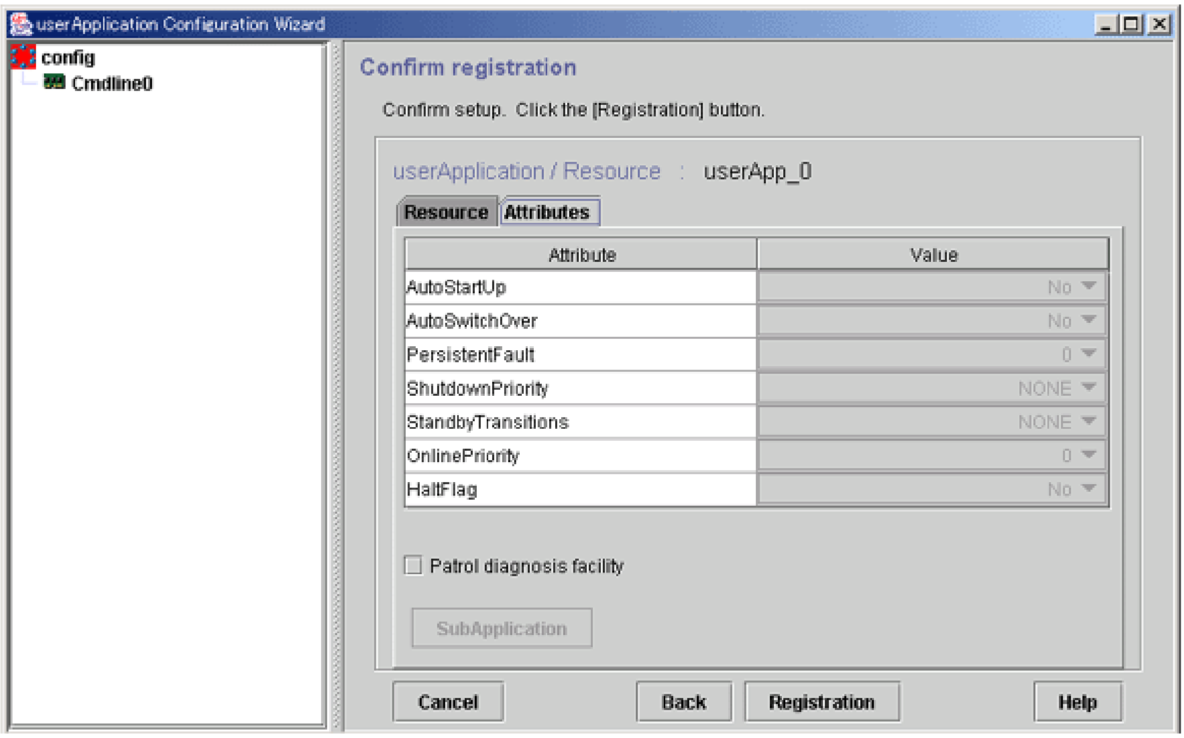
After completing the setup, click *Next* to go to the "Confirm registration" screen.

Checking the registration information of a cluster application

Check the registration information of the cluster application.

Figure 6.54 Confirm registration





Patrol diagnosis facility **checkbox**

Select this checkbox to enable the patrol diagnosis facility in the cluster application. See "6.9 Setting Up Patrol Diagnosis."

SubApplication **button**

This button is always inactive.

The attributes displayed here cannot be changed. To change the attributes, you must return to "Setting cluster application attributes."

Check the registration information, and then click *Registration*.

Note

- The patrol diagnosis facility can be used in Standby operation. To enable the patrol diagnosis facility, set "ClearFaultRequest|StartUp|SwitchRequest" to "StandbyTransitions" of the cluster application. For details on attributes, see "6.7.5 Attributes."
- The resource names of patrol diagnosis facility are displayed as follows. You cannot change these resource names:
  - userApplication Configuration Wizard

    "patrol_" + "cluster application name"
  - RMS main window of Cluster Admin or hvdisp(1M) command

    "Patrol" + "value"
- The resource names of procedure resources are displayed as follows:
  - userApplication Configuration Wizard

    Any name set when the resource is created
  - RMS main window of Cluster Admin or hvdisp(1M) command

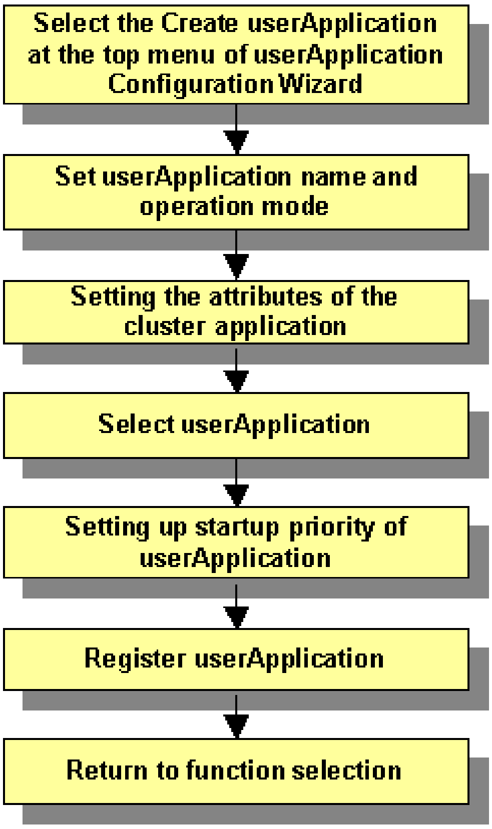
    "Pre-registered resource name on CRM" + "."+ "Resource ID on CRM"

6.7.2.2 Creating Scalable Cluster Applications

This section explains how to register scalable cluster applications to RMS by using the "userApplication Configuration Wizard."

The flow of the procedure for registering a scalable cluster application is shown below:

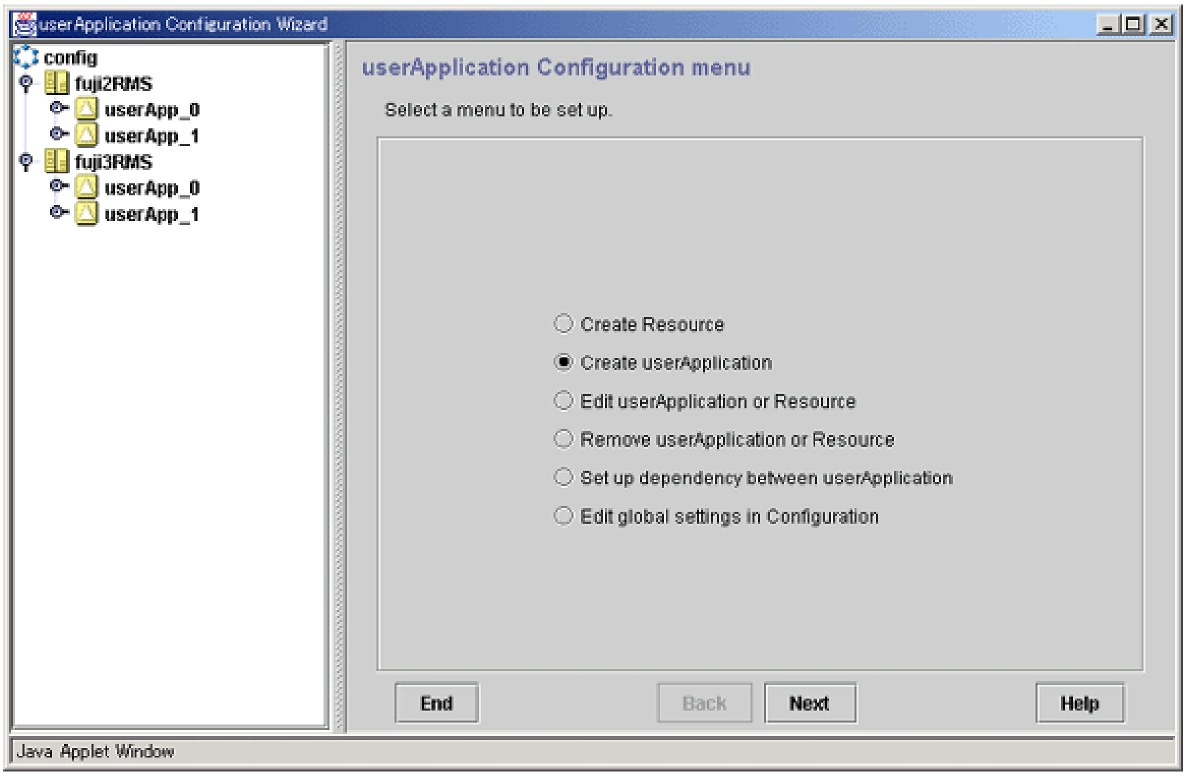
Figure 6.55 Flow of Scalable cluster application registration



Preparing standby cluster applications

The scalable cluster application performs scalable operation by interconnecting multiple standby cluster applications. Before creating a scalable cluster application, you have to create the standby cluster applications that constitute the components of the scalable cluster application.

Figure 6.56 Preparing Standby cluster applications

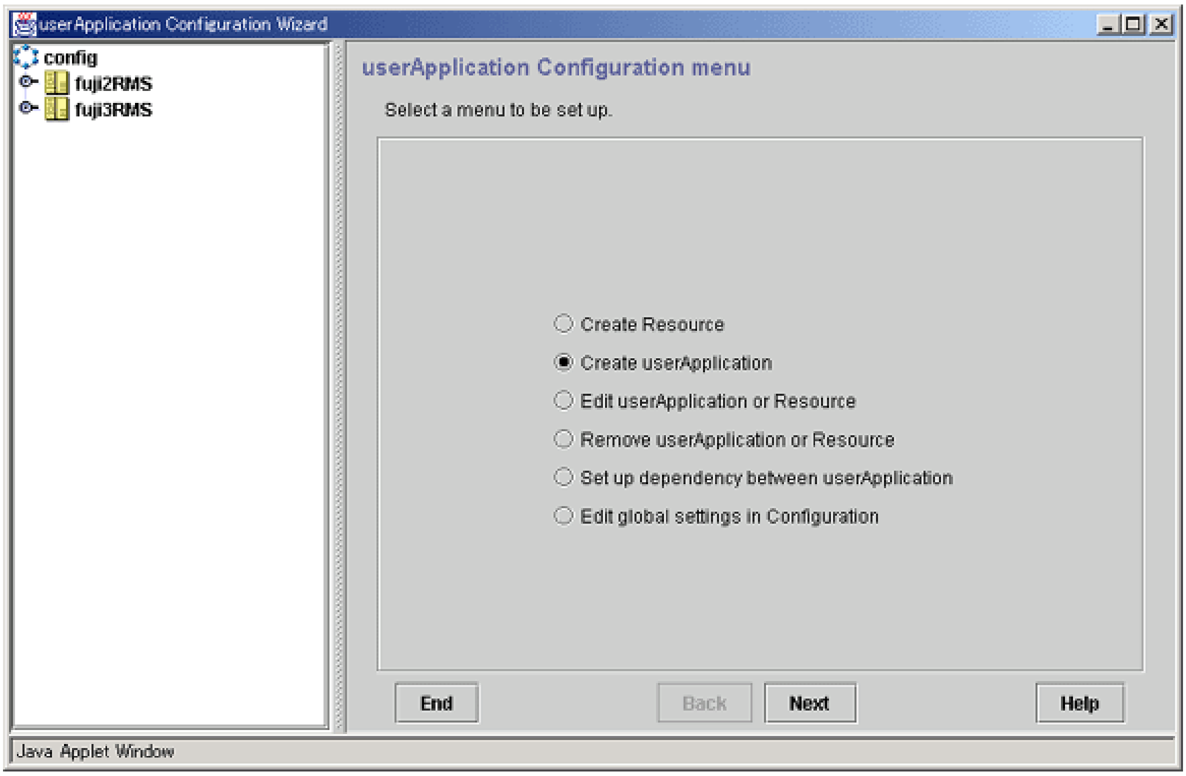


For details on how to create Standby cluster applications, see "6.7.2.1 Creating Standby Cluster Applications."

Starting up Create Cluster Applications

Specify *Create userApplication* from the top menu of "userApplication Configuration Wizard."

Figure 6.57 userApplication Configuration Wizard top menu

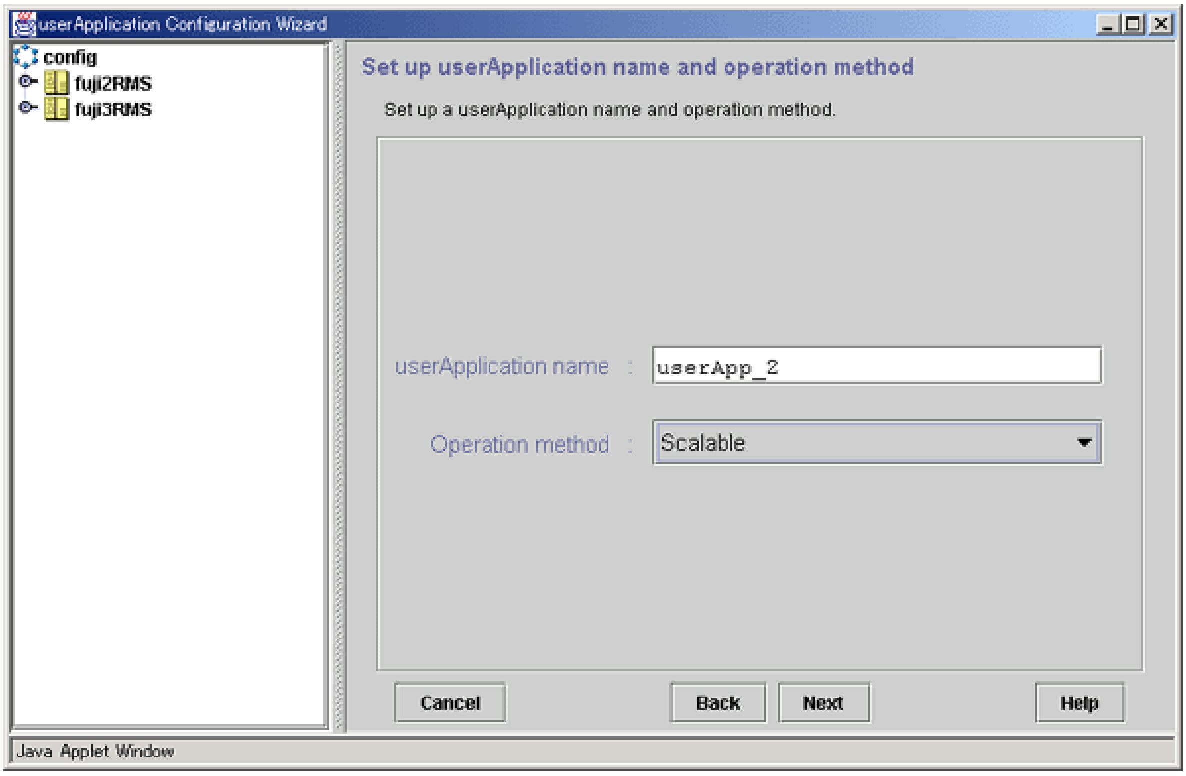


Click *Next* to go to the "Set up userApplication name and operation method" screen.

Setting userApplication name and the topology

Set up userApplication name and the topology.

Figure 6.58 Setting up userApplication name and the operation method



userApplication name

Specify a name for the cluster application that you want to set up.
Specify an input character string
Specify the input character string with up to 18 characters with an alphabetic letter and consisting of only alphanumeric characters and "_" (underscore).

Operation method

Select Scalable.

After you have made all the necessary settings, click *Next* to go to the "Set up attribute" screen.

Setting the attributes of the cluster application

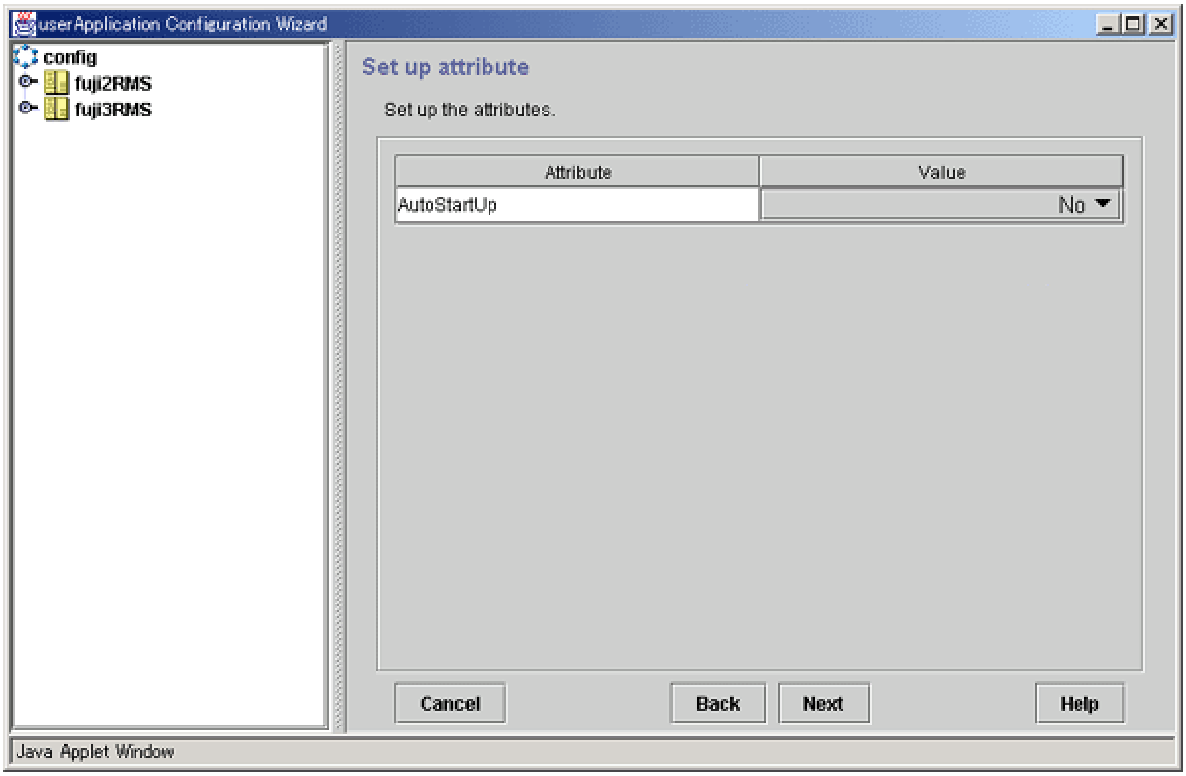
Set up the attributes of the cluster application as described below:

- Set up automatic start of the scalable cluster application when the cluster starts up.

  Set up "*AutoStartUp*" to "*Yes*."

For details on this attribute, see "6.7.5 Attributes."

Figure 6.59 Set up attributes

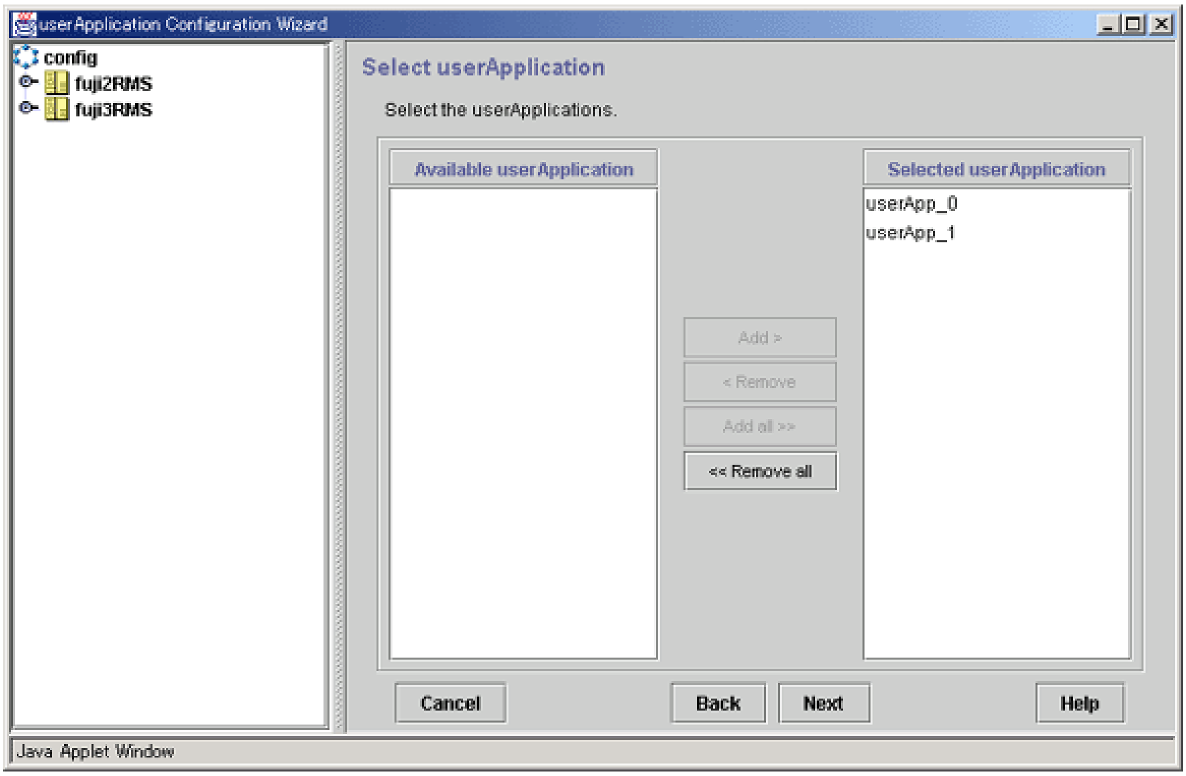


After you have made this setting, click *Next* to go to the "Select userApplication" screen.

Select userApplication

Select a cluster application for which scalable operation is to be performed.

Figure 6.60 Selecting userApplication



Available userApplication

Indicates a standby cluster application that can be used.

Selected userApplication

Indicates the cluster application that is being used for scalable operation.

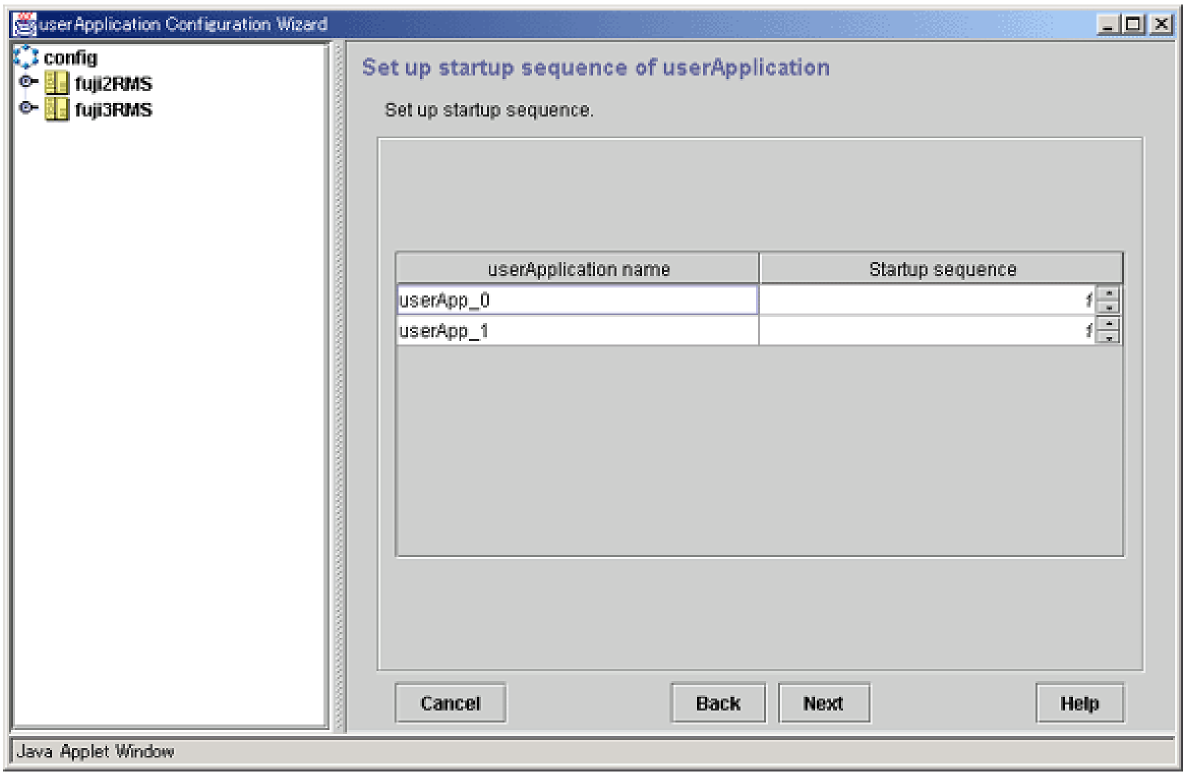
Select a cluster application for scalable operation from *Available userApplication*, and then click *Add*. If you want to add all the cluster applications, click *Add all*. To delete a cluster application from scalable operation, select the cluster application to be deleted from [*Selected userApplication*] and then click *Remove*. To delete all the cluster applications, click *Remove all*.

After you have made these settings, click *Next* and then go to the "Set up startup priority of userApplication" screen.

Setting up startup priority of userApplication

Set up the startup order of the cluster applications.

Figure 6.61 Selecting userApplication



userApplication name

Cluster applications for scalable operation.

Startup priority

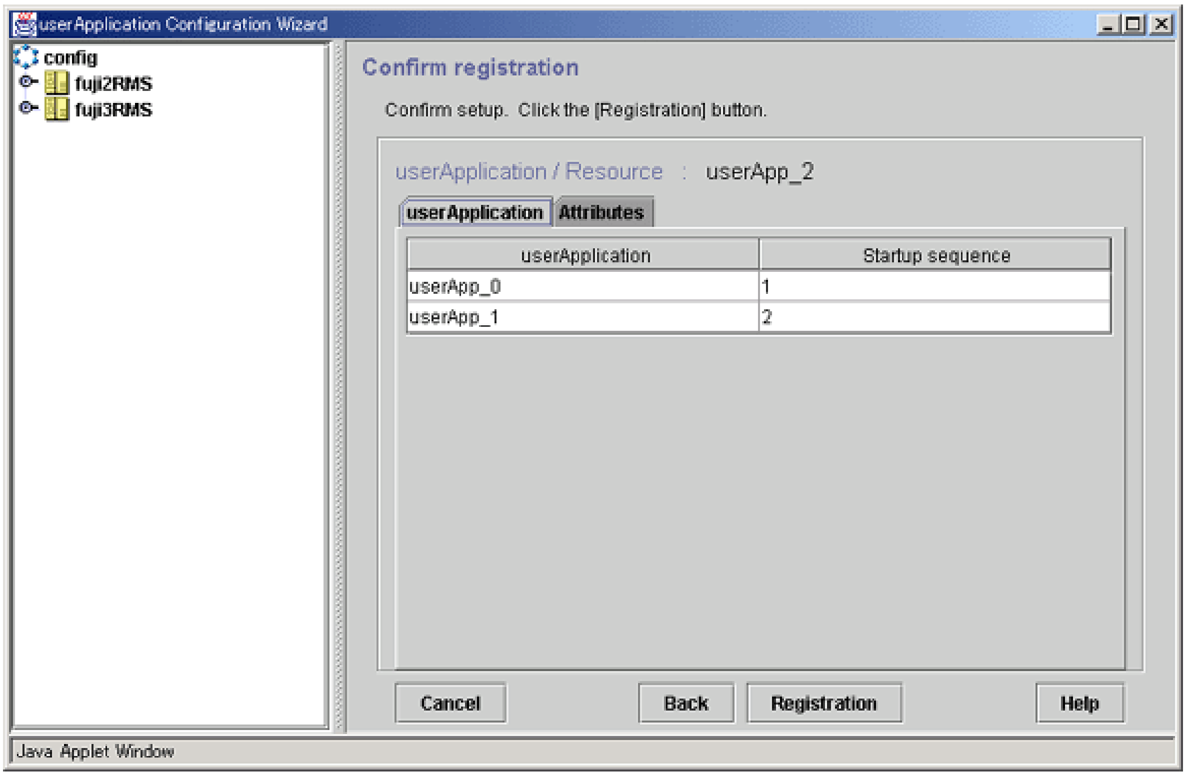
Set up a startup priority for the cluster applications. Cluster applications start up in order, starting from the smallest priority number. Conversely, cluster applications stop in order, starting from the greatest startup number. Cluster applications with the same startup priority start or stop in parallel.

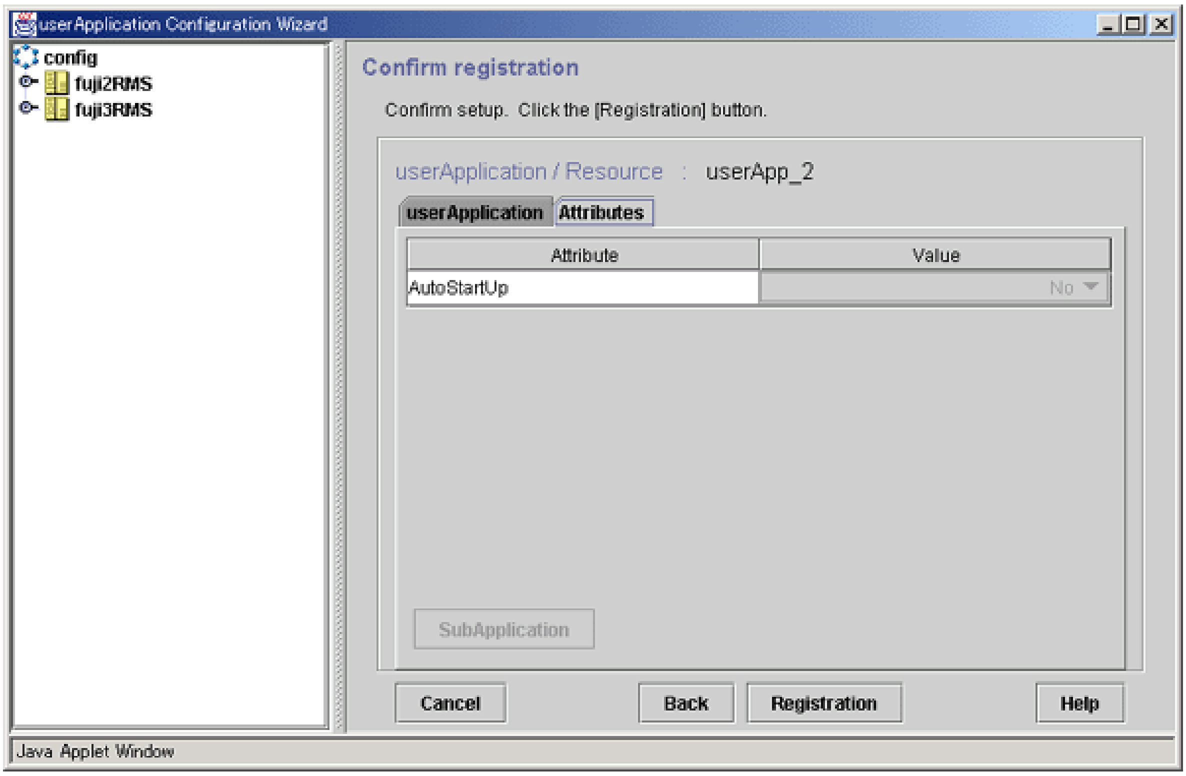
After you have made these settings, click *Next* to go to the "Set up attribute" screen.

Confirming cluster application registration information

Confirm the registration information of the cluster application.

Figure 6.62 Confirm registration





***SubApplication* button**

This button is always disabled.

After you have confirmed the registration information, click *Registration*.

6.7.3 Setting Up Dependency Relationships Between Cluster Applications

You can set up exclusivity and startup priority as part of the RMS Configuration settings.

- **Setting up exclusive relationships between cluster applications**

  Set exclusive control so that multiple cluster applications will not run concurrently in the same cluster node.

- **Setting the startup priority of a cluster application**

  Determine the sequence in which the cluster applications are to be started.

  If you omit this setup, the startup sequence of the cluster applications will be undefined.

Note

The startup priority setup function is not supported in this version.

6.7.3.1 Setting Exclusive Relationships Between Cluster Applications

Set up exclusive relationships for the cluster applications to prevent cluster applications with exclusive relationships from operating simultaneously. An exclusive relationship can be set up only between standby cluster applications. If failover occurs, determine the cluster applications that should continue operating by using the job priority of the cluster applications between which an exclusive relationship is set.

Information

- To set up an exclusive relationship, create a group of cluster applications between which an exclusive relationship is to be set. Up to 52 groups can be created. Exclusive control is established between cluster applications within a single group.
- The cluster application in which the exclusive relationship is set transits to Standby state according to the StandbyTransitions attribute.

Note

If a cluster application is in Faulted state in a node, you cannot start another cluster application with exclusive relationship to the cluster application in Faulted state on the same node.

This is because resources under the Faulted cluster application may not have been stopped.

In this case, clear the Faulted of the application and brings it to Offline. Then, start the cluster application with exclusive relationships.

For how to clear the Faulted state of a cluster application, see "7.2.2.4 Bringing Faulted Cluster Application to Available State."

The operation of cluster applications, between which an exclusive relationship is set up, during failover can be explained in the following two cases:

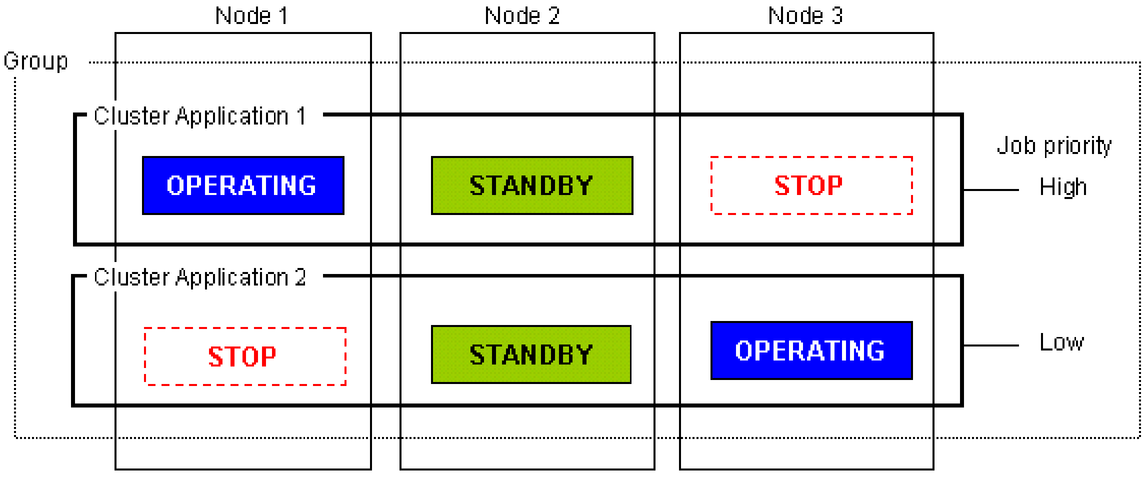
- When the job priorities are different
- When the job priorities are the same

When the job priorities of the cluster applications with an exclusive relationship are different

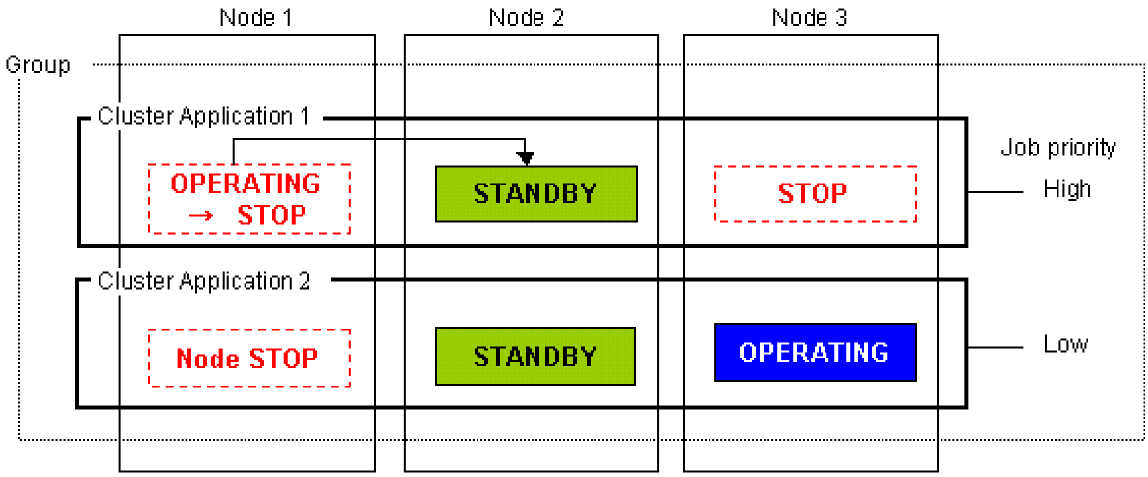
The "hvswitch(1M) -p" command is executed for a low-priority cluster application on the node where a high-priority cluster application is running, or a high-priority cluster application is switched. For the state transition of the cluster application, see the manual page of hvswitch(1M).

Failover of the cluster application with a high job priority

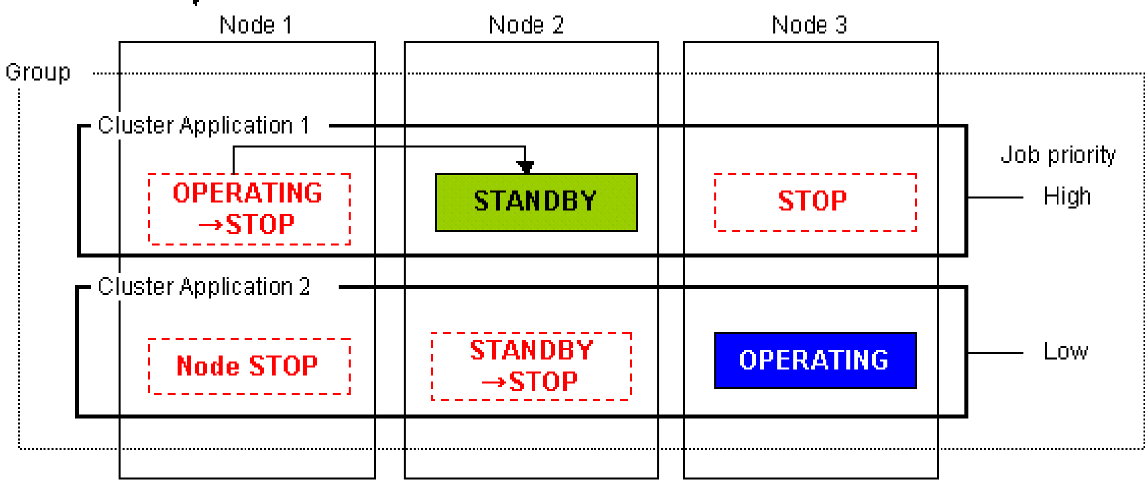
If failover occurs in a cluster application with a high job priority, the cluster application with the high job priority will always be in the operating state.



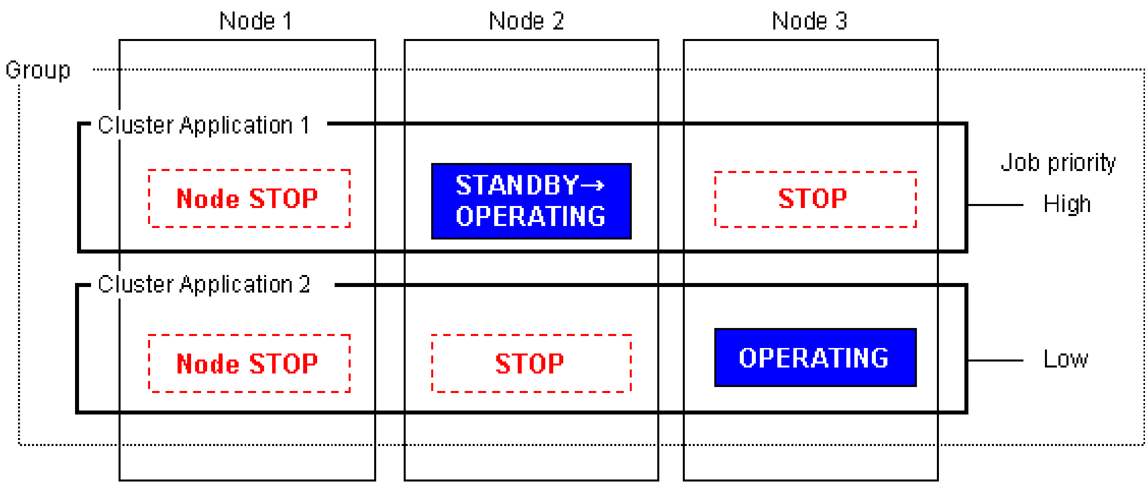
If OPERATING Node 1 using Cluster Application 1 stops, failover is attempted to STANDBY Node 2. However, Cluster Application 2 with low job priority is STANDBY on Node 2.



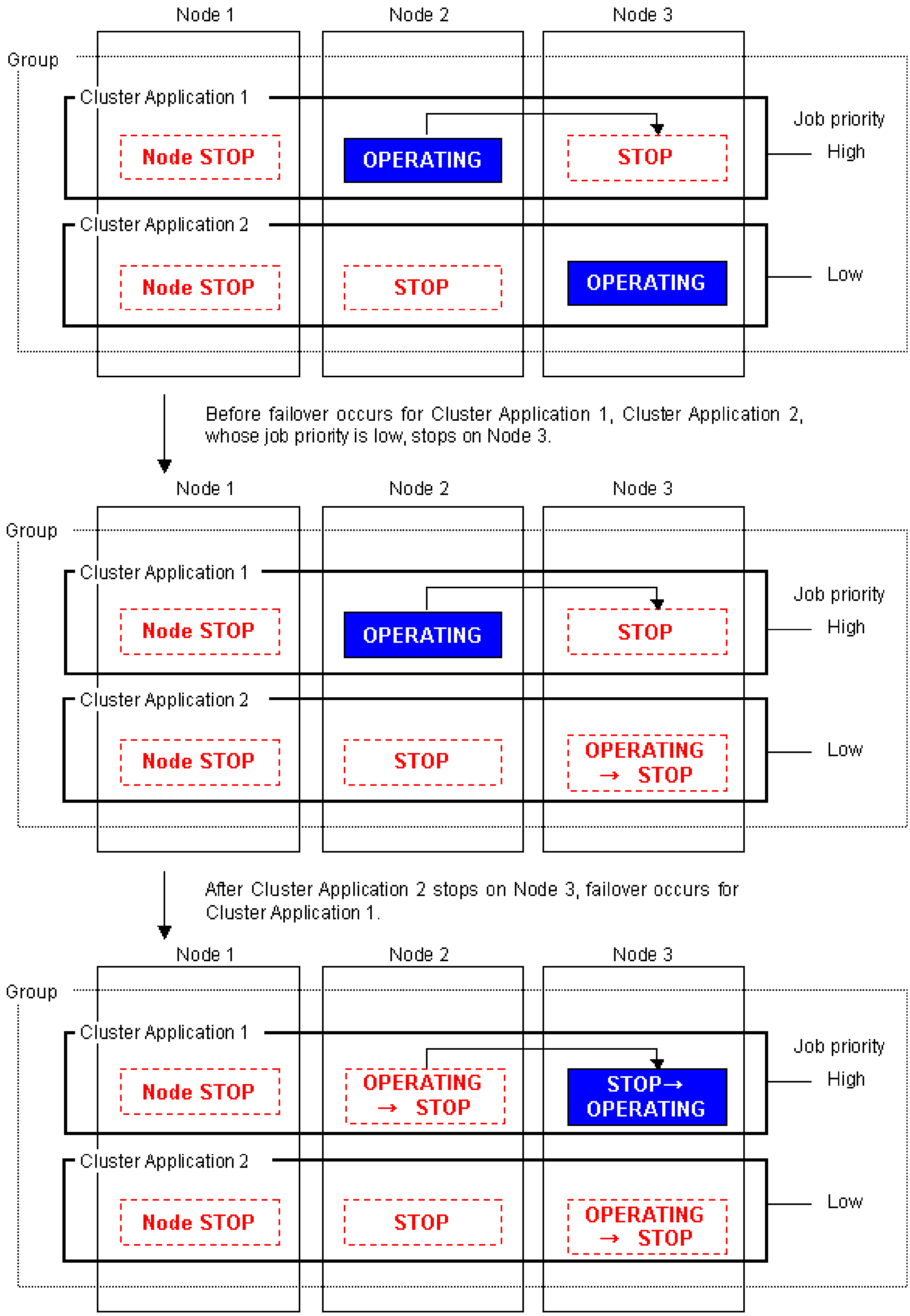
Before failover occurs for Cluster Application 1 whose job priority is high, Cluster Application 2, whose job priority is low, stops.



After Cluster Application 2 stops on Node 2, failover occurs for Cluster Application 1 whose job priority is high.

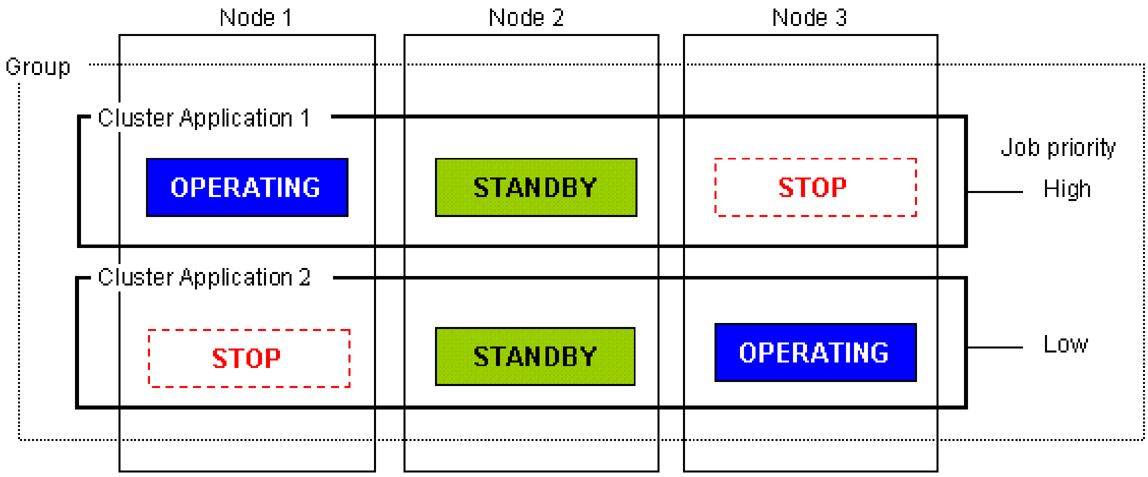


Moreover if Node 2 using Cluster Application 1 stops, failover is attempted to Node 3. However, Cluster Application 2 is OPERATING on Node 3.

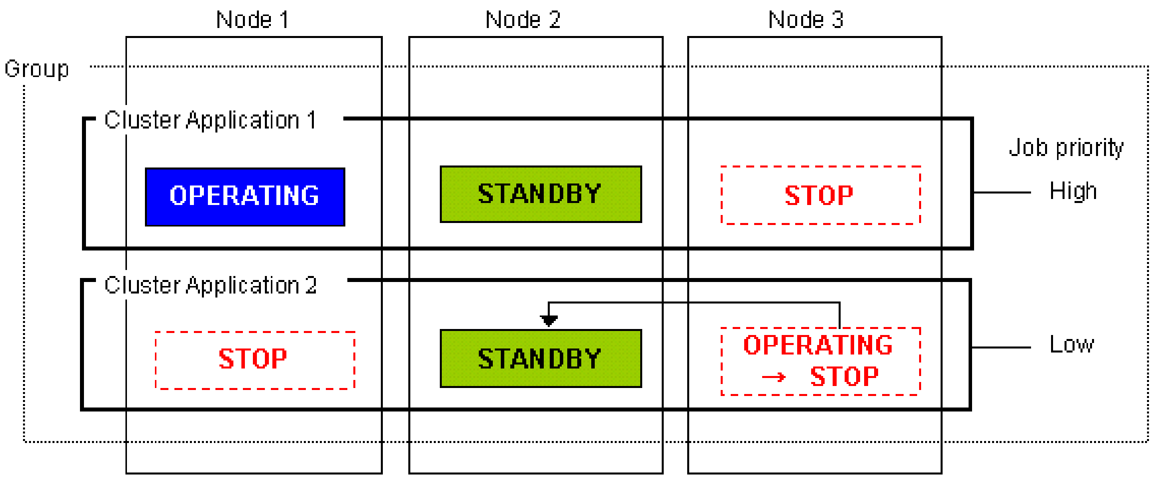


Failover of the cluster application with a low job priority

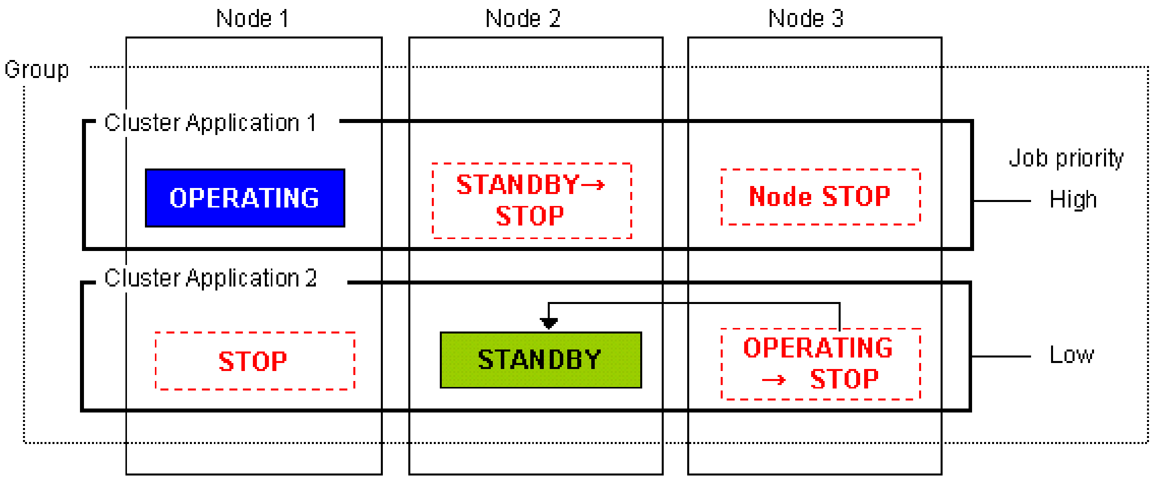
Failover occurs for a cluster application with a low job priority only when there is no cluster application with a high job priority included on the node to which the cluster application with a low job priority is to be failed over.



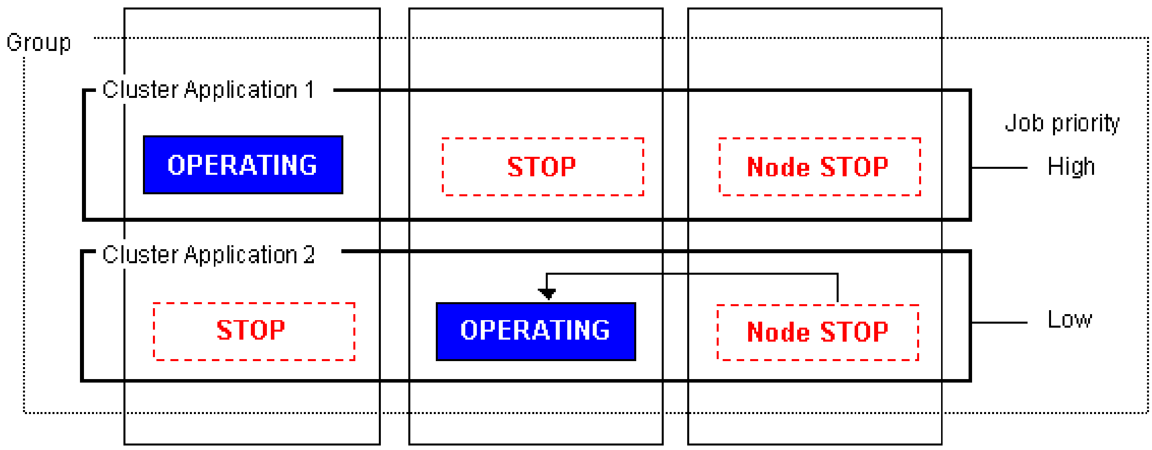
If OPERATING Node 3 using Cluster Application 2 stops, failover is attempted to STANDBY Node 2. However, Cluster Application 1 with high job priority is STANDBY on Node 2.



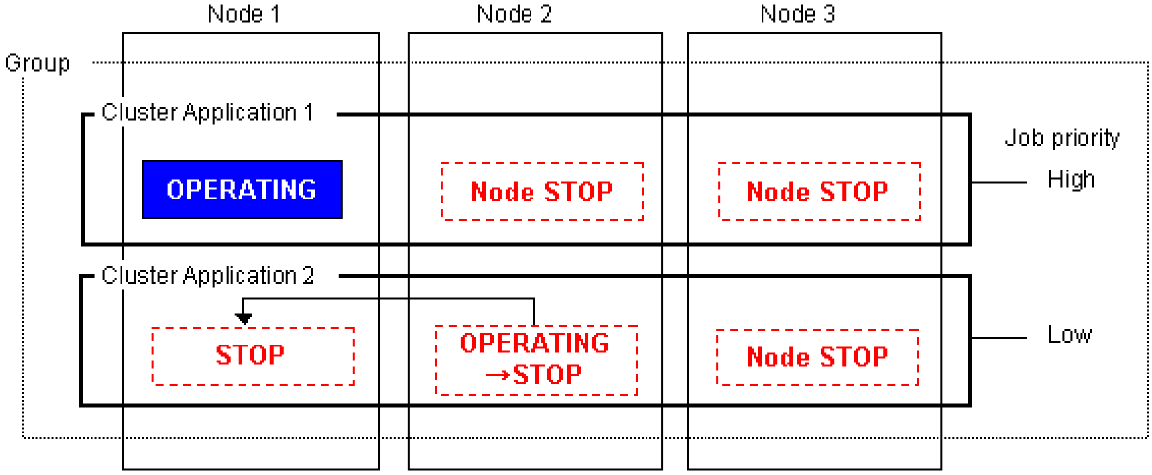
Before failover occurs for Cluster Application 2 whose job priority is low, Cluster Application 1, whose job priority is high, stops on Node 2.



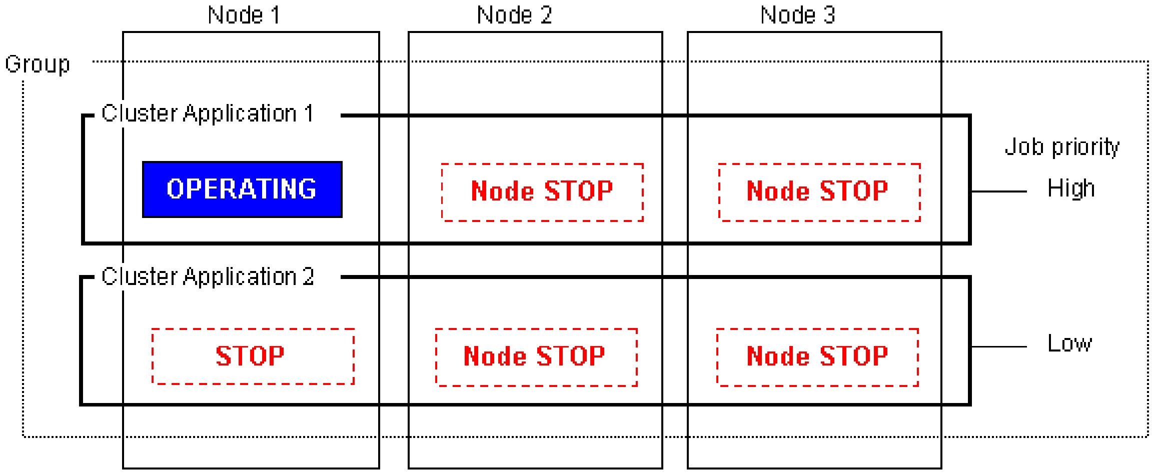
After Cluster Application 1 stops on Node 2, failover occurs for Cluster Application 2 whose job priority is low.



Moreover if Node 2 using Cluster Application 2 stops, failover is attempted to Node 1. However, Cluster Application 1, whose job priority is high, is OPERATING on Node 1.

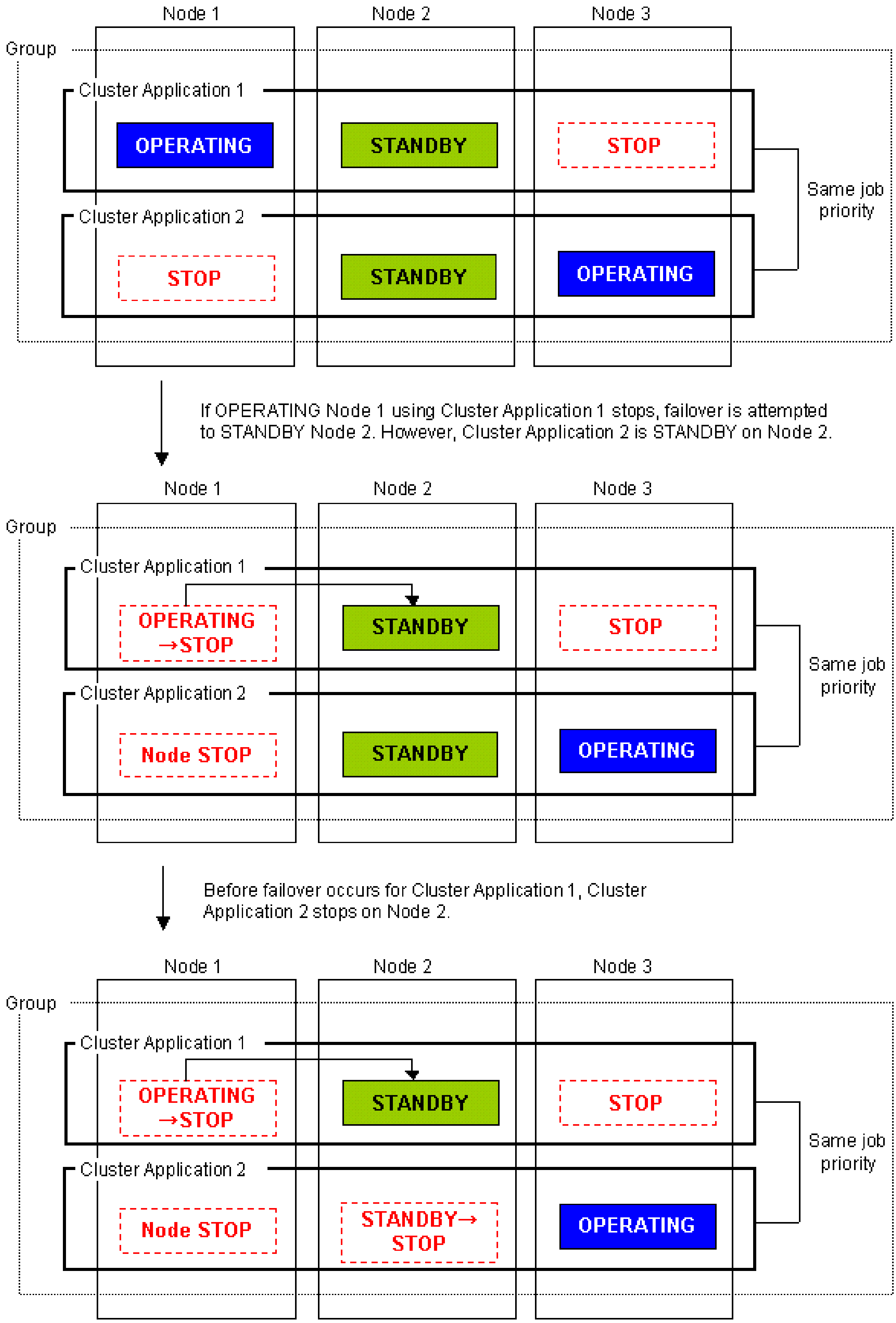


Since Cluster Application 1, whose job priority is high, is OPERATING on Node 1,Cluster Application 2, whose job priority is low, does not execute failover.

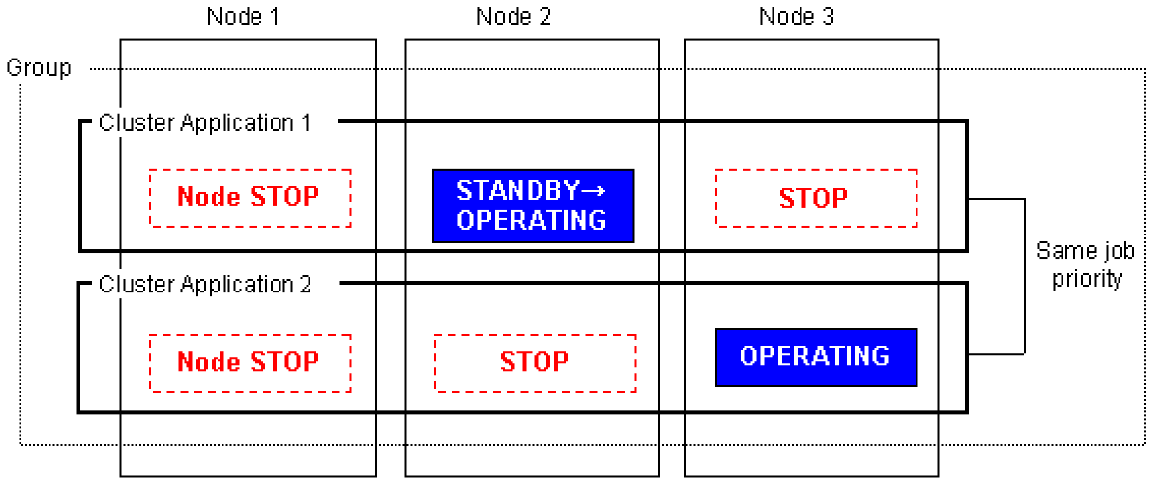


When the job priorities of cluster applications with an exclusive relationship are the same

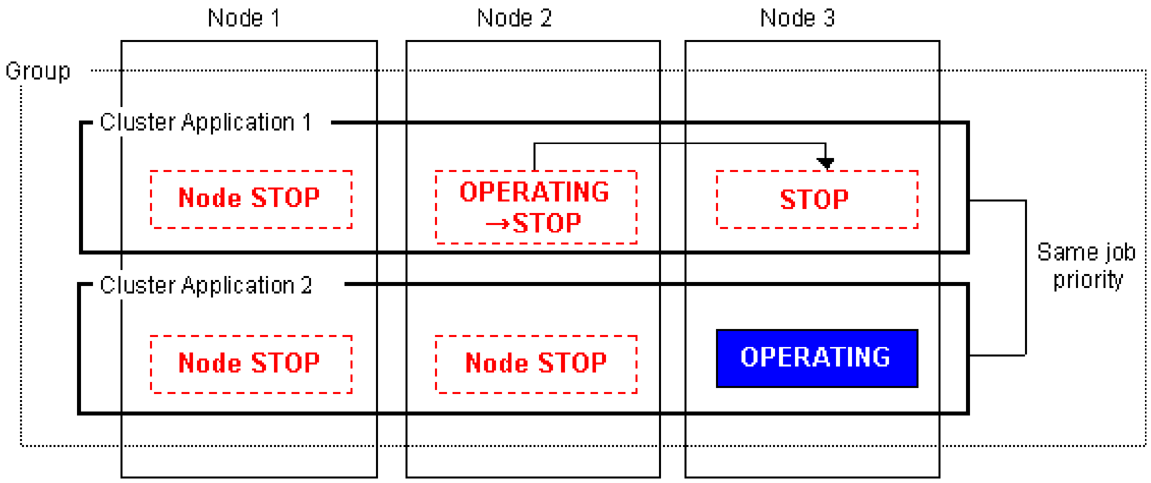
Ongoing operations of the cluster application are never disrupted. If other cluster applications are started on the node where the cluster application has been running, or they are switched to that node, the "hvswitch(1M) -p" is executed. See the manual page of hvswitch.



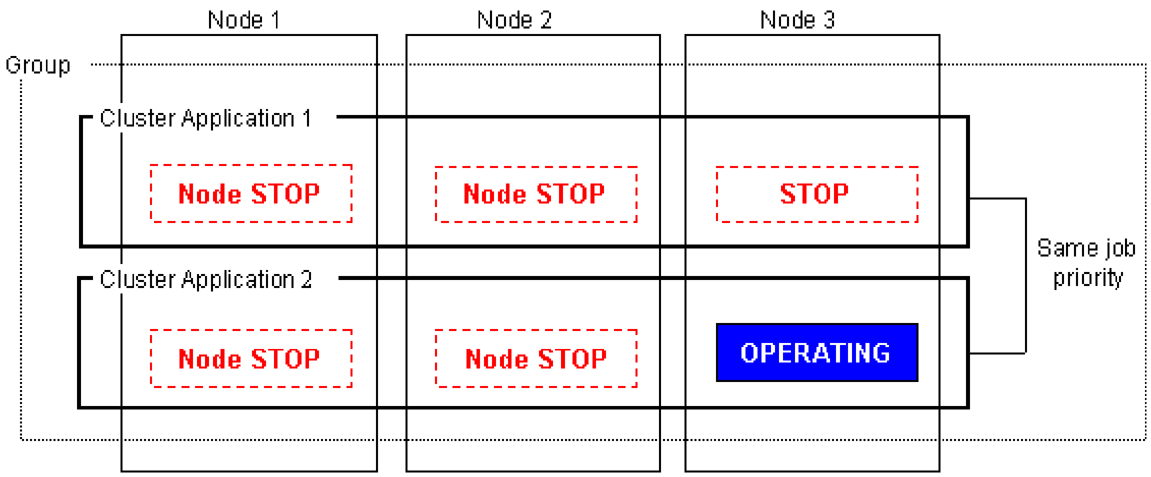
After Cluster Application 2 stops on Node 2, failover occurs for Cluster Application 1.



Moreover if Node 2 using Cluster Application 1 stops, failover is attempted to Node 3. However, Cluster Application 2 is OPERATING on Node 3.



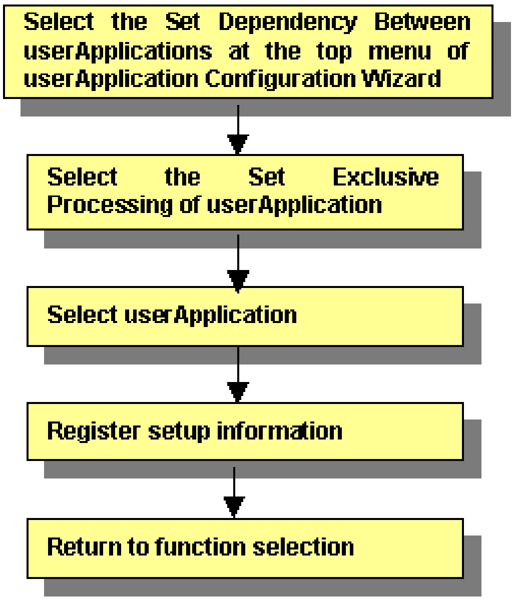
Cluster Application 2 continues to run, and Cluster Application 1 does not execute failover.



Setting up an exclusive relationship

This section explains how to set up an exclusive relationship.

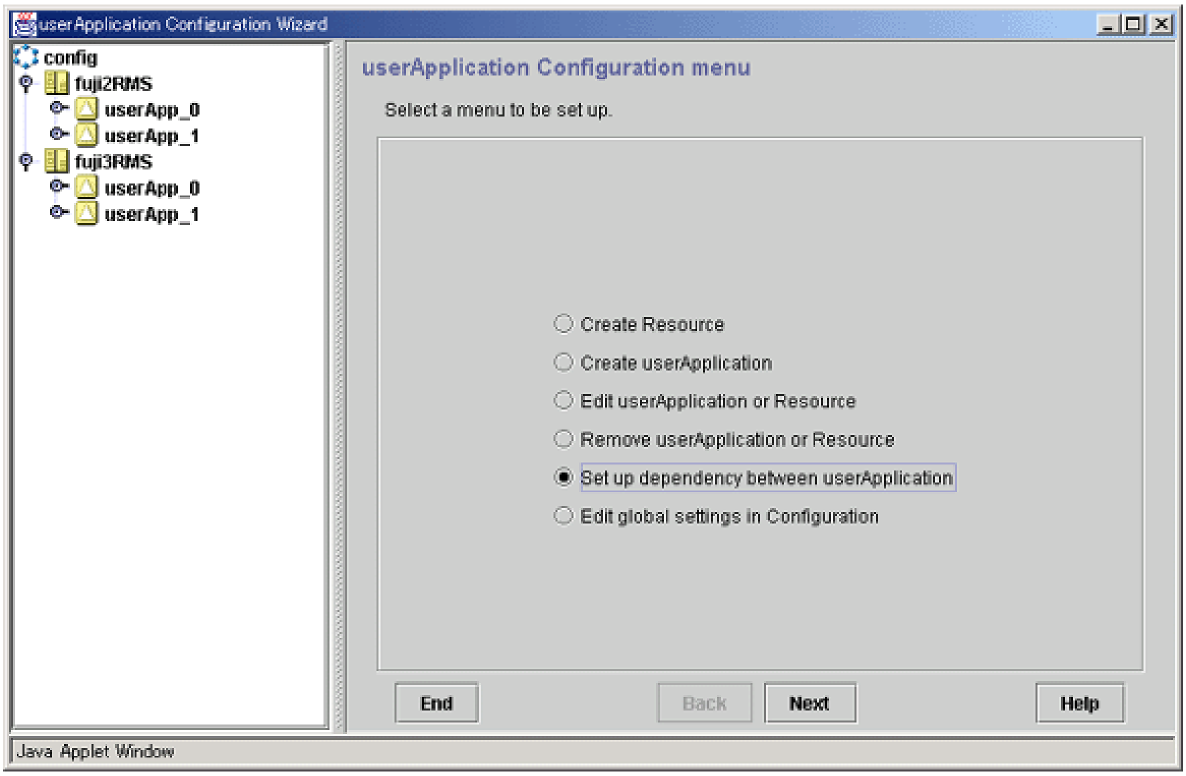
Figure 6.63 Flow for setting up an exclusive relationship



Setting up dependencies between userApplications

From the top menu of the "userApplication Configuration Wizard", select Set up dependency between userApplication.

Figure 6.64 userApplication Configuration Wizard top menu



Click Next to go to the "Set up startup priority and exclusivity of userApplication" screen.

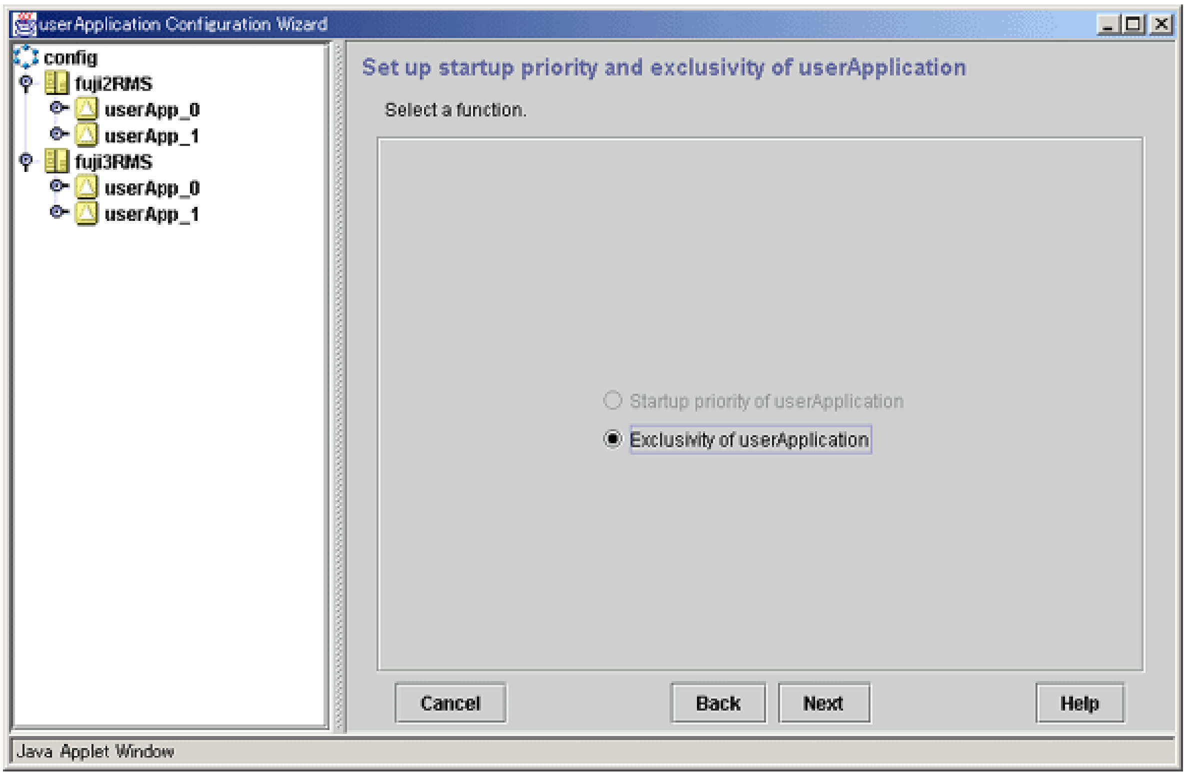
Note

The "Set up dependency between userApplication" menu can be selected only when there are two or more standby cluster applications in the cluster system.

Selecting Exclusivity of userApplication

Select Exclusivity of userApplication.

Figure 6.65 Setting up the startup priority and exclusivity of userApplication

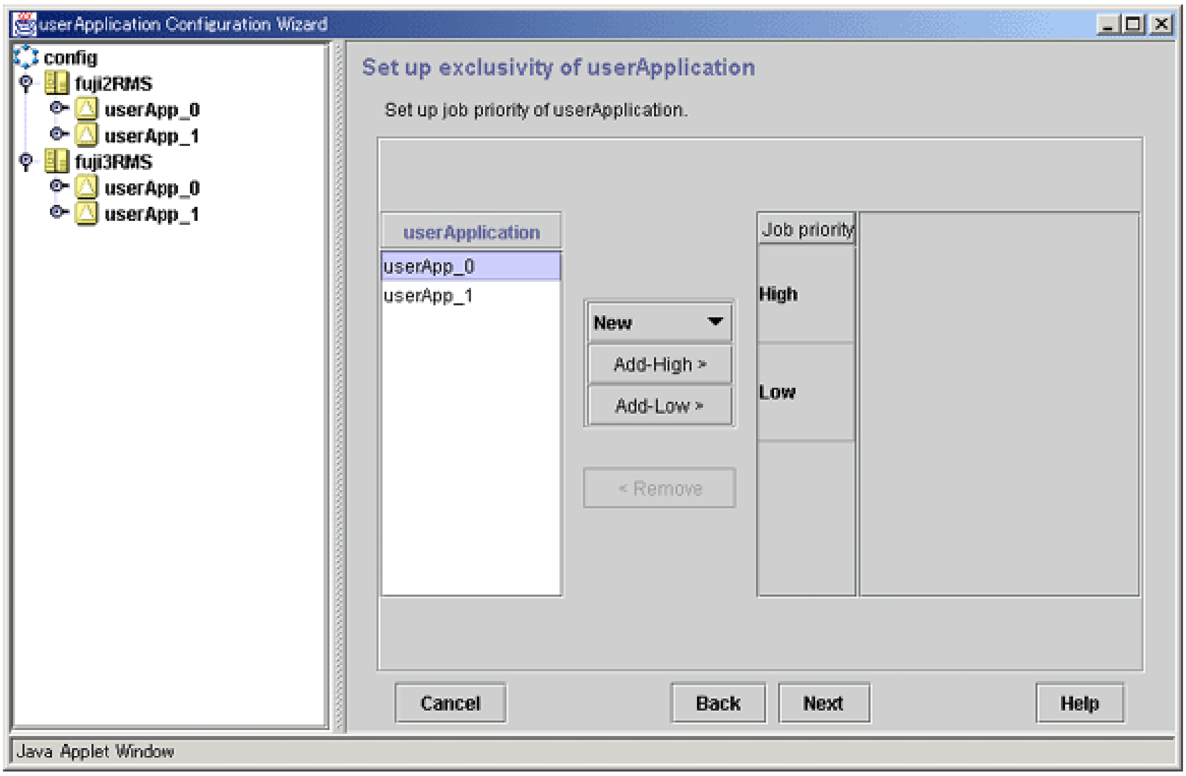


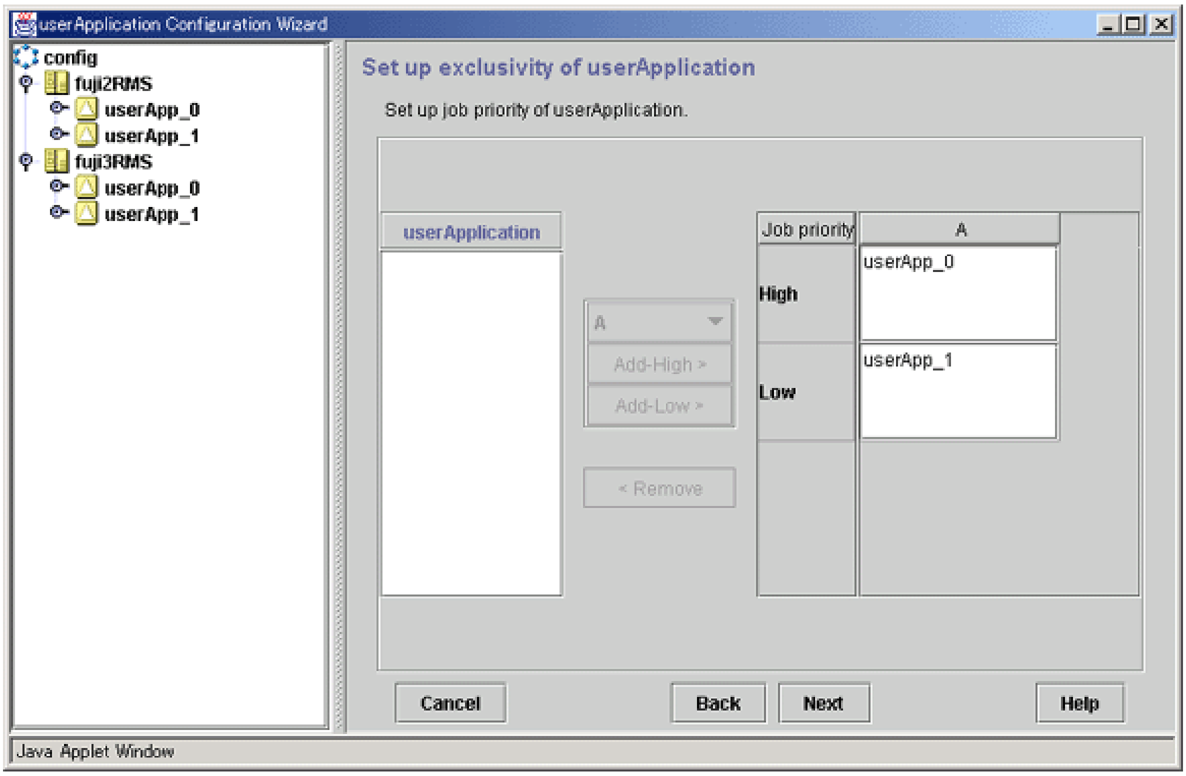
Click Next to go to the "Set up exclusivity of userApplication" screen.

Setting up exclusivity between userApplications

Create a group of cluster applications between which an exclusive relationship is to be set and then add cluster applications to the group.

Figure 6.66 Setting up exclusivity between userApplications





userApplication

Cluster applications are listed. A cluster application that has already been set up in either a group or a scalable cluster application will not be displayed.

Group selection box

From the list box in the center of the setup screen, select a group for which cluster applications are to be set up. If "New" is specified, a new group will be created. Up to 52 groups of A to Z and a to z can be created.
If a group has not been created, you can select "New" only.

***Add-High/Add-Low* buttons**

Add the cluster application that you have selected from the userApplication list to the selected group. If you make this addition by using Add-High, the job priority will be set to "high" while, if you make the addition by using Add-Low, the job priority will be set to "low." Select the addition destination group from the group selection box in advance.

***Delete* button**

Used to cancel the exclusivity setting made for cluster applications. Select cluster applications within a group, and then click Delete. You can return the cluster applications to the [useApplication] list.

After you have made this setting, click Next to go to the "Registration of the exclusive information on userApplication" screen.

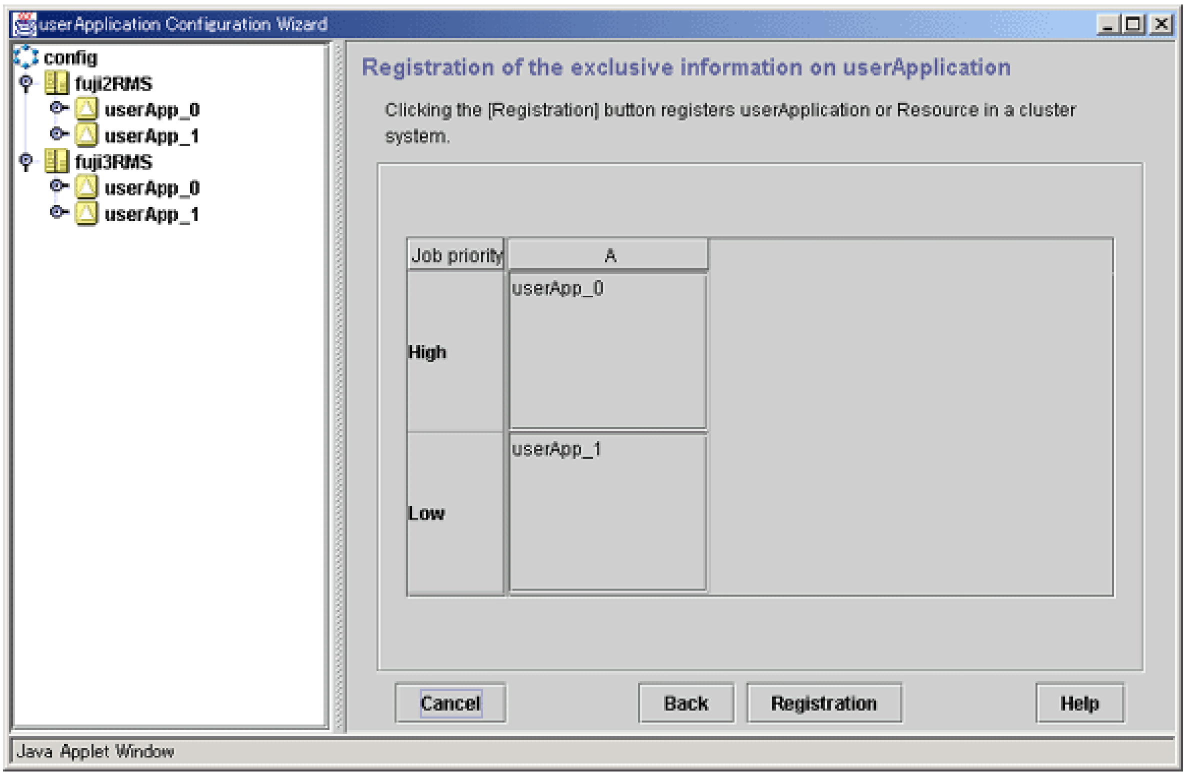
Information

If the job priorities of cluster applications are the same, the same operation will continue even when you add the cluster applications to "High" or "Low."

Confirming the registration information of exclusivity between cluster applications

Confirm the setup information for the exclusive relationship between cluster applications.

Figure 6.67 Registration of the exclusive information on userApplication



Setup information will be displayed.

After you have completed the registration information, click Registration.

6.7.4 Editing global settings in Configuration

Set global settings in the configuration for the RMS configuration.

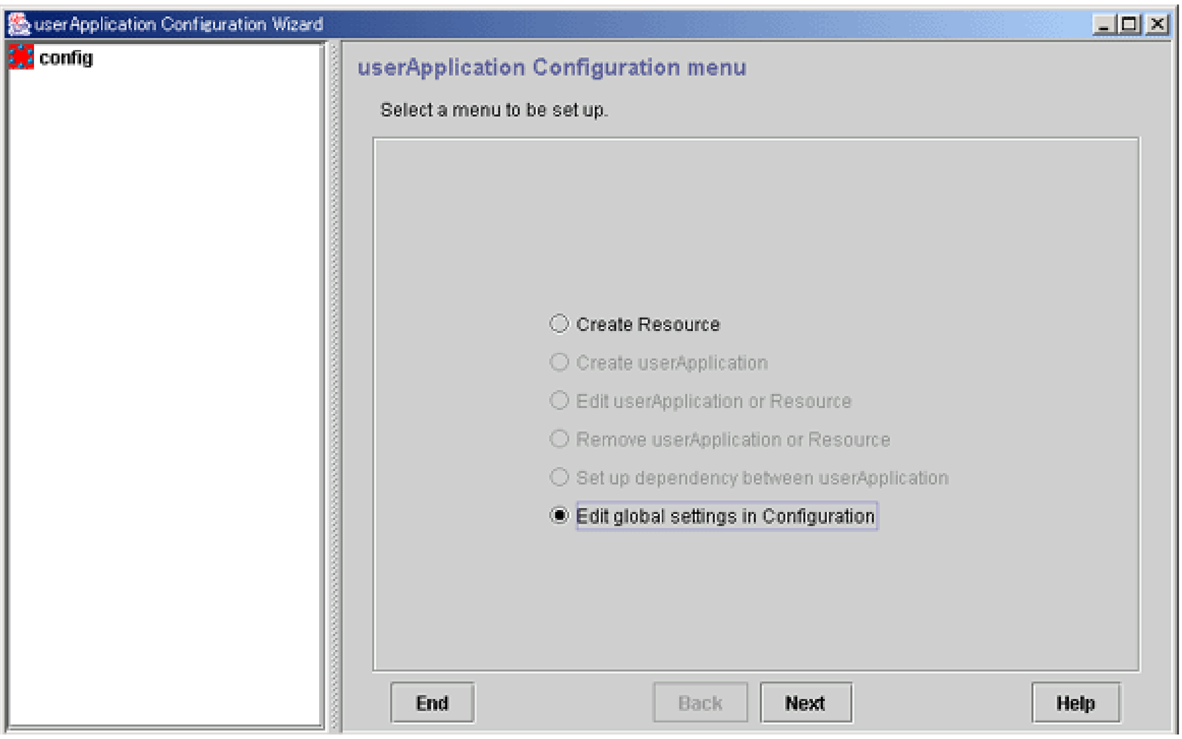
This function is not used except to change the timeout value of PreCheck or to use NFS Lock Failover.

The procedure for changing NFS Lock Failover is described below.

Starting setup of global settings in the configuration

At the top menu of the "userApplication Configuration Wizard," select *Edit global settings in Configuration*.

Figure 6.68 Top menu of userApplication Configuration Wizard

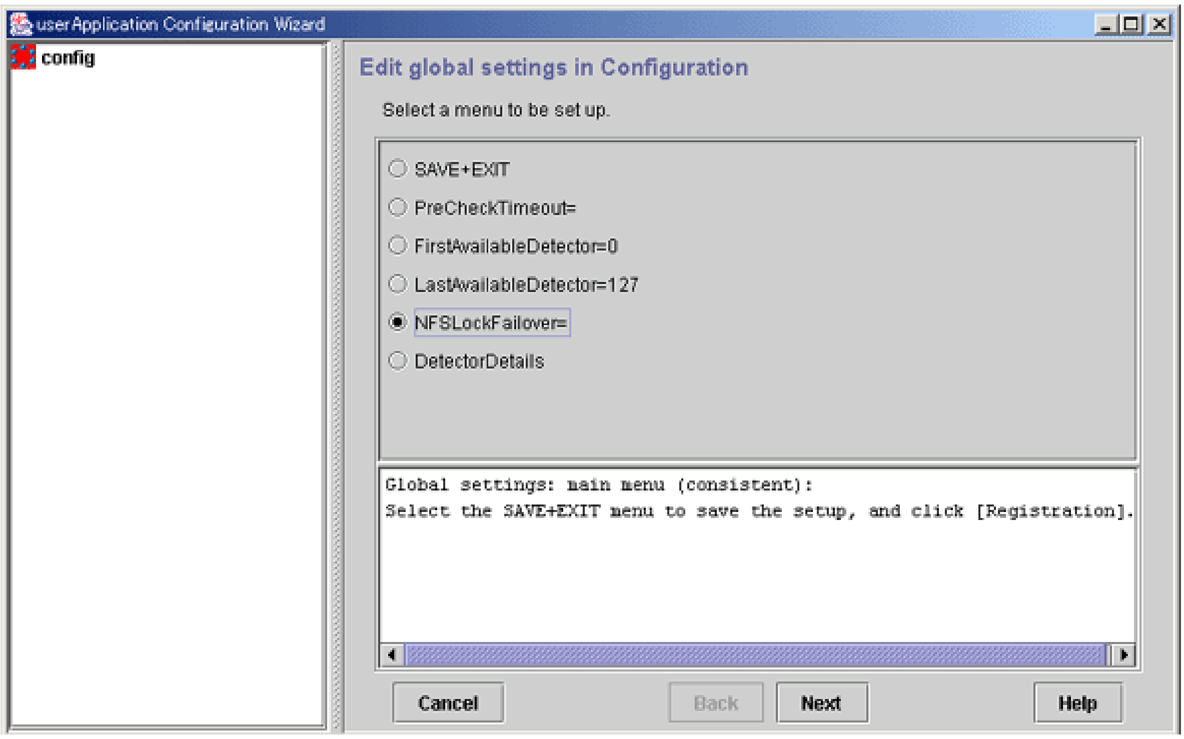


Click *Next* to go to the "Edit global settings in Configuration" screen.

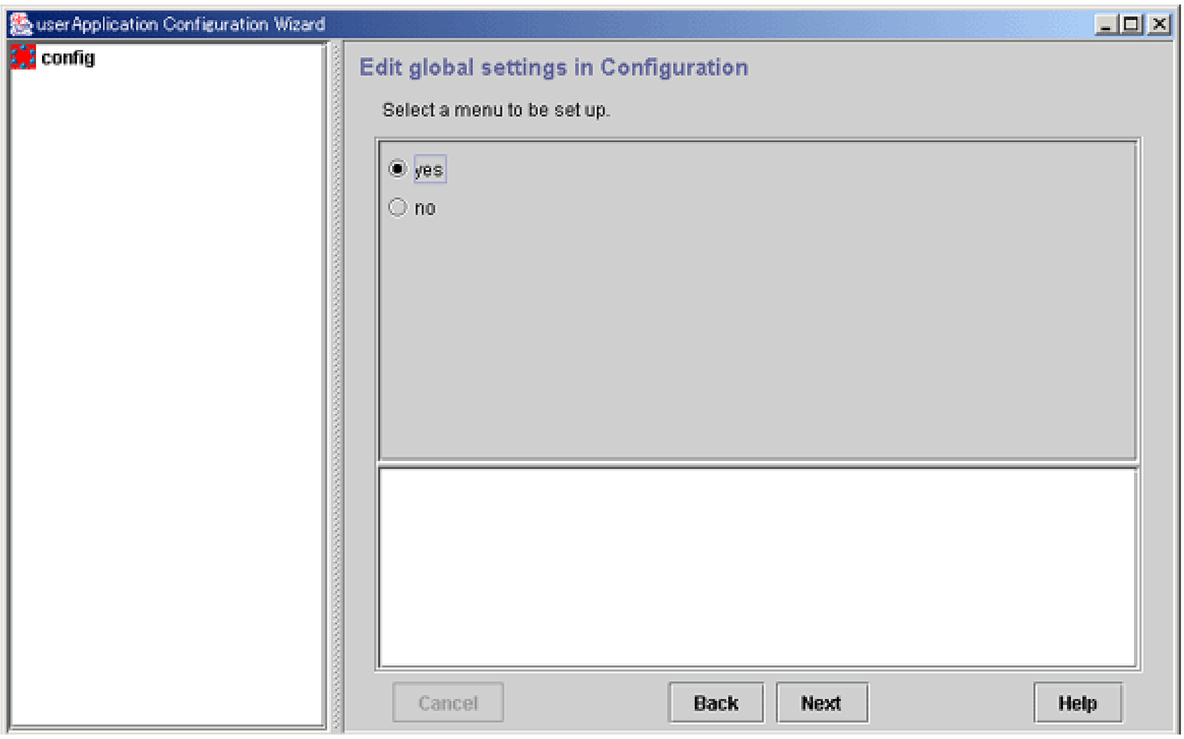
Setting up global settings in the configuration

Set NFS Lock Failover to "yes."

Figure 6.69 Edit global settings in Configuration



Select *NFS Lock Failover* and click *Next*.

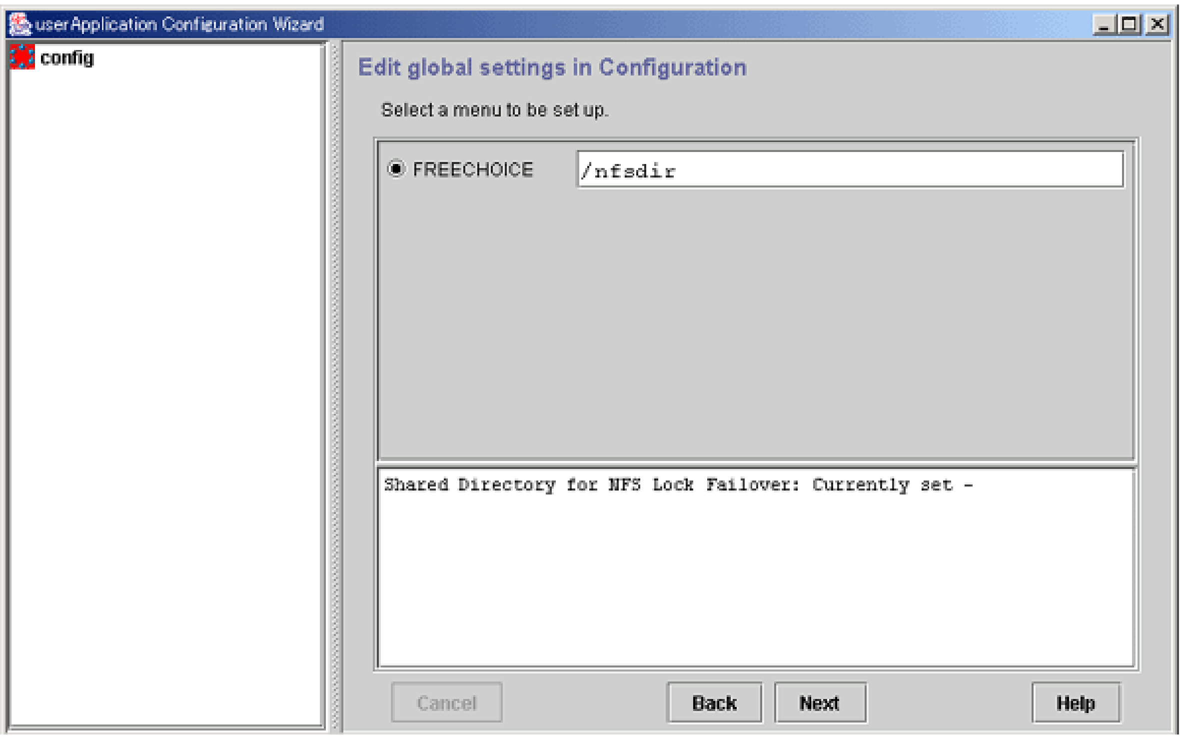


Select *yes* and click *Next* to go to the "Edit global settings in Configuration" screen.

Setting a common directory

Set a common directory.

Figure 6.70 Common directory setup



In *FREECHOICE*, enter the directory for storing NFS file lock information.
During Fsystem resource creation, the specified directory is created automatically in the file system in which the NFSLockFailover attribute was set (under the mount point).

Note

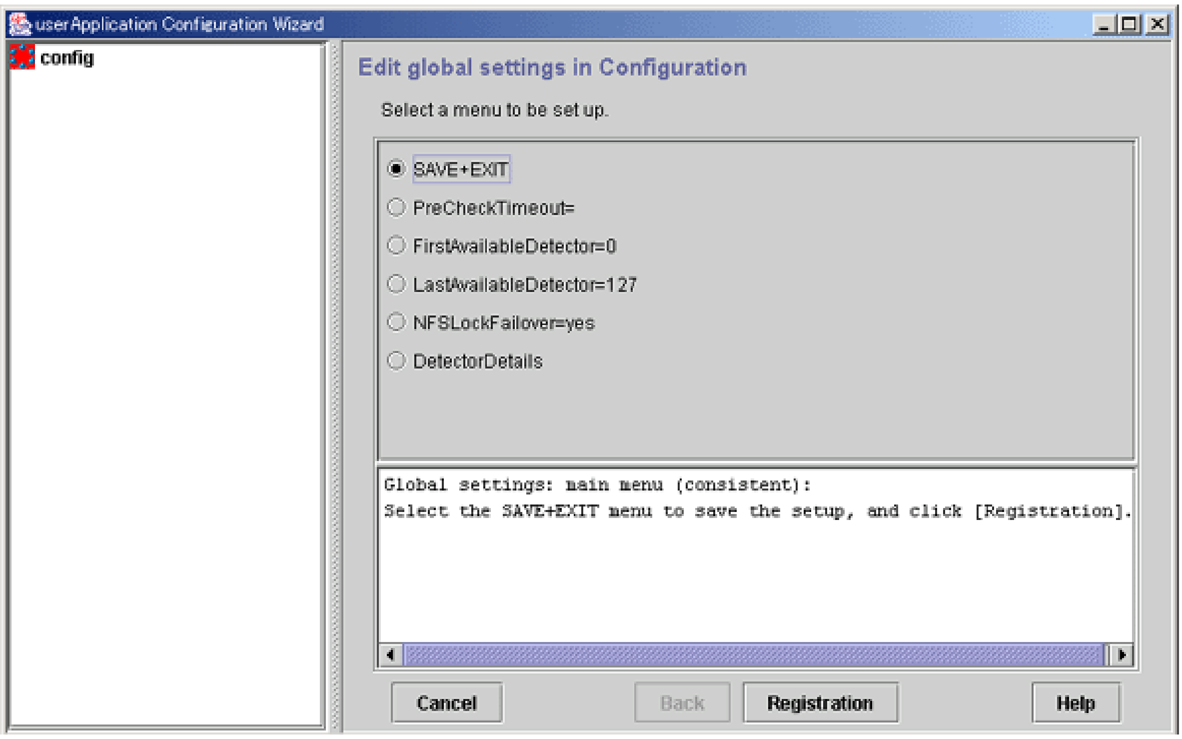
NFS file lock information is stored in this directory. This directory or any directory or file found in this directory should not be removed.

After completing the setup, click *Next* to go to the Top of "Edit global settings in Configuration" screen.

Checking the setup information

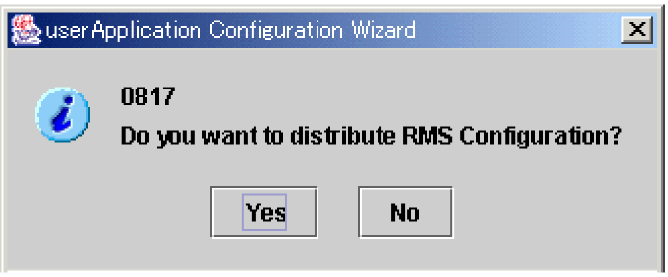
Check that NFS Lock Failover is set to *yes*.

Figure 6.71 Top of global settings in Configuration



After completing the setup, select *SAVE+EXIT*, and then click the *Registration* button.

If there is one or more cluster applications and there is no inconsistency in RMS Configuration, the following message is displayed:



To distribute the Configuration information immediately, click *Yes* to this message.
If you still need to create or change a cluster application or resource, select *No*.

Note

Make sure that you thoroughly verify and understand the operation of the Detector before changing *DetectorDetails*. If you are changing the settings for the first time, you need to follow the procedure described below because no default values have been set.

1. Select DetectorDetails and click the *Next* button.
2. Select "Return to Default Values" and click the *Next* button.
3. Click the *Back* button.

6.7.5 Attributes

Resource attributes

Attribute	Outline
InParallel	Normally, in Online or Offline processing, the programs for starting or stopping all the monitored programs run sequentially. If this attribute is set to yes, the programs run simultaneously. The default value is No (the programs are processed sequentially).
NeedAll	If this attribute is set to No, Online processing succeeds if there's at least one monitored Online program. If the attribute is set to "Yes", all defined monitoring programs must be Online to bring userApplication Online. The default value differs according to the resource. Cmdline: No Fsystem: Yes Takeover network: Yes Process monitoring: No
Timeout SCRIPTTIMEOUT	Sets the timeout duration (seconds) for resource start and stop processing. The default value differs according to the resource. Cmdline: 300 Fsystem: 180 Gds: 1800 Gls: 60 Takeover network: 60 Procedure: 1800
AutoRecover	If this attribute is set to Yes, RMS attempts to recover the faulted resources to prevent userApplication from switching to other host. This attempt is executed only once. If this attribute is set to No, userApplication switches to other host in the event of a failure according to the AutoSwitchOver setting. The default value for Gds resources is "No."

userApplication attributes

Attribute	Outline
AutoStartUp	If this attribute is set to "Yes," the userApplication starts automatically when RMS is started. In Hot-standby operation, a standby userApplication becomes Standby. The default value is "No," and automatic userApplication startup is disabled.
AutoSwitchOver	If a failure occurs in a userApplication, this attribute allows userApplication to be switched over automatically. The switchover condition can be selected from the following: No Automatic switchover by RMS is not executed. The cluster application must be switched over manually by the administrator. HostFailure If a failure occurs in the host (SysNode). ResourceFailure If a failure occurs in a resource. ShutDown If RMS shuts down.

Attribute	Outline
	Combinations of the above values can be selected. Example) HostFailure|ResourceFailure HostFailure|ShutDown ResourceFailure|ShutDown HostFailure|ResourceFailure|ShutDown The default value is "No."
PersistentFault	If this attribute is set to 1, the Faulted state of userApplication remains the same even after RMS restart.
OnlinePriority	If this attribute is set to 1, when all clusters are stopped and restarted, RMS starts userApplication in the host that was Online last. The default value is "0," and after RMS is restarted, userApplication goes Online in the node with the highest priority. The priority is determined by the node setup sequence that was specified in "SysNode Setup" when userApplication was created.
ShutdownPriority	This attribute is a weight coefficient that is assigned to userApplication. If a failure occurs in a cluster interconnect, this attribute also becomes one of the factors in determining which node has the highest priority to survive. The higher value, the higher is the priority of that userApplication. Select a value from the range of 1 to 20. The default value is "NONE," which means no weight coefficient is assigned. The node in which this userApplication is Online may be stopped because it does not contain elements that have a high priority to survive. For details on ShutdownPriority, see "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."
StandbyTransitions	Set up the condition whereby a standby resource will make the transition to the Standby state. You can select any of the following for the setup value: NONE Select this when you do not wish to set a condition for Standby state transition. ClearFaultRequest Select this when you want to clear the userApplication Faulted state and make a state transition to Offline. StartUp If AutoStartUp is set to "No," the operating userApplication will be also changed to Standby state when you start up RMS. SwitchRequest The userApplication is switched to another node. You can select the following values in combination with the above values: Example) ClearFaultRequest|StartUp ClearFaultRequest|SwitchRequest StartUp|SwitchRequest ClearFaultRequest|StartUp|SwitchRequest The default is "NONE." Note) When you use the redundant line control function of GLS, set up the following value to enable monitoring of Gls resource state on the standby node:

Attribute	Outline
	"ClearFaultRequest|StartUp|SwitchRequest"
HaltFlag	If this attribute is set to "Yes," the node is forcibly shut down when another failure occurs (double failure) during fault processing. In this case, RMS forcibly stops the node and completes failover of userApplication even if a resource failure occurs during failover processing. The default value is "No," and node elimination is not executed. Accordingly, userApplication does not execute failover. For details on HaltFlag (Halt attribute), see "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

6.7.6 Changing the RMS Configuration Name

userApplication Configuration Wizard provides a command for changing the Configuration name ("config") to be used as the default.

You can change the RMS Configuration name to be used by userApplication Configuration Wizard by executing this command in the management server of Web-Based Admin View. For detailed usage instructions, see below.

wgcnfclient-RMS Configuration name setup/reference command

Function explanation

This command allows you to change the RMS Configuration name to be displayed and operated by userApplication Configuration Wizard. As much as possible, the modified information is distributed to all clusters. Normally, the RMS Configuration name does not need to be changed.

Prerequisites:

To enable this command, the following conditions must be satisfied:

- The cluster resource manager package must be installed.

  If this package is not installed, the 3-tier configuration of Web-Based Admin View is assumed, and only the information in the local node is changed. If the 3-tier configuration of Web-Based Admin View is being used, this command must also be executed in the secondary management server of Web-Based Admin View.

- Resource database installation must be complete.

  Open the Cluster Admin CRM main window, then check all the node icons are displayed and green on the CRM tree view.

Usage examples

1. Changing the RMS Configuration name

   As shown below, specify the new RMS Configuration name in the "-o" option and execute the command.

```
# /etc/opt/FJSVwvucw/bin/wgcnfclient -s config-name -o config2
#
```

2. Inquiring the RMS Configuration name

```
# /etc/opt/FJSVwvucw/bin/wgcnfclient -g config-name
config2
#
```

   Since no value is set as the initial value of config-name, no data is displayed until you allocate one.

Supplement

If this command terminates abnormally, a system command error message may be displayed. If an error message is displayed, respond according to that message.

Note

The RMS Configuration name must be 80 character or less. If the RMS Configuration name has 81 or more characters, RMS Configuration generation and RMS activation will fail.

6.8 Setting Up the RMS Environment

When using RMS, you need to check the "A.7.1 RMS Setup Worksheet" and change the following environment variable to the value corresponding to the configuration setup.

- RMS shutdown wait time (RELIANT_SHUT_MIN_WAIT)

  The default value of the environment variable is 2147483647 (seconds) in this version.

See

For information on how to check and change the RMS environment variables, see "2.9 Environment variables" and "13 Appendix - Environment variables" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

6.9 Setting Up Patrol Diagnosis

The patrol diagnosis facility periodically diagnoses the following hardware units that are connected to the STANDBY node.

- Shared disk units

  The function diagnoses whether a shared disk unit has become unusable because the power is switched off, a cable is disconnected (adapter side or device side) or because of some other reason.

  If the diagnosis results indicate that an error was detected in a shared disk unit, a message is output to "syslogd(1M)" and the CRM main window.

- Network interface cards

  The function diagnoses whether any network interface card cannot communicate because a cable is disconnected or because of some other reason.

  If the patrol diagnosis results indicate that an error was detected in a network interface card, a message is output to "syslogd(1M)" and the CRM main window, and brings userApplication Faulted.

  If a cluster application uses a Global Link Services (GLS) resource, patrol diagnosis is disabled for the network interface card that is managed by the Gls resource. If a cluster application uses a Gls resource, use the GLS monitoring function. For details on the GLS monitoring function, see "Failure Monitoring Function" and "Standby Patrol Function" in the "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*."

When the Patrol Diagnosis Facility detects a hardware error, this error affects the Patrol Diagnosis Facility resource of a cluster application. This Patrol Diagnosis Facility resource is registered when the cluster application is created and a user enables the Patrol Diagnosis.

Set up patrol diagnosis according to the following flow:

1. Register the patrol diagnosis resource to userApplication.
2. Set up the patrol diagnosis configuration.

Note

- Set up patrol diagnosis after initial setup of the cluster resource manager is completed. If the patrol diagnosis setup command (clspconfig) is executed before initial setup of the cluster resource manager, the following message is displayed:

  ```
  6615: The cluster configuration management facility is not running.
  ```

- If you do not configure the patrol diagnosis configuration, the patrol diagnosis facility will not diagnose the hardware.
- To view the manual pages of the individual commands, add "/etc/opt/FJSVcluster/man" to the MANPATH variable.
- When specifying a shared disk unit as the hardware for the patrol diagnosis, set up the physical disk name (such as c1t4d0) of a shared disk unit to be the same in all nodes. When the physical disk name of a shared disk unit varies depending on a node, you cannot set a shared disk unit to the hardware for the patrol diagnosis.

Registering a patrol diagnosis resource to a userApplication

To register a patrol diagnosis resource to userApplication, use the userApplication Configuration Wizard (GUI) or the RMS Wizard (CUI).

For instructions on resource registration with the userApplication Configuration Wizard (GUI), see "6.7.2 Creating Cluster Applications."

For instructions on resource registration with the RMS Wizard (CUI), see "/usr/opt/reliant/htdocs.solaris/wizards.en/wizards/CRM.htm."

Setting the patrol diagnosis configuration

Before starting RMS, set up the following configuration for patrol diagnosis by executing the "clspconfig(1M)" command on an arbitrary node.

For details on the "clspconfig(1M)" command, see the manual page for "clspconfig(1M)."

Specifying the target hardware for patrol diagnosis execution

Specify the hardware product that runs patrol diagnosis if either of the following conditions applies. This hardware is used by the specified userApplication.

- If the cluster application uses a Gds (Global Disk Services) resource
- If the cluster application uses an Ipaddress (takeover network) resource

Operation Procedure: (When the cluster application uses Gds resources)

If the cluster application is using Gds resources, specify the physical disks that were defined to the Gds resources as hardware for patrol diagnosis execution. The physical disks can be obtained by the procedure below.

1. Confirmation of the disk class

   Check the disk class to be used by the cluster application. The disk class is the class that was specified in "6.7.1.3 Creating Gds Resources."

   Example) **class0001**

2. Confirmation of the physical disks

   Check the physical disks that were defined in the disk class. The physical disks are those that were defined in "6.3.2.3 Setting Up Shared Disks." Use the class configuration setup screen or the "sdxinfo(1M)" command to check the physical disks. For information on the class configuration setup screen, see "6.3.2.3 Setting Up Shared Disks." For details on the "sdxinfo(1M)" command, see the "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

   Example) **c1t4d0, mplb2048**

3. Setup of the target hardware for patrol diagnosis execution

   Set the physical disks that were defined in the disk class as hardware for patrol diagnosis execution.

   Example 1) For c1t4d0 and c1t4d1

   ```
   # /etc/opt/FJSVcluster/bin/clspconfig -u app1 TARGET_DISK=c1t4d0 TARGET_DISK=c1t4d1
   ```

   Example 2) For mplb2048

   ```
   # /etc/opt/FJSVcluster/bin/clspconfig -u app1 TARGET_DISK=mplb2048
   ```

Operation Procedure: (When the cluster application uses Ipaddress resources)

If the cluster application is using Ipaddress resources, specify the network interface card to be used in the takeover network as hardware for patrol diagnosis execution. Follow the procedure below to determine the network interface card.

1. Confirmation of the network interface card

   Check the network interface card to be used by the takeover network. The network interface card is the interface card that was selected in "Selecting an interface" of "6.7.1.5 Creating Takeover Network Resources."

   Example) hme0

2. Setup of target hardware for patrol diagnosis execution

   Configure the network interface card to be used by the takeover network as hardware for patrol diagnosis execution.

   Example)

   ```
   # /etc/opt/FJSVcluster/bin/clspconfig -u app1 TARGET_LAN=hme0
   ```

Time interval for patrol diagnosis

Specify the time interval for patrol diagnosis. If this setting is not specified, patrol diagnosis runs at 360-minute intervals.

Example) To run patrol diagnosis at 60-minute intervals

```
# /etc/opt/FJSVcluster/bin/clspconfig -u app1 INTERVAL=60
```

You can also check the current setting.

Example) Checking the current setting (INTERVAL=360)

```
    # /etc/opt/FJSVcluster/bin/clspconfig -u app1
ACTIVE=true
INTERVAL=360
TARGET_LAN=hme1
TARGET_DISK=c1t4d0
TARGET_DISK=c1t4d1
TARGET_DISK=c1t4d2
TARGET_DISK=c1t4d3
```

Stopping and restarting patrol diagnosis

Stop and restart patrol diagnosis by executing the "clspconfig(1M)" command. For details on the "clspconfig(1M)" command, see the manual page for "clspconfig(1M)."

To stop patrol diagnosis, specify "ACTIVE=false" as an operand in the "clspconfig(1M)" command.

Example)

```
# /etc/opt/FJSVcluster/bin/clspconfig -u app1 ACTIVE=false
```

To restart patrol diagnosis, specify "ACTIVE=true" as an operand in the "clspconfig(1M)" command.

Patrol diagnosis is restarted from the next time with the same configuration as when patrol diagnosis was stopped, including hardware for executing patrol diagnosis and the time interval for patrol diagnosis.

Example)

```
# /etc/opt/FJSVcluster/bin/clspconfig -u app1 ACTIVE=true
```

6.10 Checking the Cluster Environment

After making the settings for the PRIMECLUSTER environment, use the PRIMECLUSTER environment checking tool to make sure there are no mistakes in the settings.

For information on usage of and contents checked by the PRIMECLUSTER environment checking tool, refer to the following file on DVD of the PRIMECLUSTER product package:

`Tool/clchkcluster/README_EN.txt`

6.11 Setting Contents and Notes on Cluster Application

This section describes the setting contents of a cluster application depending on the operation, and notes on its setting.

6.11.1 Setting Contents of a Cluster Application

How to failover a node in the event of a double fault

Perform the following operation:

-> HaltFlag = Yes

-> AutoSwitchOver = HostFailure | ResourceFailure | Shutdown

If "No" has been set to HaltFlag, a failover is not performed even in the event of a double fault. Setting "Yes" to HaltFlag allows the Shutdown Facility to stop the failed node forcibly (PANIC, power discontinuity, and restart) in the event of a double fault. Then, a failover is performed.

Note

Even though the AutoSwitchOver attribute has been set, a failover is not performed unless HaltFlag has been set in the event of a double fault.

How to failover a userApplication in the event of a node failure, resource failure, and RMS stop

Perform the following operation:

-> AutoSwitchOver = HostFailure | ResourceFailure | Shutdown

Note

- In the event of a double fault, a failover is not performed even though this attribute value has been set.
  Set the HaltFlag attribute for performing a failover even in the event of a double fault.
- When the status of the userApplication to be switched is Fault, it cannot be switched even though AutoSwitchOver has been set.
  When performing a failover, clear the Faulted state.

How to start up userApplication automatically when RMS is started

Perform the following operation:

-> AutoStartUp = Yes

If "Yes" has been set to AutoStartUp attribute, the status of a cluster application is automatically transited to Online at RMS startup.

How to switch userApplication to Standby automatically when RMS is started, userApplication is switched, or when clearing a fault state of userApplication

Perform the following operation:

-> StandbyTransitions = Startup | SwitchRequest | ClearFaultRequest

Note

- If "Yes" has been set to AutoStartUp attribute, the status of the standby userApplication is transited to Standby when RMS is started regardless of the setting value of StandbyTransitions.

  The relationship between AutoStartUp and StandbyTransitions is as follows.

RMS Startup node		AutoStartUp=Yes StandbyTransitions		AutoStartUp=No StandbyTransitions	
		No	StartUP	No	StartUP
Multiple nodes	Operational side uap	Online	Online	Offline	Standby
	Standby side uap	Standby	Standby	Offline	Standby
One node only		Standby	Standby	Offline	Standby

If the resource which StandbyCapable attribute is set as "yes"(1) does not exist in the userApplication, the userApplication is not in the Standby state regardless of the set value of StandbyTransitions attribute.

How to set scalable cluster applications for preventing timeout of Controller resource during a state transition

When it takes time to start up and stop a cluster application that constitutes a scalable configuration, a timeout error of the Controller resource (resource to indicate the scalability) may occur during a state transition. In this case, the state transition is stopped forcibly.

In this case, the setting of Controller resource needs to be changed according to the startup and stop times for each cluster application that constitutes a scalable configuration.

Calculate the Timeout value of a scalable cluster application, and then change its setting with the following procedure:

Procedure

1. Calculating the maximum state transition time for a cluster application

   The status of the Controller resource is transited to Online when the statues of userApplications under the Controller resource are all Online. Therefore, calculate the total values of ScriptTimeouts for each resource that configures a cluster application.

   For example, if every one of the following resource; Cmdline resource, Fsystem resource, Gds resource, or Gls resource exist under the cluster application, you can calculate as follows. (The timeout value for each resource is a default value.)

   Cmdline resource 300 (sec) + Fsystem resource 180 (sec) + Gds resource 1800 (sec) + Gls resource 60 (sec) = 2340 (sec)

   This value is larger than the default value for the scalable cluster application 180 (sec), set the setting value to 2340 (sec).

   #### Information

   Default script timeout values for each resource

   Cmdline : 300
   Fsystem : 180
   Gds : 1800
   Gls : 60

2. Considering the number of SysNode

   Calculate the considered number of SysNode that configures a cluster application.

   The number of SysNode is 1

   The value calculated in Step 1 is the value where the number of SysNode is considerate.

   The number of SysNode is 2 or larger

   Minus 1 from the number of SysNode and double the value. Then, multiply it by the one calculated in Step 1.

   ```
   The maximum state transition time of a cluster application between multiple nodes

   = "1) value" x 2 ( "the number of SysNode" -1 )
   ```

Example

For example, in the case Online or Offline processing of a userApplication is assumed to be finished just before it times out when the userApplication is with a three-node configuration and the status is Online on Node1, after starting the state transition on the first Node, it takes 4 times (2 x ("the number of SysNode" - 1) for the userApplication to be Online on the final node as follows:

1. Offline processing on Node1
2. Online processing on Node2
3. Offline processing on Node2
4. Online processing on Node3

3. Calculating the total values of Step 2 for each cluster application
4. Setting the calculated value in Step 3 to the timeout value of the Controller resource

See "5.14.2 Changing timeout period of controller" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide" to change the timeout value of the Controller resource.

How to stop a standby operational system preferentially in the event of a heartbeat error

When a heartbeat error is detected, set the survival priority for the node to be stopped forcibly so that it prevents all operational and standby systems from being failed by forcibly stopping both operational and standby systems mutually. Below describes how to stop the operational system preferentially and collect the information for investigation.

Note

- The weighting of each node to set in the Shutdown Facility is defined to a node.
  If an operational and standby system is switched due to a failover or switchover, it cannot be enabled even though the setting is changed.
  As before, stop an operational system forcibly after a given time has elapsed in a standby system.
  When a cluster is switched, be sure to perform a failback.
- If a system panic, CPU load, or I/O load continues, it seems like a heartbeat has an error. In this case, the cluster node with an error is forcibly stopped regardless of the survival priority.
- A standby system with a low survival priority waits until an operational system is forcibly stopped completely. During this waiting time, if the heartbeat is recovered, some information for investigating the heartbeat error may not be collected.
  This case may occur when the CPU load or I/O load is the high in an operational system.

Procedure

Below indicates an example when the operational system is node1, and the standby system is node2.

Note

Perform the Steps 1 to 4 in the both operational and standby systems.

1. Modify the SF configuration (/etc/opt/SMAW/SMAWsf/rcsd.cfg) for the standby system (node2) with the vi editor, and so on to give a higher weight value to the standby system. Change the weight attribute value of node2 from "1" to "2."

```
node2# vi /etc/opt/SMAW/SMAWsf/rcsd.cfg
```

[Before edit]

```
node1,weight=1,admIP=x.x.x.x:agent=SA_xx,timeout=20:agent=SA_yy:timeout=20
node2,weight=1,admIP=x.x.x.x:agent=SA_xx,timeout=20:agent=SA_yy:timeout=20
```

[After edit]

```
node1,weight=1,admIP=x.x.x.x:agent=SA_xx,timeout=20:agent=SA_yy:timeout=20
node2,weight=2,admIP=x.x.x.x:agent=SA_xx,timeout=20:agent=SA_yy:timeout=20
```

Note

- Describe the setting of one node with one line in the rcsd.cfg file.
- admIP may not be described depending on the version of PRIMECLUSTER.

2. Restart the SF with the sdtool -r command.

   It takes about five seconds to execute the sdtool -r command. After that, the changed SF configuration is reflected to the SF.

```
node2# sdtool -r
```

3. Use the sdtool -C command. to check that the changed SF configuration has been reflected

   Check that the weight attribute value of node2 has become "2."

```
node2# sdtool -C

Cluster Host   Type    Weight  Admin IP        Agent List (Agent:timeout)
------------   -----   ------  --------        --------------------------
node1          CORE    1       x.x.x.x         SA_xx:20,SA_yy:20
node2          CORE    2       x.x.x.x         SA_xx:20,SA_yy:20
```

Note

"Type" may not be displayed depending on the version of PRIMECLUSTER.

4. Use the sdtool -s command to check that all the SAs defined to the SF operate properly. Moreover, check that "Test State" and "Init State" have been changed to "TestWorked" and "InitWorked respectively.

```
node2# sdtool -s
Cluster Host    Agent          SA State      Shut State  Test State  Init State
------------    -----          --------      ----------  ----------  ----------
node1           SA_xx          Idle          Unknown     TestWorked  InitWorked
node1           SA_yy          Idle          Unknown     TestWorked  InitWorked
node2           SA_xx          Idle          Unknown     TestWorked  InitWorked
node2           SA_yy          Idle          Unknown     TestWorked  InitWorked
```

Note

Perform the following Steps 5 to 8 either in the operational or standby system.

5. Check the ShutdownPriority attribute value of a cluster application (userApplication) with hvutil -W command.

   When the ShutdownPriority attribute value is other than "0," perform Steps 6 to 8.

   When it is "0," no more setting is required.

```
node1# hvutil -W
4
```

6. Stop PRIMECLUSTER (RMS).

Note

If you stop PRIMECLUSTER (RMS), the operation is also stopped.

```
node1# hvshut -a
```

7. Change the ShutdownPriority attribute value of a cluster application (userApplication) to "NONE." First, start the RMS Wizard.

   a. At the *Global Cluster Services* screen, select *userApplication Configuration Wizard*.

   b. From the tree on the left of the *userApplication Configuration Wizard* screen, select the userApplication to be changed, right-click the mouse to display the pop-up menu, and select *Edit userApplication or Resource*.

   c. Select "ShutdownPriority."

   d. After the attribute is changed, click the *Registration* button to register it to RMS Configuration.

   e. Click *Yes* when the message "0817 Do you want to distribute RMS Configuration?" is displayed.

See

For details, see "11.1 Changing the Operation Attributes of a Cluster Application."

8. Start PRIMECLUSTER (RMS).

```
node1# hvcm -a
```

Note

When a cluster is switched, be sure to perform a failback.

How to stop the operational node forcibly in the event of a subsystem hang

The following event is called a subsystem hang: the cluster does not detect that the operation is stopped (the operation seems normal from the cluster monitoring) because only some I/Os within the operational node have errors and other I/Os operate normally.

In this case, if the node is switched to a standby node, the operation may be restarted. In the event of a subsystem hang, ping may respond properly and you may be able to log in to a node.

When a subsystem hang is detected, stop the operational node with the following method and switch the operation.

If you can log in to a standby node

Stop the operational node from the standby node with the sdtool command.

```
# sdtool -k node-name
```

`node-name : CF node name of the operational node`

If you cannot log in any node

Set the operational node to panic status. For details on how to set it, see the instruction manual of a main device.

Note

It is possible to determine a subsystem hang from application failures to control a forcible stop mentioned above. In the case, it needs to be determined from multiple clients. That is, even though an error is found from one client, the error may be in the client or on the network. You need to consider such a case when controlling a forcible stop.

How to failover a system when an error occurs the line of LAN

->Set PingHost

PingHost is necessary information to check the response (ping) with the set host. You should specify two or more nodes that are not configured for the cluster system to prevent adverse effects from hub and router failures. If PingHost is not set, even if a LAN error occurs, it does not become a resource error. In this case, the system will not be switched.

How to set the PreCheckScript to a cluster application

For how to set the PreCheckScript to a cluster application, see "5.14.1 Setting PreCheckScript in cluster application" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

6.11.2 Notes on Configuration

Do not use reserved words for userApplication names and Resource names

If you use a reserved word for a userApplication or Resource name, RMS cannot be configured properly.

Do not use the following reserved words in addition to numbers and types of characters limited in this manual.

<List of reserved words>

Reserved words written in C

```
auto|break|case|char|const|continue|

default|do|double|else|enum|extern|float|

for|goto|if|int|long|main|register|return|short|

signed|sizeof|static|struct|switch|typedef|

union|unsigned|void|volatile|while
```

Reserved words written in C++

```
and|and_eq|bitand|bitor|compl|not|or|or_eq|xor|xor_eq|

asm|catch|class|delete|friend|inline|new|operator|private|

protected|public|template|try|this|virtual|throw
```

Reserved words within RMS

```
ADMIN|ADMIN_MODIFY|CONTRACT_MODIFY|ENV|ENVL|INIT_NODE|Offline|

Faulted|Online|Standby|Warning|SysNode|andOp|

assert|commdNode|contractMod|controller|env|envl|gResource|node|

object|orOp|userApp|userApplication|ScalableCtrl

abstract|attach|attribute|begin|class|consume|copy|cpp|declare|

delay|delete|error|extends|extern|hidden|implements|include|

interface|java|left|lookahead|lr|message|modify|nonassoc|node|

nosplit|notree|package|prec|private|public|reductor|repeat|right|

select|show|simple|skip|state|tree|trigger|type|used|virtual|wait|link
```

Register all Resources created by the userApplication Configuration Wizard to userApplication

When setting a cluster application, if Resource that does not belong to userApplication exists, RMS does not run. After registering all created Resources to userApplication, check that the icon of the config does not turn red.

Before using the userApplication Configuration Wizard, calculate the necessary heap memory size

If the size is more than 64 MB after the calculation, take the following actions:

- If the necessary heap memory size is 65 to 1024 MB

  Set the necessary heap memory size to Java VM used in the Web-Based Admin View client.

  However, Web-Based Admin View may not be connected when the heap memory size that was specified at Java VM startup were not secured after specifying the necessary heap memory size. In this case, take the action of "If the necessary heap memory size is more than 1024 MB" to reduce the necessary heap memory size.

  The maximum value of the heap memory size that can be connected to Web-Based Admin View depends on the memory volume that is installed in a client, and also depends on the memory status.

- If the necessary heap memory size is more than 1024 MB

  If multiple resources with the same Resource type exist in a single userApplication, take the following action to reduce the necessary heap memory size to less than 1024 MB. If the resource is not described in the following action, no action is required. If the necessary heap memory size is more than 64 MB after taking the action, set the necessary heap memory size to Java VM that is used by the Web-Based Admin View client.

  - Gds resource

    Register only one Gds resource to a single userApplication.
    When creating a Gds resource on the Select Disk Class screen, add all the disk classes to be registered in a single userApplication from *Available disk class* to create a Gds resource.
    For details, see "6.7.1.3 Creating Gds Resources."

  - Gls resource

    Register only on Gls resource to a single userApplication.
    When creating a Gls resource on the Select takeover IP address screen, add all the takeover IP addresses to be registered in a single userApplication from *Available takeover IP address* to create a Gls resource.

  - Line switching unit resource (unavailable in the Oracle Solaris 11 environment)

    Register only one line switching unit resource to a single userApplication.
    When creating a line switching unit resource on the Set up Resource screen, add all the line switching unit resources to be registered in a single userApplication by adding resources from *AdditionalResource* to create a line switching unit resource.
    For details, see "6.7.1.8 Creating Line Switching Unit Resources."

  - Procedure resource of Netcompo product

    Assign a startup order to all resources for each Netcompo product to be registered in a single userApplication. The order is optional.
    Assign a startup order on the Associate resources screen, This screen is displayed by pressing the *SubApplication* button on the "Confirm registration" screen when creating a resource of Netcompo product.
    Below is the setting example.

    If there are five procedure resources of Netcompo product

Procedure resource name of Netcompo product	Procedure resource name of Netcompo product set to SubApplication
netcompo0	Not set
netcompo1	netcompo0
netcompo2	netcompo1
netcompo3	netcompo2
netcompo4	netcompo3

    For details, see "6.7.1.6 Creating Procedure Resources."

    Note

    When the startup order is assigned to a procedure resource of Netcompo product, add the processing time of each procedure resource of Netcompo product to the startup time of userApplication.

The necessary heap memory size is calculated with the following procedure:

1. Calculate the total number of nodes, applications, and resources displayed on the userApplication Configuration Wizard screen.

    a. Create the following table for each userApplication.
    Input the number of each resource to be embedded to userApplication to "Number of resources" in the table below. If there is no resource, input 0.

Number of resources	Factors of formula	Resource type
		Cmdline Controller Nwcl NetWorker Nwst NetWorker Nwsv NetWorker Procedure(Application) Procedure(BasicApplication) Symfoware Oracle Procedure(SystemState3) Procedure(SystemState2) FNA1_FSUNvcph FNA2_FSUNvcp FNALAN_FSUNlndfc BCLine_FSUNbcdr Ims Patrol SH_SWLine Gls Ipaddress Fsystem Gds

Note

If the startup priority has been set to resources of the same type such as Cmdline or Procedure, the resources become the parent-child relationship. Therefore, modify the table.

Example: when setting the startup priority to the following Procedure(BasicApplication) to be called sequentially from the top

IS_INTERSTAGE
OTS_RMP_INTERSTAGE
ES_INTERSTAGE

Number of resources	Coefficient of formula	Resource type
		Cmdline Controller Nwcl NetWorker Nwst NetWorker Nwsv NetWorker Procedure(Application) Procedure(BasicApplication) <- ES_INTERSTAGE Procedure(BasicApplication) <- OTS_RMP_INTERSTAGE Procedure(BasicApplication) <- IS_INTERSTAGE Symfoware Oracle

Number of resources	Coefficient of formula	Resource type
		Procedure(SystemState3) Procedure(SystemState2) FNA1_FSUNvcph FNA2_FSUNvcp FNALAN_FSUNlndfc BCLine_FSUNbcdr Ims Patrol SH_SWLine Gls Ipaddress Fsystem Gds

b. Delete the line if the number of resources is 0 from the table created in step a.
Next, input the coefficient of formula to each line like R0, R1, R2, ... sequentially from the first line.

c. Calculate the total number of resources to be displayed for each application.
It is calculated using the following formula.

```
The total number of resources for each application = Node * ( 1 + R0 + R0*R1 + R0*R1*R2
+ ... )
```

```
Node         : Number of nodes
R0,R1,R2,... : Number of resources corresponding to the coefficient of formula in the table
```

d. The total number of resources to be displayed are calculated by adding the number of applications to the total number of resources that was calculated for each application.

```
The total number of resources to be displayed = the total number of resources calculated for
each application +
                           Node * The number of applications
```

2. Calculate the necessary memory volume.

Calculate the necessary memory volume from the total number of resources that was calculated in step 1 using the following formula.

```
The necessary memory volume [MB] = 0.111 [MB] * Total number of resources [Number] + 9.0 [MB]
```

Example: when the total number of resources calculated in step 1 is 1620

```
The necessary memory volume [MB] = 0.111 * 1620 + 9.0 = 188.82 [MB]
```

Set the necessary heap memory size for JavaVM of a client using the following procedure:

1) Start the Java Plug-in Control Panel.

2) Add an option to specify memory volume with a method according to the Java version.

- For Java(TM) 2 Platform Standard Edition Runtime Environment Version 7.0

Select the Java tab in Java Plug-in Control Panel and click the View button under the Java Applet Runtime Settings section to add an option to specify the memory volume to the Java Runtime Parameters box.

The maximum memory volume is added using the -Xmx option.

-Xmx

Specifies the maximum memory allocation pool in bytes.
The specified value must be more than 2 MB.
The default value is 64 MB.

Example: when setting 189 MB

```
-Xmx189m
```

Example

Example for calculating heap memory size

Below explains the examples of embedding resources when there are two nodes (uAp_0) and when there are two userApplications (uAp_1).

```
uAP_0
  Cmdline                     4 [Number]
  Procedure(Application)      8 [Number]
  Procedure(BasicApplication) 8 [Number]
  Fsystem                     2 [Number]

uAP_1
  Cmdline                     1 [Number]
  Procedure(Application)      1 [Number]
```

1. Create a table.

   uAP_0

Number of resources	Factors of formula	Resource type
4		Cmdline
0		Controller
0		Nwcl NetWorker
0		Nwst NetWorker
0		Nwsv NetWorker
8		Procedure(Application)
8		Procedure(BasicApplication)
0		Symfoware
0		Oracle
0		Procedure(SystemState3)
0		Procedure(SystemState2)
0		FNA1_FSUNvcph
0		FNA2_FSUNvcp
0		FNALAN_FSUNlndfc
0		BCLine_FSUNbcdr
0		Ims
0		Patrol
0		SH_SWLine
0		Gls
0		Ipaddress
2		Fsystem
0		Gds

   uAP_1

Number of resources	Factors of formula	Resource type
1		Cmdline
0		Controller
0		Nwcl NetWorker
0		Nwst NetWorker
0		Nwsv NetWorker
1		Procedure(Application)
0		Procedure(BasicApplication)

Number of resources	Factors of formula	Resource type
0		Symfoware
0		Oracle
0		Procedure(SystemState3)
0		Procedure(SystemState2)
0		FNA1_FSUNvcph
0		FNA2_FSUNvcp
0		FNALAN_FSUNlndfc
0		BCLine_FSUNbcdr
0		Ims
0		Patrol
0		SH_SWLine
0		Gls
0		Ipaddress
0		Fsystem
0		Gds

2. Delete the lines if the number of resources is 0 from the table created in example 1.

   uAP_0

Number of resources	Factors of formula	Resource type
4	R0	Cmdline
8	R1	Procedure(Application)
8	R2	Procedure(BasicApplication)
2	R3	Fsystem

   uAP_1

Number of resources	Factors of formula	Resource type
1	R0	Cmdline
1	R1	Procedure(Application)

3. Assign each coefficient to each formula.

```
uAp_0
2(node) * ( 1 + 4(Cmdline)
              + 4(Cmdline) * 8(Proc)
              + 4(Cmdline) * 8(Proc) * 8(Proc)
              + 4(Cmdline) * 8(Proc) * 8(Proc) * 2(Fsystem) )
= 1610

uAp_1
2(node) * ( 1 + 1(Cmdline)
              + 1(Cmdline) * 1(Proc)
= 6
```

4. Calculate the total number of resources.

```
1610 (Number of resources for uAp_0)
+
6    (Number of resources for uAp_1)
+
4    (Number of resources for userApplication)
     (It is four because two applications exist in two nodes)
= 1620
```

6.12 Notes When Setting Cmdline Resources

Users need to create a script for the following cases when: starting or stopping ISV applications and user application in line with the userApplication state transition, and switching the userApplication status in line with the stopping of the applications.
Set the created scripts as Cmdline resources, and then set those resources in the userApplication.
This chapter also describes the example of the scripts and notes when creating them.

The following three script types can be set to Cmdline:

- Start script

  is started when the status of userApplication is transited to Online or Standby.
  is a script to start user applications.

- Stop script

  is started when the status of userApplication is transited to Offline.
  is a script to stop user applications.

- Check script

  is used to monitor the status of resources (user applications) to be started or stopped with a Start or Stop script. It is executed in regular intervals after starting RMS. In addition, it is a script to report the status of user applications.

  (*) If the processing time of the Check script (time from the start to the end of the Check script) is within about 0.25 seconds, it is started in about 10-second intervals. If the processing time exceeds 0.25 seconds, it is started in about 20-second intervals.

Besides, the Start script and Stop script are called as the Online script and Offline script respectively.

For the attributes that can be set to the Cmdline resources, see "Setting up Cmdline flags" in "6.7.1.1 Creating Cmdline Resources."

Note

When PRIMECLUSTER products are not specified, do not change ReturnCodes of the Cmdline resource.

6.12.1 Scripts and State Transition

At RMS startup, the Check script is executed regardless of the setting of the AutoStartUp attribute.
In addition, during a state transition, the Stop and Start scripts are called.

In each script, you need to determine whether to perform the Online processing or Offline processing after referring to HV_LAST_DET_REPORT and HV_INTENDED_STATE.
The values of HV_LAST_DET_REPORT and HV_INTENDED_STATE set for each script during its execution are as follows.

Table 6.3 The Cmdline resource in other than Hot-standby operation

State transition			Script for execution	Value of environment variable	
Classification		State of the Cmdline resource		HV_LAST_DET_REPORT*1	HV_INTENDED_STATE
At RMS startup	Operational system	Offline->Online	Start script	Offline	Online
	Standby system	Offline->Offline	-	-	-
At RMS stop	Operational system	Online->Offline	Stop script	Online	Offline
	Standby system	Offline->Offline	Stop script *2	Offline	Offline
At switchover (operation)	Operational system	Online->Offline	Stop script	Online	Offline

State transition			Script for execution	Value of environment variable	
Classification		State of the Cmdline resource		HV_LAST_DET_REPORT*1	HV_INTENDED_STATE
	Standby system	Offline->Online	Start script	Offline	Online
At switchover (resource failure)	Operational system	Online->Faulted	-	-	-
		Faulted->Offline	Stop script	Offline	Offline
	Standby system	Offline->Online	Start script	Offline	Online
At cutting of (resource failure in standby system) *3	Operational system	Offline	-	-	-
	Standby system	Offline->Offline	Stop script *2	Offline	Offline
At exit of maintenance mode	Operational system	Online->Online	Start script *4	Online	Online
	Standby system	Offline->Offline	-	-	-

*1: The value of HV_LAST_DET_REPORT is the current resource status just before the "Script for execution" is executed.

*2: This script is executed only when the following conditions exist:

- NULLDETECTOR attribute of the resource is "Yes"; and
- Offline processing is executed when userApplication is in any state other than Offline state.

*3: When a failure of Hot-standby resources that exist under the same userApplication.

*4: This script is executed only when NULLDETECTOR attribute of the resource is "Yes".

Table 6.4 The Cmdline resource in Hot-standby operation

State transition			Script for execution	Value of environment variable	
Classification		State of the Cmdline resource		HV_LAST_DET_REPORT*1	HV_INTENDED_STATE
At RMS startup	Operational system	Offline->Online	Start script	Offline	Online
	Standby system	Offline->Standby *2	Start script	Offline	Standby
At RMS stop	Operational system	Online->Offline	Stop script	Online	Offline
	Standby system	Standby->Offline	Stop script	Standby	Offline
At switchover (operation)	Operational system	Online->Offline	Stop script	Online	Offline
	Standby system	Standby->Online	Start script	Standby	Online
At switchover (resource failure in operation system)	Operational system	Online->Faulted	-	-	-
		Faulted->Offline	Stop script	Offline or Faulted *3	Offline
	Standby system	Standby->Online	Start script	Standby	Online

State transition			Script for execution	Value of environment variable	
Classification		State of the Cmdline resource		HV_LAST_DET_REPORT*1	HV_INTENDED_STATE
At cutting of (resource failure in standby system)	Operational system	Online	-	-	-
	Standby system	Standby->Faulted	-	-	-
		Faulted->Offline	Stop script	Offline or Faulted *3	Offline
At exit of maintenance mode	Operational system	Online->Online	-	-	-
	Standby system	Standby->Standby	-	-	-

*1: The value of HV_LAST_DET_REPORT is the current resource status just before the "Script for execution" is executed.
*2: When the StandbyTransitions attribute is "Startup."
*3: When the Check script is returned to 1 (Offline) during a failure detection, the value of HV_LAST_DET_REPORT is "Offline." When the Check script is returned to 2 (Faulted) during a failure detection, the value of HV_LAST_DET_REPORT is "Faulted."

See

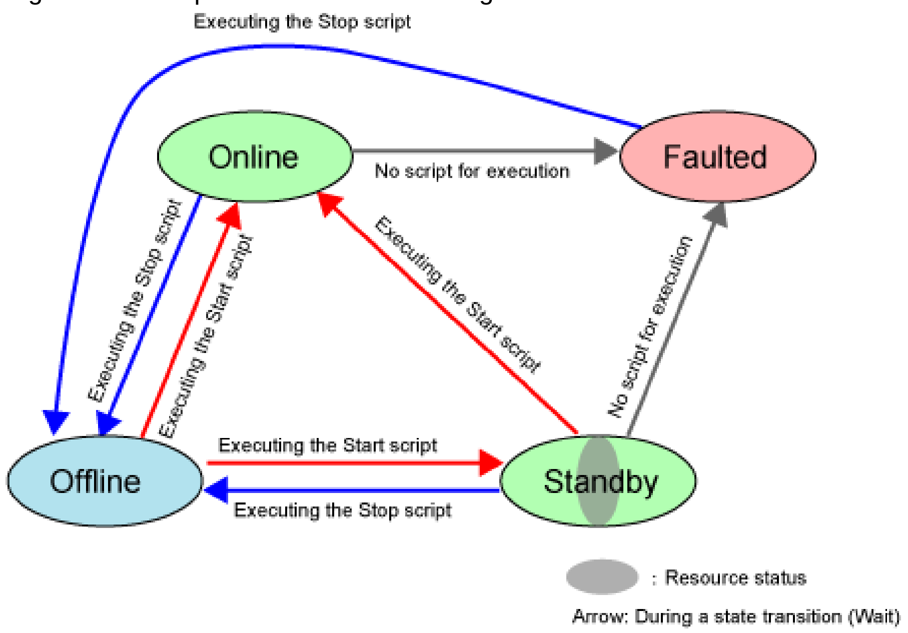
For the environment variable that can be referred to within a script, see "6.12.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts."

6.12.1.1 Scripts to be Executed in Each Resource State

Scripts to be executed in each state for the Cmdline resource during a state transition are as follows.

For the execution order for each script when the state is transited, see "6.12.1.5 Flow of the Cmdline Resource Operation."

Figure 6.72 Scripts to be executed during a state transition

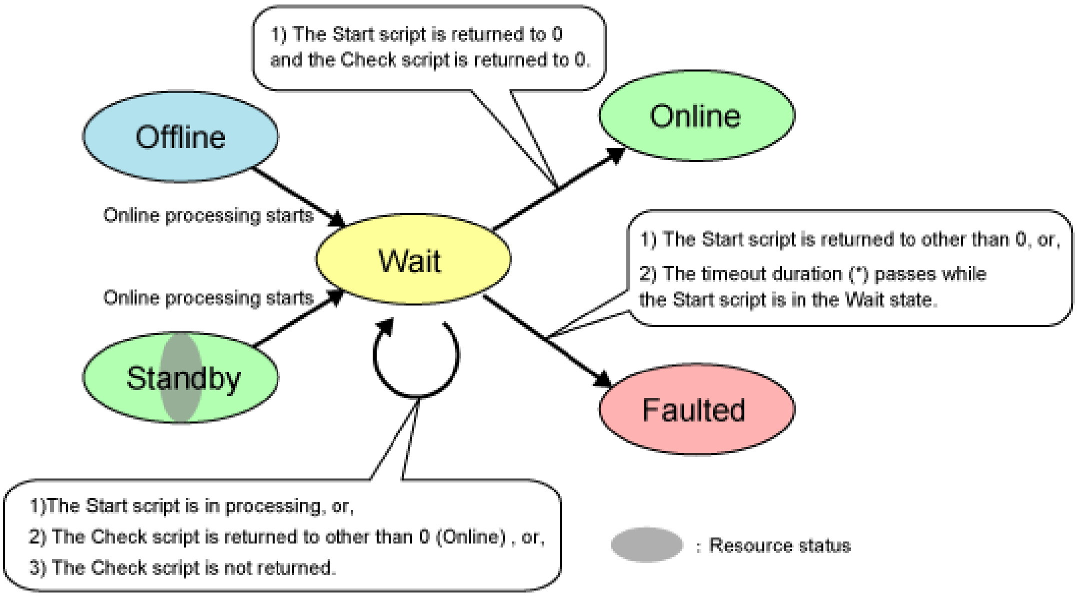


6.12.1.2 Script States When Online

When the Cmdline resource is Online, the Start script is executed only one time.
The Check script is executed in 10-second intervals. It is also executed immediately after completing the Start script.
For details, see "6.12.1.5 Flow of the Cmdline Resource Operation."

The Start script and Check script are switched based on the exit code. The states are as follows.
For details on the exit codes, see "6.12.2.2.2 Check Script Exit Code."

Figure 6.73 Scripts states when Online



(*) For a timeout, see "6.12.2.1.5 Timeout of Scripts."

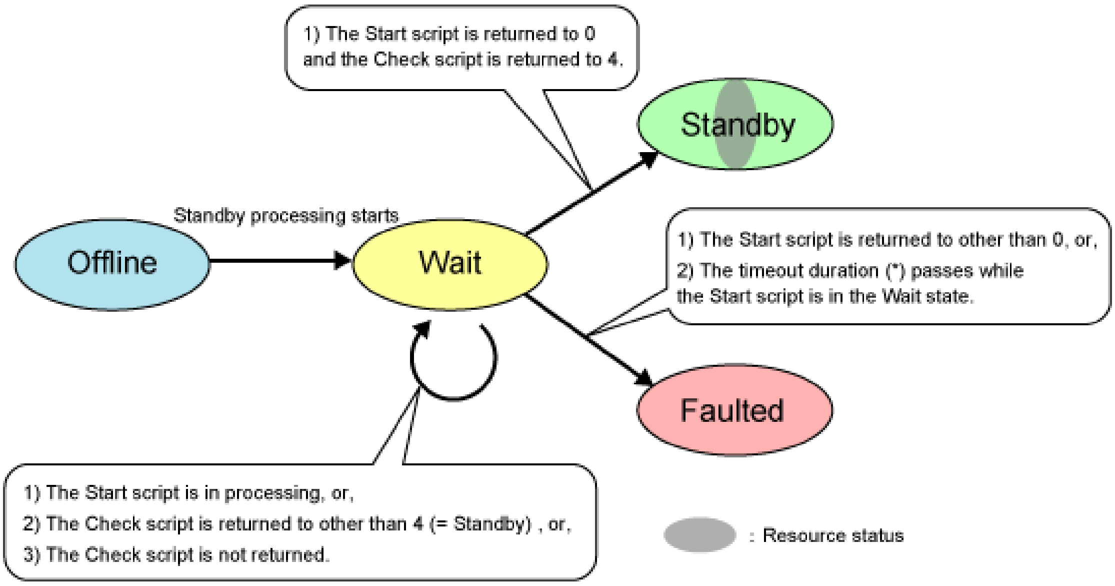
6.12.1.3 Script States When Standby

Only the Cmdline resource of Hot-standby operation becomes Standby.

When the Cmdline resource is Standby, the Start script is executed only one time.
The Check script is executed in 10-second interval. It is also executed immediately after completing the Start script.
For details, see "6.12.1.5 Flow of the Cmdline Resource Operation."

The Start script and Check script are switched based on the exit code. The states are as follows.
For details on the exit codes, see "6.12.2.2.2 Check Script Exit Code."

Figure 6.74 Script states when Standby



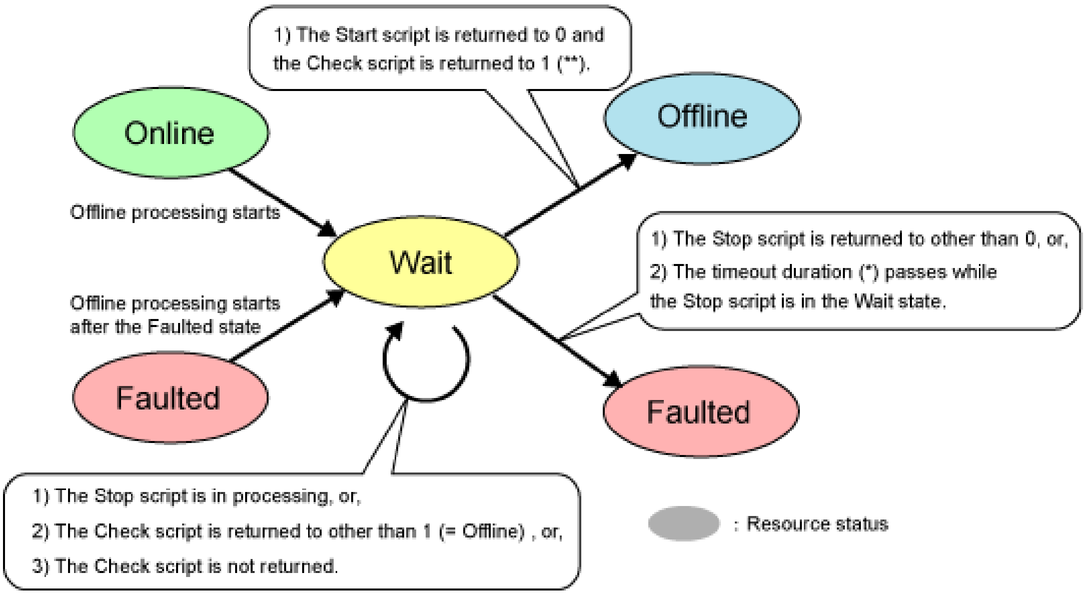
(*) For a timeout, see "6.12.2.1.5 Timeout of Scripts."

6.12.1.4 Script States When Offline

When the Cmdline resource is Offline, the Stop script is executed only one time.
The Check script is executed in 10-second intervals. It is also executed immediately after completing the Stop script.
For details, see "6.12.1.5 Flow of the Cmdline Resource Operation."

The Stop script and Check script are based on the exit code. The states are as follows.
For details on the exit codes, see "6.12.2.2.2 Check Script Exit Code."

Figure 6.75 Script states when Offline



(*) For a timeout, see "6.12.2.1.5 Timeout of Scripts."

(**) It is when ALLEXITCODES is set. For details, see the Outline of the ALLEXITCODES attribute in Table 3.1.

6.12.1.5 Flow of the Cmdline Resource Operation

The Operation of the Cmdline resource is classified as follows:

- At RMS startup
- At RMS stop
- At switchover

In addition to the Cmdline resource, the Gls resource is also described in the following figures as an example.

- At RMS startup Operational system (Offline->Online)
    - The Cmdline resource operation

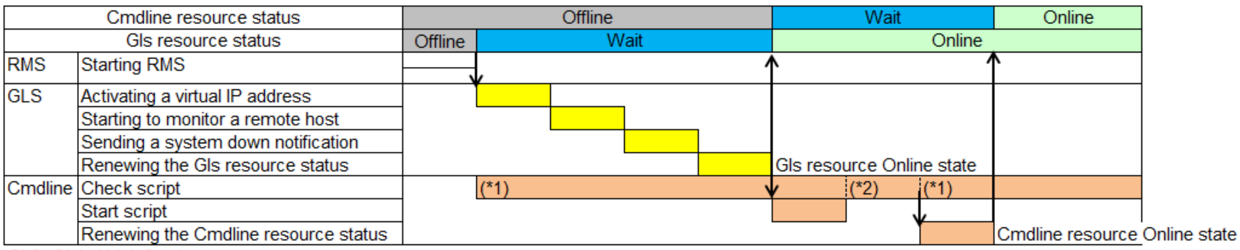
      The Check script is executed in a given interval (about 10 seconds) after starting RMS. After that, the Start script is executed. Then, the Check script is executed without waiting for the given time after returning the Start script. After the Start script is normally returned and the Check script is returned to Online, the Cmdline resource becomes Online.

      Note

      The Check script is operated before the Start script. If the Check script is returned to Online before executing the Start script, the Start script is not executed.

    - Gls resource operation

      At the same time a resource become Online after starting RMS, GLS activates a virtual IP address. In addition, to notify the location of the activated IP address, GLS sends a system down notification.



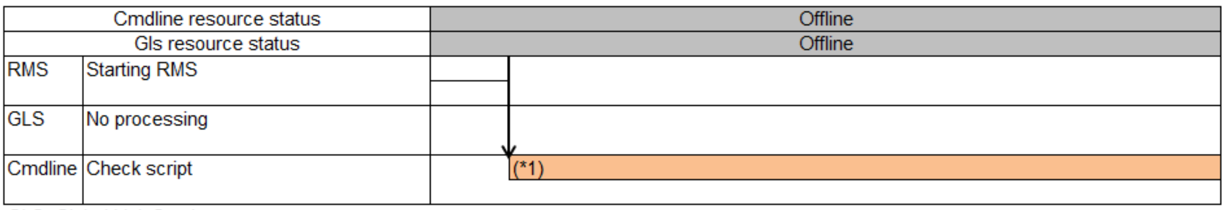
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

- At RMS startup Standby system (Offline->Offline)
    - The Cmdline resource operation

      The Check script is executed in a given interval (about 10 seconds) after starting RMS.

    - Gls resource operation

      No processing.



GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).

- At RMS startup Standby system (Offline->Standby)
    - The Cmdline resource operation

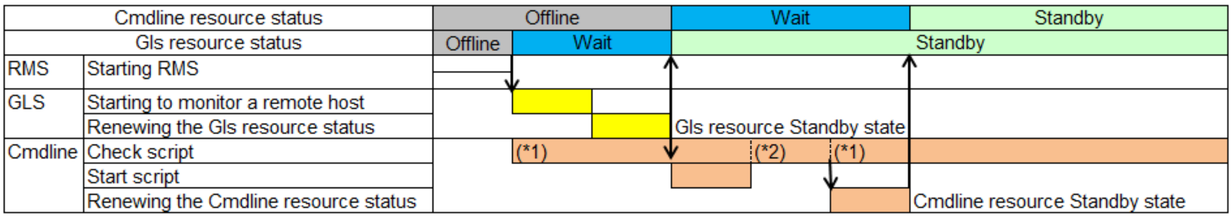
      The Check script is executed in a given interval (about 10 seconds) after starting RMS. After that, the Start script is executed. Then, the Check script is executed without waiting for the given time after returning the Start script. After the Start script is normally returned and the Check script is returned to Standby, the Cmdline resource becomes Standby.

      Note

      The Check script is operated before the Start script. If the Check script is returned to Online before executing the Start script, the Start script is not executed.

    - Gls resource operation

      In Standby state, GLS monitors a network route with the host monitoring function (ping monitoring) without activating a virtual IP address.



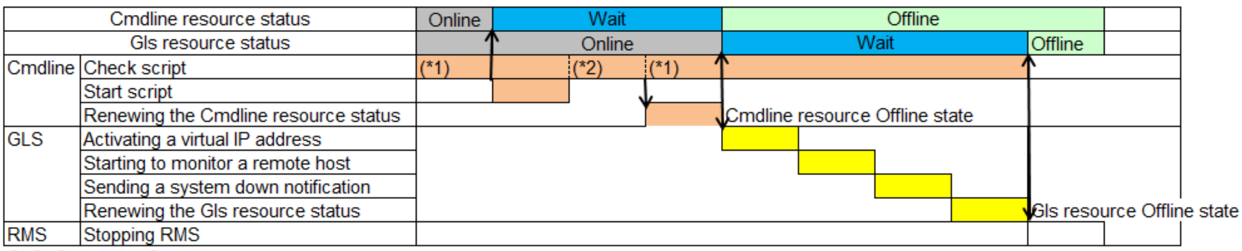
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

- At RMS stop Operational system (Online->Offline)
    - The Cmdline resource operation

      The Stop script is executed. Without waiting for the given time, the Check script is executed. After the Check script is returned to Offline, the corresponding Cmdline resource becomes Offline.

    - Gls resource operation

      Inactivate the virtual IP address that has been activated when Online state. Moreover, if the user command execution function (RESOURCE_OFFLINE) of GLS is set, execute the script.



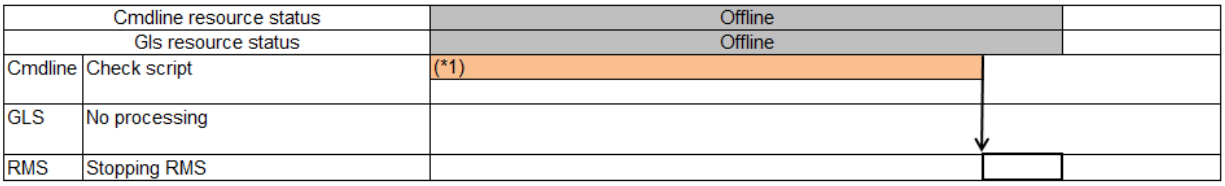
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

- At RMS stop Standby system (Offline->Offline)
    - The Cmdline resource operation

      The Cmdline resource has already Offline, the Stop script is not executed.

- Gls resource operation

  No processing.



GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).

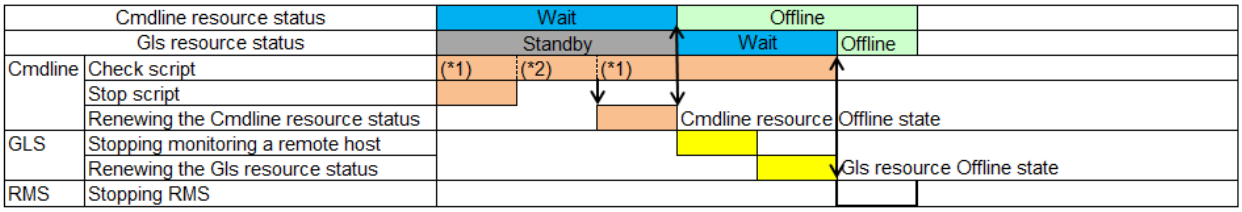
- At RMS stop Standby system (Standby->Offline)

  - The Cmdline resource operation

    The Stop script is executed. Without waiting for the given time, the Check script is executed. After the Check script is returned to Offline, the corresponding Cmdline resource becomes Offline.

  - Gls resource operation

    Stop the host monitoring function (ping monitoring) for a remote host running with Standby state.



GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

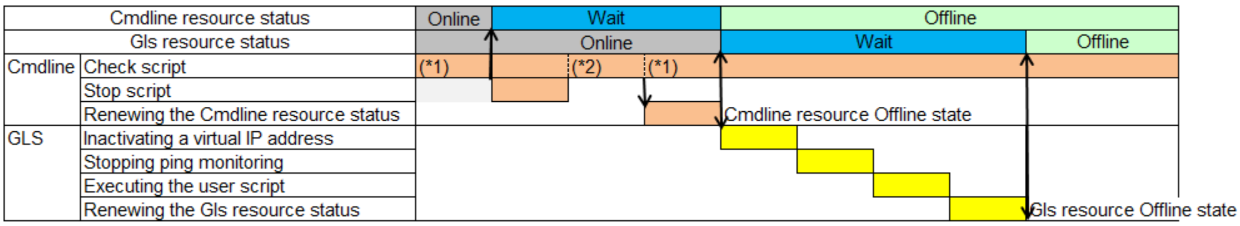
- At switchover Operational system (Online->Offline)

  - The Cmdline resource operation

    The Stop script is executed. Without waiting for the given time, the Check script is executed. After the Check script is returned to Offline, the corresponding Cmdline resource becomes Offline.

  - Gls resource operation

    Inactivate the virtual IP address that has been activated when Online state. Moreover, if the user command execution function (RESOURCE_OFFLINE) of GLS is set, execute the script.



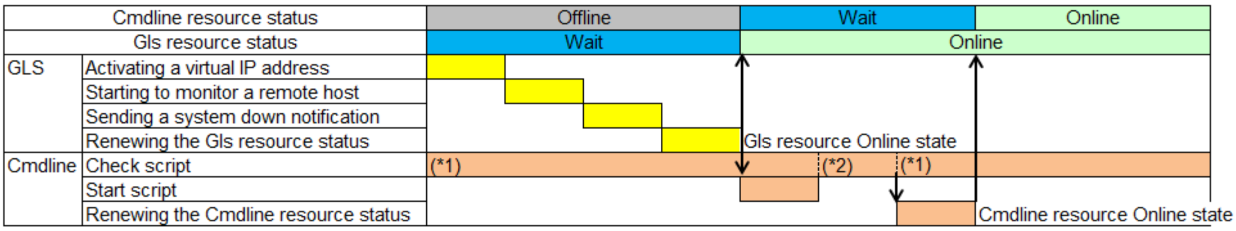
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

- At switchover Standby system (Offline->Online)
  - The Cmdline resource operation

    The Start script is executed. Without waiting for the given time, the Check script is executed. After the Check script is returned to Online, the corresponding Cmdline resource becomes Online.

  - Gls resource operation

    At the same time a resource become Online, GLS activates a virtual IP address. In addition, to notify the whereabouts of the activated IP address, GLS sends a system down notification.



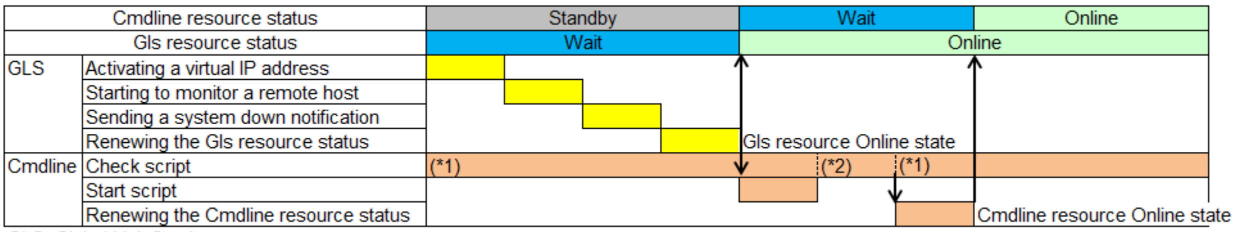
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

- At switchover Standby system (Standby->Online)
  - The Cmdline resource operation

    The Start script is executed. Without waiting for the given time, the Check script is executed. After the Check script is returned to Online, the corresponding Cmdline resource becomes Online.

  - Gls resource operation

    At the same time a resource becomes Online, GLS activates a virtual IP address. In addition, to notify the location of the activated IP address, GLS sends a system down notification.



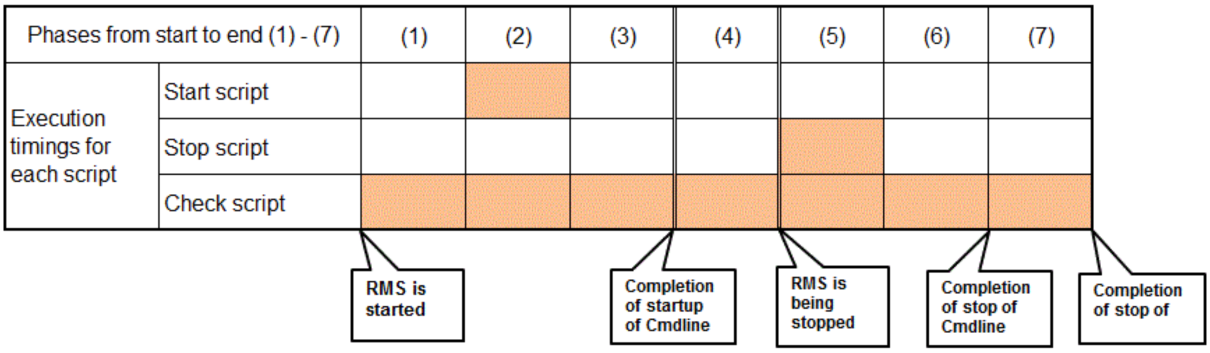
GLS: Global Link Services
(1*) The Check script is executed in a given interval (about 10 seconds).
(2*) The Check script is executed without waiting for the given time after returning the Start script.

6.12.1.6 Operation for Each Exit Code of the Check Script

This section describes the operations for each exit code of the Check script. They are classified from (1) to (7) phases from start to end as follows:

(1) Phase from RMS is started to the Cmdline resource is started.
(2) Phase the Cmdline resource is being started (the Start script is being executed.
(3) Phase the status is being checked (the Check script is being executed) after starting the Cmdline resource.
(4) Phase the Cmdline resource is being operated.
(5) Phase the Cmdline resource is being stopped (The Stop script is being executed).
(6) Phase the status is being checked (The Check script is being executed) after stopping the Cmdline resource.
(7) Phase from the Cmdline resource is stopped to RMS is stopped.

- For standby systems of the Cmdline resource other than Hot-standby operation, the Start script is not executed at RMS startup. Thus, the phases 2 and 3 do not exist.
- For standby systems of the Cmdline resource other than Hot-standby operation, the Stop script is not executed at RMS stop. Thus, the phases 5 and 6 do not exist.



- The Cmdline resource with Hot-standby operation

The Start script is executed
The status is transited to Faulted and then the Stop script is executed

			Return value of the Check script				
			0 Online	1 Offline	2 Faulted	3 Unknown	4 Standby
Phases from start to end (1) - (7)	(1)	Operational system	Online	Transitted to Offline state and then the Start script is executed	Faulted	Unknown	Transitted to Offline state and then the Start script is executed
		Standby system	Online	Transitted to Offline state and then the Start script is executed	Faulted	Unknown	Standby
	(2)	Operational system	Wait	Wait	Wait	Wait	Wait
		Standby system	Wait	Wait	Wait	Wait	Wait
	(3)	Operational system	Online	Wait	Wait	Wait	Wait
		Standby system	Online	Wait	Wait	Wait	Standby
	(4)	Operational system	Online	Transitted to Faulted state and then the Stop script is executed	Transitted to Faulted state and then the Stop script is executed	Online	Transitted to Faulted state and then the Stop script is executed
		Standby system	Online	Transitted to Faulted state and then the Stop script is executed	Transitted to Faulted state and then the Stop script is executed	Standby	Standby
	(5)	Operational system	Wait	Wait	Wait	Wait	Wait
		Standby system	Wait	Wait	Wait	Wait	Wait
	(6)	Operational system	Wait	Offline	Wait	Wait	Wait
		Standby system	Wait	Offline	Wait	Wait	Wait
	(7)	Operational system	Faulted	Offline	Faulted	Offline	Faulted
		Standby system	Faulted	Offline	Faulted	Offline	Faulted

- The Cmdline resource other than Hot-standby operation

(cyan)	The Start script is executed
(pink)	The status is transited to Faulted and then the Stop script is executed

			Return value of the Check script	
			0 Online	**Ohter than 0 Offline**
Phases from start to end (1) - (7)	**(1)**	**Operational system**	Online	Transitted to Offline state and then the Start script is executed
		Standby system	Online	Offline
	(2)	**Operational system**	Wait	Wait
		Standby system	-	-
	(3)	**Operational system**	Online	Wait
		Standby system	-	-
	(4)	**Operational system**	Online	Transitted to Faulted state and then the Stop script is executed
		Standby system	Online	Offline
	(5)	**Operational system**	Wait	Wait
		Standby system	-	-
	(6)	**Operational system**	Wait	Offline
		Standby system	-	-
	(7)	**Operational system**	Faulted	Offline
		Standby system	Online	Offline

6.12.2 Notes When Creating Scripts

This section describes notes when creating scripts.

6.12.2.1 start and stop Scripts

6.12.2.1.1 Examples of start and stop Scripts

Examples of the Start and Stop scripts other than Hot-standby operation are as follows.
The script $FULL_PATH/script.sample is an example when the following operations and settings are assumed.

[Setting]

- Start script $FULL_PATH/Script.sample -c
- Stop script $FULL_PATH/Script.sample -u

[Attribute]

- STANDBYCAPABLE: No
- AUTORECOVER: No
- CLUSTEREXCLUSIVE: Yes
- NULLDETECTOR: No
- MONITORONLY: No

[Operation]

Below is an example when assuming the operation is the same for standby and operational systems by following "Table 6.3 The Cmdline resource in other than Hot-standby operation." The same processing is executed in the lines where Start script is described. The same processing is also executed in the lines where Stop script is described.

When assuming operations other than the above, refer to the environment variable and attribute to change them.

Figure 6.76 Start script and Stop script other than Hot-standby operation

```
#!/bin/sh
#
# Script.sample
#    Sample of  Online/Offline Script
#
# Copyright(c) 2003-2006 FUJITSU LIMITED.
# All rights reserved.
#
# $1 -c: OnlineScript
#    -u: OfflineScript

if [ "$1" = "-c" ]; then
    # Start script
    # Start your application
elif [ "$1" = "-u" ]; then
    # Stop script
    # Stop your application
else
    # Default operation
    exit 1     # Error
fi

exit 0
```

Moreover, below is an example when assuming that the operation of the following sample $FULL_PATH/Script is corresponding Hot-standby operation.

[Setting]

- Start script $FULL_PATH/Script.sample -c

- Stop script $FULL_PATH/Script.sample -u

[Attribute]

- STANDBYCAPABLE: Yes
- AUTORECOVER: No
- CLUSTEREXCLUSIVE: Yes
- ALLEXITCODES: Yes
- NULLDETECTOR: No
- MONITORONLY: No

[Operation]

Below is an example of the Start script when the status is transited from Offline to Standby and also from Offline to Online. The transitions are distinguished as "Table 6.4 The Cmdline resource in Hot-standby operation."
In addition to that, another example that the Stop script distinguishes when the status is transited from Standby to Offline and also from Online to Offline is as follows.

When assuming operations other than the above, refer to the environment variable and attribute to change them.

Figure 6.77 Start script and Stop scripts with Hot-standby operation

```
#!/bin/sh
#
# Script.sample
#    Sample of  Online/Offline Script
#
# Copyright(c) 2003-2006 FUJITSU LIMITED.
# All rights reserved.
#
# $1 -c: OnlineScript
#    -u: OfflineScript

if [ "$1" = "-c" ]; then
    # Start script
    if [ "${HV_LAST_DET_REPORT}" = "Offline" ]; then
        if [ "${HV_INTENDED_STATE}" = "Standby" ]; then
            # commands for Offline -> Standby
        else
            # commands for Offline -> Online
        fi
    else
        # commands for Standby -> Online
    fi
elif [ "$1" = "-u" ]; then
    # Stop script
    if [ "${HV_LAST_DET_REPORT}" = "Standby" ]; then
        # commands for Standby -> Offline
    else
        # commands for Online -> Offline
    fi
else
    # Default operation
    exit 1    # Error
fi
exit 0
```

6.12.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts

When executing the Start script and Stop script, the following environment variables are set. You can refer to those environment variables within the scripts.
Table 3.4 indicates the environment variables set in the scripts.

Table 6.5 Environment variables can be referred to within the Start and Stop scripts

Environment variables	Outline
HV_APPLICATION	This variable sets the userApplication name that the resource belongs to. Example) app1
HV_AUTORECOVER	The value of this variable indicates whether the script is triggered by AutoRecover or not. 0: Not triggered by AutoRecover that is executed with the Online processing 1: Triggered by AutoRecover
HV_FORCED_REQUEST	This variable sets a value that indicates whether or not forced failover was requested by operator intervention. 0: Forced failover was not requested. 1: Forced failover was requested.
HV_NODENAME	This variable sets the resource name. Example) ManageProgram000_Cmd_APP1, RunScriptsAlways000_Cmd_APP1
HV_OFFLINE_REASON	This variable sets the trigger for bringing the resource Offline. SWITCH: The resource was set to Offline because of a userApplication switchover request (hvswitch) STOP: The resource was set to Offline because of a userApplication stop request (hvutil -f) FAULT: The resource was set to Offline because of a resource fault. DEACT: The resource was set to Offline because of a userApplication deactivate request (hvutil -d) SHUT: The resource was set to Offline because of an RMS stop request (hvshut)
HV_SCRIPT_TYPE	This variable sets the type of script that was executed. Online: Online script Offline: Offline script
HV_LAST_DET_REPORT	This variable sets the state of the current resources just before execution of the Start/Stop script. Online: Online state Offline: Offline state Standby: Standby state Faulted: Faulted state
HV_INTENDED_STATE	This variable sets the resource state that is expected after state transition is completed. Online: Online state Offline: Offline state Standby: Standby state Faulted: Faulted state Warning: Warning state
NODE_SCRIPTS_TIME_OUT	This variable sets the timeout duration (seconds) of the script. Example) 300

RMS has other environment variables.

See

- For details on the RMS environment variables, see "13 Appendix - Environment variables" in the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

6.12.2.1.3 Exit Code of Start and Stop Scripts

The state transition operation of a userApplication varies depending on the exit code of the Start and Stop script: Below indicates the operations of the exit code and the state transition.

0: Normal exit

The system assumes that the state transition of the Cmdline resource was processed normally, and state transition processing of the userApplication continues. If all the resources of the userApplication are processed normally, the state transition of the

userApplication is also processed normally

Other than 0: Abnormal exit

The system assumes that an error occurred during the state transition of the Cmdline resources and interrupts state transition processing of the userApplication.

6.12.2.1.4 Notes When Setting the NULLDETECTOR Flag

RMS does not monitor the state of the Cmdline resource when the NULLDETECTOR flag is enabled. In this case, Online script may be executed when the resource is already started or Offline script may be executed when the resource is already stopped. To prevent Online or Offline processing to be terminated with error, be sure to add following tasks to Online script and Offline script of the Cmdline resource when the NULLDETECTOR flag is enabled.

- Online script

  Check whether a target program has already run before starting it within the Online script. If it has already run, the Online script is immediately stopped.

- Offline script

  Check whether the target program has already stopped before stopping it within the Offline script. If it has already stopped, the Offline script is immediately stopped.

Note

If the userApplication state before the maintenance mode is started is Online or Standby, even the resource has been started by the time the maintenance mode ends, the Online script of Cmdline resource where the NULLDETECTOR flag is set is executed.

6.12.2.1.5 Timeout of Scripts

If the Start or Stop script processing is not completed within the specified time, a timeout occurs. Then, the script processing is interrupted by the SIGTERM signal and state transition ends with an error.
The timeout value can be specified with the TIMEOUT flag value of the Cmdline resources. The default value is 300 seconds.

When creating the Cmdline resource, you need to calculate the maximum processing time for each script and set a value with enough time. If a timeout occurs when the Cmdline resource is used (any one of the following messages is output: (DET, 5), (DET, 6), or (DET, 24)), change the timeout value to an appropriate value according to each operating system being used.

Select "Application-Edit" from the Main configuration menu to change the Flag of the Cmdline resource.
For details, see "10.3 Changing a Cluster Application."

Note

The processing time for each script needs to be shorter than the ScriptTimeout value of attribute that users have set.
If the processing time of scripts exceeds the ScriptTimeout value, PRIMECLUSTER determines it is a resource error and stop the startup and stop processings.

6.12.2.2 Check Script

6.12.2.2.1 Example of the Check Script

An example of the Check script other than Hot-standby operation indicates as follows.
The following example assumes that the setting has already described in 6.12.2.1.1 Examples of start and stop Scripts."

Figure 6.78 The Check script other than Hot-standby operation

```
#!/bin/sh
#
# Script.sample.check
#    Sample of Check script
#
# Copyright(c) 2003 FUJITSU LIMITED.
# All rights reserved.
#
# Check the current state of target resource.

# If status is Online:
      exit 0
# If status is not Online:
      exit 1
```

If performing Hot-standby operation in the Cmdline resource, describe the Check script, which is similar to the start and stop scripts, corresponding to Hot-standby operation

Below is an example of the Check script corresponding to Hot-standby operation.
The following example assumes that the setting has already described in 6.12.2.1.1 Examples of start and stop Scripts."

Figure 6.79 The Check script with Hot-standby operation

```
#!/bin/sh
#
# Script.sample.check
#     Sample of Check script
#
# Copyright(c) 2003 FUJITSU LIMITED.
# All rights reserved.
#

# Check the current state of target resource.
# If status is Online:
   exit 0
# If status is Standby:
   exit 4
# If status is Faulted:
  exit 2
# If status is Offline:
   exit 1
```

6.12.2.2.2 Check Script Exit Code

The exit codes of the Check script vary depending on whether performing Hot-standby operation or not.

Not performing Hot-standby operation

Use any one of the following exit codes.

Table 6.6 The exit codes other than Hot-standby operation

Exit code	Meaning
0	Online state. If a target to be monitored has started, use this exit code.
Other than 0	Offline state. If a target to be monitored has stopped, use an exit code other than 0. If it completes in Offline after it has become Online, a switchover is performed.

Performing Hot-standby operation

Use any one of the following exit codes.

Table 6.7 The exit codes for Hot-standby operation

Exit code	Meaning
0	Online state. If a target to be monitored has started, use this exit code.
1	Offline state. If a target to be monitored has stopped, use this exit code. If it completes in Offline after it has become Online, a switchover is performed.
2	Faulted state. If a target to be monitored has faulted, use this exit code. If it completes in Faulted after it has become Online, a switchover is performed.
3	Unknown state. If the state of a target to be monitored is unknown, use this exit code.
4	Standby state. If a target to be monitored has been Standby state, use this exit code.

6.12.3 Notes on Scripts

- The execute permission for each script is user: root and group: root.
- Environment variables set in each server (such as "/etc/profile") are not guaranteed to be inherited by Start, Stop, and Check scripts. Therefore, make sure to define the environment variables used with these scripts in each script.
- The Check script is called in regular intervals (10-second intervals) after starting RMS. It does not synchronize with the Start or Stop script.
  Therefore, at the time the Check script is started, the processing of the Start script has not completed or the Stop script may still be in process.
  If the Check script has started before completing the Start script, create a script so that the exit code Offline is returned.
- To register each script, make sure to check the script operation.
  If the created script does not operate properly, the startup of userApplication or a switchover may fail.
- The Cmdline resource is managed by its creator. Thus, for the operation error, the creator need to investigate the cause, modify the error, and check the operation.
  To investigate the cause of the error immediately, take some actions such as outputting a log.
- The Stop script is also executed when a resource failure occurs.
- The Cmdline resource starts the Start and Stop scripts so that the standard output and standard error output is stored in the following log.

  ```
  /var/opt/SMAWRrms/log/"user_application_name".log
  ```

  "user application name" is the user application name that the Cmdline resource has registered. If the Start or Stop script does not operate properly, you can investigate the cause from the message output in this file.
- When starting a resident process from the Start script registered in the Cmdline resource, a file descriptor of the Start script is transferred to the resident process. To output a message to a standard error or standard error output from the resident process, the message is stored in the "user application name".log file. However, the purpose of this file is to obtain a message that the Start and Stop scripts of a resource output. The messages output from the resident process all the time are not assumed. If the resident process keeps outputting messages, the "user application name".log file may weigh on its disk space.
  To start operational application which has a resident process from the Cmdline resource, perform any one of the following resolutions:
  - Change the setting of the operational application so that the resident process does not output a message to a standard output or standard error output.
  - Immediately after starting the resident process, modify the processing of the resident process so that the file descriptor of the standard output or standard error output transferred from the Start script becomes CLOSE.

    Point

    The resident process is started with taking over file descriptors other than the standard output or standard error output. There is no problem to close all the file descriptors.

  - Redirect the messages output from the resident process within the Start script to /dev/null or other files.

    Example

    If a resident process is started with the Start command; StartCommand.sh, register the Start command as follows:

    - The messages output are unnecessary for the operation (the messages are discarded with /dev/null file).

      /usr/local/bin/StartCommand.sh > /dev/null 2>&1

    - The messages are necessary for the operation and they are output to the log file /var/tmp/logfile.

      /usr/local/bin/StartCommand.sh > /var/tmp/logfile 2>&1

Note

To redirect the messages output from the resident process to other log files, you need to delete log files periodically so that they do not weigh on their disk space. You cannot delete log files during the resident process operation, copy /dev/null to log files so that the size of them becomes 0.

```
cp /dev/null /var/tmp/logfile
```

Setting the size of log files 0 periodically from the cron command allows the operation with the enough disk space.

6.13 Notes When Setting Fsystem Resource

The Fsystem resource is used when mounting a file system at userApplication startup.
To control multiple mountpoints in parent-child relationship, create the file system as a single Fsystem resource.
The Fsystem resource performs the following processing:

- Mounting or unmounting a file system
- Checking access errors (including cable disconnection)
- Publishing the network of the NFS file system
- Transferring the Lock information for NFS

6.13.1 Monitoring Fsystem

Before describing notes on the Fsystem resource, this section describes the Fsystem detector.
The Fsystem detector has the following two types:

- Monitoring the mount state of a file system (hvdet_gmount)

  This detector monitors the mount state of a file system as well as access errors.
  It performs the following processing in 10-second intervals to monitor a file system.

  - The mount state of a file system has set in line with the definition of /etc/vfstab.pcl.
  - I/O to the file system has performed properly while it is been mounted.

- Monitoring the shared state of a file system (hvdet_ckshare/hvdet_checkshare)

  This detector is operated when using the NFS server function.
  It judges whether the file system is shared or not in 10-second intervals based on the information in /etc/dfs/sharetab.

6.13.2 Notes When Using NFS Server Function

Transferring the Lock information

If you use an NFS server function, set it so that the Lock information for NFS is succeeded.

If a transfer of the Lock information for NFS is not possible, you can get multiple Lock information for the same file. Therefore, the application that uses fcnlt(2) and lockf(3C) cannot execute the exclusive control. And it becomes the factor to cause file destruction.

But the following attention is necessary on using this function. Execute the next attention whether an application uses a lock unidentified.

- On the node operation with NFS servers, do not use the NFS client function. Do not implement NFS mount.
- A file system in which NFS is shared with the operation of NFS servers must be used only from the NFS client. Do not use a file directly from an application on a node on which NFS servers are operating.

- When failover occurs, the following processing is performed. The access from the client is reserved until the NFS lock recovery processing is completed though the state of the resource becomes Online when the start of statd and lockd is completed.

  - Stopping the statd(1M) and lockd(1M) daemon
  - Starting the statd(1M) and lockd(1M) daemon
  - NFS lock recovery processing

Protocol

NFS can use TCP and UDP protocols. After using TCP for failover and failback in sequence, recovery sometimes takes longer than with UDP.
Therefore, please use the UDP protocol in PRIMECLUSTER.

6.13.3 Fsystem Resource Attribute

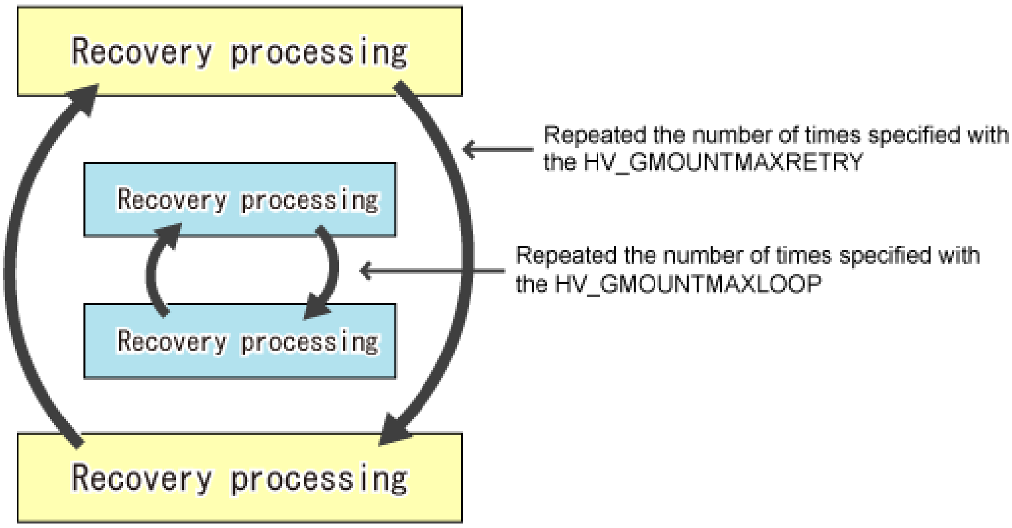
This section describes the Fsystem resource attribute (Flag).

- AUTORECOVER

  If "Yes" is set, hvdet_gmount tries to recover the failure by re-mounting when it detects a failure. If this attempt fails, the Fault processing is executed.
  The default value is "Yes."

Note

"No" is recommended to set to AUTORECOVER.
If you set "Yes," it is effective for the measures when an operator unmounts a file system mistakenly. However, it takes time for a switchover when Fsystem timeouts due to an I/O error, and so on because it tries to perform I/O again.
If an error is detected, hvdet_gmount repeats a recovery processing only for the number of times specified with the HV_GMOUNTMAXLOOP attribute as follows. Even though it cannot be recovered, perform a recovering processing specified times with the HV_GMOUNTMAXRETRY attribute.



The defaults values for HV_GMOUNTMAXLOOP and HVGMOUNTMAXRETRY are four and seven times respectively. The recovery processing for HV_GMOUNTMAXLOOP is executed in 0.5-second intervals while the recovery processing for HV_GMOUNTMAXRETRY is executed in 10-second intervals. Therefore, when a disk or path error cannot be recovered occurs, the re-try processing is executed in about 84 seconds, and then it is switched over.

Note

HV_GMOUNTMAXLOOP and HV_GMOUNTMAXRETRY are RMS environment variables. To change those values, set "export HV_GMOUNTMAXLOOP=value" and "export HV_GMOUNTMAXRETRY=value" to hvenv.local

- SHARE

  When using the NFS server function, the flag must be set to "Yes."
  Also, when using the NFS server function, you must describe an entry used by RMS to /etc/dfs/dfstab.pcl. For how to describe it, see "Preparations for sharing a file system in a network (NFS)" in "6.7.1.2.1 Prerequisites."
  The default value is "No".

- NFSLOCKFAILOVER

  The file lock information of NFS is stored in the mountpoint that is set to "Yes".
  When using NFS, the SHARE attribute in the userApplication must be one of the mountpoints that are set to "Yes" and also the NFSLOCKFAILOVER attribute must be "Yes".
  The default value is "No".

Note

When using the NFS server function, it is recommend that you set the NFS Lock Failover to "Yes" (NFS Lock Failover=Yes) except for sharing it as a read-only flag.

6.13.4 Other Notes

- In Online processing for Fsystem, fsck may be executed.

  If fsck is executed during Online processing, the processing may not be completed within time set with ScriptTimeout. As a result, the startup or switchover processing fails.
  To use Fsystem, set the ScriptTimeout value that the processing time of fsck is considered.

- Do not access mountpoint specified in Fsystem from other than a userApplication.

  During Offline processing, if accessing the mountpoint specified in Fsystem with other process, the Offline processing may fail and a switchover may not be performed.

- Do not change the mountpoint name for Fsystem with such as mv command when a userApplication is Online.

  If the mountpoint name is changed when Online, hvdet_gmount detects an error and a userApplication is switched. To change the mountpoint name temporarily, stop RMS first.

- If 31 or more mountpoints registered in a single Fsystem resource exist, you need to change the default timeout value (180 seconds).

  For the Timeout value of the Fsystem resource, "the number of mountpoints registered in single Fsystem x 6 seconds" or more needs to be set.
  For example, if 31 mountpoints are registered in a single Fsystem resource, set "31 x 6 seconds = 186 seconds" or more to the Timeout attribute of the Fsystem resource.

- The timeout value set in each Fsystem resource is the time until all processing completes for the mountpoints registered in the Fsystem resource.

  For example, if three mountpoints; /mnt1, /mnt2, and /mnt3 are registered in the Fsystem resource, and also 100 seconds is set to the timeout value, the processing times out unless the processing of all three mountpoints completes within 100 seconds.

- For the disk partition used in the Fsystem resource, it is necessary to create beforehand.

  If it has not been created, Online processing fails.

- If a shared disk cannot be accessed, double fault may occur.

  If a shared disk cannot be accessed, the Fsystem resource becomes Faulted. In this case, mountpoints cannot be unmounted in Offline processing (this processing is conducted after Faulted processing), and then double fault may occur.

6.13.5 Maintaining File Systems Controlled by the Fsystem Resource

This section describes the procedure when maintaining file systems on a shared disk registered in the Fsystem resource.

Note

To mount a file system on a shared disk manually, mount it from any one of nodes configuring a cluster system.

If you mount file systems on shared disks from multiple cluster nodes at the same time, these file systems are destroyed. Perform the operation with careful attention.

1. Stopping RMS on all cluster nodes

   Stop RMS on all cluster nodes.

   Example: Stopping RMS on all nodes configuring a cluster from any one of nodes with a command

```
# /opt/SMAW/SMAWRrms/bin/hvshut -a
```

2. Checking the mount state of a file system

   Check that a file system on a shared disk has not been mounted with the df command so that the file system cannot be mounted mistakenly from multiple cluster nodes.

   Example: Executing the df command

```
# /usr/sbin/df -k
Filesystem            kbytes    used   avail capacity   Mounted on
/dev/dsk/c0t0d0s0    6718025 4839652 1811193     73%    /
/proc                      0       0       0      0%    /proc
mnttab                     0       0       0      0%    /etc/mnttab
fd                         0       0       0      0%    /dev/fd
swap                 2244776      16 2244760      1%    /var/run
swap                 2251760    7000 2244760      1%    /tmp
```

   If the file system has already mounted, a cluster application may be in operation or the file system has already been mounted manually.

   In this case, stop the cluster application and RMS, or unmount the target file system with the umount command.

   The following procedure is performed in any one of nodes configuring a cluster.

3. Starting a GDS volume (only if necessary)

   If a file system or a file to be maintained exists in a volume managed by GDS, start the GDS volume in any one for nodes configuring a cluster.

   Example: When starting the volume volume0001 of the disk class *class* with a command

```
# /usr/sbin/sdxvolume -N -c class -v volume0001
```

4. Mounting and maintaining a file system

   If using ZFS

   When referring to a file on the ZFS file system, import the storage pool and mount the file system.

   1. Checking the ZFS storage pool controlled by a cluster

      See the /etc/vfstab.pcl file to check that the pool name of the ZFS file system exists.

      Example: when checking the contents of /etc/vfstab.pcl with the cat command

```
# /usr/bin/cat /etc/vfstab.pcl
# bdev cdev mountpoint fstype runlevel auto mount flags
#RMS#app app /app zfs - - -
```

```
#RMS#app/mp1 app/mp1 /appdata1 zfs - - -
#RMS#app/mp2 app/mp2 /appdata2 zfs - - -
```

2. Importing the ZFS storage pool

   Import the ZFS storage pool with the zpool command.

   Example: when importing the ZFS storage pool "app" that is configured in the volume of the class *class*

   ```
   # /usr/sbin/zpool import -d /dev/sfdsk/class/dsk app
   ```

   Note

   When the ZFS file system of the non-legacy file system has been set to the ZFS storage pool, the ZFS file system is mounted after this procedure.

3. Mounting the ZFS file system defined as the legacy file system (only if necessary)

   Mount the ZFS file system in which "legacy" is set to the mountpoint property.

   Example: when mounting the storage pool app/mp1 to /appdata1

   ```
   # /usr/bin/mount -F zfs app/mp1 /appdata1
   ```

4. Maintenance of a file (only if necessary)

   When the file that is used by an application is on the shared disk, refer to and update the file now.

5. Unmounting the ZFS file system defined as the legacy file system (only if necessary)

   Unmount the ZFS file system that was mounted in step 4-3.

   Example: unmounting /appdata1

   ```
   # /usr/bin/umount /appdata1
   ```

6. Exporting the ZFS storage pool

   Export the ZFS storage pool that was imported in step 4-2.

   Example: when exporting the ZFS storage pool "app"

   ```
   # /usr/sbin/zpool export app
   ```

If using UFS

When referring to a file on the UFS file system, mount the file system.

1. Restoring the UFS file system (only if necessary)

   If it is necessary to restore the file system, restore it with the fsck command. If the target UFS file system exists in the volume that is managed by GDS, execute the fsck command on the node where the GDS volume was started in step 3.

   Example: when restoring the UFS file system of the GDS volume /dev/sfdsk/class/rdsk/volume0001

   ```
   # /usr/sbin/fsck -F ufs /dev/sfdsk/class/rdsk/volume0001
   ```

2. Mounting the file system (only if necessary)

   When mounting the UFS file system, mount it with the mount command.

   The device name of the file system controlled by the Fsystem resource has been described in the /etc/vfstab.pcl file. Refer to the /etc/vfstab.pcl file to mount the file system.

   Example: when checking the contents of the /etc/vfstab.pcl file with the cat command

   ```
   # /usr/bin/cat /etc/vfstab.pcl
   #RMS#/dev/sfdsk/class0001/dsk/volume0001 /dev/sfdsk/class0001/rdsk/volume0001 /disk1 ufs
   - no -
   ```

Example: when mounting the file system of the mount point /disk1 controlled by the Fsystem resource

```
# /usr/bin/mount -F ufs /dev/sfdsk/class0001/dsk/volume0001 /disk1
```

3. Maintaining files (only if necessary)

   If files used by an operational application exist on a shared disk, refer to and update the files at this point.

4. Unmounting the file system

   If you have mounted the file system in step 4-2, unmount it with the following procedure.

   Example: when unmounting the file system unmounted in /disk1

```
# /usr/bin/umount /disk1
```

5. Stopping the GDS volume (if step 3 was performed)

   Stop the GDS volume started in step 3.

   Example: when stopping the volume *volume0001* of the disk class *class* with a command

```
# /usr/sbin/sdxvolume -F -c class -v volume0001
```

6. Starting RMS on all nodes

   Start RMS on all nodes.

   Example: when starting RMS on all nodes configuring a cluster from any one of nodes with a command

```
# /opt/SMAW/SMAWRrms/bin/hvcm -a
```

Part 3 Operations

Chapter 7 Operations

This chapter describes the functions managing PRIMECLUSTER system operations. They monitor operation statuses for PRIMECLUSTER system and operate PRIMECLUSTER system according to its operation statuses and so on. Also, notes for operating PRIMECLUSTER system are described.

The following user groups are allowed to do each specific operation:

Operation	Target
Referencing the operation management screens	All user groups
Operations	wvroot, clroot, cladmin
Monitoring	All user groups
Corrective actions for resource failures	wvroot, clroot, cladmin

7.1 Viewing the PRIMECLUSTER System Operation Management Screens

PRIMECLUSTER provides GUIs for viewing and performing cluster system operations.

- CF main window

  Use this screen to set up the configuration of the nodes that make up the cluster, manage the nodes, and display the node state.

- CRM main window

  Use this screen to identify faulted hardware and to execute hardware diagnosis.

- RMS main window

  Use this screen to monitor the state of the cluster system and to manage cluster applications and resources.

- MSG main window

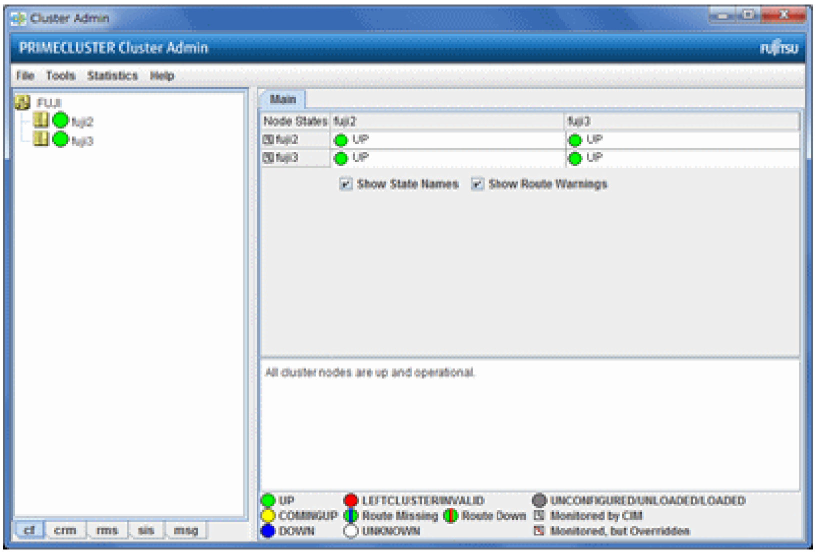
  This screen displays messages for cluster control.

See

For instructions on displaying each screen, see "4.4.3 Cluster Admin Functions."

7.1.1 CF Main Window

The CF main window allows you to set up the configuration of the cluster nodes in the cluster, manage the nodes, and display the node state.



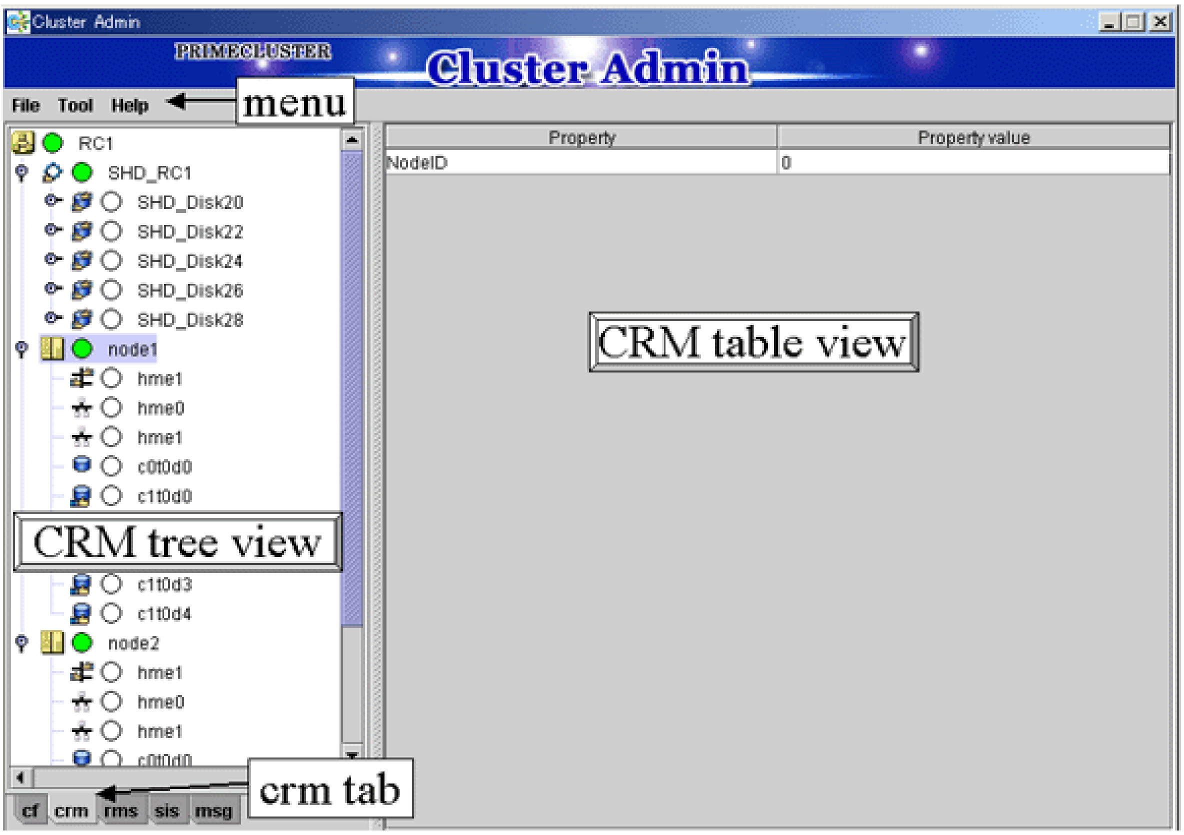
See

For details, see "5 GUI administration" in "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

7.1.2 CRM Main Window

When you monitor the state of the PRIMECLUSTER system, the CRM main window allows you to view the states of the resources managed by the cluster resource management facility and identify failed hardware.

This section describes resource states and operation methods related to hardware resources.



See

The CRM main window is a screen of the cluster resource monitoring facility. See "crm" in "4.4.3 Cluster Admin Functions."

7.1.2.1 Displayed Resource Types

This section shows the resources associated with the hardware resources that are displayed in the CRM main window.

The detailed resource information lists the icons that are used in the CRM main window.

7.1.2.1.1 Resource Icons

This section describes the icons that are displayed in the CRM tree view.

Items that are related to resources under shared resources are displayed with overlapping  .

Icon	Resource
	Shared resource
	DISK class managed by Global Disk Services
	Local disk Multipath disk
	Shared disk device Multipath disk
	IP address MAC address Network interface Takeover node name
	Node resource of a switching line used in a line switching unit Shared resource of a line switching unit Shared resource of a switching line used in a line switching unit
	Takeover network
	Resource that is not a multi-tiered resource under a shared resource Resource that is not related to a shared resource under a node and is not a multi-tiered resource
	Cluster
	Node

7.1.2.1.2 Resource States

In the CRM main window, the displayed icon types and states differ for each resource class.
For each resource, an icon that indicates a failure (OFF-FAIL or ON-FAILOVER) is displayed if there is a fault in a subordinate resource.
In the CRM tree view, expand the icons sequentially from cluster icon to node icons to subordinate resource icons, and identify the faulted resource. If the resource state is OFF-FAIL or ON-FAILOVER, see "7.4 Corrective Actions for Resource Failures," and take corrective actions.
If the resource state is ON, OFF-STOP, or UNKNOWN, you don't need to take any corrective actions.

Cluster states

The following cluster states are displayed.

Icon	Icon color	Outline	Details
	Green	ON	All nodes and shared resources are operating normally.
	Red	OFF-FAIL	One of the nodes in the state other than the ON state, or a shared resource is in the OFF-FAIL state.

Node states

The following node states are displayed.

Icon	Icon color	Outline	Details
	Green	ON	The node has been started normally.
	Green with vertical red lines	ON-FAILOVER	One of the resources under the node is in the Faulted state.
	Blue	OFF-STOP	The cluster resource management facility is stopped.
	Red	OFF-FAIL	A failure has occurred in the node.
	White	UNKNOWN	The node has not been monitored or controlled.

If a failure occurs in the node, the following icons are displayed:

Icons	Outline	Details
	Error	Immediate replacement is necessary.
	Warning	Life replacement or preventive replacement is necessary.

If either icon is displayed, see the "*Machine Administration Guide*," and take corrective actions.

Note

- If a node is LEFTCLUSTER in CF, it becomes ON in the CRM tree view.
- After you execute the "clinitreset(1M)" command to initialize the resource database, the resource states displayed in the CRM main window differ according to the actual resource states. Execute "clinitreset(1M)", reboot the nodes, and then close the browser and redisplay the screen. Execute the "clinitreset(1M)" command after removing a cluster application.

Shared resource states

The following states are displayed for shared resources.

Icon	Icon color	Outline	Details
	Green	ON	The sub-resource under the shared resource is ON, OFF-STOP, or UNKNOWN.
	Red	OFF-FAIL	One of the sub-resources under the shared resource is OFF-FAIL.

Other resource states

The following states are displayed for other resources.

Icon color	Outline	Details
Green	ON	The resource is operating normally.
Green with vertical red lines	ON-FAILOVER	The resource is operating normally, but some devices or resources that are multiplexed and managed internally are in the Faulted state.
Blue	OFF-STOP	The resource has been stopped normally.
Red	OFF-FAIL	The resource is in the Faulted state.
White	UNKNOWN	The resource has not been monitored or controlled.

7.1.2.1.3 Operations

You can perform the operations described below from the CRM main window.
In the table below, "Selection resource" is the resource class name of the selectable resource. For details on resource class names, see "7.1.2.2 Detailed Resource Information."

Table 7.1 Operations of the CRM main window

Feature	Operation method		Target group
	Menu	Selection resource	
Build CRM resource database	Tool - Initial setup	None (*1)	wvroot clroot
Configure automatically	Tool - Automatic configure	Cluster	wvroot clroot
Run patrol diagnosis immediately	Tool - Diagnosis (*2)	DISK (*3) Ethernet (*3) MPDisk(*3)	wvroot clroot cladmin
Request Resource activation	Tool - Start	SDX_DC (*4)	wvroot clroot cladmin
Request Resource deactivation	Tool - Stop	SDX_DC (*4)	wvroot clroot cladmin
Display machine information	Tool - Machine Information	Node (*5)	wvroot cladmin clmon
Exit Cluster Admin screen	File - Exit	All No selection	All
View Help	Help - Content (*6)	All No selection	All
View version	Help - About	All No selection	All

*1 Set Initial Configuration menu can be selected only if the resource database has not been set up. This menu item is not displayed in the pop-up menu.

*2 This feature is enabled only for the resources for which the patrol diagnosis has been set.

*3 Only the resources of the DISK, Ethernet, and MPDisk classes, which have been registered in hardware where the patrol diagnosis facility runs, are enabled. For the hardware settings, see "6.9 Setting Up Patrol Diagnosis."

*4 Only the disk resources that are registered to Global Disk Services are enabled.

*5 This feature is enabled only if the Enhanced Support Facility is installed.

*6 Help for the CRM main window is displayed with a separate browser from the browser that displays Help for CF, and RMS.

Note

- Only available menus are displayed in the pop-up menu.
- If there are no available menu items for the resource selected from the CRM tree view, "None" is displayed in the pop-up menu. Selecting this item will not initiate any operation.
- For information about user groups, see "4.2.1 Assigning Users to Manage the Cluster."

Initial setup

Select this item to set up the resource database to be managed by the cluster resource management facility. Select *Tool* -> *Initial setup* to display the Initial Configuration Setup screen. The initial configuration setup cannot be operated simultaneously from multiple clients. See "5.1.3.1 Initial Configuration Setup."

Automatic configure

Select this item to register the devices connected to the system to the resource database. Automatic configuration cannot be operated simultaneously from multiple clients. See "5.1.3.1 Initial Configuration Setup."

Diagnosis

Select this item to diagnose the selected resource. Execute this diagnosis operation to check whether faulted hardware has been properly recovered after hardware recovery is performed. See "7.4.2 Corrective Action when Patrol Diagnosis Detects a Fault."

- *Yes* button

  Executes resource diagnosis.

- *No* button

  Does not execute resource diagnosis.

Start

This menu item activates the selected resource. The start operation is executed during maintenance work. If the selected resource is registered to a cluster application, the start operation can be executed only when that cluster application is in the Deact state. Use the RMS main window to check the cluster application state.

Note

- After completing the maintenance work, be sure to return the resource that you worked on to its state prior to the maintenance.
- If the resource that was maintained is registered to a cluster application, be sure to stop the resource before clearing the Deact state of the application.

- *Yes* button

  Executes resource start processing.

- *No* button

  Does not execute resource start processing.

Stop

This menu item deactivates the selected resource. The stop operation is executed during maintenance work. If the selected resource is registered to a cluster application, the startup operation can be executed only when that cluster application is in the Deact state. Use the RMS main window to check the cluster application state.

Note

- After completing the maintenance work, be sure to return the resource that you worked on to its state prior to the maintenance.

- If the resource that was maintained is registered to a cluster application, be sure to stop the resource before clearing the Deact state of the application.

- *Yes* button

  Executes resource stop processing.

- *No* button

  Does not execute resource stop processing.

Machine Information

This menu item displays the machine management screen of the selected node.
This menu item can be used only if Enhanced Support Facility (ESF) is installed.
For details on the machine management screen, see the "*Machine Administration Guide*."

Note

If a message is displayed during operating at the CRM main window and the frame title of the message dialog box is "Cluster resource management facility," then see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in "PRIMECLUSTER Messages."

7.1.2.2 Detailed Resource Information

This section describes the resource attributes that are defined in the CRM main window.

Table 7.2 Detailed resource information

Icon/ resource class name	Attributes	Meaning/attribute value (Top: Meaning, Bottom: Attribute value)
Node	NodeID	Node identifier number.
		Node identifier number (0 to 127).
SH_SWLine	lsu_mask	Mask information for the LSU used by the line switching unit
		Logical sum of the mask information for the LSU used by the line switching unit LSU machine information LSU00 0x0001 LSU01 0x0002 LSU02 0x0004 ... LSU14 0x4000 LSU15 0x8000
SH_SWU	lsu_num	Number of switchover units installed in the line switching unit
		4, 16
	addr	RCI device address of the line switching unit
		RCI device address
SWLine	Port	Port number of the LSU (group) used by the line switching unit
		0, 1

Icon/ resource class name	Attributes	Meaning/attribute value (Top: Meaning, Bottom: Attribute value)
DISK	Disk_Attr	This item indicates the physical connection mode and usage mode of a disk that can be used from the cluster system.
		LOCAL Local disk that can be accessed only from one node **SHD_DISK** The disk is physically shared, but the usage mode (shared disk or switchover disk) is not specified. **SHD_SHARE** Shared disk that can be accessed from multiple nodes **SHD_SWITCH** Switching disk that is used exclusively between two nodes **Blank** The disk connection or usage mode is not set up.
MPDisk	Disk_Attr	This item indicates the physical connection mode and usage mode of a disk that can be used from the cluster system
		LOCAL Local disk that can be accessed only from one node **SHD_DISK** The disk is physically shared, but the usage mode (shared disk or switchover disk) is not specified. **SHD_SHARE** Shared disk that can be accessed from multiple nodes **SHD_SWITCH** Switching disk that is used exclusively between two nodes **Blank** The disk connection or usage mode is not set up.
	path_stat	This item indicates the stage of each path of a multipath disk.
		The state value of each path in the multipath disk is indicated in sequence for up to eight paths starting from path 0. The values are shown with the format below. (The values in parenthesis are used when there are five or more paths.) **active** (A) Active **standby** (B) Standby **discon** (D) Low-order driver is disconnected. **fail** (F) Failed **stop** (S) Stopped because a path from another node to the same controller is faulted or is undergoing maintenance
SHD_DISK, SHD_MPDisk	Disk_Attr	This item indicates the physical connection mode and usage mode of a disk that can be used from the cluster system.
		SHD_DISK The disk is physically shared, but the usage mode (shared disk or switchover disk) is not specified. **SHD_SHARE** Shared disk that can be accessed from multiple nodes **SHD_SWITCH** Switching disk that is used exclusively between two nodes
SDX_DC, SDX_SHDDC	Disk_Attr	This class indicates the physical connection mode and usage mode of a GDS-managed disk class that can be used from the cluster system.
		SHD_DISK The disk is physically shared, but the usage mode (shared disk or switchover disk) is not specified. **SHD_SHARE** Shared disk class that allows access from multiple nodes **SHD_SWITCH** Switching disk class for exclusive use between two nodes
Ethernet	node_name	This item indicates the name of the node in which this LAN board is set.
		The node name is set.

Icon/ resource class name	Attributes	Meaning/attribute value (Top: Meaning, Bottom: Attribute value)
	org_mac	This item indicates the MAC address that the operating system has assigned to this LAN board.
		The MAC address is in the format 00:11:22:33:44:55.
	WebView	This item indicates internal information used in the cluster.
		Undefined
 SHD_Host	ip_addr	This item indicates the takeover IP address.
		If the takeover IP address information is IPv4, this item is set in the format XXX.XXX.XXX.XXX. If IP address takeover has not been set, this item is blank. If the takeover IP address information is IPv6, the icon or the resource is not displayed.
	mac_addr	This item indicates the takeover MAC address.
		The takeover MAC address information is set in the format 00:11:22:33:44:55. If MAC address takeover has not been set, this item is blank.
	node_name	This item indicates the takeover node name.
		The takeover node name is set. If node name takeover has not been set, this item is blank.
	LanDev_Rid	If a takeover MAC address is used, this item indicates the resource ID of LanDev (maintenance information for customer support representative).
		If a takeover MAC address is used, the resource ID of LanDev is set.

GDS: Global Disk Services

7.1.3 RMS Main Window

The RMS main window consists of the following elements:

- RMS tree
- Configuration information or object attributes
- Switchlogs and application logs

7.1.3.1 RMS Tree

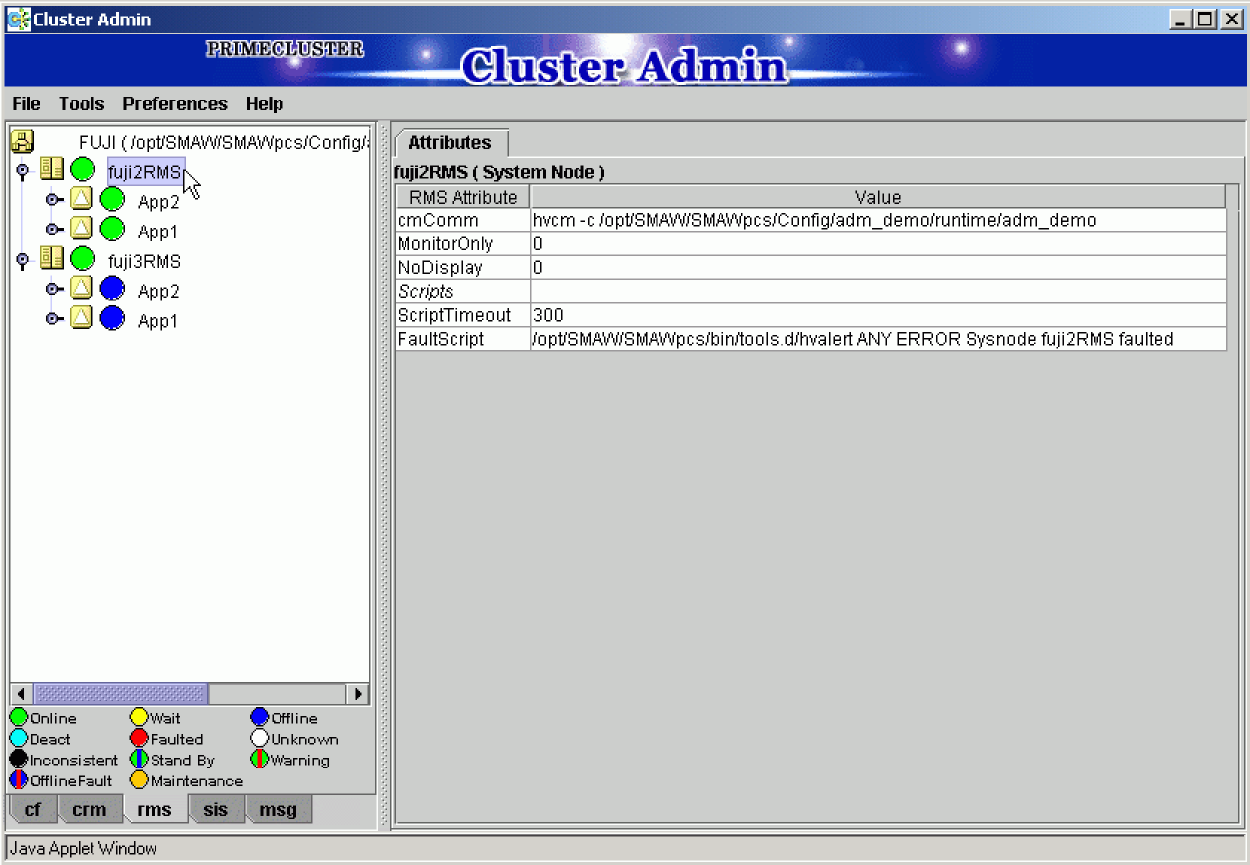
The RMS tree displays the configuration information of the cluster in a hierarchical format. The tree has the following levels:

- Root of the tree - Represents the cluster.
- First level - Represents the system nodes in the cluster.
- Second level - Represents the userApplication objects running on each of the system nodes.
- Third level - Represents the sub-applications (an aggregation of objects if any exist).
- Fourth level - Represents the resources required for each of the sub-applications.

If an application has sub-applications, the fourth level represents resources used by that sub-application. If an application does not have sub-applications, then the third level represents all the resources used by userApplication.

Dependency relationships between cluster applications are indicated by controller objects in the RMS tree.

Figure 7.1 RMS main window



Meanings of object icons

Icon	Meaning
	Represents the cluster.
	Represents a node.
	Represents a parent object (cluster application) that has a child object.
	Represents a child object (cluster application or resource).
	Represents a leaf object (cluster application or resource). A leaf object is an object that cannot have a child object.
	Represents a controller object (cluster application). This object controls an object of another cluster application.

Meanings of state display icons

On the right side of the object icons shown above is displayed a color-coded circle that indicates the state of each object. This section describes the meaning of the colored circles (state display icons).

Information

State display icons are not displayed in cluster icons. Instead, the RMS cluster table can be displayed. For details, see "7.3.3 Concurrent Viewing of Node and Cluster Application States."

State indication of node

The state icons that are displayed in nodes are shown below.

Icon	Icon color	Outline	Details
Node	Green	Online	Node is enabled and ready for use.
	Blue	Offline	Node is enabled but RMS is disabled.
	Red	Faulted	Node is disabled. If the node is shutdown normally, "Shutdown" appears in the SysNode state detailed information (as the value of the StateDetails attribute). If the node is shutdown abnormally, "Killed" appears in the SysNode state detailed information (as the value of the StateDetails attribute).
	Yellow	Wait	Node is undergoing a state transition.

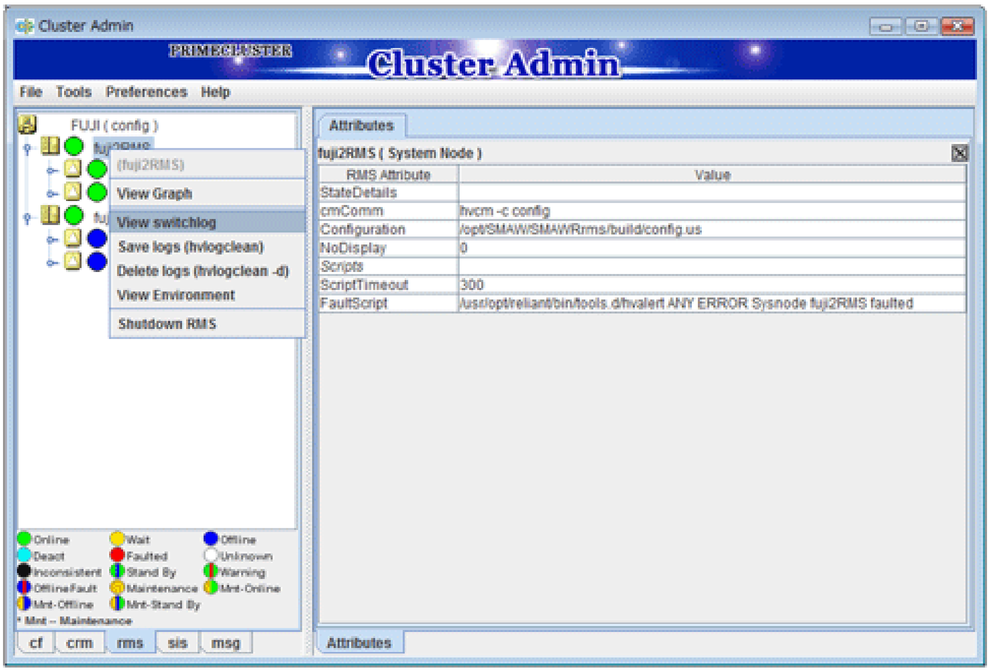
State indication of other objects

The following state icons are displayed with object icons that fall under parent objects.

Icon	Icon color	Outline	Details
Parent object Child object Controller object Leaf object	Green	Online	Object is enabled and ready for use.
	Green with vertical red lines	Warning	Object has exceeded some warning threshold.
	Blue	Offline	Object is disabled and should not be used.
	Red	Faulted	Object encountered an error condition.
	White	Unknown	Monitoring and control are not being executed for the object.
	Yellow	Wait	Node is undergoing a state transition.
	Sky blue	Deact	Node is deactivated because of maintenance, etc.
	Black	Inconsistent	Node state is inconsistent.
	Green with vertical blue lines	Stand By	Object is in such a state that it can be quickly brought Online when needed
	Blue with vertical red lines	OfflineFault	Object is Offline, but a fault has occurred before and is not cleared yet.
	Orange	Maintenance	Object is in maintenance mode.
	Orange in the left and green in the right	Maintenance-Online	Object is in maintenance mode and must be Online when exiting maintenance mode.
	Orange in the left and blue in the right	Maintenance-Offline	Object is in maintenance mode and must be Offline when exiting maintenance mode.
	Orange in the left and green in the right with vertical blue lines	Maintenance-Stand By	Object is in maintenance mode and must be Stand By when exiting maintenance mode.

Pop-up menu

If you right-click an object in the RMS tree, a pop-up menu lists the operations for that object. You can also use this menu for monitoring the state.



Note

- The following icons may be displayed in the userApplication object or the gResource object.

   : This icon is displayed at the right side of the userApplication object state icon. It means that only some resources under the userApplication are started. For details, see "7.2.3 Resource Operation."

   : This icon is displayed at the right side of the gResource object. It means that a resource fault occurred in the past. For details, see "8.3.5 Fault Traces of Resources" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

   : This icon is displayed at the right side of the userApplication object state icon. It means that the status of some resources in the userApplication has changed from the status just before the start of maintenance mode. To exit the maintenance mode, all the resource status in userApplication must be changed back to the original status just before the start of maintenance mode. For more information, refer to "7.2.2.6 Entering maintenance mode for Cluster Application."

-  : Though this icon indicates that the resource fault occurred in the past, it has nothing to do with the current state of the resource. For this reason, this icon is subsequently shown as "Fault Traces of Resources."
  If you want to check the current state of the resource, check the resource object state.

  This icon is hidden in any of the following cases:

  - After executing the Online processing of the resource.
  - After clearing the fault trace resources manually. For details, see "7.2.3.3 Clearing Fault Traces of Resources." Even when the icon is shown, unlike the Faulted state of cluster applications(*), there is no influence on switchover activities of cluster applications. Therefore, if you do not need to display the icon, clear it manually.

  (*) When the cluster application is in the Faulted state, you need to clear the Faulted state if you specify the cluster application for switchover again.

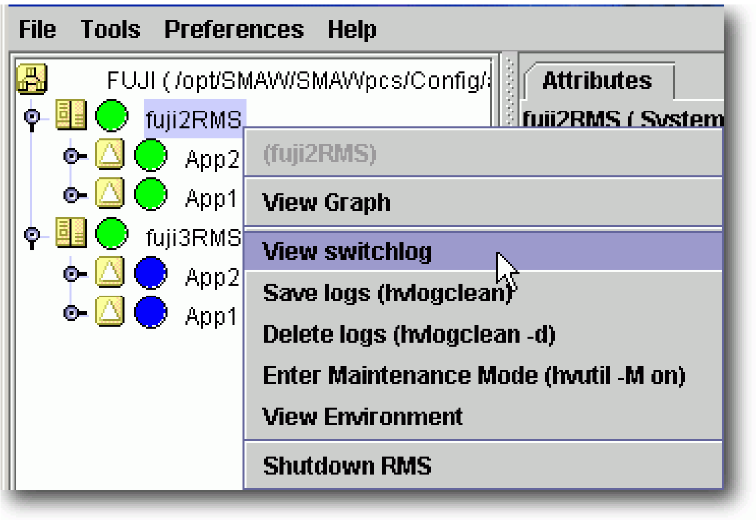
7.1.3.2 Configuration information or object attributes

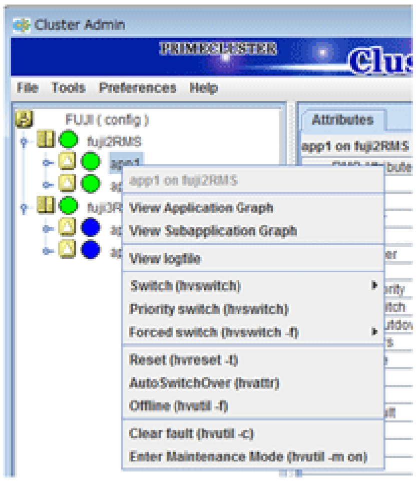
View the configuration information for the individual objects by left-clicking the object in the RMS tree. The properties are displayed in a tabular format on the right-hand side panel of the RMS main window.

7.1.3.3 Switchlogs and application logs

Each node has a log file referred to as the switchlog. In this file, switchover requests and node failures are recorded. The switchlog is displayed in a tab on the right-side panel.

Display the application log by right-clicking on an application on the RMS tree and choosing *View Application Log*.





7.2 Operating the PRIMECLUSTER System

7.2.1 RMS Operation

To monitor RMS, RMS needs to be activated.

To stop multiple nodes at the same time, you must stop the user applications and RMS.

Note

To stop two or more nodes at the same time, it is necessary to first stop RMS.

Note that the user application is also stopped when you stop RMS. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

The sections below explain how to start and stop RMS.

7.2.1.1 Starting RMS

This section explains how to start RMS.

Operation Procedure:

From the top screen of Web-Based Admin View, open Cluster Admin according to the following procedure:

1. Select Global Cluster Services.
2. Click *Cluster Admin* to switch to the cluster menu.
3. Select the *rms* tab.
4. Start RMS.
    1. Use the *Tools* pull-down menu of the RMS main window, and choose *Start RMS*.
    2. Choose *all available nodes* or *one node from the list* from the pop-up window.

See

See "8.1.1 Starting RMS" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

7.2.1.2 Stopping RMS

This section explains how to stop RMS.

Operation Procedure:

1. Use the *Tools* pull-down menu of the RMS main window, and choose *Shutdown RMS*.
2. Choose *all available nodes* or *one node from the list* from the pop-up window.
    - When selecting *all available nodes*:

      Select Stop all Apps or Keep local Apps.
    - When selecting *one node from the list*:

      Choose one of Stop all Apps, Keep local Apps, or Forced shutdown.

See

See "8.1.3 Stopping RMS" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

7.2.2 Cluster Application Operations

This section explains how to change the operation state of the PRIMECLUSTER system. The operations include starting and stopping a cluster application.

7.2.2.1 Starting a Cluster Application

The procedure for starting a cluster application is described below.

Operation Procedure:

1. On the RMS tree in the RMS main window, right-click the cluster application to be started, and select *Online* from the pop-up menu that is displayed.

   The cluster application will start.

Information

You can also display the pop-up menu by right-clicking the target icon in an RMS graph or the RMS cluster table. For details on RMS graphs and the RMS cluster table, see "7.3.5 Viewing Detailed RMS Object Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

Note

To start a cluster application manually, check that the cluster application and resources under it are stopped on all nodes other than the node on which the cluster application is to be started. You can check whether they are stopped by the Offline or Standby state. With the state other than Offline or Standby, they may be running. In this case, stop them and then start the cluster application on the target node.

7.2.2.2 Stopping a Cluster Application

The procedure for stopping a cluster application is described below.

Operation Procedure:

1. On the RMS tree in the RMS main window, right-click the cluster application to be stopped, and select *Offline* from the displayed pop-up menu.

   The cluster application will stop.

Information

You can also display the pop-up menu by right-clicking the target icon in an RMS graph or the RMS cluster table. For details on RMS graphs and the RMS cluster table, see "7.3.5 Viewing Detailed RMS Object Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

7.2.2.3 Switching a Cluster Application

The procedure for switching a cluster application is described below.

Operation Procedure:

1. Right-click on the application object and select the Switch menu option.

   A pull-down menu appears listing the available nodes for switchover.

2. Select the target node from the pull-down menu to switch the application to that node.

Information

You can also display the pop-up menu by right-clicking the target icon in an RMS graph or the RMS cluster table. For details on RMS graphs and the RMS cluster table, see "7.3.5 Viewing Detailed RMS Object Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

7.2.2.4 Bringing Faulted Cluster Application to Available State

The procedure for bringing a Faulted cluster application to available state is described below.

Operation Procedure:

1. Right-click on the cluster application object in the RMS tree, and select *Clear Fault*.

Information

You can also display the pop-up menu by right-clicking the target icon in an RMS graph or the RMS cluster table. For details on RMS graphs and the RMS cluster table, see "7.3.5 Viewing Detailed RMS Object Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

7.2.2.5 Clearing the Wait State of a Node

The procedure for clearing the Wait state of a node is described below.

Operation Procedure:

1. Check that the node in the Wait state has been stopped. If not, stop the node manually.
2. Check that the CF state is DOWN in the CF main window. If the CF state is LEFTCLUSTER, clear LEFTCLUSTER in the CF main window and make sure the node state is changed from LEFTCLUSTER to DOWN.
3. If the Wait state of the node has not been cleared after performing 2, right-click on the system node in the RMS graph and select the "*Clear Wait & shutdown (hvutil -u)*" from the menu.

Note

If you clear the Wait state of a system node manually, RMS and CF assume that you have already checked that the target node had stopped. Therefore, if you clear the Wait state when the node has not been stopped, this may lead to the data corruption.

Information

You can also display the pop-up menu by right-clicking the target icon in an RMS graph or the RMS cluster table. For details on RMS graphs and the RMS cluster table, see "7.3.5 Viewing Detailed RMS Object Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

7.2.2.6 Entering maintenance mode for Cluster Application

The maintenance mode is a specific mode to temporarily restrict a specified cluster application switching.

If a cluster application enters maintenance mode, it cannot be switched.

Note that cluster nodes and resources are monitored during maintenance mode. In this case, when the resource state is changed, the resource state of the cluster application that is viewed on the RMS tree is also changed.

If the state of a cluster application resource has changed while in maintenance mode, since switching is not carried out, it enters a state in which consistency with the resource registered in the cluster application is collapsed. (Example: Some resources are in the Offline state while others are in the Online state.) Therefore, before exiting the maintenance mode, it is necessary to revert the resource state of the cluster application to the same state as before starting the maintenance mode.

For using maintenance mode, see "8.4 Using maintenance mode" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide".

Note

Please note the following for using maintenance mode.

- Perform maintenance mode to the cluster application of the standby operation containing resources for which the maintenance is necessary.

- Since the resources for which the maintenance is necessary during the operation are not contained, it is not necessary to make the cluster application of the scalable operation into maintenance mode.

- To start maintenance mode, a cluster application must be in the Online, Standby, or Offline state.

- To exit maintenance mode, a cluster application and each resource must be returned in the same state before starting maintenance mode.

- Do not stop RMS or the system with cluster applications in maintenance mode. Be sure to exit maintenance mode of all cluster applications before stopping RMS or the system.

- Use maintenance mode only when applicable products are specified in the environment that uses PRIMECLUSTER products.

- When the cluster application that includes Cmdline resource that sets the NULLDETECTOR flag is in maintenance mode, the script that was set to the Cmdline resource must correspond to the maintenance mode. For details, see "6.12.2.1.4 Notes When Setting the NULLDETECTOR Flag."

For details, see "8.4.2 Maintenance mode operating notes" or "3.1.7.1 Restrictions during maintenance mode" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

7.2.3 Resource Operation

This section describes how to set resources Online/Offline individually.

Note

- It is assumed that this function is used when you check the behavior of resources during cluster application configuration. Do not perform any business operations while cluster applications are partially Online.

  If you want to carry out business operations without starting a resource, delete that resource from the cluster application. For instructions on deleting a cluster application, see "10.5 Deleting a Resource."

  After using this function, restart the application by the following procedure before starting any business operation, and make sure that all resources become Online.

  1. Stop userApplication.

     ```
     # hvutil -f userApplication
     ```

  2. Check that all resources controlled by userApplication are stopped.

     ```
     # hvdisp -a
     ```

  3. Start userApplication.

     ```
     # hvswitch userApplication SysNode
     ```

  4. Check that all resources controlled by userApplication are started.

     ```
     # hvdisp -a
     ```

- Stop cluster applications in scalable operation whenever you start/stop a resource with scalable configuration individually. After that, execute the operation on the cluster applications in standby operation that constitute the cluster applications in scalable operation.

- For details, see "8.3 Managing resources" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

7.2.3.1 Starting Resources

The following describes how to start resources.

Operation Procedure:

1. On the RMS tree in the RMS main window, right-click the resource to be started, and select [Resource Online] from the pop-up menu.

   The resource will start.

2. Click "Yes" when the confirmation pop-up is displayed.

Information

Also, the pop-up menu can be displayed by right-clicking on the icon of the RMS graph. For instructions on the RMS graph, see "7.3.5 Viewing Detailed RMS Object Information."

7.2.3.2 Stopping Resources

The following describes how to stop resources.

Operation Procedure:

1. On the RMS tree in the RMS main window, right-click the resource to be stopped, and select [Resource Offline] from the pop-up menu.

   The resource will stop.

2. Click "Yes" when the confirmation pop-up screen is displayed.

Information

Also, the pop-up menu can be displayed by right-clicking on the icon of the RMS graph. For instructions on the RMS graph, see "7.3.5 Viewing Detailed RMS Object Information."

7.2.3.3 Clearing Fault Traces of Resources

If a resource fault occurred in the past, the icon ( ) is displayed in the right side of the state icon of the failed RMS resource.

Check the state of the failed resource first, and then clear the fault trace according to the procedure below.

Operation procedure:

1. Right-click the failed resource in the RMS tree of the RMS main window, and then select [Clear fault trace (hvutil -c)] from the pop-up menu.

2. A pop-up confirmation dialog appears. Click "Yes."

Point

In addition to the hvutil -c command can clear the fault trace, it can be also cleared automatically when the resource becomes Online next time.

Information

For details on the icon of fault traces of resource, see "7.1.3.1 RMS Tree."

For the method of displaying fault traces of resources, see "8.3.5 Fault Traces of Resources" in the "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

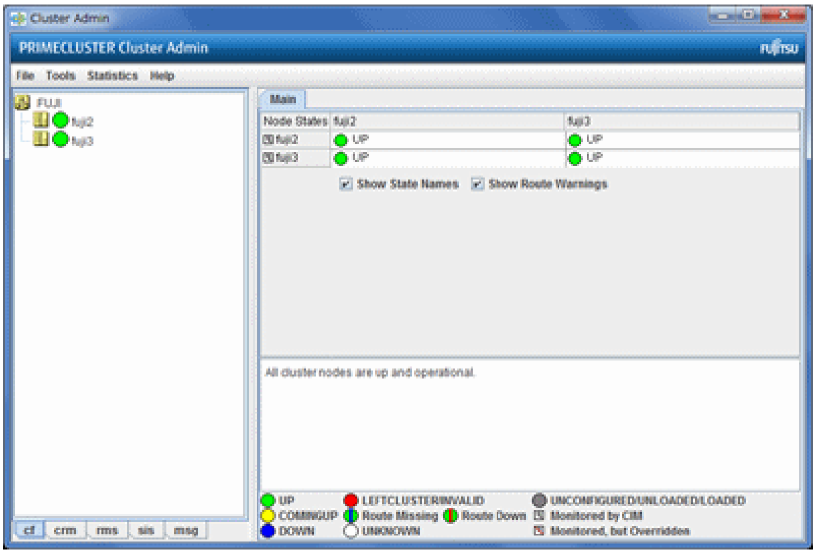
The pop-up context menu can be displayed by right-clicking the icon of the RMS graph. For details on the RMS graph, see "7.3.5 Viewing Detailed RMS Object Information."

7.3 Monitoring the PRIMECLUSTER System

This section describes how to monitor the state of a cluster application or node from the Cluster Admin screen.

7.3.1 Monitoring the State of a Node

Click an icon of the cluster name on the CF tree. The node state will appear in the right panel.

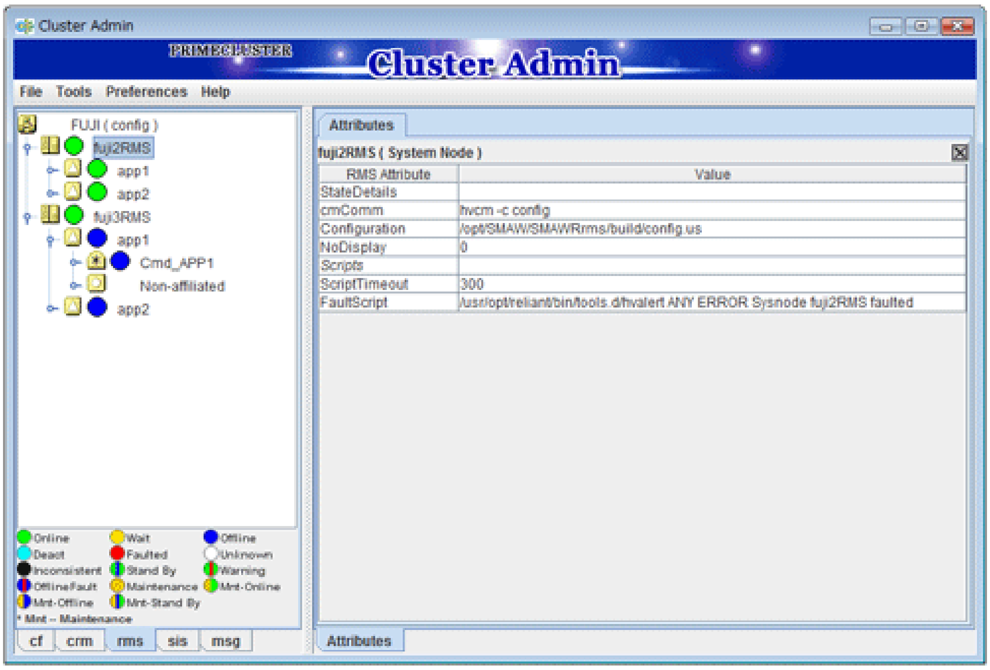


The node indicates one of the following states:

CF state		Description
Green	UP	The node is up and part of this cluster.
Yellow	COMINGUP	The node is joining the cluster.
Blue	DOWN	The node is down and not in the cluster.
Red	LEFTCLUSTER / INVALID	The node has left the cluster unexpectedly, probably from a crash. To ensure cluster integrity, it will not be allowed to rejoin until marked DOWN.
Green with vertical blue lines	Route Missing	Some cluster interconnects have not been recognized on startup.
White	UNKNOWN	The reporting node has no opinion on the reported node.
Green with vertical blue lines	Route Down	Some cluster interconnects are not available.
Gray	UNCONFIGURED /UNLOADED /LOADED	This icon shows any of the following status: - CF has not been set. - The CF driver has not been loaded. - The CF driver has been loaded but CF is not started.

7.3.2 Monitoring the State of a Cluster Application

When you display a cluster application in an RMS tree, the cluster application state appears to the right of the icon. The right panel displays the properties of the cluster application.



The cluster application shows one of the following states:

- Online
- Wait
- Offline
- Deact
- Faulted
- Unknown
- Inconsistent
- Stand By
- Warning
- OfflineFault
- Maintenance
- Maintenance-Online
- Maintenance-Offline
- Maintenance-Stand By

See

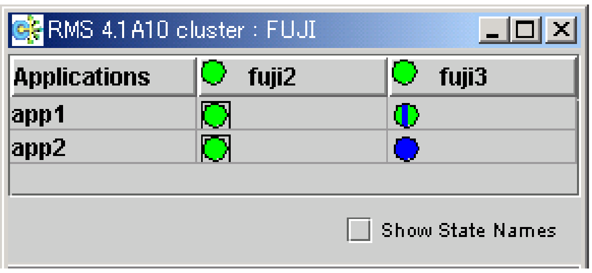
See "State indication of other objects" in "7.1.3.1 RMS Tree."

7.3.3 Concurrent Viewing of Node and Cluster Application States

To view the states of the nodes and cluster applications concurrently, display the RMS cluster table.

Right-click the cluster icon on the RMS tree, and select *View Cluster Wide Table* from the displayed pop-up menu. The RMS cluster table is displayed as shown below.

RMS clusterwide table



The first line shows the names of the nodes that RMS is managing (fuji2 and fuji3 in the example above). To the left of each node name is a state display icon that shows the state of that node.

The second and subsequent lines show the names of all cluster applications that RMS is managing and the states of those applications. The RMS cluster table enables you to display the states of nodes and cluster applications in one table.

Viewing the RMS Cluster Table

If the background color of the cluster application name is the same as that of the background of the window

It indicates that the cluster application is online.

If the background of the cluster application name is pink

This condition indicates that the cluster application is in the Faulted state and a failure has occurred in one or more SysNode.

If the background of the cluster application name is sky blue

This condition indicates that the cluster application is in the Offline state.

If the state display icon of a cluster application is enclosed in a rectangle

This condition indicates that the node has the highest priority among those nodes that configure the cluster application. If the cluster application is started after creating the cluster application, the node in a rectangle will be in the Online state.

Displaying/hiding state names

Select the *Show State Names* checkbox to display state names to the right of the state display icons.

See

For details on the RMS cluster table, see "7.1 Using the RMS clusterwide table" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

7.3.4 Viewing Logs Created by the PRIMECLUSTER System

There are two types of logs that can be viewed in the PRIMECLUSTER system:

- Switchlog

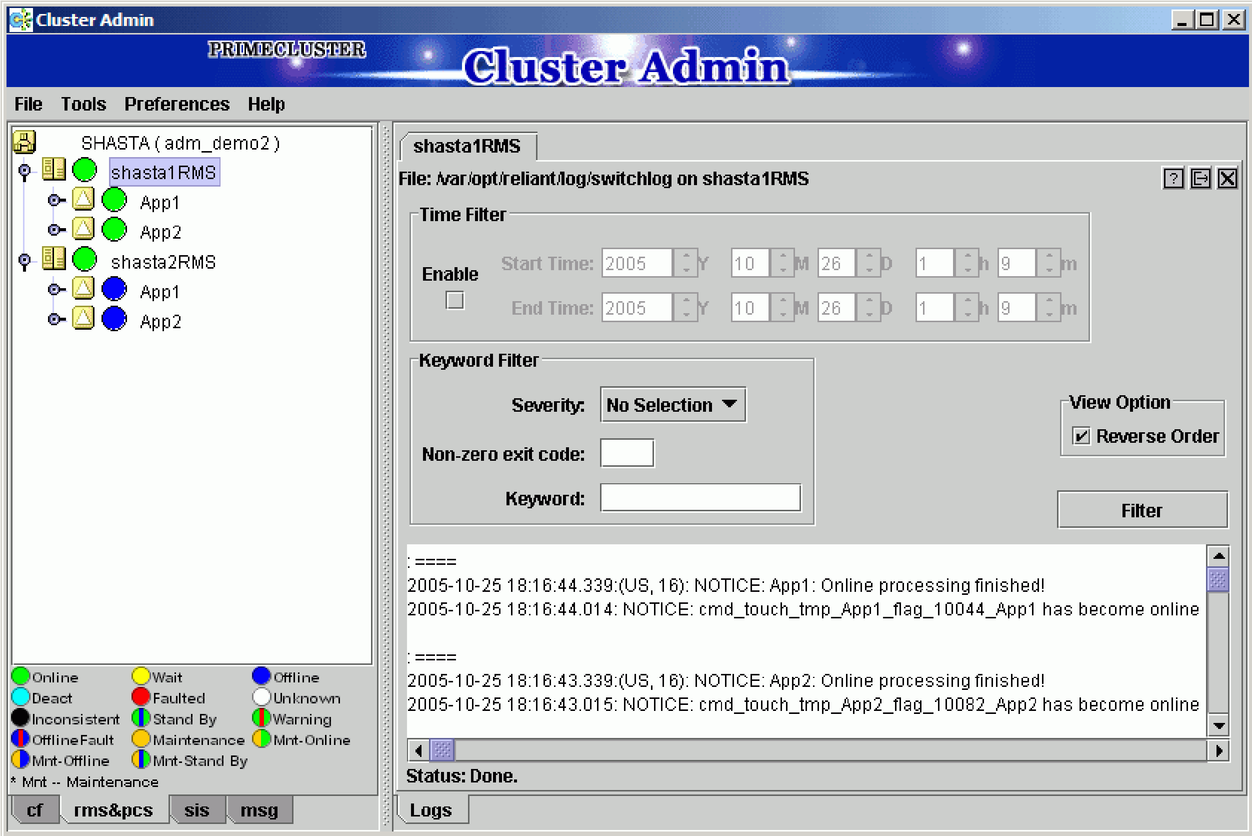
  The switchover requests or failures that occur in nodes are displayed.

- Application log

  The operation log of the cluster application is displayed.

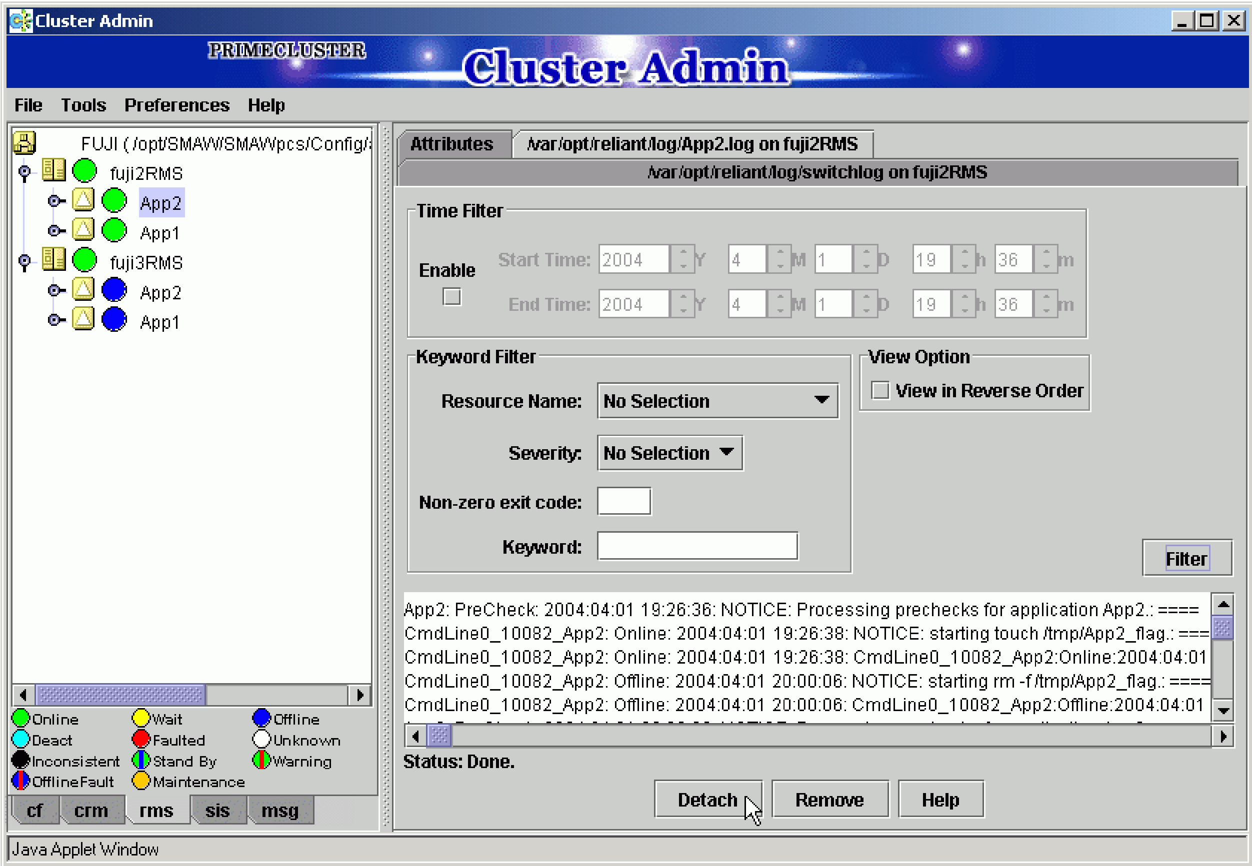
7.3.4.1 Viewing switchlogs

Right-click the system node and select the *View Switchlog* option from the pop-up menu. The switchlog is displayed on the right side of the screen.



7.3.4.2 Viewing application logs

Right-click an application on the RMS tree and choose *View Log File*. The application log for that application will be displayed on the right side of the screen.



Information

The following display formats are enabled for the log. For details, see "7.4 Viewing RMS log messages" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

- Narrow the log by date or keyword.
- Scroll or jump to any entry.
- Search by keyword, date range, error message significance, or exit code other than 0 to exclude unrelated entries.

7.3.5 Viewing Detailed RMS Object Information

Use RMS graphs to display detailed object information for each cluster application.

There are four types of RMS graphs. Each type can be displayed from the menu when you right-click an object on the RMS tree.

- Full graph

  Displays the configuration of the entire cluster system in which RMS is running.

- Application graph

  Shows all objects used by the specified application. You can check the details of the specific object using this graph.

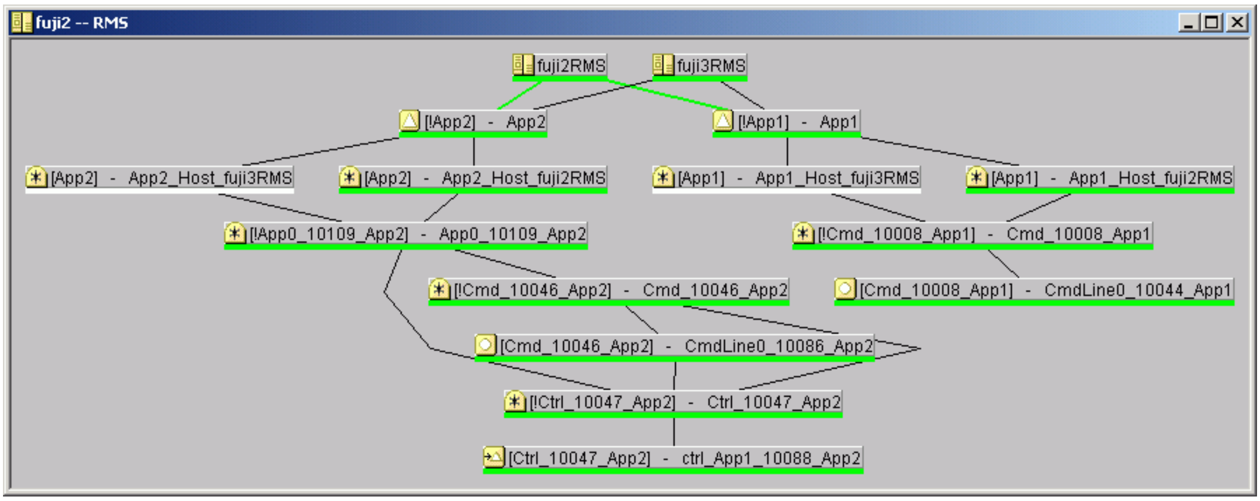
- Sub-application graph

  Lists all sub-applications used by a given application and shows the connections between the sub-applications.

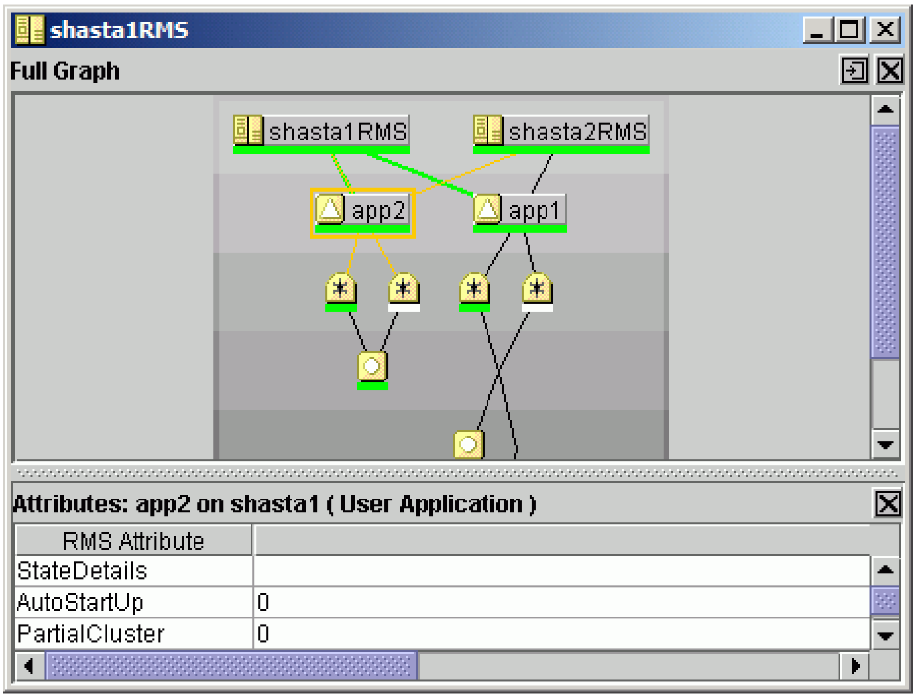
- Composite sub-applications graph

  Shows all sub-applications that the application depends on directly or indirectly.

RMS graphs



Clicking on the object brings up a window with further details such as the object's attributes.

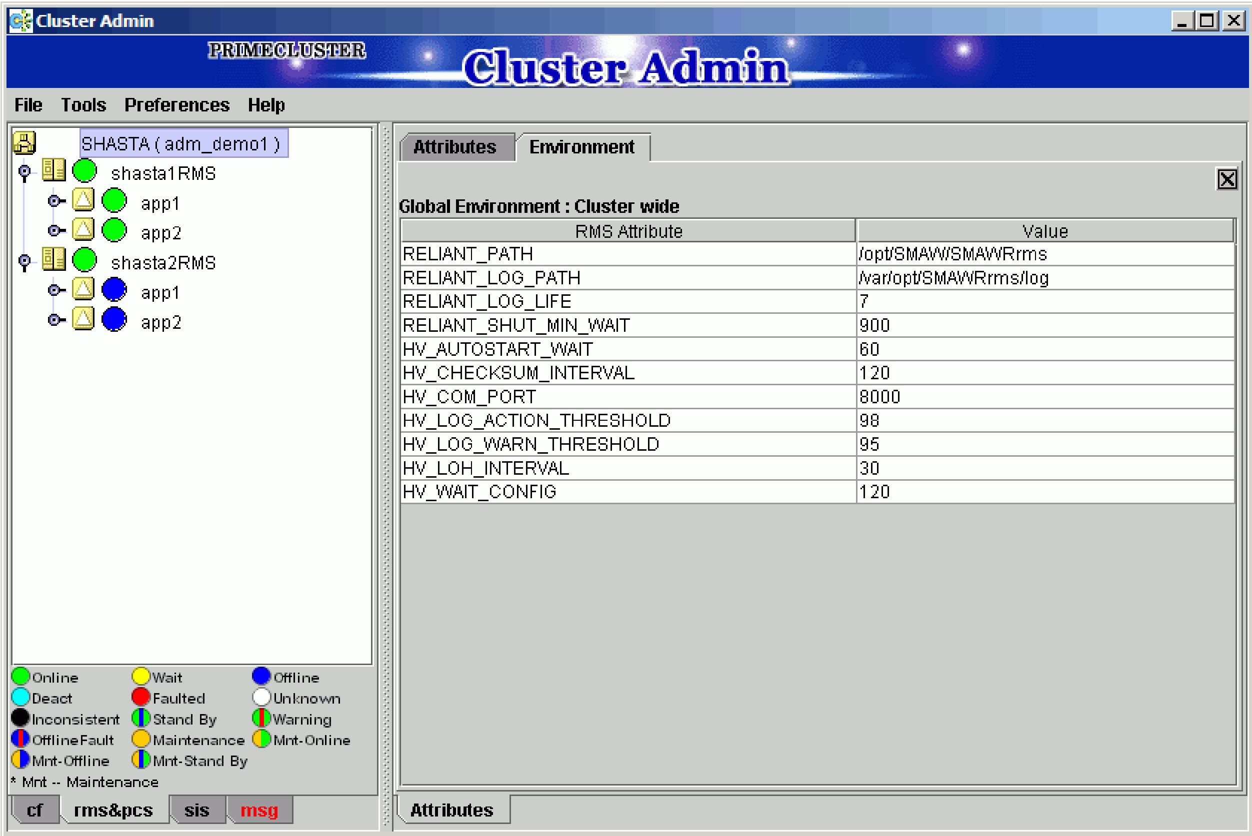


See

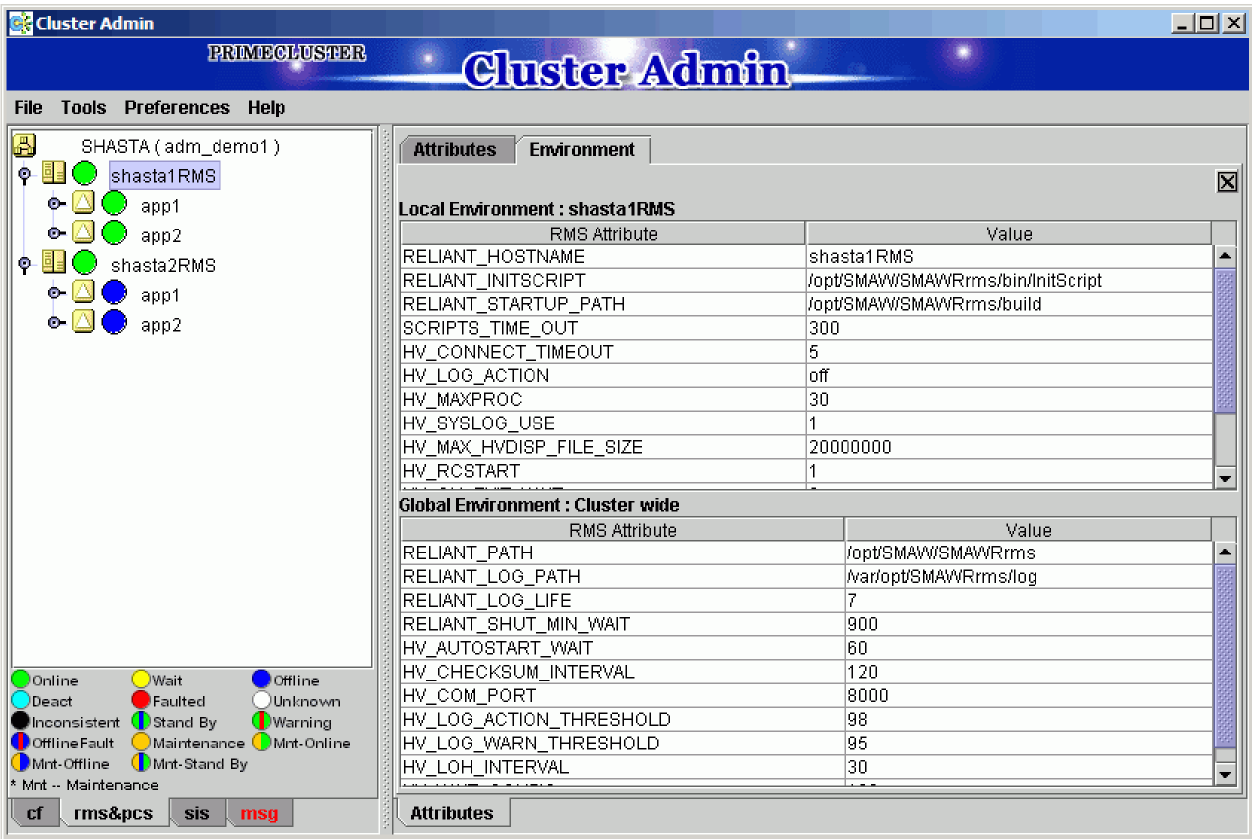
See "7.2 Using RMS graphs" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

7.3.6 Displaying environment variables

Right-click a cluster in the RMS tree window and select *View Environment*. The global variable is displayed.



Right-click a node in the RMS tree, and select *View Environment*. The local and global variables are displayed.



7.3.7 Monitoring Cluster Control Messages

Select the *msg* tab, which is found at the bottom of the tree panel. If a new message was added to the text area since the last time the area was displayed, this tab is displayed in red.

You can clear the message text area or isolate it from the main panel.

7.4 Corrective Actions for Resource Failures

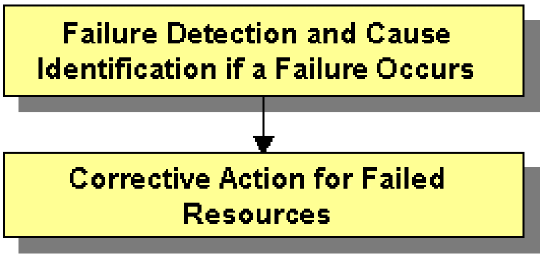
A hardware or software failure might occur while the cluster system is running. If a failure occurs in a resource, a message indicating that a failure occurred is displayed in the Cluster admin pop-up screen. Based on this message, you need to identify the faulted resource using the CF, CRM, and RMS main window and take corrective actions to maintain high availability in the cluster system.

This section describes the actions to be taken for the following cases:

- If the resource state became Faulted.
- If the Patrol Diagnosis Facility detects a failure.

7.4.1 Corrective Action when the resource state is Faulted

This section describes the corrective actions to take when the resource state became Faulted.



7.4.1.1 Failure Detection and Cause Identification if a Failure Occurs

If a failure occurs in a resource, you can use the functions of PRIMECLUSTER and the operating system to detect the failure and identify the faulted resource that caused the failure.

The descriptions given in (a) to (k) below are relevant to the "Failure confirmation features list" given below:

Failure detection

Normally, the RMS main window (b) is used to monitor the cluster applications.

- If a failure occurs in a resource or the system

  Failover of the userApplication or node panic will occur.

  In such a case, you can detect the failure by observing the following conditions:

  - A pop-up message screen (a) is displayed.
  - The color of the icons in the RMS main window (b) changes.
  - A message is output to the msg main window (g), Syslog(j), and the console (k).

- If a warning-level failure occurs in the system

  If a warning-level failure (for example, insufficient disk space or insufficient swap area) occurs in the system, you can detect the failure by observing the following conditions:

  - The node icon in the CRM main window (d) changes.

- A message is output to Syslog(j) and the console (k).

- If RMS fails to start on all the nodes, the userApplication will not start. You can start the userApplication by executing the "clreply" command.
    - By executing the "clreply" command, you can confirm an operator intervention request to which no response has been entered and start up the userApplication by responding to it. For information on the "clreply" command, see the manual pages.
    - The operator intervention request message will be output to Syslog(j) and the console (k). By responding to the operator intervention request message, you can start the userApplication.

  For further details, see "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

Note

If there are multiple operator intervention request messages for which no response has yet been entered, you need to respond to each of them.

In addition, you can use the features described in "Failure confirmation features" to detect the failure.

Cause identification

You can also use the function that detected the failure and the features listed in "Failure confirmation features" below to identify the faulted resource that caused the failure.

Failure confirmation features list

Failure confirmation features		Manual reference
(a)	Message screen	C.3.1 Failed Resource Message
(b)	RMS main window The RMS tree and the RMS cluster table can be used from this screen.	7.1.3 RMS Main Window
(c)	CF main window The CF tree can be used from this screen.	7.1.1 CF Main Window
(d)	CRM main window The CRM tree can be used from this screen. This screen is useful in detecting hardware resource faults.	7.1.2 CRM Main Window
(e)	"Resource Fault History" screen This screen is useful in detecting hardware resource faults.	C.3.2 Resource Fault History
(f)	Current list of resources in which a failure has occurred	C.3.3 Fault Resource List
(g)	MSG main window The cluster control messages can be viewed in this screen. To display this screen, select the msg tab in the Cluster Admin screen.	-
(h)	Application log	7.3.4.2 Viewing application logs
(i)	switchlog	7.3.4.1 Viewing switchlogs
(j)	Syslog	-
(k)	Console Messages that are displayed on the console can be checked. Viewing the "console problem" information on the console can help you identify the fault cause.	PRIMECLUSTER Messages
(l)	Machine management GUI	Machine Administration Guide

Failure confirmation features		Manual reference
(m)	MultiPathDisk view	Multipath Disk Control Load Balance option x.x Guide
(n)	GDS GUI	PRIMECLUSTER Global Disk Services Configuration and Administration Guide

GDS: Global Disk Services

7.4.1.2 Corrective Action for Failed Resources

Take the following steps for failed resources;

1. Correct the faulted resource

   Correct the problem in the failed resource. For details, see "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

   If an error message of patrol diagnosis is displayed, see "7.4.2 Corrective Action when Patrol Diagnosis Detects a Fault."

   "hvdet_sptl" is displayed in the name of the program that outputs the patrol diagnosis message.

   Note

   If you are using an operation management product other than a PRIMECLUSTER product, you may need to take corrective actions prescribed for that product.

   For details, see the manual provided with each operation management product.

   [Examples] Machine Administration, MultiPathDisk view, GDS

2. Recover the cluster application

   At the RMS main window, check the state of the cluster application to which the corrected resource is registered. If the cluster application is in the Faulted state, execute the Fault clear operation.

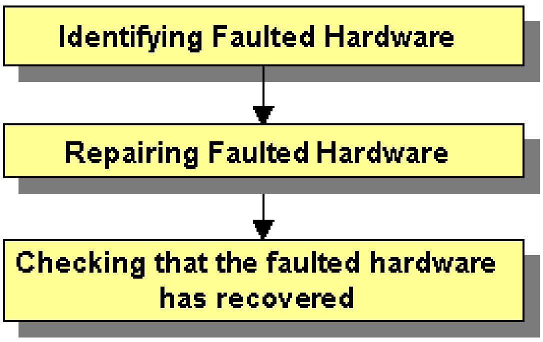
   For details on the Fault clear operation, see "7.2.2.4 Bringing Faulted Cluster Application to Available State."

3. Clear the fault trace of the failure resource

   Clear the fault trace of the failure resource. For more information, refer to "7.2.3.3 Clearing Fault Traces of Resources."

7.4.2 Corrective Action when Patrol Diagnosis Detects a Fault

This section explains the actions to take when the patrol diagnosis facility detects a failure



7.4.2.1 Identifying Faulted Hardware

Use one of the following methods to identify the faulted hardware:

- Message text output to the CRM main window or syslogd(1M)

  See "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in "PRIMECLUSTER Messages."

- CRM main window

  The CRM main window displays the OFF-FAIL state for the faulted hardware. See "7.1.2 CRM Main Window."

- "clgettree(1)" command

  The "clgettree(1)" command displays the OFF-FAIL state for the faulted hardware. See the manual page for clgettree(1).

7.4.2.2 Corrective Action for Faulted Hardware

Correct the faulted hardware according to the operation procedure below.

Operation Procedure:

1. Stop the node to which the faulted hardware is connected.

2. Repair the faulted hardware.

3. Start the node.

   Note

   When a disk unit that is registered with GDS is to be exchanged, follow the steps described in the GDS disk replacement procedure. For information on GDS disk replacement, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

4. Check that the faulted hardware has recovered using one of the following methods:

   1. Use the CRM main window.

   2. Execute the "clgettree(1)" command.

   If the above procedure shows that the fault was not corrected, you need to continue the following procedure:

   1. Execute the diagnosis operation for the faulted hardware from the CRM main window

      Then, use the CRM main window to check whether the fault was corrected. If the fault was corrected, the ON state is displayed.

   2. Execute the "clsptl(1M)" command to initiate the diagnosis operation.

      The "clsptl(1M)" command has two functions. One function allows you to specify a faulted hardware unit and diagnoses only the specified device. The other function runs batch diagnosis of all shared disk units or all network interface cards. If faults occur in multiple hardware units, it is convenient to use the batch diagnosis function.

      - Example in which a faulted shared disk unit is specified and diagnosis is executed:

      ```
      # /etc/opt/FJSVcluster/bin/clsptl -u generic -n c1t4d4
      ```

      - Example in which batch diagnosis is executed for all shared disk units:

      ```
      # /etc/opt/FJSVcluster/bin/clsptl -a DISK
      ```

      Execute the "clgettree(1)" command to check whether the fault was corrected. If the fault was corrected, the ON state will be displayed for the hardware.

5. Bring the Faulted cluster application Online.

   Confirm that the state of the cluster application to which the recovered hardware is registered, either in the RMS main window or with the "hvdisp(1M)" command.
   If the cluster application is Faulted, switch the cluster application from the failed to the active state, either in the RMS main window or with the "hvutil(1M)" command. For information on the procedures related to the CRM main window, see "7.2.2.4 Bringing Faulted Cluster Application to Available State."

   If operator intervention request is enabled, a message will be displayed with the "syslogd(1M)" command and Cluster Admin when RMS is started. By entering a response to this message, you can switch the state of the cluster application from the failed state to active. For information on the setup procedure for operator intervention requests, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

An example of an operator intervention request is shown below. For details on the messages requesting operator intervention, see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in "PRIMECLUSTER Messages."

```
1422 On the SysNode "node1RMS", the userApplication "app0" is the Faulted state due to a fault in
the resource "apl1".
Do you want to clear fault? (yes/no)
Message number: 1001
```

Note

If "Yes" is set for the "AutoStartUp" attribute, an operator intervention request message will be displayed at node startup. You need to respond to the operator intervention message after executing step 4. of the procedure.

7.5 Notes on Operation

This section describes notes when operating PRIMECLUSTER system.

Do not stop RMS while RMS is being started

Heartbeats between nodes are interrupted and the node where RMS is stopped may be forcibly shut down.

Stop RMS after completing its startup processing (completing the state transition processing of a cluster application).

Use hvshut -a to stop RMS on all nodes simultaneously

When executing the hvshut -l command on all nodes simultaneously, RMS will not be stopped and occasionally the timeout and hvshut command times out or hangs up.
When stopping RMS on all nodes, execute the hvshut -a command on any one of the nodes that configures a cluster system.
When stopping RMS on each node, execute the hvshut -l command on the node which stops RMS.

If mistakenly executing the hvshut -l command on all nodes simultaneously and the hvshut command times out, stop or reboot all the nodes. In addition, if the hvshut command hangs up, stop RMS forcibly using the hvshut -f command, and then stop or reboot all the nodes.

Do not stop operating system services after stopping RMS

Even if RMS is stopped using the hvshut command, other PRIMECLUSTER services (CF, SF, CRM, and so on) run.

Therefore, if you stop or reboot operating system services to modify its information (such as network information), heartbeat monitoring by CF fails and unexpected switchover will be occurred.

When modifying operating system information, be sure to do it after stopping all PRIMECLUSTER services (unloading CF) or in a single-user mode.

Create cluster applications used in RMS before starting RMS

If starting RMS without creating cluster applications, an error message (CML,14) will be output and RMS will not start.

The overview and the methods for creating cluster applications, "Chapter 6 Building Cluster Applications."

If operating systems hang up or slow down on a node in a cluster, a healthy node may be forcibly stopped.

If operating systems hang up or slow down on a node in a cluster due to system load, and so on, CF or RMS detects LEFTCLUSTER and stop the Shutdown Facility stops the node forcibly.

The Shutdown Facility forcibly stops a node according to the survival priority. Therefore, when the hang-up and slowdown of operating systems on the failed node are recovered before a healthy node forcibly stops the failed node, the healthy node may be forcibly stopped first.

When a system volume on a disk device cannot be referred to because all paths failed in a SAN boot configuration, the PRIMECLUSTER failure detection function cannot be operated depending on the status of the system.

Because the node which cannot refer to the system volume is unstable, set the node to panic status with the following method.

When you can log in cluster nodes other than the relevant node

Stop the relevant node using the sdtool command.

```
# sdtool -k <the relevant node>
```

When you cannot log in any nodes

Set the operational node to panic status. For details on how to set it, see the instruction manual of a main device.

Do not use the ipadm command for starting and stopping CIP as well as for changing its configuration.

If you start or stop CIP as well as to change its configuration using the ipadm command, an error message will be output and the command will fail.

Instead of that, use the cipconfig command or the ciptool command.

Do not use the service command of SMF for checking the status of PRIMECLUSTER and operating its system.

PRIMECLUSTER services are managed by SMF. Take note of the following two points:

- The state of PRIMECLUSTER services cannot be checked by the service command (such as (svcs(1)). Check it by the system operation screen (Cluster Admin) or a command of PRIMECLUSTER as before.
- Do not change the status of PRIMECLUSTER services using the service operation command (such as svcadm(1M)). Use the system operation screen (Cluster Admin) or a command of PRIMECLUSTER as before.

7.5.1 Notes on Switching a Cluster Application Forcibly

When Forced switch request (Forced startup) of a cluster application or resource is issued, RMS overrides all safety checks and starts the cluster application or resource. So if shared resources which require exclusive control between nodes become Online on the multiple nodes simultaneously, it could result in data corruption or other inconsistencies.

Note

A node where RMS is not running could be forcibly killed before the cluster application or the resource is forcibly started on another node to reduce the risk of data corruption.

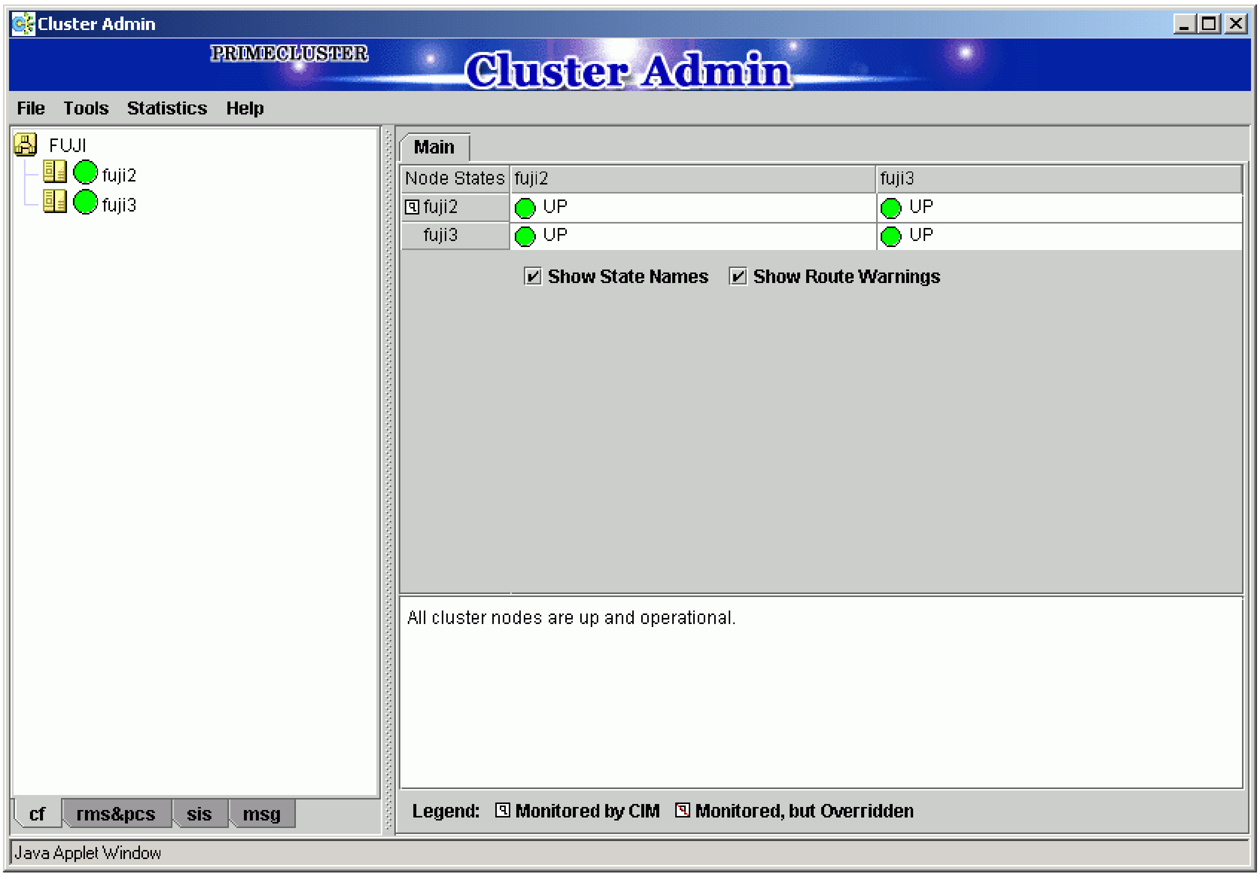
To perform forced startup of a cluster application or resource safely, check whether RMS is running on all nodes in the cluster before starting forced startup according to the following procedure, and if there are the nodes on which RMS is not running, then shut down the nodes.

1. Check the node state by one of the following methods:

   - Execute the cftool -n command on all nodes.

   ```
   fuji2# cftool -n
   Node  Number State       Os      Cpu
   fuji2 1      UP          Solaris Sparc
   fuji3 2      UP          Solaris Sparc
   ```

   - Check the CF tree of the Cluster Admin.



1. . Check the following contents for the node states, and take corrective actions if necessary:

   - Check the node states are all UP.
   - If a LEFTCLUSTER node exists, recover CF from the LEFTCLUSTER state.

     For details, see "PRIMECLUSTER Cluster Foundation Configuration and Administration."

   - If a node with DOWN or UNKNOWN exists, or if a node for which the state is not displayed exists, check whether the operating system of the node has stopped. If the operating system is running, shut down the operating system or restart the system in single-user mode.

2. Check whether some nodes on which RMS is not running exist among the nodes on which cluster applications or resources will be forcibly started by one of the following methods:

   - Execute the hvdisp -a command on nodes where cluster applications or resources will be started and check that the state of objects whose Type is SysNode is Online.

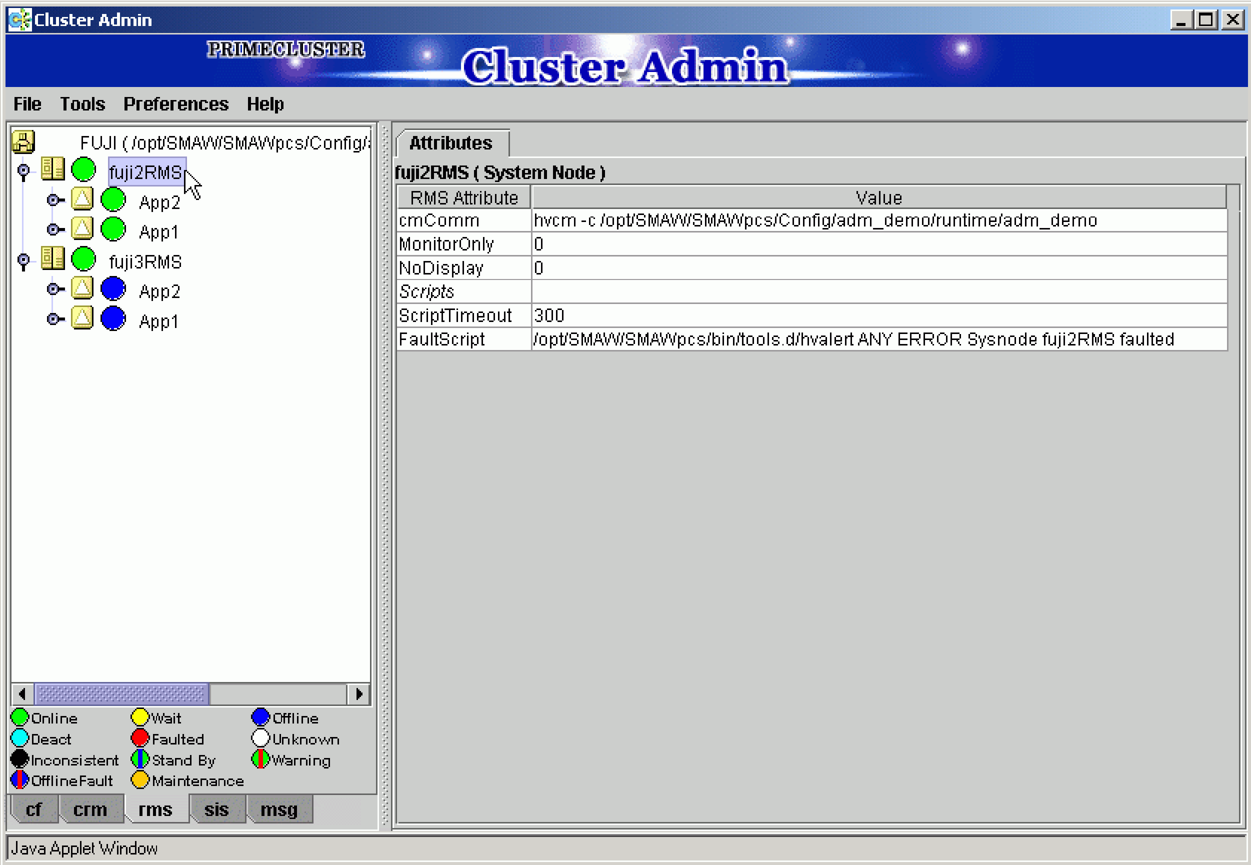
```
fuji2# hvdisp -a

Local System:  fuji2RMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us

Resource              Type    HostName              State        StateDetails
-----------------------------------------------------------------------------
fuji3RMS              SysNode                       Online
fuji2RMS              SysNode                       Online
app2                  userApp                       Offline
Machine001_app2       andOp   fuji3RMS
Machine000_app2       andOp   fuji2RMS              Offline
ManageProgram000_Cmd_APP2 gRes                            Offline
app1                  userApp                       Offline
Machine001_app1       andOp   fuji3RMS
```

```
Machine000_app1      andOp    fuji2RMS            Offline
ManageProgram000_Cmd_APP1 gRes                            Offline
```

- Check that the states of all SysNode displayed in the RMS tree of the Cluster Admin are Online.



3. If nodes which satisfy the following conditions exist, shut down the operating system of the nodes, or restart the system in single-user mode.

   - The node state is UP, and

   - The state of SysNode is not Online.

4. Execute the Forced switch (hvswitch -f) to forcibly start the cluster application or resource.

7.6 CF and RMS Heartbeats

PRIMECLUSTER sends heartbeats to CF and RMS. Each type of heartbeat failure that is detected from CF and RMS respectively and its detection time (default) are as follows.

Table 7.3 Failures detected with a heartbeat and its detection time of heartbeat timeout (CF and RMS))

	Failure type detected with a heartbeat	Detection time of heartbeat timeout (default)
CF	- System hangs on the kernel layer level - All paths failure of cluster interconnects - Remote node panics or reset (*1)	10 seconds
RMS	- System hangs on the user layer (application layer) level - RMS abnormal stop of a remote node(*2 and *3)	600 seconds

(*1): When using the monitoring agent of PRIMECLUSTER, the monitoring agent detects it immediately

(*2): The ELM heartbeat (RMS heartbeat) detects it immediately.

(*3): As an example, there is a double fault.

Note

The error detected by a CF heartbeat effects well on the operation. Therefore, the detection time of heartbeat timeout (detection time) is set shorter than RMS detection time.

If you set the detection time of CF shorter than that of RMS, the following warning message is output during RMS startup.

```
(BM, 4) The CF cluster timeout <cftimeout> exceeds the RMS timeout <rmstimeout>. This may result in RMS node elimination request before CF timeout is exceeded. Please check the CF timeout specified in /etc/default/cluster.config and the RMS heartbeat miss time specified by hvcm '-h' option.
```

7.7 cron Processing

This section describes the processing which PRIMECLUSTER performs with the cron command of a root user.

For details on each environment variable, see "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

cron entry name	Execution interval (default setting value)	Contents
hvlogcron	Once a day (at night)	Delete all RMS related log files from directories specified in RELIANT_LOG_PATH. The target files had been specified earlier than the period set in RELIANT_LOG_LIFE. RELIANT_LOG_LIFE: is a period for deleting RMS related log files Setting value any numbers of days Default value 7 (days) RELIANT_LOG_PATH: is where RMS and wizard tools log files are stored Setting value any valid path Default value /var/opt/SMAWRrms/log For the system that the operation is being stopped at the period of time hvlogcron is executed, logs are not deleted. Thus, the log volume may be increased. Change the cron configuration so that hvlogcron is executed once a day.
hvlogcontrol	in 15-minute intervals	Prevent RMS related log files from occupying the disk space. If the disk usage rate is more than HV_LOG_ACTION_THRESHOLD, delete all sub directories under RELIANT_LOG_PATH. If the disk usage rate is still more than HV_LOG_ACTION_THRESHOLD even if deleting sub directories, follow HV_LOG_ACTION and delete all current log files (when HV_LOG_ACTION is "on") or the WARNING message is output (HV_LOG_ACTION is "off") every time hvlogcontrol is executed. HV_LOG_ACTION_THRESHOLD: is the disk usage rate determined that log files occupy the disk space Setting value 0 - 100 Default value 98 (%) HV_LOG_ACTION: is the operation when the disk space is determined to be occupied Setting value on / off Default value off

cron entry name	Execution interval (default setting value)	Contents
sflogcontrol	in 15-minute intervals	Prevent SF related log files from occupying the disk space When the disk usage rate is 98 % or more, delete all sub directories under /var/opt/SMAWsf/log, delete the current log files when their size are 1 M bite or larger. Then, a WARNING message is output.
sflogcontrol midnight	Once a day (at night)	Delete SF related log files created 7 days ago or earlier from all sub directories under /var/opt/SMAWsf/log. For the system that the operation is being stopped at the period of time sflogcontrol midnight is executed, logs are not deleted. Thus, the volume of the logs is expected to increase. Change the cron configuration so that sflogcontrol midnight is executed once a day.
hvcleanupnfs	Once a day (at night)	Execute a recovery processing required for the RFS (NFS file system) resource. Use this cron in the Wizard for NAS (RFS) environment.

Note

Do not delete the entries which PRIMECLUSTER registered to the root user's cron, and do not move them to another user's cron as well.

Part 4 System Configuration Modification

Chapter 8 Changing the Cluster System Configuration

This chapter explains some configuration nodes of PRIMECLUSTER system, and how to add, delete, and change hardware.

After changing the cluster system configuration, use the PRIMECLUSTER environment checking tool to check the PRIMECLUSTER environment.

For details on checking the PRIMECLUSTER environment, see "6.10 Checking the Cluster Environment."

8.1 Adding, Deleting, and Changing Hardware

This section describes how to add, delete, and change the following hardware in the existing configuration:

- Shared disk device
- Network interface card used for the public LAN and the administrative LAN
- XSCF (only for change)

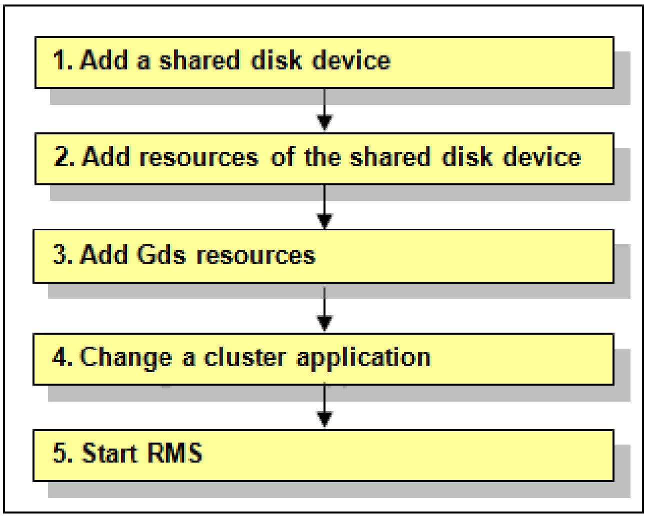
8.1.1 Adding Hardware

This section describes how to add hardware.

8.1.1.1 Adding a Shared Disk Device

To add a shared disk device, you need to change a cluster application related to the added shared disk device.

Figure 8.1 Procedure to add a shared disk device



Information

You must stop RMS during performing "4. Change a cluster application."

However, you do not need to stop RMS if all the following conditions are met because performing "4. Change a cluster application" is not necessary under the condition:

- The added shared disk device is registered with the existing class of GDS.
- The added shared disk device is no used as Fsystem resource.

1. Add a shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to add a shared disk device.

2. Add resources of the shared disk device.

   Register resources corresponding to the added shared disk device.

   See

   To register resources, see "5.1.3.2 Automatic Configure."

3. Add Gds resources.

   To use Global Disk Services (GDS), set up GDS and create Gds resources.

   If you register the added shared disk device with the existing class of GDS, you do not need to set Gds resources.

   See

   For information on how to set up GDS and create Gds resources, see "6.3.2 GDS Configuration Setup" and "6.7.1.3 Creating Gds Resources."

4. Change a cluster application.

   Add the following resources related to the added shared disk device to a cluster application.

   - Fsystem resource
   - Gds resource

   See

   For more details, see "10.3.1 Changing the Cluster Application Configuration."

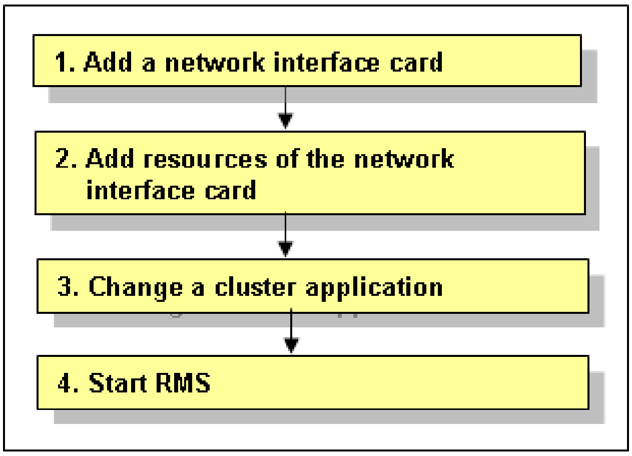
5. Start RMS.

   If RMS is not running, see "7.2.1.1 Starting RMS" and start RMS on all nodes.

8.1.1.2 Adding a Network Interface Card Used for the Public LAN and the Administrative LAN

To add a network interface card used for the public LAN and the administrative LAN, you need to change the cluster application related to the added network interface card according to the procedure below.

Figure 8.2 Procedure to add a network interface card



1. Add a network interface card.

   See "12.2 Maintenance Flow" and ask field engineers to add a network interface card.

2. Add resources of the network interface card.

   Register resources corresponding to the added network interface card.

   See

   To register resources, see "5.1.3.2 Automatic Configure."

3. Change a cluster application.

   Add the following resources related to the added network interface card to a cluster application.

   - Takeover network resource
   - Gls resource

   See

   For more details, see "10.3.1 Changing the Cluster Application Configuration."

4. Start RMS.

   If RMS is not running, see "7.2.1.1 Starting RMS" and start RMS on all nodes.

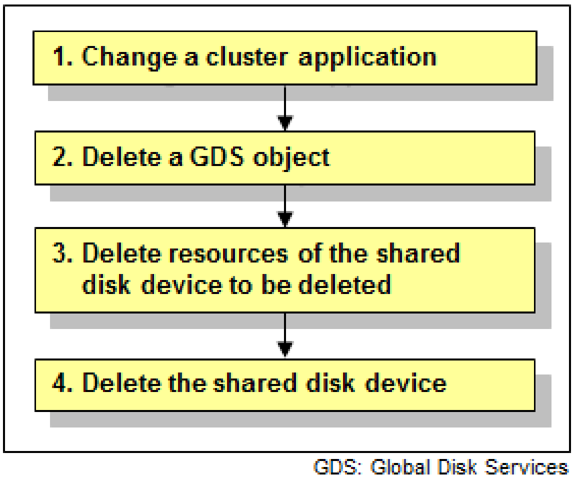
8.1.2 Deleting Hardware

This section describes how to delete hardware.

8.1.2.1 Deleting a shared disk device

To delete a shared disk device, you need to change a cluster application which includes resources of the shared disk device to be deleted beforehand.

Figure 8.3 Procedure to delete a shared disk device



1. Change a cluster application.

   Delete the following resources using the shared disk device to be deleted from a cluster application:

   - Fsystem resource
   - Gds resource

See

To change the configuration of a cluster application and delete a resource, see "10.3.1 Changing the Cluster Application Configuration."

2. Delete a GDS object.

   Delete a GDS object related to the shared disk device to be deleted.

See

To delete a GDS object, see "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide.*"

3. Delete resources of the shared disk device to be deleted.

   Delete resources of the registered shared disk device by using the "cldeldevice" command.

   For details on the "cldeldevice" command, see the manual page.

Note

When resources of the shared disk device to be deleted are registered to a GDS class, delete the shared disk device from the GDS class first, and then delete resources of the shared disk device. To delete the shared disk device from the GDS class, see the "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide.*"

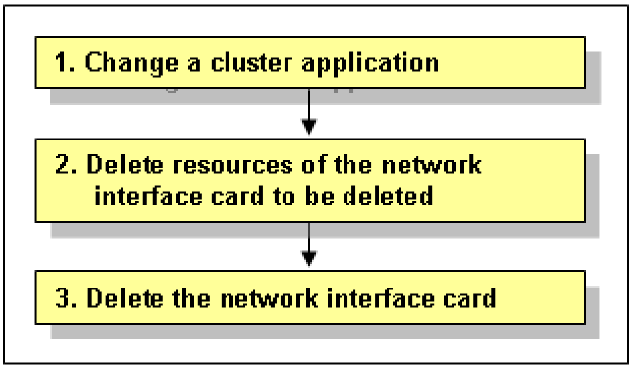
4. Delete the shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to delete the shared disk device.

8.1.2.2 Deleting a network interface card used for the public LAN and the administrative LAN

To delete a network interface card used for the public LAN and the administrative LAN, you need to change a cluster application which includes resources of the network interface card to be deleted beforehand.

Figure 8.4 Procedure to delete a network interface card



1. Change a cluster application.

   Delete the following resources using the network interface card to be deleted from a cluster application:

   - Takeover network resource
   - Gls resource

See

To change the configuration of a cluster application and delete a resource, see "10.3.1 Changing the Cluster Application Configuration."

2. Delete resources of the network interface card.

   Delete resources of the registered network interface card by using the "cldelrsc" command.

   For details on the "cldelrsc" command, see the manual page.

3. Delete the network interface card.

   See "12.2 Maintenance Flow" and ask field engineers to delete the network interface card.

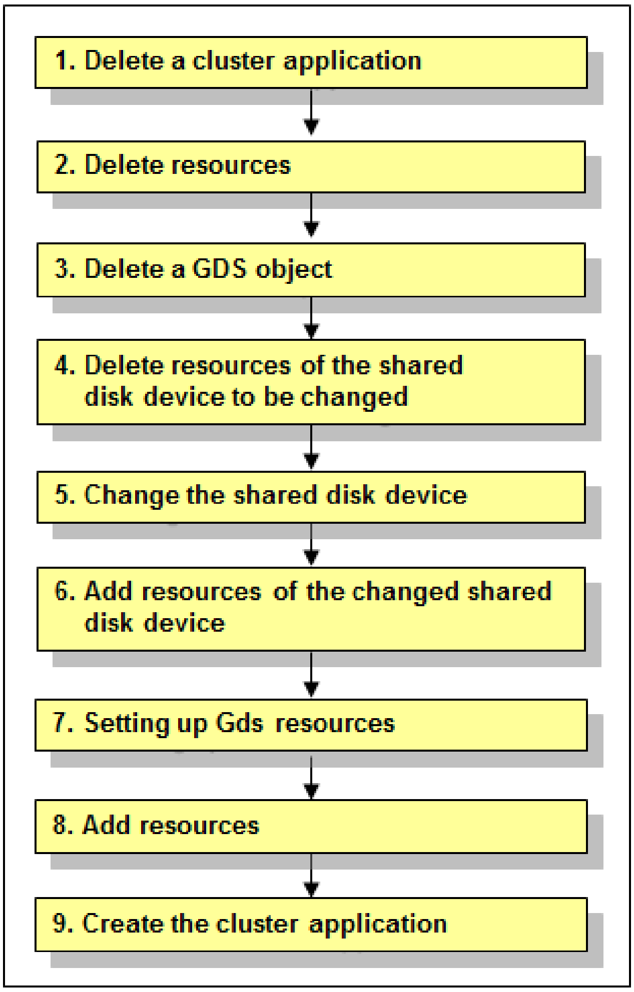
8.1.3 Changing Hardware

This section describes how to change hardware.

8.1.3.1 Changing a shared disk device

To change a shared disk device, you need to delete a cluster application which includes resources of the shared disk device to be changed beforehand, and then create the cluster application again after changing the shared disk device.

Figure 8.5 Procedure to change a shared disk device



1. Delete a cluster application.

   Note

   To delete a cluster application, you must stop RMS beforehand. If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

   Delete the cluster application which includes the following resources related to the shared disk device to be changed:

   - Fsystem resource
   - Gds resource

   See

   To delete a cluster application, see "10.2 Deleting a Cluster Application."

2. Delete resources.

   Delete Fsystem resources and Gds resources related to the shared disk device to be changed.

   See

   To delete resources, see "10.5 Deleting a Resource."

3. Delete a GDS object.

   Delete a GDS object related to the shared disk device to be changed.

   See

   To delete a GDS object, see "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

4. Delete resources of the shared disk device to be changed.

   Delete resources of the registered shared disk device by using the "cldeldevice(1M)" command.

   For details on the "cldeldevice(1M)" command, see the manual page.

   Note

   When resources of the shared disk device to be deleted are registered to a GDS class, delete the shared disk device from the GDS class first, and then delete resources of the shared disk device. To delete the shared disk device from the GDS class, see the "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

5. Change the shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to change the shared disk device.

6. Add resources of the changed shared disk device.

   Register resources corresponding to the changed shared disk device to the resource database with the automatic configuration facility.

   See

   For information on automatic configuration of resources, see "5.1.3.2 Automatic Configure."

7. Setting up Gds resources.

   To use Global Disk Services (GDS), set up GDS and create Gds resources.

See

For information on how to set up GDS and create Gds resources, see "6.3.2 GDS Configuration Setup" and "6.7.1.3 Creating Gds Resources."

8. Add resources.

   If you have deleted Fsystem resources in Step 2, add Fsystem resources.

   See

   To add resources, see "6.7.1 Setting Up Resources."

9. Create the cluster application.

   Create the cluster application deleted in Step 1 again.

   If cluster application dependency relationships have been set up, they must be set up again.

   After creating the cluster application, execute "RMS startup" from the Cluster Admin screen. If necessary, also start the cluster application.

   See

   To create a cluster application, see "6.7.2 Creating Cluster Applications."

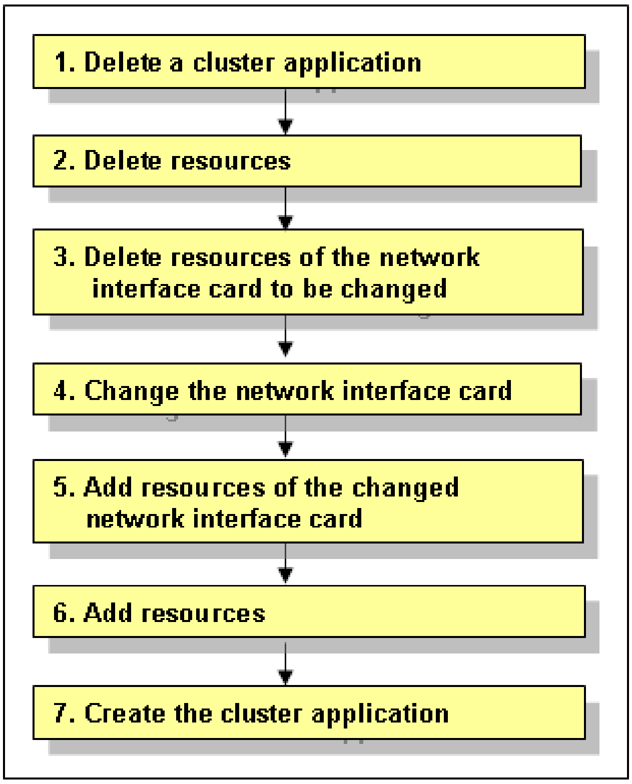
   For information on how to set up cluster application dependency relationships, see "6.7.3 Setting Up Dependency Relationships Between Cluster Applications."

   For instructions on starting RMS, see "7.2.1.1 Starting RMS." For instructions on starting the cluster application, see "7.2.2.1 Starting a Cluster Application."

8.1.3.2 Changing a network interface card used for the public LAN and the administrative LAN

To change a network interface card used for the public LAN and the administrative LAN, you need to delete a cluster application which includes resources of the network interface card to be changed beforehand, and then create the cluster application again after changing the network interface card.

Figure 8.6 Procedure to change a network interface card



1. Delete a cluster application.

   Note

   To delete a cluster application, you must stop RMS beforehand. If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

   Delete the cluster application which includes the following resources related to the network interface card to be changed:

   - Takeover network resource
   - Gls resource

   See

   To delete a cluster application, see "10.2 Deleting a Cluster Application."

2. Delete resources.

   Delete takeover network resources and Gls resources related to the network interface card to be changed.

   See

   To delete resources, see "10.5 Deleting a Resource."

3. Delete resources of the network interface card to be changed.

   Delete resources of the registered network interface card by using the "cldelrsc" command.

   For details on the "cldelrsc" command, see the manual page.

4. Changing the network interface card.

   See "12.2 Maintenance Flow" and ask field engineers to change the network interface card.

5. Add resources of the changed network interface card.

   Register resources corresponding to the changed network interface card to the resource database with the automatic configuration facility.

   See

   For information on automatic configuration of resources, see "5.1.3.2 Automatic Configure."

6. Add resources.

   If you have deleted takeover network resources and Gls resources in Step 2, add takeover network resources and Gls resources.

   See

   To add resources, see "6.7.1 Setting Up Resources."

7. Create the cluster application.

   Create the cluster application deleted in Step 1 again.

   If cluster application dependency relationships have been set up, they must be set up again.

   After creating the cluster application, execute "RMS startup" from the Cluster Admin screen. If necessary, also start the cluster application.

   See

   To create a cluster application, see "6.7.2 Creating Cluster Applications."

   For information on how to set up cluster application dependency relationships, see "6.7.3 Setting Up Dependency Relationships Between Cluster Applications."

   For instructions on starting RMS, see "7.2.1.1 Starting RMS." For instructions on starting the cluster application, see "7.2.2.1 Starting a Cluster Application."

8.1.3.3 Changing a network interface card used for CIP

This section describes how to change a network interface card used for CIP from /dev/hme3 to /dev/hme4 using an example flow.

Note

Specify a physical device name to "device" (for example, /dev/hme0). You cannot specify a link name (for example, net0) in the Solaris 11 environment.

Procedure when changing from /dev/hme3 to /dev/hme4

1. Stop CF on all the nodes configuring a cluster.

   For how to stop CF, see "PRIMECLUSTER Cluster Foundation Configuration and Administration Guide."

2. Check the currently used interfaces by executing the following command on all the nodes.

```
# cfconfig -g
The own node name the cluster name /dev/hme3
```

3. Delete CF configuration by executing the following command on all the nodes.

```
# cfconfig -d
```

4. Configure CF by executing the following command on all the nodes.

```
# cfconfig -S <the own node name> <the cluster name> /dev/hme4
```

5. Make sure that the interfaces currently used have been changed by executing the following command on all the nodes.

```
# cfconfig -g
<The own name> <the cluster> /dev/hme4 (Check that /dev/hme4 has been displayed).
```

6. Start CF on all the nodes configuring a cluster.

    1. Log in to Web-Based Admin View.
    2. See "5.7.1 Starting CF" in "PRIMECLUSTER Cluster Foundation Configuration and Administration Guide" to start CF on all the nodes configuring a cluster.
    3. Make sure that all the nodes are Online on "cf" in Cluster Admin. In addition, make sure that each interconnect is UP.
    4. Finish Cluster Admin.
    5. Log out from Web-Based Admin View.

8.1.3.4 Replacing XSCF

This section describes the necessary procedure for replacing XSCF.

1. Replacing XSCF

   Follow the procedure in "12.2 Maintenance Flow" to ask field engineers to replace XSCF. (Perform the stop maintenance.)

2. Checking the XSCF information

   Check the XSCF settings according to your server machine type and the environment.

   For SPARC M10, see "5.1.2.1.1 Checking XSCF Information."

   For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000, see "5.1.2.2.1 Checking Console Configuration."

   For the Oracle Solaris Kernel Zones environment, see "5.1.2.5.1 Checking XSCF Information."

3. Setting SNMP

   Perform this step only for SPARC M10.

   Follow the procedure in "5.1.2.1.2 Setting SNMP" to set various settings relating SNMP.

4. Setting the IP address of XSCF

   Perform this step only when changing the IP address of XSCF.

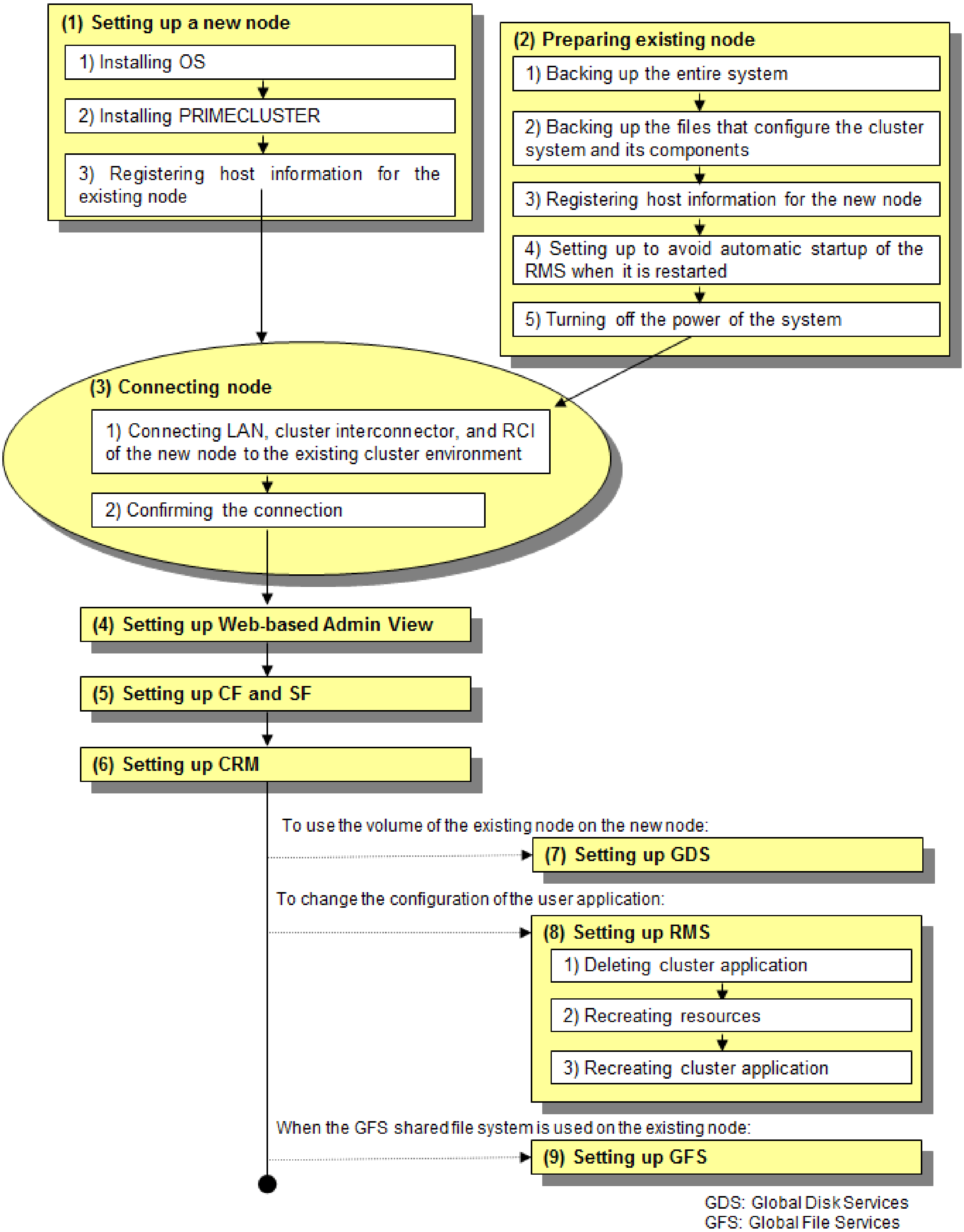
   Follow Steps 11 and 13 in "9.2.1 Changing an IP Address on the Public LAN" to set the IP address of XSCF again.

8.2 Adding a Node

Node expansion means adding on an extra node to an operating cluster system. The purpose of node expansion is to provide high availability and extend available business operations.

8.2.1 Procedure for Node Expansion

The node expansion procedure is explained below:



Note

- Use the same OS version, collective updates, and patches as those of the existing cluster nodes.
- Stop operation when nodes are to be added.

- Two or more nodes cannot be added at the same time. When multiple nodes are to be added, add them one by one.
- The nodes to be added must be of the same model as the existing cluster nodes.

In the following explanation, node1 and node2 are used as the node names of the existing cluster nodes while node3 is used as the node name of the new node.

8.2.1.1 Setting Up a New Node

Take the following steps to set up the new node.

The new node should be prepared such that the operating system, PTFs, FibreChannel, and packages such as the multipath software have already been installed.

Procedure

1. Install PRIMECLUSTER on the new node.

   This must be done in a single user mode.
   For details, see the "*PRIMECLUSTER Installation Guide*."

2. Configure NTP.

   Configure NTP for the new node to match the NTP of the existing nodes.

3. Define the following information in the "/etc/inet/hosts" file of the new node.

   - The IP address of the existing cluster nodes and the host name that is associated with the IP address
   - The IP address of the remote console that is connected to the existing cluster nodes and the host name that is associated with the IP address
   - The IP address of the CIP interface of the existing cluster nodes and the CIP name which uses that IP address

4. Turn off the power to the new node.

8.2.1.2 Preparing the Existing Nodes

Take the following steps to prepare the existing nodes.

Procedure

1. Preparing for unexpected failures, you need to back up the entire system of all existing cluster nodes, the PRIMECLUSTER system and the configuration files of each component.

   1. Back up the entire system.

      1. Stop RMS by executing the following command on any one of the existing cluster nodes.

         ```
         node1# hvshut -a
         ```

      2. Reboot all the existing cluster nodes from a single user mode.

         ```
         node1# /usr/sbin/shutdown -g0 -i0 -y
           ....
         ok boot -s
           ....

         Type control-d to proceed with normal startup,
          (or give root password for system maintenance):
            ....
         ```

      3. Mount the file system on all the existing cluster nodes.

         ```
         node1# mountall -l
         node1# zfs mount -a
         ```

4. Back up the entire system or property in the shared disk by executing the "ufsdump(1M)" or "dd(1M)" command.

2. Back up the PRIMECLUSTER system and the configuration files of each component.

   1. Back up the configuration files of the PRIMECLUSTER system on all existing cluster nodes. See "Chapter 13 Backing Up and Restoring a PRIMECLUSTER System".

   2. Back up the configuration files that are used for GLS on all existing cluster nodes.
      To back up the configuration files for GLS (redundant line control), use the following command (For details on the "hanetbackup" command, see the " PRIMECLUSTER *Global Link Services Configuration and Administration Guide: Redundant Line Control Function* "):

```
node1# /opt/FJSVhanet/usr/sbin/hanetbackup -d /var/tmp/backup
```

      To back up the configuration files for the GLS multipath function:

```
node1# cd /etc/opt/FJSVmpnet
node1# tar cvf - conf | compress > /var/tmp/backup/mpnetfile.tar.Z
```

2. Define the following information in the /etc/inet/hosts file of all the existing cluster nodes.

   - The IP address of the node to be added and the name of the host that is associated with the IP address

   - The IP address of the remote console that is connected to the node to be added and the host name that is associated with the IP address

3. Edit the "/opt/SMAW/SMAWRrms/bin/hvenv.local" file as shown below so that RMS does not start automatically on any of the existing cluster nodes, even when a cluster node is rebooted.

```
node1# vi /opt/SMAW/SMAWRrms/bin/hvenv.local
```

```
export HV_RCSTART=0
```

4. To add a new node, all the existing cluster nodes must be turned off.

8.2.1.3 Connecting a Node

Join a new node with the existing cluster nodes.

Procedure

1. Connect the LAN, Cluster Interconnect, and the RCI of the new node to the existing cluster environment.

   At this time, configure the RCI address for the new node.
   (This operation is done by your Fujitsu CE.)

2. After setting the RCI address, boot up the existing cluster nodes and check that no error message is output to the console or syslog.

3. Boot the new node and confirm that the new node and its remote console are correctly configured in the network by executing the "ping(1M)" command.

4. Confirm that the RMS is stopped on any one of the existing cluster nodes, and then stop the SF by executing the following commands on each existing cluster node.

   Confirm that RMS is stopped.

```
node1# hvdisp -a
hvdisp: RMS is not running
```

   Stop SF. Execute the following command on all the existing cluster nodes.

```
node1# sdtool -e
node1# sdtool -s
(SMAWsf, 30, 13) : The RCSD is not running
```

5. If the GFS shared file system is used in an existing node, take the following steps to stop the GFS operation.

    1. Execute the following command for the entire GFS shared file system on any one of the existing cluster nodes, and then unmount the file system.

    ```
    node1# sfcumntgl <mount point>
    ```

    2. Execute the following command, and then stop the GFS daemon on all cluster nodes.

    ```
    node1# sfcfrmstop
    ```

8.2.1.4 Configure Web-Based Admin View

This section explains how to configure Web-Based Admin View.

The nodes on which you need to configure Web-Based Admin View vary depending on the following cases;

Target node:

- When the existing management server is used

  The management server must be defined on the new node.

  Configure Web-Based Admin View on the new node.

- When the new node is used as the management server

  The definition of the new management server must be defined on all the nodes.

  Configure Web-Based Admin View on all the nodes.

Procedure

1. Set up Web-Based Admin View on the node.

   See "4.2.3.1 Initial setup of the operation management server."

2. Confirm that Web-Based Admin View is running correctly.

   For confirmation, use any one of the cluster nodes as explained in "4.2.3.2 Confirming Web-Based Admin View Startup."

8.2.1.5 Setting Up CF and SF

Make the CF and SF configuration by using Cluster Admin. This section explains how to configure CF and SF. See "2.1.4 Example of creating a cluster" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

Procedure

1. Start the Web browser on a client and connect to the primary management server to display Web-Based Admin View.
2. Log on to Web-Based Admin View and then select the "Global Cluster Services" menu (see Figure 3).
3. Select a node name to be added on the node selection screen (see Figure 6).
4. Click the "*Configure*" button in the left-hand side panel on the screen, to start the CF wizard (see Figure 8).
5. Select an existing cluster system name from the "Create/Join Cluster" screen and then click the "*Add local node to an existing CF Cluster*" option button. Then, choose the "*Next*" button (see Figure 10).
6. Select a node that you want to add from [Available Nodes] on the "Selecting cluster nodes and the cluster name" screen, and then add the node to [Clustered Nodes] (see Figure 11).
7. When the Cluster Interconnect confirmation screen appears, confirm that the combination of network interface cards is correct on all nodes. Then, click the "*Next*" button (see Figure 14).
8. Check the "*For RMS*" checkbox for the CIP subnet settings (note that the RMS cannot use the CIP if this is not set.)
9. When the "*Complete Configuration*" dialog box appears, close that screen and click the "*Finish*" button. This completes the CF settings.

10. Configure SF.
    For details, see "5.1.2 Configuring the Shutdown Facility."

11. Confirm that the CF and SF have been configured correctly by executing the following commands on the GUI screen or on any one of the cluster nodes.

```
node1# cftool -n
Node     Number State     Os      Cpu
node1    1      UP        Solaris Sparc
node2    2      UP        Solaris Sparc
node3    3      UP        Solaris Sparc
```

```
node1# sdtool -s
Cluster Host    Agent                SA State      Shut State  Test State  Init State
------------    -----                --------      ----------  ----------  ----------
node1           SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
node1           SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
node1           SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
node1           SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
node3           SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
node3           SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
node3           SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
node3           SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
```

8.2.1.6 Setting Up CRM (Resource Database)

This section explains how to set up the Customer Resource Management (CRM) resource database.

Procedure

1. Reconfigure the resource database on the existing nodes.

   Confirm the following:

   - All of the existing nodes have been started.
   - CIP entry of the new node is in /etc/cip.cf of the existing nodes.

     The resource database uses /etc/cip.cf to associate CF node names with CIP names.

   Reconfigure the resource database.

   The procedure is shown below:

   1. Log into any one of existing nodes using a system administrator access privilege.
   2. Specify the "-a" and the "-g" options in the "clsetp(1M)" command.

      ```
      # /etc/opt/FJSVcluster/bin/clsetup -a node -g file
      ```

      As a result of this operation, a new node will be added to the resource database of the existing nodes.

      The configuration information on the resource database, created after the execution of the "clsetup(1M)" command, is used when the resource database of the new node is configured. Therefore, do not specify a directory that will be automatically deleted when rebooting the node with the "-g" option (for example: /tmp).

      Specify the CF node name in *node*, and a full path name of the file name of the resource database configuration information. "tar.Z" extension will be appended to the resource database configuration information.

      For example, to add a new node which has a CF node name of fuji4, and a configuration information file name of /mydir/rdb, the command to be executed will be as shown below:

```
# /etc/opt/FJSVcluster/bin/clsetup -a fuji4 -g /mydir/rdb
```

The configuration information file of the resource database will be created as /mydir/rdb.tar.Z.

3. Confirm that the new node has been added to the resource database.

   Execute the "clgettree(1)" command, and then confirm that the new node is displayed on the output result. At this time, the state of the new node is displayed as UNKNOWN.

2. Set up the resource database of the new node.

   Before setting up the resource database of the new node, confirm the following:

   - The content of /etc/cip.cf of the new node must be the same as that of the existing nodes.

     Confirm that the CIP of the new node is in /etc/cip.cf of the new node and that the content of /etc/cip.cf is the same as that of the existing nodes.

   - Communication must be enabled in CIP

     Confirm that the new node is connected to all of the existing nodes by CIP using the "ping(1M)" command.

     If two or more CIPs are configured in the new nodes, use the first CIP for the resource database. Then, connection will be enabled. An example using "fuji4RMS" as the new node is shown below:

```
# ping fuji4RMS
```

   After confirming the above, set up the resource database of the new node.
   The procedure is as follows:

   1. Log into the new node using a system administrator access privilege.

   2. Copy the "resource database configuration information" file created in Step 2) in "Recreate the settings for the resource database of the existing nodes" to the new node.

   3. Specify the -s option in the clsetup(1M) command, and execute it.

```
# /etc/opt/FJSVcluster/bin/clsetup -s file
```

   Specify *file* with a full path name of the resource database configuration file.

   When the resource database configuration information file "rdb.tar.Z" is copied to /mydir, the command to be executed will be as shown below:

```
# /etc/opt/FJSVcluster/bin/clsetup -s /mydir/rdb.tar.Z
```

   4. Confirm that the resource database of the new node is configured.

      Execute the clgettree(1) command for the new node and confirm the following:

      - The new node is displayed.

      - The state of the new node is displayed as ON.

      - The output result is the same as that of the existing nodes.

3. Register the hardware, which is connected to the new node, to the resource database.

   Log into any one of nodes using a system administrator access privilege, and execute the command shown below:

```
# /etc/opt/FJSVcluster/bin/clautoconfig -r
```

Setting up synchronization with the resource database

If the individual nodes are restarted at different times after node expansion, the tuning parameter must be set up to be synchronized with the resource database. For details, see "4.5.1 Start up synchronization and the new node" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

8.2.1.7 Setting Up GDS

If a new node was added to the GDS configuration, it is necessary to change the class scope by executing the sdxattr command.

For information on the change procedure, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

8.2.1.8 Setting Up RMS

This section explains how to register a new node (SysNode) in a userApplication that has been running on an existing node.

Procedure

1. Configuration for each resource

   Take the following steps depending upon the resources in the existing userApplication:

   - Cmdline

     Create the Start, Stop, and Check scripts in the new node or copy them from the existing node. If "*Path enter*" is selected from "*Creation method*" when creating the existing Cmdline resource, specify the paths to the scripts. If "New" is selected, the scripts under /opt/FJSVwvucw/scripts/start, /opt/FJSVwvucw/scripts/stop, and /opt/FJSVwvucw/scripts/check must be stored in the same directory of the new node. You also need to add the access privilege by executing the "chmod(1)" command.

   - Gds

     Take the following steps to expand the class scope:

     1. Expand the class scope.

        See "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

     2. Execute the following command on any one of the cluster nodes:

        ```
        # /opt/SMAW/SMAWRrms/bin/hvgdsetup -a class
        hvgdsetup with -a option performs the following tasks on
        nodes to which the specified disk class belongs.
          1) Make GDS disk class on resource database not activated
             automatically when the node boots. If this operation has
              been done before, nothing will be performed anymore.
          2) Next make volumes of the specified disk class
             enabled manual online on a node on which an application
             is offline or faulted while the application is online
             or standby on another node.
          3) Then make volumes of the specified disk class stopped
             immediately.
             This process is executed on the nodes to which the disk
             class belongs.
        Do you want to continue with these processes ? [yes/no] yes
        ```

        The following message might appear after executing the "hvgdsetup" command. This does not disrupt ongoing operation.

        ```
        FJSVcluster: error: clrmd: 7516: An error occurred in the resource deactivation
        processing. (resource:resource rid:rid detail:detail)
        WARNING !!
        Failed to control 'dc_class' in the following node(s).
          node(s) node_name:
        Check the state of the nodes. If any nodes have failed, you may ignore this message.
        ```

   - Fsystem

     Add the mount point entry to /etc/vfstab.pcl on the new node.

   - Gls

     Take the following steps for Gls:

1. Set up the virtual interface for the takeover IP address on the new node and register it as a cluster resource. For details, see the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

2. Restart Gls by executing the following command:

```
node3# /opt/FJSVhanet/usr/sbin/resethanet -s
```

- Takeover network

  Nothing needs be done at this time. In Procedure 3, however, it is necessary to recreate the resources.

- Procedure

  Create a state transition procedure on the new node and register the procedure resource with the cluster resource manager. For more details, see "E.1 Registering a Procedure Resource".

- Process monitoring

  Add the startup command to the new node. Also, you need to add the access privilege by using the "chmod(1)" command. Then, recreate the resources in Procedure 3.

2. Deleting userApplication

   Delete the existing userApplication by using the userApplication Configuration Wizard. At this time, select "*Delete only userApplication*."

   For more details, see "10.3.1 Changing the Cluster Application Configuration."

3. Recreating the takeover network and process monitoring resources

   If the takeover network resource and the process monitoring resource are registered in the cluster system, first delete and then recreate those resources.

   See "10.5 Deleting a Resource," "6.7.1.5 Creating Takeover Network Resources," and "6.7.1.7 Creating Process Monitoring Resources."

4. Recreating userApplication

   Recreate the userApplication that was deleted in Procedure 2, using the same procedure as that used to create it. Note that the new node must be registered when SysNode is to be selected. For details, see "6.7.2 Creating Cluster Applications."

5. Copy /opt/SMAW/SMAWRrms/bin/hvenv.local of the existing node to /opt/SMAW/SMAWRrms/bin/ of the new node.

6. Edit /opt/SMAW/SMAWRrms/bin/hvenv.local in each node with the "vi" editor, and delete the following entry:

   ```
   export HV_RCSTART=0
   ```

8.2.1.9 GFS Shared Settings

If the GFS shared file system is used in an existing node, set up the GFS shared file system on the new node by using the following procedure:

Procedure

1. Confirm the GFS daemon (sfcfrmd) is not running by executing the "ps" command on all cluster nodes. If GFS daemon is running, see Step 5 of "8.2.1.3 Connecting a Node" when stop the GFS daemon.

2. Execute sfcsetup on the new node, and then register the node information in the management partition.

3. Execute sfcfrmstart and then start up the GFS daemon on all cluster nodes.

4. Execute sfcnode on any one of the cluster nodes, and then add the node configuration information of the new node.

5. Create a mount point and set up /etc/vfstab on the new node.

6. Execute sfcmntgl on any one of the cluster nodes and then mount the GFS shared file system.

See

For information on how to use each command, see "PRIMECLUSTER Global File Services Configuration and Administration Guide."

8.2.2 Recovering the Original Cluster Configuration at Node Expansion

When the original cluster configuration must be restored due to a fault or interruption during node expansion, take the following steps for recovery:

Procedure

1. Cancel the configuration changes in GFS.

   Perform this operation in multi-user mode.

   If a new node was added to the GFS configuration at node expansion, it is necessary to cancel the GFS configuration in multi-user mode before deleting the new node.

   Execute sfcnode on any one of the cluster nodes and delete the node configuration information. Then, execute sfcfrmstop and stop GFS on all the nodes. In this state, execute sfcsetup on the node to be deleted, and then delete the node information from the management partition.

   See

   For information on how to use each command, see "PRIMECLUSTER Global File Servers Configuration and Administration Guide."

2. Change the class scope.

   If a new node was added to the GDS configuration, it is necessary to change the class scope by executing the sdxattr command. For information on the change procedure, see the "*Global Disk Services Configuration and Administration Guide*."

3. Recover the PRIMECLUSTER system configuration files in single user mode.

   See "Chapter 13 Backing Up and Restoring a PRIMECLUSTER System."

4. Recover the GLS configuration files.

   Restore the environment definition files used by the GLS multipath function by using the following procedure:

   ```
   node1# cd /etc/opt/FJSVmpnet
   node1# uncompress /var/tmp/backup/mpnetfile.tar.Z
   node1# tar xvf /var/tmp/backup/mpnetfile.tar
   ```

5. Restore the original cluster configuration.

   Restore the cluster interconnect, LAN cable, and RCI cable to the original cluster configuration. At this time, restore the original RCI settings. This should be done by field engineers.

Chapter 9 Changing the Cluster System Environment

This chapter describes how to change the configuration information and environmental settings of PRIMECLUSTER system.

After changing the cluster system environment, use the PRIMECLUSTER environment checking tool to check the PRIMECLUSTER environment.

For details on checking the PRIMECLUSTER environment, see "6.10 Checking the Cluster Environment."

9.1 Changing the Cluster Configuration Information

9.1.1 Changing a Cluster Node Name

The following explains how to change the node name after building a PRIMECLUSTER system.

Note

Changing a node name may have a serious impact on the system. Therefore, make this change only when it is absolutely necessary.

Use the "sys-unconfig(1M)" command to change a node name. The "sys-unconfig(1M)" command is used to reset the system. All settings such as the name service, time zone, IP address, IP subnet mask, and root password are all erased and are then reset when the node next boots.

See

For details on the "sys-unconfig(1M)" command, see the manual page describing sys-unconfig(1M).

Operation Procedure:

1. Stop CF on the node whose node name is to be changed.

   For details on how to stop CF, see "5.7 Starting and stopping CF" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

2. Set up the shutdown facility.

   When changing the node name assigned to the IP address of the Administrative LAN that is to be used in the shutdown facility, perform the following according to the definition of the /etc/opt/SMAW/SMAWsf/rcsd.cfg file.

   - When the IP address is defined in "admIP" of the rcsd.cfg file

     You do not need to change the rcsd.cfg file.

   - When the node name registered in the hosts(4) file is defined "admIP" of the rcsd.cfg file

     1. Execute the following command on each cluster node to stop the shutdown facility.

        ```
        # sdtool -e
        ```

     2. Define the node name to be changed to "admIP" of the /etc/opt/SMAW/SMAWsf/rcsd.cfg file.

   See

   For details on the rcsd.cfg file, see the manual page of rcsd.cfg(4M).

3. Define "export HV_RCSTART=0" in the hvenv.local file of the node whose node name is to be changed.

   This disables automatic startup RMS when the node is booted after changing the node name.

   For details on the hvenv.local file and the HV_RCSTART attribute, see the manual page describing hvenv.local.

4. Change the node name.
   For Solaris 10, the "sys-unconfig(1M)" command is used. For Solaris 11, the "sysconfig(1M)" command is used. After the node names are changed, edit the hosts(4) file if necessary. For details, see Oracle Solaris documents.

   Note

   Do not attempt to change information other than node names.

5. Restart the system.

```
# /usr/sbin/shutdown -y -i6 -g0
```

6. Start RMS.

   For information on how to start up RMS, see "8.1.1 Starting RMS" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

7. After confirming that RMS has started up, restore the definition of the HV_RCSTART environment variable corrected in Step 3.

9.1.2 Changing a CF Node Name

This section describes how to change the CF node name after building a PRIMECLUSTER system.

Operation Procedure:

The following explains the operation procedure using an example when changing the CF name from fuji2 and fuji3 to fuji4 and fuji5.

1. When using RMS, stop the automatic RMS startup.
   Check the current automatic RMS startup, and then execute the following regarding the setting. Perform this operation on all nodes where RMS is used.

```
# hvsetenv HV_RCSTART
1 <- Check this value.
```

   - If "0" is set, RMS has been stopped. Go to Step 2.
   - If "1" is set, execute the following to stop the automatic RMS startup.

```
# hvsetenv HV_RCSTART 0
# hvsetenv HV_RCSTART
0 <- Check the "0" is output.
```

2. When using the GFS Shared File System, back up the GFS configuration information with the following procedure:

   1. Back up the management partition information of the GFS Shared File System on the node before change.

      Execute the following command on any running node.

```
# sfcgetconf _backup_file_
```

      In the above example, sfcgetconf(1M) generates a shell script named _backup_file_ in the current directory.

   2. Check the setup of the startup procedure of the sfcfrmd daemon.

```
# sfcsetup -m
wait_bg
```

      Record the output value.

      This value is used when restoring the GFS configuration information after changing the CF node name.

   3. Check the management partition information of GFS.

```
# sfcinfo -a
/dev/sfdsk/gfs01/dsk/volume01:
```

```
FSID special                                      size  Type  mount
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)  14422  META  -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)   5116  LOG   -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)  95112  DATA  -----
```

```
# sfcrscinfo -m -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID  MDS/AC  STATE  S-STATE   RID-1   RID-2   RID-N  hostname
   1  MDS(P)  stop   -             0       0       0  host2
   1  AC      stop   -             0       0       0  host2
   1  MDS(S)  stop   -             0       0       0  host3
   1  AC      stop   -             0       0       0  host3
```

Save the output result.

This information is used for checking the restored configuration information of GFS after changing the CF node name.

3. Start all nodes in single-user mode.

4. Change the CF node name and the CIP/Sysnode name.

   Perform the following operation on all nodes.

   Note

   For the naming convention of the CF node name, see "5.1.1 Setting Up CF and CIP."

   1. Change the string of the CF node name within the CF node name and the CIP/Sysnode name that are described in /etc/cip.cf.

      [Before change]

```
fuji2       fuji2RMS:netmask:255.255.255.0
fuji3       fuji3RMS:netmask:255.255.255.0
```

      [After change]

```
fuji4       fuji4RMS:netmask:255.255.255.0
fuji5       fuji5RMS:netmask:255.255.255.0
```

   2. Change the string of the CF node name within the CIP/Sysnode name that is described in /etc/inet/hosts.

      [Before change]

```
192.168.0.1     fuji2RMS
192.168.0.2     fuji3RMS
```

      [After change]

```
192.168.0.1     fuji4RMS
192.168.0.2     fuji5RMS
```

   3. Change the CF node name described in /etc/default/cluster.

      [Before change]

```
nodename fuji2
clustername PRIMECLUSTER1
device /dev/hme2
device /dev/hme3
```

      [After change]

```
nodename fuji4
clustername PRIMECLUSTER1
device /dev/hme2
device /dev/hme3
```

5. Delete the file of the Cluster Resource Management Facility.

   If the /etc/opt/FJSVcluster/FJSVcldbm/config/shmno file exists, delete it.

```
# rm /etc/opt/FJSVcluster/FJSVcldbm/config/shmno
```

6. Change the node name of the Cluster Resource Management Facility.

   Note

   This procedure is unnecessary when the Cluster Resource Management Facility is not being set.

   Execute the following command to change the node name of the Cluster Resource Management Facility.

```
# /etc/opt/FJSVcluster/bin/clchgnodename
```

7. Delete the information in the management partition of GFS.

   Note

   This procedure is unnecessary when the GFS Shared File System is not being used.

   Delete the information in the management partition of the GFS Shared File System. Execute the following command on all nodes.

```
# rm /var/opt/FJSVgfs/sfcfsrm.conf
```

8. Delete the /etc/opt/SMAW/SMAWsf/rcsd.cfg file of the shutdown facility.

   Execute the following on all nodes.

```
# rm /etc/opt/SMAW/SMAWsf/rcsd.cfg
```

9. Start all nodes in multi-user mode.

10. Set up the Cluster Integrity Monitor (CIM).

    Delete the CF node names used before change, and set the CF node names to be used after change.

    Perform this setting on any node that configures the cluster system.

    Example: The CF node names used before change are fuji2 and fuji3, and those to be used after change are fuji4 and fuji5.

```
# rcqconfig -d fuji2 fuji3
# rcqconfig -a fuji4 fuji5
```

11. Checking the CF setting item

    Check if the changed CF node name and CIP/SysNode name are correct.

    a. Checking the CF node name

       Execute the "cfconfig -g" command on each node to check if the set CF node name is correct.

       Example: When the changed is fuji4

```
# cfconfig -g
fuji4 PRIMECLUSTER1 /dev/hme2 /dev/hme3
```

    b. Checking the CIP/Sysnode name

       Check that all the CIP/Sysnode names set in the remote host are enabled to communicate. Check the communication status on all nodes.

       Example: When the Sysnode name set in the remote host is fuji5RMS

```
# ping fuji5RMS
```

If an error occurs in the above step a or b, check if the CF node name and CIP/SysNode name set in /etc/cip.cf, /etc/default/cluster or /etc/inet/hosts are correct.

If an error occurs, take the procedure below:

1. Start the system in single-user mode.
2. Perform "4. Change the CF node name and the CIP/Sysnode name" and start the node again.
3. Perform "10. Set up the Cluster Integrity Monitor (CIM)" again.

12. Set up the shutdown facility.

    1. Delete the information of the asynchronous monitoring that was used before changing the CF node name.

       For SPARC M10

       Execute the following command on any node to check that the information of the SNMP asynchronous monitoring that was used before changing the CF node name is displayed.

       ```
       # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
       ```

       Execute the following commands on any node to delete the information of the SNMP asynchronous monitoring that was used before changing the CF node name.

       ```
       # /etc/opt/FJSVcluster/bin/clsnmpsetup -d fuji2
       # /etc/opt/FJSVcluster/bin/clsnmpsetup -d fuji3
       ```

       Execute the following command on any node to check that information of the SNMP asynchronous monitoring is not displayed.

       ```
       # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
       ```

       For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000, and
       for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series

       Execute the following command on any one node to check that the information of the console asynchronous monitoring that was used before changing the CF node name is displayed.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -l
       ```

       Execute the following command on any node to delete the information of the console asynchronous monitoring that was used before changing the CF node name.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -d fuji2
       # /etc/opt/FJSVcluster/bin/clrccusetup -d fuji3
       ```

       Execute the following command on any node to check that the information of the console asynchronous monitoring is not displayed.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -l
       ```

       For SPARC Enterprise T1000 and T2000

       Delete the /etc/opt/SMAW/SMAWsf/SA_sunF.cfg file.

       ```
       # rm /etc/opt/SMAW/SMAWsf/SA_sunF.cfg
       ```

    2. See "5.1.2 Configuring the Shutdown Facility" to set the shutdown facility again.

13. When using the GFS Shared File System, restore the GFS configuration information with the following procedure:

    1. Reinitialize the management partition on the one node of the copy destination servers.

       Example: Initializing the /dev/sfdsk/gfs/rdsk/control file as the management partition.

```
# sfcsetup -cf /dev/sfdsk/gfs/rdsk/control
```

2. Reregister the information of the configuration node on each node.

```
# sfcsetup -a /dev/sfdsk/gfs/rdsk/control
```

3. Redo the settings for the startup method of the sfcfrmd daemon as recorded in Step 2-2 on any node.

   Example: when the startup method of sfcfrmd daemon is wait_bg

```
# sfcsetup -m wait_bg
```

Note

This procedure is required when changing the startup method of the sfcfrmd daemon from the default value wait.

4. Confirm that the management partition is reinitialized.

   The path name of the management partition for which the settings were made can be confirmed by executing the"sfcsetup(1M)" command with the -p option.

```
# sfcsetup -p
/dev/sfdsk/gfs/rdsk/control
```

   The registered node information can be confirmed by executing the "sfcsetup(1M)" command without any option.

   Check that the value of CIPNAME is the CIP/Sysnode name that was changed in Step 4-2.

```
# sfcsetup
HOSTID          CIPNAME         MP_PATH
80380000        fuji4RMS        yes
80380001        fuji5RMS        yes
```

   The startup method of the sfcfrmd daemon can be confirmed by executing the "sfcsetup(1M)" command with the -m option.

```
# sfcsetup -m
wait_bg
```

5. Start the sfcfrmd daemon by executing the following command on all nodes.

```
# sfcfrmstart
```

6. Restore the information of the management partition.

   Execute the shell script generated in Step 2-1 on any node.

```
# sh _backup_file_
get other node information start ... end
```

   Confirm that restoration of the management partition of GFS was successful by running the "sfcinfo(1M)" command and the "sfcrscinfo(1M)" command.

   Moreover, confirm that the information is the same as the one in Step 2-3.

```
# sfcinfo -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID special                                        size  Type  mount
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)    14422  META  -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)     5116  LOG   -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)    95112  DATA  -----
```

```
# sfcrscinfo -m -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID  MDS/AC  STATE  S-STATE   RID-1   RID-2   RID-N  hostname
```

```
1  MDS(P)  stop    -             0       0       0  host2
1  AC      stop    -             0       0       0  host2
1  MDS(S)  stop    -             0       0       0  host3
1  AC      stop    -             0       0       0  host3
```

7. Mount the GFS Shared File System on all nodes.

```
# sfcmntgl <mount point>
```

Perform the following procedures when the cluster application is being set. If not, it is unnecessary.

14. Obtain the RMS Configuration name. Follow the procedure below to output the file content. The character string after -c will be the RMS Configuration name. Perform this operation on any one of the nodes constituting a cluster.

Example

If the RMS Configuration name is 'config'

```
# cat /opt/SMAW/SMAWRrms/etc/CONFIG.rms
hvcm -c config <- RMS Configuration name
```

15. Collect the backup of the configuration file. Perform this operation on all cluster nodes.

Execute the following command.

```
cd /opt/SMAW/SMAWRrms/build/wizard.d
cp -rp <RMS Configuration name> <RMS Configuration name>.backup
```

When the takeover network resource is being used, execute the following command on all cluster nodes.

```
cp -p /opt/SMAW/SMAWRrms/etc/hvipalias /opt/SMAW/SMAWRrms/etc/hvipalias.backup
```

Example

If <RMS Configuration name> is config

```
# cd /opt/SMAW/SMAWRrms/build/wizard.d
# cp -rp config config.backup
# cp -p /opt/SMAW/SMAWRrms/etc/hvipalias /opt/SMAW/SMAWRrms/etc/hvipalias.backup
```

16. Modify the SysNode name of the configuration file. Perform this operation on the same node where the Step 14 was performed.

1. Move it to the directory where the configuration file is stored.

```
# cd /opt/SMAW/SMAWRrms/build/wizard.d/<RMS Configuration name>
```

Example

If <RMS Configuration name> is config

```
# cd /opt/SMAW/SMAWRrms/build/wizard.d/config
```

2. Search the target file with the following command.

```
grep -l HvpMachine *
```

Example

```
# grep -l HvpMachine *
userApp_0.m
userApp_1.m
    :
```

The displayed ".m" file is the target file.

3. Search and change the line starting "HvpMachine" in the displayed file. Open the file with the vi command.

```
# vi userApp_0.m
HvpApplication=userApp_0
HvpAutoBreak=yes
      :
    (Omitted)
      :
HvpMachine000=fuji2RMS  <- Target line to be changed
HvpMachine001=fuji3RMS  <- Target line to be changed
      :
   (Omitted)
```

Change the SysNode name that is set in the right side of "=" to the new SysNode name.

The SysNode name is set by adding RMS (capital letters) to the CF node name.

The following example indicates when changing the CF node name from fuji2 and fuji3 to fuji4 and fuji5.

[Before change]

```
HvpMachine000=fuji2RMS
HvpMachine001=fuji3RMS
```

[After change]

```
HvpMachine000=fuji4RMS
HvpMachine001=fuji5RMS
```

Perform this procedure on all the ".m" files obtained in Step 16-2.

4. Check that the old SysNode does not exist.

Example

```
# grep HvpMachine *.m
userApp_0.m:HvpMachine000=fuji4RMS
userApp_0.m:HvpMachine001=fuji5RMS
userApp_1.m:HvpMachine000=fuji4RMS
userApp_1.m:HvpMachine001=fuji5RMS
```

5. Rename the <SysNode name>.s file under the directory where the configuration file is stored with the mv command. Change the file with the old SysNode name to the new SysNode name.

Example

```
# mv fuji2RMS.s fuji4RMS.s
# mv fuji3RMS.s fuji5RMS.s
```

17. Modify the CF node name of the procedure resource, line switching unit resource, and patrol diagnosis. Perform this operation on the same node where the Step 16 was performed.

    This procedure is unnecessary if the procedure resource, line switching unit resource, and patrol diagnosis are not being used.

    1. Move it to the directory where the configuration file is stored.

    ```
    # cd /opt/SMAW/SMAWRrms/build/wizard.d/<RMS Configuration name>
    ```

    Example

    If <RMS Configuration name> is config

    ```
    # cd /opt/SMAW/SMAWRrms/build/wizard.d/config
    ```

    2. Search the target file wit the following command.

    ```
    grep -l HvpCrmScopeFilter *
    ```

    Example

    ```
    # grep -l HvpCrmScopeFilter *
    Procedure0.m
    Procedure1.m
       :
    ```

    The displayed ".m" file is the target file.

    3. Search and change the line starting "HvpCrmScopeFilter" in the displayed file. Open the file with the vi command.

    ```
    # vi Procedure0.m
    HvpApplType=RESOURCE
    HvpApplication=Procedure0
    HvpClassNameFilter=BasicApplication
    HvpConsistent=consistent
    HvpCrmFlags000=OT1800
    HvpCrmResourceId000=33
    HvpCrmResourceName000=SDISK
    HvpCrmScopeFilter=fuji2:fuji3  <- Target line to be changed.
    HvpPlugin=BasicApplication
    HvpPreCheckHeritageIn=''
       :
    (Omitted)
    ```

    Change the CF node name that is set in the right side of "=" to the new CF node name.

    The following example indicates when changing the CF node name from fuji2 and fuji3 to fuji4 and fuji5.

    [Before change]

    ```
    HvpCrmScopeFilter=fuji2:fuji3
    ```

    [After change]

    ```
    HvpCrmScopeFilter=fuji4:fuji5
    ```

    Perform this procedure on all the ".m" files displayed in Step 17-2.

    4. Check that the old CF node name does not exist.

    ```
    grep HvpCrmScopeFilter *.m
    ```

Example

```
# grep HvpCrmScopeFilter *.m
Procedure0.m:HvpCrmScopeFilter=fuji4:fuji5
Procedure1.m:HvpCrmScopeFilter=fuji4:fuji5
   :
```

18. When the takeover network resource is being used, change the CF node name in the hvipalias file.

    This file is stored in /opt/SMAW/SMAWRrms/etc/hvipalias. If the file does not exist or there is no CF node name in the file, this procedure is unnecessary because the takeover network resource is not being used.

    For each line, the CF node name is set in the first field and the host name of the takeover network is set in the second field. Modify the CF node name of each line to the new name. Perform this operation on all the nodes constituting a cluster.

Example

When the CF node names are changed from fuji2 and fuji3 to fuji4 and fuji5.

[Before change]

```
fuji2  hostname    net0  0xffffff00        # sh_rid=34 rid=32 192.168.100.233
fuji3  hostname    net0  0xffffff00        # sh_rid=34 rid=33 192.168.100.233
```

[After change]

```
fuji4  hostname    net0  0xffffff00        # sh_rid=34 rid=32 192.168.100.233
fuji5  hostname    net0  0xffffff00        # sh_rid=34 rid=33 192.168.100.233
```

19. Execute the following command on the same node where the Step 14 was performed.

```
hvw -F Configuration-Activate -xj -n <RMS Configuration name>
```

Example

```
# hvw -F Configuration-Activate -xj -n config
```

In the command output result, check the following contents of (1) and (2).

```
Testing for RMS to be up somewhere in the cluster ... done.

Arranging sub applications topologically ... done.

Check for all applications being consistent ... done.

Running overall consistency check ... done.

Generating pseudo code [one dot per (sub) application]: ..... done.

Generating RMS resources [one dot per
resource]: ............................................................ done


hvbuild using /usr/opt/reliant/build/wizard.d/config/config.us
About to distribute the new configuration data to hosts: fuji4RMS,fuji5RMS (1)

The new configuration was distributed successfully.

About to put the new configuration in effect ... done.
```

```
The activation has finished successfully. (2)
#
```

In (1), check that the new SysNode name is output.

In (2), check that "The activation has finished successfully." is displayed.

If the output results are different, check the following and take a necessary action:

- Check if RMS is in operation on any node that constitutes a cluster. If it is, stop RMS.
- Check that all nodes that constitute a cluster run in multi-user mode.
- Check that the host name in the /etc/hosts file or the CF node name has been change on all nodes.

After taking the necessary action, perform the operation in Step 19 again.

20. If the automatic RMS startup was changed in Step 1, return it with the following procedure. Perform this operation on all nodes where RMS is used.

```
# hvsetenv HV_RCSTART 1
# hvsetenv HV_RCSTART
1   <- Check that the returned value is output.
```

9.2 Changing the Network Environment

This section explains how to change the network environment, for example, how to change the IP address setting if the IP address of a public LAN, administrative LAN, or CIP is changed after the PRIMECLUSTER system is installed. Note that when you change the IP address, do not change the host name.

9.2.1 Changing an IP Address on the Public LAN

This section explains how to change the IP address setting if the IP address of a public LAN or administrative LAN are changed after the PRIMECLUSTER system is installed.

Operation Procedure:

1. Execute the following command on one of the cluster nodes to stop RMS operation:

```
# hvshut -a
```

2. Execute the following command in all nodes to start the system in single-user mode:

```
# /usr/sbin/shutdown -y -g0 -i0
     :
ok> boot -s
```

3. Execute the following command in all nodes to mount the file system:

```
# /usr/sbin/mountall -l
# /usr/sbin/zfs mount -a
```

4. Edit the /etc/inet/hosts file, and change the IP address of each node.
For Solaris 11, additional settings by using the "ipadm" command are required.

5. On each node, change the IP address of the public LAN used by Web-Based Admin View.

See

For details, see "7.1 Network address," "7.3 Management server," and "7.5 Multi-network between server and client by classified use" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

6. If a takeover IP address must be changed, correct the IP address being used as the takeover IP address in the "/etc/inet/hosts" file of each node.

7. If you created an Ipaddress resource by using the userApplication Configuration Wizard, the RMS automatic startup needs to be disabled.

   Execute the following steps after checking the current state value of the RMS automatic startup.

```
# /opt/SMAW/SMAWRrms/bin/hvsetenv HV_RCSTART
1 <- Check this value.
```

   - If the value is "0", the RMS automatic startup has been disabled. Go to Step 14.

   - If the value is "1", execute the following command to change "1" to "0".

```
# /opt/SMAW/SMAWRrms/bin/hvsetenv HV_RCSTART 0
# /opt/SMAW/SMAWRrms/bin/hvsetenv HV_RCSTART
0 <- Make sure this value is "0".
```

8. When changing the IP address used by the shutdown facility, you must decide whether the /etc/opt/SMAW/SMAWsf/rcsd.cfg file must be changed on each cluster node.

   - If the host name registered to the /etc/inet/hosts file is defined in "admIP" of the rcsd.cfg file, you do not need to change the rcsd.cfg file.

   - If an IP address is defined in "admIP" of the rcsd.cfg file, you need to change the rcsd.cfg file. Change the definition of "admIP."

   For details on the rcsd.cfg file, see the rcsd.cfg(4M) manual page.

9. Restart the system on all nodes.

```
# /usr/sbin/shutdown -y -g0 -i6
```

10. Execute the following command on each cluster node to stop the shutdown facility:

```
# sdtool -e
```

11. Change the IP address of XSCF, ALOM, or ILOM described in the /etc/inet/hosts file on each node.

12. For SPARC M10, perform the following steps. For other than SPARC M10, proceed to the next step.

    To change the IP address used in the shutdown facility for SPARC M10, change the setting of the SNMP agents on all the XSCFs which constitutes a cluster.

    Execute the command below to check the current setting value.

```
XSCF> showsnmp
```

    Delete the IP address to be changed, and then re-set the changed IP address.

```
XSCF> setsnmp remtraphost -t v2 -s FJSVcldev -p 9385 [IP address before the change]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address after the change]
```

    Execute the command below to check if the setting has been changed.

```
XSCF> showsnmp
```

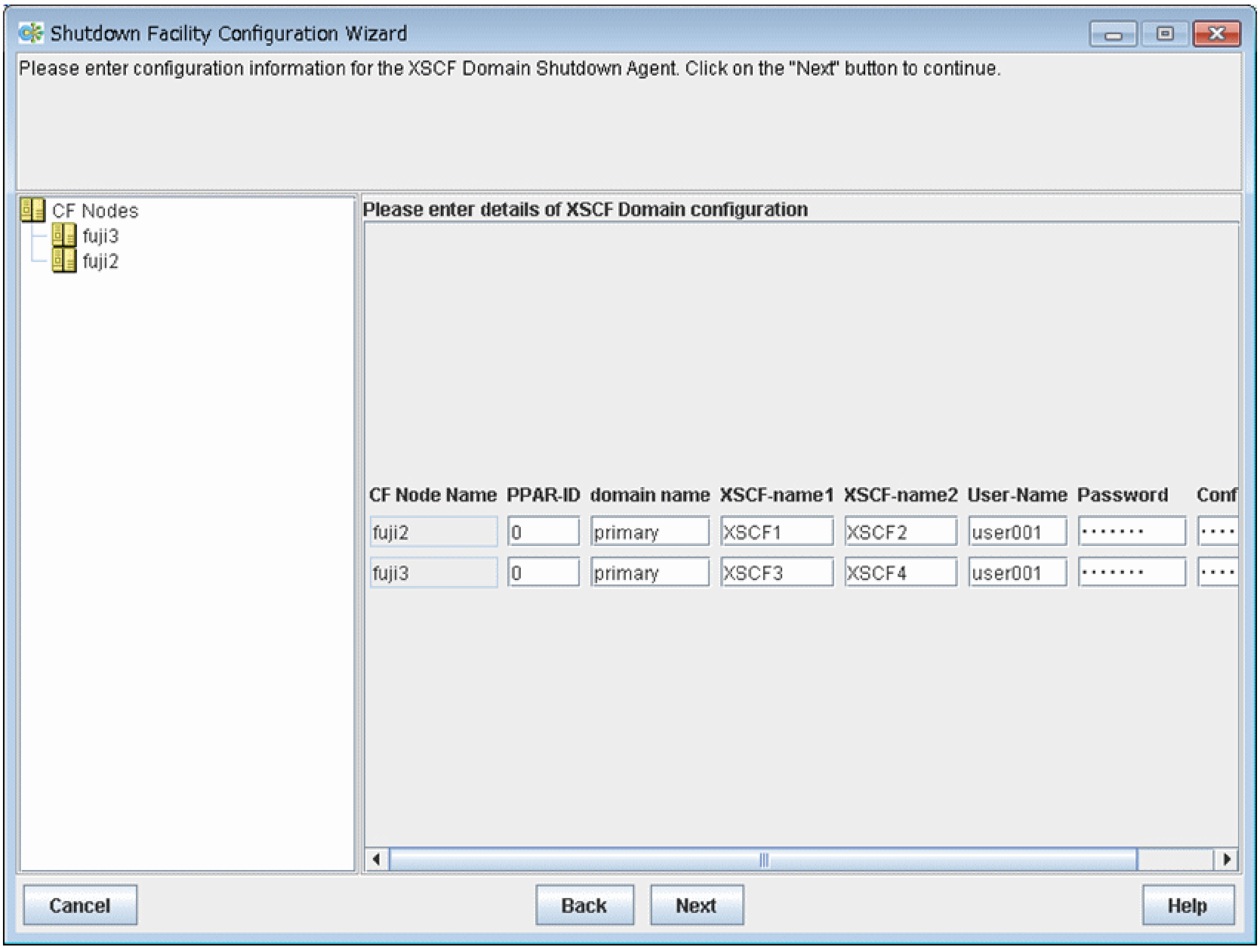
13. Change the following IP addresses registered to the shutdown facility as necessary.

    a. IP address of XSCF (for SPARC M10)

       For XSCF, see "5.1.2.1.3 Using the Shutdown Configuration Wizard" and reconfigure the IP address of XSCF.

The IP address does not need to be changed if the host name registered to the /etc/inet/hosts file is displayed for the XSCF name.

Figure 9.1 XSCF (SPARC M10)

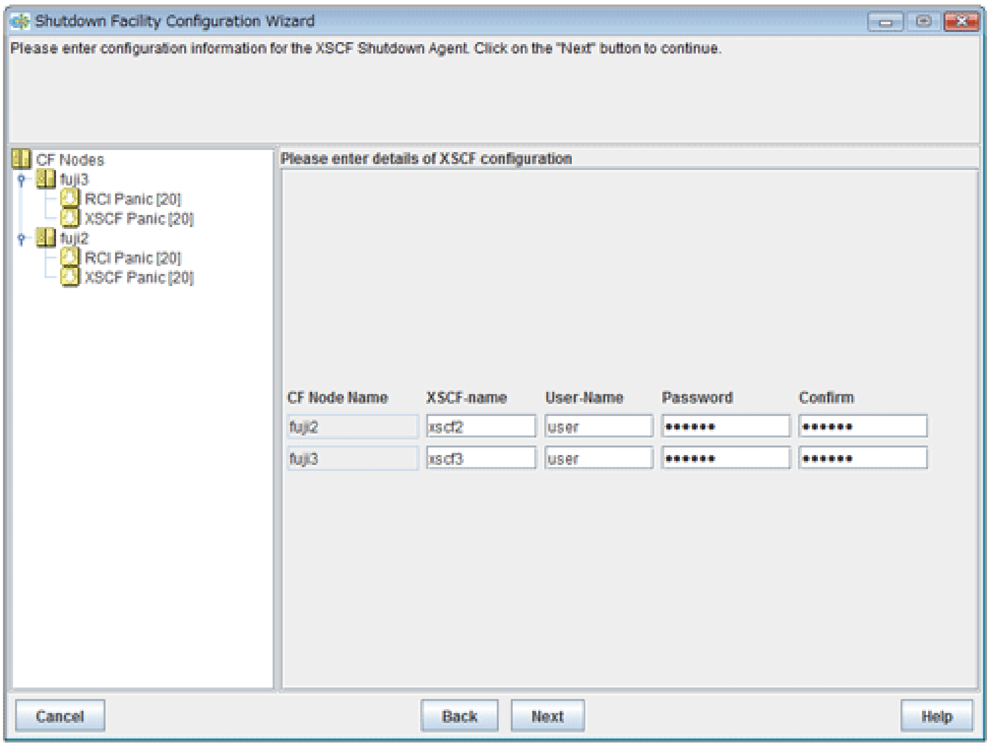


b. IP address of XSCF (for SPARC Enterprise M3000, M4000, M5000, M8000, and M9000)

For XSCF, see "5.1.2.2.2 Using the Shutdown Configuration Wizard" and reconfigure the IP address of XSCF.

The IP address does not need to be changed if the host name registered to the /etc/inet/hosts file is displayed for the XSCF name.

Figure 9.2 XSCF (SPARC Enterprise M3000, M4000, M5000, M8000, and M9000)

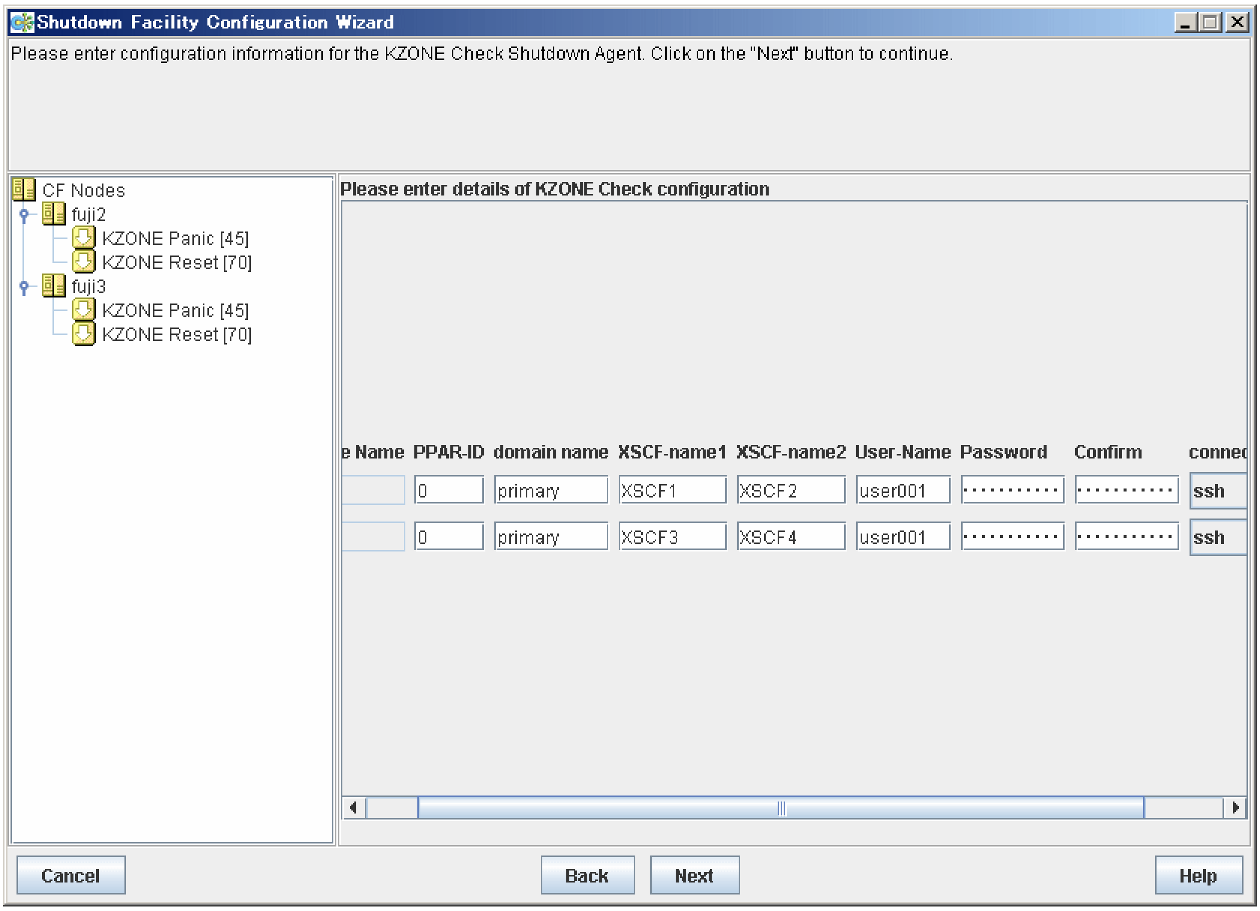


c. IP address of XSCF (for Oracle Solaris Kernel Zones)

For XSCF, see "5.1.2.5.4 Using the Shutdown Configuration Wizard" and reconfigure the IP address of XSCF.

The IP address does not need to be changed if the host name registered to the /etc/inet/hosts file is displayed for the XSCF name.

Figure 9.3 XSCF (Oracle Solaris Kernel Zones)

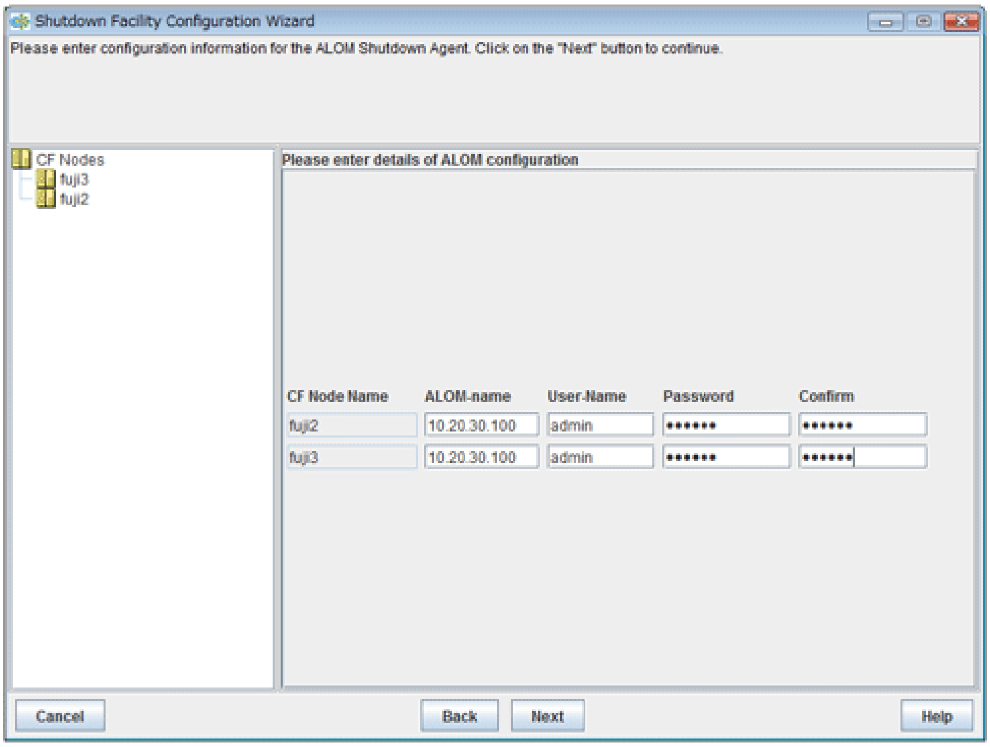


d. IP address of ALOM

For ALOM, see "5.1.2.4.2 Using the Shutdown Configuration Wizard" and reconfigure the IP address of ALOM.

Specify the changed IP address for the ALOM name.

Figure 9.4 ALOM (SPARC Enterprise T1000, T2000)

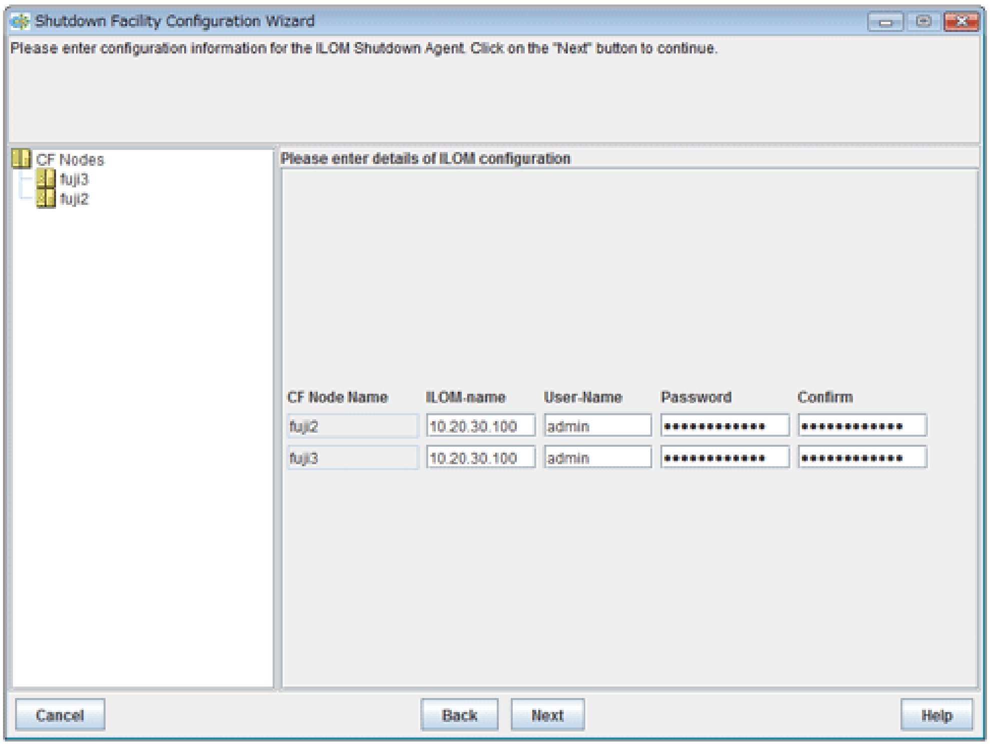


e. IP address of ILOM (for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series)

For ILOM, see "5.1.2.3.2 Using the Shutdown Configuration Wizard" and reconfigure the IP address of ILOM.

The IP address does not need to be changed if the host name registered to the /etc/inet/hosts file is displayed for the ILOM name.

Figure 9.5 ILOM (SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series)

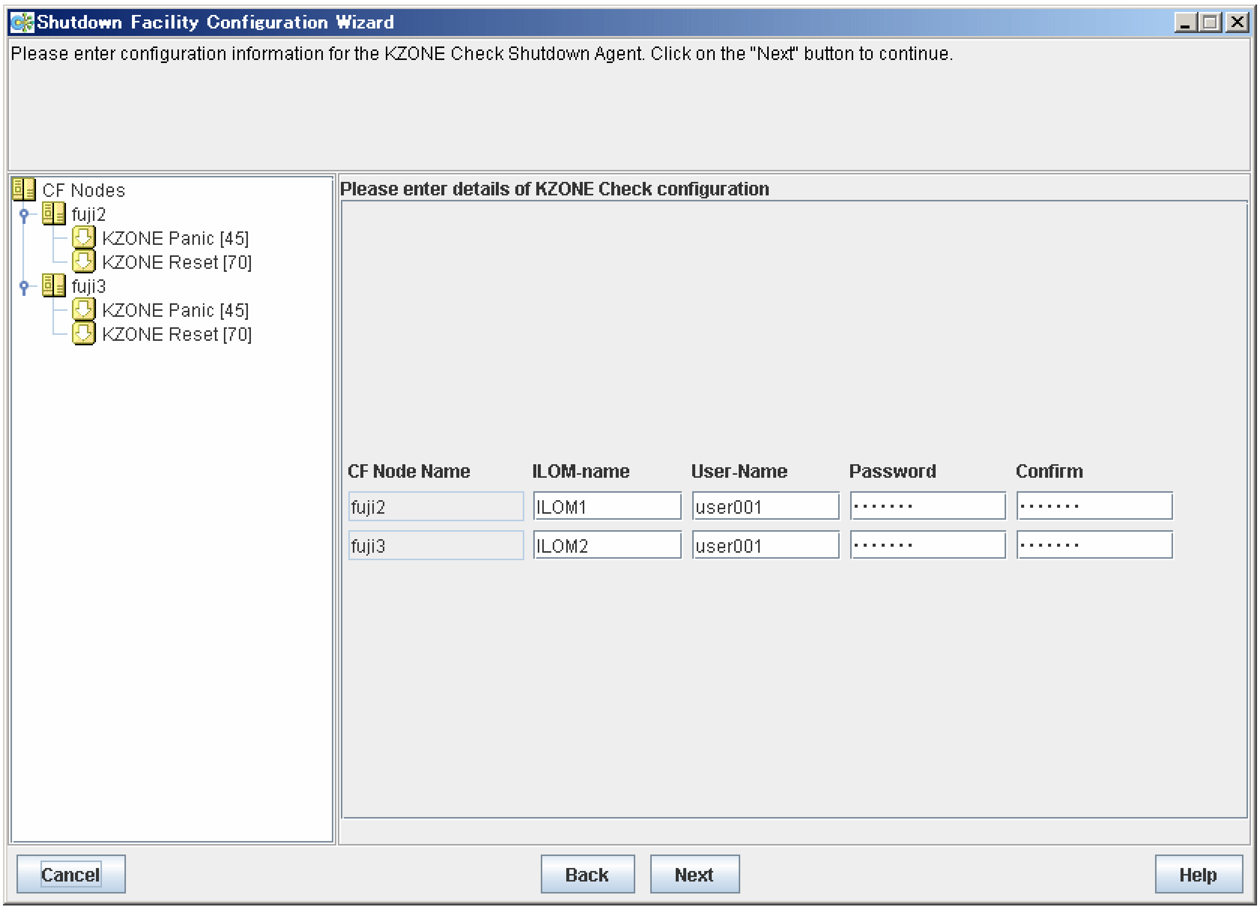


f. IP address of ILOM (for Oracle Solaris Kernel Zones)

For ILOM, see "5.1.2.5.4 Using the Shutdown Configuration Wizard" and reconfigure the IP address of ILOM.

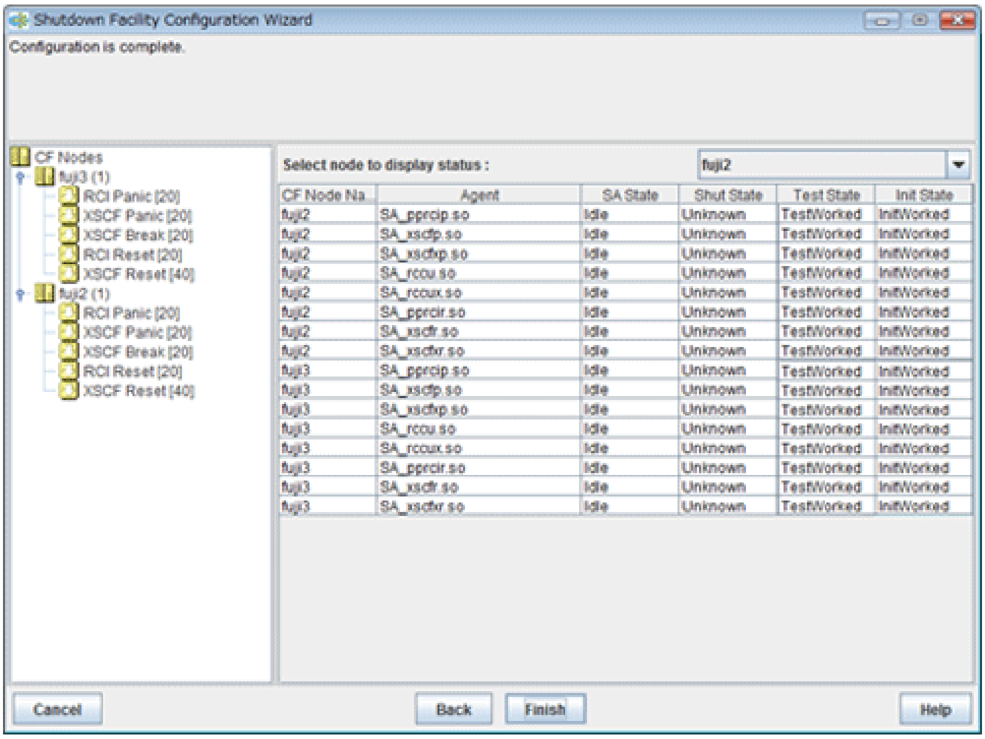
The IP address does not need to be changed if the host name registered to the /etc/inet/hosts file is displayed for the ILOM name.

Figure 9.6 ILOM (Oracle Solaris Kernel Zones)



14. After the IP address registered to the shutdown facility is changed, check the state of configuration of shutdown facility for each node by the display screen for the state of configuration of the shutdown facility.

Figure 9.7 Displaying the state of configuration of shutdown facility



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

- On this screen, confirm that the shutdown facility is operating normally.
  If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the IP address registered to the shutdown facility may contain an error. Reconfigure the IP address.

- When the connection to XSCF or ILOM is SSH connection, check that user inquiries of the first SSH connection (such as generation of the RSA key) have been completed by connecting to XSCF or ILOM from the cluster nodes via SSH using the log in user account for the shutdown facility after changing the IP address. In this case, if the host name is used for specifying XSCF name or ILOM name, specify this host name to check the SSH connection.

15. If you used the userApplication Configuration Wizard to set up an Ipaddress resource, take the following steps to change the takeover IP address in the resource database.

Note

Perform this step only when IPv4 Ipaddress resources are set through the userApplication Configuration Wizard. This step is not necessary when IPv6 Ipaddress resources are set.

1. Identify the takeover IP address resource to be changed in the resource database.

   Execute the "clgettree(1)" command on one of the cluster nodes. The takeover IP address resource has the resource class name for IP addresses.

   Example) In the following example, the resources with resource IDs 56 and 57 are the takeover IP address resources.

```
# /etc/opt/FJSVcluster/bin/clgettree
Cluster 1 cluster
     Domain 2 RC2
             Shared 7 SHD_RC2
                     SHD_Host 58 config_Ipaddress0_ip1 UNKNOWN
                             IP_Address 56 hme0:1 UNKNOWN narcissus
                             IP_Address 57 hme0:1 UNKNOWN sweetpea
             Node 3 narcissus ON
                     Ethernet 21 hme0 UNKNOWN
                             IP_Address 56 hme0:1 UNKNOWN
                     Application 31 proc1 UNKNOWN
                     DISK 19 c0t0d0 UNKNOWN
             Node 5 sweetpea ON
                     Ethernet 22 hme0 UNKNOWN
                             IP_Address 57 hme0:1 UNKNOWN
                     Application 32 proc1 UNKNOWN
                     DISK 20 c0t0d0 UNKNOWN
```

2. Identify the shared resources of the takeover IP address.

   To identify the shared resources, execute the "clgettree(1)" command on one cluster node.

   The shared resources of the takeover IP address are resources that have the SHD_Host resource class name.

   Example) For the example shown in Step 1), the resource that has resource ID 58 is the shared resource of the takeover IP address.

3. Change the takeover IP address.

   To change the address, execute the "clsetrsc(1M)" command. The takeover IP address is defined in the attributes of the shared resource for the takeover IP address that was identified in Step 2.

   On each cluster node, execute the "clsetrsc(1M)" command with the following format.

```
# /etc/opt/FJSVcluster/sys/clsetrsc -A ip_addr=new-IP-address resource-ID-of-shared-resource-having-IP-address
```

   Example) To change the takeover IP address of a shared resource (resource with resource ID=58) to 10.10.10.10

```
# /etc/opt/FJSVcluster/sys/clsetrsc -A ip_addr=10.10.10.10 58
```

4. Execute the "clgetrsc(1)" command to check that the takeover IP address was changed.

   Using the command format below, execute the "clgetrsc(1)" command on one of the cluster nodes and check that the takeover IP address was changed.

```
# /etc/opt/FJSVcluster/sys/clgetrsc -a ip_addr resource-ID-of-shared-resource-having-IP-address
```

   Example) If the takeover IP address of a shared resource (resource with resource ID=58) was changed to 10.10.10.10

```
# /etc/opt/FJSVcluster/sys/clgetrsc -a ip_addr 58
ip_addr 10.10.10.10
```

16. If you changed "1" to "0" in Step 7, change "0" to "1" again as follows.

```
# /opt/SMAW/SMAWRrms/bin/hvsetenv HV_RCSTART 1
# /opt/SMAW/SMAWRrms/bin/hvsetenv HV_RCSTART
1 <- Make sure this value is "1".
```

17. When the Migration function is used in Oracle VM Server for SPARC Environments, and the IP address of the control domain is changed at the same time, execute the following command for all the target guest domains to be migrated.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest-name guest-clustername guest-cfname
guest_ip guest-user
```

See

For details, see Step 2 in "17.2.1 Performing Live Migration of the Cluster on the Control Domains."

18. When the Migration function is used in Oracle VM Server for SPARC Environments, and the IP address of the guest domain is changed at the same time, take the following procedure.

    1. Execute the following command on any control domain.

    ```
    # /etc/opt/FJSVcluster/bin/clovmmigratesetup -m guest-name -p guest_ip
    ```

    2. Check that the target guest domain to be changed can be connected from all the control domains by using ssh. You need to complete the user inquiry of the first SSH connection (RSA key generation).
    For details, see Step 2 in "17.2.1 Performing Live Migration of the Cluster on the Control Domains."

Note

- An interface or IP address to which MAC address takeover has already been configured cannot be used in Web-Based Admin View.

- If a takeover IP address has been defined by GLS and if the takeover IP address is specified directly with an IP address value, not a host name, in the GLS environment settings and the Gls resource settings, first delete the Gls resource. Then change /etc/inet/hosts and the GLS environment settings, and register the Gls resource.
  For details, see "10.5.2 Supplement on Cluster Application and Resource Deletion," "6.2.1 GLS Setup," and "6.7.1.4 Creating Gls Resources."

9.2.2 Changing a CIP Address

This section describes how to change the IP address after the PRIMECLUSTER system is installed.

Operation Procedure:

1. Start all the nodes that constitute the cluster system.

2. Stop CF on all nodes that constitute the cluster system.

   For information on how to stop CF, see "5.7 Starting and stopping CF" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

3. While referring to the cip.cf file, confirm the CIP name to change the IP address.

   For details on the cip.cf file, see "2.2 CIP configuration file" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*" and the manual page describing cip.cf.

4. For the IPv6 address, edit the cip.cf file and change the IP address corresponding to the CIP name.

   When the original address and the modified address are both IPv4, you do not need to change it.

   Perform this procedure on all the nodes constituting the cluster system.

5. Change the IP address of the CIP name that is defined in the hosts(4) file.

   Perform this procedure on all the nodes constituting the cluster system.

6. Start CF on all the nodes constituting the cluster system.

   For instructions on how to start CF, see "5.7 Starting and stopping CF" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide.*"

7. Use the "ciptool(1M)" command to confirm that the IP address of CIP was changed.

```
# /opt/SMAW/SMAWcf/bin/ciptool -a
```

See

For details on the "ciptool(1M)" command, see the manual page describing ciptool(1M).

9.2.3 Changing the Subnet Mask of CIP

To change a subnet mask of CIP, perform the following procedure:

1. Migrate all nodes in single-user mode.

2. Change the subnet mask of CIP controlled in the cip.cf(4) file.

   This task is performed on all nodes configuring a cluster. For details on the cip.cf(4) file.

   Note

   Do not change anything other than a subnet mask for this file.

3. Start all nodes in multi-user mode.

9.2.4 Changing Port Numbers for SNMP

This section describes the procedures to change port numbers for SNMP.

Preparation

Decide the new port number after change.

Procedure

1. Execute the following command on one of the cluster nodes to stop RMS operation:

```
# /opt/SMAW/bin/hvshut -a
```

2. Execute the following command on each cluster node to stop the shutdown facility:

```
# /opt/SMAW/bin/sdtool -e
```

3. On each cluster node, change the port number of the SNMP trap receiving daemon (snmptrapd) used in the shutdown facility.

   Edit the /etc/inet/services file and change the port number of the "sftrap" service.

```
sftrap <Changed port number>/udp
```

4. Change the port number of the SNMP agent in XSCF on each cluster node.

   a. When changing the port number

      Execute the following commands in XSCF on each node to set the changed port number:

```
XSCF> setsnmp remtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN]
XSCF> setsnmp remtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
```

```
monitoring sub-LAN]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p <Changed port number> [IP address of the
administrative LAN]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p <Changed port number> [IP address of the
asynchronous monitoring sub-LAN]
```

b. When setting back the changed port number to 9385 used in the shutdown facility

```
XSCF> setsnmp remtraphost -t v2 -s FJSVcldev -p <Changed port number> [IP address of the
administrative LAN]
XSCF> setsnmp remtraphost -t v2 -s FJSVcldev -p <Changed port number> [IP address of the
asynchronous monitoring sub-LAN]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the administrative LAN]
XSCF> setsnmp addtraphost -t v2 -s FJSVcldev -p 9385 [IP address of the asynchronous
monitoring sub-LAN]
```

5. Execute the following command on each node to restart the system:

```
# /usr/sbin/shutdown -y -i6 -g0
```

9.3 Changing the Operation Environment for Hardware

9.3.1 Changing Settings for the Shared Device Connection Confirmation Feature

For instructions on changing the operation environment for shared disk connection confirmation, see "5.3 Setting Up Shared Disk Connection Confirmation."

9.3.2 Changing the Operation Environment for Patrol Diagnosis

To change the operation environment for patrol diagnosis, execute the "clspconfig(1M)" command. For details on the "clspconfig(1M)" command, see the manual page for clspconfig(1M).

9.4 Changing Option Hardware Settings

9.4.1 Changing the User Name and Password to Control the Console

This section explains how to change the user name and password to control the console after installation of the PRIMECLUSTER system.

Operation Procedure:

1. Execute the following command on one of the cluster nodes to stop RMS operation.

```
# /opt/SMAW/bin/hvshut -a
```

2. Execute the following command on each cluster node to stop the shutdown facility.

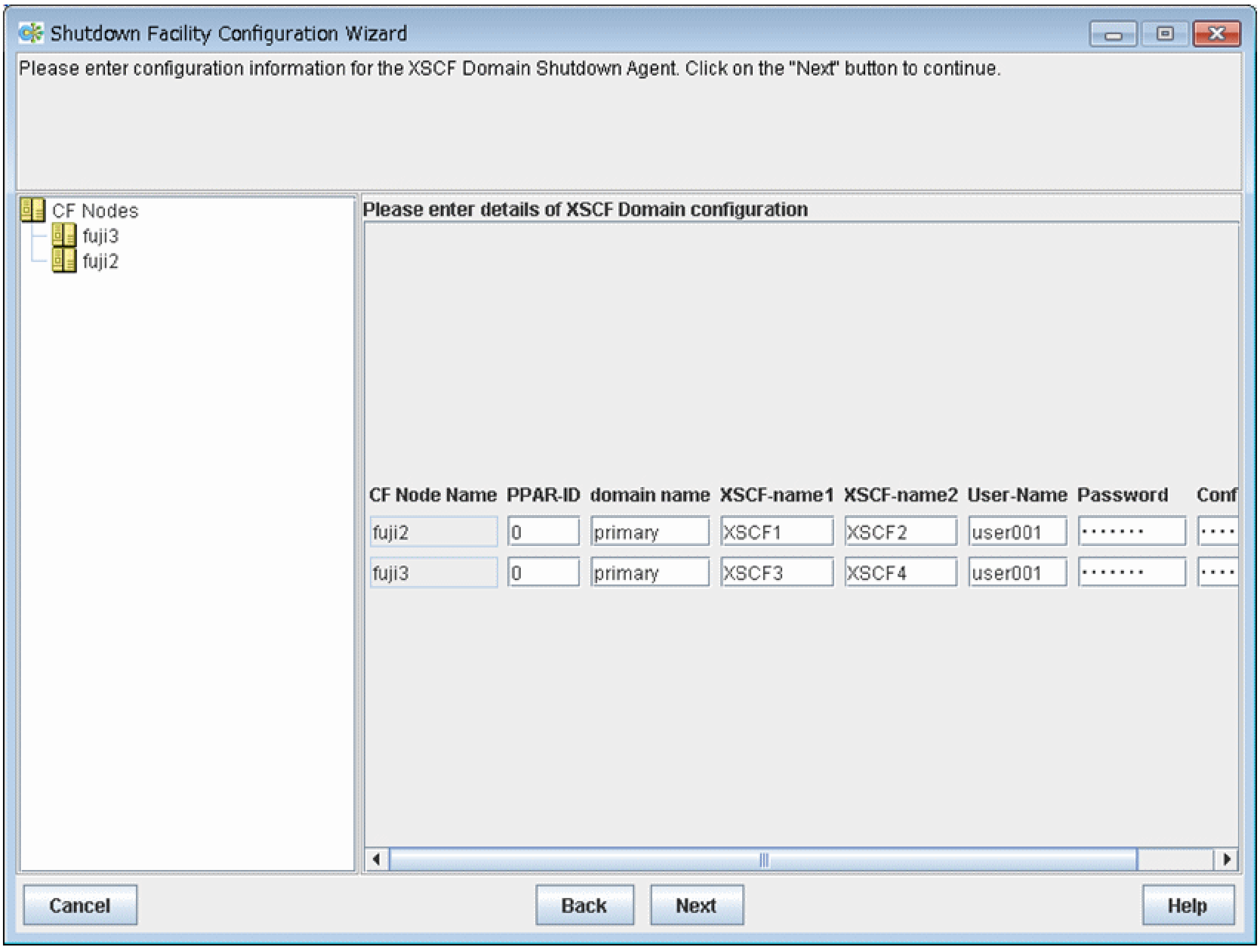
```
# /opt/SMAW/bin/sdtool -e
```

3. Change the following user names and passwords registered to the shutdown facility as necessary:

a. User name and password to login to XSCF (for SPARC M10)

For XSCF, see "5.1.2.1.3 Using the Shutdown Configuration Wizard" and reconfigure the user name and password to log in to XSCF.

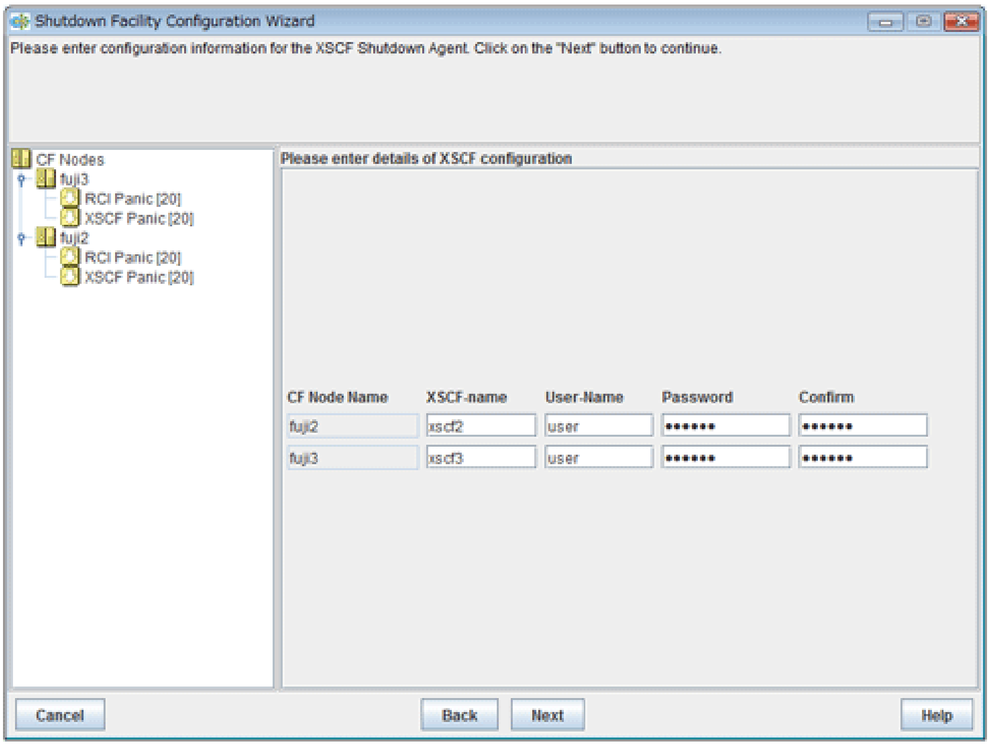
Figure 9.8 XSCF (SPARC M10)



b. User name and password to login to XSCF (for SPARC Enterprise M3000, M4000, M5000, M8000, and M9000)

For XSCF, see "5.1.2.2.2 Using the Shutdown Configuration Wizard" and reconfigure the user name and password to login to XSCF.

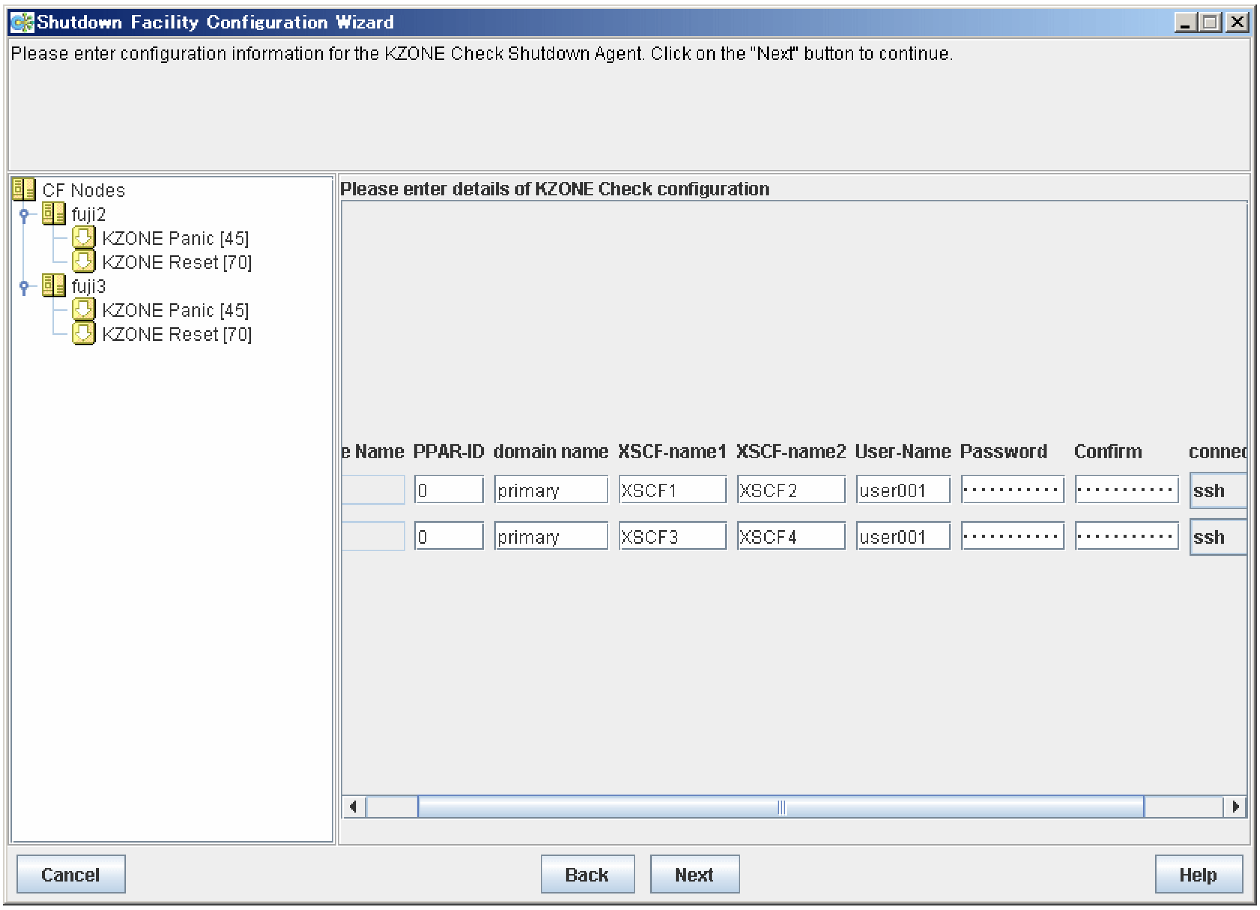
Figure 9.9 XSCF (SPARC Enterprise M3000, M4000, M5000, M8000, and M9000)



c. User name and password to login to XSCF (for Oracle Solaris Kernel Zones)

For XSCF, see "5.1.2.5.4 Using the Shutdown Configuration Wizard" and reconfigure the user name and password to login to XSCF.

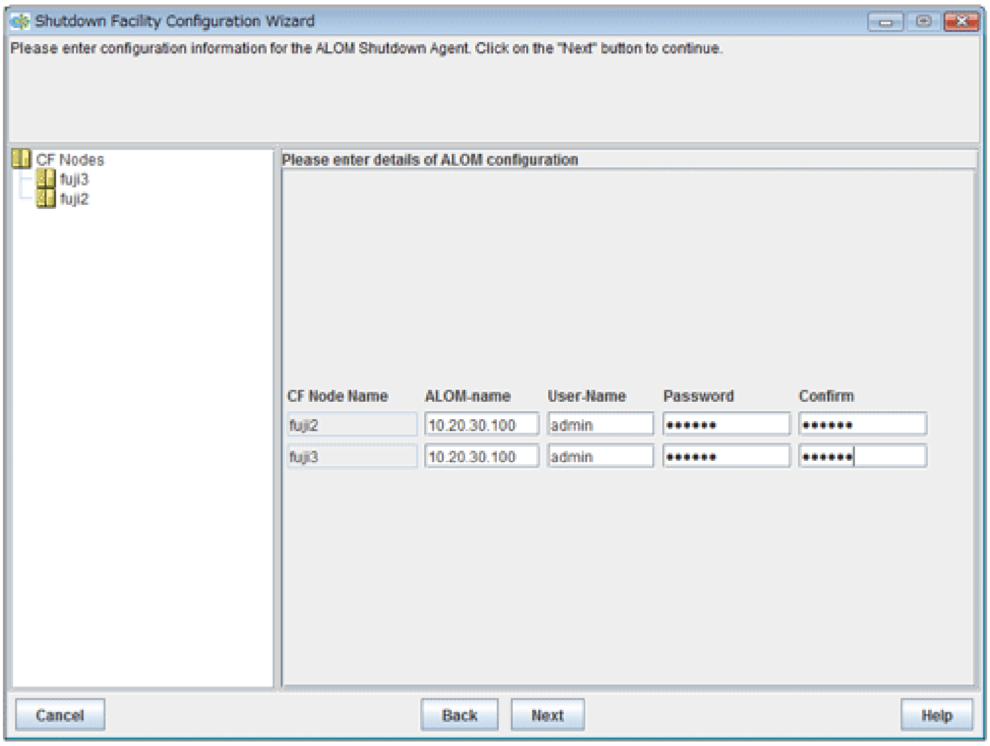
Figure 9.10 XSCF (Oracle Solaris Kernel Zones)



d. User name and password to login to ALOM

For ALOM, see "5.1.2.4.2 Using the Shutdown Configuration Wizard" and reconfigure the username and password to login to ALOM.

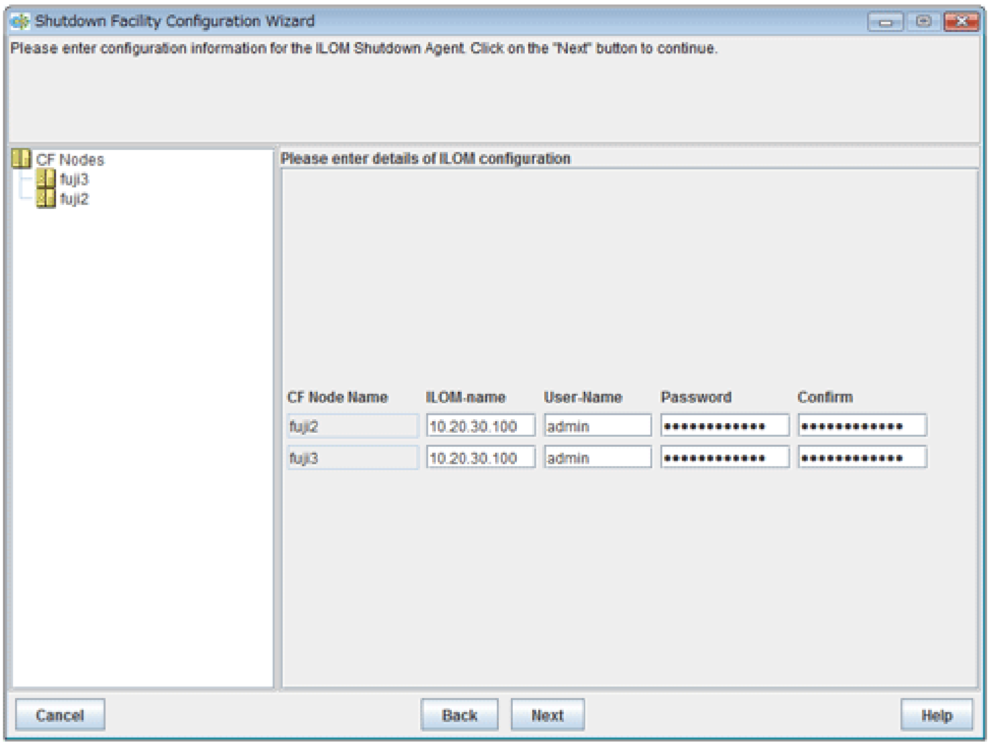
Figure 9.11 ALOM



e. User name and password to login to ILOM (for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series).

For ILOM, see "5.1.2.3.2 Using the Shutdown Configuration Wizard" and reconfigure the user name and password to login to ILOM.

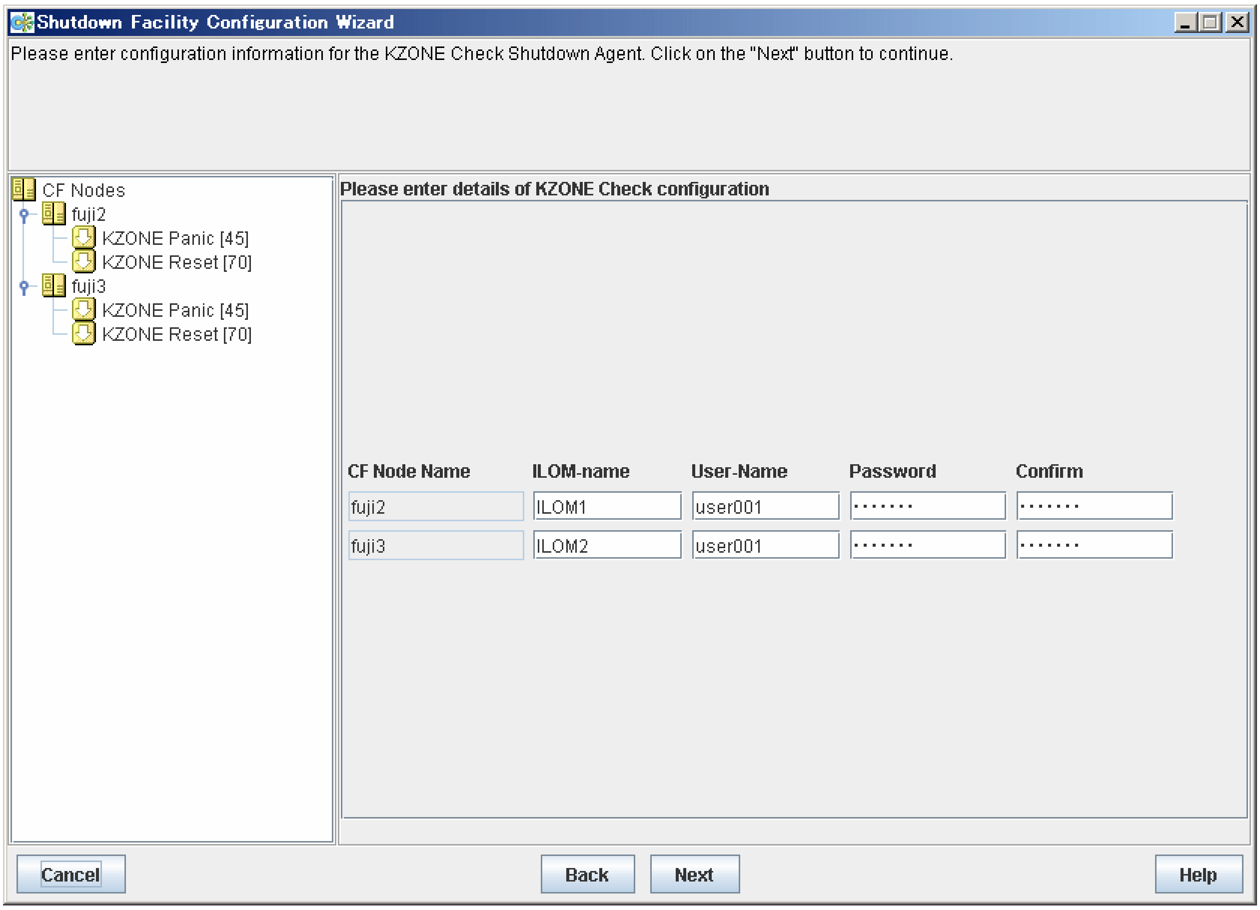
Figure 9.12 ILOM (SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series)



f. User name and password to login to ILOM (for Oracle Solaris Kernel Zones)

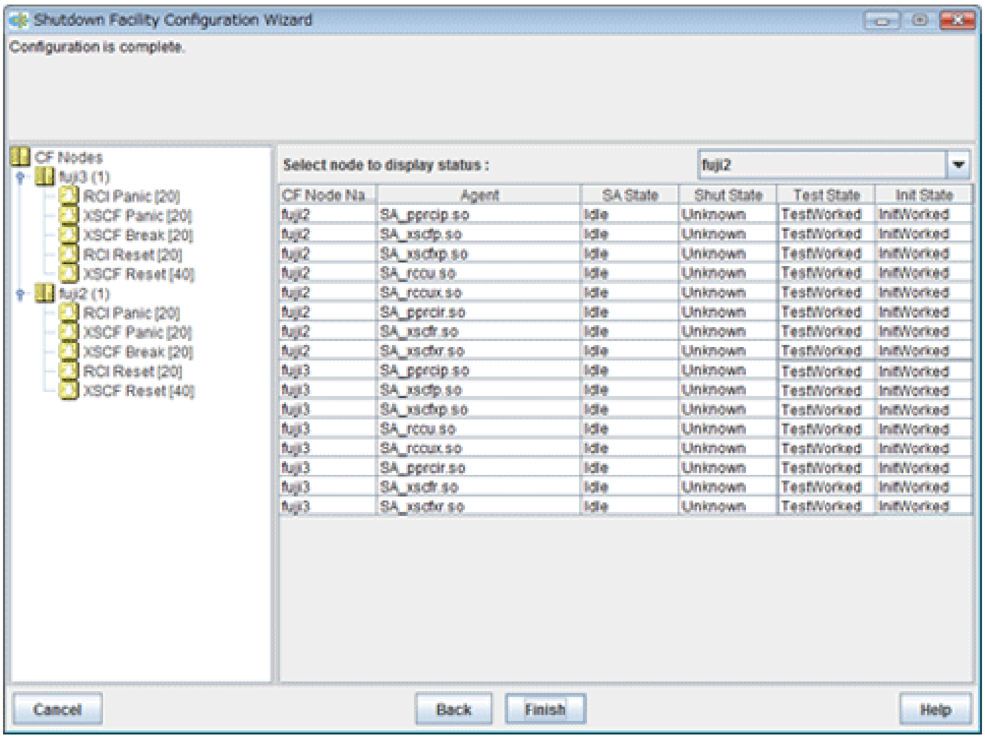
For ILOM,"5.1.2.5.4 Using the Shutdown Configuration Wizard" to reconfigure the user name and password to login to ILOM.

Figure 9.13 ILOM(Oracle Solaris Kernel Zones)



4. After the IP address registered to the shutdown facility is changed, check the state of configuration of shutdown facility for each node by the display screen for the state of configuration of shutdown facility.

Figure 9.14 Displaying the state of configuration of shutdown facility



Shut State

"Unknown" is shown during normal system operation. If an error occurs and the shutdown facility stops the relevant node successfully, "Unknown" will change to "KillWorked".

Test State

Indicates the state in which the path to shut down the node is tested when a node error occurs. If the test of the path has not been completed, "Unknown" will be displayed. If the configured shutdown agent operates normally, "Unknown" will be changed to "TestWorked".

Init State

Indicates the state in which the shutdown agent is initialized.

To exit the configuration wizard, click *Finish*. Click *Yes* in the confirmation popup screen that appears.

Note

On this screen, confirm that the shutdown facility is operating normally.

If "InitFailed" is displayed in the Initial state even when the configuration of the shutdown facility has been completed or if "Unknown" is displayed in the Test state or "TestFailed" is highlighted in red, the user name and password registered to the shutdown facility may contain an error. Reconfigure the user name and password.

5. Execute the following command in one of the cluster nodes to start RMS operation:

```
# /opt/SMAW/bin/hvcm -a
```

Chapter 10 Changing the Cluster Application Configuration

This chapter describes how to add, delete, or change a cluster application, a resource, or so on.

10.1 Adding a Cluster Application

For how to add a cluster application, see "6.7.2 Creating Cluster Applications."

Note

Be sure to stop RMS before adding a cluster application. If the "userApplication Configuration Wizard" screen is started while RMS is operating, message 0833 will be displayed. If this happens, stop RMS, and then start the "userApplication Configuration Wizard" screen again. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

10.2 Deleting a Cluster Application

This section explains how to delete a cluster application without deleting its resources.

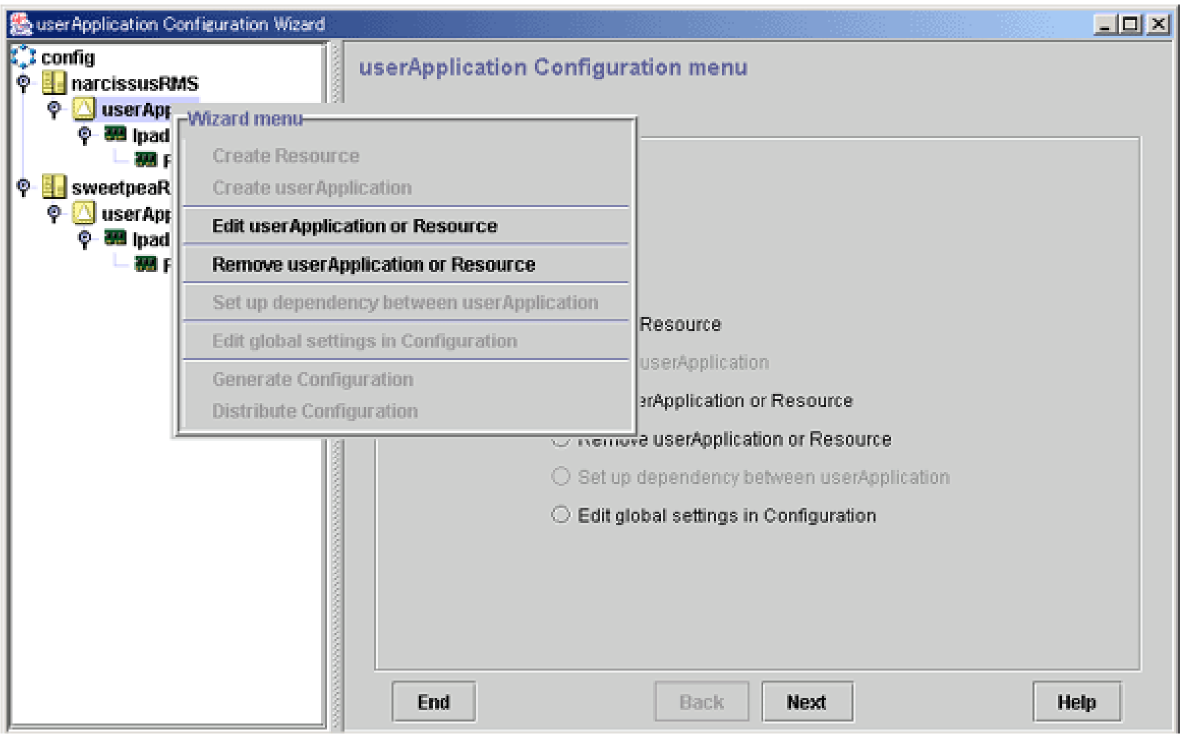
Note

Be sure to stop RMS before deleting a cluster application. If the "userApplication Configuration Wizard" screen is started while RMS is operating, message 0833 will be displayed. If this happens, stop RMS, and then start the "userApplication Configuration Wizard" screen again. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

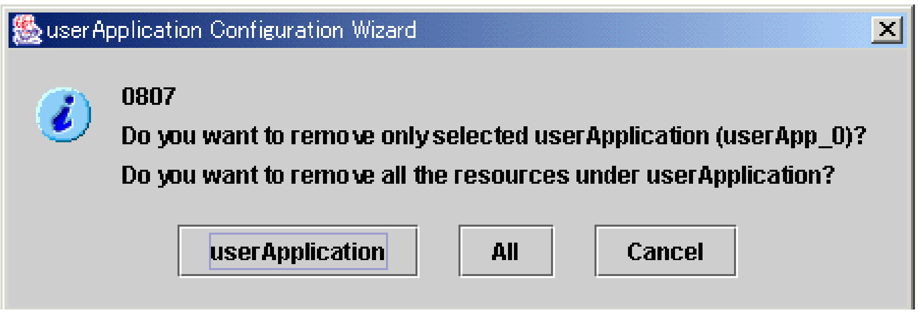
Operation Procedure:

1. At the top menu of the *userApplication Configuration Wizard*, select userApplication to be deleted from the tree shown on the left of the screen.

2. Click the right mouse button to display the pop-up menu, and select *Remove userApplication or Resource*.



3. At the displayed dialog box, select *userApplication*. Only the userApplication will be deleted.



userApplication

Select this button to delete the specified userApplication.

Note

The resources that are specified to userApplication will not be deleted.

The hierarchical relationships of the resources will be initialized. However, levels that were set with "Resource association" will not be initialized.

All

Select this button to delete the specified userApplication and all related resources.
If resources are to be deleted, delete as described in the "10.5.2 Supplement on Cluster Application and Resource Deletion."

Cancel

Select this button to cancel the deletion.

You can confirm that userApplication has been deleted by observing that the userApplication no longer appears in the tree on the "userApplication Configuration Wizard" screen.

10.3 Changing a Cluster Application

This section describes how to change a cluster application.

Note

- Be sure to stop RMS before changing a cluster application. If you start the "userApplication Configuration Wizard" screen while RMS is running, the message 0833 will be displayed. If this happens, stop RMS, and then restart the "userApplication Configuration Wizard" screen. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."
- If there are resources to be removed from the monitoring target by the standby patrol function after a cluster configuration is changed, delete resources from the target of the standby patrol function by using the following command after stopping RMS:
    - For a physical disk defined by a disk class

```
# /etc/opt/FJSVcluster/bin/clspconfig -u <cluster application name> -d TARGET_DISK=<target
resource>
```

    - For a network interface card

```
# /etc/opt/FJSVcluster/bin/clspconfig -u <cluster application name> -d TARGET_LAN=<target
resource>
```

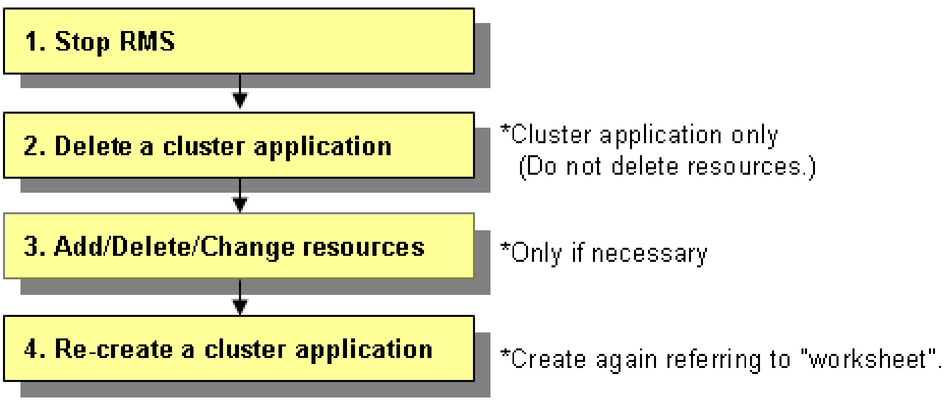
Information

After you finish changing the cluster application, execute "RMS startup" from the Cluster Admin screen. If necessary, also start the cluster application. For instructions on starting RMS, see "7.2.1.1 Starting RMS." For instructions on starting the cluster application, see "7.2.2.1 Starting a Cluster Application."

10.3.1 Changing the Cluster Application Configuration

To change the configuration of a cluster application, you must first delete the cluster application. The procedure is shown below.

Operation flow



Operation Procedure:

1. Stop RMS.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Delete the cluster application.

   Delete only the target cluster application.

For details on how to make this deletion, see "10.2 Deleting a Cluster Application."

Note

Do not delete a resource that is necessary to configure the cluster application.

3. Add, delete, or change the resource.

   Add a new resource and delete any unnecessary resources.

   See

   For details on how to perform the above operations, see the following:

   When a resource is to be added: 6.7.1 Setting Up Resources

   When a resource is to be deleted: 10.5 Deleting a Resource

   When a resource is to be changed: 10.6 Changing a Resource

   Note

   If you have changed a resource, check that all resources are registered to the cluster application. If there are resources that are not registered to the cluster application, delete them or add them to the cluster application.

4. Create a cluster application.

   Create the same cluster application as that deleted in Step 2.

   For details on how to create the cluster application, see "6.7.2 Creating Cluster Applications." If cluster application dependency relationships have been set up, they must be set up again.
   For details on how to make these settings, see "6.7.3 Setting Up Dependency Relationships Between Cluster Applications."

   Note

   If it is not necessary to change the operation attributes of the cluster application, leave the attribute unchanged as explained in the "A.9 Cluster Application Worksheets."

10.4 Adding a Resource

For how to add a resource, see "6.7.1 Setting Up Resources."

Note

Be sure to stop RMS before adding a resource. If the "userApplication Configuration Wizard" screen is started while RMS is operating, message 0833 will be displayed. If this happens, stop RMS, and then start the "userApplication Configuration Wizard" screen again. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

10.4.1 Adding Fsystem Resource Dynamically

This section describes how to add Fsystem resources without stopping jobs.

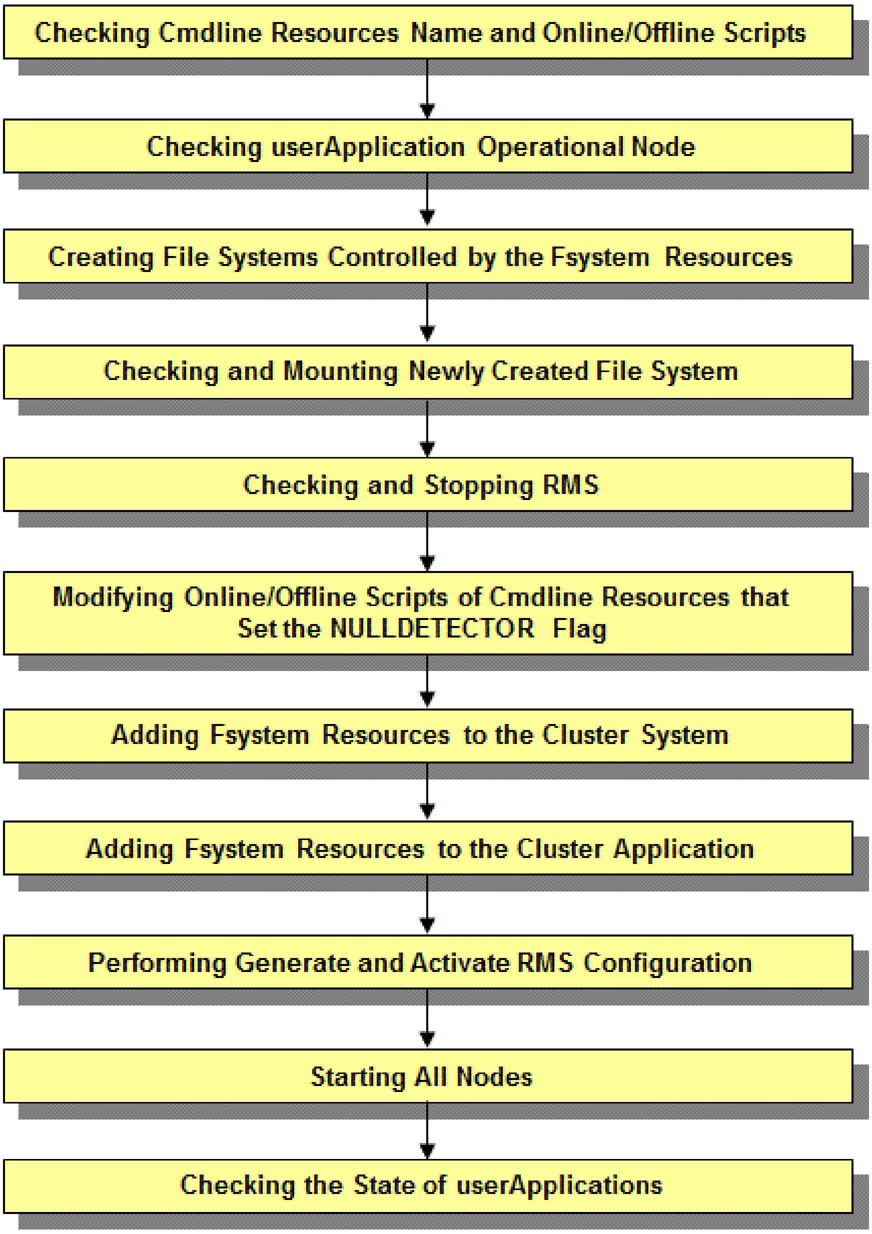
Note

In the dynamic changing configuration, RMS is stopped with the cluster application operating.

When RMS is disabled, a cluster application is not failed over if an error occurs in the cluster application. In this case, to minimize the shutdown time of RMS, check the following operation procedure carefully, then investigate and sort out the necessary operating steps.

Moreover, disable the failover report function or take another action if necessary when using middleware that notifies an error when RMS is stopped.

Operation flow



Note

This procedure is necessary when performing the following operations:

- Newly adding the ZFS storage pool
- Newly adding the legacy ZFS file systems to the existing ZFS storage pool
- Newly adding the UFS file systems

When newly adding the non-legacy file systems to the existing ZFS storage pool, the procedure explained in this section is unnecessary.

In this case, by setting the cluster application to the maintenance mode in advance, the ZFS file system can be added dynamically while RMS is in operation.

For how to use the maintenance mode, refer to "8.4 Using maintenance mode" in "Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Operation Procedure:

1. Check Cmdline Resources Name and Online/Offline Script.

   Check the resource name of the Cmdline resource by "hvdisp -T gResource" command when the Cmdline resource is included in the cluster application.

   If the Cmdline resource name contains the resource name that starts with "RunScriptsAlways", the NULLDETECTOR flag is set to that resource.

   #### Example

   When the execution result of the hvdisp command is the following, it can be judged that the NULLDETECTOR flag is set to the Cmdline resource RunScriptsAlways001_Cmd_APP1 and the Cmdline resource RunScriptsAlways001_Cmd_APP2.

```
# hvdisp -T gResource
Local System:  node01RMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us

Resource            Type    HostName            State        StateDetails
-----------------------------------------------------------------------------
RunScriptsAlways001_Cmd_APP2 gRes                         Online
ManageProgram000_Cmd_APP2 gRes                         Online
RunScriptsAlways001_Cmd_APP1 gRes                         Offline
ManageProgram000_Cmd_APP1 gRes                         Offline
```

   It is necessary to add the processing described in "6.12.2.1.4 Notes When Setting the NULLDETECTOR Flag" to the Online/Offline scripts of the Cmdline resource when the NULLDETECTOR flag is enabled.

   Modify the script after stopping RMS according to the following procedure when the necessary processing is not included.

2. Check userApplication Operational Node.

   Check that the standby userApplication operates in which node in the cluster (Which node is the operational node?) by the hvdisp -T userApplication command.

   #### Example

   When the execution result of the hvdisp command is the following, the operational node of app1 is node02 and the operational node of app2 is node01.

```
# hvdisp -T userApplcation
Local System:  node01RMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us

Resource            Type    HostName            State        StateDetails
-----------------------------------------------------------------------------
app2                userApp                     Online
app1                userApp                     Standby
app1                userApp node02RMS           Online
```

   When determining the node that mounts the file system manually according to the following procedure, information of the operation node of the cluster application is necessary.

3. Create File Systems Controlled by the Fsystem Resources.

   When the mount point controlled by the Fsystem resource is created on the new volume of GDS, create the file system after starting the volume of GDS on operating node.

   Information

   For details on the procedure for using ZFS as Fsystem resource, see "6.4.1 If using ZFS."

4. Check and Mount Newly Created File System.

   On the operation node of userApplication that adds the Fsystem resources according to Step 2, mount the newly created file system and check that the mount is correctly done.

   Example

   If using ZFS

   Below is an example when adding the ZFS storage pool: app2 and also mounting the legacy ZFS file systems to /appdata.

   Execute the following command on the operational node to import the ZFS storage pool (when the ZFS storage pool: app2 is created on the volume of disk class *class2* of GDS).

```
# /usr/sbin/zpool import -d /dev/sfdsk/class2/dsk -R "/" app2
```

   Execute the following command on the operational node to mount the ZFS file system (if the ZFS file system of the non-legacy file systems is set to the ZFS storage pool, the ZFS file system is mounted when importing the ZFS storage pool. The following procedure is unnecessary).

```
# /usr/sbin/mount -F zfs app2/mp /appdata
```

   After mounting, execute the following command to check that the ZFS storage pool and the mount point are displayed (the file system is mounted).

```
# /usr/sbin/zfs list -r app2
NAME          USED  AVAIL  REFER  MOUNTPOINT
app2          148K   976M    31K  none
app2/mp        31K   976M    31K  legacy
# /usr/bin/df -k | /bin/grep "app2/mp "
app2/mp              999424          31       999276      1%    /appdata
```

   Additionally, check that the file system is not mounted on the standby node.

   If using UFS

   Below is an example when mounting the UFS file systems on the GDS volume to /disk2.

   Execute the following command on the operational node to mount the UFS file systems.

```
# /usr/sbin/mount -F ufs /dev/sfdsk/class0002/dsk/volume0001 /disk2
```

   After mounting, execute the following command to check that the mount point is displayed (the file system is mounted).

```
# /usr/bin/df -k | /bin/grep "/disk2"
/dev/sfdsk/class0002/dsk/volume0001     999424     31      999276      1%    /disk2
```

   Additionally, check that the file system is not mounted on the standby node.

5. Stop RMS.

   Execute the hvshut -L command on all nodes to stop RMS when cluster application is still operating.

   Enter 'yes' in response to the warning message when the hvshut -L command is executed.

```
# hvshut -L
                        WARNING
                        -------
The '-L' option of the hvshut command will shut down the RMS
software without bringing down any of the applications.
In this situation, it would be possible to bring up the same
application on another node in the cluster which *may* cause
data corruption.

Do you wish to proceed ? (yes = shut down RMS / no = leave RMS running).
yes
```

See

For details on the hvshut command, see the manual page of the "hvshut(1M)" command.

6. Check the Stop of RMS.

   Execute the hvdisp -a command on all nodes. If RMS has stopped, the command outputs the standard error output "hvdisp: RMS is not running".

   ```
   # hvdisp -a
   hvdisp: RMS is not running
   ```

7. Modify the Online/Offline Scripts of the Cmdline Resources when NULLDETECTOR Flag is Enabled if Necessary.

   As a result of the check of Step 1, if the correction is necessary for the Online/Offline scripts of the Cmdline resources when NULLDETECTOR flag is enabled, see "6.12.2.1.4 Notes When Setting the NULLDETECTOR Flag" to modify the scripts.

8. Add Fsystem Resources to the Cluster System.

   Perform the following procedures that are described in "6.7.1.2 Creating Fsystem Resources".

   Example

   - When adding the ZFS storage pool: app2 and also mounting the legacy ZFS file systems to /appdata, write /etc/vfstab.pcl as follows.

     ```
     #RMS#app2 app2 /app2 zfs - - -
     #RMS#app2/mp app2/mp /appdata zfs - - -
     ```

   - When adding the UFS file systems, write the /etc/vfstab.pcl file as follows.

     ```
     #RMS#/dev/sfdsk/class0002/dsk/volume0001 /dev/sfdsk/class0002/rdsk/volume0001 /disk2 ufs -
     no -
     ```

   When the mount point controlled by the Fsystem resource is created on the new class of GDS, see "6.7.1.3 Creating Gds Resources."

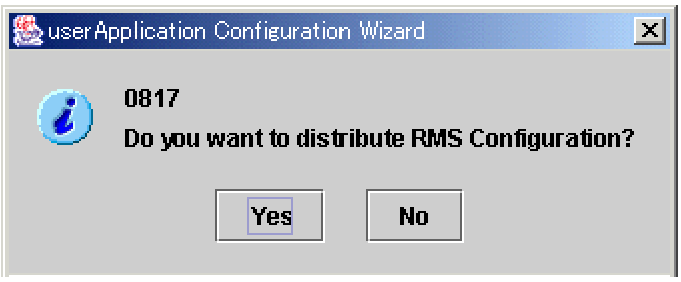
9. Add Fsystem Resources to the Cluster Application.

   When adding the Fsystem Resources added in Step 8 to the existing cluster application, see "10.3.1 Changing the Cluster Application Configuration" to delete the target cluster application, and then create a new cluster application again.

   When newly creating a cluster application, see "6.7.2 Creating Cluster Applications" to create a cluster application.

10. Perform Generate and Activate of RMS Configuration.

    After the cluster application was registered to RMS Configuration in Step 9, if it is judged that RMS Configuration can be distributed, the following message is displayed.



If there are neither other resource to be added nor cluster applications to be changed, click *Yes*.

11. Start RMS on All Nodes.

    Execute the hvcm - a command on any one node to start RMS on all nodes.

```
# hvcm -a
```

12. Check the State of userApplications.

    Execute the hvdisp -a command on all nodes, and check that the state of userApplication is Online on operational node and the state of userApplication is Offline or Standby on standby node according to Step 2.

    Note

    UserApplication will be Inconsistent state on any of the nodes or both the nodes after starting RMS in Step 11 when the mount of file system is not correctly operated according to Step 4. In this case, perform the following procedures.

    1. Execute the hvutil -f command on the standby node so that the state of userApplication on the standby node becomes Offline.
    2. When userApplication on the standby node is transited to Standby, execute the hvutil -s command on the standby node.
    3. Execute the hvswitch command on the operational node so that the state of userApplication on the operational node becomes Offline.

10.5 Deleting a Resource

This section describes how to delete a resource.

Note

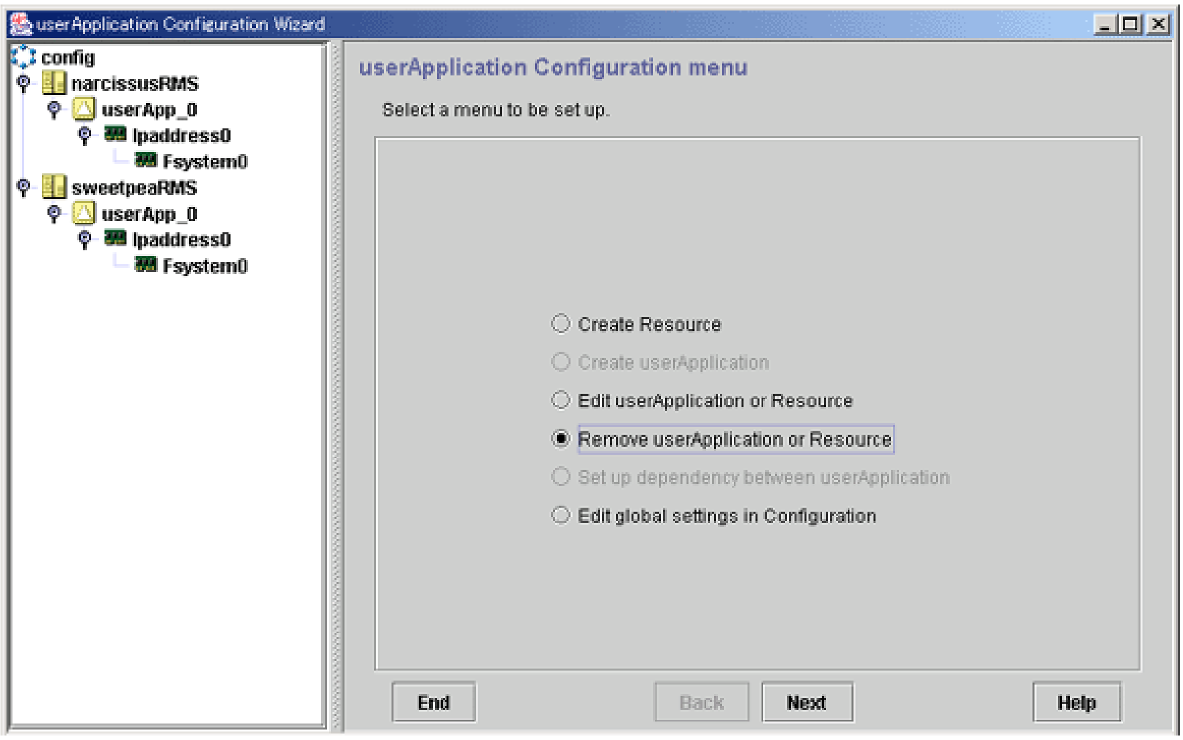
Be sure to stop RMS before deleting a resource. If the "userApplication Configuration Wizard" screen is started while RMS is operating, message 0833 will be displayed. If this happens, stop RMS, and then start the "userApplication Configuration Wizard" screen again. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

10.5.1 Deleting a Resource to Be Used by the Cluster Application

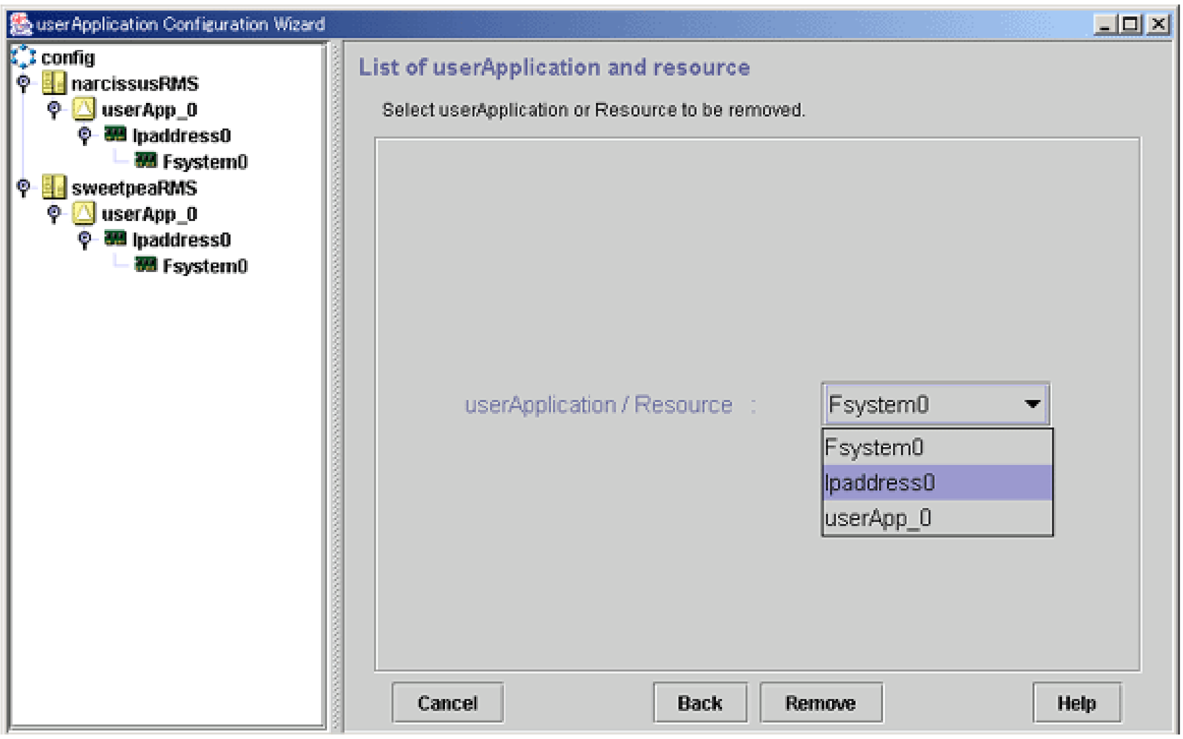
This section describes how to delete a resource that is to be used by the cluster application.

Operation Procedure:

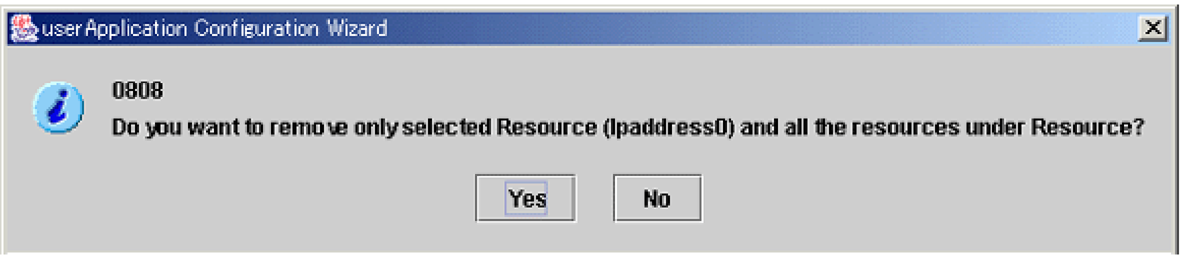
1. At the top menu of the *userApplication Configuration Wizard*, select *Remove userApplication or Resource*.



2. Select the resource to be deleted, and then select *Remove*.



3. Respond to the displayed dialog box.



Yes

Select this button to delete the specified resource and all subordinate resources.
If deleting the selected resources deletes all the resources that construct userApplication, message 0835 will confirm with you whether you want to delete all the resources. To continue the deletion, select "*Yes*".

No

Select this button to cancel the deletion.

You can confirm the deletion by checking that the resource disappears in the tree on the "userApplication Configuration Wizard" screen.

Note

When deleting resources, refer to the "10.5.2 Supplement on Cluster Application and Resource Deletion."

10.5.2 Supplement on Cluster Application and Resource Deletion

This section explains the operations to be done after "10.2 Deleting a Cluster Application" and "10.5.1 Deleting a Resource to Be Used by the Cluster Application."

10.5.2.1 Supplement on Cmdline resource deletion

If the "userApplication Configuration Wizard" screen is used to create new script files for Cmdline resource creation, follow the procedures below to delete each script which is used for the Cmdline resource.

Note

Before deleting scripts, you need to delete the Cmdline resources.

Each script is stored in the following directories:

- /etc/opt/FJSVwvucw/scripts/start
- /etc/opt/FJSVwvucw/scripts/stop
- /etc/opt/FJSVwvucw/scripts/check

Execute the deletion command as follows:

```
# cd /etc/opt/FJSVwvucw/scripts/start

# rm <script name to be deleted>

# cd /etc/opt/FJSVwvucw/scripts/stop

# rm <script name to be deleted>

# cd /etc/opt/FJSVwvucw/scripts/check
```

```
# rm <script name to be deleted>
```

10.5.2.2 Supplement on Fsystem resource deletion

Restore the environment files related to the mount point to the original state.

Note

Before deleting the environment files, you need to delete the Fsystem resources.

Deleting mount point entries that were added at prerequisites

Delete the relevant entry from the following environment files that were edited as part of the "6.7.1.2.1 Prerequisites."

- /etc/vfstab.pcl
- /etc/dfs/dfstab.pcl

If a dummy entry is deleted from /etc/dfs/dfstab.pcl, delete the actual directory.

When an NFS Lock Failover was used

If the NFS Lock Failover functions are no longer used due to deletion of the Fsystem resources, take the following steps:

Procedure:

1. Change the settings for shared information with the configuration.

   Change the value of NFSLockFailover to "no."

See

See "6.7.4 Editing global settings in Configuration."

10.5.2.3 Supplement on Gds resource deletion

Delete the disk class if it is no longer necessary.

See "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

10.5.2.4 Supplement related to Gls resource deletion

Delete the takeover network definition if it is no longer necessary.

See "5.4 Deleting configuration for Cluster System" in the "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*."

10.5.2.5 Supplement on takeover network resource deletion

Restore the environment files related to the takeover network to the original state.

When node name takeover was used

Change the node name in the cluster system used for node name takeover. To bring the system state back to the state before the change, take the following steps on all the nodes where node name takeover resources were created.

Operation

1. Execute the command on all the nodes, as follows:

```
# setuname -n fuji2
```

fuji2 is used as the node name in this example.

2. Reboot the system.

```
# /usr/sbin/shutdown -g0 -i6 -y
```

10.5.2.6 Supplement on procedure resource deletion

Delete the procedure resource from the cluster resource management facility when it is no longer necessary.

See "E.3 Deleting a Procedure Resource."

10.5.2.7 Supplement on process monitoring resource deletion

Delete scripts that were created automatically if the stop command of the process monitoring resource was used.

Note

Before deleting the created scripts, you need to delete the process monitoring resources.

Each script is stored in the /etc/opt/FJSVwvucw/scripts/process_stop directory.

Execute the deletion command as shown below.

```
# cd /etc/opt/FJSVwvucw/scripts/process_stop
# rm ./*
```

10.5.2.8 Supplement on line switching unit resource deletion

Delete the line switching unit resource from the cluster resource management facility if it is not necessary.

See the manual pages of the "cldelswursc(1M)" command.

10.5.2.9 ISV resource deletion procedure (supplement)

For information on the deletion procedure, see the manuals provided with each product.

10.5.3 Deleting the Hardware Resource

Delete the hardware resource that was registered with the automatic configuration facility using the following procedure.

Procedure

1. Stop RMS.

   If RMS is activated, stop RMS as explained in "7.2.1.2 Stopping RMS."

2. Delete the hardware resource.

   Use each command to delete the hardware resource that was registered with the automatic configuration facility.

   See the relevant manual pages for details on each command.

Resource	Command
Shared disk device	cldeldevice(1M)
Network interface card	cldelrsc(1M)
Line switching unit (only in Oracle Solaris 10 environment)	cldelswursc(1M)

Note

If you have deleted an available network interface card by mistake, reregister the resources for the accidentally deleted network interface card by executing the "clautoconfig(1M)" command.

10.6 Changing a Resource

The resource change includes the following:

1. Change the interface used by a resource.
2. Change the attributes to the resources and resource interfaces.

Note

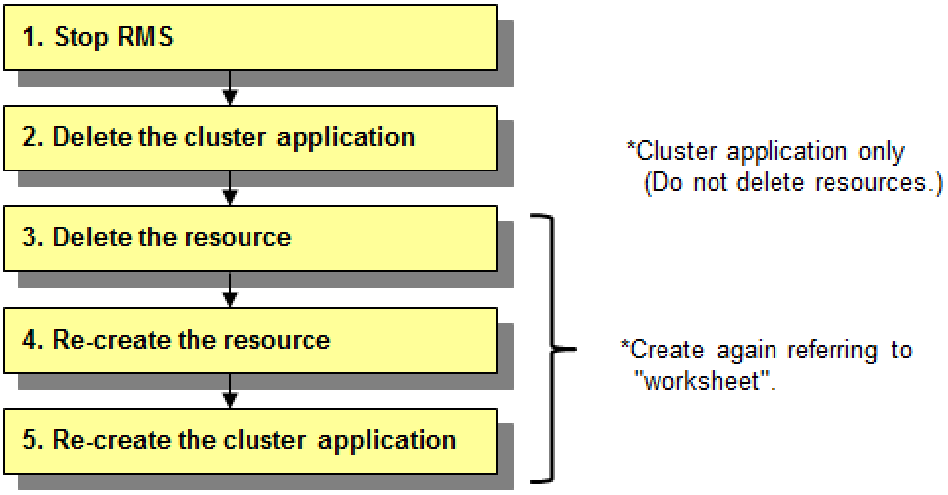
Be sure to stop RMS before changing a resource. If the "userApplication Configuration Wizard" screen is started while RMS is operating, message 0833 will be displayed. If this happens, stop RMS, and then start the "userApplication Configuration Wizard" screen again. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

10.6.1 Changing the Interface Used by a Resource

To change the interface used by a resource, you must first delete the target resource and cluster application, then create them again. Interface examples include a takeover network for an Ipaddress and a disk class for a Gds resource.

The procedure is as follows.

Operation flow



Operation Procedure:

Use the "userApplication Configuration Wizard" screen to change a resource used by a cluster application.

1. Stop RMS.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

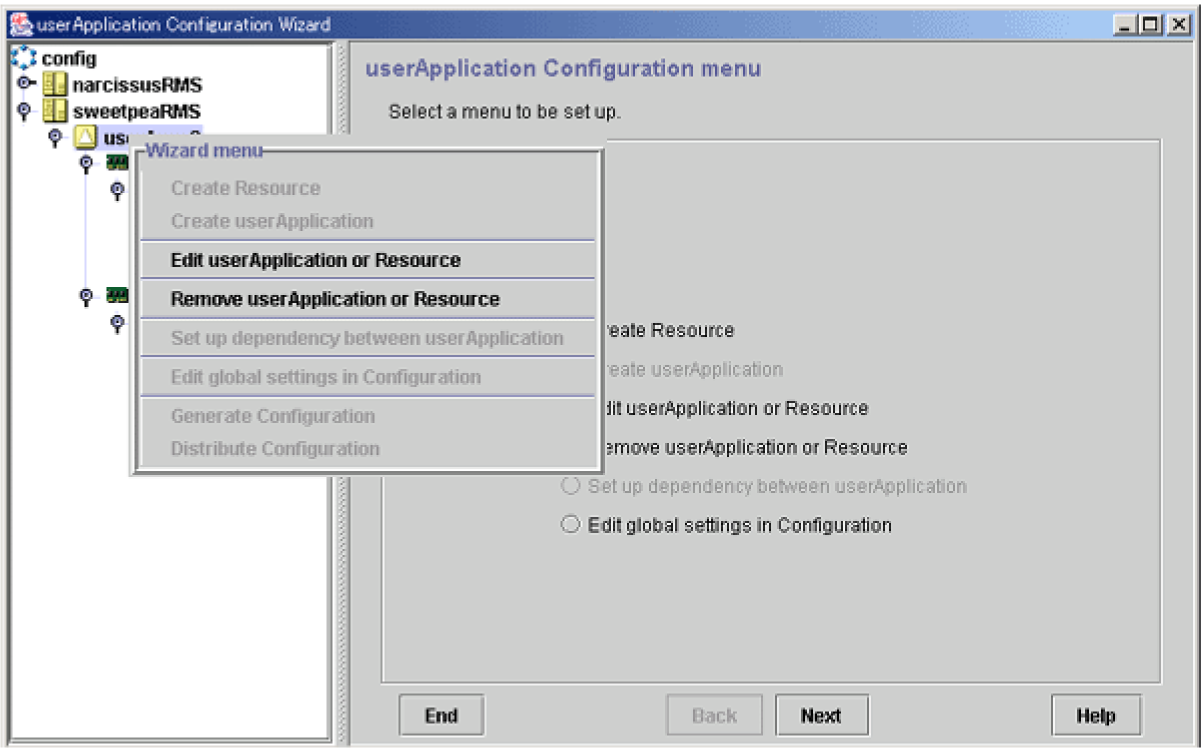
2. Delete the cluster application.

   Delete only the cluster application that contains the target resource.

   For how to delete it, see "10.2 Deleting a Cluster Application."

Note

Make sure not to delete the resource that constitute a cluster application.



Note

If you delete a resource included in the cluster application, the other resources found under the target resource are also deleted.

In other words, the entire hierarchy of objects found under the resource on the tree in the left panel will be deleted regardless of those objects being related to other resources.

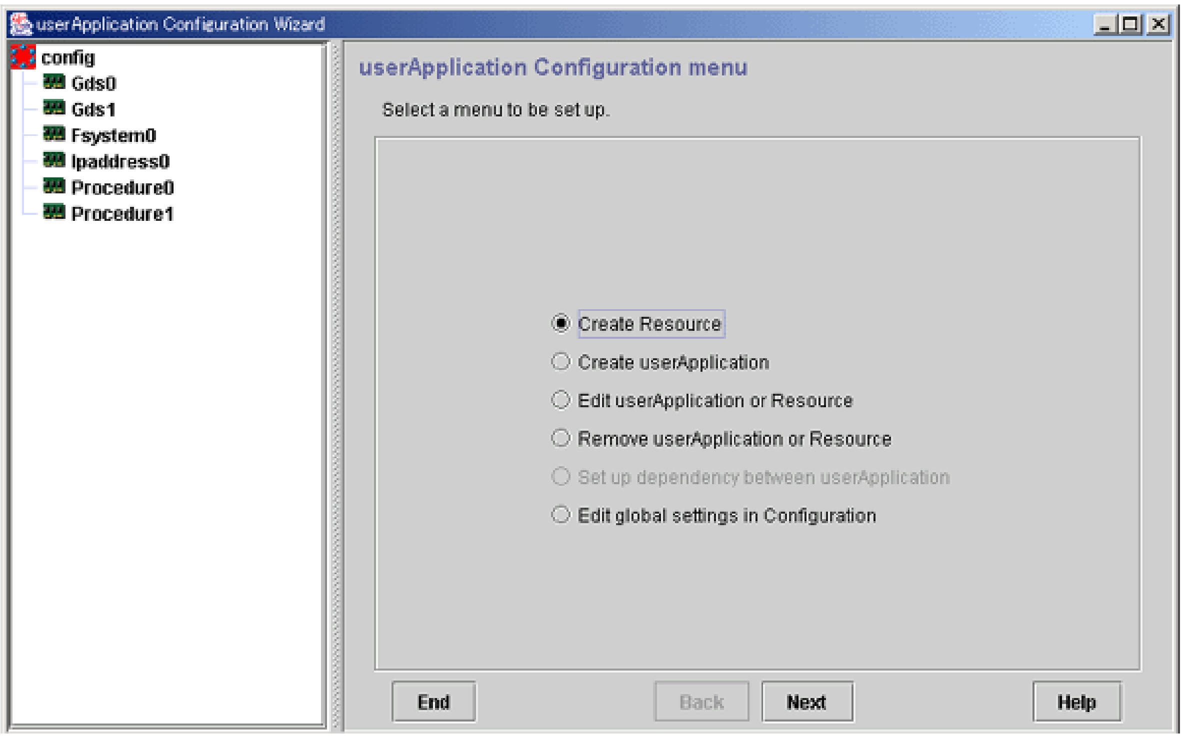
Therefore, to change the interface used in the resource, you must delete the cluster application once, and then delete only the target resource.

3. Delete the target resource once.

   From the tree on the left of the "userApplication Configuration Wizard" screen, select the userApplication to be changed, right-click the mouse to display the pop-up menu, and select *Remove userApplication or Resource*.

4. Re-create the resource.

   Select *Create Resource* from the "userApplication Configuration Wizard" menu, and then re-create the deleted resource with the step above.



   Note

   If changes are not required, refer to the "A.9 Cluster Application Worksheets" and set up the resource so that the setup information is the same as before.

   See

   For setup instructions, see "6.7.1 Setting Up Resources."

5. Re-create the cluster application.

   Select Create *userApplication* from the "userApplication Configuration Wizard" menu, and then re-create the deleted cluster application with the step above.

   Note

   If changes are not required, refer to the "A.9 Cluster Application Worksheets" and set up the cluster application so that the setup information is the same as before.

   See

   For details on how to create the cluster application, refer to "6.7.2 Creating Cluster Applications."

   If the dependency relationships between cluster applications have been set, you must set them again.

   For setup instructions, see "6.7.3 Setting Up Dependency Relationships Between Cluster Applications."

10.6.2 Changing the Attributes Used by a Resource or a Resource Interface

Change the attributes used by a resource or a resource interface according to the following procedure.

Operation Procedure:

Use the "userApplication Configuration Wizard" screen to change a resource used by the cluster application.

1. Stop RMS.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

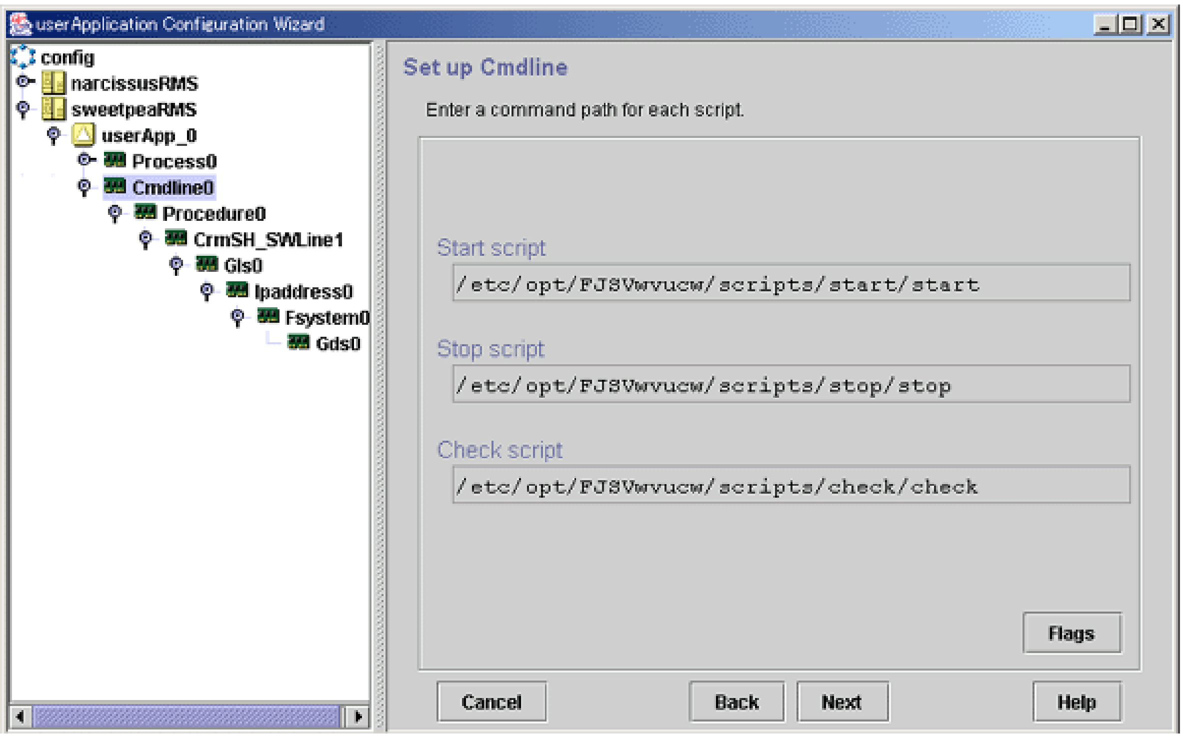
2. Select the resource.

   Return to the "Global Cluster Services" screen, and select *userApplication Configuration Wizard*.
   From the tree on the left of the "userApplication Configuration Wizard" screen, select the resource to be changed, right-click the mouse to display the pop-up menu, and select *Edit userApplication or Resource*.

Point

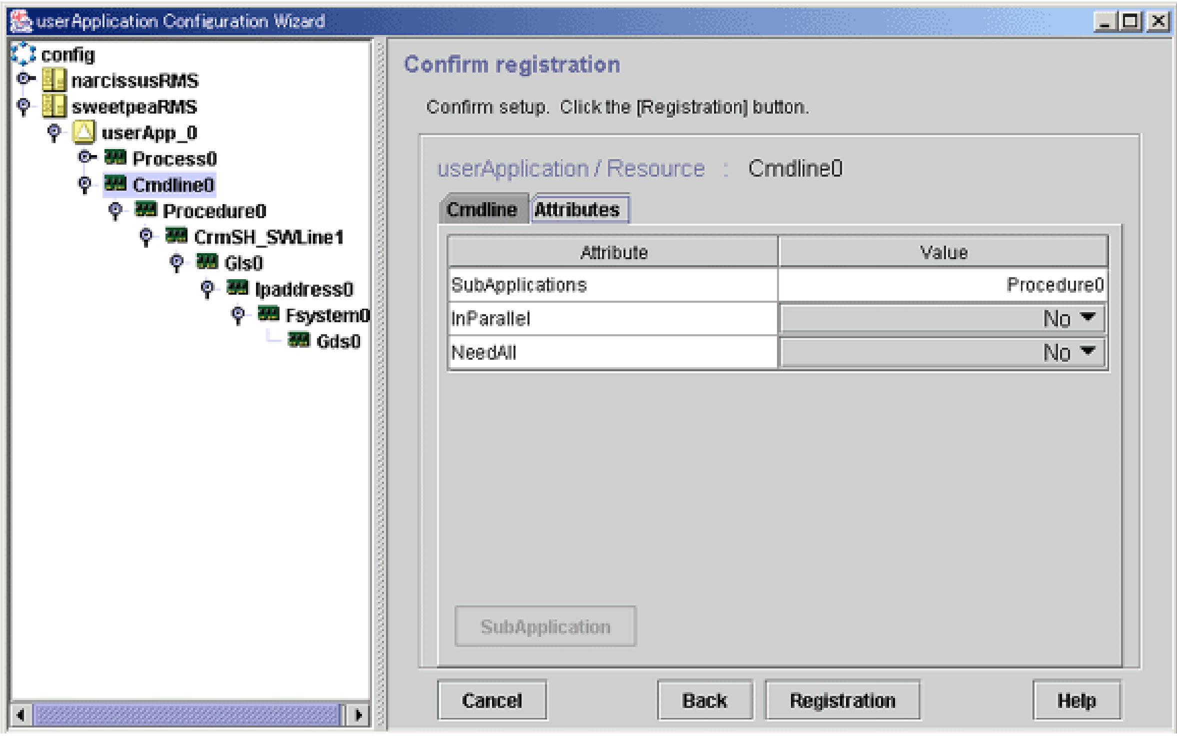
To change the attributes used by the resource or an interface of the resource, you can also select *Edit userApplication or Resource* from the top menu of the userApplication Configuration Wizard, and then select *Next*.

Cmdline



Flags button

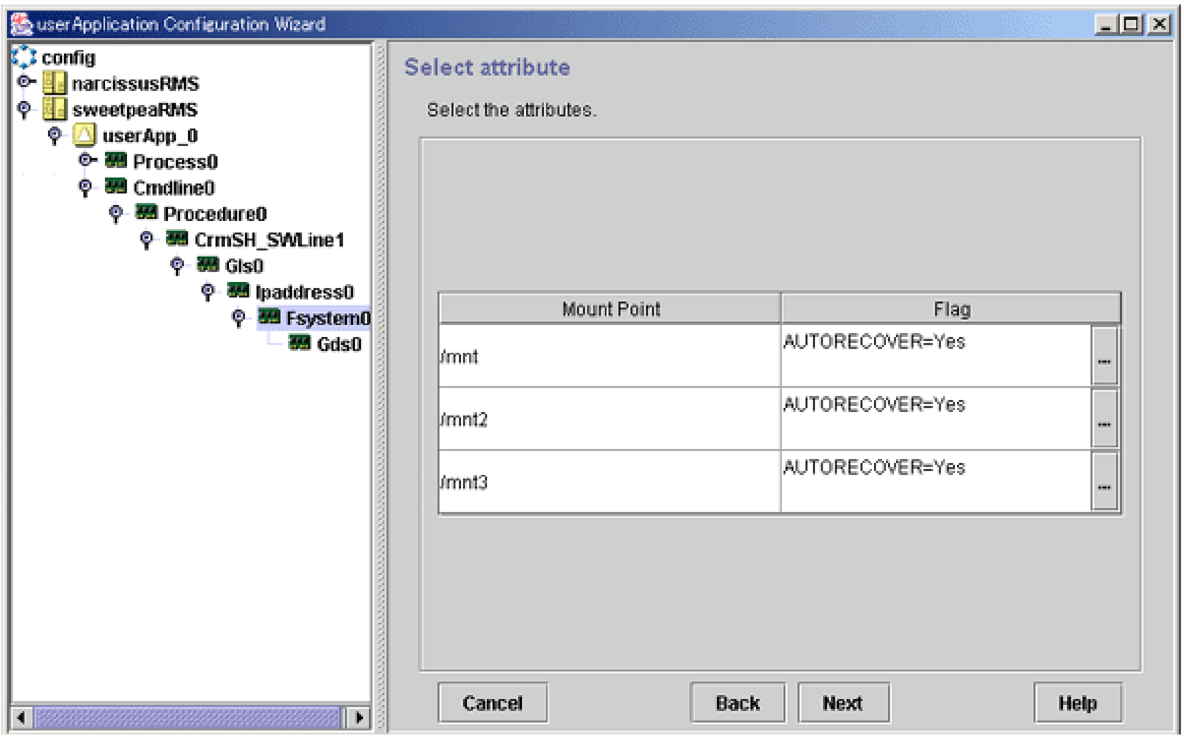
Click this button to change the script attributes, if necessary.
A dialog box for setting the attributes of the created script will be displayed. For details, see "Setting up Cmdline flags" in "6.7.1.1 Creating Cmdline Resources."



SubApplication button

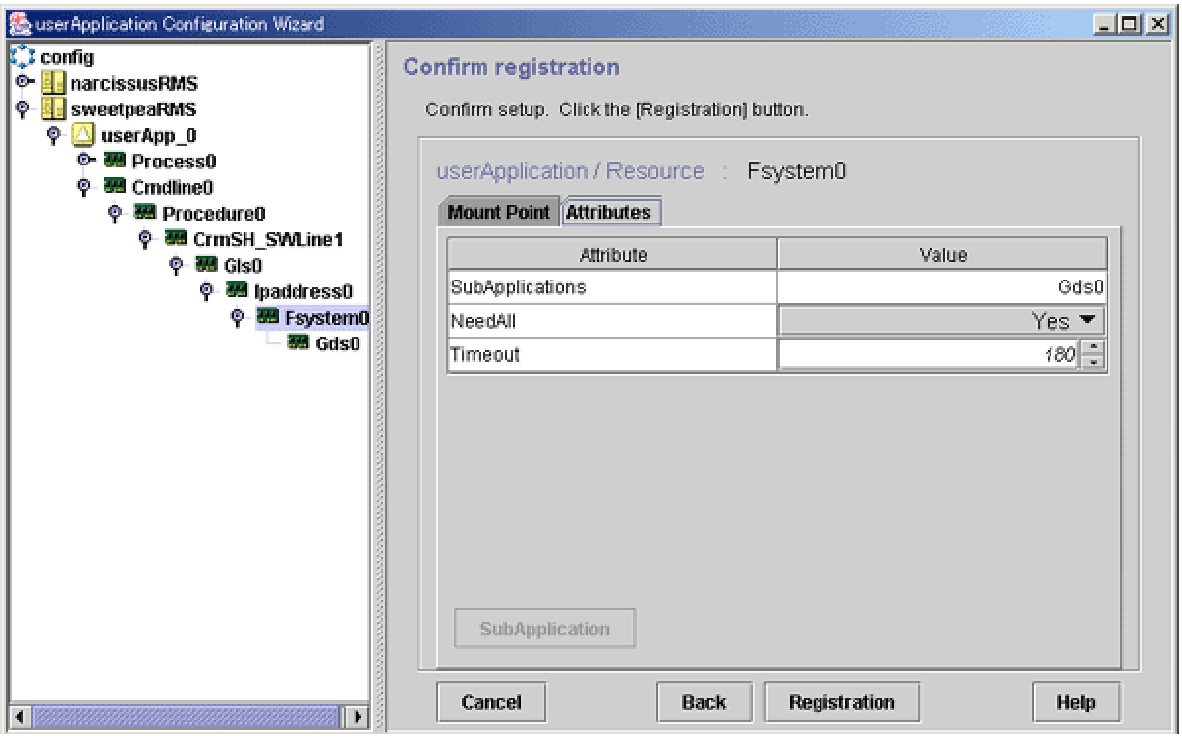
Click this button to associate a previously created Cmdline resource or processing monitoring resource found under the current Cmdline resource. This button cannot be selected if the resource belongs to a userApplication.

Fsystem



[...] button

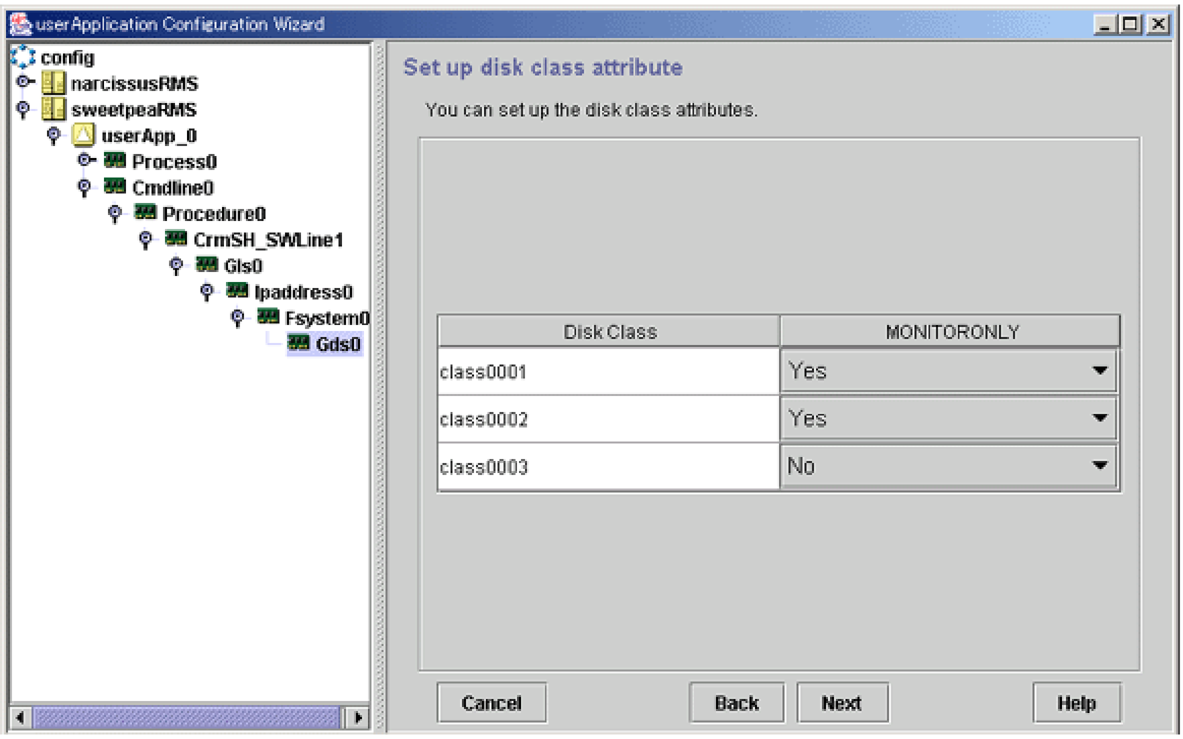
Press this button to change attributes for controlling mount point management. When this button is clicked, the Set Mount Point Attributes screen is displayed.



SubApplication button

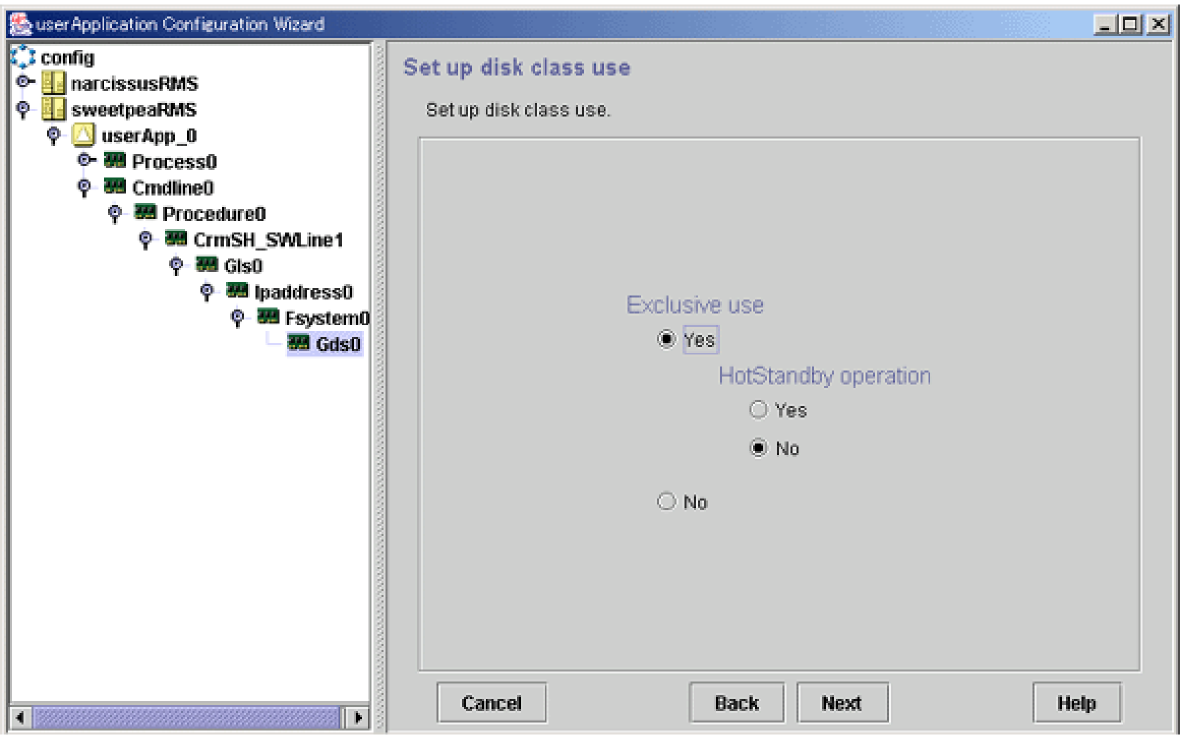
This button is for specifying another resource that depends on the current resource. However, in Fsystem, this button is disabled.

Gds



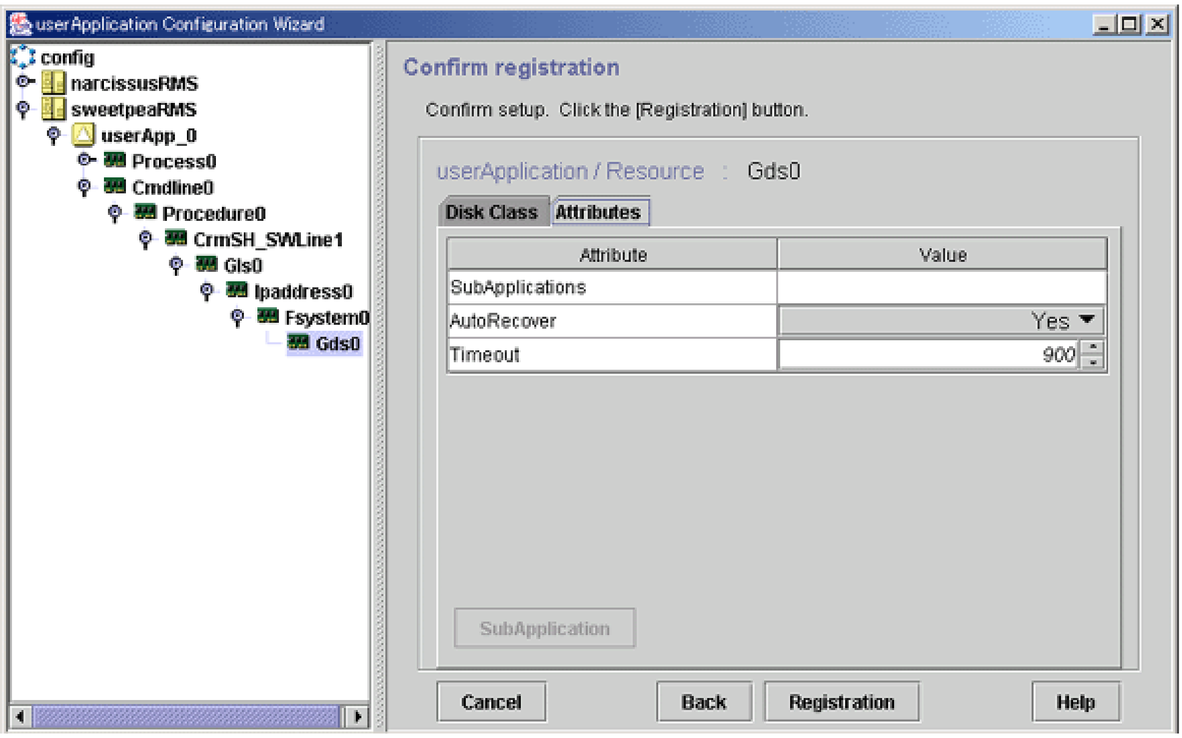
Flag	Outline
MONITORONLY	This flag determines whether disk class failures are to be reported to the userApplication. If "Yes" is set and a disk class failure occurs, the disk class

Flag	Outline
	is switched to faulted state but the Gds resources remain online, and userApplication failover does not occur.



Exclusive use

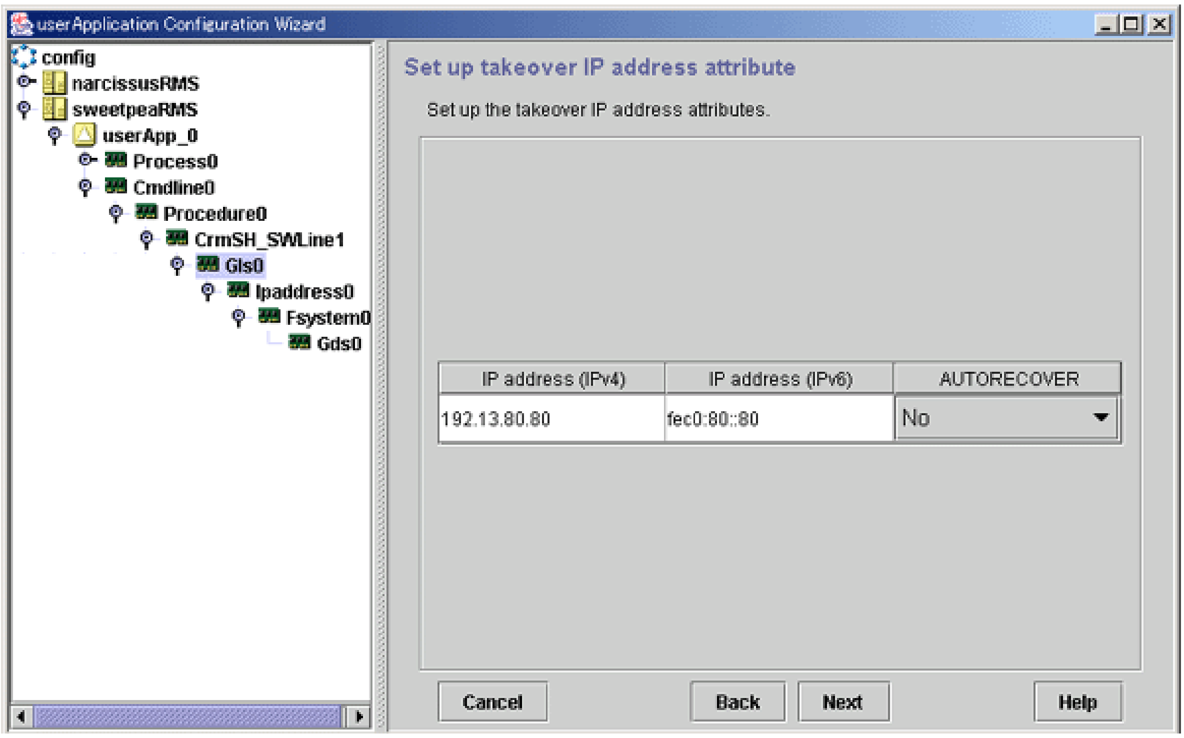
Select "*Yes*" or "*No*" for *Exclusive use*. If you select "*Yes*", also select "*Yes*" or "*No*" for *Hot-Standby operation*. For information on the shared disk, see "Shared disk uses" in "6.7.1.3 Creating Gds Resources." Note that if you change *Exclusive use* from "*No*" to "*Yes*," there is the risk of disk damage if another Gds resource is using the same disk class.



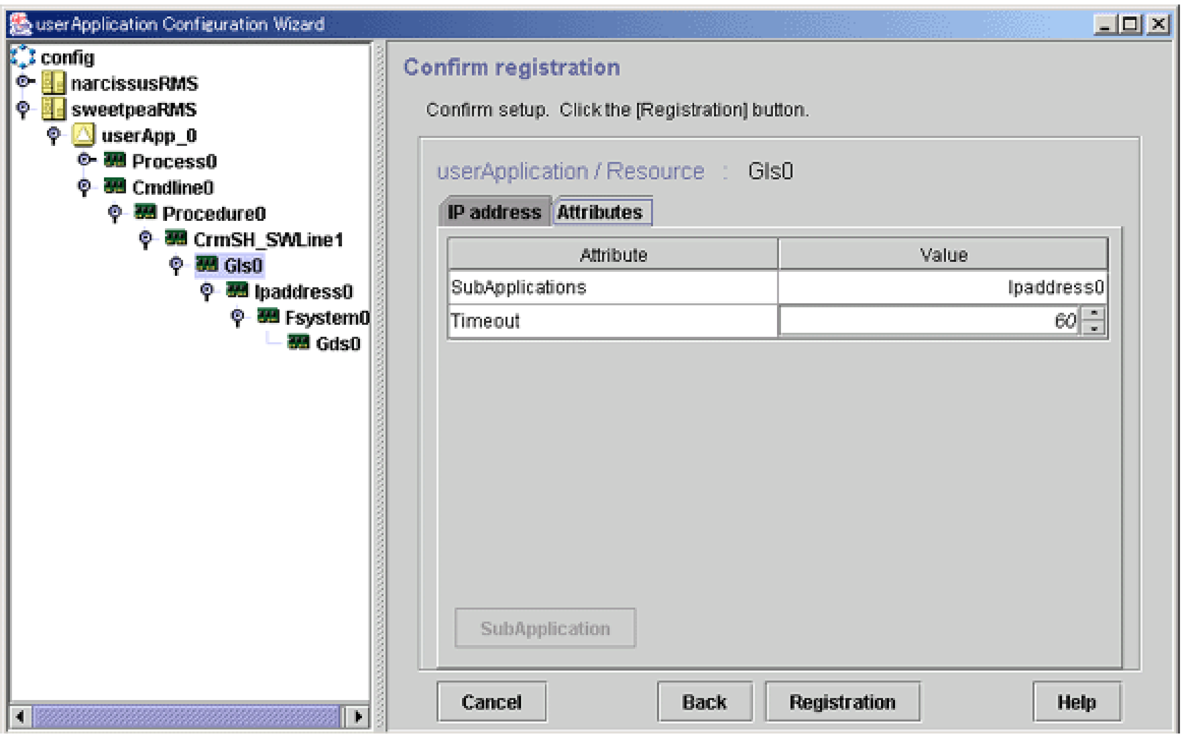
SubApplication button

Click this button to associate the other Gds resources with the Gds resource that has been created above. This button cannot be selected if the resource belongs to userApplication.

Gls



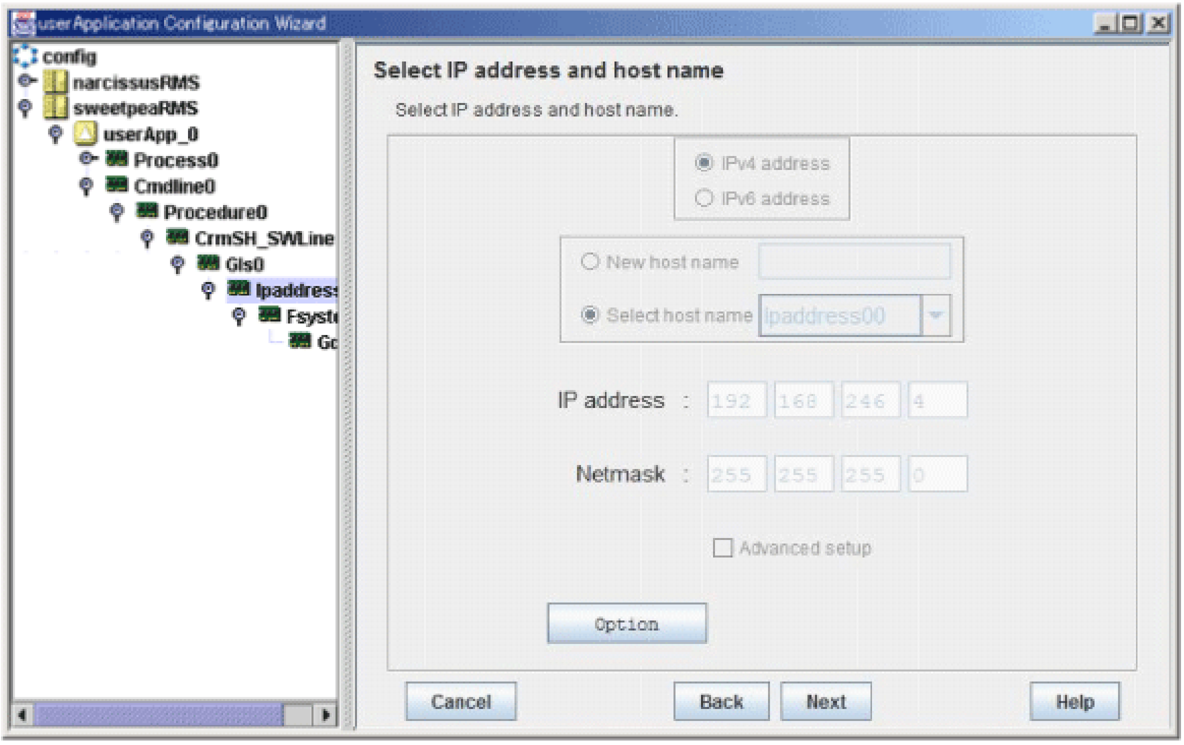
Flag	Outline
AUTORECOVER	If this is set to "Yes," RMS attempts to restore the faulted resource to prevent userApplication from being switched to other host. For GLS, this flag must be set to "No."



SubApplication button

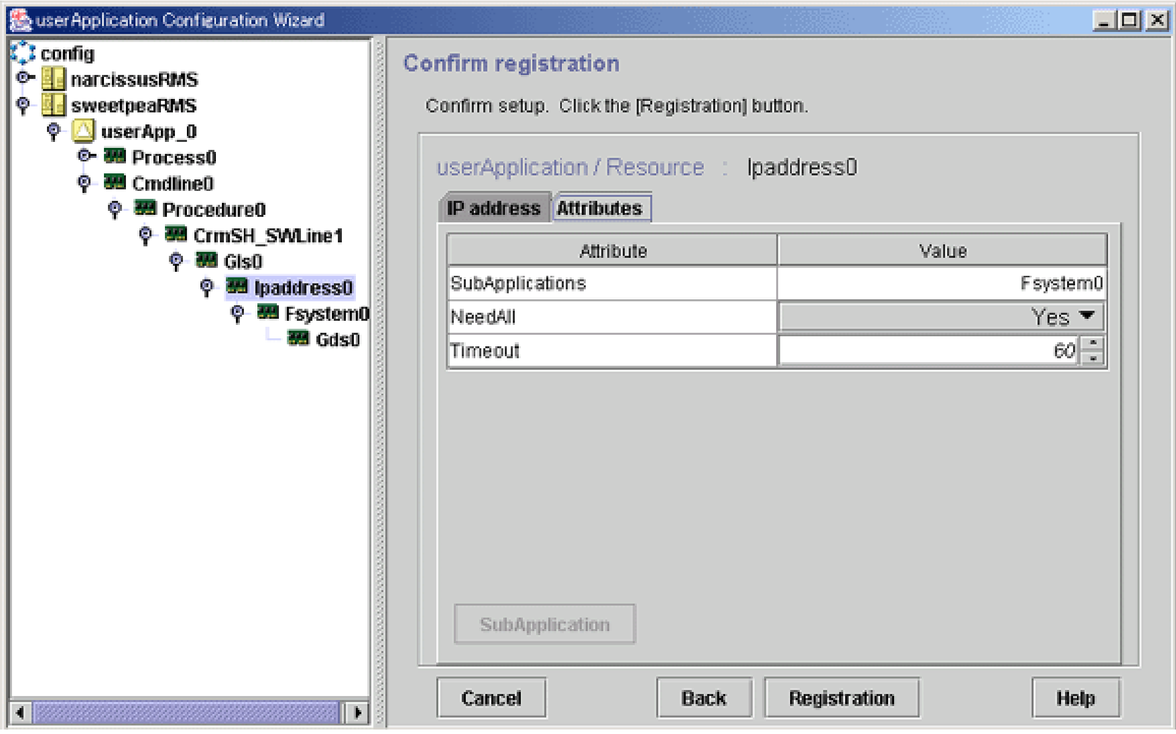
Click this button to associate the other Gls resources with the Gls resource that has been created above. This button cannot be selected if the resource belongs to userApplication.

Ipaddress (the following figure shows when using an IPv4 address for a takeover IP address)



Options button

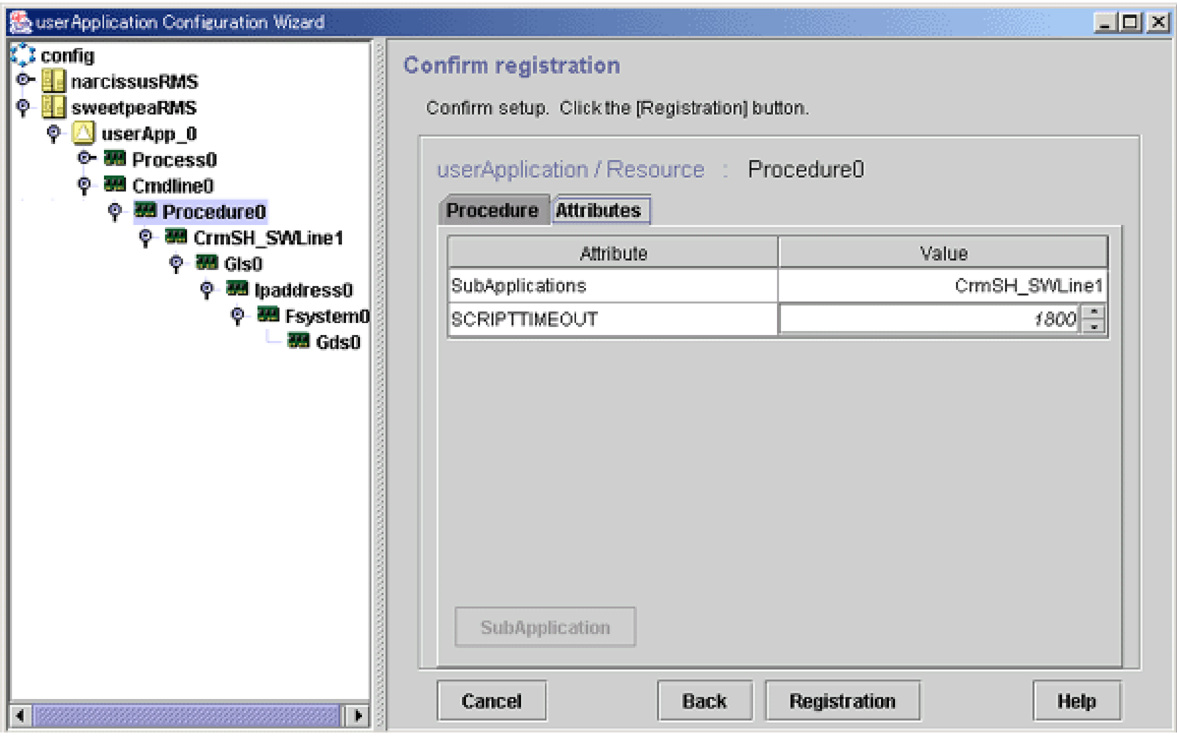
Use this button to change attributes of a takeover IP address. For operation instructions, see "Setting up Takeover IP Address Attributes."



SubApplication button

Click this button to associate the other Ipaddress resources with the Ipaddress resource that has been created above. This button cannot be selected if the resource belongs to userApplication.

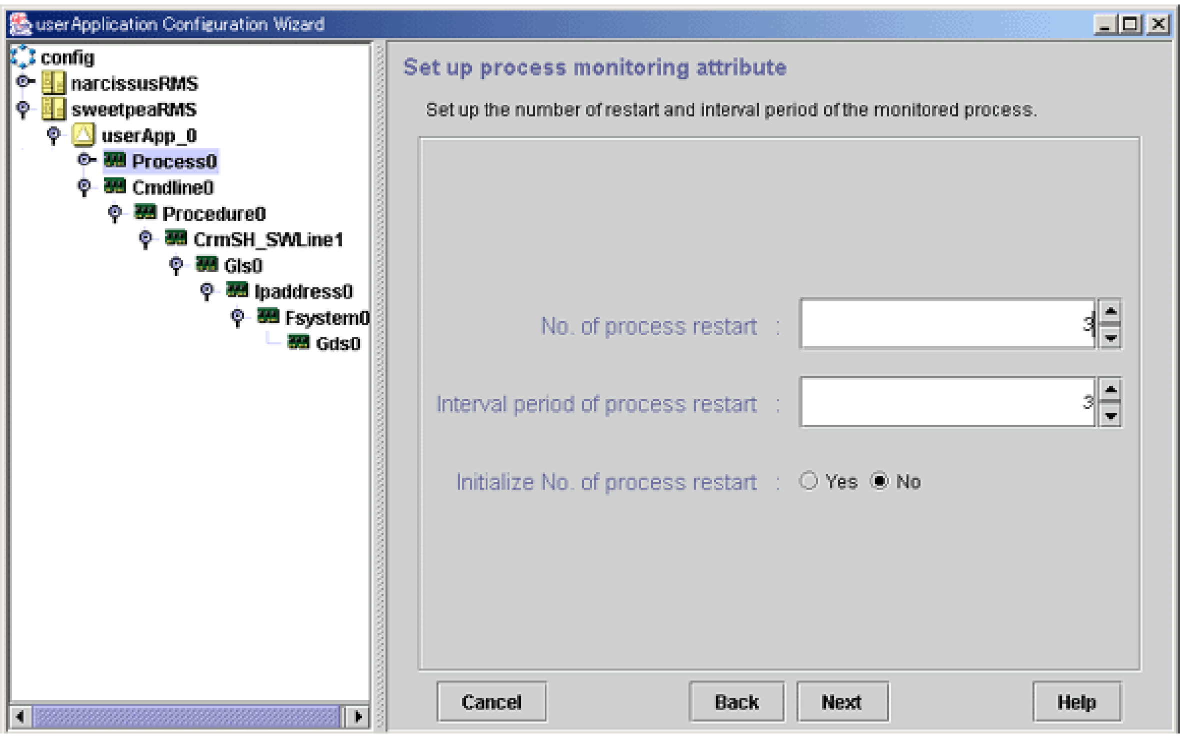
Procedure



SubApplication button

Click this button to associate the other Procedure resources with the Procedure resource that has been created above. This button cannot be selected if the resource belongs to userApplication.

Process



No. of process restart

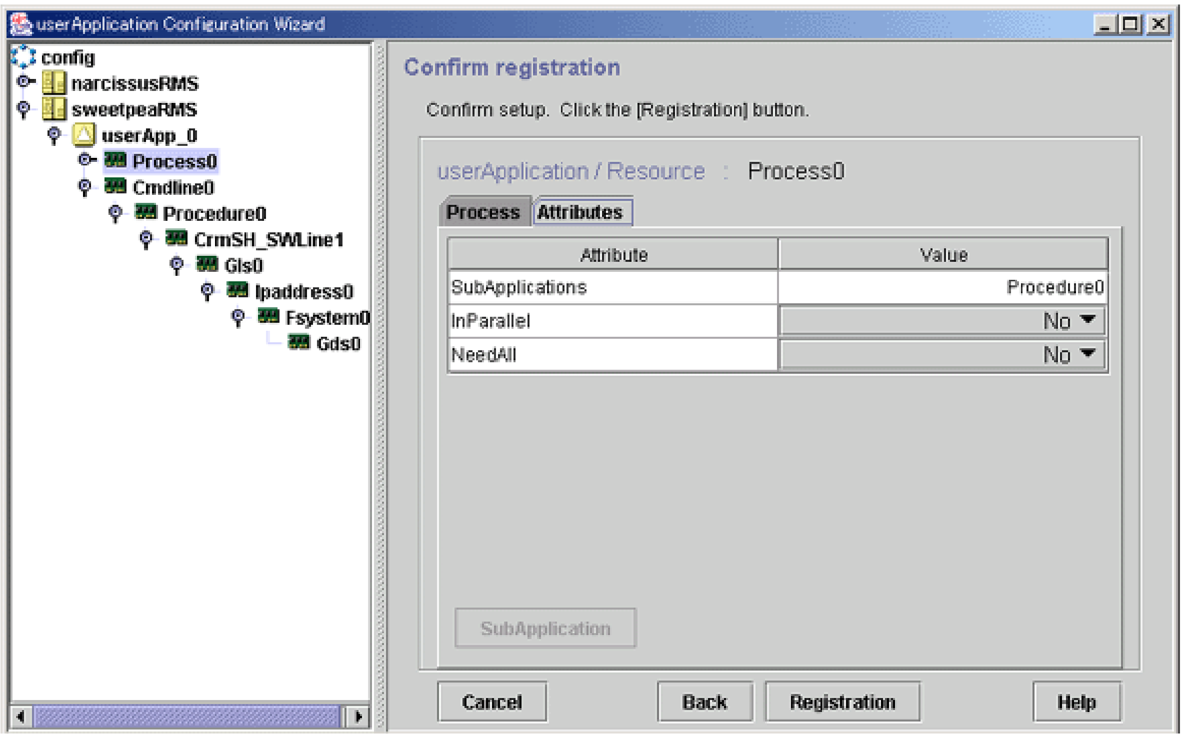
Specify the number of times the monitored process is to be restarted between 0 and 99 (default is 3). If 0 is specified and the monitored process stops, the process is not restarted and becomes Faulted.

Interval period of process restart

Specify the interval period when the process monitoring facility determines that the process has stopped until the facility executes restart. The specification range is between 0 and 3600 seconds (default is 3).

Initialize No. of process restart

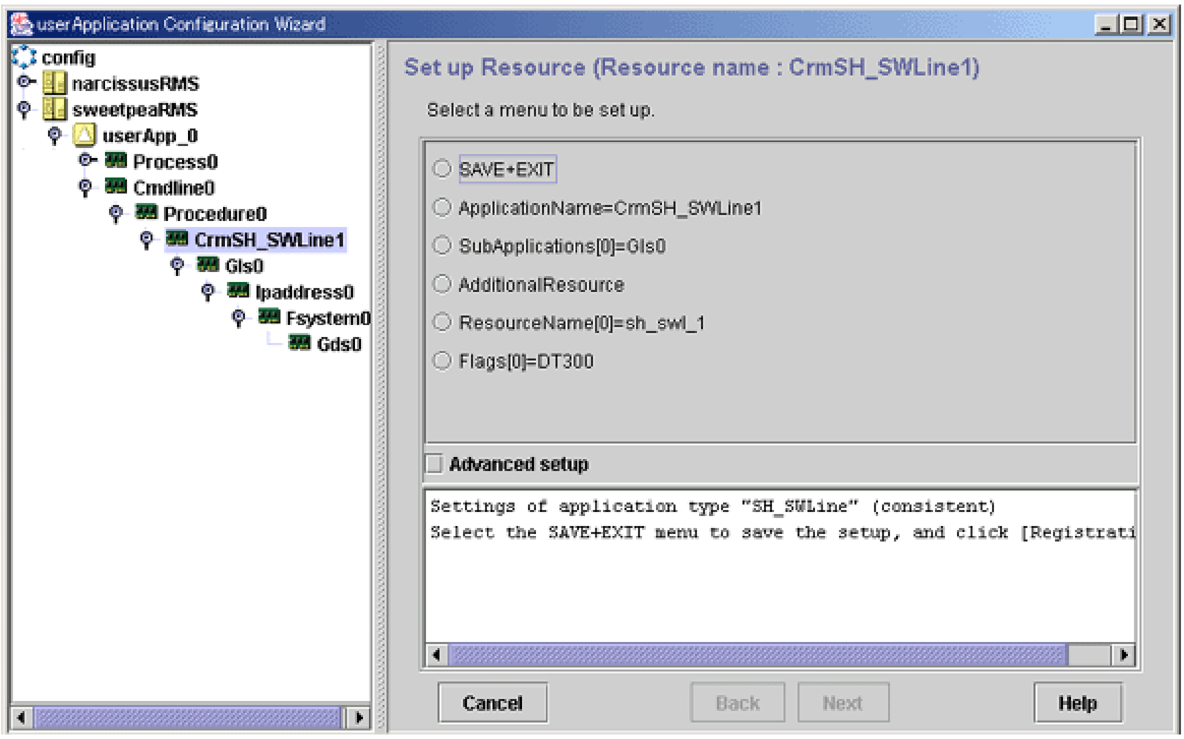
Specify whether or not the counter that has the specified *No. of process restart* value in the process monitoring facility as its maximum value is to be initialized periodically. If you select "*Yes*," the counter is initialized every "No. of process restart value x 60 seconds." If you select "*No*," the counter is not initialized periodically.



SubApplication button

Click this button to associate the Process resources or Cmdline resources with the Process resource that has been created above. This button cannot be selected if the resource belongs to userApplication.

SH_SWLine



Advanced setup

Check this box to display attributes that do not need to be changed, during resource creation. Use this checkbox to display items such as the setup menu for the resource timeout value.

Note

Specify "Application name" of the resource only if the resource is not associated with userApplication. If this resource is associated with userApplication, you must delete and then re-create the resource.

ISV

For information on changing an ISV resource, see the manuals for the product.

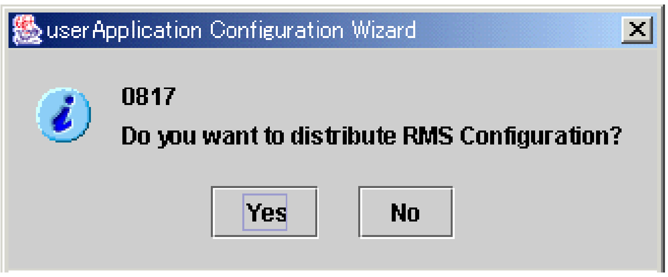
Note

Specify "Application name" of the resource only if the resource is not associated with userApplication. If this resource is associated with userApplication, you must delete and then re-create the resource.

RMS Configuration distribution messages

After you finish changing the operation attributes, click the *Registration* button on the screen to register the new information to RMS Configuration.

If the system judges that registration to RMS Configuration is completed and RMS Configuration can be distributed, it displays the following message screen:



If you have completed making changes, click *Yes*.

10.7 Changing the Timeout Value for the ZFS Detector

If an error occurs in the shared disk where the ZFS storage pool is placed or in the FC cable and the access to the ZFS storage pool cannot be recovered, the ZFS detector detects the error. The default timeout to detect the error is 300 seconds.

This timeout value can be changed by the RMS environment variable "HV_ZFS_CMD_TIMEOUT." However, if the value is too small, the ZFS detector may detect the error mistakenly caused by due to the temporary high load. In addition, if the value is too large, it may take time to detect the error. The value is set in seconds and can be set in the range of 30 - 600. If the value that is smaller than 30 or larger than 600 is set, 30 or 600 is assumed specified.

Note that this value is effective only for the ZFS detector and does not affect the behavior of other resource detectors.

An example is shown below.

Add the following definition to /opt/SMAW/SMAWRrms/bin/hvenv.local on all nodes constituting a cluster. If this file does not exist, create it.

```
export HV_ZFS_CMD_TIMEOUT=<Timeout value>
```

After the setting, restart RMS. The setting change becomes effective after restarting RMS.

Chapter 11 Changing the Operation Attributes of a Cluster System

This chapter describes how to change the operation attributes of a cluster application.

11.1 Changing the Operation Attributes of a Cluster Application

The procedure for changing the operation attributes of a cluster application is described below.

Operation Procedure:

1. Stop RMS.

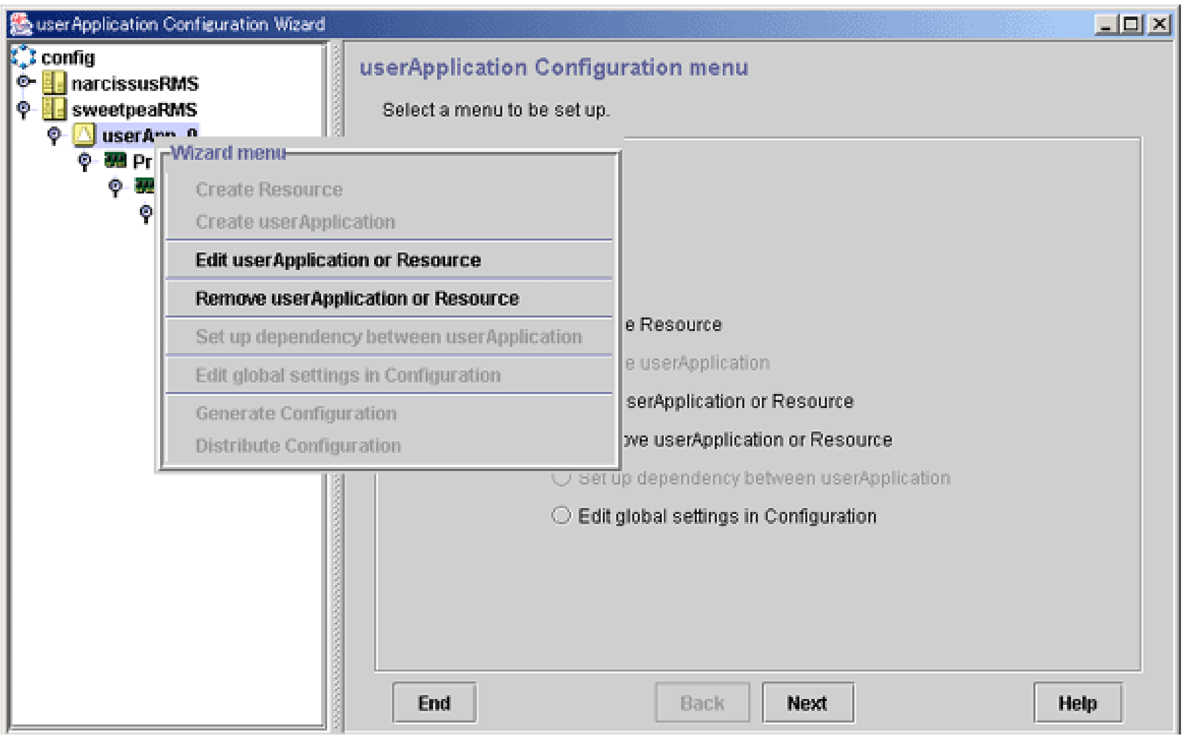
   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

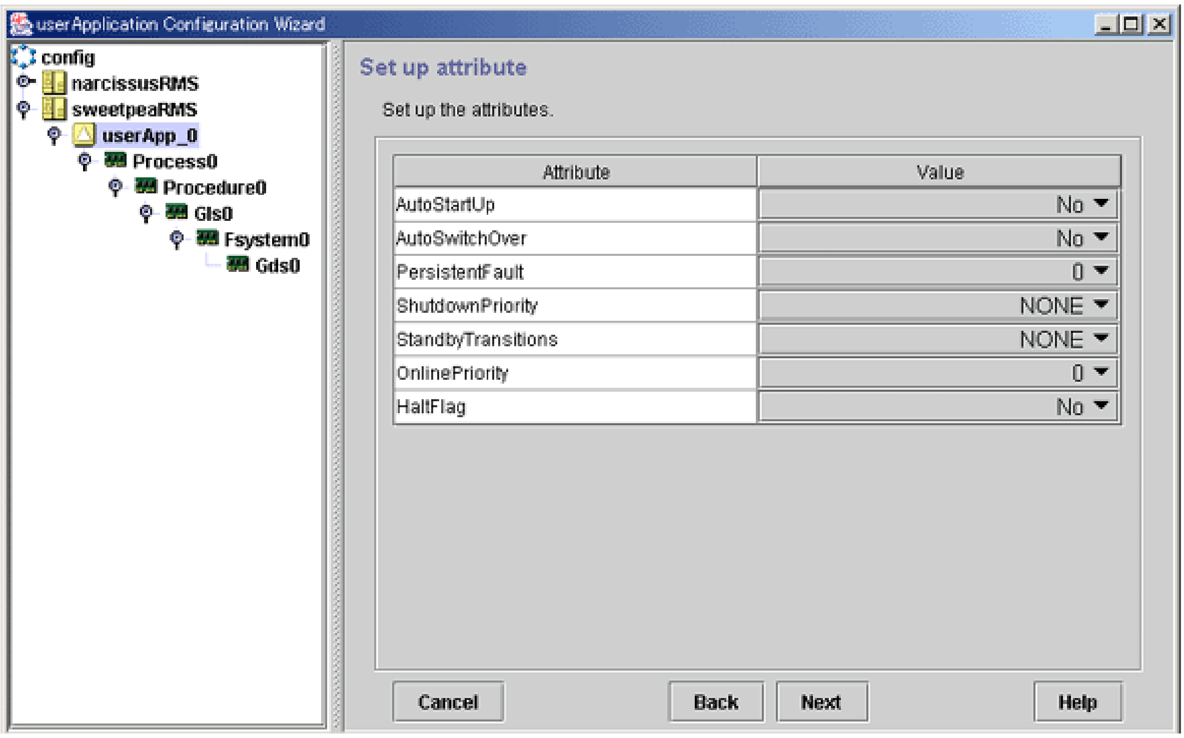
2. Change the operation attributes of the target userApplication.

   At the "Global Cluster Services" screen, select *userApplication Configuration Wizard*.
   From the tree on the left of the "userApplication Configuration Wizard" screen, select the userApplication to be changed, right-click the mouse to display the pop-up menu, and select *Edit userApplication or Resource*.

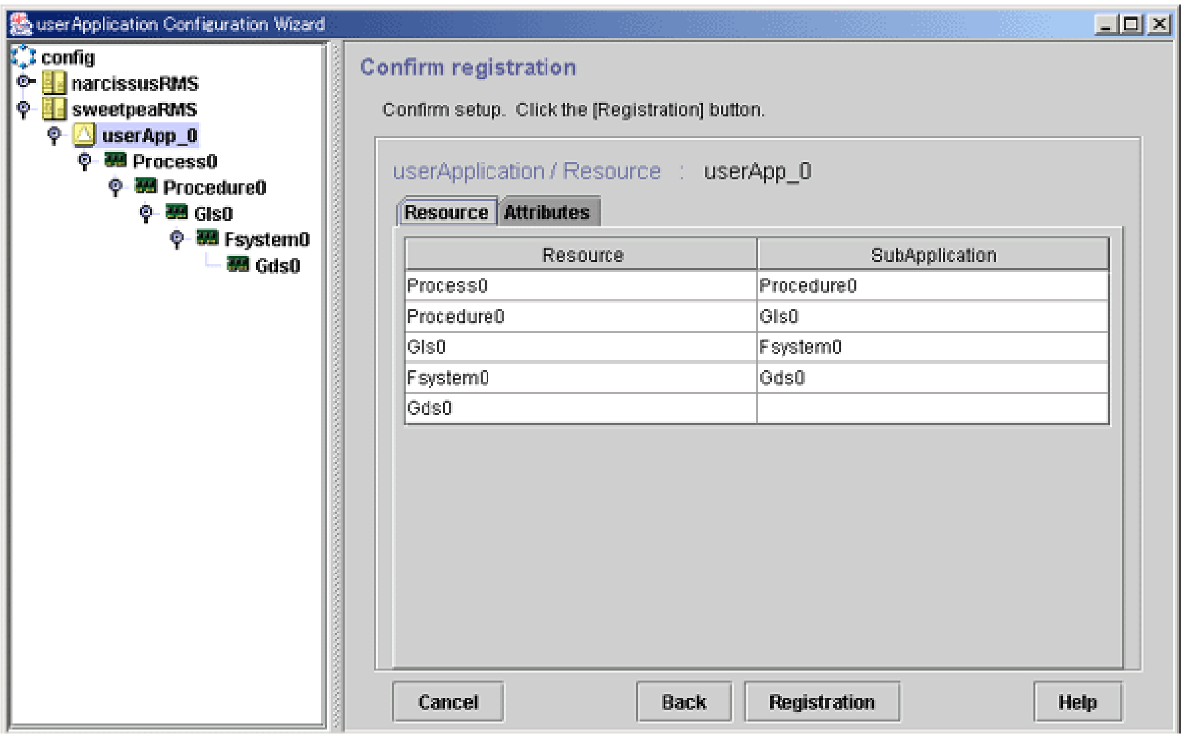
Point

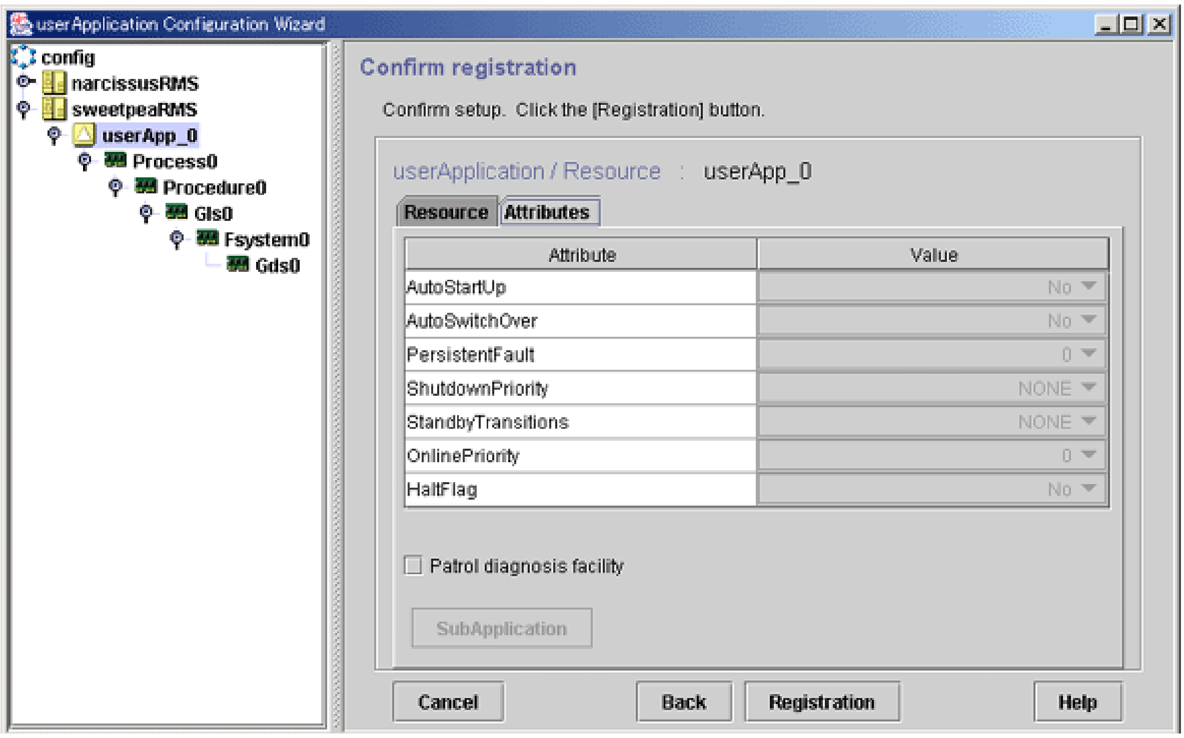
To change the cluster application, you can also select *Edit userApplication or Resource* at the top menu of the userApplication Configuration Wizard, and click *Next*.





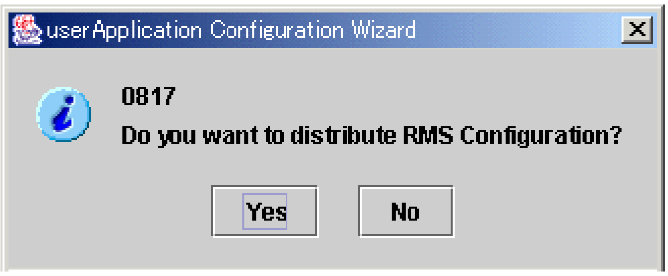
Change the operation attributes. For information on the individual operation attributes, see "6.7.5 Attributes."





After you finish changing the operation attributes, register the attributes. Click the *Registration* button to register the attributes to RMS Configuration.

After registration to RMS Configuration is completed, the following message screen is displayed if the system judges that distribution of RMS Configuration is possible.



If you are not changing the operation attributes or resource attributes of another cluster application, click *Yes*.

Information

The table below shows the operation attributes and environment variables of userApplication objects and the values that can be set for those attributes and variables.

Table 11.1 Operation attributes of userApplication objects

Operation	userApplication attribute/environment variable	Setting value (Bold: Default)	Description
Failover operation	AutoSwitchOver	**No**	Automatic failover is disabled.
		HostFailure	Failover occurs if a node fails.
		ResourceFailure	Failover occurs if a resource fails.
		Shutdown	Failover occurs if RMS stops.
Operation if resource inactivation fails	HaltFlag	**No**	The node is not forcibly stopped.

Operation	userApplication attribute/environment variable	Setting value (Bold: Default)	Description
		Yes	The node is forcibly stopped.
Operation for cluster application startup	OnlinePriority	**0**	The highest node of the PriorityList attribute becomes the active system.
		1	The node that was the active system last becomes the active system.
Operation of failover function for node startup	HV_RCSTART environment variable in hvenv.local file	**0**	The failover function is not started automatically.
		1	The failover function is started automatically.
Handling of failed nodes	PersistentFault	**0**	The failed node is not eliminated after the node is restarted.
		1	The failed node is eliminated even after the node is restarted.
Hot-standby operation	StandbyTransitions	**No**	Hot-standby is not adopted.
		StartUp	Hot-standby is adopted when the failover function is started.
		SwitchRequest	Hot-standby is adopted after the switchover operation.
		ClearFaultRequest	Hot-standby is adopted after the fault clear operation.

11.2 Changing the RMS Environment Variables

The PRIMECLUSTER operation attributes are set with the RMS environment variables.

The environment variables can be changed by editing /opt/SMAW/SMAWRrms/bin/hvenv.local of RMS to set or change the values.

See

- For information on hvenv.local, see "2.9 Environment Variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."
- For details on the RMS environment variables, see "13 Appendix - Environment Variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."
- Create the hvenv.local file as required.

11.2.1 Changing Timeout Period during RMS Stop Processing

Once the hvshut command is executed, RMS performs an Offline processing of a cluster application being started, and then performs RMS stop processing.

Therefore, set the total time of the following in second to an environment variable RELIANT_SHUT_MIN_WAIT specifying the time until the hvshut command times out:

1. The maximum required time to finish the Offline processing of a cluster application
2. The maximum required time to stop BM (base monitor) (30 seconds)

Note

If the value of RELIANT_SHUT_MIN_WAIT is too small, the hvshut may time out often before finishing the Offline processing of a cluster application. Tune RELIANT_SHUT_MIN_WAIT carefully.

See

For details on RELIANT_SHUT_MIN_WAIT, see "RELIANT_SHUT_MIN_WAIT" of "13.2 Global environment variables" in the "PRIMECLUSTER RMS Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

For how to refer to or change the RMS environment variable, see "6.3.4 Displaying environment variables" or "13.1 Setting environment variables" in the "PRIMECLUSTER RMS Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

11.3 Changing Time to Detect Heartbeat Timeout

11.3.1 Changing Time to Detect CF Heartbeat Timeout

If CF cannot perform a heartbeat for 10 minutes, it is determined that a heartbeat fails. In the environment where the system is overloaded and a heartbeat failure occurs often, tune the heartbeat time.

Note

If you set the heartbeat time long, it takes long to detect an error. Therefore, tune the heartbeat time carefully.

To tune the heartbeat time (10 seconds), perform the following procedure:

1. Add the following to the end of the "/etc/default/cluster.config" file on all nodes configuring a cluster system. To restore the older file version, take a note of the contents before changing it.

   CLUSTER_TIMEOUT "second"
   Example: Changing it to 30 seconds
   CLUSTER_TIMEOUT "30"

2. To enable the setting value, you need to execute cfset -r at the same time on all nodes configuring a cluster system.

```
# cfset -r
```

3. Execute cfset -a to confirm the setting value.

```
# cfset -a
From cfset configuration in CF module:

KEY: CFCP       VALUE: cfcp
KEY: CFSH       VALUE: cfsh
KEY: CLUSTER_TIMEOUT    VALUE: 30
```

11.3.2 Changing Time to Detect RMS Heartbeat Timeout

If RMS cannot perform a heartbeat for 600 seconds, it is determined that a heartbeat fails. In the environment where the system is overloaded and a heartbeat failure occurs often, tune the heartbeat time.

Note

- If you set the heartbeat time long, it takes long to detect an error. Therefore, tune the heartbeat time carefully.

- If you set the heartbeat time shorter than CF heartbeat time, a warning message is output during RMS startup. For details, see the notes on "7.6 CF and RMS Heartbeats."

To tune the heartbeat time, perform the following procedure:

Default value: 600 seconds

Setting value: set it from 45 to 3600 seconds

1. Stop a cluster application and RMS on all the nodes.

```
# hvshut -a
```

2. Change /usr/opt/reliant/etc/CONFIG.rms on all the nodes as follows.

```
hvcm -c config -h waiting time (seconds)
```

Example

To change the default value from 600 to 800 seconds

-h monitoring timeout (Maximum: 3600)

```
hvcm -c config -h 800
```

3. Start RMS on all the nodes.

```
# hvcm -a
```

4. Check if RMS has started with the option specified in Step 2.

```
# hvdisp -h
```

5. Check if hvcm -c config -h waiting time (seconds) (Check that the waiting time is the value set above).

Part 5 Maintenance

This part explains the procedure for maintaining the PRIMECLUSTER system as well as the procedure for backing up and restoring the PRIMECLUSTER operation environment.

Chapter 12 Maintenance of the PRIMECLUSTER System

This chapter explains items and procedures related to maintenance of the PRIMECLUSTER system.

12.1 Maintenance Types

The maintenance of the PRIMECLUSTER system is divided as described below, depending on whether maintenance is performed while the job is stopped:

Stop maintenance

Maintenance that is performed while the entire cluster system is stopped.

Hot maintenance

Maintenance that is performed while the job continues to operate without state transition of the cluster application (failover or degeneration).
Hot maintenance is executed mainly after disk replacement of the disk array unit.
Also, the PCI Hot Plug function of SPARC Enterprise M series can be used for hot maintenance.

Job hot maintenance

Maintenance that is performed while the maintenance target node is detached from the cluster by state transition of the cluster application (failover or degeneration), while the job is allowed to continue operating.

Of these, the type to be performed depends on the location and contents of the failure. Determine the maintenance that is to be performed after consulting with field engineers.

12.2 Maintenance Flow

Field engineers shall perform parts replacement. The flow of maintenance is as follows:

Identifying the location of a failure

After an error is detected, identify the location of that error by using the CRM main window or RMS main window. For details on how to identify the location of the error, see "7.4.1.1 Failure Detection and Cause Identification if a Failure Occurs."

When stop maintenance is to be performed

1. The running PRIMECLUSTER system shall be stopped by administrator of the PRIMECLUSTER system.
2. Pass the operation over to field engineers.
3. Field engineers shall then perform maintenance of the erroneous location (repair or replacement). Confirm that the system operates normally by running a test program, etc.
4. After the completion of maintenance by field engineers, check the relevant equipment and then boot the PRIMECLUSTER system.

When job hot maintenance is to be performed

1. The administrator of the PRIMECLUSTER system shall shut down the node that contains the target equipment, so as to separate it from the operation, and then pass the operation over to field engineers.

   For details on how to separate the node from the operation, see "12.2.1 Detaching Resources from Operation."

2. Field engineers shall confirm the target equipment and shall perform maintenance of the erroneous equipment (repair or replacement). Operation shall be confirmed by using a test program, etc.
3. After field engineers complete the maintenance and confirm the operation of the relevant equipment, boot the node and then execute standby restoration for the operation.

   For details on standby restoration for the operation, see "12.2.2 Executing Standby Restoration for an Operating Job."

12.2.1 Detaching Resources from Operation

Execute the following for the node that you are going to shut down.

Cluster application failover

If the relevant node is operating, you must first execute failover operation with the hvswitch(1M) command.

See

For details on how to determine whether the relevant node is operating, see "7.1.3.1 RMS Tree."

Stopping RMS

After confirming that the relevant node is in either the Offline or Standby state, stop RMS running on the relevant node by executing the hvshut(1M) command.

See

For details on how to stop RMS, see "8.1.3 Stopping RMS" in "*PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide*."

Stopping a node

Execute the shutdown(1M) command to stop the relevant node.

12.2.2 Executing Standby Restoration for an Operating Job

Perform standby restoration for an operating job, as described below.

Procedure

1. Power on the relevant node.
2. Perform standby restoration for the relevant node (if necessary, subsequently execute failback).

See

For details on how to execute cluster application standby restoration, see "7.2.2.1 Starting a Cluster Application." For details on how to execute failback, see "7.2.2.3 Switching a Cluster Application."

12.3 Software Maintenance

This section provides notes on batch correction, emergency repair, and the application of patches to the PRIMECLUSTER system.

12.3.1 Notes on Applying Corrections to the PRIMECLUSTER System

Note the following when you apply batch corrections to the cluster system.

- Back up the system environment before you attempt to apply a correction.
- The software version to be installed on each node must be the same on all nodes in the cluster system. Also, the corrections must be the same on all the nodes constituting the system. Note, however, that this is not always true when rolling update, described below, is allowed.
- To apply a batch correction, you must stop the node temporarily. This means that the job must be stopped, albeit temporarily. You should consider a maintenance plan to ensure that the maintenance is completed within a specified period. You must also examine the time and duration of the maintenance to minimize the impact on a job.
- **Rolling update** is a method by which software is updated while the job continues to operate by executing job failover for a node in a cluster to separate the standby node from the operation in order to apply corrections to the node one by one.
  If you apply this method, the job stop time required for software update can be minimized. To perform update with this method,

however, you must satisfy the prerequisites for rolling update (the items to be corrected must be correctable with rolling update). To apply this method, you must confirm the contents of the README file for the relevant patch and then contact field engineers.

- To apply SRU, note the following points:
  - After the application of SRU, do not change the PRIMECLUSTER settings and configuration until the operating system is restarted.
  - If SRU was applied at RMS startup, the cluster application may become Faulted. In this case, follow the procedure in "7.2.2.4 Bringing Faulted Cluster Application to Available State" to clear the Faulted cluster application to Online.

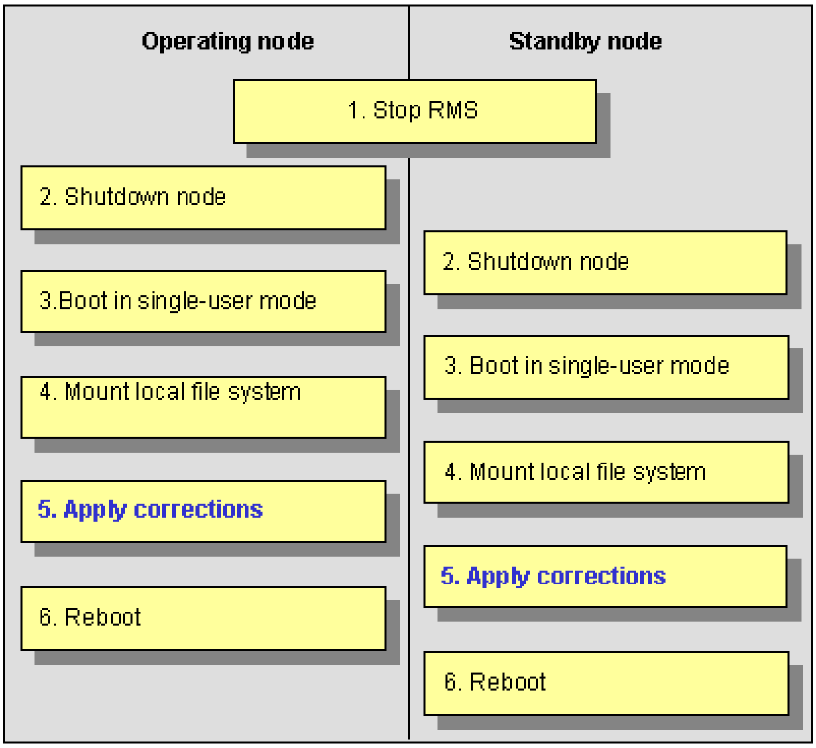
12.3.2 Overview of the Correction Application Procedure

This section provides an overview of the procedure for applying corrections such as a batch correction to a cluster system. The procedure explained here is a general procedure.

12.3.2.1 Procedure for Applying Corrections by Stopping an Entire System

This section explains the procedure for applying corrections by stopping the entire cluster system. An example of a two-node 1:1 standby configuration is used here.

Flow of operation



Procedure

Copy the correction to be applied to each node to the local file system in advance.

1. Stop RMS.

   Execute **hvshut -a** on either cluster node to stop the operation of RMS.

2. Shut down all nodes.

   Shut down all nodes.

3. Boot in single-user mode.

   Boot all the nodes that were shut down in single-user mode.

4. Mount the local file system.

   Execute **mountall -1** and **zfs mount -a** on all the nodes to mount the local file system.

5. Apply corrections.

   Apply the corrections that were copied to the local file system in advance.

6. Reboot.

   After applying the corrections, boot the nodes by using **shutdown -i6**.

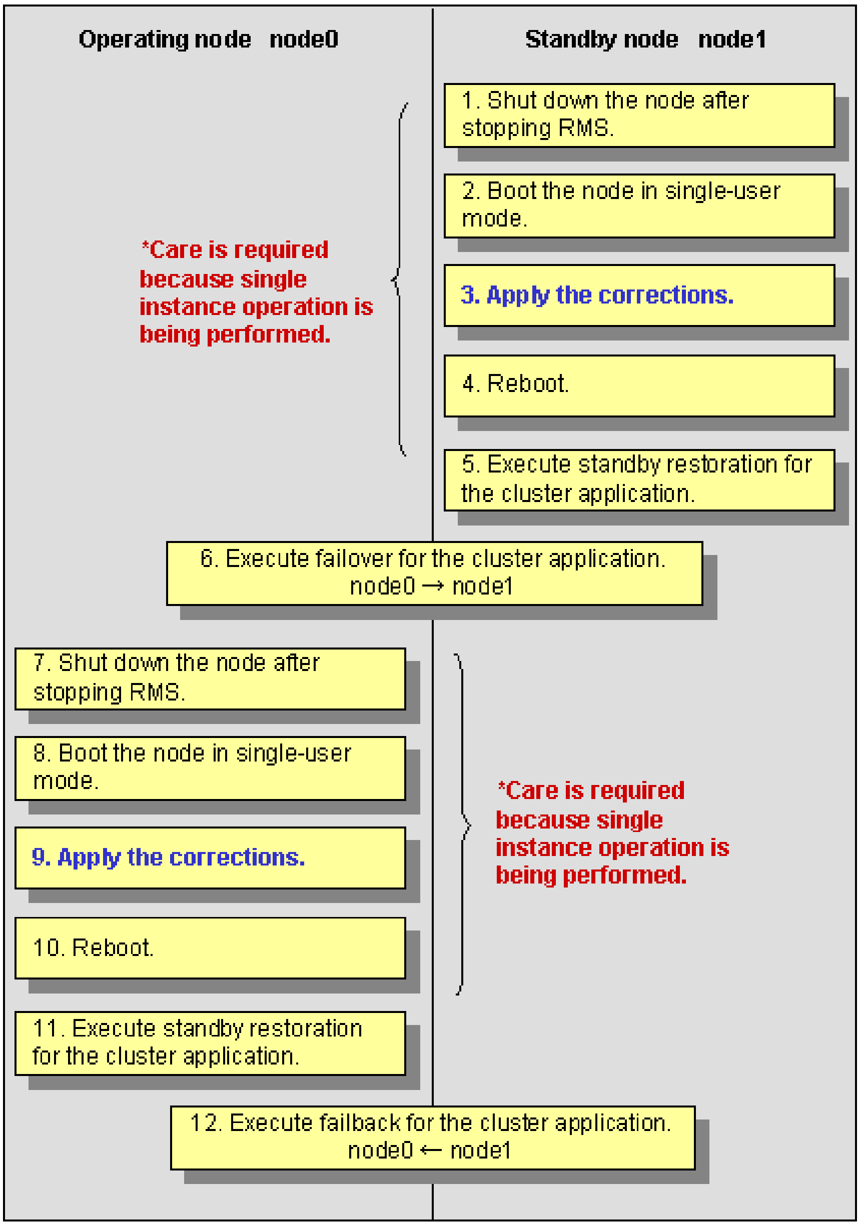
Note

- For details on the corrections, refer to the manuals provided with the corrections.
- For details on the standby restoration of cluster applications, see "7.2.2.1 Starting a Cluster Application." For details on failback, see "7.2.2.3 Switching a Cluster Application."

12.3.2.2 Procedure for Applying Correction by Rolling Update

This section explains the procedure for applying corrections by rolling update. An example of two-node 1:1 standby configuration is used for this explanation.

Flow of operation



Procedure

1. Shut down the standby node (node1).

   To apply corrections to the standby node (node1), shut down the node after stopping RMS.

   Note that, as a result of this shutdown, a cutoff state transition occurs and dual instance operation is disabled until standby restoration is performed.

2. Boot the standby node (node1) in single-user mode.

   Boot in single-user mode.

3. Apply corrections.

   Apply the necessary corrections.

4. Reboot the standby node (node1).

   Reboot the node by using shutdown -i6.

5. Execute standby restoration for the standby node (node1).

   Since the shutdown node (node1) to which corrections have been applied is cut off from the cluster system, execute standby restoration for the node.

6. Execute failover of the cluster application.

   To apply corrections to the operating node (node0), execute the **hvswitch**(1M) command and switch all the cluster applications to the standby node (node1).

   When you switch the "scalable" userApplication, switch all the "standby" userApplications under the "scalable" userApplication first, then switch the "scalable" userApplication.

7. Shut down the operating node (node0).

   After the completion of failover, stop RMS, and then shut down the node.

8. Boot the operating node (node0) in single-user mode.

   Boot the node in single-user mode.

9. Apply corrections.

   Apply the necessary corrections.

10. Reboot the operating node (node0).

    Reboot the node to apply the corrections to the system by using shutdown -i6.

11. Execute standby restoration for the operating node.

    Since the shut-down node (node0) to which corrections were applied is cut off from the cluster system, execute standby restoration for the node.

12. Execute failback.

    Restore the state of the standby layout defined at installation by executing failback operation, as required.

Note

- For details on the corrections, refer to the manuals provided with the corrections.
- For details on standby restoration of cluster applications, see "7.2.2.1 Starting a Cluster Application." For details on failback, see "7.2.2.3 Switching a Cluster Application."

Chapter 13 Backing Up and Restoring a PRIMECLUSTER System

The PRIMECLUSTER system supports the collective backup and restoration of a variety of environmental files that constitute PRIMECLUSTER, for the purpose of recovering from problems caused by operating errors during upgrading of, or configuration changes to, the system.
The following properties of the PRIMECLUSTER system can be backed up and restored with this function.

- CF configuration
- SF (Shutdown Facility) configuration
- RMS configuration
- GUI (including Web-Based Admin View) configuration
- GLS (redundant line control function) configuration

Note

- The hardware configuration must not change between back up and the restoring .
- Among the RMS configuration files, special commands and utilities that can be invoked by the user from the configuration are backed up.
- The PRIMECLUSTER backup/restore function cannot backup or restore a GLS (multipath function), GFS, or GDS configuration. For backing up and restoring GLS (multipath function), GFS, and GDS configurations, please refer to the corresponding manuals.
- For details on the backup and restoration of PRIMECLUSTER-related products, refer to the corresponding manual.
- If you collect system backup on the environment where 1 is set to OnlinePriority attribute in cluster application, the automatic start of application may be restrained right after restoring. In this case, use the hvswitch command to change userApplication to Online manually.

Prerequisites for backup and restoration

The following conditions must be satisfied to use the backup and restoration functions of PRIMECLUSTER:

a. PRIMECLUSTER 4.1A10 or later.
   The backup/restore function including GLS (redundant line control function) will be available on PRIMECLUSTER 4.1A20 or later.

b. The version to be used for restoration must be of the same PRIMECLUSTER version as that was used for backup, or later.

c. Execute the backup procedure on all the nodes in the PRIMECLUSTER operation environment. It is preferable to execute this procedure simultaneously. You can back up in either multi-user or single user mode.

   Note

   If backup is not executed for all nodes simultaneously, the machines may contain backup data with a different configuration. If the configuration of the backup data differs between machines, the consistency of the cluster configuration cannot be maintained at restoration.

d. Execute the restoration procedure in single user mode.

   Note

   - If the OS is newly installed, and the PRIMECLUSTER environment is restored, the normal operations of a cluster system cannot be guaranteed.

- If restoration is not done on all nodes but only on some nodes, the consistency of the cluster configuration might not be maintained between nodes, and failures might occur.

e. When restoration is to be performed, the hardware configuration must be the same as that used for backup.

f. With the restoration function, Solaris standard system files cannot be restored. The backup function also cannot be used to back up all the system files in a PRIMECLUSTER system. You need to manually back up or restore the system files, if necessary.
For details on the method of restoring system files that have been backed up with the backup function, refer to Procedure 3) in the explanation of the restoration procedure.

Information

You should back up a cluster configuration each time it is modified.

13.1 Backing Up the PRIMECLUSTER Operation Environment

Execute the following command to back up the PRIMECLUSTER operation environment:

```
/opt/SMAW/bin/cfbackup [ -f ] [ n ]
```

See

For details on the command parameters, refer to the description of the "cfbackup(1M)" command.

Example

Back up the PRIMECLUSTER environment.

```
# cfbackup
08/07/02 11:17:55 cfbackup 1 started
...
08/07/02 11:21:23 cfbackup 1 ended
#
```

Example

Confirm the backup files.

```
# ls -l /var/spool/SMAW/SMAWccbr
-rw-rw-rw-   1 root     other      2860 Aug  7 11:21 fuji2_ccbr1.tar.Z
```

13.2 Restoring the PRIMECLUSTER Operation Environment

Take the following steps to restore the PRIMECLUSTER operation environment:

Procedure

1. Boot the machine to be restored in single-user mode.

Note

After booting the machine in single-user mode, mount the file system by executing the "mountall -l" command and the "zfs mount -a command".

Example: Reboot in single-user mode.

```
# shutdown -g0 -i0 -y
...
ok boot -s
...
# mountall -l
# zfs mount -a
```

2. To restore the PRIMECLUSTER operation environment, execute the following command.

```
/opt/SMAW/bin/cfrestore [ -f ] [ -p ] [ -y ] [ -M ] [ n ]
```

See

For details on the command parameters, refer to the explanation of the cfrestore(1M) command.

Example: Restore the PRIMECLUSTER operation environment.

```
# cfrestore
08/07/02 11:17:55 cfrestore 1 started
...
08/07/02 11:21:23 cfrestore 1 ended
#
```

3. Restore a single configuration file.

Note

The restoration function does not restore backed-up files. The system files mentioned here in are provided by the operation system (environment). Restore the backed-up files manually if necessary.

Example: Pick up the "/etc/vfstab.pcl" file from the backup data "fuji2_ccbr1.tar.Z".

1. Confirm the backup data "fuji2_ccbr1.tar.Z" and then check for the backed-up system file.

```
# ls /var/spool/SMAW/SMAWccbr/*.tar.Z
fuji2_ccbr1.tar.Z
# zcat /var/spool/SMAW/SMAWccbr/fuji2_ccbr1.tar.Z | tar tvf - ./OS
drwxrwxrwx   0/1        0 Nov 20 13:58 2002 ./OS/
drwxrwxrwx   0/1        0 Nov 20 13:57 2002 ./OS/etc/
...
-rw-r--r--   0/1      909 Nov 20 13:57 2002 ./OS/etc/vfstab.pcl
...
#
```

2. Put the "/etc/vfstab.pcl" file in the "/var/tmp" directory.

```
# cd /var/tmp
# zcat /var/spool/SMAW/SMAWccbr/fuji2_ccbr1.tar.Z | tar xvf - ./OS
x ./OS, 0 bytes, 0 tape blocks
x ./OS/etc, 0 bytes, 0 tape blocks
x ./OS/etc/hosts, 401 bytes, 1 tape blocks
...
```

```
x ./OS/etc/vfstab.pcl, 909 bytes, 2 tape blocks
...
#
```

3. Restore part or all of the file contents by executing the "cp(1)" or "vi(1)" command.

Note

If an invalid correction is made to the system file, login may be disabled and the system may fail to boot.

4. If the AutoStartUp function of the RMS has been validated, invalidate it.

   For details, see the "13 Appendix - Environment variables" in the *"PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."*

5. Reboot the machine.

   Example: The machine is rebooted.

```
# shutdown -g0 -i6 -y
```

6. Distribute the RMS configuration.

   After completing procedures up to 5 using all cluster nodes, log in with system administrator authority and execute the command below using one of the nodes which configure the clusters.

```
# /opt/SMAW/SMAWRrms/bin/hvw -xj -F Configuration-Activate
```

7. Enable the AutoStartup function if it is disenabled.

   For details, see the "13 Appendix - Environment variables" in the *"PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."*

8. Start up the RMS.

   For details, see the "8.1.1 Starting RMS" in the *"PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."*

13.3 Environment Setup File

The two environment configuration files for backup and restoration are "ccbr.conf" and "ccbr.gen". The storage destination of the data to be backed up or restored can be changed by editing these files using the text editor "vi(1)".

Environmental variables and environment configuration files used for backup and restoration are shown below:

- /opt/SMAW/ccbr/ccbr.conf

  This file defines variables for PRIMECLUSTER backup and restoration. The variables include the following:

  CCBRHOME :

  This variable indicates a directory. The backup data is stored into this directory.
  The following is a default value:

  `/var/spool/SMAW/SMAWccbr`

- /opt/SMAW/ccbr/ccbr.gen

  This file defines the number of records for PRIMECLUSTER backup. This number increases one by one whenever the backup process is done.

Information

The number of generations can be also changed with the arguments of cfbackup (1M) / cfrestore(1M). For details, refer to the explanation of each command.

Example

The backup configuration is as follows:

<Setup contents>

- Host name: fuji2
- CCBRHOME variable: /var/spool/SMAW/SMAWccbr
- The value of the "ccbr.gen" file: 1

In the case of above, the following file will be created as the backup data:

```
/var/spool/SMAW/SMAWccbr/fuji2_ccbr1.tar.Z
```

Configuration procedure example

Configure the backup data storage destination to "/var/spool/SMAW/SMAWccbr" and the number of generations to "1".

1. Edit the "/opt/SMAW/ccbr/ccbr.conf" file using the "vi(1)" text editor, and then change the value of the CCBRHOME file to "/var/spool/SMAW/SMAWccbr".
   If the value has already been entered, go to Procedure 2.

   Confirm the contents.

   ```
   # cat /opt/SMAW/ccbr/ccbr.conf
   ...
   CCBRHOME=/var/spool/SMAW/SMAWccbr
   export CCBRHOME
   ...
   #
   ```

2. Edit the "/opt/SMAW/ccbr/ccbr.gen" file using the "vi(1)" text editor, and then change the value to "1". If the value has already been configured, go to Procedure 3.

   Confirm the contents.

   ```
   # cat /opt/SMAW/ccbr/ccbr.gen
   1
   #
   ```

3. Back up the PRIMECLUSTER configuration.

   ```
   # cfbackup
   08/07/02 11:17:55 cfbackup 1 started
   ...
   08/07/02 11:21:23 cfbackup 1 ended
   #
   ```

   Confirm the contents of the backup file.

   ```
   # ls /var/spool/SMAW/SMAWccbr
   fuji2_ccbr1
   ```

Example

The restoration configuration is as follows:

<Setup contents>

- Host name: fuji2
- CCBRHOME variable: /var/spool/SMAW/SMAWccbr

- Value of the "ccbr.gen" file: 2

In the above case, the following file will be used as the data to be restored:

`/var/spool/SMAW/SMAWccbr/fuji2_ccbr1`

Configuration procedure example

Configure the destination of the restoration data to "/var/spool/SMAW/SMAWccbr" and the number of generations to "1".

1. Edit the "/opt/SMAW/ccbr/ccbr.conf" file using the "vi(1)" text editor, and then make the value of the CCBRHOME variable "/var/spool/SMAW/SMAWccbr".
   If the value has already been configured, proceed to Procedure 2.

   Confirm the contents.

   ```
   # cat /opt/SMAW/ccbr/ccbr.conf
   ...
   CCBRHOME=/var/spool/SMAW/SMAWccbr
   export CCBRHOME
   ...
   #
   ```

2. Edit the "/opt/SMAW/ccbr/ccbr.gen" file using the "vi(1)" text editor, and then change the value to "2" (specify a value that is 1 greater than the target number of generations).
   If the value has already been configured, proceed to Procedure 3.

   Confirm the contents.

   ```
   # cat /opt/SMAW/ccbr/ccbr.gen
   2
   #
   ```

3. Restore the PRIMECLUSTER environment.

   ```
   # cfrestore
   08/07/02 11:17:55 cfrestore 1 started
   ...
   08/07/02 11:21:23 cfrestore 1 ended
   #
   ```

Part 6 Virtualized Environments

This part describes Oracle VM Server for SPARC environment and Oracle Solaris Zones environment as virtualized environments where you can build PRIMECLUSTER systems.

Chapter 14 Using PRIMECLUSTER in Oracle VM Server for SPARC Environment

This chapter provides explanations on applying PRIMECLUSTER in Oracle VM Server for SPARC Environments.

- Procedure for configuration of PRIMECLUSTER in Oracle VM Server for SPARC Environments
- Precautions on using cluster systems in Oracle VM Server for SPARC Environments
- Maintenance of cluster systems in Oracle VM Server for SPARC Environments
- Collection of Diagnostic Data in Oracle VM Server for SPARC Environments
- Recommended configuration

See

For further details on Oracle VM Server for SPARC, refer to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide."

14.1 Procedure for Configuration of PRIMECLUSTER in Oracle VM Server for SPARC Environments

Below section explains the procedure for configuring control domain clusters in Oracle VM Server for SPARC Environments.

14.1.1 Software Installation and Configuration of Cluster Environment

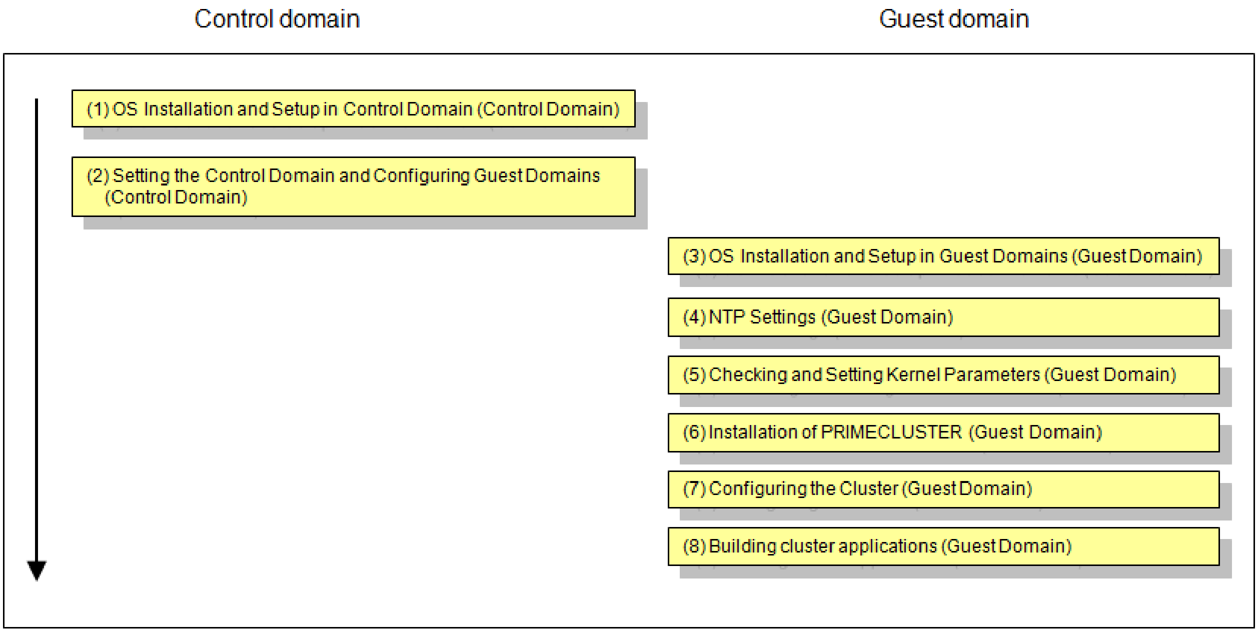
Install the Solaris software and the software required for PRIMECLUSTER in the control and guest domains of all nodes. Then make the OS and hardware settings required for actual implementation and operation.

For further details on configuring Oracle VM Server for SPARC, refer to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide."

The building procedure is different according to the following cluster systems you are using:

- Cluster system between guest domains within a same physical partition. (Supported only for SPARC M10)
- Cluster system between guest domains among different physical partitions. (Supported only for SPARC M10)
- Cluster system between control domains.

14.1.1.1 Cluster System Between Guest Domains Within a Same Physical Partitions



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   Disk-Related Settings

   For using the shared disk device, installation and setup of related software products are required.
   Carry out these settings in the control domain before installing PRIMECLUSTER.
   For details, refer to "3.2.2 Setting Up Disk Units".

   Do not register the shared disk device on the control domain.

2. Setting the Control Domain and Configuring Guest Domains

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

   Sample Configuration of Control and Guest Domains

   As the virtual disk used as the system disk, specify the file name or the block device.

   For a disk used as the switching disk in the cluster, make sure to specify the block device and export the same number of disks as guest domains which share the disk for each disk.

   When doing so, pay attention to the following points

   - Specify a disk as the full disk without specifying the slice option

   - When exporting one disk more than once, specify the -f option at the second time or later.

   Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device of the system disk.
ldm add-vdiskserverdevice <file name of disk 1> <volume name 1>@<virtual disk service name>
ldm add-vdiskserverdevice < file name of disk 2> <volume name 2>@<virtual disk service name>
# Define the virtual disk server device of the shared disk.
```

```
ldm add-vdiskserverdevice <block name> <volume name 3>@<virtual disk service name>
ldm add-vdiskserverdevice -f <block name> <volume name 4>@<virtual disk service name>
```

Configuration Script for Guest Domain

```
VDISK0=<volume name 1>@<virtual disk service name>
VDISK1=<volume name 2>@<virtual disk service name>
VSHDISK0=<volume name 3>@<virtual disk service name>
VSHDISK1=<volume name 4>@<virtual disk service name>
DOMAIN1=<domain name 1>
DOMAIN2=<domain name 2>
[...]
ldm add-vdisk <virtual disk name 1> $VDISK0 $DOMAIN1
ldm add-vdisk <virtual disk name 2> $VDISK1 $DOMAIN2
ldm add-vdisk timeout=360 <virtual disk name 3> $VSHDISK0 $DOMAIN1
ldm add-vdisk timeout=360 <virtual disk name 4> $VSHDISK1 $DOMAIN2
```

Example:

Scripts for the following configurations

file name of disk 1:/LDoms/Vol1/vdisk0.img

file name of disk 2:/LDoms/Vol2/vdisk0.img

block name:/dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2

volume name 1: vol1_ldom1

volume name 2: vol1_ldom2

volume name 3: vol2_ldom1

volume name 4: vol2_ldom2

virtual disk service name: primary-vds0

domain name 1: ldom1

domain name 2: ldom2

virtual disk name 1: vdisk0

virtual disk name 2: vdisk1

virtual disk name 3: vshdisk0

virtual disk name 4: vshdisk1

Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device of the system disk.
ldm add-vdiskserverdevice /LDoms/Vol1/vdisk0.img vol1_ldom1@primary-vds0
ldm add-vdiskserverdevice /LDoms/Vol2/vdisk0.img vol1_ldom2@primary-vds0
# Define the virtual disk server device of the shared disk.
ldm add-vdiskserverdevice /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2 vol2_ldom1@primary-
vds0
ldm add-vdiskserverdevice -f /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2
vol2_ldom2@primary-vds0
```

Configuration Script for Guest Domain

```
VDISK0=vol1_ldom1@primary-vds0
VDISK1=vol1_ldom2@primary-vds0
VSHDISK0=vol2_ldom1@primary-vds0
VSHDISK1=vol2_ldom2@primary-vds0
DOMAIN1=ldom1
DOMAIN2=ldom2
```

```
[...]
ldm add-vdisk vdisk0 $VDISK0 $DOMAIN1
ldm add-vdisk vdisk1 $VDISK1 $DOMAIN2
ldm add-vdisk timeout=360 vshdisk0 $VSHDISK0 $DOMAIN1
ldm add-vdisk timeout=360 vshdisk1 $VSHDISK1 $DOMAIN2
```

3. OS Installation and Setup in Guest Domains (Guest Domain)

   Install the operating system in each guest domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete installation of operating systems in all guest domains.

4. NTP Settings (Guest Domain)

   This setup serves to synchronize the clocks of every node in the cluster system. Make sure to make this setting when configuring a cluster.
   Carry out these settings in the guest domain before installing PRIMECLUSTER.

5. Checking and Setting Kernel Parameters (Guest Domain)

   For running PRIMECLUSTER-related software, the OS kernel parameters need to be adjusted to the environment.
   Carry out these settings in the guest domain before rebooting the system after "Step 6 Installation of PRIMECLUSTER (Guest Domain)".
   For details, refer to "3.2.3 Checking the Kernel Parameters."

6. Installation of PRIMECLUSTER (Guest Domain)

   For details on installing PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation."

7. Configuring the Cluster (Guest Domain)

   Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster", make the initial cluster settings in the guest domain.

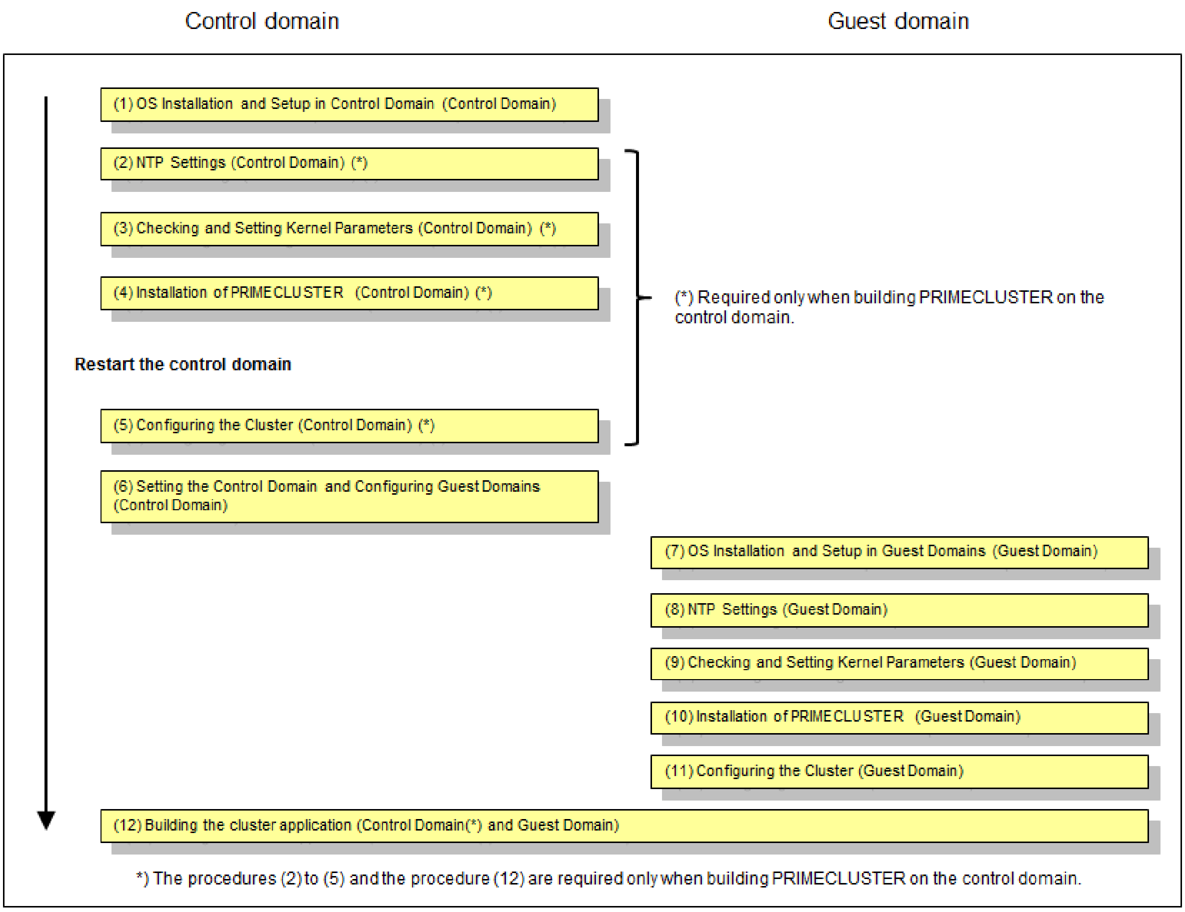
   Note

   For using the Migration function of Oracle VM Server for SPARC, see "Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment."

8. Building cluster applications (Guest Domain)

   Build cluster applications. For details on how to build cluster applications on guest domains, see "Chapter 6 Building Cluster Applications."

14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC Environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   Disk-Related Settings

   For using the shared disk device, installation and setup of related software products are required.
   Carry out these settings in the control domain before installing PRIMECLUSTER.
   For details, refer to "3.2.2 Setting Up Disk Units".

   At this time, do not set GDS yet. Set GDS in "Step 12. Building the cluster application (Control Domain and Guest Domain)."

2. NTP Settings (Control Domain)

   This setup serves to synchronize the clocks of every node in the cluster system. Make sure to make this setting when configuring a cluster.
   Carry out these settings in the control domain before installing PRIMECLUSTER.

3. Checking and Setting Kernel Parameters (Control Domain)

   For running PRIMECLUSTER-related software, the OS kernel parameters need to be adjusted to the environment.
   Carry out these settings in the control domain before rebooting the system after "Step 4 Installation of PRIMECLUSTER (Control Domain)".
   For details, refer to "3.2.3 Checking the Kernel Parameters".

4. Installation on PRIMECLUSTER (Control Domain)

   For details on installing PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation".

5. Configuring the Cluster (Control Domain)

   Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster", make the initial cluster settings in the control domain. To build clusters, perform procedures from "5.1.1 Setting Up CF and CIP" through "5.1.2 Configuring the Shutdown Facility" and also perform "5.2 Setting Up Power Supply Linkage" as needed.

   Note

   - Set the timeout value of the cluster system on the control domain in CF settings as follows:

     Timeout value: 20 seconds (Change the timeout value by using the cfset command.)

   - When a failure of the control domain (including the cluster application error) is detected and the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

   - When the virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.

6. Setting the Control Domain and Configuring Guest Domains

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

   Note

   When performing the migration of the cluster on the control domains, the network must be set so that the control domains can access the guest domains via SSH.

   For more information, see "Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment."

   Sample Configuration of Control and Guest Domains

   As the virtual disk used as the system disk, specify the file name or the block device.
   For the disk used as the switching disk in the cluster, specify the block device, for example, /dev/dsk/cXtXdXs2. When doing so, specify the disk as the full disk without specifying the slice option.

   Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice <disk name> <volume name>@<virtual disk service name>
ldm add-vdiskserverdevice <block name> <volume name>@<virtual disk service name>
```

   Configuration Script for Guest Domain

```
VDISK0=<volume name 1>@<virtual disk service name>
VSHDISK0=<volume name 2>@<virtual disk service name>
[...]
ldm add-vdisk <virtual disk service name 1> $VDISK0 $DOMAIN1
ldm add-vdisk timeout=360 <virtual disk service name 2> $VSHDISK0 $DOMAIN1
```

   Example:

   Script example of the following configuration

   Disk name 1: /LDoms/Vol1/vdisk0.img

   Block name: /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2

   Volume name 1: vol1

   Volume name 2: vol2

   Virtual disk service nam: primary-vds0

Virtual disk name 1: vdisk0

Virtual disk name 2: vshdisk0

Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice /LDoms/Vol1/vdisk0.img vol1@primary-vds0
ldm add-vdiskserverdevice /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2 vol2@primary-vds0
```

Configuration Script for Guest Domain

```
VDISK0=vol1@primary-vds0
VSHDISK0=vol2@primary-vds0
[...]
ldm add-vdisk vdisk0 $VDISK0 $DOMAIN
ldm add-vdisk timeout=360 vshdisk0 $VSHDISK0 $DOMAIN
```

7. OS Installation and Setup in Guest Domains (Guest Domain)

   Install the operating system in each guest domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete installation of operating systems in all guest domains.

8. NTP Settings (Guest Domain)

   This setup serves to synchronize the clocks of every node in the cluster system. Make sure to make this setting when configuring a cluster.
   Carry out these settings in the guest domain before installing PRIMECLUSTER.

9. Checking and Setting Kernel Parameters (Guest Domain)

   For running PRIMECLUSTER-related software, the OS kernel parameters need to be adjusted to the environment.
   Carry out these settings in the guest domain before rebooting the system after "Installation of PRIMECLUSTER (Guest Domain)".
   For details, refer to "3.2.3 Checking the Kernel Parameters".

10. Installation of PRIMECLUSTER (Guest Domain)

    For details on installing PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation".

11. Configuring the Cluster (Guest Domain)

    Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster", make the initial cluster settings in the guest domain.

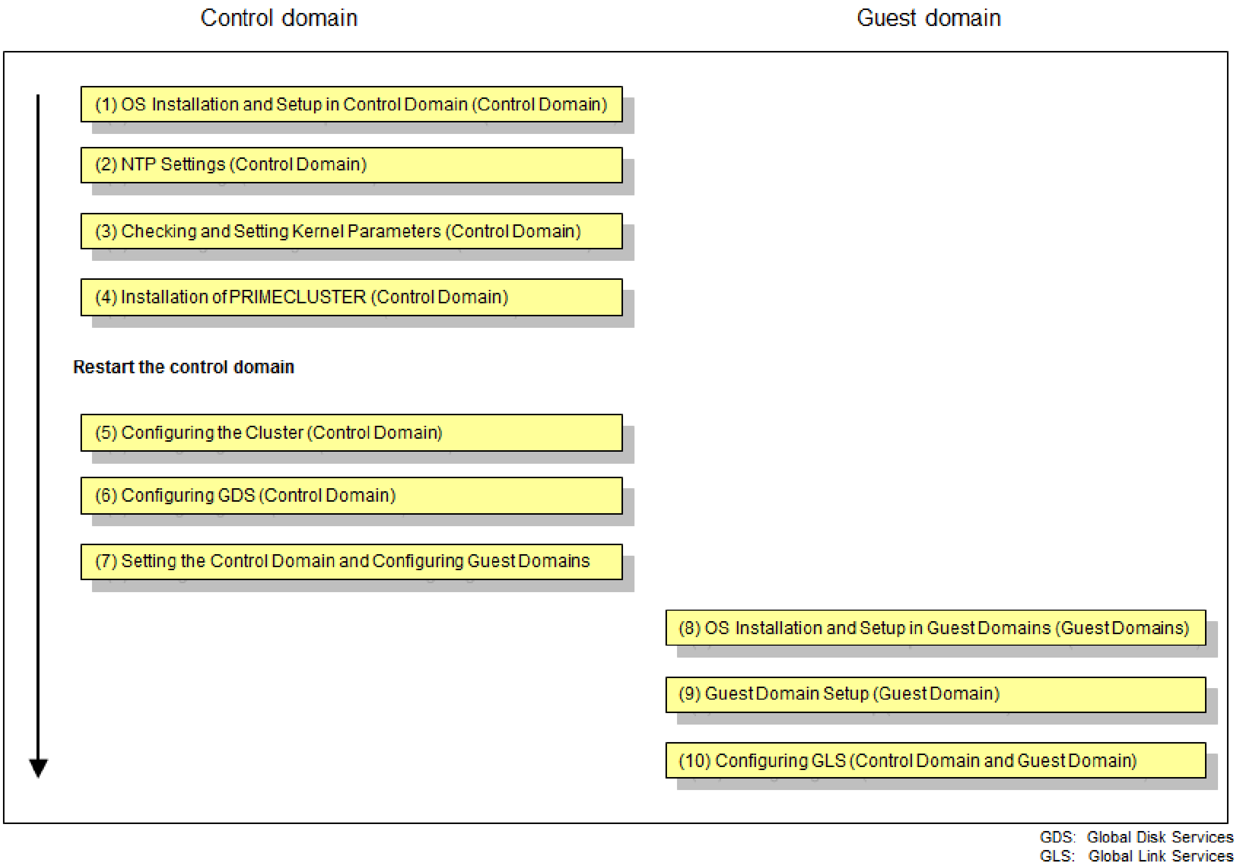
    Note

    - When building PRIMECLUSTER also on the control domain, do not change the timeout value on guest OSes from 10 seconds to a different value in CF settings.
    - For using the Migration function of Oracle VM Server for SPARC, see "Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment."

12. Building the cluster application (Control Domain and Guest Domain)

    Build the cluster application. For details on how to build the cluster application on the guest domain, see "Chapter 6 Building Cluster Applications."

14.1.1.3 Cluster System Between Control Domains



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC Environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   Disk-Related Settings

   For using the shared disk device, installation and setup of related software products are required.
   Carry out these settings in the control domain before installing PRIMECLUSTER.
   For details, refer to "3.2.2 Setting Up Disk Units".

   At this time, do not set GDS yet. Set GDS in "Step 6. Configuring GDS (Control Domain)."

2. NTP Settings (Control Domain)

   This setup serves to synchronize the clocks of every node in the cluster system. Make sure to make this setting when configuring a cluster.
   Carry out these settings in the control domain before installing PRIMECLUSTER.

3. Checking and Setting Kernel Parameters (Control Domain)

   For running PRIMECLUSTER-related software, the OS kernel parameters need to be adjusted to the environment.
   Carry out these settings in the control domain before rebooting the system after "Installation of PRIMECLUSTER (Control Domain)".
   For details, refer to "3.2.3 Checking the Kernel Parameters".

4. Installation of PRIMECLUSTER (Control Domain)

   For details on installing PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation".
   After installing PRIMECLUSTER, make the settings for the OBP auto-boot? variable and in /etc/system.

Setting the OBP auto-boot? Variable

In order to prevent the control domain from rebooting automatically after an enforced stop, set the OBP auto-boot? variable to "false". Otherwise, the suspended control domain may be automatically rebooted, possibly causing data loss due to restored virtual in- and output from guest domains that were stopped.

Example:

```
{1} ok setenv auto-boot? false
auto-boot? =          false
{1} ok reset-all
```

After setting the OBP auto-boot? variable, reboot the node to make the OBP settings effective.

Settings in /etc/system

To prevent an automatic reboot of a panicked node, perform the following settings for /etc/system. Then restart it.

```
set halt_on_panic=1
```

After the panic occurs, the node is stopped in OBP mode with this setting.

5. Configuring the Cluster (Control Domain)

   Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster", make the initial cluster settings in the control domain.

6. Configuring GDS (Control Domain)

   Referring to "6.3 Initial GDS Setup", make the GDS settings in the control domain.
   For further details on how to make GDS settings, refer to the "PRIMECLUSTER Global Disk Services Configuration and Administration Guide".

Note

- For enhanced availability, we recommend allocating virtual disks to mirror volumes if they are to be used as guest domain system disks.
  When configuring GDS, create a root-class or local-class mirror volume on which the files for the virtual disks shall be located.
  Do not deploy any files for virtual disks on shared-class volumes.
- When using multiple guest domains, create shared classes separately for each guest domain.

7. Setting the Control Domain and Configuring Guest Domains

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

Sample Configuration of Control and Guest Domains

As the virtual disk used as the system disk, specify the file name or the block device.
As the shared-class volume to be used as the switching disk in the cluster, specify a block device like /dev/sfdsk/class0001/dsk/volume0001.
In such cases, make sure to specify the "slice" option.

Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice <disk name> <volume name>@<virtual disk service name>
ldm add-vdiskserverdevice <block name> <volume name>@<virtual disk service name>
```

Configuration Script for Guest Domain

```
VDISK0=<volume name 1>@<virtual disk service name>
VSHDISK0=<volume name 2>@<virtual disk service name>
[...]
```

```
ldm add-vdisk <virtual disk service name 1> $VDISK0 $DOMAIN1
ldm add-vdisk timeout=360 <virtual disk service name 2> $VSHDISK0 $DOMAIN1
```

Example:

Script example of the following configuration

Disk name 1: /LDoms/Vol1/vdisk0.img

Block name: /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2

Volume name 1: vol1

Volume name 2: vol2

Virtual disk service nam: primary-vds0

Virtual disk name 1: vdisk0

Virtual disk name 2: vshdisk0

Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice /LDoms/Vol1/vdisk0.img vol1@primary-vds0
ldm add-vdiskserverdevice options=slice /dev/sfdsk/class0001/dsk/volume0001 vol2@primary-vds0
```

Configuration Script for Guest Domain

```
VDISK0=vol1@primary-vds0
VSHDISK0=vol2@primary-vds0
[...]
ldm add-vdisk vdisk0 $VDISK0 $DOMAIN
ldm add-vdisk timeout=360 vshdisk0 $VSHDISK0 $DOMAIN
```

Creation of Virtual Network

For administrative/user LAN communication via virtual network (vnet) devices in the guest domain, create a virtual switch service (vsw) in the control domain and add the vnet devices to the guest domain.

Besides this, for creating network redundancy and for switching clusters after a network failure in environments using virtual network devices for user LAN communication of guest domains, GLS must be installed in both control and guest domains. For using GLS, create a virtual switch service and a virtual network device for multiple physical network devices.

See

For details on how to create a virtual network, refer to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide."

Note

In control domain clusters, the only redundancy method supported by GLS is NIC switching.

8. OS Installation and Setup in Guest Domains (Guest Domains)

   With PRIMECLUSTER, in order to enable cluster switching, an operating system must be installed in each guest domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete installation of operating systems in all guest domains.
   After installing the operating systems, install GLS in the guest domains.

See

For details on installing GLS, refer to the "PRIMECLUSTER GLS Installation Guide".

Note

The guest domains recognize the GDS volume created in the control domain as a virtual disk. At this point, the number "0" is allocated as the slice, rendering the device name into a format like, e.g., /dev/dsk/c0d1s0. As this name is different from the device name in the control domain, be careful not to confuse the disk to be used with another one.

Example) The following is an example for using the virtual disk /dev/dsk/c0d1s0 in a guest domain.

```
# prtvtoc /dev/rdsk/c0d1s0
* /dev/rdsk/c0d1s0 (volume "vdisk") partition map
*
* Dimensions:
*     512 bytes/sector
*    2048 sectors/track
*      16 tracks/cylinder
*   32768 sectors/cylinder
*     874 cylinders
*     874 accessible cylinders
*
* Flags:
*   1: unmountable
*  10: read-only
*
* Unallocated space:
*       First     Sector    Last
*       Sector     Count    Sector
*           0         1         0
*
*                          First     Sector    Last
* Partition  Tag  Flags    Sector     Count    Sector  Mount Directory
 0       0    00          1  28639232  28639232
# newfs /dev/dsk/c0d1s0
```

9. Guest Domain Setup (Guest Domain)

   In order to reboot the OS automatically after switching guest domains, set the OBP auto-boot? variable to "true".

   Example:

```
{1} ok setenv auto-boot? true
auto-boot? =          true
{1} ok reset-all
```

   After setting the OBP auto-boot? variable, reboot the guest domain to make the OBP settings effective.

Note

When the power of a node is turned on, guest domains need to stay stopped until they get started by PRIMECLUSTER. Configure the domain configuration in system controller so that all guest domains controlled by PRIMECLUSTER remain stopped when their control domain gets started. For details, refer to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide."

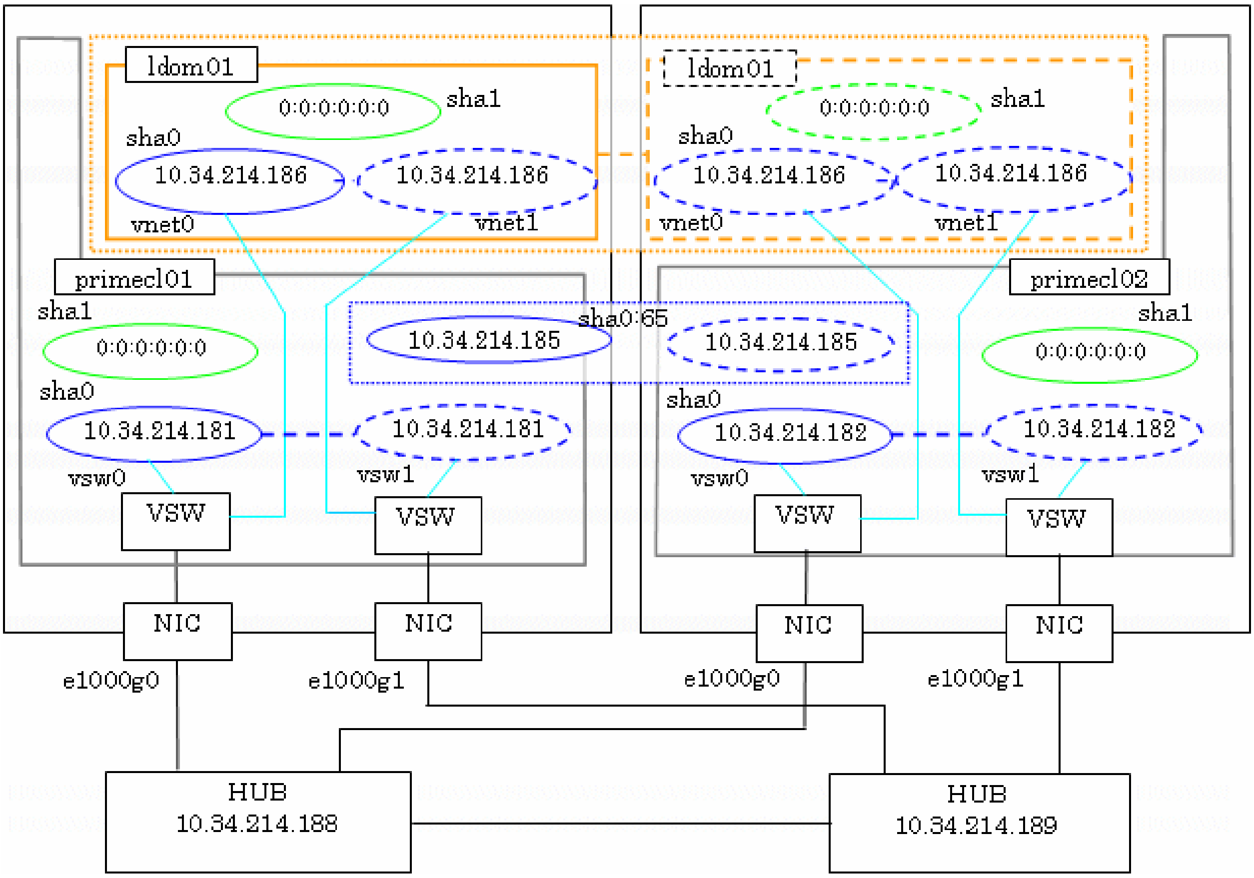
10. Configuring GLS (Control Domain and Guest Domain)

    For information on the initial GLS setup, refer to "Chapter 5 Operation on Cluster System" in the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function".

This section uses below example to explain the setup procedure for GLS (redundant line control function) for a 1:1 operational/ standby setup of a control domain cluster with IPv4 NIC switching in an Oracle VM Server for SPARC Environment. Interface names listed in examples of setting up vary depending on the environment. Replace interface names according to the Environment. For Solaris 11 OS, the default interface name is netX (X means the instance number).

Note

In control domain clusters, the only redundancy method supported by GLS is NIC switching.
The setting values for each category correspond to those in the "GLS Setup Worksheet".



In this sample configuration, control domains primecl01 and primecl02 are forming a cluster, whereas guest domain ldom01 runs as the cluster resource. The control and guest domains share the same user LAN network lines via a virtual switch service.
When user LAN transmissions on the operational node are disabled, GLS in the control domains detects a failure in the transmission route and switches clusters. When clusters are switched, the guest domains in the operational node are stopped, and, instead, the guest domains in the standby node are started.
The IP addresses of each control domains (10.34.214.181 for primecl01, 10.34.214.182 for primecl02) can be accessed independently of their operational statuses; however, setting takeover IP addresses allows you to automatically access the control domains currently in use, even if you are not sure which one it is.

Operating Procedure

- If the operating node is [HOST-primecl01]

1. Make the system settings.

    1. Define the IP addresses and host names to be used in the /etc/inet/hosts file.

```
10.34.214.185    takeoverIP    # virtual IP
10.34.214.181    primecl01     # physical IP of primecl01
10.34.214.182    primecl02     # physical IP of primecl02
10.34.214.186    ldom01        # physical IP of ldom01
```

```
10.34.214.188   swhub01      # IP of primary monitoring target hub
10.34.214.189   swhub02      # IP of secondary monitoring target hub
```

   2. Define the subnet masks in the /etc/inet/netmasks file.

```
10.34.214.0     255.255.255.0
```

   3. Set up interface to use.

      - For Solaris 10

      Add the host names from above definitions to the /etc/hostname.vsw0 file.

      Contents of /etc/hostname.vsw0:

```
primecl01
```

      - For Solaris 11

      Set up with the interface and the host name defined above using ipadm(1M) command.

      Interface net0 (corresponding to vsw0 in the figure)

```
# /usr/sbin/ipadm create-ip net0
# /usr/sbin/ipadm create-addr -T static -a primecl01/24 net0/v4
```

   4. Create the virtual network to be used in the user LAN.

```
# ldm add-vsw net-dev=e1000g0 primary-vsw0 primary
# ldm add-vsw net-dev=e1000g1 primary-vsw1 primary
# ldm add-vnet vnet0 primary-vsw0 ldom01
# ldm add-vnet vnet1 primary-vsw1 ldom01
# touch /reconfigure
```

2. Reboot the system.

   Run the following command and reboot the system. After rebooting the system, verify the interface set using ifconfig command is enabled.

```
# /usr/sbin/shutdown -y -i6 -g0
```

3. Make the virtual interface settings.

   The underlined portion specifies the physical IP address.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.34.214.185 -e 10.34.214.181 -
t vsw0,vsw1
```

   After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.181     vsw0,vsw1

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+--------------------------+
```

4. Make the settings for HUB monitoring.

   The underlined portions specify the IP addresses of the HUB to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+---------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
         hub-hub poll       =    OFF
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES
```

5. Make the settings for the GLS standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.181     vsw0,vsw1
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+---------------------------+
```

6. Create the takeover virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------+
 sha0:65    10.34.214.185    -
```

- If the standby node is [HOST-primecl02]

1. Make the system settings.

    1. Define the IP addresses and host names to be used in the /etc/inet/hosts file.

       The contents to be defined are the same as for primecl01.

    2. Define the subnet masks in the /etc/inet/netmasks file.

       The contents to be defined are the same as for primecl01.

3. Set up interface to use.

   - For Solaris 10

   Add the host names from above definitions to the /etc/hostname.vsw0 file.

   Contents of /etc/hostname.vsw0:

```
primecl02
```

   - For Solaris 11

   Set up with the interface and the host name defined above using ipadm(1M) command.

   Interface net0 (corresponding to vsw0 in the figure)

```
# /usr/sbin/ipadm create-ip net0
# /usr/sbin/ipadm create-addr -T static -a primecl02/24 net0/v4
```

4. Create the virtual network to be used in the user LAN.

```
# ldm add-vsw net-dev=e1000g0 primary-vsw0 primary
# ldm add-vsw net-dev=e1000g1 primary-vsw1 primary
# ldm add-vnet vnet0 primary-vsw0 ldom01
# ldm add-vnet vnet1 primary-vsw1 ldom01
# touch /reconfigure
```

2. Reboot the system.

   Run the following command and reboot the system. After rebooting the system, verify the interface set using ifconfig command is enabled.

```
# /usr/sbin/shutdown -y -i6 -g0
```

3. Make the virtual interface settings.

   The underlined portion specifies the physical IP address.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.34.214.185 -e 10.34.214.182 -
t vsw0,vsw1
```

   After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.182     vsw0,vsw1

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+------------------------+
```

4. Make the settings for HUB monitoring.

   The underlined portions specify the IP addresses of the HUB to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

   After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)     =      5( 60) sec
         times              =      5 times
```

```
        max_retry          =      5 retry
        repair_time        =      5 sec
        link detection     =    NO
        failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+--------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
        hub-hub poll       =    OFF
        interval(idle)     =      5( 60) sec
        times              =      5 times
        max_retry          =      5 retry
        repair_time        =      5 sec
        link detection     =    NO
        failover mode      =    YES
```

5. Make the settings for the GLS standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.185    d   10.34.214.182     vsw0,vsw1
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+---------------------------+
```

6. Create the takeover virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------+
 sha0:65    10.34.214.185    -
```

- For guest domain [HOST-ldom01]

Make the GLS settings for guest domains on one node after another, first on the nodes that are designated to be operational, and then on the nodes designated to be standby nodes, making sure not to assign duplicate IP addresses.

1. Make the system settings.

    1. Define the IP addresses and host names to be used in the /etc/inet/hosts file.

    ```
    10.34.214.186   ldom01       # physical IP of ldom01
    10.34.214.188   swhub1       # IP of primary monitoring target hub
    10.34.214.189   swhub2       # IP of secondary monitoring target hub
    ```

    2. Define the subnet masks in the /etc/inet/netmasks file.

    ```
    10.34.214.0     255.255.255.0
    ```

3. Set up interface to use.

   - For Solaris 10

   Add the host names from above definitions to the /etc/hostname.vnet0 file.

   Contents of /etc/hostname.vnet0:

```
ldom01
```

   - For Solaris 11

   Set up with the interface and the host name defined above using ipadm(1M) command.

   Interface net0 (corresponding to vnet0 in the figure)

```
# /usr/sbin/ipadm create-ip net0
# /usr/sbin/ipadm create-addr -T static -a ldom01/24 net0/v4
```

2. Reboot the system.

   Run the following command and reboot the system. After rebooting the system, verify the interface set using ifconfig command is enabled.

```
# /usr/sbin/shutdown -y -i6 -g0
```

3. Make the virtual interface settings.

   The underlined portion specifies the physical IP address.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m e -i 10.34.214.186 -t vnet0,vnet1
```

   After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.186    e                     vnet0,vnet1

[IPv6]

 Name        Hostname/prefix                  Mode Interface List
+-----------+---------------------------------+----+---------------------------+
```

4. Make the settings for HUB monitoring.

   The underlined portions specify the IP addresses of the HUB to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

   After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
        interval(idle)     =      5( 60) sec
        times              =      5 times
        max_retry          =      5 retry
        repair_time        =      5 sec
        link detection     =    NO
        failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+---------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
```

```
        hub-hub poll       =    OFF
        interval(idle)     =      5( 60) sec
        times              =      5 times
        max_retry          =      5 retry
        repair_time        =      5 sec
        link detection     =    NO
        failover mode      =    YES
```

5. Make the settings for the GLS standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.186    e                     vnet0,vnet1
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                  Mode Interface List
+-----------+---------------------------------+----+--------------------------+
```

6. Start hub monitoring.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll on
```

7. Activate the virtual interface.

```
# /opt/FJSVhanet/usr/sbin/strhanet
```

14.1.2 Building Cluster Applications

Build cluster applications on guest domains or control domains.

For details on how to build cluster applications on guest domains, see "Chapter 6 Building Cluster Applications."

14.1.2.1 Building Cluster Applications on Clusters Between Control Domains

In the cluster between control domain in Oracle VM Server for SPARC Environments, create one cluster application for one guest domain to monitor the state of the guest domain. Configuration of each cluster application that monitors the state of the guest domain must include at least one Cmdline resource, and one or more Gls resources and Gds resources.
This section explains how to build the cluster application that monitors the state of the guest domain.
For details on the cluster application, see "Chapter 6 Building Cluster Applications".

- Cmdline resource for monitoring guest domain statuses

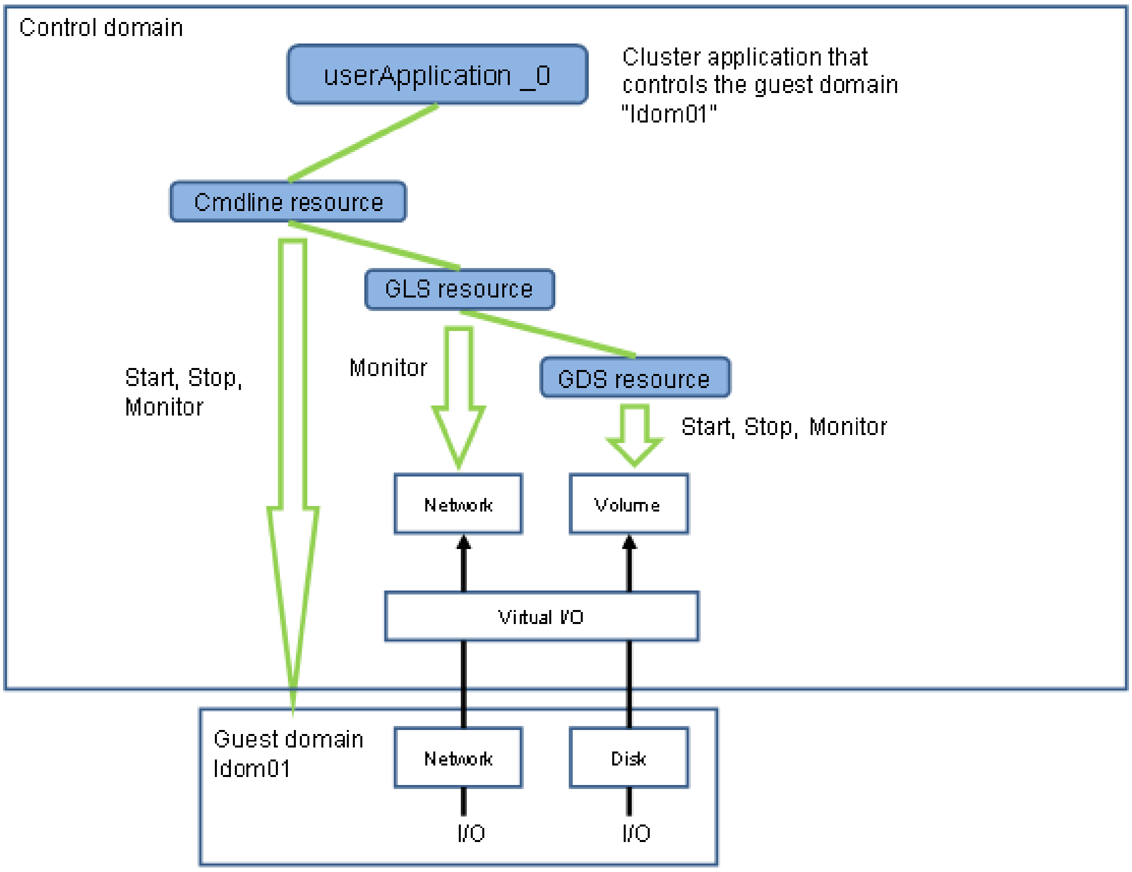
  Starts, stops and monitors the corresponding guest domain.

- Gls resource

  Monitors the networks used by the corresponding guest domain via virtual in- and output from the control domain.

- Gds resource

  Starts, stops, and monitors the volumes used by the corresponding guest domain via virtual in- and output from the control domain.

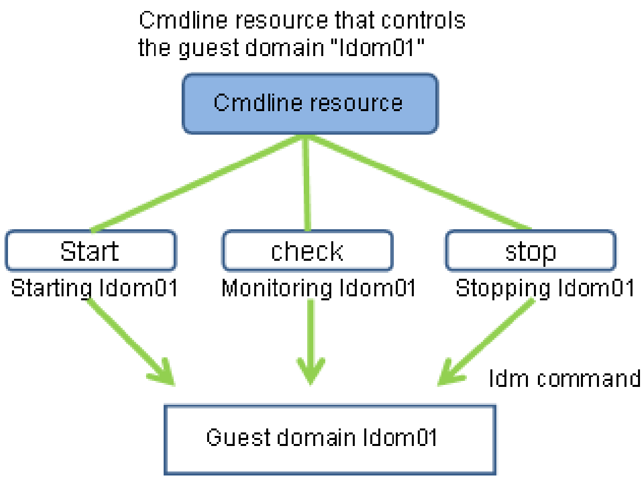


See below for how to configure the cluster application that monitors the state of the guest domain in the cluster between the control domain:

1. Creation of the Cmdline resource
2. Setup of the cluster application

14.1.2.1.1 Creation of Cmdline Resources for Monitoring Guest Domain Statuses

Set up Online, Offline and Check scripts so that the corresponding guest domain starts/stops in conjunction with userApplication.
Set the commands to start the corresponding guest domain to the Online script. Set the commands to stop the corresponding guest domain to the Offline script. In the Check script, set the commands to verify the corresponding guest domain's ability to start its OS.



Sample Scripts

This section presents sample scripts for Online (Start), Offline (Stop) and Check scripts.

Specify the target guest domain name for the argument of each script. Also if setting the script to the Cmdline resource, the target guest domain name is required to be the same name for each node.

Set a Cmdline resource with the scripts to userApplication. For details, refer to "6.7.1.1 Creating Cmdline Resources."

- Start script: starts the corresponding guest domains. Returns "0" when execution was successful, or any other value when failed.

```
#!/bin/sh
#
#  Sample online script for a guest domain
#       Exit with  0: Success   1: Fail
#
#  Copyright(c) 2008 FUJITSU LIMITED. All rights reserved.
#

LDOMSNAME=$1
LDM="/opt/SUNWldm/bin/ldm"

$LDM bind-domain $LDOMSNAME

$LDM start-domain $LDOMSNAME

# Exits on "0" (success), only when logical domain is active
$LDM list-domain $LDOMSNAME | tail -1 | awk '{print $2}' | grep "^active" && exit 0

exit 1
```

- Stop script: stops the corresponding guest domains. Returns "0" when execution was successful, or any other value when failed.

```
#!/bin/sh
#
#  Sample offline script for a guest domain
#       Exit with  0: Success   1: Fail
#
#  Copyright(c) 2008 FUJITSU LIMITED. All rights reserved.
#

LDOMSNAME=$1
LDM="/opt/SUNWldm/bin/ldm"

$LDM stop-domain $LDOMSNAME
[ 0 -ne $? ] && $LDM stop-domain -f $LDOMSNAME

$LDM unbind-domain $LDOMSNAME

# Exits on "0" (success), only when logical domain is inactive
$LDM list-domain $LDOMSNAME | tail -1 | awk '{print $2}' | grep "^inactive" && exit 0

exit 1
```

- Check script: monitors "active" statuses of corresponding guest domains. Returns "0" when domain is active, or any other value otherwise.

```
#!/bin/sh
#
#  Sample check script for a guest domain
#       Exit with  0: Online   1: Offline
#
#  Copyright(c) 2008 FUJITSU LIMITED. All rights reserved.
#
```

```
LDOMSNAME=$1
LDM="/opt/SUNWldm/bin/ldm"

# Exits on "0" (success), only when logical domain is active
$LDM list-domain $LDOMSNAME | tail -1 | awk '{print $2}' | grep "^active" && exit 0

exit 1
```

14.1.2.1.2 Setup of Cluster Applications

Make the settings for Cmdline resource, Gds resource and Gls resource cluster applications.

For details on Gds and Gls resource settings, refer to "Creating Gds Resources" and "Creating Gls Resources" in "6.7.1 Setting Up Resources". For details on cluster application settings, refer to "6.7 Setting Up Cluster Applications".

14.2 Precautions on Using Cluster Systems in Oracle VM Server for SPARC Environments

For operating cluster systems in Oracle VM Server for SPARC Environments, be careful about the following matters:

14.2.1 Notes on Clusters Between Control Domains

14.2.1.1 During Installation

Installation of PRIMECLUSTER

After installing PRIMECLUSTER, set the OBP auto-boot? variable to "false" and specify "set halt_on_panic=1" in /etc/system in order to prevent the control domain forcibly stopped from rebooting automatically.

For details, refer to "4. Installation of PRIMECLUSTER (Control Domain)" in "14.1.1.3 Cluster System Between Control Domains".

Setting of Guest Domain

PRIMECLUSTER on the control domain controls only startup and shutdown of guest domains. It does not control the operating systems and applications on guest domains. Therefore, set up your guest domains so that operating systems and applications are automatically booted in extension of each guest domain's startup. For details on how to automatically start operating systems in guest domains, refer to "9. Guest Domain Setup (Guest Domain)" in "14.1.1.3 Cluster System Between Control Domains".

Registering Cmdline resources for monitoring the guest domain status

The import of ZFS storage pool may fail on the switch destination guest domain if all of the following conditions are met:

- Each guest domain has different host identifier values (output of hostid).
- The ZFS storage pool is located on a disk (other than the system disk) shared between guest domains.

In the environment where all of the above conditions are met, the following settings are necessary to prevent this event.

Perform the procedure below when the guest domain is in "bound" state.

1. Set the same host identifier to each guest domain that shares a disk.

   Execute the following command on the control domain of each guest domain that shares the disk.

   The host identifier must be other than 00000000, and it must be the same value on each guest domain that shares the disk.

   ```
   # ldm set-domain hostid=<hhhhhhhh> <guest_domain_name>
   ```

   (<*hhhhhhhh*>: Specify the host identifier in 8-digit hexadecimal number.)

   <Execution example>

   ```
   # ldm set-domain hostid=1234abcd ldom01
   ```

2. Make sure the same host identifier is set to each guest domain.

   Execute the following command on the control domain and confirm that the same host identifier is set to each guest domain that shares the disk.

```
# ldm list-bindings <guest_domain_name> | awk '/HOSTID/,/0x/'
```

   <Execution example>

```
# ldm list-bindings ldom01 | awk '/HOSTID/,/0x/'
HOSTID
    0x1234abcd
```

14.2.1.2 During Operation

Controlling Guest Domains

Never use the ldm(1M) command immediately to start or stop a guest domain or to bind its resources.
Guest domains have to be started and stopped from PRIMECLUSTER.
Otherwise, you risk loss or damage of data.

Stopping the System While Guest Domains Are Operational

- Control Domains

  Run the shutdown(1M) or the init(1M) command to stop the system. Using reboot(1M), uadmin(1M) or similar commands may possibly cause data loss as the control domain is rebooted without suspended guest domains.

- Guest Domains

  To stop guest domains, run the shutdown(1M) command with the "i0" option.
  After that, put the cluster applications of the suspended guest domains in the control domain offline.

When Control Domain Is Stopped

Guest domains on a forcibly stopped node or partition, which have been monitored by the cluster system, are sometimes not stopped although the control domain has been forced to stop. Data loss may occur when restoring the control domain while any of the guest domains are not stopped.
In order to stop absolutely all input and output from guest domains, make sure to turn off the electrical power of nodes or partitions by either of the following methods before restoring the stopped control domain. After that, reboot the disconnected nodes or partitions. Also note that no kernel crash dump is collected for operating systems in guest domains.

- On the ILOM remote control, select "Immediate Power Off".
- Stop partitions with the poweroff command of XSCF.
- Press and hold the power button of the rack for a while.

See

- For details on the remote control, refer to the "Integrated Lights Out Manager User's Guide".
- For details on XSCF, refer to the "SPARC M10 Systems System Operation and Administration Guide".

Manual Switching of Cluster Applications

When guest domains are switched to a standby system, PRIMECLUSTER uses the ldm(1M) command to stop those guest domains.
Therefore, whenever PRIMECLUSTER suspends a guest domain, the OS and all applications in that guest domain stop abruptly.
In order to avoid such abrupt stops when switching cluster applications manually, log in to the guest domain OS and stop the OS by running the shutdown(1M) or a similar command before switching the cluster application.

Shutting Down the Control Domains

Before shutting down a control domain, stop RMS on the control domain to set Cmdline resources which control guest domains offline. For how to stop RMS, refer to "7.2.1.2 Stopping RMS".

14.3 Maintenance of Cluster Systems in Oracle VM Server for SPARC Environments

This section describes maintenance methods for cluster systems in Oracle VM Server for SPARC Environments.

14.3.1 Cluster Between Control Domains

14.3.1.1 Maintenance of the Control Domain

Carry out maintenance of the control domain in the same way as for normal clusters.

For further details, refer to "Chapter 12 Maintenance of the PRIMECLUSTER System".

14.3.1.2 Maintenance of Guest Domains

For doing maintenance work on guest domains, the statuses of cluster applications which correspond to the guest domains have to be "online". Follow the following procedure:

1. Check that the status of each cluster application corresponding to a guest domain in which to carry out maintenance is "online". If there are any cluster applications that are not online, shutdown the target guest domains with shutdown(1M) command, and then switch the guest domains to the intended node with hvswitch(1M) command.

   See

   For details on how to determine whether cluster applications are online, refer to "7.1.3.1 RMS Tree".

2. Log in to the guest domain in which to carry out maintenance and boot it in single-user mode.

3. Implement maintenance of the guest domain as necessary.

4. Switch back cluster applications if you switched them in above Step 1. If you have not switched any cluster applications in Step 1, boot the guest domain in multi-user mode by using shutdown -i6.

   See

   For details on how to switch back, refer to "7.2.2.3 Switching a Cluster Application".

14.4 Collection Troubleshooting Information in Oracle VM Server for SPARC Environment

After any trouble in the PRIMECLUSTER system, collect the following data, necessary for diagnosis, from all clustered control domains, from all guest domains specified for cluster applications, and from the cluster management server.
After that, contact our customer support.

- Control Domains

  Referring to "C.1 Collecting Troubleshooting Information", collect the following data:

  - PRIMECLUSTER diagnostic data

  - If the malfunction is reproducible, documentation describing procedures for reproduction

- Guest Domains

  Run fjsnap or FJQSS to collect the data needed for error diagnosis.
  For details on fjsnap, refer to "C.1.1 Executing the fjsnap Command". For details on FJQSS, refer to "C.1.2 Collecting Information by FJQSS(Information Collection Tool)."

Information

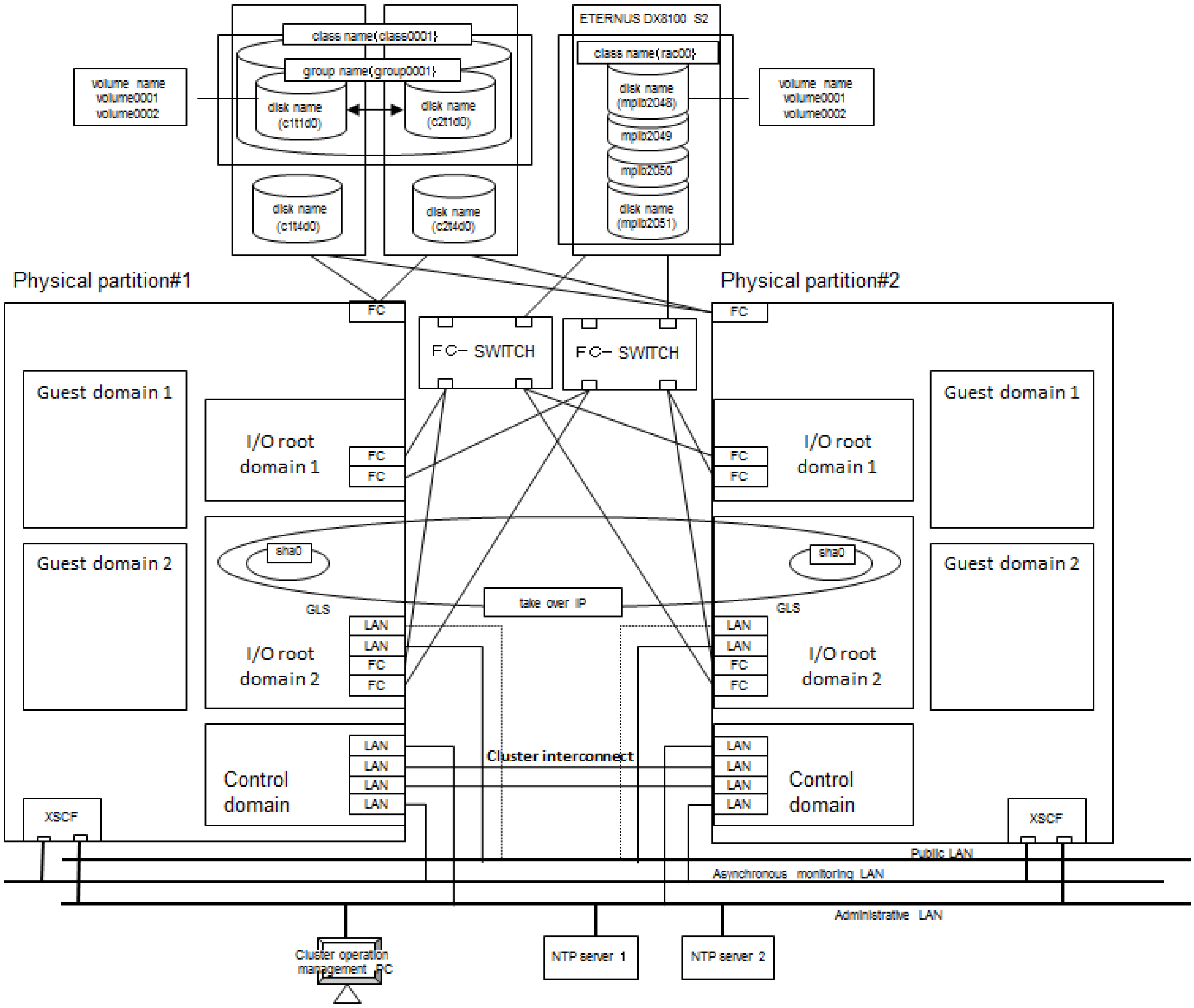
- For reporting a problem, first collect the necessary error diagnosis data. If you do not provide information enabling us to confirm the problem and reproduce the error, it may take us longer or even be impossible to simulate and diagnose the problem, hence, making it impossible to help you at all with troubleshooting.
- Collect the diagnostic material from all nodes in the PRIMECLUSTER System as soon as possible. Especially in the case of data to be collected by fjsnap or FJQSS, necessary information may get lost if too much time elapses after the error.
- For nodes that have been suspended by force, run "sync" in OBP mode in order to collect the system dump.

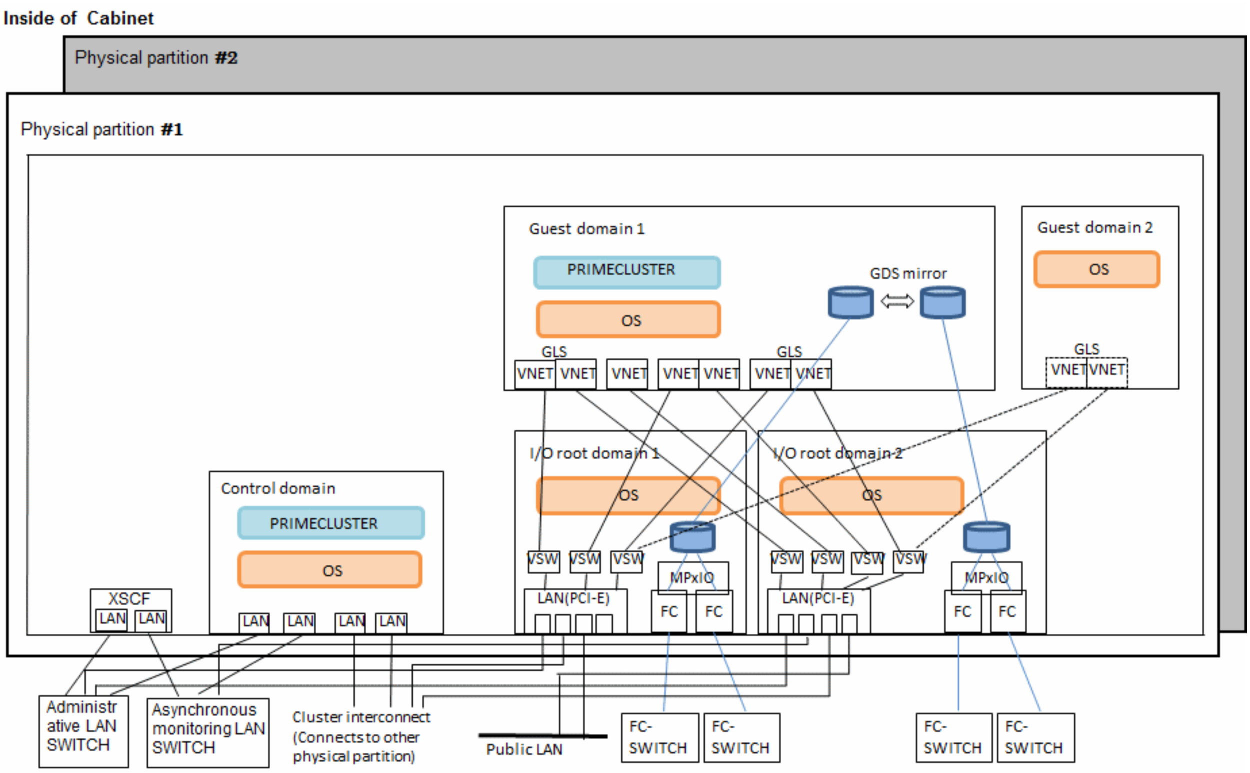
14.5 Recommended Configuration

This section presents a sample layout for a PRIMECLUSTER system in an Oracle VM Server for SPARC Environment. Use it for reference when designing your system.

System Configuration

Figure 14.1 Sample System Configuration





Utilized Hardware

- SPARC M10-4

Utilized Software

- Solaris 11
- Enhanced Support Facility 3.2
- PRIMECLUSTER 4.3
- Oracle VM Server for SPARC Manager Software 1.0.3
- Internet Explorer 9 (installed on administrator PC)
- Java(TM) 2 Platform Standard Edition Runtime Environment Version 6 (installed on administrator PC)

Shared Disk Storage System Settings

Device Name	LUs	RAID Groups	RAID Level	Size
ETERNUS DX8100 S2 (FC)	4	1	5	248.2 GB

Note

For using the disk storage system with UNIX, the disks must be formatted beforehand with the "format(1M)" command.
When you plan to format the disk storage system as part of your cluster system setup works, take sufficient time for formatting into consideration.

Sample Settings for Virtual Network

Node	Name of Control Domain	Name of Physical NIC	Name of Virtual Switch	Name of Guest Domain	Name of Virtual NIC
primecl01	primary	e1000g0	primary-vsw0	guest01	vnet0
		e1000g1	primary-vsw1	guest01	vnet1
primecl02	primary	e1000g0	primary-vsw0	guest02	vnet0
		e1000g1	primary-vsw1	guest02	vnet1

In the control domain, create virtual switches primary-vsw0 and primary-vsw1 using physical NICs e1000g0 and e1000g1, and add vnet0 and vnet1 as virtual NICs to be used in the guest domains. Use GLS to create redundant virtual NICs in each domain. Specify vsw0 and vsw1 in the control domain as well as vnet0 and vnet1 in the guest domains as virtual NICs to be made redundant by GLS.

Note

Settings for the virtual network are made by running the ldm(1M) command in the control domain.
Network adapters used in virtual networks must be compliant with GLDv3.

14.5.1 Cluster Configuration Worksheet

Below section presents a sample worksheet for cluster configuration.

	Item			Setting	Remarks
Shared disk unit: Type 1	Machine model			ETERNUS DX8100 S2	
	Device name			ETERNUSDX8000#003	
	IP address			10.33.66.215	
	Subnet mask			255.255.255.0	
	Interface			fibre channel	
	LUN number			0 - 0F	
	Ports			CM0CA2P0, CM1CA2P0	
	Connection paths			2	
	FC-SW		Use of fibre channnel switching hub	■ yes □ no	
			Machine model	ETERNUS SN200 model 630	
			Device name	ETERNUSSN200#003	
			IP address	10.33.66.203	
			Subnet mask	255.255.255.0	
	File system selection (type of file system used in cluster)			UFS	
NTP	Operation mode			□ NTP server (integrated in cluster) ■ NTP client	
	Protocol (only when broadcast is specified)				
	Server	Primary	Host name	pclntp1	
			IP address	10.124.95.11	
			Subnet mask	255.255.255.0	
		Secondary	Host name	pclntp2	
			IP address	10.21.8.3	
			Subnet mask	255.255.255.0	
Web-Based Admin View (Management view) For control domain	Operation mode			□ 1 ■ 2 □ 3 layers	
	Server	Primary	Host name	primecl01	
			IP address	10.33.66.81	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.81:8081/Plugin.html	
		Secondary	Host name	primecl02	
			IP address	10.33.66.82	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.82:8081/Plugin.html	
	User ID		wvroot group	pclwv	
			clroot group	pclcl	
			cladmin group	pcladm	
			clmon group	pclmon	
			sdxroot group	pclsdx	

<table>
<tr><td rowspan="14">Web-Based Admin View (Management view) For guest domain</td><td colspan="3">Operation mode</td><td>☐ 1 ■ 2 ☐ 3 layers</td><td></td></tr>
<tr><td rowspan="8">Server</td><td rowspan="4">Primary</td><td>Host name</td><td>guestl01</td><td></td></tr>
<tr><td>IP address</td><td>10.33.66.91</td><td></td></tr>
<tr><td>Subnet mask</td><td>255.255.255.0</td><td></td></tr>
<tr><td>URL</td><td>http://10.33.66.91:8081/Plugin.html</td><td></td></tr>
<tr><td rowspan="4">Secondary</td><td>Host name</td><td>guest02</td><td></td></tr>
<tr><td>IP address</td><td>10.33.66.92</td><td></td></tr>
<tr><td>Subnet mask</td><td>255.255.255.0</td><td></td></tr>
<tr><td>URL</td><td>http://10.33.66.92:8081/Plugin.html</td><td></td></tr>
<tr><td colspan="2" rowspan="5">User ID</td><td>wvroot group</td><td>pclwv</td><td></td></tr>
<tr><td>clroot group</td><td>pclcl</td><td></td></tr>
<tr><td>cladmin group</td><td>pcladm</td><td></td></tr>
<tr><td>clmon group</td><td>pclmon</td><td></td></tr>
<tr><td>sdxroot group</td><td>pclsdx</td><td></td></tr>
<tr><td rowspan="9">Cluster Foundation (CF)</td><td colspan="3">Cluster name</td><td>SLV002</td><td></td></tr>
<tr><td colspan="3">Cluster nodes</td><td>2</td><td></td></tr>
<tr><td colspan="3">Cluster node names</td><td>primecl01, primecl02</td><td></td></tr>
<tr><td colspan="3">Subnets</td><td>1</td><td></td></tr>
<tr><td colspan="3">Subnet number</td><td>192.168.1.0</td><td></td></tr>
<tr><td colspan="3">Subnet mask</td><td>255.255.255.0</td><td></td></tr>
<tr><td colspan="3">Use of CF remote service</td><td>■ cfcp (CF file copy)
■ cfsh (CF remote command execution)</td><td></td></tr>
<tr><td colspan="3">IP interconnect</td><td>◆ [no]
◇ yes (the number of IP interconnects)</td><td></td></tr>
<tr><td colspan="3">Use by RMS</td><td>◆ [yes] ◇ no</td><td></td></tr>
<tr><td rowspan="3">Cluster resource management facility (CRM)</td><td colspan="3" rowspan="3">Hardware devices stored in resource database</td><td>☐ Network unit</td><td>Required for using takeover of public LAN (IP) and node names.</td></tr>
<tr><td>☐ Line switching unit</td><td>Required for using line takeover function.</td></tr>
<tr><td>■ Disk unit</td><td>Required when using GDS and GFS.</td></tr>
</table>

Node 1	Machine model			SPARC M10-4	
	Device name			DBSV#001	
	Control domain	Domain ID / Domain name		—	
		Node name (uname -n)		primecl01	
		CF node name		primecl01	
		Interconnect	IP address	192.168.1.1	
			Subnet mask	255.255.255.0	
			Device name	e1000g3, e1000g4	
		Node in CF quorum set		◆[yes] ◇ No	Do not change the default setting.
		CF Custer timeout (CLUSTER_TIMEOUT)		"20"	Do not change the default setting. It must be "20."
		Use by RMS		◆ [yes] ◇ no Suffix : primecl01RMS	
		Weight		2	Node weight (priority for any partitioned clusters) used by SF.
		RCI	Address	—	Only when using RCI asynchronous monitoring.
		System console	IP address 1	10.33.66.71	XSCF, ALOM, and ILOM data
			Subnet mask 1	255.255.255.0	
			IP address 2	10.33.76.71	
			Subnet mask 2	255.255.255.0	
			User ID	root	
			User authority	—	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □ no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
		SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11 - sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43 - sd74	
		Management LAN	IP address	10.33.66.81	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Asynchronous monitoring sub-LAN	IP address	10.33.76.81	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN 1	IP address	10.34.214.181	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 2	IP address	10.34.215.21	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	
		Shutdown facility (SF)	Administrative IP addresses	10.33.66.81	XSCF, ALOM, and ILOM data
			Weight	2	
			Shutdown agent	■ XSCF (SPARC M10) □ XSCF (SPARC Enterprise M series) □ ILOM □ ALOM	
			User ID	xuser	
			User authority	platadm	
			Console LAN#1	10.33.66.71	
			Console LAN#2	10.33.76.71	
			PPAR-ID	0	
			Domain name	primary	
			Connection type	■ SSH □telnet	

		Domain ID / Domain name		domain1	
		Node name (uname -n)		guest01	
		CF node name		guest01	
		Interconnect	IP address	192.168.2.1	
			Subnet mask	255.255.255.0	
			Device name	e1000g3,e1000g4	
		Node in CF quorum set		◆[Yes] ◇ No	Do not change the default setting.
		CF Custer timeout (CLUSTER_TIMEOUT)		Not specified.	Do not change the default setting.
		Use by RMS		◆ [yes] ◇ no Suffix : guest01RMS	
		Weight		2	Node weight (priority for any partitioned clusters) used by SF.
		RCI	Address	–	Only when using RCI asynchronous monitoring.
		System console	IP address 1	10.33.66.71	XSCF, ALOM, and ILOM data
			Subnet mask 1	255.255.255.0	
			IP address 2	10.33.76.71	
			Subnet mask 2	255.255.255.0	
			User ID	root	
			User authority	–	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □ no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
	Guest domain	SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11～sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43～sd74	
		Management LAN	IP address	10.33.66.91	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Asynchronous monitoring sub-LAN	IP address	10.33.76.91	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN 1	IP address	10.34.224.181	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 2	IP address	10.34.225.21	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	
		Shutdown facility (SF)	Administrative IP addresses	10.33.66.91	
			Weight	2	
			Shutdown agent	■ XSCF (SPARC M10) □ XSCF (SPARC Enterprise M series) □ ILOM □ ALOM	
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.71	
			Console LAN#2	10.33.76.71	
			PPAR-ID	0	
			Domain name	domain1	
			Connection type	■ SSH □telnet	

Node 2	Machine model			SPARC M10-4	
	Device name			DBSV#002	
	Control domain	Domain ID / Domain name		—	
		Node name (uname -n)		primecl02	
		CF node name		primecl02	
		Interconnect	IP address	192.168.1.2	
			Subnet mask	255.255.255.0	
			Device name	e1000g3, e1000g4	
		Node in CF quorum set		◆[yes] ◇ No	Do not change the default setting.
		CF Custer timeout (CLUSTER_TIMEOUT)		"20"	Do not change the default setting.
		Use by RMS		◆ [yes] ◇ no Suffix : primecl02RMS	
		Weight		1	Node weight (priority for any partitioned clusters) used by SF.
		RCI	Address	—	Only when using RCI asynchronous monitoring.
		System console	IP address 1	10.33.66.72	XSCF, ALOM, and ILOM data
			Subnet mask 1	255.255.255.0	
			IP address 2	10.33.76.72	
			Subnet mask 2	255.255.255.0	
			User ID	root	
			User authority	—	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □ no	
		scsi-initiator-id (OBP variable #eeprom		8	
		SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11 - sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43 - sd74	
		Management LAN	IP address	10.33.66.82	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN 1	IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 2	IP address	10.34.215.22	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	
		Shutdown facility (SF)	Administrative IP addresses	10.33.66.82	XSCF, ALOM, and ILOM data
			Weight	1	
			Shutdown agent	■ XSCF (SPARC M10) □ XSCF (SPARC Enterprise M series) □ ILOM □ ALOM	
			User ID	xuser	
			User authority	platadm	
			Console LAN#1	10.33.66.72	
			Console LAN#2	10.33.76.72	
			PPAR-ID	1	
			Domain name	primary	
			Connection type	■ SSH □telnet	

		Domain ID / Domain name		—	
		Node name (uname -n)		primecl02	
		CF node name		primecl02	
		Interconnect	IP address	192.168.1.2	
			Subnet mask	255.255.255.0	
			Device name	e1000g3, e1000g4	
		Node in CF quorum set		◆[yes] ◇ No	Do not change the default setting.
		CF Cluster timeout (CLUSTER_TIMEOUT)		"20"	Do not change the default setting. It must be "20."
		Use by RMS		◆ [yes] ◇ no Suffix : primecl02RMS	
		Weight		1	Node weight (priority for any partitioned clusters) used by SF.
		RCI	Address	—	Only when using RCI asynchronous monitoring.
		System console	IP address 1	10.33.66.72	XSCF, ALOM, and ILOM data
	Guest domain		Subnet mask 1	255.255.255.0	
			IP address 2	10.33.76.72	
			Subnet mask 2	255.255.255.0	
			User ID	root	
			User authority	—	
			Connection type	■ SSH ☐telnet	
		Use of shared disk		■ yes ☐ no	
		scsi-initiator-id (OBP variable #eeprom		8	
		SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11 - sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43 - sd74	
		Management LAN	IP address	10.33.66.92	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN 1	IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 2	IP address	10.34.215.22	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	
		Shutdown facility (SF)	Administrative IP addresses	10.33.66.82	XSCF, ALOM, and ILOM data
			Weight	1	
			Shutdown agent	■ XSCF (SPARC M10) ☐ XSCF (SPARC Enterprise M series) ☐ ILOM ☐ ALOM	
			User ID	xuser	
			User authority	platadm	
			Console LAN#1	10.33.66.72	
			Console LAN#2	10.33.76.72	
			PPAR-ID	1	
			Domain name	primary	
			Connection type	■ SSH ☐telnet	

Edit the /etc/inet/ntp.conf file for every node as follows:

```
server 10.34.214.100
server 10.34.214.101
server 127.127.1.0

fudge 127.127.1.0 stratum 9

enable auth monitor
disable pll
driftfile /etc/ntp/ntp.drift
statsdir /var/ntp/ntpstats/
filegen peerstats file peerstats type day enable
```

```
filegen loopstats file loopstats type day enable
filegen clockstats file clockstats type day enable
```

For details on how to edit the /etc/inet/ntp.conf file, run the following command to find online reference on xntpd(1M):

```
# man xntpd
```

14.5.2 GLS Setup Worksheet

This worksheet serves to organize necessary input items beforehand for using GLS (Global Link Services) in Oracle VM Server for SPARC Environments.

This section presents a sample worksheet for the following redundancy method:

- NIC Switching (logical IP takeover in control domain, physical IP takeover in guest domains)

NIC Switching (Logical IP Takeover in Control Domain, Physical IP Takeover in Guest Domains)

Item				Setting
GLS settings	Takeover virtual interface name			sha0:65
		Takeover virtual IP address (or host name)		10.34.214.185
		Subnet mask		255.255.255.0
	Node name (1)			guest01
		Control domain configuration information	Virtual interface name	sha0
			Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
			Primary physical interface name	vsw0
			Secondary physical interface name	vsw1
			Physical IP address (or host name)	10.34.214.181
			Logical IP address (or host name)	10.34.214.185
		Guest domain configuration information	Virtual interface name	sha0
			Switching mode	◇NIC switching mode (logical IP takeover) ◆NIC switching mode (physical IP takeover)
			Primary physical interface name	vnet0
			Secondary physical interface name	vnet1
			Physical IP address (or host name)	10.34.214.186
		Monitoring destination information	IP address (or host name) of primary monitoring destination	10.34.214.188
			IP address (or host name) of secondary monitoring destination	10.34.214.189
		Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)			guest02
		Control domain configuration information	Virtual interface name	sha0
			Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
			Primary physical interface name	vsw0
			Secondary physical interface name	vsw1
			Physical IP address (or host name)	10.34.214.182
			Logical IP address (or host name)	10.34.214.185
		Guest domain configuration information	Virtual interface name	sha0
			Switching mode	◇NIC switching mode (logical IP takeover) ◆NIC switching mode (physical IP takeover)
			Primary physical interface name	vnet0
			Secondary physical interface name	vnet1
			Physical IP address (or host name)	10.34.214.186
		Monitoring destination information	IP address (or host name) of primary monitoring destination	10.34.214.188
			IP address (or host name) of secondary monitoring destination	10.34.214.189
		Optional function	Standby NIC patrol	◇Disable ◆Enable

14.5.3 GDS Configuration Worksheet

Below section presents a sample worksheet for GDS configuration.

Item					Setting
GDS configuration	Class 1	Class name			class0001
		Class scope (node name)	Node 1		guest01
			Node 2		guest02
		Spare disk 1	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
		Single disk 1	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
			Single volume 1	Volume name	
				Size	
				Virtual disk server device	
				Virtual disk	
			Single volume 2	Volume name	
				Size	
				Virtual disk server device	
				Virtual disk	
		Single disk 2	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
			Single volume 1	Volume name	
				Size	
				Virtual disk server device	
				Virtual disk	
			Single volume 2	Volume name	
				Size	
				Virtual disk server device	
				Virtual disk	
		Disk 1 to be connected to group	SDX disk name		diskmplb0001
			Physical disk name in node 1		mplb2048
			Physical disk name in node 2		mplb2048
		Disk2 to be connected to group	SDX disk name		diskmplb0002
			Physical disk name in node 1		mplb2049
			Physical disk name in node 2		mplb2049
		Highest-order group 1	Group name		group0001
			Group type		mirror
			Stripe width (*6)		
			Disk / low order group name	Disk / low-order group 1	
				Disk / low-order group 2	
			Volume 1	Volume name	volume0001
				Size	8380799
				Virtual disk server device	vol00@primery-vds0
				Virtual disk	vshdisk00
			Volume 2	Volume name	
				Size	
				Virtual disk server device	
				Virtual disk	

Chapter 15 Using PRIMECLUSTER in Oracle Solaris Kernel Zones Environment

This chapter explains how to use PRIMECLUSTER in Oracle Solaris Kernel Zones environment.

- Procedure for the configuration of PRIMECLUSTER in Oracle Solaris Kernel Zones environment
- Maintenance of cluster systems in Oracle Solaris Kernel Zones environment
- Collecting troubleshooting information in Oracle Solaris Kernel Zones environment
- Recommended configuration

See

For details on the Oracle Solaris Zones, see Oracle Solaris documents.

15.1 Procedure for the configuration of PRIMECLUSTER in Oracle Solaris Kernel Zones Environment

This section explains the procedure for building PRIMECLUSTER in Oracle Solaris Kernel Zones environment.

The procedure for building Kernel Zones in a physical environment is similar to the following cases only for the control domain in Oracle VM Server for SPARC environment:

- Cluster system between Kernel Zones within a same physical partition (Control domain)
- Cluster system between Kernel Zones among different physical partitions (Control domain)

15.1.1 Software Installation and Configuration of Cluster Environment

Install the Solaris software and the software required for PRIMECLUSTER in the control and guest domains of all nodes. Then perform the OS and hardware settings required for installation and operation.

For further details on configuring Oracle VM Server for SPARC or creating Oracle Solaris Kernel Zones, see "Creating and Using Oracle Solaris Kernel Zones."

The building procedure is different depending on the following cluster systems:

- Cluster system between Kernel Zones within a same physical partition (Control domain)
- Cluster system between Kernel Zones within a same physical partition (Guest domain)
- Cluster system between Kernel Zones among different physical partitions (Control domain)
- Cluster system between Kernel Zones among different physical partitions (Guest domain)

15.1.1.1 Cluster system between Kernel Zones within a same physical partition (Control domain/Guest domain)

This section explains how to build a cluster system between Kernel Zones within a same physical partition (Control domain/Guest domain).

Figure 15.1 Sample configuration of a cluster system between Kernel Zones within a same physical partition (Control domain)

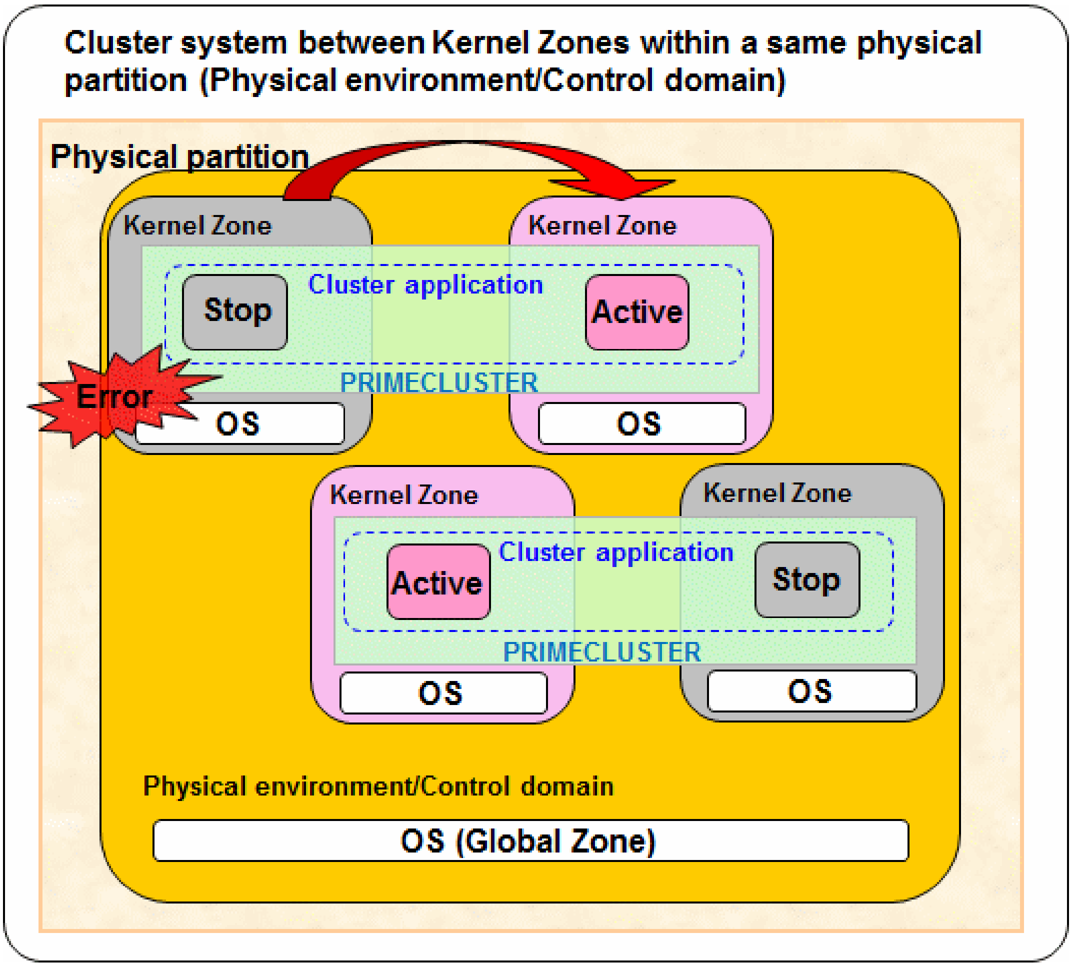


Figure 15.2 Sample configuration of a cluster system between Kernel Zones within a same physical partition (Guest domain)

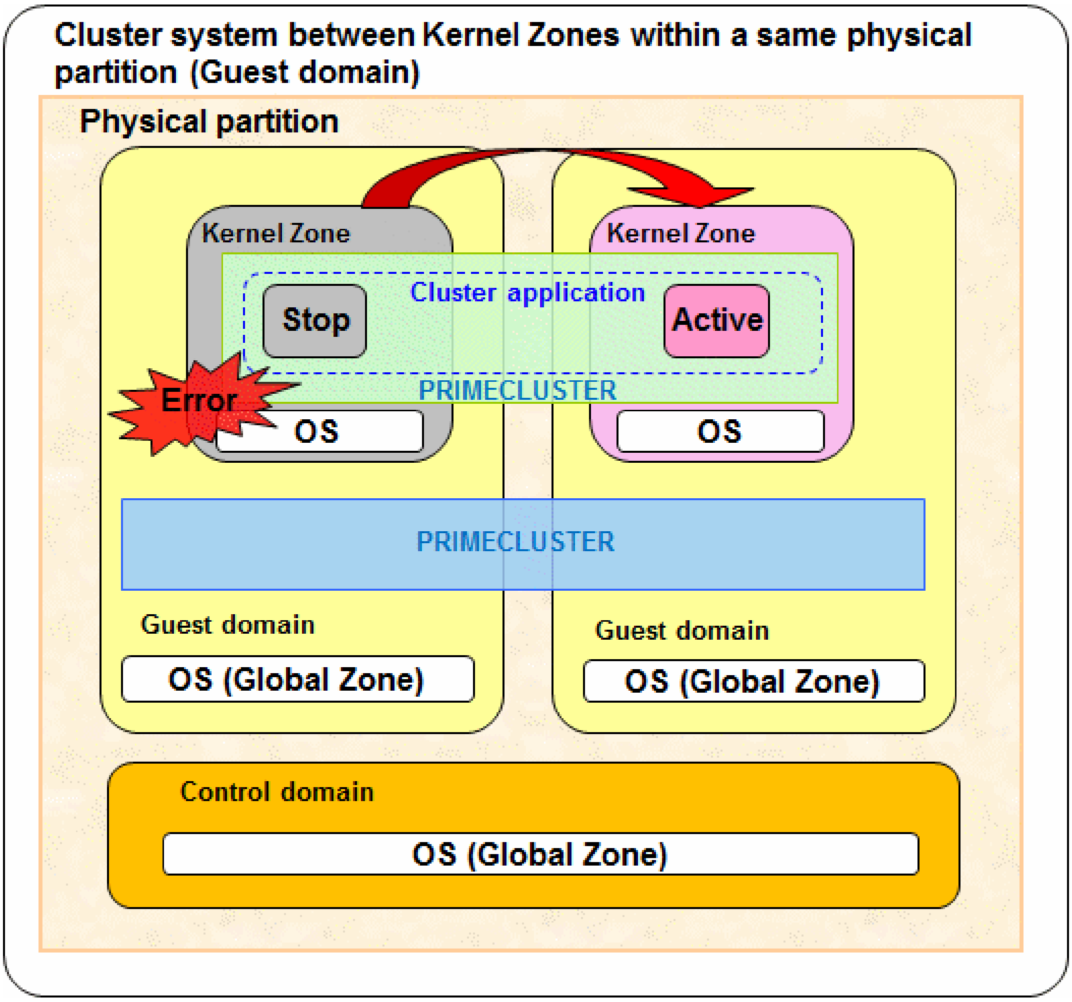
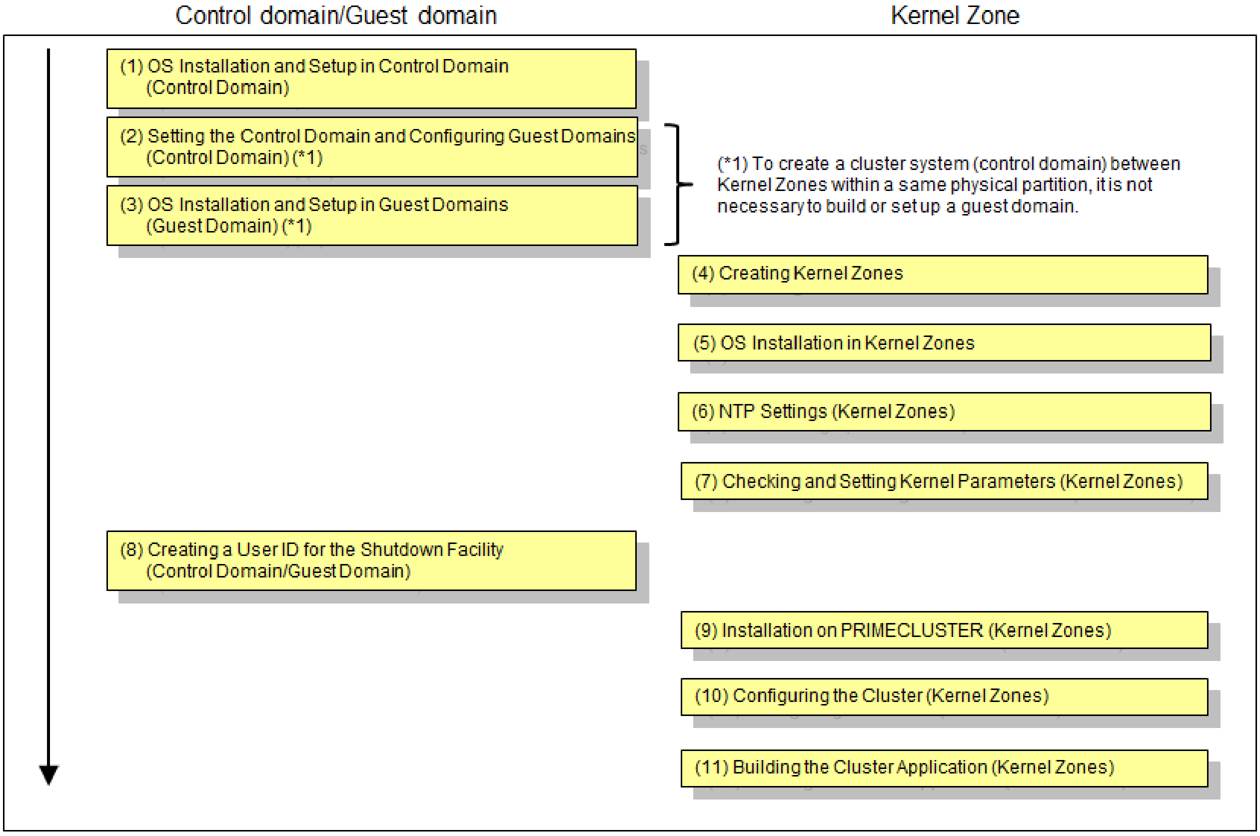


Figure 15.3 Building Procedure



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   If a guest domain is not built, the preparation for configuring an Oracle VM Server environment is unnecessary.

   Disk-Related Settings

   When using a shared disk unit, you need to install and set up the related software product.

   Carry out these settings in the control domain before installing PRIMECLUSTER.

   For details, see "3.2.2 Setting Up Disk Units."

   When building a cluster system between Kernel Zones within a same physical partition (Control domain), you must enable MPxIO to use a shared disk unit in Kernel Zones. Execute the following command on the control domain.

   Note that once this command is executed, the system is automatically restarted.

   - For the fibre channel connection

   ```
   # stmsboot -D fp -e
   ```

   - For the ISCSI connection

   ```
   # stmsboot -D iscsi -e
   ```

2. Setting the Control Domain and Configuring Guest Domains (Control domain)

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

Note

For "Cluster system between Kernel Zones within a same physical partition (Control domain)," it is not necessary to configure the guest domain.

Sample Configuration of Control and Guest Domains

As the virtual disk used as the system disk, specify a file in the format /LDoms/Vol1/vdisk0.img.
For the disk used as the switching disk in the cluster, specify a block device like /dev/dsk/cXtXdXs2. When doing so, specify a disk as the full disk without specifying the slice option.

Example: Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice /LDoms/Vol1/vdisk0.img vol1@primary-vds0
ldm add-vdiskserverdevice /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2 vol2@primary-vds0
```

Example: Configuration Script for Guest Domain

```
VDISK0=vol1@primary-vds0
VSHDISK0=vol2@primary-vds0
...
ldm add-vdisk vdisk0 $VDISK0 $DOMAIN
ldm add-vdisk timeout=360 vshdisk0 $VSHDISK0 $DOMAIN
```

Setting change of Virtual Network Device (vnet)

When using a Kernel Zone on the guest domain, you must assign the MAC address for the Kernel Zone to the virtual network device (vnet) that is used in the Kernel Zone. For the virtual network device (vnet) that is used as a public LAN, administrative LAN, or cluster interconnect in the Kernel Zone from the control domain, execute the following command.

- Setting the MAC address automatically

```
# ldm set-vnet alt-mac-addrs=auto <virtual network device name> <domain name>
```

- Setting the MAC address manually

```
# ldm set-vnet alt-mac-addrs=<MAC address> <virtual network device name> <domain name>
```

When creating multiple Kernel Zones on the guest domain, separate alt-mac-addrs by ", (comma)" to specify "auto" for the number of Kernel Zones or MAC addresses.

Example: using vnet0, vnet1, and vnet2 in three Kernel Zones on the guest1 domain (MAC address is automatically set)

```
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet0 guest1
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet1 guest1
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet2 guest1
```

3. OS Installation and Setup in Guest Domains (Guest Domain)

   Install the operating system in each guest domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete installation of operating systems in all guest domains.

   To use a shared disk unit in Kernel Zones, you must enable MPxIO. Execute the following command on the control domain.

   Note that once this command is executed, the system is automatically restarted.

   - For the fibre channel connection

```
# stmsboot -D fp -e
```

- For the ISCSI connection

```
# stmsboot -D iscsi -e
```

Note

For "Cluster system between Kernel Zones within a same physical partition (Control domain)," this procedure is not necessary.

4. Creating Kernel Zones

You must perform this section's procedure the same number of times as the number of Kernel Zones you need. For how to create Kernel Zones, see "Creating and Using Oracle Solaris Kernel Zones."

Using the zonecfg command, create Kernel Zones. Refer to the following example.

```
# zonecfg -z kzone01    *1
```

*1: "kzone01" is the Kernel Zone name (hereafter, the name is the same)

```
kzone01: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:kzone01> create -t SYSsolaris-kz
zonecfg kzone01> set autoboot=false
zonecfg:kzone01> info virtual-cpu
zonecfg:kzone01> add virtual-cpu
zonecfg:kzone01:virtual-cpu> set ncpus=2
zonecfg:kzone01:virtual-cpu> end
zonecfg:kzone01> info virtual-cpu
virtual-cpu:
        ncpus: 2
zonecfg:kzone01> select capped-memory
zonecfg:kzone01:capped-memory> set physical=4G
zonecfg:kzone01:capped-memory> end
```

```
[Deleting a default network]
zonecfg:kzone01> remove anet id=0
```

```
[Assigning a network]
zonecfg:kzone01> add net
zonecfg:kzone01:net> set physical=net1   *2
zonecfg:kzone01:net> end
```

*2: If the physical interface is multiplexed within the Kernel Zone, it is necessary to specify two or more physical interfaces.

```
[Using a disk device in the Kernel Zone]
zonecfg:kzone1> add device
zonecfg:kzone1:device> set match=/dev/rdsk/c9t0d0    *3
zonecfg:kzone1:device> end
```

*3: Enter the disk device added to the Kernel Zone using a full path.

```
zonecfg:kzone01> verify
zonecfg:kzone01> commit
zonecfg:kzone01> exit
```

See

For details, see the manual of the zonecfg command and also Oracle Solaris documents.

5. OS Installation in Kernel Zones

   Install the OS in Kernel Zones.

   Install Solaris in Kernel Zones using the zoneadm install command. Below is an example of installing the Kernel Zone.

```
# zoneadm -z kzone01 install
Preparing to install zone <kzone01>.
Creating list of files to copy from the global zone.
Copying <155078> files to the zone.
Initializing zone product registry.
Determining zone package initialization order.
Preparing to initialize <1282> packages on the zone.
Initialized <1282> packages on zone.
Zone <kzone01> is initialized.
Installation of <51> packages was skipped.
The file </zone-a-system/root/var/sadm/system/logs/install_log> contains a log of the zone
installation.
```

6. NTP Settings (Kernel Zones)

   This setup synchronizes the time in all of the nodes of the cluster system. Make sure to make this setting when configuring a cluster.

   Carry out these settings in Kernel Zones before installing PRIMECLUSTER.

7. Checking and Setting Kernel Parameters (Kernel Zones)

   When operating the software products related to PRIMECLUSTER, you need to adjust the OS kernel parameters according to the hardware configuration and applications.
   Carry out these settings in Kernel Zones before rebooting the system after "Installation on PRIMECLUSTER (Kernel Zones)".
   For details, see "3.2.3 Checking the Kernel Parameters."

8. Creating a User ID for the Shutdown Facility (Control Domain/Guest Domain)

   In Kernel Zones environment, you must set a user ID for the shutdown facility because PRIMECLUSTER has to control the Kernel Zone OS.

   Perform the following procedure on the host (Control domain/Guest domain) where Kernel Zones operates.

   Point

   The user ID to be created in this procedure is the user to login the global zone host (Control domain/Guest domain) for forcibly stopping a node by the shutdown facility.

   Use this user ID and password when setting the shutdown facility.

   1. Creating a General User ID (optional)

      Create a general user ID (optional) for the shutdown facility in the global zone host.

```
# useradd -m -d <Home directory> <User ID>
```

   2. Setting Up the sudo Command

      You must set up the sudo command so that the general user ID (optional) for the shutdown facility can execute the command as the root user.

      Use the visudo command to add the following setting so that the general user created in Step 1 can execute the command without entering the password.

      Since the shutdown facility uses the sudo command, a log of the sudo command is output to a syslog in 10 seconds.

      For outputting a log of the sudo command to an arbitrary file not via the syslog, add the setting together with allowing the command to be executed.

```
# visudo
```

Example of how to allow the command to be executed

```
<User ID>  ALL=(root) NOPASSWD:   ALL
```

Example of how to disable log output to a syslog

```
Defaults !syslog
```

Example of how to output a sudo log to an arbitrary file

```
Defaults logfile=<Log file full path name>
```

On all Kernel Zone, you must complete the user inquiry of the first SSH connection (RSA key generation) for all hosts where Kernel Zones operate. Follow the procedures detailed in "5.1.2.5.3 Logging in to Global Zone Host" to log in as the user for the shutdown facility.

9. Installation on PRIMECLUSTER (Kernel Zones)

   For details on installing PRIMECLUSTER, see "3.1 PRIMECLUSTER Installation."

10. Configuring the Cluster (Kernel Zones)

    Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," make the initial cluster setting in Kernel Zones.

    Note

    Do not change the timeout value on Kernel Zones from 10 seconds to a different value in CF settings.

11. Building the Cluster Application (Kernel Zones)

    Build the cluster application in Kernel Zones.

    Referring to "15.1.2 Building the cluster application (Control domain, Guest domain, and Kernel Zones)," build the cluster application in Kernel Zones.

15.1.1.2 Cluster system between Kernel Zones among different physical partitions (Control domain)

This section explains how to build a cluster system between Kernel Zones among different physical partitions (Control domain)

Figure 15.4 Sample configuration of a cluster system between Kernel Zones among different physical partitions (Control domain)

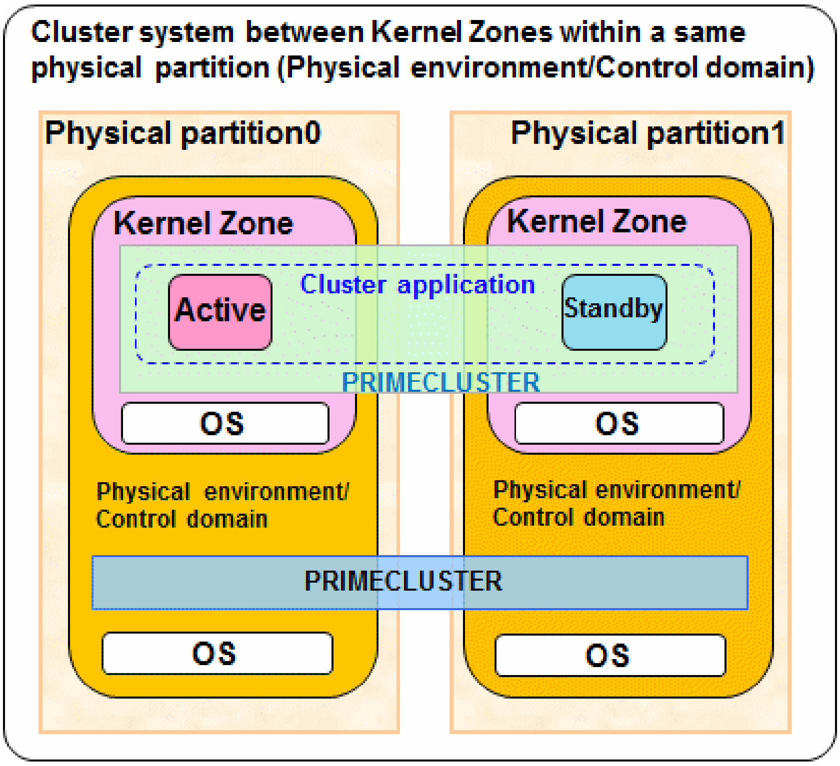
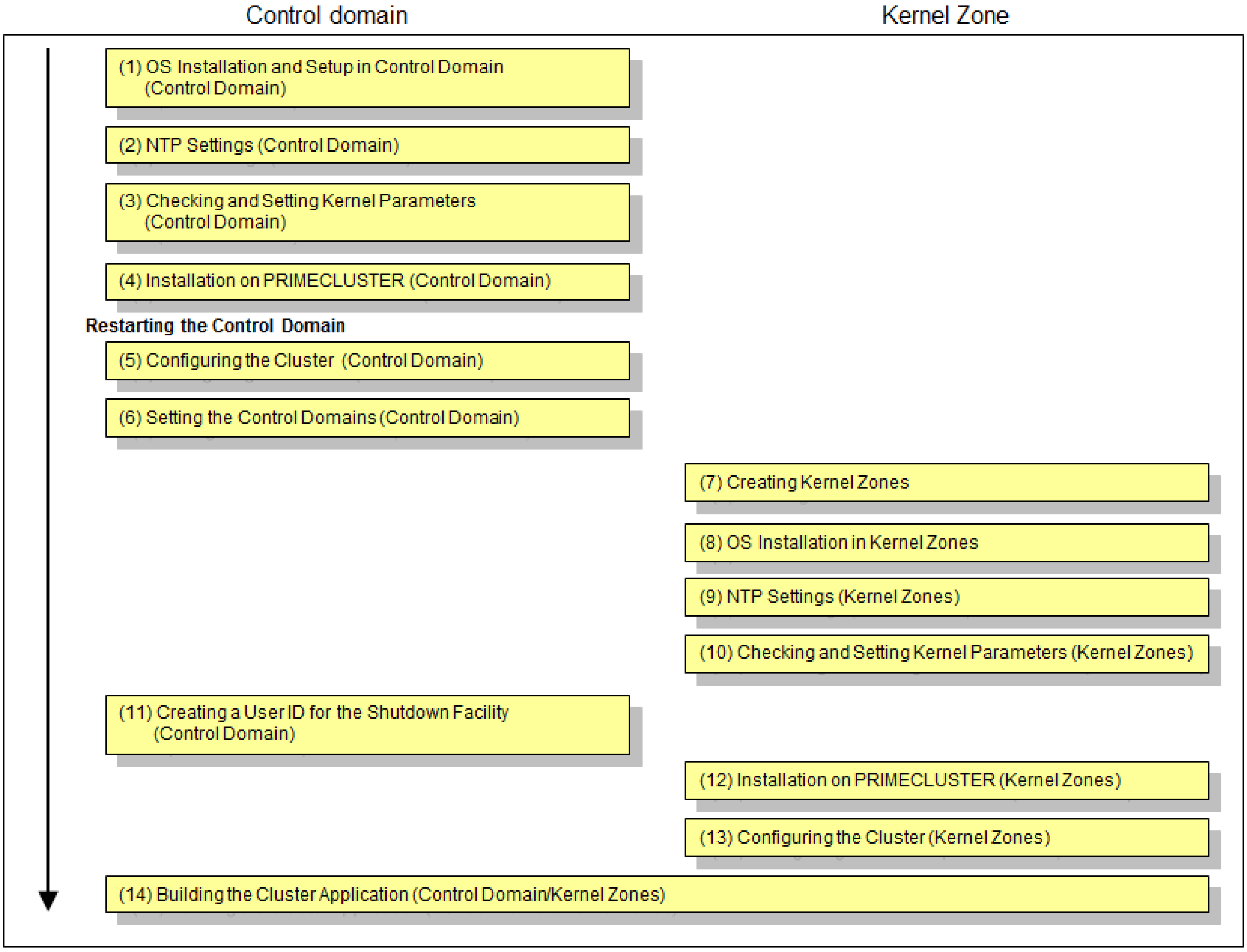


Figure 15.5 Building Procedure



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   If a guest domain is not built, the preparation for configuring an Oracle VM Server environment is unnecessary.

   Disk-Related Settings

   When using a shared disk unit, you need to install and set up the related software product.

   Carry out these settings in the control domain before installing PRIMECLUSTER.

   For details, see "3.2.2 Setting Up Disk Units."

   To use a shared disk unit in Kernel Zones, you must enable MPxIO. Execute the following command on the control domain.

   Note that once this command is executed, the system is automatically restarted.

   - For the fibre channel connection

   ```
   # stmsboot -D fp -e
   ```

   - For the ISCSI connection

   ```
   # stmsboot -D iscsi -e
   ```

2. NTP Settings (Control Domain)

   This setup synchronizes the time in all of the nodes of the cluster system. Make sure to make this setting when configuring a cluster.

Carry out these settings in the control domain before installing PRIMECLUSTER.

3. Checking and Setting Kernel Parameters (Control Domain)

   When operating the software products related to PRIMECLUSTER, you need to adjust kernel parameters according to the hardware configuration and applications.

   Carry out these settings in the control domain before rebooting the system after "Step 4. Installation on PRIMECLUSTER (Control Domain)".

   For details, see "3.2.3 Checking the Kernel Parameters."

4. Installation on PRIMECLUSTER (Control Domain)

   For details on installing PRIMECLUSTER, see "3.1 PRIMECLUSTER Installation."

5. Configuring the Cluster (Control Domain)

   Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," make the initial cluster settings in the control domain. To build clusters, perform procedures from "5.1.1 Setting Up CF and CIP" through "5.1.2 Configuring the Shutdown Facility" and also perform "5.2 Setting Up Power Supply Linkage" as needed.

   Note

   - Set the timeout value of the cluster system on the control domain in CF settings as follows.

     Timeout value: 20 seconds (Change the timeout value by using the cfset command.)

   - When a failure of the control domain (including the cluster application error) is detected and the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

   - When a virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.

6. Setting the Control Domains (Control Domain)

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

7. Creating Kernel Zones

   Create Kernel Zones. You must perform this section's procedure the same number of times as the number of Kernel Zones you need. For how to create Kernel Zones, see "Creating and Using Oracle Solaris Kernel Zones."

   Using the zonecfg command, create Kernel Zones. Refer to the following example.

```
# zonecfg -z kzone01    *1
```

   *1: "kzone01" is the Kernel Zone name (hereafter, the name is the same)

```
kzone01: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:kzone01> create -t SYSsolaris-kz
zonecfg kzone01> set autoboot=false
zonecfg:kzone01> info virtual-cpu
zonecfg:kzone01> add virtual-cpu
zonecfg:kzone01:virtual-cpu> set ncpus=2
zonecfg:kzone01:virtual-cpu> end
zonecfg:kzone01> info virtual-cpu
virtual-cpu:
        ncpus: 2
zonecfg:kzone01> select capped-memory
zonecfg:kzone01:capped-memory> set physical=4G
zonecfg:kzone01:capped-memory> end
```

```
[Deleting a default network]
zonecfg:kzone01> remove anet id=0
```

```
[Assigning a network]
zonecfg:kzone01> add net
zonecfg:kzone01:net> set physical=net1   *2
zonecfg:kzone01:net> end
```

*2: If the physical interface is multiplexed within the Kernel Zone, it is necessary to specify two or more physical interfaces.

```
[Using a disk device in the Kernel Zone]
zonecfg:kzone1> add device
zonecfg:kzone1:device> set match=/dev/rdsk/c9t0d0    *3
zonecfg:kzone1:device> end
```

*3: Enter the disk device added to the Kernel Zone using a full path.

```
zonecfg:kzone01> verify
zonecfg:kzone01> commit
zonecfg:kzone01> exit
```

See

For details, see the manual of the zonecfg command and also Oracle Solaris documents.

8. OS Installation in Kernel Zones

   Install the OS in Kernel Zones.

   Install Solaris in Kernel Zones using the zoneadm install command. Below is an example of installing the Kernel Zone.

```
# zoneadm -z kzone01 install
Preparing to install zone <kzone01>.
Creating list of files to copy from the global zone.
Copying <155078> files to the zone.
Initializing zone product registry.
Determining zone package initialization order.
Preparing to initialize <1282> packages on the zone.
Initialized <1282> packages on zone.
Zone <kzone01> is initialized.
Installation of <51> packages was skipped.
The file </zone-a-system/root/var/sadm/system/logs/install_log> contains a log of the zone
installation.
```

9. NTP Settings (Kernel Zones)

   This setup synchronizes the time in all of the nodes of the cluster system. Make sure to make this setting when configuring a cluster.

   Carry out these settings in Kernel Zones before installing PRIMECLUSTER.

10. Checking and Setting Kernel Parameters (Kernel Zones)

    When operating the software products related to PRIMECLUSTER, you need to adjust the OS kernel parameters according to the hardware configuration and applications.
    Carry out these settings in Kernel Zones before rebooting the system after "Installation on PRIMECLUSTER (Kernel Zones)".
    For details, see "3.2.3 Checking the Kernel Parameters."

11. Creating a User ID for the Shutdown Facility (Control Domain)

    In Kernel Zones environment, you must set a user ID for the shutdown facility because PRIMECLUSTER has to control the Kernel Zone OS.

    Perform the following procedure on the host (Control domain) where Kernel Zones operates.

Point

The user ID to be created in this procedure is the user to login the global zone host (Control domain) for forcibly stopping a node by the shutdown facility.

Use this user ID and password when setting the shutdown facility.

1. Creating a General User ID (optional)

   Create a general user ID (optional) for the shutdown facility in the global zone host.

   ```
   # useradd -m -d <Home directory> <User ID>
   ```

2. Setting Up the sudo Command

   You must set up the sudo command so that the general user ID (optional) for the shutdown facility can execute the command as the root user.

   Use the visudo command to add the following setting so that the general user created in Step 1 can execute the command without entering the password.

   Since the shutdown facility uses the sudo command, a log of the sudo command is output to a syslog in 10 seconds.

   For outputting a log of the sudo command to an arbitrary file not via the syslog, add the setting together with allowing the command to be executed.

   ```
   # visudo
   ```

   Example of how to allow the command to be executed

   ```
   <User ID>  ALL=(root) NOPASSWD:   ALL
   ```

   Example of how to disable log output to a syslog

   ```
   Defaults !syslog
   ```

   Example of how to output a sudo log to an arbitrary file

   ```
   Defaults logfile=<Log file full path name>
   ```

On all Kernel Zones, you must complete the user inquiry of the first SSH connection (RSA key generation) for all hosts where Kernel Zones operates.. Follow the procedures detailed in "5.1.2.5.3 Logging in to Global Zone Host" to log in as the user for the shutdown facility.

12. Installation on PRIMECLUSTER (Kernel Zones)

    For details on installing PRIMECLUSTER, see "3.1 PRIMECLUSTER Installation."

13. Configuring the Cluster (Kernel Zones)

    Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," make the initial cluster setting in Kernel Zones.

Note

Do not change the timeout value on Kernel Zones from 10 seconds to a different value in CF settings.

14. Building the Cluster Application (Control domain/Kernel Zones)

    Build the cluster application in the control domain and Kernel Zones.

    Referring to "15.1.2 Building the cluster application (Control domain, Guest domain, and Kernel Zones)," build the cluster application in the control domain and Kernel Zones.

15.1.1.3 Cluster system between Kernel Zones among different physical partitions (Guest domain)

This section explains how to build a cluster system between Kernel Zones among different physical partitions (Guest domain).

Figure 15.6 Sample configuration of a cluster system between Kernel Zones among different physical partitions (Guest domain)

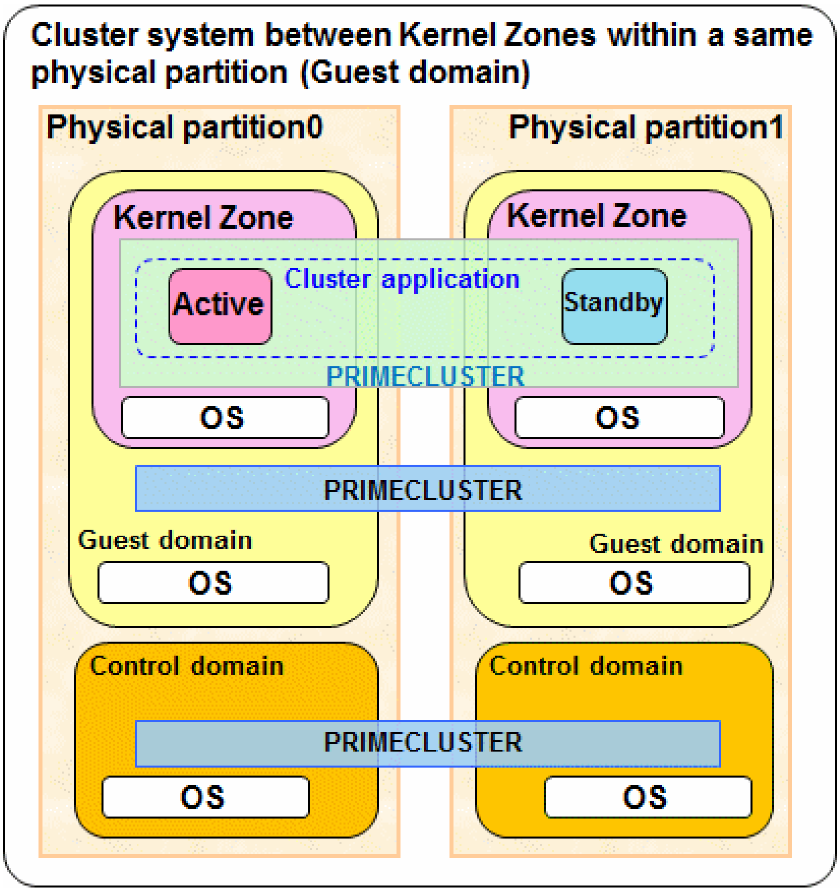
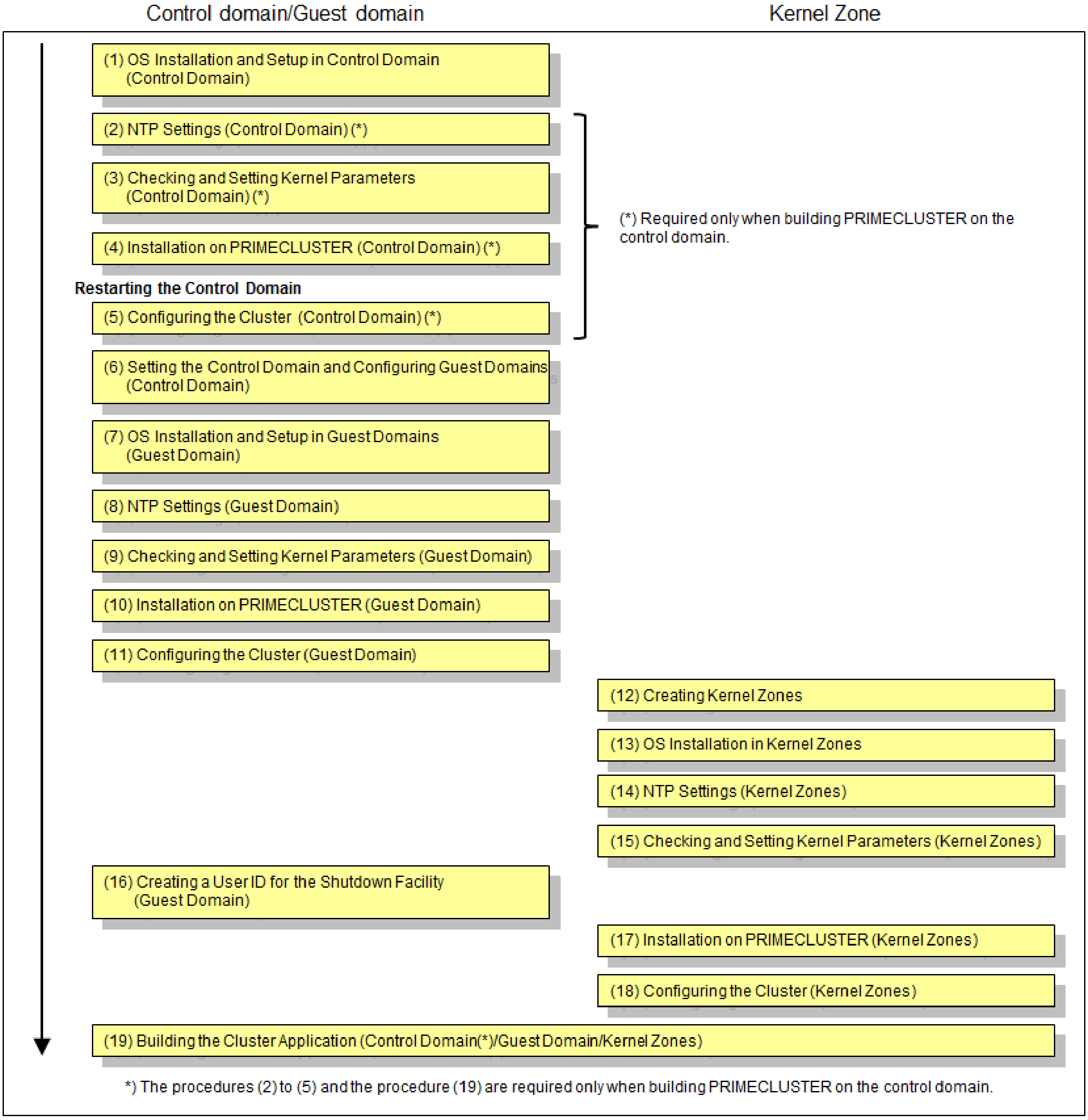


Figure 15.7 Building Procedure



1. OS Installation and Setup in Control Domain (Control Domain)

   Complete all preparations for configuring your Oracle VM Server for SPARC environment by installing the operating system and the Oracle VM Server for SPARC package in the control domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide" for the details, complete all preparations through to the files for the virtual disk server devices.

   If a guest domain is not built, the preparation for configuring an Oracle VM Server environment is unnecessary.

   Disk-Related Settings

   When using a shared disk unit, you need to install and set up the related software product.

   Carry out these settings in the control domain before installing PRIMECLUSTER.

   For details, see "3.2.2 Setting Up Disk Units."

2. NTP Settings (Control Domain)

   This setup synchronizes the time in all of the nodes of the cluster system. Make sure to make this setting when configuring a cluster.

   Carry out these settings in the control domain before installing PRIMECLUSTER.

3. Checking and Setting Kernel Parameters (Control Domain)

   When operating the software products related to PRIMECLUSTER, you need to adjust kernel parameters according to the hardware configuration and applications.

   Carry out these settings in the control domain before rebooting the system after "Step 4. Installation on PRIMECLUSTER (Control Domain)".

   For details, see "3.2.3 Checking the Kernel Parameters."

4. Installation on PRIMECLUSTER (Control Domain)

   For details on installing PRIMECLUSTER, see "3.1 PRIMECLUSTER Installation."

5. Configuring the Cluster (Control Domain)

   Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," make the initial cluster settings in the control domain. To build clusters, perform procedures from "5.1.1 Setting Up CF and CIP" through "5.1.2 Configuring the Shutdown Facility" and also perform "5.2 Setting Up Power Supply Linkage" as needed.

   Note

   - Set the timeout value of the cluster system on the control domain in CF settings as follows.

     Timeout value: 20 seconds (Change the timeout value by using the cfset command.)

   - When a failure of the control domain (including the cluster application error) is detected and the control domain cannot be forcibly stopped, all the guest domains or all the I/O domains within the failed physical partition are stopped regardless of whether a cluster system exists. This is because of stopping the physical partition forcibly.

   - When a virtual I/O is set on the control domain, the guest domain within the failed physical partition may be stopped regardless of whether a cluster system exists.

6. Setting the Control Domain and Configuring Guest Domains

   Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete all configurations for the control and guest domains.

   Sample Configuration of Control and Guest Domains

   As the virtual disk used as the system disk, specify a file in the format /LDoms/Vol1/vdisk0.img.
   For the disk used as the switching disk in the cluster, specify a block device like /dev/dsk/cXtXdXs2. When doing so, specify a disk as the full disk without specifying the slice option.

   Example: Configuration Script for Control Domain

```
# Define the virtual disk server.
ldm add-vdiskserver primary-vds0 primary
# Define the virtual disk server device.
ldm add-vdiskserverdevice /LDoms/Vol1/vdisk0.img vol1@primary-vds0
ldm add-vdiskserverdevice /dev/dsk/c0t6000B5D0006A0000006A0FB800130000d0s2 vol2@primary-vds0
```

   Example: Configuration Script for Guest Domain

```
VDISK0=vol1@primary-vds0
VSHDISK0=vol2@primary-vds0
...
ldm add-vdisk vdisk0 $VDISK0 $DOMAIN
ldm add-vdisk timeout=360 vshdisk0 $VSHDISK0 $DOMAIN
```

   Setting change of Virtual Network Device (vnet)

When using a Kernel Zone on the guest domain, you must assign the MAC address for the Kernel Zone to the virtual network device (vnet) that is used in the Kernel Zone. For the virtual network device (vnet) that is used as a public LAN, administrative LAN, or cluster interconnect in the Kernel Zone from the control domain, execute the following command.

- Setting the MAC address automatically

```
# ldm set-vnet alt-mac-addrs=auto <virtual network device name> <domain name>
```

- Setting the MAC address manually

```
# ldm set-vnet alt-mac-addrs=<MAC address> <virtual network device name> <domain name>
```

When creating multiple Kernel Zones on the guest domain, separate alt-mac-addrs by ", (comma)" to specify "auto" or MAC addresses for the number of Kernel Zones.

Example: using vnet0, vnet1, and vnet2 in three Kernel Zones on the guest1 domain (MAC address is automatically set)

```
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet0 guest1
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet1 guest1
# ldm set-vnet alt-mac-addrs=auto,auto,auto vnet2 guest1
```

7. OS Installation and Setup in Guest Domains (Guest Domain)

   Install the operating system in each guest domain. Referring to "Oracle VM Server for SPARC Guide" or "SPARC M10 Systems Domain Configuration Guide," complete installation of operating systems in all guest domains.

   To use a shared disk unit in Kernel Zones, you must enable MPxIO. Execute the following command on the control domain.

   Note that once this command is executed, the system is automatically restarted.

   - For the fibre channel connection

   ```
   # stmsboot -D fp -e
   ```

   - For the ISCSI connection

   ```
   # stmsboot -D iscsi -e
   ```

8. NTP Settings (Guest Domain)

   This setup serves to synchronize the clocks of every node in the cluster system. Make sure to make this setting when configuring a cluster.
   Carry out these settings in the guest domain before installing PRIMECLUSTER.

9. Checking and Setting Kernel Parameters (Guest Domain)

   For running PRIMECLUSTER-related software, the OS kernel parameters need to be adjusted to the environment.
   Carry out these settings in the guest domain before rebooting the system after "Step 10. Installation on PRIMECLUSTER (Guest Domain)".
   For details, refer to "3.2.3 Checking the Kernel Parameters".

10. Installation on PRIMECLUSTER (Guest Domain)

    For details on installing PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation."

11. Configuring the Cluster (Guest Domain)

    Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster", make the initial cluster settings in the guest domain. To build clusters, perform procedures from "5.1.1 Setting Up CF and CIP" through "5.1.2 Configuring the Shutdown Facility" and also perform "5.2 Setting Up Power Supply Linkage" as needed.

    Note

    When stopping a guest domain forcibly after detecting an error (including the cluster application error) on the guest domain, stop all Kernel Zones within the failed guest domain regardless of whether a cluster system exists.

12. Creating Kernel Zones

Create Kernel Zones. You must perform this section's procedure the same number of times as the number of Kernel Zones you need. For how to create Kernel Zones, see "Creating and Using Oracle Solaris Kernel Zones."

Using the zonecfg command, create Kernel Zones. Refer to the following example.

```
# zonecfg -z kzone01    *1
```

*1: "kzone01" is the Kernel Zone name (hereafter, the name is the same)

```
kzone01: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:kzone01> create -t SYSsolaris-kz
zonecfg kzone01> set autoboot=false
zonecfg:kzone01> info virtual-cpu
zonecfg:kzone01> add virtual-cpu
zonecfg:kzone01:virtual-cpu> set ncpus=2
zonecfg:kzone01:virtual-cpu> end
zonecfg:kzone01> info virtual-cpu
virtual-cpu:
        ncpus: 2
zonecfg:kzone01> select capped-memory
zonecfg:kzone01:capped-memory> set physical=4G
zonecfg:kzone01:capped-memory> end
```

```
[Deleting a default network]
zonecfg:kzone01> remove anet id=0
```

```
[Assigning a network]
zonecfg:kzone01> add net
zonecfg:kzone01:net> set physical=net1   *2
zonecfg:kzone01:net> end
```

*2: If the physical interface is multiplexed within the Kernel Zone, it is necessary to specify two or more physical interfaces.

```
[Using a disk device in the Kernel Zone]
zonecfg:kzone1> add device
zonecfg:kzone1:device> set match=/dev/rdsk/c9t0d0    *3
zonecfg:kzone1:device> end
```

*3: Enter the disk device added to the Kernel Zone using a full path.

```
zonecfg:kzone01> verify
zonecfg:kzone01> commit
zonecfg:kzone01> exit
```

See

For details, see the manual of the zonecfg command and also Oracle Solaris documents.

13. OS Installation in Kernel Zones

Install the OS in Kernel Zones.

Install Solaris in Kernel Zones using the zoneadm install command. Below is an example of installing the Kernel Zone.

```
# zoneadm -z kzone01 install
Preparing to install zone <kzone01>.
Creating list of files to copy from the global zone.
Copying <155078> files to the zone.
Initializing zone product registry.
```

```
Determining zone package initialization order.
Preparing to initialize <1282> packages on the zone.
Initialized <1282> packages on zone.
Zone <kzone01> is initialized.
Installation of <51> packages was skipped.
The file </zone-a-system/root/var/sadm/system/logs/install_log> contains a log of the zone
installation.
```

14. NTP Settings (Kernel Zones)

    This setup synchronizes the time in all of the nodes of the cluster system. Make sure to make this setting when configuring a cluster.

    Carry out these settings in Kernel Zones before installing PRIMECLUSTER.

15. Checking and Setting Kernel Parameters (Kernel Zones)

    When operating the software products related to PRIMECLUSTER, you need to adjust the OS kernel parameters according to the hardware configuration and applications.
    Carry out these settings in Kernel Zones before rebooting the system after "Installation on PRIMECLUSTER (Kernel Zones)".
    For details, see "3.2.3 Checking the Kernel Parameters."

16. Creating a User ID for the Shutdown Facility (Guest Domain)

    In Kernel Zones environment, you must set a user ID for the shutdown facility because PRIMECLUSTER has to control the Kernel Zone OS.

    Perform the following procedure on the host (Guest domain) where Kernel Zones operates.

    Point

    The user ID to be created in this procedure is the user to login the global zone host (Guest domain) for forcibly stopping a node by the shutdown facility.

    Use this user ID and password when setting the shutdown facility.

    1. Creating a General User ID (optional)

       Create a general user ID (optional) for the shutdown facility in the global zone host.

       ```
       # useradd -m -d <Home directory> <User ID>
       ```

    2. Setting Up the sudo Command

       You must set up the sudo command so that the general user ID (optional) for the shutdown facility can execute the command as the root user.

       Use the visudo command to add the following setting so that the general user created in Step 1 can execute the command without entering the password.

       Since the shutdown facility uses the sudo command, a log of the sudo command is output to a syslog in 10 seconds.

       For outputting a log of the sudo command to an arbitrary file not via the syslog, add the setting together with allowing the command to be executed.

       ```
       # visudo
       ```

       Example of how to allow the command to be executed

       ```
       <User ID>  ALL=(root) NOPASSWD:   ALL
       ```

       Example of how to disable log output to a syslog

       ```
       Defaults !syslog
       ```

       Example of how to output a sudo log to an arbitrary file

       ```
       Defaults logfile=<Log file full path name>
       ```

On all Kernel Zones, you must complete the user inquiry of the first SSH connection

(RSA key generation) for all hosts where Kernel Zones operates. Follow the procedures detailed in "5.1.2.5.3 Logging in to Global Zone Host" to log in as the user for the shutdown facility.

17. Installation on PRIMECLUSTER (Kernel Zones)

    For details on installing PRIMECLUSTER "3.1 PRIMECLUSTER Installation."

18. Configuring the Cluster (Kernel Zones)

    Referring to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," make the initial cluster setting in Kernel Zones.

    Note

    Do not change the timeout value on Kernel Zones from 10 seconds to a different value in CF settings.

19. Building the Cluster Application (Control Domain/Guest Domain/Kernel Zones)

    Build the cluster application in the control domain, guest domain, and Kernel Zones.

    Referring to "15.1.2 Building the cluster application (Control domain, Guest domain, and Kernel Zones)," build the cluster application in the control domain, guest domain, and Kernel Zones.

15.1.2 Building the cluster application (Control domain, Guest domain, and Kernel Zones)

For details on how to build the cluster application on the guest domain, see "Chapter 6 Building Cluster Applications."

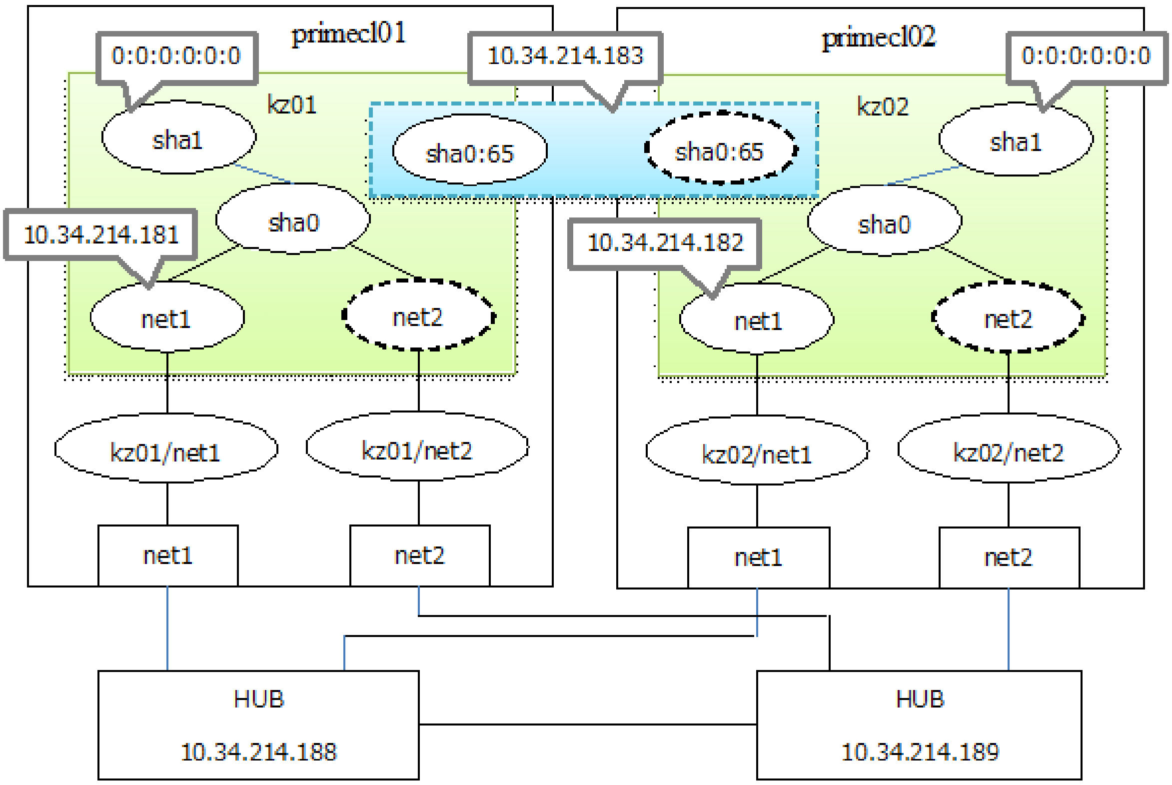
When building the cluster application in the environment where GLS is used in Kernel Zones, you must build it with the procedure described in "15.1.2.1 Configuring GLS (Kernel Zones)" instead of "6.2.1 GLS Setup."

15.1.2.1 Configuring GLS (Kernel Zones)

This section uses below example to explain the setup procedure for GLS (redundant line control function) for Kernel Zone clusters (1:1 operational/standby) of the NIC switching mode (Ipv4) in the Kernel Zones environment.

For information on the initial GLS setup, refer to "Chapter 5 Operation on Cluster System" in the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

Figure 15.8 Sample GLS Configuration



In this configuration, a cluster is configured using the Kernel Zone kz01 on primecl01 as an operational node and using the Kernel Zone kz02 on primecl02 as a standby node. When user LAN transmissions on the operational node are disabled, GLS in Kernel Zones detects an error in the transmission route and switches clusters.

15.1.2.1.1 Network Setup in a Global Zone

primecl01

Set up the network setting of kz01 with the zonecfg command.

Below is an example when adding a new anet resource used in the virtual interface of the NIC switching mode to Kernel Zones.

```
# zonecfg -z kz01
zonecfg:kz01> add anet  <- Create the anet interface.
zonecfg:kz01:anet> set lower-link=net1  <- Specify the physical NIC net1 to connect.
zonecfg:kz01:anet> end
zonecfg:kz01> add anet  <- Create the anet interface.
zonecfg:kz01:anet> set lower-link=net2  <- Specify the physical NIC nets to connect.
zonecfg:kz01:anet> end
zonecfg:kz01> export <- Check the setting.
create -b
set autoboot=false
set autoshutdown=shutdown
set hostid=0x48bfa8db
add anet
set lower-link=auto
set configure-allowed-address=true
set link-protection=mac-nospoof
set mac-address=auto
set id=0
end
add anet
set lower-link=net1
set configure-allowed-address=true
```

```
set link-protection=mac-nospoof
set mac-address=auto
set id=1
end
add anet
set lower-link=net2
set configure-allowed-address=true
set link-protection=mac-nospoof
set mac-address=auto
set id=2
end
add device
set storage=dev:/dev/zvol/dsk/%{global-rootzpool}/VARSHARE/zones/%{zonename}/disk%{id}
set bootpri=0
set id=0
end
add capped-memory
set physical=2G
end
add keysource
set raw="{base64}qj3mgMZcH8ssmk7WYB7TjA=="
end
zonecfg:kz01> commit  <- Confirm the zone configuration.
zonecfg:kz01> exit    <- Zone setting is completed.
```

primecl02

Set up the network setting of kz02 with the zonecfg command.

The setting procedure is the same as primecl01

15.1.2.1.2 GLS Setup in Kernel Zones

Operational node [kz01]

1. Setting up the system.

    1. Define the IP address and Host name in /etc/inet/hosts file.

    ```
    10.34.214.183   takeoverIP   # Virtual IP
    10.34.214.181   kz01         # Physical IP of kz01
    10.34.214.182   kz02         # Physical IP of kz02
    10.34.214.188   swhub1       # HUB IP for the primary monitoring destination
    10.34.214.189   swhub2       # HUB IP for the secondary monitoring destination
    ```

    2. Define the subnet masks in the /etc/inet/netmasks file.

    ```
    10.34.214.0   255.255.255.0
    ```

    3. Setting up interface to use.

    ```
    # /usr/sbin/ipadm create-ip net1
    # /usr/sbin/ipadm create-addr -T static -a kz01/24 net1/v4
    ```

2. Reboot the zone.

    Execute the following command to reboot the zone. After rebooting the zone, verify the interface set by using ifconfig command is enabled.

    ```
    # /usr/sbin/shutdown -y -i6 -g0
    ```

3. Creating of virtual interface.

    For the underlined parameter, specify the specific IP address of the node.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.32.214.183 -e 10.34.214.181 -t
net1,net2
```

Check that the resources have been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname       Mode MAC Adder/Phys ip  Interface List
+----------+--------------+----+-----------------+------------------+
 sha0        10.34.214.183  d    10.34.214.181       net1,net2

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+----------+---------------------------------+----+------------------------+
```

4. Setting up the HUB monitoring function.

   For the underlined parameter, specify the IP addresses of the hubs to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+--------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
         hub-hub poll       =    OFF
         interval(idle)     =      5( 60) sec
         times              =      5 times
         max_retry          =      5 retry
         repair_time        =      5 sec
         link detection     =    NO
         failover mode      =    YES
```

5. Make the settings for the GLS standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name         Hostname       Mode MAC Adder/Phys ip Interface List
+-----------+--------------+----+-----------------+------------------+
 sha0         10.34.214.183   d    10.34.214.181     net1,net2
 sha1         -               p    00:00:00:00:00:00 sha0

[IPv6]
```

```
 Name         Hostname/prefix                   Mode Interface List
+-----------+--------------------------------+----+------------------------+
```

6. Creating the takeover virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+---------------+--------------------------------------------------+
 sha0:65    10.34.214.183    -
```

Standby node [kz02]

1. Setting up the system.
    1. Define the IP address and Host name in /etc/inet/hosts file.

       The definition content is the same as kz01.
    2. Define the subnet masks in the /etc/inet/netmasks file.

       The definition content is the same as kz01.
    3. Setting up interface to use.

       ```
       # /usr/sbin/ipadm create-ip net1
       # /usr/sbin/ipadm create-addr -T static -a kz02/24 net1/v4
       ```
2. Reboot the zone.

   Execute the following command to reboot the zone. After rebooting the zone, verify the interface set by using ifconfig command is enabled.

```
# /usr/sbin/shutdown -y -i6 -g0
```

3. Creating of virtual interface.

   For the underlined parameter, specify the specific IP address of the node.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.32.214.183 -e 10.34.214.182 -t
net1,net2
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.183    d   10.34.214.182      net1,net2

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+--------------------------------+----+------------------------+
```

4. Setting up the HUB monitoring function.

   For the underlined parameter, specify the IP addresses of the hubs to be monitored.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.34.214.188,10.34.214.189 -b off
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll print
[ Standard Polling Parameter ]
         interval(idle)      =      5( 60) sec
         times               =      5 times
         max_retry           =      5 retry
         repair_time         =      5 sec
         link detection      =    NO
         failover mode       =    YES

[ Polling Parameter of each interface ]
 Name    Hostname/Polling Parameter
+-------+---------------------------------------------------------------+
 sha0    10.34.214.188,10.34.214.189
         hub-hub poll        =    OFF
         interval(idle)      =      5( 60) sec
         times               =      5 times
         max_retry           =      5 retry
         repair_time         =      5 sec
         link detection      =    NO
         failover mode       =    YES
```

5. Make the settings for the GLS standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+-------------------+
 sha0        10.34.214.183    d   10.34.214.181      net1,net2
 sha1        -                p   00:00:00:00:00:00  sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+-------------------------+
```

6. Creating the takeover virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

After executing above command, check that the settings are correct.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+---------------------------------------------------+
 sha0:65    10.34.214.183    -
```

15.2 Maintenance of cluster systems in Oracle Solaris Kernel Zones Environment

This section describes maintenance methods for the cluster systems in Oracle Solaris Kernel Zones environment.

Carry out maintenance of the cluster system of Oracle Solaris Kernel Zones environment on all domains that constitute the cluster system in the same way as for normal cluster systems.

For further details, see "Chapter 12 Maintenance of the PRIMECLUSTER System."

15.3 Collecting Troubleshooting Information in Oracle Solaris Kernel Zones Environment

After any trouble in the PRIMECLUSTER system, collect the following data, necessary for diagnosis, from all physical environments, control domains, guest domains, Kernel Zones, and the cluster management server that constitute a cluster system.

After that, contact our customer support.

- Physical Environment/Control Domains

  Referring to "C.1 Collecting Troubleshooting Information," collect following data.

  - PRIMECLUSTER diagnostic data
  - If the malfunction is reproducible, documentation describing procedures for reproduction

- Kernel Zones/Guest Domains

  Run fjsnap or FJQSS to collect the data needed for error diagnosis.

  For details on fjsnap, see "C.1.1 Executing the fjsnap Command." For details on FJQSS, see "C.1.2 Collecting Information by FJQSS(Information Collection Tool)."

Information

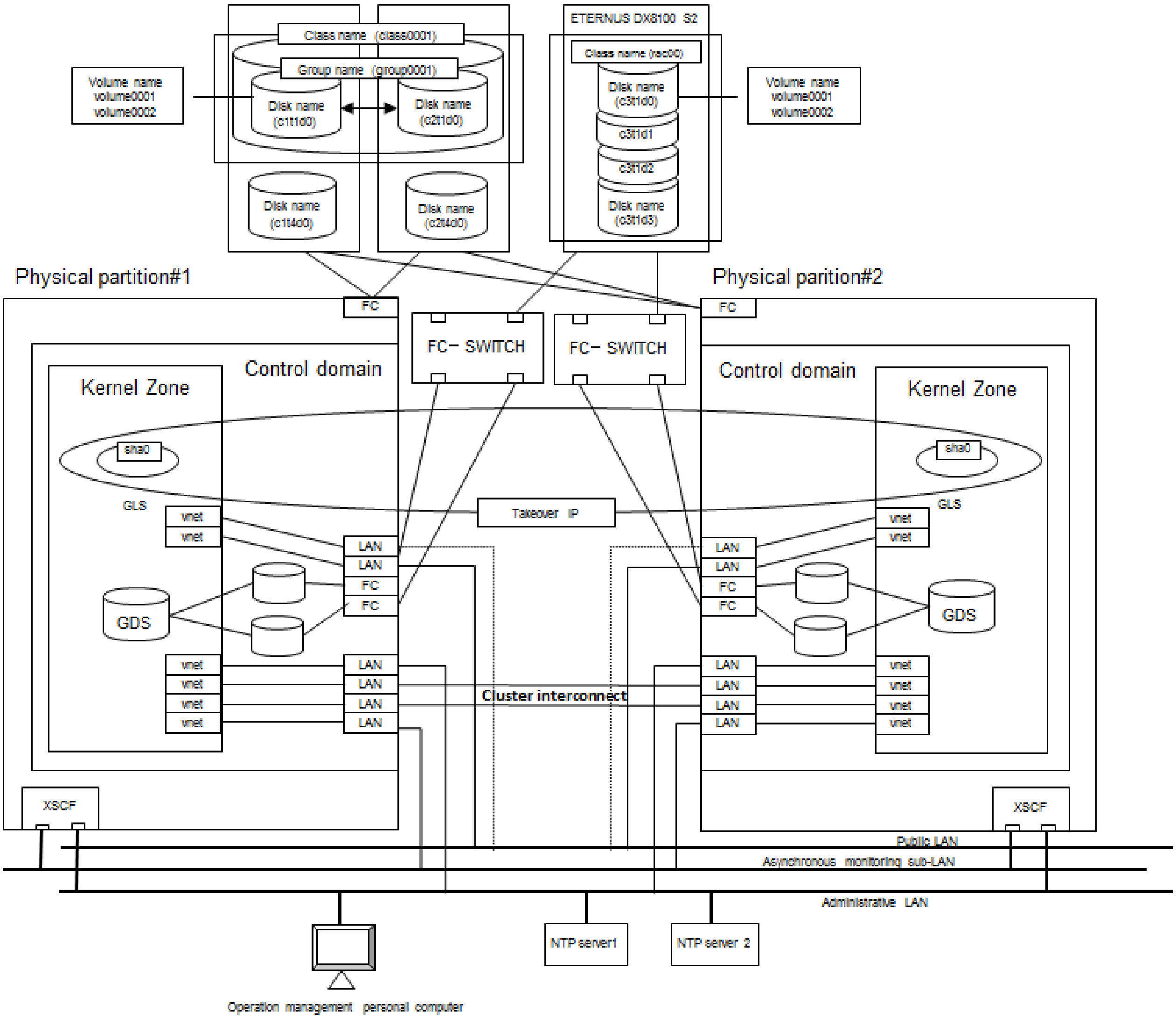
- For reporting a problem, first collect the necessary error diagnosis data. If you do not provide information for problem checking and error reproduction execution, it may take a long time to reproduce and diagnose the problem or it may become impossible to do so.
- Collect the diagnostic material from all nodes in the PRIMECLUSTER System as soon as possible. Especially in the case of data to be collected by FJQSS or fjsnap, necessary information may get lost if too much time elapses after the error.
- For nodes that have been suspended by force, run "sync" in OBP mode in order to collect the system dump.

15.4 Recommended configuration

This section presents a sample layout when creating Kernel Zones in the control domain in Oracle VM Server for SPARC environment and building a PRIMECLUSTER system. Use it for reference when designing your system.

System Configuration

Figure 15.9 Sample System Configuration



15.4.1 Cluster Configuration Worksheet

Below section presents a sample worksheet for cluster configuration.

	Item			Setting	Remarks
Shared disk unit: Type 1	Machine model			ETERNUS DX8100 S2	
	Device name			ETERNUSDX8000#003	
	IP address			10.33.66.215	
	Subnet mask			255.255.255.0	
	Interface			fibre channel	
	LUN number			0 to 0F	
	Ports			CM0CA2P0,CM1CA2P0	
	Connection paths			2	
	FC-SW		Use of fibre channel switching hub	■ yes □ no	
			Machine model	ETERNUS SN200 model 630	
			Device name	ETERNUSSN200#003	
			IP address	10.33.66.203	
			Subnet mask	255.255.255.0	
	File system selection (type of file system used in cluster)			ZFS	
NTP	Operation mode			□ NTP server (integrated in cluster) ■ NTP client	
	Protocol (only when broadcast is specified)				
	Server	Primary	Host name	pclntp1	
			IP address	10.124.95.11	
			Subnet mask	255.255.255.0	
		Secondary	Host name	pclntp2	
			IP address	10.21.8.3	
			Subnet mask	255.255.255.0	
Web-Based Admin View (management view) for control domain	Operation mode			□ 1 ■ 2 □ 3 layers	
	Server	Primary	Host name	primecl01	
			IP address	10.33.66.81	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.81:8081/Plugin.html	
		Secondary	Host name	primecl02	
			IP address	10.33.66.82	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.82:8081/Plugin.html	
	User ID		wvroot group	pclwv	
			clroot group	pclcl	
			cladmin group	pcladm	
			clmon group	pclmon	
			sdxroot group	pclsdx	

Web-Based Admin View (managemen t view) for Kernel Zones	Operation mode			☐ 1 ■ 2 ☐ 3 layers	
	Server	Primary	Host name	pclkz01	
			IP address	10.33.66.91	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.91:8081/Plugin.html	
		Secondary	Host name	pclkz02	
			IP address	10.33.66.92	
			Subnet mask	255.255.255.0	
			URL	http://10.33.66.92:8081/Plugin.html	
	User ID		wvroot group	pclwv	
			clroot group	pclcl	
			cladmin group	pcladm	
			clmon group	pclmon	
			sdxroot group	pclsdx	
Cluster Foundation (CF)	Cluster name			PCLKZONE1	
	Cluster nodes			2	
	Cluster node names			primecl01,primecl02	
	Subnets			1	
	Subnet number			192.168.1.0	
	Subnet mask			255.255.255.0	
	Use of CF remote services			■ cfcp (CF file copy) ■ cfsh (CF remote command execution)	
	IP interconnect			◆ [no] ◇ yes (the number of IP interconnects)	
	Use by RMS			◆ [yes] ◇ no	
Cluster resource management facility (CRM)	Hardware devices stored in resource database			☐ Network unit	Required for using takeover of public LAN (IP) and node names
				☐ Line switching unit	Required for using line takeover function
				■ Disk unit	Required when using GDS and GFS

Node 1	Machine model			SPARC M10-4	
	Device name			DBSV#001	
	Control domain	Domain ID / domain name		–	
		Node name (uname -n)		primecl01	
		CF node name		primecl01	
		Interconnect	IP address	192.168.1.1	
			Subnet mask	255.255.255.0	
			Device name	e1000g3,e1000g4	
		Node in CF quorum set		◆[Yes] ◇ No	Do not change the default setting.
		Cluster timeout of CF (CLUSTER_TIMEOUT)		"20"	Make sure to set "20".
		Use by RMS		◆ [yes] ◇ no Suffix: primecl01RMS	
		RCI	Address	–	Only when using RCI asynchronous monitoring
		Kernel Parameter	scfd:scf_rdctrl_sense_wait	_	
		System console	IP address#1	10.33.66.71	XSCF, ALOM, and ILOM data
			Subnet mask#1	255.255.255.0	
			IP address#2	10.33.76.71	
			Subnet mask#2	255.255.255.0	
			User ID	root	
			User authority	–	
			Connection type	■ SSH □telnet	
		Shutdown Facility (SF)	IP address of management LAN	10.33.66.81	
			Weight	2	
			Shutdown agent	■ XSCF(SPARC M10) □ XSCF(SPARC Enterprise M-series) □ ILOM □ ALOM	
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.71	
			Console LAN#2	10.33.76.71	
			PPAR-ID	0	
			Domain name	primary	
			Connection type	■ SSH □ telnet	
		Use of shared disk		■ yes □no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
		SCSI / fibre channel	Circuits (paths)	2	
			Device path name#1	fjpfca#0	
			sd device name	sd11~sd42	
			Device path name#2	fjpfca#1	
			sd device name	sd43~sd74	
		Management LAN	IP address	10.33.66.81	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Asynchronous monitoring sub-LAN	IP address	10.33.76.81	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN#1	IP address	10.34.214.181	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN#2	IP address	10.34.215.21	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	

	Kernel Zone	Kernel Zone name		kzone01	
		Node name (uname -n)		pclkz01	
		CF node name		pclkz01	
		Interconnect	IP address	192.168.2.1	
			Subnet mask	255.255.255.0	
			Device name	e1000g3,e1000g4	
		Node in CF quorum set		◆[Yes] ◇ No	Do not change the default setting.
		Cluster timeout of CF (CLUSTER_TIMEOUT)		Not set	Do not change the default setting.
		Use by RMS		◆ [yes] ◇ no Suffix: pclkz01RMS	
		RCI	Address	–	Only when using RCI asynchronous monitoring
		Kernel Parameter	scfd:scf_rdctrl_sense_wait	–	
		System console	IP address#1	10.33.66.71	XSCF, ALOM, and ILOM data
			Subnet mask#1	255.255.255.0	
			IP address#2	10.33.76.71	
			Subnet mask#2	255.255.255.0	
			User ID	root	
			User authority	–	
			Connection type	■ SSH □telnet	
		Shutdown Facility (SF)	IP address of management LAN	10.33.66.91	
			Weight	2	
			Shutdown agent	□ XSCF(SPARC M10) □ XSCF(SPARC Enterprise M-series) □ ILOM □ ALOM ■ KZONE(SPARC M10) □ KZONE(SPARC T4)	
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.71	
			Console LAN#2	10.33.76.71	
			PPAR-ID	0	
			Domain name	domain1	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
		SCSI / fibre channel	Circuits (paths)	2	
			Device path name#1	fjpfca#0	
			sd device name	sd11~sd42	
			Device path name#2	fjpfca#1	
			sd device name	sd43~sd74	
		Management LAN	IP address	10.33.66.91	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Asynchronous monitoring sub-LAN	IP address	10.33.76.91	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN#1	IP address	10.34.224.181	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN#2	IP address	10.34.225.21	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	

Node 2	Machine model			SPARC M10-4	
	Device name			DBSV#002	
	Control domain	Domain ID / domain name		–	
		Node name (uname -n)		primecl02	
		CF node name		primecl02	
		Interconnect	IP address	192.168.1.2	
			Subnet mask	255.255.255.0	
			Device name	e1000g3,e1000g4	
		Node in CF quorum set		◆[Yes] ◇ No	Do not change the default setting.
		Cluster timeout of CF (CLUSTER_TIMEOUT)		"20"	Make sure to set "20".
		Use by RMS		◆ [yes] ◇ no Suffix: primecl02RMS	
		RCI	Address	–	Only when using RCI asynchronous monitoring
		Kernel Parameter	scfd:scf_rdctrl_sense_wait	–	
		System console	IP address#1	10.33.66.72	XSCF, ALOM, and ILOM data
			Subnet mask#1	255.255.255.0	
			IP address#2	10.33.76.72	
			Subnet mask#2	255.255.255.0	
			User ID	root	
			User authority	–	
			Connection type	■ SSH □telnet	
		Shutdown Facility (SF)	IP address of management LAN	10.33.66.82	
			Weight	1	
			Shutdown agent	■ XSCF(SPARC M10) □ XSCF(SPARC Enterprise M-series) □ ILOM □ ALOM	
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.72	
			Console LAN#2	10.33.76.72	
			PPAR-ID	1	
			Domain name	primary	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □no	
		scsi-initiator-id (OBP variable #eeprom command)		8	
		SCSI / fibre channel	Circuits (paths)	2	
			Device path name#1	fjpfca#0	
			sd device name	sd11~sd42	
			Device path name#2	fjpfca#1	
			sd device name	sd43~sd74	
		Management LAN	IP address	10.33.66.82	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN#1	IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN#2	IP address	10.34.215.22	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	

	Kernel Zone	Kernel Zone name		kzone02	
		Node name (uname -n)		pclkz02	
		CF node name		pclkz02	
		Interconnect	IP address	192.168.2.2	
			Subnet mask	255.255.255.0	
			Device name	e1000g3,e1000g4	
		Node in CF quorum set		◆[Yes] ◇ No	Do not change the default setting.
		Cluster timeout of CF (CLUSTER_TIMEOUT)		Not set	Do not change the default setting.
		Use by RMS		◆ [yes] ◇ no Suffix: pclkz02RMS	
		RCI	Address	–	Only when using RCI asynchronous monitoring
		Kernel Parameter	scfd:scf_rdctrl_sense_wait	–	
		System console	IP address#1	10.33.66.72	XSCF, ALOM, and ILOM data
			Subnet mask#1	255.255.255.0	
			IP address#2	10.33.76.72	
			Subnet mask#2	255.255.255.0	
			User ID	root	
			User authority	–	
			Connection type	■ SSH □telnet	
		Shutdown Facility (SF)	IP address of management LAN	10.33.66.92	
			Weight	1	
			Shutdown agent	□ XSCF(SPARC M10) □ XSCF(SPARC Enterprise M-series) □ ILOM □ ALOM ■ KZONE(SPARC M10) □ KZONE(SPARC T4)	
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.72	
			Console LAN#2	10.33.76.72	
			PPAR-ID	1	
			Domain name	domain2	
			Connection type	■ SSH □telnet	
		Use of shared disk		■ yes □no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
		SCSI / fibre channel	Circuits (paths)	2	
			Device path name#1	fjpfca#0	
			sd device name	sd11~sd42	
			Device path name#2	fjpfca#1	
			sd device name	sd43~sd74	
		Management LAN	IP address	10.33.66.92	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Public LAN#1	IP address	10.34.224.182	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN#2	IP address	10.34.225.22	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	

15.4.2 GLS Setup Worksheet

Below section presents a sample worksheet for GLS configuration.

■NIC switching mode (logical IP takeover) worksheet

Item				Setting
GLS setup	Takeover virtual interface name			sha0:65
		Takeover virtual IP address (or host name)		10.34.214.183
		Subnet mask		255.255.255.0
	Node name (1)			primecl01
		Kernel Zone configuration	Virtual interface name	sha0
			Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
			Primary interface name	net1
			Secondary interface name	net2
			Physical IP address (or host name)	10.34.214.181
			Logical IP address (or host name)	10.34.214.183
		Monitoring destination	Primary monitoring destination IP address (or host name)	10.34.214.188
			Secondary monitoring destination IP address (or host name)	10.34.214.189
		Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)			primecl02
		Kernel Zone configuration	Virtual interface name	sha0
			Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
			Primary interface name	net1
			Secondary interface name	net2
			Physical IP address (or host name)	10.34.214.182
			Logical IP address (or host name)	10.34.214.183
		Monitoring destination	Primary monitoring destination IP address (or host name)	10.34.214.188
			Secondary monitoring destination IP address (or host name)	10.34.214.189
		Optional function	Standby NIC patrol	◇Disable ◆Enable

15.4.3 GDS Setup Worksheet

Below section presents a sample worksheet for GDS configuration.

Item					Setting
GDS configuration	Class 1	Class name			class0001
		Class scope (node name)	Node name 1		primecl01
			Node name 2		primecl02
		Disk 1 connected to Group	SDX disk name		disk0001
			Physical disk name in node1		c1t1d0
			Physical disk name in node2		c1t1d0
		Disk 2 connected to Group	SDX disk name		disk0002
			Physical disk name in node1		c2t1d0
			Physical disk name in node2		c2t1d0
		Highest-order group1	Group name		group0001
			Group type		mirror
			Disk/low-order group name	Disk/low-order group1	disk0001
				Disk/low-order group2	disk0002
			Single volume1	Volume name	volume0001
				Size	20971520
			Single volume2	Volume name	volume0002
				Size	41943040
	Class 2	Class name			rac00
		Class scope (node name)	Node name 1		primecl01
			Node name 2		primecl02
		Single disk1	SDX disk name		disk0001
			Physical disk name in node1		c3t1d0
			Physical disk name in node2		c3t1d0
			Single volume1	Volume name	volume0001
				Size	41943040
			Single volume2	Volume name	volume0002
				Size	41943040

Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment

This chapter explains the overview, design, building, operation, maintenance, and recommended configuration for implementing PRIMECLUSTER to an Oracle Solaris Zones environment. Oracle Solaris Zones environments are available on guest domains in Oracle VM Server for SPARC Environments as well as on physical servers.

16.1 Design

This section explains the design for PRIMECLUSTER operating on Oracle Solaris Zones environments.

The table below shows examples of PRIMECLUSTER system building in Oracle Solaris Zones environments.

	Migration (building) destination	Building PRIMECLUSTER	PRIMECLUSTER medium to use	Building OS/ Middleware
Newly building	Oracle Solaris 10 Zones environment Oracle Solaris 11 Zones environment	Newly building	Medium of PRIMECLUSTER 4.3A20 or later	Newly building
Migrating from an environment with PRIMECLUSTER operation	Oracle Solaris 10 Zones environment Oracle Solaris 11 Zones environment	Rebuilding	Medium of PRIMECLUSTER 4.3A20 or later	Migrating by creating a flash archive (*1)
	Oracle Solaris 8 Containers (OSLC) environment Oracle Solaris 9 Containers (OSLC) environment	Rebuilding	Medium of PRIMECLUSTER used in the migration source	Migrating by creating a flash archive (*1)
Migrating from an environment without PRIMECLUSTER operation	Oracle Solaris 10 Zones environment Oracle Solaris 11 Zones environment	Newly building	Medium of PRIMECLUSTER 4.3A20 or later	Migrating by creating a flash archive (*1)
	Oracle Solaris 8 Containers (OSLC) environment Oracle Solaris 9 Containers (OSLC) environment	Newly building	Medium of PRIMECLUSTER 4.3A20 or later	Migrating by creating a flash archive (*1)

*1: See each middleware manual to check the compatibility of the middleware in a new environment.

Range of support and supported resources by each middleware are different in Oracle Solaris Zones environments compared to physical environments. Besides, newly building PRIMECLUSTER is required in Oracle Solaris Zones environments. Rebuilding PRIMECLUSTER is required even for the migration from an environment with PRIMECLUSTER operation as the above example shows.

In the migration from an environment with PRIMECLUSTER operation to Oracle Solaris Legacy Containers (OSLC) environment, migratable PRIMECLUSTER versions are as follows:

- Solaris 8

  PRIMECLUSTER 4.1A30 and 4.1A40

- Solaris 9

  PRIMECLUSTER 4.1A30, 4.1A40, and 4.2A00

See

- See "16.2 Building" for the complete building procedure to newly build PRIMECLUSTER in Oracle Solaris 10 Zones environment or Oracle Solaris 11 Zones environment, or to migrate PRIMECLUSTER to Oracle Solaris 10 Zones environment or Oracle Solaris 11 Zones environment.
- See "16.3 Configuration for Using OSLC" for the complete building procedure due to a migration to Oracle Solaris Legacy Containers(OSLC) environments.
- For how to migrate the environment where GFS is used to the Oracle Solaris Legacy Containers(OSLC) environment, see "Appendix E Migrating GFS by Using OSLC" in "PRIMECLUSTER Global File Services Configuration and Administration Guide."

Note

The following function is not available in an OSLC environment.

- IPv6 on a public LAN

16.1.1 Range of Support

The configurations which support PRIMECLUSTER operating on an Oracle Solaris Zones environment are as follows.

- Number of cluster applications which can be created to the non-global zone: 1
- Non-global zone configuration

  Whole root zone (a zone not sharing the global zones and system files)
- Non-global zone type

Non-global zone OS	Non-global zone type
Oracle Solaris 10	Oracle Solaris 8 Containers (OSLC) Oracle Solaris 9 Containers (OSLC) Oracle Solaris 10 Zones
Oracle Solaris 11	Oracle Solaris 10 Zones Oracle Solaris 11 Zones

- Number of CPU cores assigned to the global zone: 2 or more

16.1.2 Design Items

If using PRIMECLUSTER with Oracle Solaris Zones environments, select the operation mode and configuration for each of the following items. For the criteria in selecting for each item, refer to the sections below.

Table 16.1 Design Items

Item	Choices
Operation mode	Warm-standby operation Cold-standby operation Single-node cluster operation
Non-global zone image allocation	Non-shared Shared
Network mode	Exclusive IP zone Shared IP zone

Item	Choices
Application monitoring	Yes No

16.1.2.1 Operation Mode

The PRIMECLUSTER operations are on multiple nodes and on one node.

The features are as follows:

Table 16.2 The features of operations on multiple nodes and one node

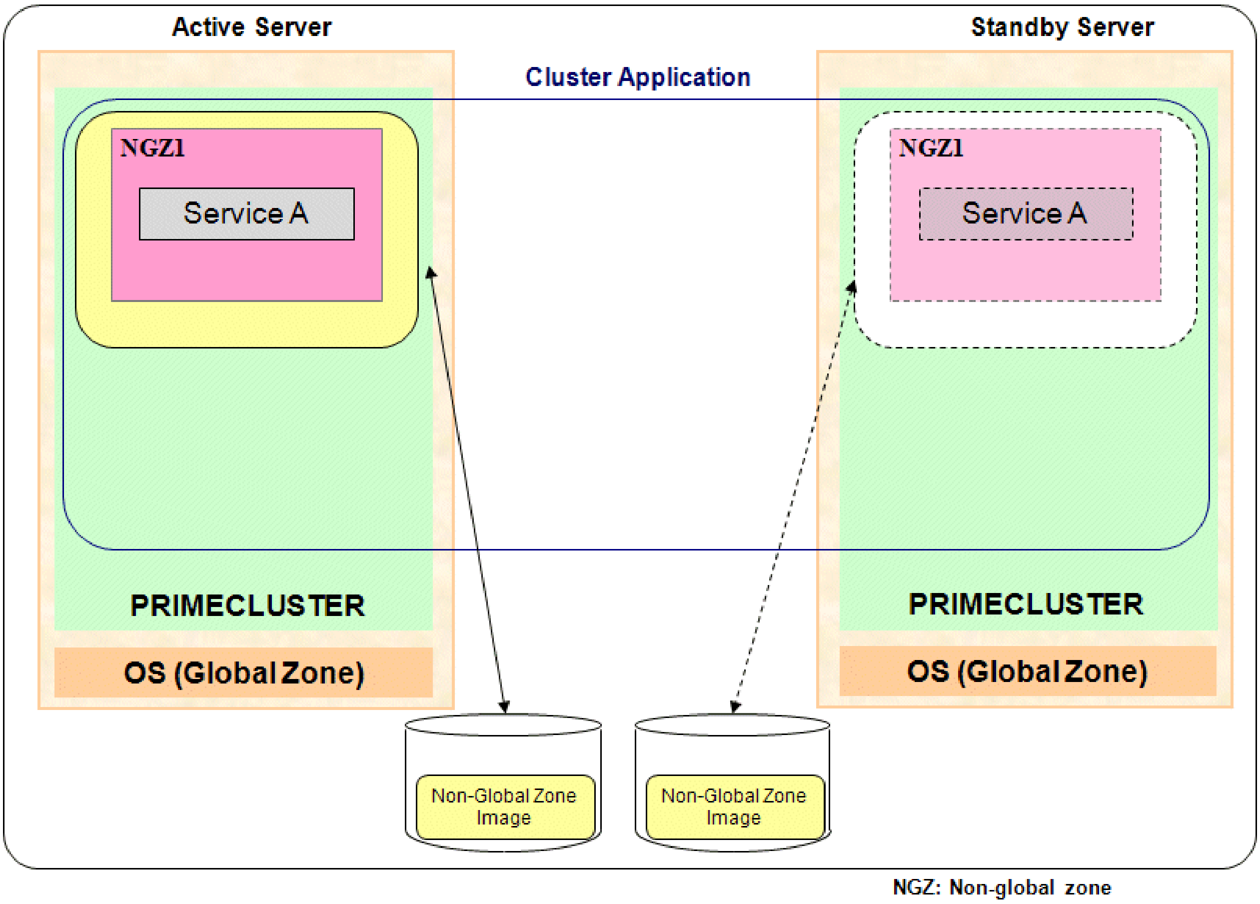
Item	Multiple nodes	One node
	Warm-standby operation Cold-standby operation	Single-node cluster operation
Operation recovery from soft errors	Available	Available
Operation recovery from hard errors of the server	Available	Not available
Enhancement of the availability by one node	Not available	Available

When operating on multiple nodes, operation server enables the continuation of service by taking over the application running on the non-global zone to a standby server. There are two types of operations as follows:

- Warm-standby Operation

  With the non-global zones started up on both the operating server and standby server as is, this operation switches over only the applications operating within the non-global zone, and takes over services. Since the standby system's non-global zone OS enters a startup status, a faster switchover is possible.

  Figure 16.1 Warm-standby Operation

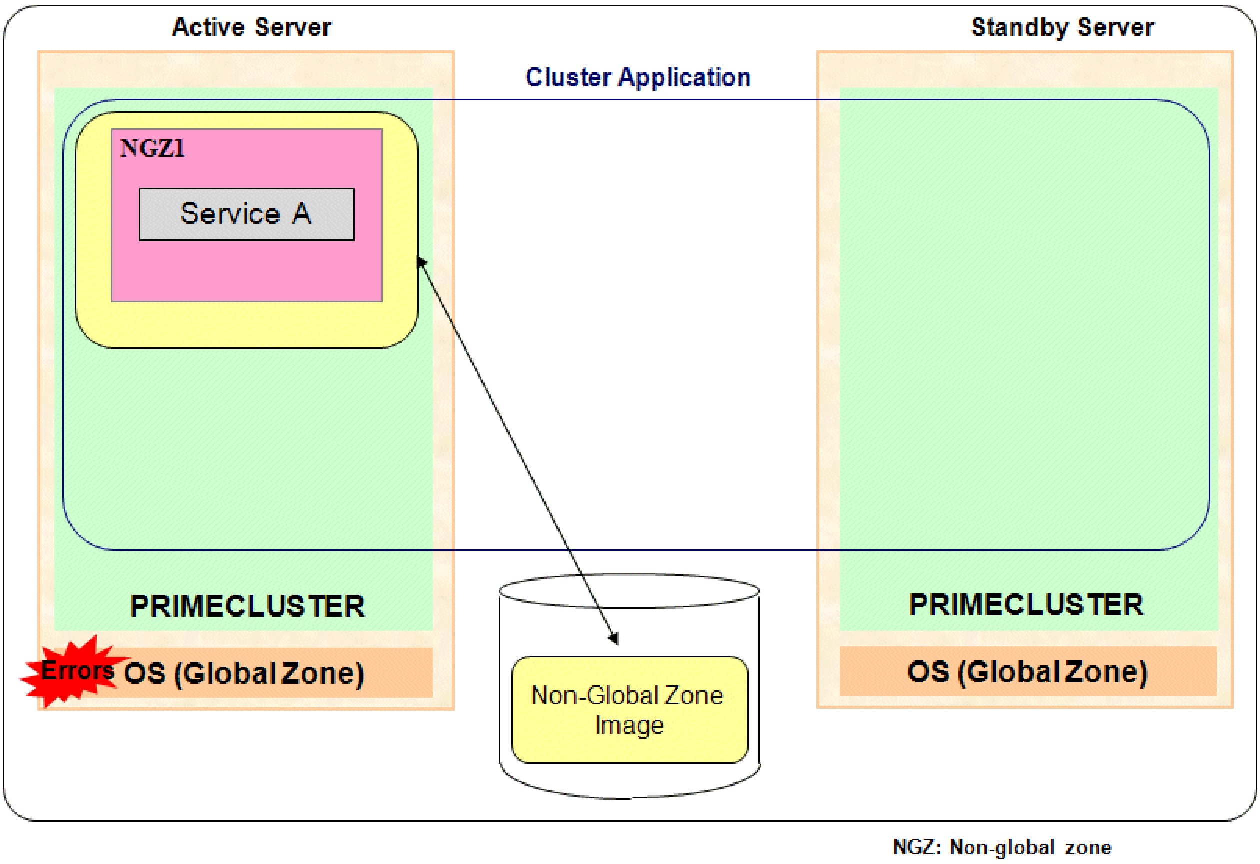


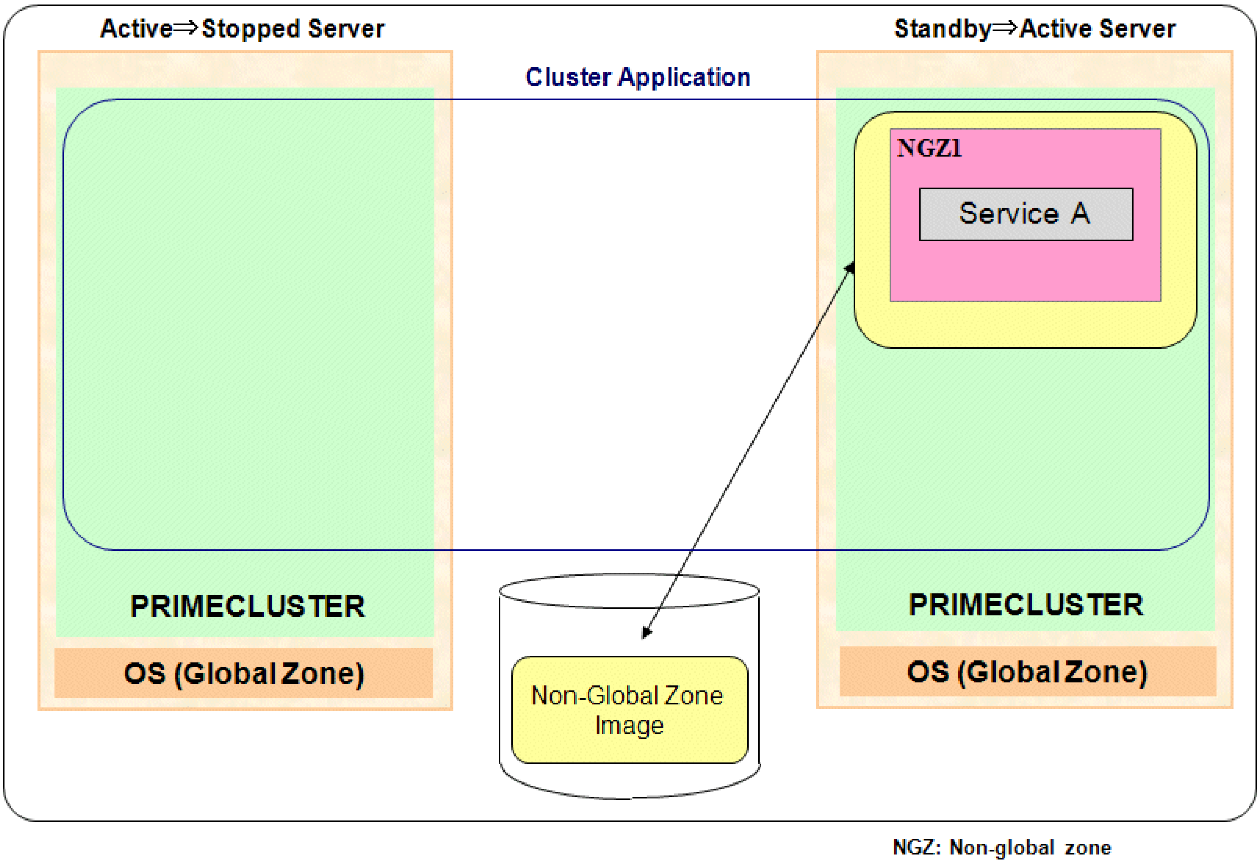
- Cold-Standby Operation

  This operation takes over services between servers by starting the non-global zones on the standby server after the non-global zones on the operating server have been stopped. A configuration sharing non-global zone images between the cluster nodes is possible for this operation.

During operation, since this operation does not start non-global zones on the standby server, it does not use up CPU and memory resources.

Figure 16.2 Cold-Standby Operation





The characteristics for each operation mode are as follows.

Table 16.3 Characteristics of the Warm-standby Operation and Cold-Standby Operation

Item	Warm-standby Operation	Cold-Standby Operation
High-speed switchover	Available	Not available
Conserving use of standby server CPU and memory	Not available	Available
Non-global zone images not shared	Available	Available
Non-global zone images shared	Not available	Available

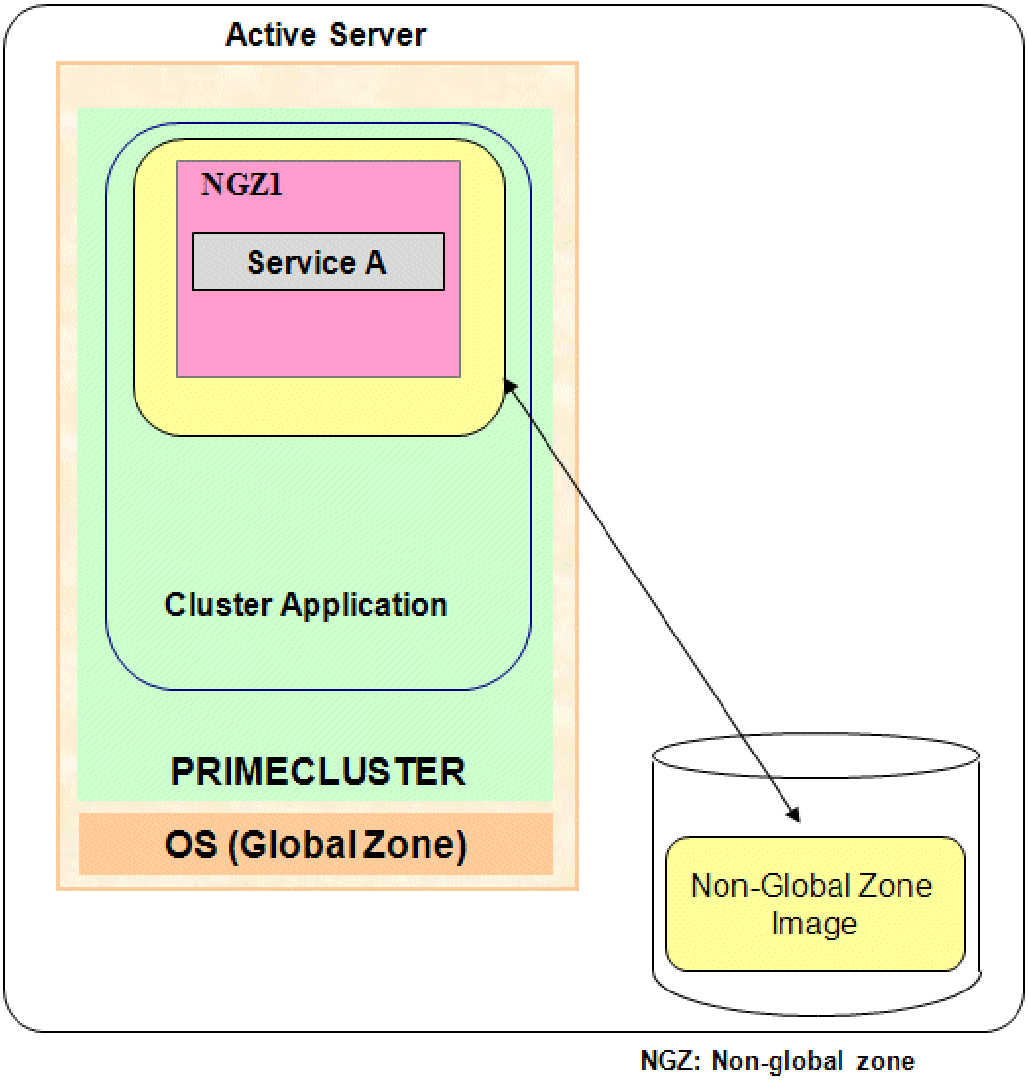
When operating on one node, the operation type is as follows:

- Single-Node Cluster Operations

  Monitor the status of the OS and applications on the non-global zones. The availability is increased by restarting the non-global zone or an application on the non-global zone automatically to perform recovery when an error is detected.

In this configuration, PRIMECLUSTER on the global zone monitors the status of the non-global zone. Also, PRIMECLUSTER on the non-global zone monitors the cluster applications on the non-global zone.

Figure 16.3 Single-Node Cluster Operations



16.1.2.2 Allocation of Non-Global Zone Images

It is possible to share non-global zone images between cluster nodes or, alternatively, allocate them to each node.

The characteristics for each configuration are as follows.

Table 16.4 Characteristics for Allocation of Non-Global Zone Images

Item	Non-shared	Shared
Warm-standby operation	Available	Not available
Cold-standby operation	Available	Available
Rolling update (maintenance of the standby system non-global zones while service operations are in progress)	Available	Not available
Continuation of services through switchover when destruction of non-global zone images occurs	Available	Not available
A shared disk for takeover of non-global zone images is not needed	Available	Not available
Maintenance work is required only for one node	Not available	Available

Item	Non-shared	Shared
The non-global zones are identical in content for the operational system and standby system	Not available	Available

Allocate non-global zone images to the following disk areas:

- For Non-Shared

  Allocate non-global zone images to volumes of the root class or local classes of GDS, or the disks which have not been registered with GDS.

- For Shared

  Allocate non-global zone images to the volume of GDS's shared class.

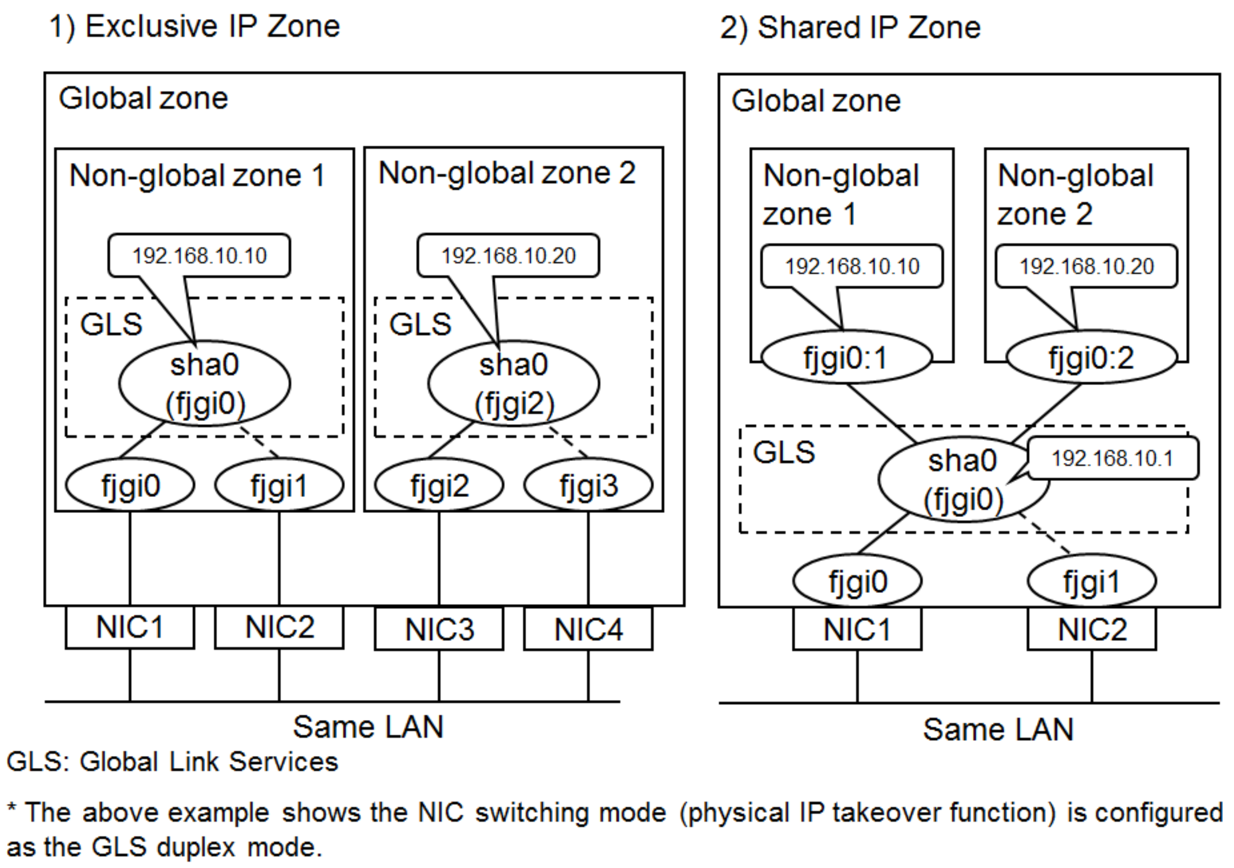
16.1.2.3 Network Mode

For the non-global zones' network mode, there are exclusive IP zone and shared IP zone.

- Exclusive IP Zone
  This is a network mode wherein the identified non-global zone has exclusive possession of the network interface. By splitting the global zone and non-global zone network interfaces, it is possible to split the network design between the zones.

- Shared IP Zone
  This is a network mode which shares with the non-global zone the network interface configured with the global zone. It is possible to share the network interface between a global zone and multiple non-global zones.

Figure 16.4 Network Mode and GLS Setup



The characteristics for each network mode are as follows.

Table 16.5 Characteristics of Exclusive IP Zones and Shared IP Zones

Item	Exclusive IP Zone	Shared IP Zone
Splitting the network design between zones	Available	Not available
Sharing the network interface between a global zone and multiple non-global zones	Not available	Available

Information

If performing non-global zone transmission with GLS fast switching mode or GS/SURE linkage mode, select the shared IP zone mode. Transmission which has used these modes is not possible with an exclusive IP zone. If using NIC switching mode, transmission is possible with either mode.

16.1.2.4 Application Monitoring

For Oracle Solaris Zones environments, PRIMECLUSTER, in addition to the global zone monitoring, also provides the following monitoring functions to the non-global zones.

- Non-global zone status monitoring

  PRIMECLUSTER confirms that the non-global zones have not stopped (this is confirmed with the status displayed using the zoneadm command)

- Non-global zone OS error monitoring

  Confirm that no error occurs when logging in (zlogin command) to non-global zone so that hang-ups in the non-global zones can be detected.

Additionally, if making the applications operating on the non-global zones the target for monitoring, the following type of monitoring is possible.

- Non-global zone application monitoring

  Provides an application monitoring function equivalent to monitoring on the global zone using the RMS operating on the non-global zones.

The characteristics of application monitoring are as follows.

Table 16.6 Characteristics of Application Monitoring

Item	Application Monitoring	
	Yes	No
Detects stop to non-global zones	Available	Available
Detects hang-ups with non-global zones	Available	Available
Application monitoring for non-global zones	Available	Not available

Note

- To make applications operating on the non-global zones the target for monitoring, it is necessary to install PRIMECLUSTER to the non-global zones and create a single cluster configuration.

- If performing application monitoring, for the resources which can be used in the non-global zone, refer to "16.1.3 Resource Configuration."

- Be aware that the state of the non-global zone will not be detected as a resource fault when the abnormal state of the non-global zone does not continue for 10 seconds or more because the non-global zone is monitored at 10-second intervals by the Cmdline resource. For example, if the non-global zone is down abnormally but it is rebooted within 10 seconds, it will not be detected as a resource fault.

16.1.3 Resource Configuration

The supported operation mode and configuration combinations are as follows.

Table 16.7 Operation Mode and Configuration Combinations

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring	Support when using OSLC
1	Warm-standby Operation	Non-shared	Exclusive IP Zone	Yes	Yes
2	Warm-standby Operation	Non-shared	Shared IP Zone	Yes	Yes
3	Cold-Standby Operation	Shared/Non-shared	Exclusive IP Zone	Yes	No
4	Cold-Standby Operation	Shared/Non-shared	Shared IP Zone	Yes	Yes
5	Cold-Standby Operation	Shared/Non-shared	Exclusive IP Zone	No	No
6	Cold-Standby Operation	Shared/Non-shared	Shared IP Zone	No	Yes
7	Single-Node Cluster Operation	-	Exclusive/ Shared IP Zone	Yes/No	Yes

The resources which are necessary and the resources which can be set up differ according to configuration. The necessary resources and the resources which can be set up for each configuration are as follows.

Table 16.8 Resource Configuration

Zone	Resource	Description	Configuration						
			1	2	3	4	5	6	7
Global Zone	Cmdline (for control of non-global zones)	Status monitoring and control of the applications within the non-global zones and the global zone *1: The IP addresses set up to the non-global zones (i.e., the IP addresses set up with zonecfg) are taken over per each non-global zone.	A	A	A *1	A *1	A *1	A *1	A
	Cmdline (for control of shared IP)	Takeover of non-global zone IP addresses	D	C	D	D	D	D	D
	Gls	Detects non-global zone network errors Note: If using NIC switching mode, the takeover IP address is set, but this IP address is not used for non-global zone transmission.	D	C	D	C	D	C	C
	Gds Note: It is possible to use one GDS shared class for both purposes given at right	GDS shared class for the non-global zone images *2: Necessary if sharing non-global zone images	E	E	B *2	B *2	B *2	B *2	E

Zone	Resource	Description	Configuration						
			1	2	3	4	5	6	7
		GDS shared class for data takeover between non-global zones *3: Necessary if performing data takeover between non-global zones by using a shared disk device (raw access).	C *3	C *3	C *3	C *3	C *3	C *3	E
	Fsystem	Mount control of non-global zone images *4: Necessary if sharing non-global zone images	D	D	B *4	B *4	B *4	B *4	C
	Optional resources		C	C	C	C	C	C	C
Non-Global Zone	Gls Note: Setup possible only for NIC switching mode	IP address takeover for non-global zones Detects errors to the non-global zone network	C	E	C	E	E	E	D
	Fsystem	Control of mounting of switching file system between non-global zones	C	C	D	D	E	E	C
	Cmdline		C	C	C	C	E	E	C
	Oracle		C	C	C	C	E	E	C
	NetWorker Note: Setup impossible for Solaris 8/9 Containers		C	C	C	C	E	E	C
	NetApp		C	C	C	C	E	E	C
	NetBackup Note: Setup impossible for Solaris 8/9 Containers		C	C	C	C	E	E	C
	Systemwalker		C	C	C	C	E	E	C
	Status migration procedures Note: Setup impossible for Solaris 8/9 Containers		C	C	C	C	E	E	C

A:Required, B:Required with Conditions, C:Setup Possible, D:Setup Unrequired, E:Setup Impossible
GDS: Global Disk Services

Information

The processing of the Cmdline resource (shared IP control) which achieves IP address takeover between the shared IP zones may differ depending on the duplexing mode for the combined GLS. If using NIC switching mode, the Cmdline resource (shared IP control) is achieved by newly adding an IP address and performing zone setup. However, if using fast switching mode or GS/SURE linkage mode, the Cmdline resource (shared IP control) is achieved by performing zone setup to the takeover IP address which the GLS manages.

See

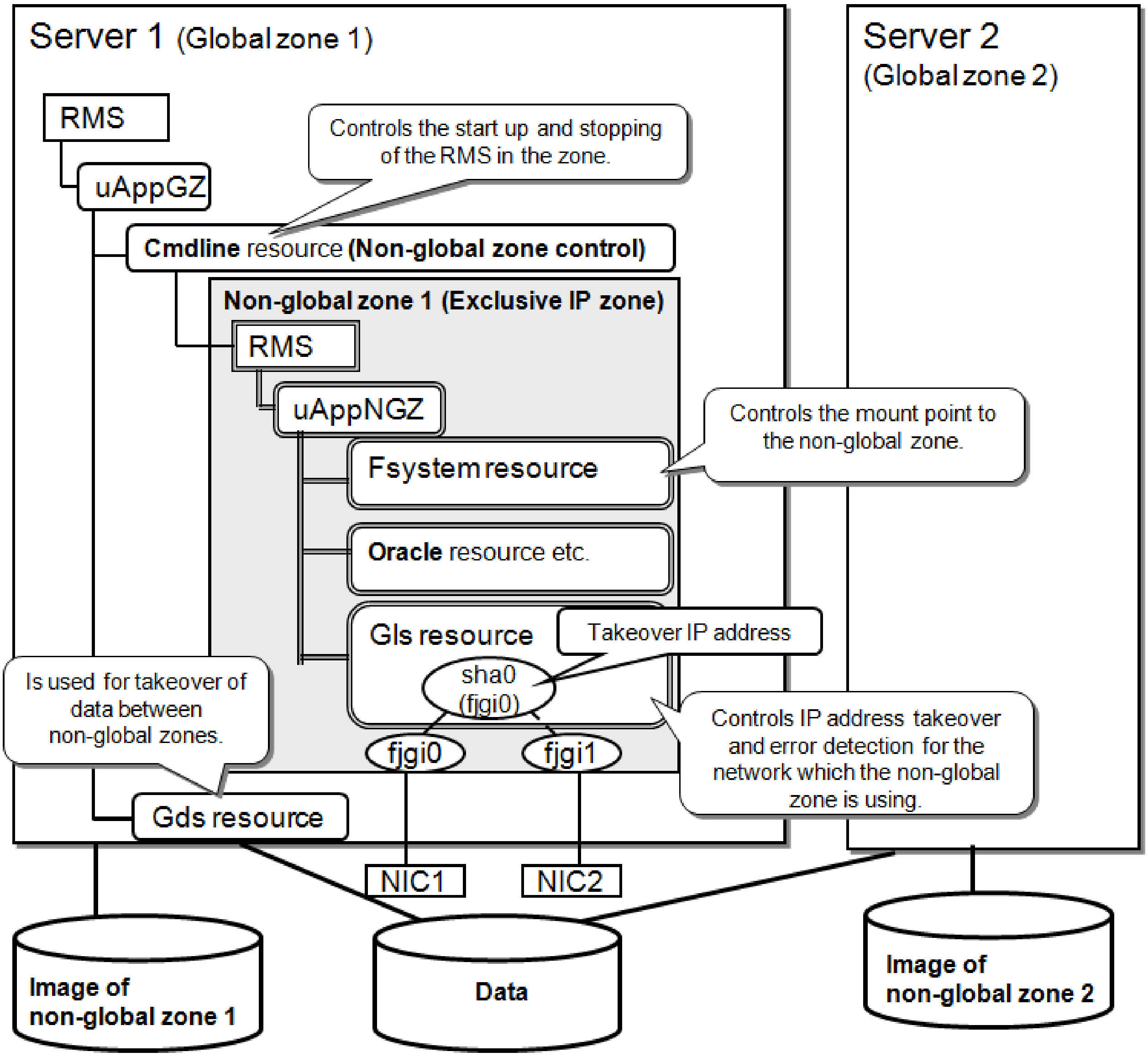
For combining with Wizard products, refer to each Wizard product manual.

16.1.3.1 Configuration 1

Table 16.9 Operation Mode and Configuration for Configuration 1

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
1	Warm-standby operation	Non-shared	Exclusive IP Zone	Yes

Figure 16.5 Resource Configuration for Configuration 1



16.1.3.2 Configuration 2

Table 16.10 Operation Mode and Configuration for Configuration 2

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
2	Warm-standby Operation	Non-shared	Shared IP Zone	Yes

Figure 16.6 Resource Configuration for Configuration 2 (When Using NIC Switching Mode)

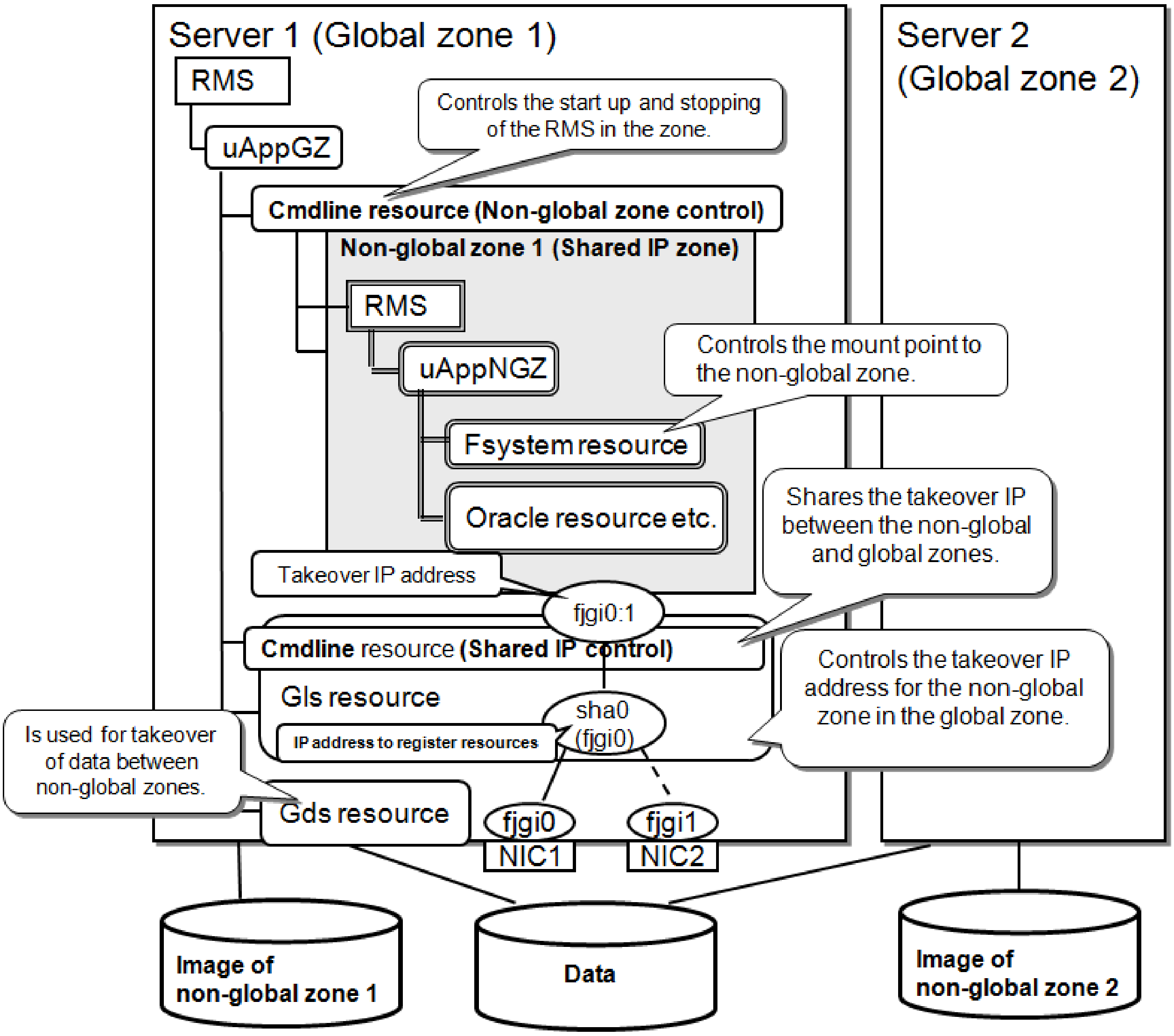
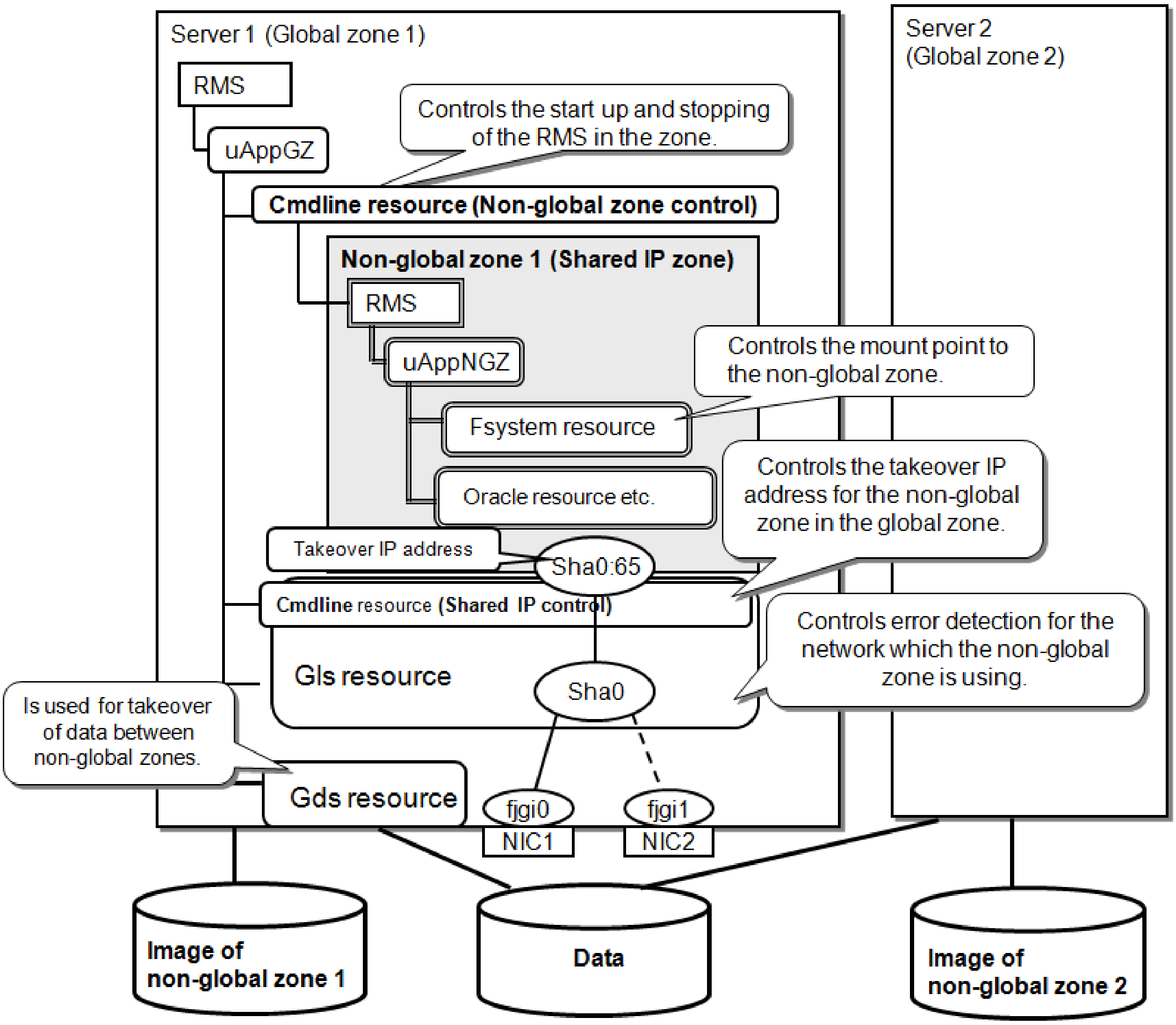


Figure 16.7 Resource Configuration for Configuration 2 (When Using Fast Switching Mode or GS/SURE Linkage Mode)

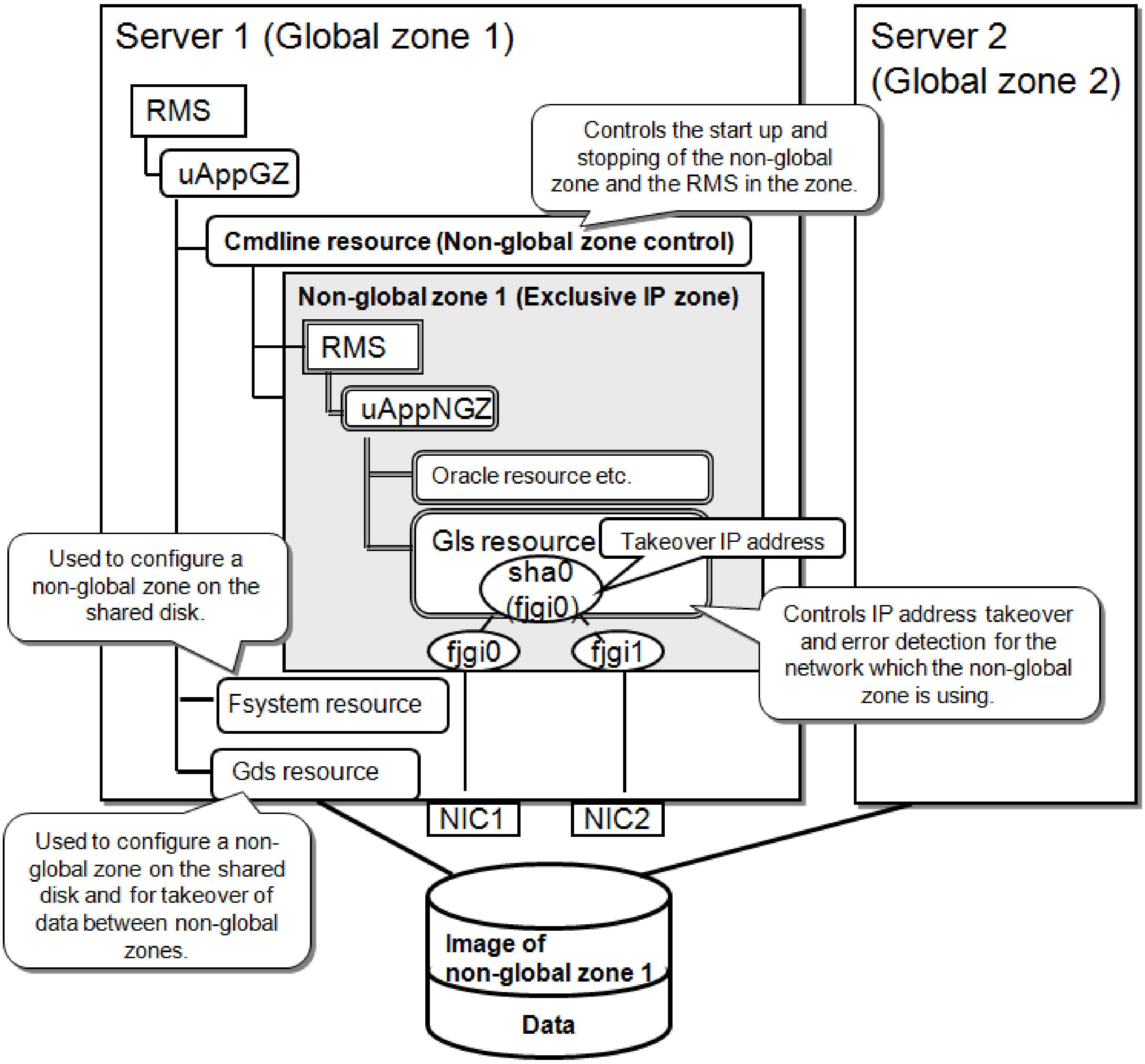


16.1.3.3 Configuration 3

Table 16.11 Operation Mode and Configuration for Configuration 3

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
3	Cold-Standby Operation	Shared/Non-Shared	Exclusive IP Zone	Yes

Figure 16.8 Resource Configuration for Configuration 3 (If Sharing Non-Global Zone Images)

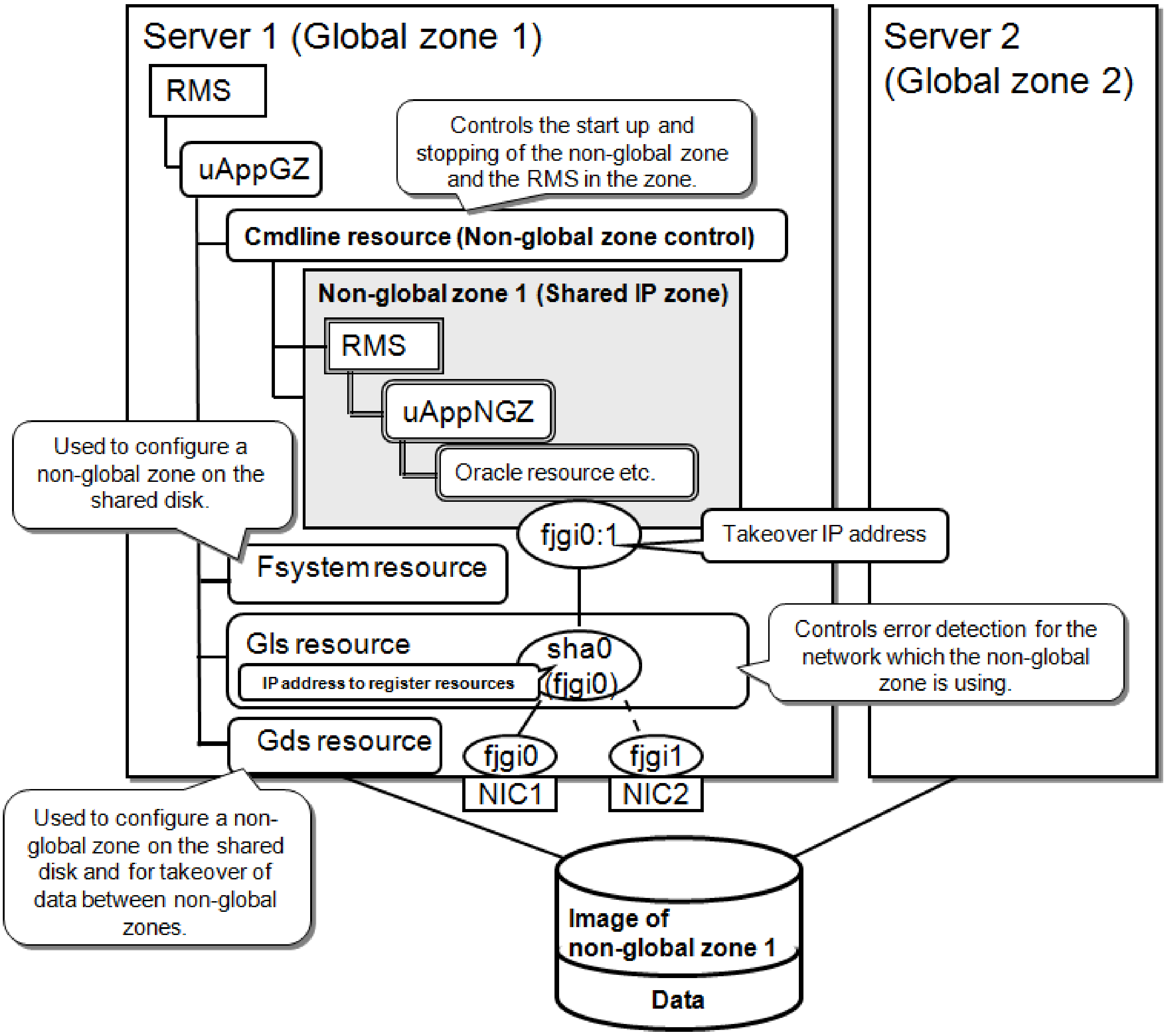


16.1.3.4 Configuration 4

Table 16.12 Operation Mode and Configuration for Configuration 4

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
4	Cold-Standby Operation	Shared/Non-shared	Shared IP Address	Yes

Figure 16.9 Resource Configuration for Configuration 4 (If Sharing Non-Global Zone Images)

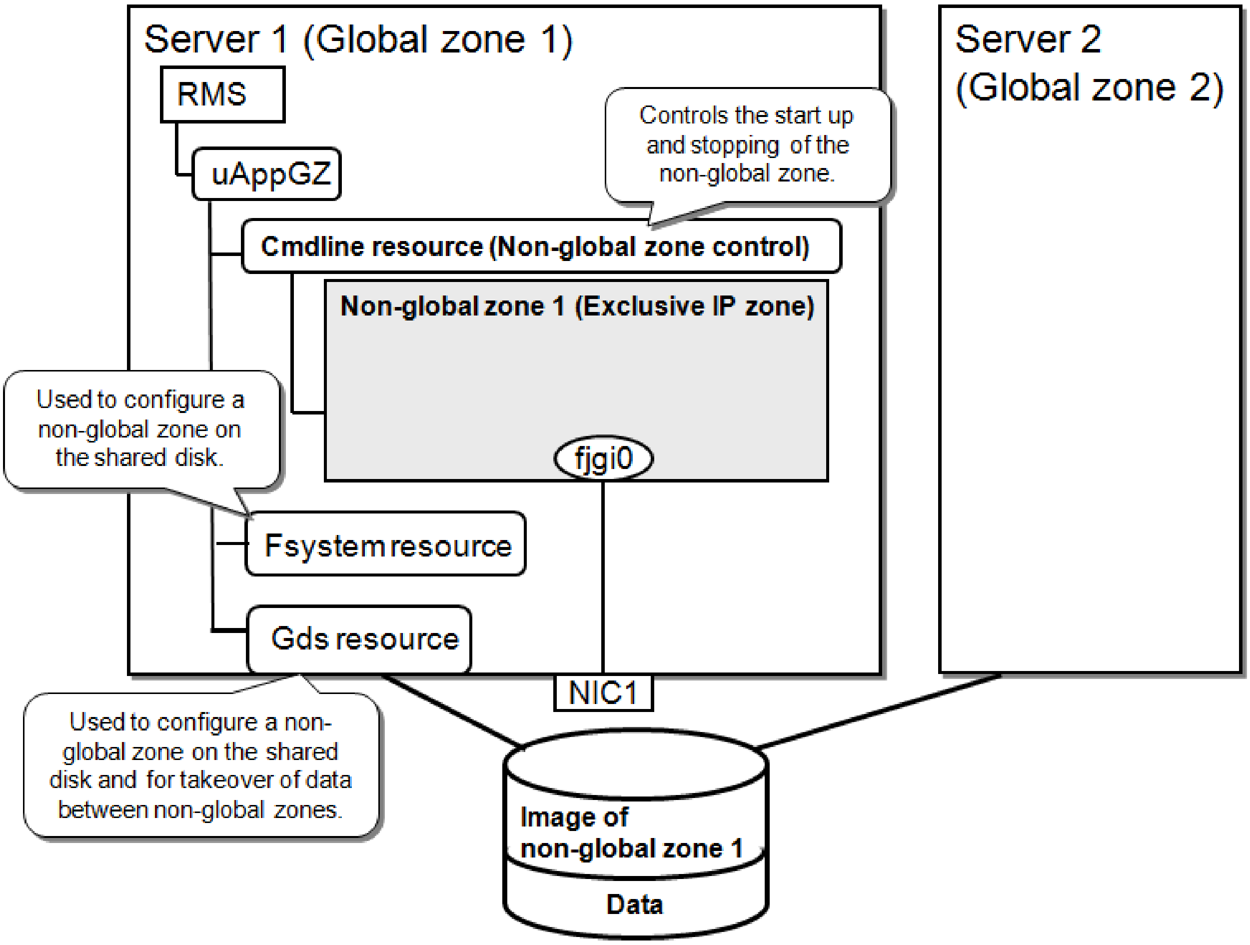


16.1.3.5 Configuration 5

Table 16.13 Operation Mode and Configuration for Configuration 5

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
5	Cold-Standby Operation	Shared/Non-Shared	Exclusive IP Zone	No

Figure 16.10 Resource Configuration for Configuration 5 (If Sharing Non-Global Zone Images)

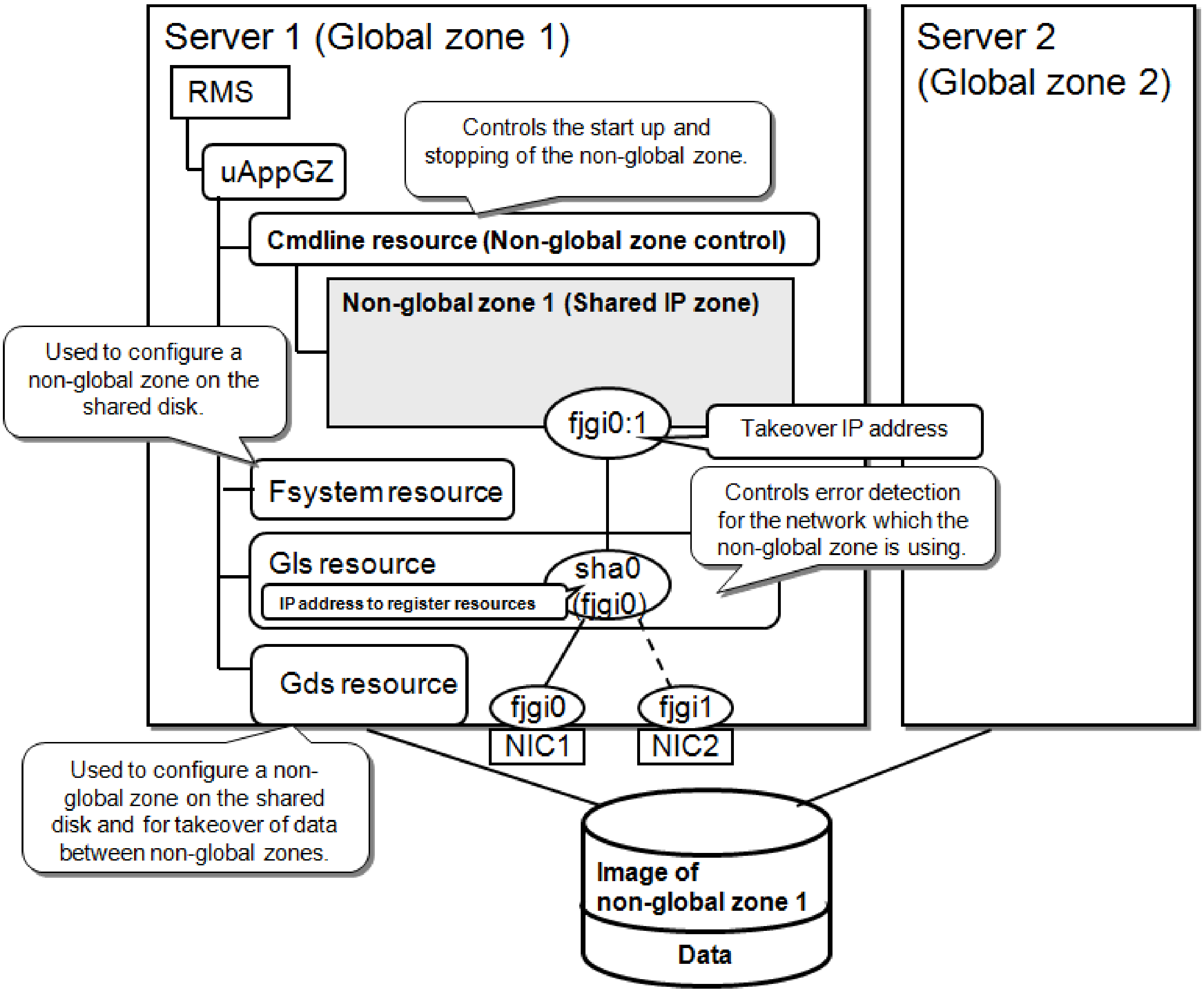


16.1.3.6 Configuration 6

Table 16.14 Operation Mode and Configuration for Configuration 6

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
6	Cold-Standby Operation	Shared/Non-Shared	Shared IP Zone	No

Figure 16.11 Resource Configuration for Configuration 6 (If Sharing Non-Global Zone Images)

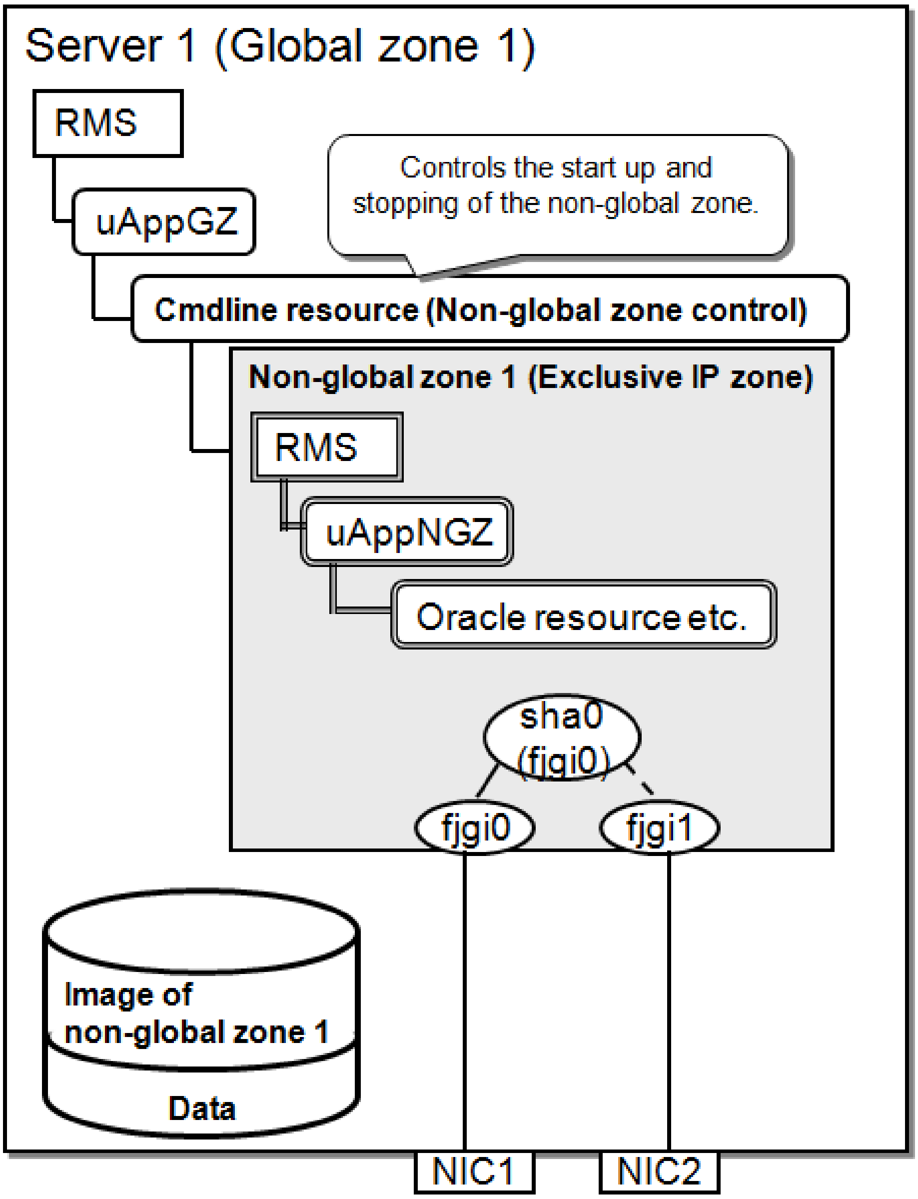


16.1.3.7 Configuration 7

Table 16.15 Operation Mode and Configuration for Configuration 6

Configuration	Operation Mode	Allocation of Non-Global Zone Images	Network Mode	Application Monitoring
7	Single-Node Cluster Operation	Not available	Exclusive/Shared IP Zone	Yes/No

Figure 16.12 Resource Configuration for Configuration 7 (When there are Exclusive IP Zone and Application Monitoring)



16.1.4 Different Specifications when Installing a New PRIMECLUSTER in OSLC Environments

This section describes different specifications between PRIMECLUSTER in the non-global zone and PRIMECLUSTER in the global zone.

- When migrating PRIMECLUSTER (installing a new PRIMECLUSTER) from the old Solaris 8 or Solaris 9 environment where PRIMECLUSTER was not used to new OSLC environment.

However, this section does not cover the contents described from "16.1.1 Range of Support" to "16.1.3 Resource Configuration."

Note

When migrating PRIMECLUSTER from the 4.2A00 or earlier version to OSLC environment, see the manual of the PRIMECLUSTER version used in the migration source.

16.1.4.1 Setup of the Fsystem Resource

Difference

The files that define the mount point of the file system used by the Fsystem resource and the entry of NFS are different between in the non-global zone and in the global zone.

Description

Replace /etc/vfstab.pcl file and /etc/dfs/dfstab.pcl file, which are described in this manual, with etc/vfstab file and /etc/dfs/dfstab file instead for PRIMECLUSTER in the non-global zone.

16.1.4.2 hvshut Command

Difference

The default value of the environment variable RELIANT_SHUT_MIN_WAIT, which specifies the timeout interval of the hvshut command, is different between in the non-global zone and in the global zone.

Description

The default value of this environment variable RELIANT_SHUT_MIN_WAIT, which specifies the timeout interval of the hvshut command, is 900 (seconds) instead of INTMAX (seconds) for PRIMECLUSTER in the non-global zone. The command times out in 900 (seconds) if this environment variable is unchanged from the default value.

16.1.4.3 HV_CONNECT_TIMEOUT

Difference

The default value of the RMS local environment variable HV_CONNECT_TIMEOUT is different between in the non-global zone and in the global zone.

Description

The default value of HV_CONNECT_TIMEOUT is 5 (seconds) instead of 30 (seconds) for PRIMECLUSTER in the non-global zone.

16.1.4.4 RMS Message

Difference

The RMS message (SYS, 8) that is logged in the syslog file is different between in the non-global zone and in the global zone.

Description

The RMS message (SYS, 8) that is logged in the syslog file is as follows.

PRIMECLUSTER in the global zone:

(SYS, 8): ERROR: RMS failed to shut down the host <host> via a Shutdown Facility, no further kill functionality is available.

The cluster is now hung. An operator intervention is required.

PRIMECLUSTER in the non-global zone:

(SYS, 8): ERROR: RMS failed to shut down the host <host> via a Shutdown Facility, no further kill functionality is available. The cluster is now hung.

16.1.4.5 Severity of the RMS Wizard Message

Difference

Severity of the following RMS wizard message is different between in the non-global zone and in the global zone.

"cannot grab mount lock for dostat() check_getbdev(), returning previous state"

Description

The message "cannot grab mount lock for dostat() check_getbdev(), returning previous state" is displayed as WARNING instead of NOTICE for PRIMECLUSTER in the non-global zone.

16.1.4.6 Port Number Used by RMS

Difference

The port number used by RMS is changed.

Description

The port number 11111 is used for PRIMECLUSTER in the global zone while the number is not used for PRIMECLUSTER in the non-global zone.

16.2 Building

This section explains below the procedure for building clusters in an Oracle Solaris Zones environment.

For the building when using Oracle Solaris Legacy Containers(OSLC) environments, see "16.3 Configuration for Using OSLC."

For details on building the Oracle Solaris Zones, see Oracle Solaris documents.

For details on building a single-node cluster, see Note in "5.1.1 Setting Up CF and CIP" and "6.7.2.1 Creating Standby Cluster Applications."

To build Oracle Solaris Zones on guest domains on Oracle VM Server for SPARC Environments, create guest domains on Oracle VM Server for SPARC Environments beforehand, and then follow the procedure below.

The overview of the environment building procedure explained in this section is shown below.

Figure 16.13 Flow for Building PRIMECLUSTER

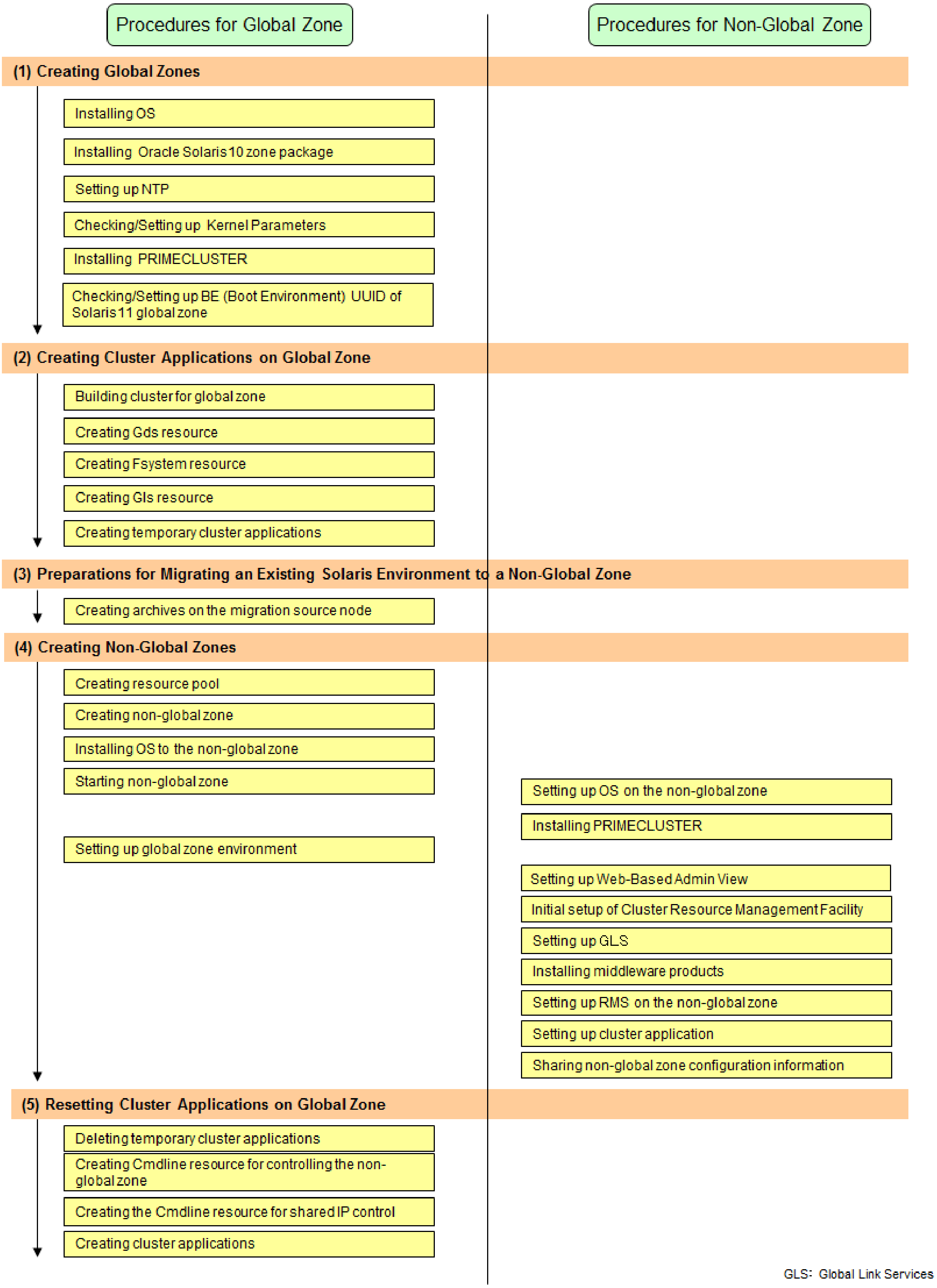


Table 16.16 Building Procedure

No	Task	Explanatory Chapter	Task necessity for Each Configuration (Configuration number of 13.1.3)						
			1	2	3	4	5	6	7
(1) Create Global Zone		16.2.1							
1	OS Installation and Setup to the Global Zone	16.2.1.1	A	A	A	A	A	A	A
2	Installing Oracle Solaris 10 Zone Package	16.2.1.2	B	B	B	B	B	B	B
3	Setup of NTP to Global Zone	16.2.1.3	A	A	A	A	A	A	B
4	Check/Setup of Kernel Parameters to Global Zone	16.2.1.4	A	A	A	A	A	A	A
5	Installing PRIMECLUSTER to the Global Zone	16.2.1.5	A	A	A	A	A	A	A
6	Check/Setup of BE (Boot Environment) UUID of a Solaris 11 Global Zone	16.2.1.6	B	B	B	B	B	B	B
(2) Create Cluster Applications on Global Zone		16.2.2							
7	Building Cluster	16.2.2.1	A	A	A	A	A	A	A
8	Creating Gds Resource	16.2.2.2	B	B	B	B	B	B	C
9	Creating Fsystem Resource	16.2.2.3	C	C	B	B	B	B	C
10	Creating Gls Resource	16.2.2.4	C	B	C	B	C	B	C
11	Creating Temporal Cluster Applications	16.2.2.5	A	A	A	A	A	A	C
(3) Preparations for Migrating an Existing Solaris Environment to a Non-Global Zone		16.2.3							
12	Creating an Archive on the Migration Source Node	16.2.3.1	B	B	B	B	B	B	B
(4) Creating Non-Global Zones		16.2.4							
13	Creating Resource Pool	16.2.4.1	A	A	A	A	A	A	A
14	Creating Non-Global Zones	16.2.4.2	A	A	A	A	A	A	A
15	OS Installation to Non-Global Zones	16.2.4.3	A	A	A	A	A	A	A
16	Non-Global Zone Startup and OS Setup	16.2.4.4	A	A	A	A	A	A	A
17	Installing PRIMECLUSTER to Non-Global Zones	16.2.4.5	A	A	A	A	C	C	B
18	Global Zone Environment Setup (After Installation of PRIMECLUSTER to Non-Global Zones)	16.2.4.6	A	A	A	A	C	B	B
19	Setup of Web-Based Admin View for Non-Global Zones	16.2.4.7	A	A	A	A	C	C	B
20	Initial Setup of Non-Global Zone Cluster Resource Management Facility	16.2.4.8	A	A	A	A	C	C	B
21	Setup of GLS in Non-Global Zone	16.2.4.9	B	C	B	C	C	C	B
22	Installing Middleware Products to Non-Global Zones	16.2.4.10	B	B	B	B	C	C	B
23	RMS Setup for Non-Global Zones	16.2.4.11	A	A	A	A	C	C	A
24	Building Non-Global Zone Cluster Applications	16.2.4.12	A	A	A	A	C	C	B

No	Task	Explanatory Chapter	Task necessity for Each Configuration (Configuration number of 13.1.3)						
			1	2	3	4	5	6	7
25	Sharing Non-Global Zone Configuration Information	16.2.4.13	C	C	A	A	B	B	C
(5) Reconfiguration of Cluster Applications on Global Zone		16.2.5							
26	Deleting Temporal Cluster Applications	16.2.5.1	B	B	B	B	B	B	C
27	Creating the Cmdline Resource for Non-Global Zone Control	16.2.5.2	A	A	A	A	A	A	A
28	Creating the Cmdline Resource for Shared IP Control	16.2.5.3	C	B	C	C	C	C	C
29	Creating Cluster Applications	16.2.5.4	A	A	A	A	A	A	A

A:Required, B:Perform as Required, C:Unrequired
GLS: Global Link Services

16.2.1 Creating Global Zone

Perform this procedure with all of the nodes which comprise the cluster system.

16.2.1.1 OS Installation and Setup to the Global Zone

Install the OS to the global zone. For details, see Oracle Solaris documents.

- Disk-related settings

  If using a shared disk device, it will be necessary to install and set up the related software products.

  Perform this setup with the global zone before installing PRIMECLUSTER.

  For details, refer to "3.2.2 Setting Up Disk Units."

16.2.1.2 Installing Oracle Solaris 10 Zone Package to the Global Zone

If migrating an existing Solaris 10 environment to Zones on Solaris 11, use Oracle Solaris 10 zone package. Check if the system/zones/brand/brand-solaris10 package is installed. If not, install the package to the global zone for all nodes. For details on how to install it, see Oracle Solaris documents.

If migrating an existing Solaris 10 environment to Zones on Solaris 10, this task is not required.

16.2.1.3 Setup of NTP to the Global Zone

This setup is for synchronizing the time for each node which comprises the cluster system. It is absolutely necessary that this be performed when building a cluster of two nodes or more.

Perform this setup on the global zone before installing PRIMECLUSTER.

16.2.1.4 Check/Setup of Kernel Parameters to the Global Zone

If operating PRIMECLUSTER-related software, it is necessary to take into account the environment and adjust the OS kernel parameters.

Perform this setup with the global zone before performing the restart after "16.2.1.5 Installing PRIMECLUSTER to the Global Zone."

For details, refer to "3.2.3 Checking the Kernel Parameters."

16.2.1.5 Installing PRIMECLUSTER to the Global Zone

For details on the installation of PRIMECLUSTER, refer to "3.1 PRIMECLUSTER Installation."

16.2.1.6 Check/Setup of BE (Boot Environment) UUID of a Solaris 11 Global Zone

When you are using a Solaris 11 global zone, set the same UUID value for BE of the global zone on both active nodes and standby nodes.

1. Check UUID of BE on an active node.

```
# beadm list -H
UPDATE_BE;1ca3c7e2-fcdb-4d1e-ce92-c567dfd64969;NR;/;8292676096;static;1344328401
solaris;10857262-1e20-e712-ef07-fd9f2765f822;;;12295168;static;1344300070
newbe;04b7be53-779d-40aa-aeda-e7a39c212ebd;;;61440;static;1347415726
```

2. Check UUID of BE on a standby node.

```
# beadm list -H
UPDATE_BE;f4d8160d-4904-4857-c183-f2e47c28fc09;NR;/;8620887552;static;1344330302
solaris;10857262-1e20-e712-ef07-fd9f2765f822;;;12285952;static;1344298030
```

3. Compare UUID of BE of an active node with a standby node. If the value is different, set UUID of the active node in ZFS of the standby node.

```
# zfs set org.opensolaris.libbe:uuid=1ca3c7e2-fcdb-4d1e-ce92-c567dfd64969 rpool/ROOT/UPDATE_BE
# beadm list -H
UPDATE_BE;1ca3c7e2-fcdb-4d1e-ce92-c567dfd64969;NR;/;8620887552;static;1344330302
solaris;10857262-1e20-e712-ef07-fd9f2765f822;;;12285952;static;1344298030
```

4. If there is no BE which exits in an active node, create BE with the same name as the active node in the standby node, and then set UUID of the active node in ZFS of the standby node.

```
# beadm create newbe
# beadm list -H
UPDATE_BE;1ca3c7e2-fcdb-4d1e-ce92-c567dfd64969;NR;/;8620893696;static;1344330302
solaris;10857262-1e20-e712-ef07-fd9f2765f822;;;12285952;static;1344298030
newbe;32d4c688-fab2-6f2b-8f93-89c13be14144;;;61440;static;1347416270
# zfs set org.opensolaris.libbe:uuid=04b7be53-779d-40aa-aeda-e7a39c212ebd rpool/ROOT/newbe
# beadm list -H
UPDATE_BE;1ca3c7e2-fcdb-4d1e-ce92-c567dfd64969;NR;/;8621014528;static;1344330302
solaris;10857262-1e20-e712-ef07-fd9f2765f822;;;12285952;static;1344298030
newbe;04b7be53-779d-40aa-aeda-e7a39c212ebd;;;61440;static;1347416270
```

16.2.2 Creating Cluster Applications on the Global Zone

16.2.2.1 Building the Cluster

Refer to "Chapter 4 Preparation Prior to Building a Cluster" and "Chapter 5 Building a Cluster," and perform the initial setup of the cluster to the global zone.

Also, create the Gds resource, Fsystem resource, Gls resource, and cluster applications on the global zone as detailed below. For single-node cluster operations, creating these resources is not necessary. Perform the settings only for GDS or GLS as needed.

Creating temporal cluster applications is not necessary as well. Proceed to "16.2.3 Preparations for Migrating an Existing Solaris Environment to a Non-Global Zone."

16.2.2.2 Creating Gds Resources

This setup is necessary if sharing non-global zone images or if performing data takeover between non-global zones. Create on any one node a GDS shared class for each non-global zone and create to the shared classes a volume for non-global zone images and a volume for data takeover.

If using the system as a switchover file system, edit the /etc/vfstab.pcl for all nodes which comprise the cluster.

Example

If the GDS class for non-global zone zone-a is class0001, the volume for the non-global zone's root file system is volume 0001 (with the mount point being /zone-a-system, and the file system type being UFS), the volume for allocation of Oracle data being volume0002 (with the mount point being /zone-a-oracle, and the file system type being UFS), then write the /etc/vfstab.pcl as follows. Make sure the beginning of each line always starts with "#RMS#."

```
#RMS#/dev/sfdsk/class0001/dsk/volume0001  /dev/sfdsk/class0001/rdsk/volume0001 /zone-a-system ufs -
no -
#RMS#/dev/sfdsk/class0001/dsk/volume0002  /dev/sfdsk/class0001/rdsk/volume0002 /zone-a-oracle ufs -
no -
```

16.2.2.3 Creating Fsystem Resources

This setup is necessary if sharing GDS shared classes between non-global zones. If allocating non-global zones to a shared disk, also create Fsystem resources corresponding to the file systems into which the non-global zones are allocated. Perform this setup from any one node.

Note

To create non-global zones, set the mount point mode for the non-global zone root file system volume to 700 with all nodes. This setup is necessary for creating the non-global zone's root file system.

Example)

```
# chmod 700 /zone-a-system
```

16.2.2.4 Creating Gls Resources

If you want to notify the error to a cluster when it occurs with a network in a shared IP zone, set up a takeover virtual interface to the GLS and then register it as a Gls resource to the global zone user application.

Create the same number of takeover virtual interfaces as you want to notify a failure among virtual interfaces used in each shared IP zone.

Example

An example to meet the following two requirements is shown below:

- Three shared IP zones (zone-a, zone-b, and zone-c) hold two LANs: management LAN (192.168.10.0/24) and a public LAN (192.168.20.0/24).
- When a fault occurs to each LAN, the error is sent to the cluster.

In this case, you must perform the following procedure.

1. Create a total of six takeover virtual interfaces of GLS.

```
[Setup example of the takeover virtual interface]

# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+--------------------------------------------------+
 sha11:65   192.168.10.11    -
 sha12:65   192.168.10.12    -
 sha13:65   192.168.10.13    -
 sha21:65   192.168.20.11    -
 sha22:65   192.168.20.12    -
 sha23:65   192.168.20.13    -
```

2. Register Gls resources of management and public LANs for three cluster applications where each zone is registered.

```
[Resource registration for the user application]

 User application(uApp1)
   -Cmdline resource(controls non-global zone(zone-a))
   -Gls resource(management LAN(sha11: 192.168.10.11 of 65))
   -Gls resource(public LAN(sha21: 192.168.20.11 of 65))
 User application(uApp2)
   -Cmdline resource(controls non-global zone(zone-b))
   -Gls resource(management LAN(sha12:192.168.10.12 of 65))
   -Gls resource(public LAN(sha22: 192.168.20.12 of 65))
 User application(uApp3)
   -Cmdline resource(controls non-global zone(zone-c))
   -Gls resource(management LAN(sha13: 192.168.10.13 of 65))
   -Gls resource(public LAN(sha23: 192.168.20.13 of 65))
```

Note

- It is not possible to use the IP address for the takeover virtual interface registered as a Gls resource for shared IP zone transmission. Use for transmission the IP address allocated for the non-global zone by the Cmdline resource (shared IP control).
- If specifying an interface coordinated with the NIC switching mode to the non-global zone network setup, it is necessary to change the standby interface's deactivation method from "Unplumb" to "Plumb" using the hanetparam command's -d option.

```
# /opt/FJSVhanet/usr/sbin/hanetparam -d plumb

# /opt/FJSVhanet/usr/sbin/hanetparam print
   Line monitor interval(w)           :5
   Line monitor message output (m)    :0
   Cluster failover (l)               :5
   Standby patrol interval(p)         :15
   Standby patrol message output(o)   :3
   NIC switching mode(d)              :Plumb
   Cluster failover in unnormality (c):OFF
   Line status message output (s)     :OFF
```

16.2.2.5 Creating Temporal Cluster Applications

On the global zone, create temporal cluster applications in the same number as that of the number of non-global zones.

For the temporal cluster applications corresponding to a given non-global zone, register a file system resource (Fsystem) which corresponds to each volume included in the Gds resource, Gls resource, and GDS class which in turn correspond to that non-global zone.

Start each of the temporal cluster applications created in this section and set their status to Online status in the operational system.

16.2.3 Preparations for Migrating an Existing Solaris Environment to a Non-Global Zone

If migrating an existing Solaris 10 or Solaris 11 environment to the non-global zone, perform the following preparations. For details on the procedures, see Oracle Solaris documents.

16.2.3.1 Creating an Archive on the Migration Source Node

Create an archive in the migration source's Solaris 10 or Solaris 11 environment. For the method on creating the archive, the type of the supported archive, and notes on the migration, see Oracle Solaris documents. The created archive is used for inputting the zoneadm command when installing the OS to the non-global zone.

Example) To archive the UFS root file system in Solaris 10 environment.

```
# cd /
# flarcreate -S -n zone-a-system /var/tmp/zone-a-system.flar
current filter settings
Creating the archive...
8303358 blocks
Archive creation complete.
```

Transfer the created archive (zone-a-system.flar) to the desired folder in the global zone.

Information

When mirroring the system disk using GDS on the migration source node, it is not necessary to cancel the system disk mirroring when creating an archive.

16.2.4 Creating Non-Global Zones

This section describes procedures for building non-global zones. You need to perform this section's procedure the same number of times as the number of necessary non-global zones. To create non-global zones, the following conditions must be satisfied.

- Must be Whole Root Zone (a zone not sharing system files with the global zone)
- If using warm-standby or single-node cluster, the non-global zone autoboot (the setting to automatically start up the non-global zone during startup of the global zone; the default setting is false) must be set to true, whereas, if using cold-standby, it must be set to false.
- If using a configuration where non-global zone images are not shared between cluster nodes, the zone names (the names of zones as defined by zonecfg and zoneadm -z) must match for all nodes.

If not otherwise specified in the subsequent parts of this section, implement the procedures for only the operational system if one is using a configuration which shares non-global zone images between cluster nodes. If using a configuration which does not share non-global zone images between cluster nodes, implement the procedures in all nodes.

16.2.4.1 Creating the Resource Pool

For each creation of a non-global zone, create a resource pool beforehand. For details on the procedure, see Oracle Solaris documents.

If building a cluster with a Solaris Zones environment, make the number of CPU cores allocated to the global zone two or more.

See

If using ZFS in local classes, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

16.2.4.2 Creating the Non-Global Zone

Using the zonecfg command, create the non-global zones. Create them by referring to the following example.

```
# zonecfg -z zone-a    *1
```

*1: "zone-a" is the zone name (it is the same below)

```
zone-a: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:zone-a> create (if the global zone and non-global zone types are the same)
zonecfg:zone-a> create -t SYSsolaris10 (If using Oracle Solaris 10 Zones on Oracle Solaris 11)
zonecfg:zone-a> set zonepath=/zone-a-system    *2
zonecfg:zone-a> set autoboot=true (for warm-standby or single-node cluster)
zonecfg:zone-a> set autoboot=false (for cold-standby)
```

*2: For /zone-a-system, specify the directory to which the zone-a images are allocated.

If not sharing images, specify the file system on the local system.

If sharing images, specify the mountpoint registered as an Fsystem resource.

```
zonecfg:zone-a> set limitpriv="default,proc_priocntl"
zonecfg:zone-a> add fs
zonecfg:zone-a:fs> set dir=/oracle-data
zonecfg:zone-a:fs> set special=/zone-a-oracle    *3
zonecfg:zone-a:fs> set type=lofs
zonecfg:zone-a:fs> end
```

*3: For /zone-a-oracle, specify the directory of the Fsystem resource corresponding to the volume for zone-a Oracle data allocation.

```
zonecfg:zone-a> remove inherit-pkg-dir dir=/lib    *4
zonecfg:zone-a> remove inherit-pkg-dir dir=/platform   *4
zonecfg:zone-a> remove inherit-pkg-dir dir=/sbin *4
zonecfg:zone-a> remove inherit-pkg-dir dir=/usr   *4
```

*4: If creating non-global zones on Solaris 10 global zone, use "remove inherit-pkg-dir" and set things such that system files will not be inherited from the global zone and make this a whole root zone. For Solaris 11 global zone, this procedure is not required.

```
[If making the non-global zone's network mode a shared IP zone configuration]
zonecfg:zone-a> set ip-type=shared
zonecfg:zone-a> remove anet   *5
zonecfg:zone-a> add net (If making it a shared IP zone configuration)
zonecfg:zone-a:net> set physical=e1000g0   *6
```

*5: If creating a shared IP zone on Solaris 11 global zone, the anet needs to be removed after changing the jp-type. For Solaris 10 global zone, this procedure is not required.

*6: If specifying a network interface multiplexed with GLS, specify the Primary interface for the corresponding Gls resource.

```
zonecfg:zone-a:net> set address=10.20.30.40/24
zonecfg:zone-a:net> end
```

```
[If making the non-global zone's network mode an exclusive IP zone configuration]
zonecfg:zone-a> set ip-type=exclusive
zonecfg:zone-a> add net
zonecfg:zone-a:net> set physical=e1000g0   *7
zonecfg:zone-a:net> end
```

*7: Specify the physical interface exclusive to Zones. Perform the IP address setup and physical interface multiplexing from within Zones.

```
zonecfg:zone-a> add net
zonecfg:zone-a:net> set physical=e1000g1   *8
zonecfg:zone-a:net> end
```

*8: If multiplexing the physical interface within Zones, it is necessary to specify two or more physical interfaces.

```
zonecfg:zone-a> verify
zonecfg:zone-a> commit
zonecfg:zone-a> exit
```

See

For details, see the manual for the zonecfg command and also Oracle Solaris documents.

Note

If using a shared IP zone configuration

For the IP address set up to the non-global zone, set up and IP address which is not being used with GLS. Perform the setup of the default gateway to the zone on the global zone. If one has set up the default gateway using the zonecfg command, the paths will be disabled when performing NIC switching with GLS.

16.2.4.3 OS Installation to the Non-Global Zone

Install the OS to the non-global zone.

If newly installing Solaris 10 or Solaris 11 to the non-global zone, perform Step 1 as below. If creating the non-global zone from an archive, perform Step 2 as below. For details, see Oracle Solaris documents.

1. Newly Installing Solaris 10 or Solaris 11

    1. Check that the IPS package repository is set (only for Solaris 11)

        When installing Solaris to the non-global zone, the IPS package repository must have been set. Below is an example of checking that the IPS package repository has been set.

        ```
        # pkg publisher
        PUBLISHER                   TYPE     STATUS P LOCATION
        solaris                     origin   online F http://localhost/
        ```

    2. Install Solaris to the non-global zone using the zoneadm install command. Below is an example of installation of the non-global zone.

        ```
        # zoneadm -z zone-a install
        Preparing to install zone <zone-a>.
        Creating list of files to copy from the global zone.
        Copying <155078> files to the zone.
        Initializing zone product registry.
        Determining zone package initialization order.
        Preparing to initialize <1282> packages on the zone.
        Initialized <1282> packages on zone.
        Zone <zone-a> is initialized.
        Installation of <51> packages was skipped.
        The file </zone-a-system/root/var/sadm/system/logs/install_log> contains a log of the zone
        installation.
        ```

1. If creating the non-global zone from an archive

    Create the non-global zone from an archive using the zoneadm install command. Below is an example for creating the non-global zone.

    ```
    # zoneadm -z zone-a install -u -a /var/tmp/zone-a-system.flar
         Log File: /var/tmp/ zone-a-system.install.1987.log
           Source: /Etude/zone-a-system.flar
       Installing: This may take several minutes...
    ```

    If PRIMECLUSTER is not installed to the migration source environment, proceed to "16.2.4.4 Non-Global Zone Startup and OS Setup."

    If PRIMECLUSTER is installed to the migration source environment, uninstall the PRIMECLUSTER and perform the procedure below.

    Startup the non-global zone under single-user mode.

    ```
    # zoneadm -z zone-a boot -s
    ```

    Mount the medium of PRIMECLUSTER with the same version as the migration source in the global zone. For 4.2A00 or earlier, the mountpoint needs to be the directory which can be referred from the non-global zone.

    Log in to the non-global zone.

    ```
    # zlogin -C zone-a
    ```

Prepare for PRIMECLUSTER deletion in the non-global zone.

1. Check if files other than the class.db exist under the /etc/opt/FJSVsdx/sysdb.d. Moreover, check that the inside of the class.db is empty.

```
# cd /etc/opt/FJSVsdx/sysdb.d
# ls
class.db ...
```

   If files other than the class.db exist, delete by the rm command.

   If the information exists in the class.db file, delete the line.

   (The line which starts with # is a comment line.)

2. Check if files other than the _adm, _diag exist under the /dev/sfdsk.

```
# cd /dev/sfdsk
# ls
_adm _diag ...
```

   If files other than the _adm, _diag exist, delete by the rm-rf command.

3. If 4.2A00 or previous version of PRIMECLUSTER is installed at the migration source, remove the FJSVsdx (GDS Basic Software) package.

```
# pkgrm FJSVsdx
```

4. Back up the preremove and postremove files of SMAWcf package.

```
# cd /var/sadm/pkg/SMAWcf/install
# cp preremove /var/tmp/preremove.orig
# cp postremove /var/tmp/postremove.orig
```

5. Update the preremove and postremove files of SMAWcf package with the following procedure.

```
# cat /var/tmp/preremove.orig | sed -e 's,$CFCONFIG -u,echo " ",' | \
  sed -e 's,/opt/SMAW/bin/cipconfig -u,echo " ",' \
  > preremove
# cat /var/tmp/postremove.orig | sed -e 's,^module_id.*,module_id="",' | \
  sed -e 's,out=`rem_drv.*,out=" "; echo " ",' \
  > postremove
```

6. Delete the following descriptions from the /var/sadm/pkg/SMAWcf/install/preremove file for Solaris 11.

```
for drv in cip cf_drv mc1x
do
      drv_uninstall $drv
done
```

Uninstall PRIMECLUSTER from the non-global zone. Follow the "PRIMECLUSTER Installation Guide" to uninstall the PRIMECLUSTER with the same version as the migration source.
For 4.2A00 or earlier, execute the uninstallation script in the non-global zone. For the uninstallation script, use the uninstallation script of the medium of PRIMECLUSTER mounted beforehand.

If PRIMECLUSTER Wizard for Oracle, PRIMECLUSTER Wizard for NAS, and PRIMECLUSTER Wizard for NetWorker have been installed in the migration source environment, uninstall them from the non-global zone before uninstalling PRIMECLUSTER. For the procedure for uninstalling each Wizard product, see the installation guide of the same version as each Wizard product in the migration source.

Post-processing after the PRIMECLUSTER uninstallation is performed in the non-global zone.

Delete the backups of the preremove and postremove files created in the above procedure.

```
# rm /var/tmp/preremove.orig /var/tmp/postremove.orig
```

Stop the non-global zone.

```
# shutdown -y -g0 -i0
```

Unmount the medium of PRIMECLUSTER in the global zone.

16.2.4.4 Non-Global Zone Startup and OS Setup

Using the zoneadm boot command, start up the zone for which installation was performed. After that, acquire the console and perform Solaris OS setup (setup of node names, time zones, etc.).

```
# zoneadm -z zone-a boot
# zlogin -C zone-a
```

See

For details, see the manuals for the zoneadm command and zlogin command and also Oracle Solaris documents.

If monitoring applications in the non-global zone, proceed to "16.2.4.5 Installation of PRIMECLUSTER to the Non-Global Zone."

If not monitoring applications in the non-global zone, proceed to "16.2.4.13 Sharing Non-Global Zone Configuration Information."

16.2.4.5 Installation of PRIMECLUSTER to the Non-Global Zone

Perform installation of PRIMECLUSTER to the non-global zone. For details, refer to the "PRIMECLUSTER Installation Guide."

16.2.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone)

1. Revising Kernel Parameters

   This task is unnecessary if one is not installing PRIMECLUSTER to the non-global zone.

   Add the number of non-global zones to be created and make this the value for the kernel parameter detailed in the table below to the /etc/system for all of the **global zones** which comprise the cluster system. Then restart the global zone.

```
# shutdown -y -g0 -i6
```

Table 16.17 Kernel Parameters Requiring Revision

Kernel Parameters	Attribute	Value	Remarks
shmsys:shminfo_shmmni	Add	30	Performed in the resource database; the necessary value is per Zones
semsys:seminfo_semmni	Add	20	Performed in the resource database; the necessary value is per Zones

Note

- Do not delete the following definitions written in the **non-global zones'** /etc/system.

  set semsys:seminfo_semmni=30

  set shmsys:shminfo_shmmni=130

  set in_sync=1

- It will not cause any problems if the following definitions written in the non-global zones' /etc/system are deleted.

  msgsys:msginfo_msgmnb

  msgsys:msginfo_msgmni

  msgsys:msginfo_msgtql

2. Registering the GDS shared class volume

   This procedure is necessary if attempting to access the GDS shared class volume from the non-global zone.

   The procedure is different for Solaris 11 and Solaris 10.

   [For Solaris 11]

   Add the GDS shared class volume created in the global zone to the non-global zone, and then restart the non-global zone.

   Execute the following commands in the global zone.

```
# zonecfg -z zone-a
# zonecfg:zone-a>add device
# zonecfg:zone-a:device>set match=/dev/sfdsk/class0001/rdsk/volume0001
# zonecfg:zone-a:device>end
# zonecfg:zone-a>add device
# zonecfg:zone-a:device>set match=/dev/sfdsk/class0001/dsk/volume0001
# zonecfg:zone-a:device>end
# zonecfg:zone-a>verify
# zonecfg:zone-a>commit
# zonecfg:zone-a>exit
# zlogin zone-a shutdown -y -g0 -i6
```

   (If the zone name is zone-a, the class name is class0001, and the volume name is volume0001)

   [For Solaris 10]

   Copy the special file of the GDS shared class volume under /<zonepath>/dev.

   Execute the following commands in the global zone.

```
# cd /dev
# tar cvf /var/tmp/dsk.tar sfdsk/class0001/dsk/volume0001
# tar cvf /var/tmp/rdsk.tar sfdsk/class0001/rdsk/volume0001
# cd /zone-a-system/dev
# tar xvf /var/tmp/dsk.tar
# tar xvf /var/tmp/rdsk.tar
```

   (If the zonepath is /zone-a-system, the class name is class0001, and the volume name is volume0001)

   Note

   If you detached a non-global zone once, and attach it again by using attach subcommand of zoneadm command without specifying -F option, the GDS special file copied in the procedure above may be deleted by the OS specification. In this case, perform the above procedure again after attaching a non-global zone. When attaching a non-global zone by using attach -F subcommand of zoneadm command, the GDS special file is not be deleted.

3. Creating the file system

   For Solaris 11, specify the file system type to be mounted in the non-global zone, and then restart the non-global zone.

   Execute the following commands in the global zone. For Solaris 10, do not execute these commands.

```
# zonecfg -z zone-a
# zonecfg:zone-a> set fs-allowed=hsfs,nfs,ufs,zfs
# zonecfg:zone-a> verify
# zonecfg:zone-a> commit
# zonecfg:zone-a> exit
# zlogin zone-a shutdown -y -g0 -i6
```

   (If the zone name is zone-a, and the file system type is hsfs, nfs, ufs, or zfs)

   Regardless of the OS version, create the file system to the volume in the non-global zones.

   Execute the following command in the non-global zone.

```
# newfs /dev/sfdsk/class0001/rdsk/volume0001
```

(If the class name is class0001, the volume name is volume0001, and the file system is UFS)

Note

Perform the creation of the above file system only from the one node first used.

4. Setting the IP address of CIP

   When performing application monitoring, set the IP address of CIP according to the example below:

   - For shared IP zone

     Set up the following in the global zone.

     ```
     # zonecfg -z zone-a
     zonecfg:zone-a> add net
     zonecfg:zone-a:net> set address=127.0.0.2 *1
     zonecfg:zone-a:net> set physical=lo0
     zonecfg:zone-a:net> end
     zonecfg:zone-a> verify
     zonecfg:zone-a> commit
     zonecfg:zone-a> exit
     ```

   - For exclusive IP (Solaris 11)

     Execute the following command on the non-global zone. Set up the following in all non-global zones.

     ```
     # ipadm create-addr -T static -a local=127.0.0.2/8 lo0/cip *1
     ```

   - For exclusive IP (Solaris 10)

     Create /etc/hostname.lo0:1(*2) and enter the following.

     ```
     127.0.0.2 *1
     ```

     *1) Specify a loopback address which is not used by the system.
     The same loopback address cannot be used by multiple non-global zones that exist on one global zone. Use different loopback addresses for each non-global zone. (Example: 127.0.0.3, 127.0.0.4)

     *2) Use a non-existent file name.

   Add the address specified in /etc/inet/hosts after finishing above settings.

   ```
   127.0.0.2 xxxRMS
   ```

   "xxx" is the CF node name of non-global zone, which can be checked with cftool -l command on the non-global zone.

16.2.4.7 Setup of Web-Based Admin View for the Non-Global Zone

Perform this task in the non-global zone.

Refer to "4.2.3 Initial Setup of Web-Based Admin View," and perform the setup and startup for Web-Based Admin View. When doing so, specify the same non-global zone IP addresses as those for both the primary management server and the secondary management server specified with "4.2.3.1 Initial setup of the operation management server."

(Example: If the non-global zone IP address is 10.20.30.40)

```
# /etc/init.d/fjsvwvcnf stop
# /etc/init.d/fjsvwvbs stop
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server 10.20.30.40
```

```
# /etc/opt/FJSVwvbs/etc/bin/wvCntl start
# /etc/init.d/fjsvwvcnf start
```

After setup, use the procedure "4.3 Starting the Web-Based Admin View Screen" to confirm that one is able to start up the GUI screen.

16.2.4.8 Initial Setup of the Non-Global Zone Cluster Resource Management Facility

When connecting to the non-global zone set up with 16.2.4.7 Setup of Web-Based Admin View for the Non-Global Zone and starting up the Web-Based Admin View screen, refer to "5.1.3 Initial Setup of the Cluster Resource Management Facility" and "5.1.3.1 Initial Configuration Setup," and perform the initial configuration setup for the cluster resource management facility.

It is not necessary to perform CF and CIP setup, shutdown facility setup, or automatic configuration for the non-global zone.

Note

If performing initial configuration setup for the cluster resource management facility, the message below will be output onto the non-global zone console, but this will not be a problem for its operation.

/dev/rdsk/*: No such file or directory

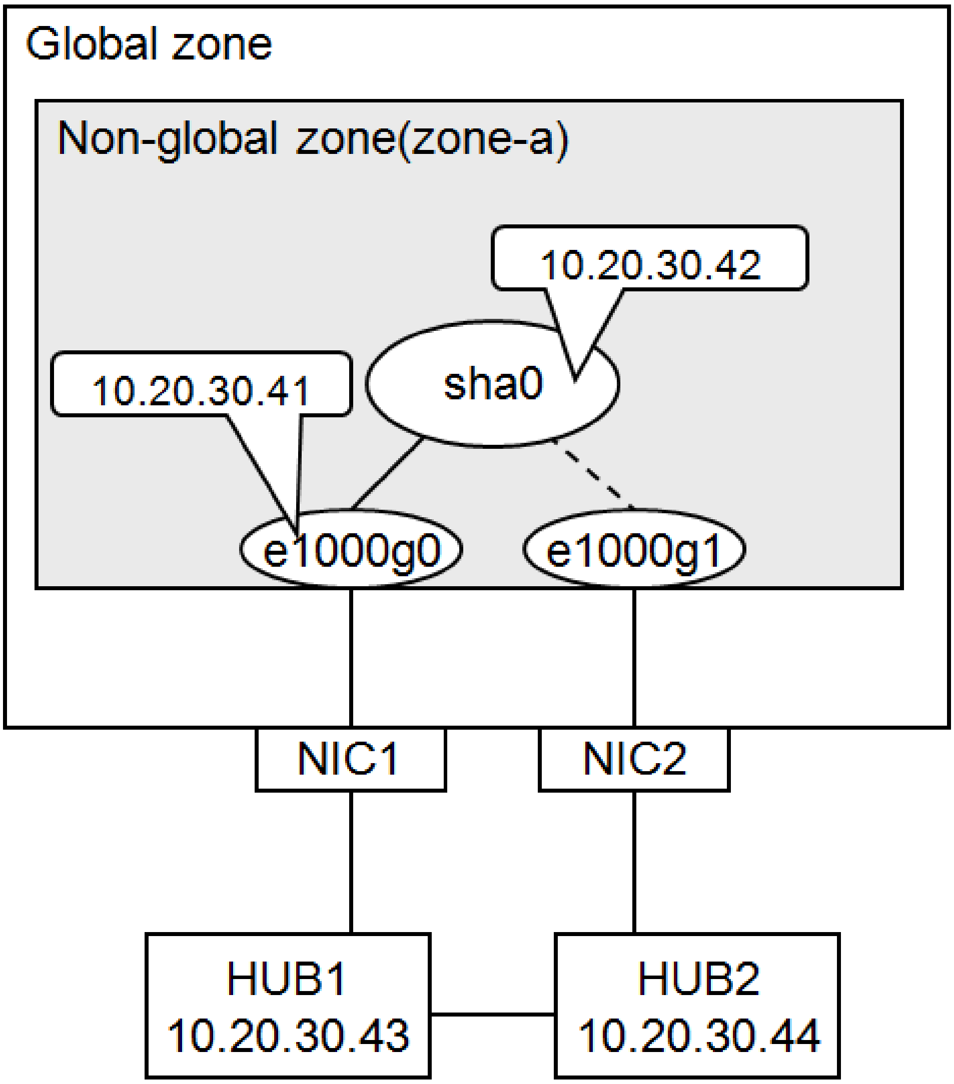
Also, if initial configuration setup failed, it is possible that the non-global zone kernel parameters were insufficient. Refer to the "A.5 Kernel Parameter Worksheet" and correct the kernel parameter value. After restarting the non-global zone, perform resource database initialization using the clinitreset (1M) command and re-perform the initial configuration setup.

16.2.4.9 Setup of GLS in a Non-Global Zone

This procedure is necessary only if one is using the NIC switch mode with an exclusive IP zone configuration. If setting up GLS on a non-global zone, refer to the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function" and perform the setup for multiplexing the physical interface.

Perform this section's tasks in all of the non-global zones which are to build the cluster system.

Figure 16.14 Example of an Environment Setup for if Configuring Between the Non-Global Zones with a Warm-standby Configuration



1. System settings

   1-1) Define the IP address to be used and the host name to the /etc/inet/hosts file.

```
10.20.30.42 zone-a0    # zone-a virtual IP(takeover IP)
10.20.30.41 zone-a01   # zone-a physical IP
10.20.30.43 swhub1     # primary monitoring destination HUB IP
10.20.30.44 swhub2     # secondary monitoring destination HUB IP
```

   Note

   Setup the zone-a physical IP address such that it does not overlap with other non-global zone physical IP addresses.

   1-2) Write the host name defined above to the /etc/hostname.e1000g0 file.

   Content of /etc/hostname.e1000g0

```
zone-a01
```

   1-3) Define the subnet mask to the /etc/inet/netmasks file.

```
10.20.30.0 255.255.255.0
```

2. Reboot

   Execute the following command and reboot the non-global zone. Perform this command from the global zone. After reboot, execute the ifconfig command to confirm that the e1000g0 is activated.

```
# /usr/sbin/zlogin zone-a shutdown -y -g0 -i6
```

3. Creating the virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.20.30.42 -e 10.20.30.41 -t e1000g0,e1000g1
```

   Note

   Always be sure that the physical IP address defined to the option "-e" matches with the physical IP address set up to the /etc/hostname.e1000g0.

4. Setup of standby patrol function

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -t sha0
```

   Information

   For GLS4.3A10 or later, the -a option can be omitted. In that case, the settings below are performed automatically.

   - The environment the MAC addresses of the active NIC and standby NIC are the same.

     The local MAC address based on the global address.

   - The environment MAC addresses of the active NIC and standby NIC are different.

     0:0:0:0:0:0

5. Setup of HUB monitoring function

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.20.30.43,10.20.30.44 -b off
```

6. Creating the takeover virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

   Note

   This settings are not necessary for single-node cluster operations.

7. Starting HUB monitoring

```
# /opt/FJSVhanet/usr/sbin/hanetpoll on
```

16.2.4.10 Installing Middleware Products to Non-Global Zones

For the installation procedure and points of caution for each middleware product, refer to the respective middleware product manual.

16.2.4.11 Setup of Non-Global Zone RMS

Edit the file "/opt/SMAW/SMAWRrms/bin/hvenv.local" as follows. If the file "/opt/SMAW/SMAWRrms/bin/hvenv.local" does not exist, create the file (create the file access privilege in 644), and write in the line shown in [After changes].

- For cluster of two or more nodes in the global zone

```
export HV_CF_PACKAGE=none
export HV_USE_ELM=0
export HV_RCSTART=0
export RELIANT_HOSTNAME=<CF node name>RMS
```

- For single node cluster in the global zone

```
export HV_CF_PACKAGE=none
export HV_USE_ELM=0
export HV_RCSTART=1
export RELIANT_HOSTNAME=<CF node name>RMS
```

You can check the CF node name with the cftool -n command.

Example

When the CF node name is "zone-a"

```
# cftool -n
Node    Number State       Os      Cpu
zone-a 1       UP          Solaris Sparc
```

16.2.4.12 Setup of Non-Global Zone Cluster Applications

This section explains the procedure for creating cluster applications on the non-global zone.

Perform the following procedure taking into account the cluster resources that are to be set up.

No.	Task Overview	Procedure necessary to configuration						
		1	2	3	4	5	6	7
1	Setup of the Cmdline resource	A	A	A	A	B	B	A
2	Setup of the Oracle resource	A	A	A	A	B	B	A
3	Setup of the NetWorker resource	A	A	A	A	B	B	A
4	Setup of the Netapp resource	A	A	A	A	B	B	A
5	Setup of the state transition procedure resources	A	A	A	A	B	B	A
6	Setup of the Gls resource	A	B	A	B	B	B	B
7	Setup of the Fsystem resource	A	A	B	B	B	B	A
8	Creation of the cluster applications	A	A	A	A	B	B	A

A: Perform as required, B: Unrequired

1. Setup of the Cmdline resource

   For the method for setting up the Cmdline resource, refer to "6.7.1.1 Creating Cmdline Resources."

2. Setup of the Oracle resource

   Refer to the "PRIMECLUSTER Wizard for Oracle Configuration and Administration Guide" and perform the setup of the cluster resource.

3. Setup of the Netapp resource

   Refer to the " PRIMECLUSTER Wizard for NetWorker 4.2 Configuration and Administration Guide" and perform the setup of the cluster resource.

4. Setup of the Netapp resource

   Refer to the " PRIMECLUSTER Wizard for NAS Configuration and Administration Guide" and perform the setup of the cluster resource.

5. Setup of the state transition procedure resources

   Refer to the middleware manual and set up the state transition procedure resources. As for the availability of middleware products and PRIMECLUSTER in combination in a non-global zone, contact field engineers.

6. Setup of the Gls resource

   This procedure is necessary only if one is using an exclusive IP zone configuration.

   Refer to "6.7.1.4 Creating Gls Resources" and perform the setup of the Gls resource.

7. Setup of the Fsystem resource

   This procedure is necessary if using a switching file system with a non-global zone.

   Refer to "6.7.1.2 Creating Fsystem Resources" and perform the setup of the Fsystem resource.

   Note that you cannot set ZFS for Fsystem resources in non-global zones.

8. Creation of the cluster applications

   Create the cluster applications on non-global zones.

   For the method for creating the cluster applications, follow "6.7.2.1 Creating Standby Cluster Applications." However, there are the following differences in procedure:

   - Cluster application attributes

     - Set No to AutoStartUp, AutoSwitchOver, and HaltFlag. However, when the global zone is operated on a single-node cluster, set Yes to AutoStartUp.

     - If on a warm-standby configuration and wishing to put the cluster applications on the standby system's non-global zone into Standby mode, set the ClearFaultRequest to StandbyTransitions. For all other circumstances, set No to it.

     - Set NONE to Shutdown Priority.

     - Set 0 to Online Priority.

16.2.4.13 Sharing Non-Global Zone Configuration Information

If using cold-standby, stop the non-global zones in the operational system nodes.

```
# zlogin zone-a shutdown -i0 -g0 -y
```

If sharing non-global zone images in cold-standby operation, make it so that one is able to use the information for the non-global zones created thus far from the standby system's nodes as well.

Export the non-global zone configuration information with the operational system node.

```
# zonecfg -z zone-a export -f /var/tmp/zone-a.exp
```

Copy the output file (in the example above /var/tmp/zone-a.exp) to the standby system nodes.

Import the non-global zone into the standby system nodes.

```
# zonecfg -z zone-a -f /var/tmp/zone-a.exp
```

Note

When performing import, since it is not necessary to access the non-global zone's file system, do not perform an operation with the standby system nodes making the cluster application Online. Also, do not perform an operation which attaches or starts up the non-global zone.

16.2.5 Reconfiguration of Cluster Applications on Global Zone

Build on the global zone the cluster applications which correspond to each non-global zone. Perform this section's procedure to each non-global zone.

16.2.5.1 Deleting Temporal Cluster Applications

Stop the RMS if it is running.

If it is not single-node cluster operations, delete the temporal cluster applications created with "16.2.2.5 Creating Temporal Cluster Applications." Do not perform deletion of the resources when deleting the cluster applications.

16.2.5.2 Creating the Cmdline Resource for Non-Global Zone Control

Create the Cmdline resource for controlling from the global zone the non-global zone and the cluster applications configured to the non-global zone.

- If performing application monitoring within the non-global zone. (except single node cluster operations)

  Create the Cmdline resource.

  Select "Path Input" from the "Creation Method" for creating the Cmdline and set up the Start script, Stop script, and Check script as follows.

  - Start script

    ```
    /opt/SMAW/bin/hvzone -c -z <zone_name> -a <app_name> {-s|-n} -t <timeout>
    ```

  - Stop script

    ```
    /opt/SMAW/bin/hvzone -u -z <zone_name> -a <app_name> {-s|-n} -t <timeout>
    ```

  - Check script

    ```
    /opt/SMAW/bin/hvzone -m -z <zone_name> -a <app_name> {-s|-n} -t <timeout>
    ```

    The differences above are only with the -c, -u, and -m options. Specify the name of the zone that is the target for control and the name of the cluster application set up in the non-global zone for <zone_name> and <app_name> respectively.

  Specify in seconds the shutdown process time out value for <timeout>. During Offline processing, this script performs RMS stop and non-global zone stop, but if the processing does not stop even after the time designated here has elapsed, stop the non-global zone using the halt command for zoneadm (zoneadm -z <zone_name> halt).

  If sharing non-global zones between cluster nodes, specify the -s option. If not sharing, specify the -n option.

  An example is given below. In this example, the configuration is as follows: the zone name is zone1, the cluster application name in the non-global zone is userApp_0, the timeout value is 200 seconds, and the non-global zone is shared between cluster nodes in the global zone.

  - Start script

    ```
    /opt/SMAW/bin/hvzone -c -z zone1 -a userApp_0 -s -t 200
    ```

  - Stop script

    ```
    /opt/SMAW/bin/hvzone -u -z zone1 -a userApp_0 -s -t 200
    ```

  - Check script

    ```
    /opt/SMAW/bin/hvzone -m -z zone1 -a userApp_0 -s -t 200
    ```

  If one specifies the -s option, the non-global zone on the standby system side will not start and moreover the non-global zone will enter a status of being detached from that node.

  After that, set up the script attributes. Click the "Flag" button and set the following values.

Flag	Overview
ALLEXITCODES	Change this to "Yes."
AUTORECOVER	When "Yes" is specified, do not set the following resources. - Gls resource on the non-global zone - Cmdline resource (shared IP control) on the global zone (For details, see "16.2.5.3 Creating the Cmdline Resource for Shared IP Control.")
STANDBYCAPABLE	If using warm-standby, change this to "Yes."
TIMEOUT	The default value is 300 seconds. Set a value larger than the total of the following values. - The time for starting up the non-global zone (the time for the entire startup sequence to finish) - The time it takes for the cluster applications defined in the non-global zone to enter Online status

- If not conducting application monitoring within the non-global zone or single-node cluster is operated;

  Create the Cmdline resource.

  Create resource script files. This script is created for each Cmdline resource. Moreover, script files are created for all nodes which use Cmdline resources. The following examples show that Start script is set for"/var/tmp/PCL/rmstools/start_zone.sh", Stop script is set for "/var/tmp/PCL/rmstools/stop_zone.sh", and check script is set for "/var/tmp/PCL/rmstools/check_zone.sh". The contents can be modified according to the elements. If the /var/tmp/PCL/rmstools directory does not exist, create the directory.

  - Create script file

```
# vi /var/tmp/PCL/rmstools/start_zone.sh

(Paste the following the content of Start script file)

# vi /var/tmp/PCL/rmstools/stop_zone.sh

(Paste the following the content of Stop script file)

# vi /var/tmp/PCL/rmstools/check_zone.sh

(Paste the following the content of Check script file)

# chmod +x /var/tmp/PCL/rmstools/start_zone.sh
# chmod +x /var/tmp/PCL/rmstools/stop_zone.sh
# chmod +x /var/tmp/PCL/rmstools/check_zone.sh
```

  - Start script

```
#!/bin/sh

MYZONE=$1

zoneadm -z $MYZONE list -p | grep :configured:
if [ $? -eq 0 ]; then
    zoneadm -z $MYZONE attach -F || exit $?
fi

zoneadm -z $MYZONE list -p | grep :running:
if [ $? -eq 0 ]; then
    zoneadm -z $MYZONE reboot
    RET=$?
else
    zoneadm -z $MYZONE boot
    RET=$?
```

```
fi

exit $RET
```

- Stop script

```
#!/bin/sh

MYZONE=$1

RET=0
RET2=0

zoneadm -z $MYZONE list -p | grep :running:
if [ $? -eq 0 ]; then
    zoneadm -z $MYZONE halt
    RET=$?
fi

zoneadm -z $MYZONE list -p | grep :installed:
if [ $? -eq 0 ]; then
    zoneadm -z $MYZONE detach
    RET2=$?
fi

if [ $RET -eq 0 ]; then
  exit $RET2
fi
exit $RET
```

- Check script

```
#!/bin/sh

# Return Offline if zlogin to the NGZ does not end in 30 seconds (Please change if needed)
TIMEOUT=30
MYZONE=$1

zoneadm -z $MYZONE list -p | grep :running: > /dev/null 2>&1
RET=$?

if [ $RET -ne 0 ]; then
        exit $RET
fi

/usr/sbin/zlogin $MYZONE "/usr/bin/ls >/dev/null 2>&1" 2>/dev/null &
PID=$!

i=0
while [ $i -lt $TIMEOUT ]
do
        ps -p $PID > /dev/null 2>&1
        if [ $? -ne 0 ]; then
                wait $PID
                exit $?
        fi
        sleep 1
        i=`expr $i + 1`
done
exit 1
```

Select "Path Input" from the "Creation Method" for creating the Cmdline and set up the Start script, Stop script, and Check script as follows. (When the zone name is zone1.)

- Start script

```
/var/tmp/PCL/rmstools/start_zone.sh zone1
```

- Stop script

```
/var/tmp/PCL/rmstools/stop_zone.sh zone1
```

- Check script

```
/var/tmp/PCL/rmstools/check_zone.sh zone1
```

Then set attributes for the scripts. Click the *Flag* button and set the value below.

Flag	Overview
AUTORECOVER	- For single-node cluster operation, set to "Yes." - When "Yes" is specified, do not set the following resources. These resources are used to take over the IP address among several nodes, so they are not required for the single-node cluster. - Cmdline resource (shared IP control) on the global zone (For details, see "5-3. Creating the Cmdline Resource for Shared IP Control."

16.2.5.3 Creating the Cmdline Resource for Shared IP Control

If using a shared IP zone non-global zone but not sharing non-global zone images, create a resource for the takeover IP address in the global zone.

1. Create the script file for all nodes which use Cmdline resource. If the /var/tmp/PCL/rmstools directory does not exist, create the directory.

   - Create hvzonelip.sh script file (when using GLS NIC switching mode, or not using GLS)

   ```
   # vi /var/tmp/PCL/rmstools/hvzonelip.sh

   (Paste the content of the following hvzonelip.sh script file)

   # chmod +x /var/tmp/PCL/rmstools/hvzonelip.sh
   ```

   - Create hvzonelifrsc.sh script file (when using GLS fast switching mode or GS/SURE linkage mode)

   ```
   # vi /var/tmp/PCL/rmstools/hvzonelifrsc.sh

   (Paste the following the content of hvzonelifrsc.sh script file)

   # chmod +x /var/tmp/PCL/rmstools/hvzonelifrsc.sh
   ```

   - Content of hvzonelip.sh script file (if being used with an NIC switching mode for GLS or if not using GLS)

   ```
   #!/bin/sh
   #
   # hvzonelip.sh 1.1
   #

   LIP=""
   IPMASK=""
   IFLIST=""
   IPTYPE=""
   ZONENAME=""
   ZONEWAIT=""

   #------------------------------------------------
   ```

```
# Add Logical IP
#---------------------------------------------------

TAG=`/usr/bin/basename $0`
CONFFILE=$2
RSCNAME=${HV_NODENAME:-Unknown}
PHASE=${HV_INTENDED_STATE:-Unknown}

Start()
{
    ADDINGIP=$LIP
    IFATTR="netmask + broadcast + -trailers"
    IFCFG_RETRY_CNT=4

    if [ "$IPTYPE" = "inet6" ]
    then
        ADDINGIP="$LIP"/"$IPMASK"
        IFATTR=""
    fi

    for PIF in $IFLIST
    do
        /usr/sbin/ifconfig $PIF $IPTYPE 2>/dev/null | /bin/grep "UP," >/dev/null 2>&1
        if [ $? -eq 0 ]
        then
            break
        fi
    done

    CheckIp $PIF $ADDINGIP $ZONENAME
    if [ $? != 0 ]
    then
       Pmsg INFO "ip address has already been assigned. if=$CHKIF LIP=$LIP ZONENAME=$ZONENAME"
        return 0
    fi

    RET=1
    CNT=0
    while [ $RET -ne 0 ] && [ $CNT -le $IFCFG_RETRY_CNT ]

    do
        CNT=`/usr/bin/expr $CNT + 1`
        CMD="/usr/sbin/ifconfig $PIF $IPTYPE addif $ADDINGIP $IFATTR up zone $ZONENAME"
        Pmsg TRACE "$CMD"
        $CMD
        RET=$?
        Pmsg TRACE "exitcode = $RET"
    done

    if [ $RET -ne 0 ]
    then
        Pmsg ERROR "ifconfig failed. cmd=$CMD"
    fi

    return $RET
}

CheckIp()
{
    CHKPIF=$1
    CHKLIP=$2
    CHKZONE=$3
    CHKIF=
```

```
    CHKIFCAU=`/usr/sbin/ifconfig -au $IPTYPE 2>/dev/null`
    echo "$CHKIFCAU" | /bin/grep "$IPTYPE $CHKLIP " >/dev/null 2>&1
    if [ $? != 0 ]
    then
        return 0
    fi

    CHKIFLIST=`echo "$CHKIFCAU" | /bin/grep "^$CHKPIF:[0-9]*: " | /usr/bin/sed -e "s/: .*//"`
    for CHKIF in $CHKIFLIST
    do
        CHKIFC=`/usr/sbin/ifconfig $CHKIF $IPTYPE 2>/dev/null`
        echo "$CHKIFC" | /bin/grep "$IPTYPE $CHKLIP " >/dev/null 2>&1
        if [ $? != 0 ]
        then
            continue
        fi
        echo "$CHKIFC" | /bin/grep "zone $CHKZONE$" >/dev/null 2>&1
        if [ $? != 0 -a "$CHKZONE" != "global" ]
        then
            continue
        fi
        return 1
    done

    return 0
}

#--------------------------------------------------
# Remove Logical IP
#--------------------------------------------------

Stop()
{
    IFCFG_RETRY_CNT=9

    for PIF in $IFLIST
    do
        RET=0
        CNT=0
        while [ $RET -eq 0 ] && [ $CNT -le $IFCFG_RETRY_CNT ]
        do
            CNT=`/usr/bin/expr $CNT + 1`
            CMD="/usr/sbin/ifconfig $PIF $IPTYPE removeif $LIP"
            Pmsg TRACE "$CMD"
            $CMD 2>/dev/null
            RET=$?
            Pmsg TRACE "exitcode = $RET"
        done
    done

    return 0
}

#--------------------------------------------------
# Common functions
#--------------------------------------------------

WaitForZoneBoot()
{
    MSGFLAG=0

    while [ 1 ]
```

```
    do
       ZONESTATUS=`/usr/sbin/zoneadm -z "$ZONENAME" list -p 2>/dev/null | /bin/grep ":shared"
| /usr/bin/nawk -F: '{ print $3 }'`
        if [ "$ZONESTATUS" = "ready" -o "$ZONESTATUS" = "running" ]
        then
            break
        fi
        if [ "$ZONESTATUS" = "" ]
        then
            Pmsg ERROR "shared-ip zone does not exist. ZONENAME=$ZONENAME"
            return 1
        fi
        if [ "$MSGFLAG" = 0 ]
        then
            Pmsg INFO "wait for zone boot. stat=$ZONESTATUS ZONENAME=$ZONENAME"
            MSGFLAG=1
        fi
        sleep 1
    done

    if [ "$MSGFLAG" = 1 ]
    then
        Pmsg INFO "finished waiting. stat=$ZONESTATUS ZONENAME=$ZONENAME"
    fi

    return 0
}

Pmsg()
{
    TSTAMP=`/usr/bin/date +"%Y-%m-%d %H:%M:%S"`
    echo "$RSCNAME: $PHASE: $TSTAMP: $1: $TAG($$) $2"
    return 0
}

ReadConf()
{
    if [ -r "$CONFFILE" ]
    then
        . $CONFFILE
    fi
    if [ -z "$LIP" -o -z "$IPMASK" -o -z "$IFLIST" -o -z "$IPTYPE" -o -z "$ZONENAME" ]
    then
       Pmsg ERROR "configuration error. LIP=$LIP IPMASK=$IPMASK IFLIST=$IFLIST IPTYPE=$IPTYPE
ZONENAME=$ZONENAME"
        return 1
    fi
    return 0
}

#---------------------------------------------------
# Main
#---------------------------------------------------

if [ x"$1" = x"start" ]
then
    ReadConf
    if [ $? != 0 ]
    then
        exit 1
    fi
    if [ "$ZONEWAIT" != 0 ]
    then
```

```
        WaitForZoneBoot
        if [ $? != 0 ]
        then
            exit 1
        fi
    fi
    Start
    exit $?
elif [ x"$1" = x"stop" ]
then
    ReadConf
    if [ $? != 0 ]
    then
        exit 0
    fi
    Stop
    exit $?
else
    Pmsg ERROR "usage: $TAG {start|stop} [conffile]. cmd=$0 $*"
fi

exit 1
```

- Content of the hvzonelifrsc.sh script file (if being used with fast switching mode for GLS or GS/SURE linkage mode)

```
#!/bin/sh
#
# hvzonelifrsc.sh 1.0
#

LIF=""
ZONENAME=""
ZONEWAIT=""

#---------------------------------------------------
# Place the logical interface in non-global zone
#---------------------------------------------------

TAG=`/usr/bin/basename $0`
CONFFILE=$2
RSCNAME=${HV_NODENAME:-Unknown}
PHASE=${HV_INTENDED_STATE:-Unknown}

Start()
{
    IFCFG_RETRY_CNT=4

    RET=1
    CNT=0
    while [ $RET -ne 0 ] && [ $CNT -le $IFCFG_RETRY_CNT ]
    do
        CNT=`/usr/bin/expr $CNT + 1`
        CMD="/usr/sbin/ifconfig $LIF zone $ZONENAME"
        Pmsg TRACE "$CMD"
        $CMD
        RET=$?
        Pmsg TRACE "exitcode = $RET"
    done

    if [ $RET -ne 0 ]
    then
        Pmsg ERROR "ifconfig failed. cmd=$CMD"
    fi
```

```
    return $RET
}

#-------------------------------------------------
# Place the logical interface in global zone
#-------------------------------------------------

Stop()
{
    IFCFG_RETRY_CNT=4

    RET=1
    CNT=0
    while [ $RET -ne 0 ] && [ $CNT -le $IFCFG_RETRY_CNT ]
    do
        CNT=`/usr/bin/expr $CNT + 1`
        CMD="/usr/sbin/ifconfig $LIF -zone"
        Pmsg TRACE "$CMD"
        $CMD
        RET=$?
        Pmsg TRACE "exitcode = $RET"
    done

    return 0
}

#-------------------------------------------------
# Common functions
#-------------------------------------------------

WaitForZoneBoot()
{
    MSGFLAG=0

    while [ 1 ]
    do
       ZONESTATUS=`/usr/sbin/zoneadm -z "$ZONENAME" list -p 2>/dev/null | /bin/grep ":shared"
|\
/usr/bin/nawk -F: '{ print $3 }'`
        if [ "$ZONESTATUS" = "ready" -o "$ZONESTATUS" = "running" ]
        then
            break
        fi
        if [ "$ZONESTATUS" = "" ]
        then
            Pmsg ERROR "shared-ip zone does not exist. ZONENAME=$ZONENAME"
            return 1
        fi
        if [ "$MSGFLAG" = 0 ]
        then
            Pmsg INFO "wait for zone boot. stat=$ZONESTATUS ZONENAME=$ZONENAME"
            MSGFLAG=1
        fi
        sleep 1
    done

    if [ "$MSGFLAG" = 1 ]
    then
        Pmsg INFO "finished waiting. stat=$ZONESTATUS ZONENAME=$ZONENAME"
    fi

    return 0
```

```
}

Pmsg()
{
    TSTAMP=`/usr/bin/date +"%Y-%m-%d %H:%M:%S"`
    echo "$RSCNAME: $PHASE: $TSTAMP: $1: $TAG($$) $2"
    return 0
}

ReadConf()
{
    if [ -r "$CONFFILE" ]
    then
        . $CONFFILE
    fi
    if [ -z "$LIF" -o -z "$ZONENAME" ]
    then
        Pmsg ERROR "configuration error. LIF=$LIF ZONENAME=$ZONENAME"
        return 1
    fi
    return 0
}

#---------------------------------------------------
# Main
#---------------------------------------------------

if [ x"$1" = x"start" ]
then
    ReadConf
    if [ $? != 0 ]
    then
        exit 1
    fi
    if [ "$ZONEWAIT" != 0 ]
    then
        WaitForZoneBoot
        if [ $? != 0 ]
        then
            exit 1
        fi
    fi
    Start
    exit $?
elif [ x"$1" = x"stop" ]
then
    ReadConf
    if [ $? != 0 ]
    then
        exit 0
    fi
    Stop
    exit $?
else
    Pmsg ERROR "usage: $TAG {start|stop} [conffile]. cmd=$0 $*"
fi

exit 1
```

2. Create a new configuration file for the allocated script. Any file names can be specified. Moreover, set the following items for the created files.

   - Create configuration files of hvzonealip.sh (when using NIC switching mode of GLS, or GLS is not used)
     Configure the five items; LIP, IPMASK, IFLIST, IPTYPE, and ZONENAME.

     - /var/tmp/PCL/rmstools/hvzonelip_zone-net.conf

```
LIP=
IPMASK=
IFLIST=
IPTYPE=
ZONENAME=
```

   - Create configuration files of hvzonelifrsc.sh (fast switching mode of GLS or GS/SURE linkage mode is used)
     Configure two items, LIF and ZONENAME.

     - /var/tmp/PCL/rmstools/hvzonelifrsc_zone-net.conf

```
LIF=
ZONENAME=
```

Item	Setting Details
Takeover IP Address (LIP=)	In the non-global zone, set up the takeover IP address. Set up for this IP address one that is not being used with GLS, takeover network resources, or other non-global zones. Since you need to consider that address characters may be omitted if you set an IPv6 address, set the same address characters displayed when the ifconfig is executed while the IPv6 address is set for a network interface. Example: "192.168.10.10", "fd00:100::2e0:edff:fe0a:f30a"
Netmask Length (IPMASK=)	Set the netmask length of LIP (for IPv4) or the prefix length (for IPv6). Example: "24", "64
List of Network Interface Names (IFLIST=)	Set up the name of the network interface which is on the global zone being used by the non-global zone. If linking multiple network interfaces with GLS, write both the primary and secondary network interfaces with a space separating them. Example: "fjgi0 fjgi1," "fjgi0"
Type of IP Address (IPTYPE=)	Set inet if LIP is IPv4 and set inet6 if LIP is IPv6. Example: "inet", "inet6"
Takeover Logical Virtual Interface (LIF=)	In the non-global zone, set up the logical virtual interface (the interface with colons) for which the takeover IP address (IPv4 address) is allocated. Example: "sha0:65"
Non-Global Zone Name (ZONENAME=)	Set up the name of the non-global zone. Example: "zone1"
Non-Global Zones Wait (ZONEWAIT=)	When allocating an IP address to a non-global zone, specify to check the status of the target non-global zone or not. If nothing is set, starting non-global zone is postponed until TIMEOUT for Cmdline resource passes at a maximum. Without waiting, to make resources errors immediately, add "ZONEWAIT="0"" on the configuration file. Example: "0"

GLS: Global Link Services

Information

- If using the NIC switching mode or GLS is not used, execute IP address takeover between the non-global zones by adding an IP address which differs from the takeover IP address registered to the Gls resource or the takeover network resource. However, if using fast switching mode or GS/SURE linkage mode, execute IP address takeover between the non-global zones by sharing with the global zone the takeover IP address registered to the Gls resource.

- If using fast switching mode or GS/SURE linkage mode, set up the takeover logical virtual interface by adding the -i option to the hanethvrsc create command for GLS and executing this command. Also, by using the hanethvrsc print command, it is displayed as an ifname item.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------------+
 sha0:65    192.168.100.101  -
```

- If using the GS/SURE linkage mode, it is not possible to use the takeover IP address in non-global zone if one has performed setup without adding the -i option to the hanethvrsc create command for GLS. The takeover IP is for allocation to the virtual interface (shaX), not the logical virtual interface that can be shared between the non-global and global zones (shaX: 65). Determine if it was set up as a logical virtual interface by checking for a discrepancy between the takeover-ipv4 item displayed by executing hanethvrsc print and the virtual IP address for the Hostname item displayed by executing hanetconfig print.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------------+
 sha0:65    192.168.100.101  -

# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+----------------------------+
 sha1        192.168.10.1     n                     fjgi0
 sha2        192.168.20.1     n                     fjgi1
 sha0        192.168.100.100  c                     sha1,sha2
```

Example

- Example of setup of the NIC switching mode script
  The setup example below is for if setting up the takeover address "192.168.10.10" in the non-global zone "zone1" by using the network interfaces "fjgi0 fjgi1" which the GLS NIC switching mode multiplied in the global zone.

  - /var/tmp/PCL/rmstools/hvzonelip_zone1-net1.conf

```
LIP="192.168.10.10"
IPMASK="24"
IFLIST="fjgi0 fjgi1"
IPTYPE="inet"
ZONENAME="zone1"
```

- Example of setup of the script for if not using GLS

  The setup example below is for if setting up the takeover IP address "fd00:100::2e0:edff:fe0a:f30a" in the non-global zone "zone1" by using the global zone network interface "fjgi0."

  - /var/tmp/PCL/rmstools/hvzonelip_zone1-net1.conf

```
LIP="fd00:100::2e0:edff:fe0a:f30a"
IPMASK="64"
IFLIST="fjgi0"
IPTYPE="inet6"
ZONENAME="zone1"
```

- Example of setup of the fast switching script or GS/SURE linkage script
  The setup example below is for if setting up the takeover logical virtual interface "sha0:65" in the non-global zone "zone1" by using the network interface multiplied by the GLS fast switching mode or GS/SURE linkage mode.

  ```
  LIF="sha0:65"
  ZONENAME="zone1"
  ```

3. The created files for all nodes (global zone) which use Cmdline resources are copied.

   Note

   If a configuration file is not allocated for each node, Cmdline resources may not be started.

4. Register Cmdline resources.
   Select "Path Input" from "Creation Method" for creating the Cmdline to configure the Start script and Stop script. No input is needed for Check script.

   - When using NIC switching mode of GLS, or not using GLS;

     - Start script

       ```
       /var/tmp/PCL/rmstools/hvzonelip.sh start [config file]
       ```

     - Stop script

       ```
       /var/tmp/PCL/rmstools/hvzonelip.sh stop [config file]
       ```

     - Check script

       No input.

   - When using fast switching mode of GLS or using GS/SURE linkage mode;

     - Start script

       ```
       /var/tmp/PCL/rmstools/hvzonelifrsc.sh start [config file]
       ```

     - Stop script

       ```
       /var/tmp/PCL/rmstools/hvzonelifrsc.sh stop [config file]
       ```

     - Check script

       No input.

   The example are shown below.

   - Start script

     ```
     /var/tmp/PCL/rmstools/hvzonelip.sh start /var/tmp/PCL/rmstools/hvzonelip_zone1-net1.conf
     ```

   - Stop script

     ```
     /var/tmp/PCL/rmstools/hvzonelip.sh stop /var/tmp/PCL/rmstools/hvzonelip_zone1-net1.conf
     ```

   - Check script

     No input.

5. Set attributes for scripts. Click the Flag button and set the following values.

Flag	Overview
NULLDETECTOR	Since there is no Check script, change this to "Yes."

Flag	Overview
TIMEOUT	Change the default value from 300 seconds to 60 seconds. If making the TIMEOUT shorter than 60 seconds, set a value upon consideration of the execution time for the Start script and Stop script registered above. For details, refer to "6.7.1.1 Creating Cmdline Resources Creating Cmdline Resources."

Note

- Under the status of the non-global zones having been started up, make the Cmdline resource (shared IP control) into Online status. If the non-global zone is stopped, the shared zone IP zone setup using the ifconfig will fail.
- If one wishes to start up the takeover IP address before the applications operating with the RMS on the non-global zone enter Online status, set up the associating of the Resource to the Cmdline resource (non-global zone control). Set up the Cmdline resource (shared IP control) to the subApplication for the Cmdline resource (non-global zone control). For details, refer to "6.7 Setting Up Cluster Applications."

16.2.5.4 Creating Cluster Applications

In addition to the Gds resource, Gls resource, and Fsystem resource previously registered to the global zone, add the Cmdline resource created with Procedure "5-2. Creating the Cmdline Resource for Non-Global Zone Control" and create the cluster application corresponding to the target non-global zone.

Non-global zone should be stopped to create the cluster application. When the non-global zone is active, take the procedure below to stop the non-global zone that is controlled by the cluster application on all the cluster nodes, and then create the cluster application.

```
# zlogin zone-a shutdown -i0 -g0 -y
```

Check the cluster environment after creating the cluster applications. For details see "6.10 Checking the Cluster Environment."

Note

For the procedure for creating the cluster applications, follow "6.7.2.1 Creating Standby Cluster Applications." However, there is the following difference in procedure.

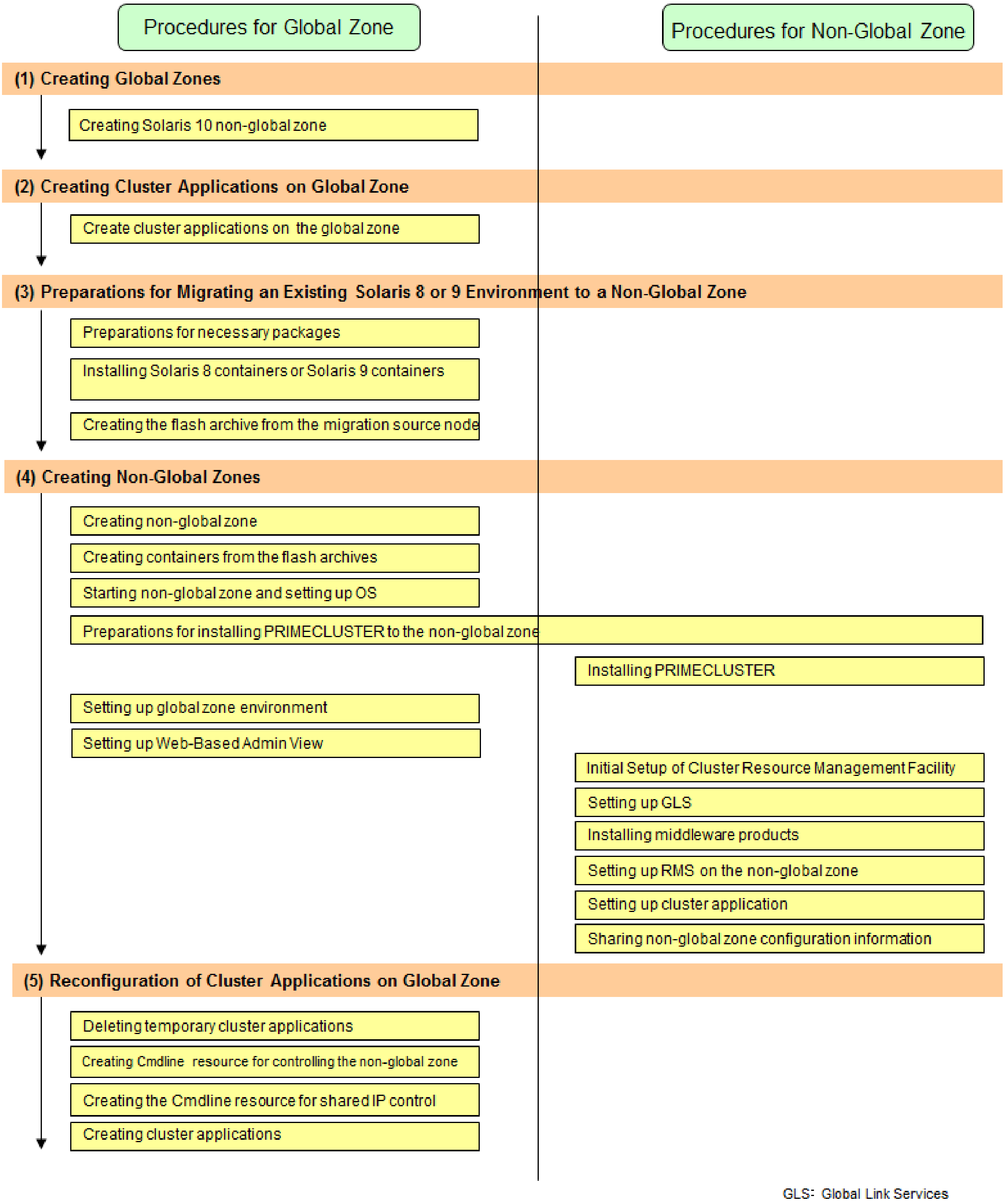
- Cluster application attributes

  If using warm-standby, be sure to always set the "Standby Transitions" to "ClearFaultRequest|StartUp|SwitchRequest."

16.3 Configuration for Using OSLC

This section explains the procedure for building clusters when using an Oracle Solaris Zones environment.

Figure 16.15 Flow of building PRIMECLUSTER



Shown below is an overview of building procedures. Select a configuration stated in "Table 13.7 Operation Mode and Configuration Combinations," and perform a task corresponding to the configuration.

Table 16.18 Building procedure

No	Task	Explanatory Chapter	Task Necessity for Each Configuration (Configuration number of 13.1.3)				
			1	2	4	6	7
(1) Creating Global Zones		16.3.1					
1	Creating Solaris 10 Non-Global Zone	16.3.1.1	B	B	B	B	B
(2) Creating Cluster Applications on Global Zone							
2	Creating Cluster Applications on Global Zone	16.3.2	B	B	B	B	C
Preparations for Migrating an Existing Solaris 8 or 9 Environment to a Non-Global Zone		16.3.3					
3	Preparations for Necessary Packages	16.3.3.1	A	A	A	A	A
4	Installing Solaris 8 Containers or Solaris 9 Containers	16.3.3.2	A	A	A	A	A
5	Creating a Flash Archive from the Migration Source Node	16.3.3.3	A	A	A	A	A
(4) Creating Non-Global Zones		16.3.4					
6	Creating Non-Global Zones	16.3.4.1	A	A	A	A	A
7	Creating Containers from a Flash Archive	16.3.4.2	A	A	A	A	A
8	Non-Global Zone Startup and OS Setup	16.3.4.3	A	A	A	A	A
9	Preparations for Installing PRIMECLUSTER to the Non-Global Zone	16.3.4.4	A	A	A	C	B
10	Installing PRIMECLUSTER to Non-Global Zones	16.3.4.5	A	A	A	C	B
11	Global Zone Environment Setup (After Installation of PRIMECLUSTER to Non-Global Zones)	16.3.4.6	A	A	A	C	B
12	Setup of Web-Based Admin View for Non-Global Zones	16.3.4.7	A	A	A	C	B
13	Initial Setup of Non-Global Zone Cluster Resource Management Facility	16.3.4.8	A	A	A	C	B
14	Setup of GLS in Non-Global Zone	16.3.4.9	B	B	C	C	B
15	Installing Middleware Products to Non-Global Zones	16.3.4.10	B	B	B	C	B
16	RMS Setup for Non-Global Zones	16.3.4.11	C	C	C	C	A
17	Building Non-Global Zone Cluster Applications	16.3.4.12	A	A	A	C	B
18	Sharing Non-Global Zone Configuration Information	16.3.4.13	C	C	A	B	C
(5) Reconfiguration of Cluster Applications on Global Zone		16.3.5					
19	Deleting Temporal Cluster Applications	16.3.5.1	A	A	A	A	C
20	Creating the Cmdline Resource for Non-Global Zone Control	16.3.5.2	A	A	A	A	A
21	Creating the Cmdline Resource for Shared IP Control	16.3.5.3	C	B	C	C	C
22	Creating Cluster Applications	16.3.5.4	A	A	A	A	A

A:Required, B:Perform as Required, C:Unrequired
GLS: Global Link Services

16.3.1 Creating Global Zones

Perform this procedure with all of the nodes which comprise the cluster system.

16.3.1.1 Creating Solaris 10 Non-Global Zones

Install Solaris 10 to the global zone. For details, see "16.2.1 Creating Global Zone."

Be sure to apply the PRIMECLUSTER patch below:

- T012013SP-01 or later

16.3.2 Creating Cluster Applications on the Global Zone

For details on this procedure, see "16.2.2 Creating Cluster Applications on the Global Zone."

16.3.3 Preparations for Migrating an Existing Solaris 8 or 9 Environment to a Non-Global Zone

For migrating an existing Solaris 8 or Solaris 9 environment to the non-global zone, perform the following preparations.

There are the following two types of procedures to create the non-global zone:

- Migrating an existing Solaris 8 environment to Containers
- Migrating an existing Solaris 9 environment to Containers

Before migrating it, perform the following steps. For details, see "System Administration Guide: Oracle Solaris 9 Containers" or "System Administration Guide: Oracle Solaris 8 Containers."

16.3.3.1 Preparations for Necessary Packages

Perform the following in the global zone on each node.

Mount the DVD medium of PRIMECLUSTER and convert the format of packages.

<*DVD_DIR*> is a directory that the DVD is mounted and <*PKG_DIR*> is a given directory to store packages.

```
# pkgtrans <DVD_DIR>/Solaris10/products/CF_NGZ/sparc <PKG_DIR>/FSUNclnet.str FSUNclnet
```

16.3.3.2 Installing Solaris 8 Containers or Solaris 9 Containers

If migrating an existing Solaris 8 or Solaris 9 to Containers, Solaris 8 Containers or Solaris 9 Containers are used. With the following procedure, check that a run time environment of Solaris 8 Containers or Solaris 9 Containers has been installed to the global zone. After that install packages below (download them from Oracle download site).

- For installing Solaris 8 Containers

  Checking a run time environment for Solaris 8 Containers. Execute the following commands to check that the package information is displayed.

  ```
  # pkginfo SUNWs8brandr
  # pkginfo SUNWs8brandu
  ```

  If the package information is not displayed, check that the operating system in the global zone has been installed correctly.

  Install the SUNWs8brandk package for Solaris 8 Containers. <DIR> is used for placing the package.

  ```
  # pkgadd -d <DIR> SUNWs8brandk
  ```

- For installing Solaris 9 Containers

  Checking a run time environment for Solaris 9 Containers. Execute the following commands to check that the package information is displayed.

  ```
  # pkginfo SUNWs9brandr
  # pkginfo SUNWs9brandu
  ```

  If the package information is not displayed, check that the operating system in the global zone has been installed correctly.

Install the SUNWs9brandk package for Solaris 9 Containers. <DIR> is used for placing the package.

```
# pkgadd -d <DIR> SUNWs9brandk
```

16.3.3.3 Creating a Flash Archive from the Migration Source Node

Create a flash archive from the migration source's Solaris 8 environment or Solaris 9 environment. Execute the following commands in the migration source environment to create the flash archive.

```
# flarcreate -S -n zone-a-system /var/tmp/zone-a-system.flar
current filter settings
Creating the archive...
8303358 blocks
Archive creation complete.
```

Transfer the created flash archive (zone-a-system.flar) to the desired folder in the global zone.

Note

When migrating the environment that used the other cluster software in the migration source to the new PRIMECLUSTER environment, follow the specification of the cluster software and delete it from the environment in the migration source if necessary. Then, create a Flash Archive.

16.3.4 Creating Non-Global Zones

This section describes procedures for building non-global zones. Implement the procedures for only the operational system if one is using a configuration which shares non-global zone images between cluster nodes. If using a configuration which does not share non-global zone images between cluster nodes, implement the procedures in all nodes. To create non-global zones, the following conditions must be satisfied.

- Must be Whole Root Zone (a zone not sharing system files with the global zone)
- The non-global zone autoboot (the setting to automatically start up the non-global zone during startup of the global zone; the default setting is false) must be set to false.
- If using a configuration where non-global zone images are not shared between cluster nodes, the zone names (the names of zones as defined by zonecfg and zoneadm -z) must match for all nodes.

If not otherwise specified in the subsequent parts of this section, implement the procedures for only the operational system if one is using a configuration which shares non-global zone images between cluster nodes. If using a configuration which does not share non-global zone images between cluster nodes, implement the procedures in all nodes.

16.3.4.1 Creating Non-Global Zones

For each creation of a non-global zone, create a resource pool beforehand. Create it on all nodes. For details on the procedure, see "System Administration Guide: Oracle Solaris 9 Containers" or "System Administration Guide: Oracle Solaris 8 Containers."

If building a cluster with a Solaris Containers environment, make the number of CPU cores to be allocated to the global zone two or more.

Using the zonecfg command, create the non-global zones. Create them by referring to the following example.

```
# zonecfg -z zone-a  *1
```

*1: "zone-a" is the zone name (it is the same below).

```
zone-a: No such zone configured
Use 'create' to begin configuring a new zone.
zonecfg:zone-a> create -t SUNWsolaris9  (For Solaris 9 Container)
zonecfg:zone-a> create -t SUNWsolaris8  (For Solaris 8 Container)
zonecfg:zone-a> set zonepath=/zone-a-system  *2
zonecfg:zone-a> set autoboot=true (for warm-standby)
zonecfg:zone-a> set autoboot=false (for cold-standby)
```

*2: For /zone-a-system, specify the directory to which the zone-a images are allocated.

If sharing images, specify the mountpoint registered as an Fsystem resource. If not sharing images, specify the file system on the local system.

```
zonecfg:zone-a> set limitpriv="default,proc_priocntl"
zonecfg:zone-a> add fs
zonecfg:zone-a:fs> set dir=/oracle-data
zonecfg:zone-a:fs> set special=/zone-a-oracle  *3
zonecfg:zone-a:fs> set type=lofs
zonecfg:zone-a:fs> end
```

*3: For /zone-a-oracle, specify the directory of the Fsystem resource corresponding to the volume for zone-a Oracle data allocation.

```
[If making the non-global zone's network mode a shared IP zone configuration]
zonecfg:zone-a> add net   (If making it a shared IP zone configuration)
zonecfg:zone-a:net> set physical=e1000g0  *4
zonecfg:zone-a:net> set address=10.20.30.40/24
zonecfg:zone-a:net> end
```

*4: If specifying a network interface multiplexed with GLS, specify the Primary interface for the corresponding Gls resource.

```
[If making the non-global zone's network mode an exclusive IP zone configuration]
zonecfg:zone-a> set ip-type=exclusive
zonecfg:zone-a> add net
zonecfg:zone-a:net> set physical=e1000g0  *5
zonecfg:zone-a:net> end
```

*5: Specify the physical interface exclusive to Containers. Perform the IP address setup and physical interface multiplexing from within Containers. If the physical interface is activated, perform deactivation.

```
zonecfg:zone-a> add net
zonecfg:zone-a:net> set physical=e1000g1  *6
zonecfg:zone-a:net> end
```

*6: If the physical interface is multiplexed within Containers, it is necessary to specify two or more physical interfaces.

```
zonecfg:zone-a> verify
zonecfg:zone-a> commit
zonecfg:zone-a> exit
```

Note

If using a shared IP zone configuration

For the IP address set up to the non-global zone, set up the IP address which is not being used with GLS. Perform the setup of the default gateway to the zone on the global zone. If one has set up the default gateway using the zonecfg command, the paths will be disabled when performing NIC switching with GLS.

16.3.4.2 Creating Containers from a Flash Archive

Create Solaris 8 Containers or Solaris 9 Containers from the flash archive created in the migration source, Solaris 8 environment or Solaris 9 environment.

Note

When expanding a flash archive to the GDS volume on the shared disk, activate the GDS volume and then mount the file system on it.

Example: When expanding a flash archive to the file system (ufs) on GDS volume

```
# sdxvolume -N -c class0001 -v volume0001
# mount -F ufs /dev/sfdsk/class0001/dsk/volume0001 /zone-a-system
# zoneadm -z zone-a install -u -a /var/tmp/zone-a-system.flar
     Log File: /var/tmp/ zone-a-system.install.1987.log
       Source: /Etude/zone-a-system.flar
   Installing: This may take several minutes...
```

For details, see "System Administration Guide: Oracle Solaris 9 Containers" or "System Administration Guide: Oracle Solaris 8 Containers."

If PRIMECLUSTER has been installed to the migration source environment, uninstall the PRIMECLUSTER in the migration source with the following procedure:

1. Startup the non-global zone under single-user mode.

```
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)
# zoneadm -z zone-a boot -s
```

2. Mount the medium of PRIMECLUSTER with the same version as the migration source in the global zone. The mountpoint needs to be the directory which can be referred from the non-global zone.

3. Log in to the non-global zone.

```
# zlogin -C zone-a
```

4. Prepare for PRIMECLUSTER deletion in the non-global zone. When a flash archive is created from a server where GDS has not been installed, perform the procedure after Step 8.

5. Check if files other than the class.db exist under the /etc/opt/FJSVsdx/sysdb.d. Moreover, check that the inside of the class.db is empty.

```
# cd /etc/opt/FJSVsdx/sysdb.d
# ls
class.db
```

   * If files other than the class.db exist, delete them with the rm command.
   * If the information exists in the class.db file, delete the line.
   (The line which starts with # is a comment line.)

6. Check if files other than the _adm, _diag exist under the /dev/sfdsk.

```
# cd /dev/sfdsk
# ls
_adm _diag ...
```

   * If files other than the _adm, _diag exist, delete them with the rm-rf command.

7. Remove the GDS Basic Software package (FJSVsdx).

```
# pkgrm FJSVsdx
```

8. Back up the preremove and postremove files of SMAWcf package.

```
# cd /var/sadm/pkg/SMAWcf/install
# cp preremove /var/tmp/preremove.orig
# cp postremove /var/tmp/postremove.orig
```

9. Update the preremove and postremove files of SMAWcf package with the following procedure.

```
# cat /var/tmp/preremove.orig | sed -e 's,$CFCONFIG -u,echo " ",' | \
sed -e 's,/opt/SMAW/bin/cipconfig -u,echo " ",' \
> preremove
```

```
# cat /var/tmp/postremove.orig | sed -e 's,^module_id.*,module_id="",' | \
sed -e 's,out=`rem_drv.*,out=" "; echo " ",'\
> postremove
```

10. Uninstall PRIMECLUSTER from the non-global zone.

    Follow the "PRIMECLUSTER Installation Guide" which is the same version as the migration source to uninstall the PRIMECLUSTER. If the uninstallation procedure varies depending on the server models, follow the procedure for the server you were using before the migration for uninstallation.

    For 4.2A00 or earlier, execute the uninstallation script in the non-global zone. Use the uninstallation script of the medium of PRIMECLUSTER mounted beforehand.

    If PRIMECLUSTER Wizard for Oracle and PRIMECLUSTER Wizard for NAS have been installed in the migration source environment, uninstall them from the non-global zone before uninstalling PRIMECLUSTER. For the procedure for uninstalling each Wizard product, see the installation guide of the same version as each Wizard product in the migration source.

11. Delete the backups of the preremove and postremove files created in Step 8.

```
# rm /var/tmp/preremove.orig /var/tmp/postremove.orig
```

16.3.4.3 Non-Global Zone Startup and OS Setup

Using the zoneadm boot command, start up the zone for which installation was performed. After that, acquire the console and perform Solaris OS setup (setup of node names, time zones, and so on).

```
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)
# zoneadm -z zone-a boot
# zlogin -C zone-a
```

From now on, set up the Solaris OS according to the instructions on the screen.

For details, see the manual for the zoneadm and zlogin commands, and also "System Administration Guide: Oracle Solaris 9 Containers" or "System Administration Guide: Oracle Solaris 8 Containers."

If monitoring applications in the non-global zone, proceed to Procedure "16.3.4.4 Preparations for Installing PRIMECLUSTER to the Non-Global Zone."

If not monitoring applications in the non-global zone, proceed to Procedure "16.3.4.13 Sharing Non-Global Zone Configuration Information."

16.3.4.4 Preparations for Installing PRIMECLUSTER to the Non-Global Zone

Before installation, preparation work is required in the non-global zone created above and in the global zone. This procedure varies depending on which environment you use: Solaris 8 Containers environment or Solaris 9 Containers environment.

1. Preparations on the Non-Global Zone (for Solaris 8 Containers)

   Mount the DVD or CD medium of PRIMECLUSTER set in the global zone to the file system under a non-global zone.

   Note

   If the non-global zone is in the Solaris 8 Containers environment and that PRIMECLUSTER was used in the migration source environment, the installation CD medium used in the Containers environment in the migration source is required.

   Log in to the non-global zone.

```
# zlogin zone-a
```

   Mount the DVD medium of this software or the CD1 medium of PRIMECLUSTER in case that PRIMECLUSTER was used in the migration source on the non-global zone, or copy the contents of the medium, so that the contents can be accessed from the non-global zone. In the following example, the directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris8/CD1.)

For Solaris 8 Containers environments, install the SMAWskel package with the following procedure.

If you are asked to input some value while executing the pkgadd(1M) command, input the default value. If there is no default value, select "y."

```
# cd <CDROM_DIR>/Solaris9/products/CF/sparc
# pkgadd -d . SMAWskel
```

2. Preparations on the Non-Global Zone (for Solaris 9 environment)

Mount the DVD or CD medium of PRIMECLUSTER set in the global zone to the file system under a non-global zone.

Note

If the non-global zone is in the Solaris 9 Containers environment and that PRIMECLUSTER was used in the migration source environment, the installation CD medium used in the Containers environment in the migration source is required.

Log in to the non-global zone.

```
# zlogin zone-a
```

Mount the DVD medium of this software or the CD1 medium of PRIMECLUSTER in case that PRIMECLUSTER was used in the migration source on the non-global zone, or copy the contents of the medium, so that the contents can be accessed from the non-global zone. In the following example, the directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris9/CD1.)

For Solaris 9 Containers environments, install the SMAWskel package with the following procedure.

If you are asked to input some value while executing the pkgadd(1M) command, input the default value. If there is no default value, select "y."

```
# cd <CDROM_DIR>/Solaris9/products/CF/sparc
# pkgadd -d . SMAWskel
```

3. Preparations in the Global Zone (for both Solaris 8 and 9 environments)

Perform this task in the global zone.

Execute the cfzinstall script. If an error message is output and it requires you to restart the non-global zone, select "yes" and restart the non-global zone. Then, proceed to the next step.

```
# /opt/SMAW/SMAWcf/bin/cfzinstall zone-a
CF configuration set to: zone-a ZONE-A
/usr/sbin/zlogin zone-a /etc/init.d/cf start
```

Execute the hvzonesetup script.

```
# hvzonesetup -z zone-a -c
NOTICE: User can now install PCL in zone zone-a.
        After installing PCL run "hvzonesetup -z zone-a -u".
```

Copy the FSUNclnet.str package prepared in "16.3.3.1 Preparations for Necessary Packages" to any one of the directories in the non-global zone.

Information

The following process is performed by the cfzinstall script:

- Installation and setup of CF for the non-global zone

The following processes are performed by the hvzonesetup script:

- Changing the zonename command temporarily (This command is restored by hvzonesetup -u to be executed after installing PRIMECLUSTER)
- Configuring the RMS environment variables for the non-global zone (hvenv.local file setting)
- Adding PATH settings to /etc/profile

Note

When migrating the environment that used the other cluster software in the migration source to the new PRIMECLUSTER environment, and also when the cluster software used in the migration source was not deleted before creating a Flash Archive, follow the specification of the cluster software used in the migration source and delete it from the Containers environment.

16.3.4.5 Installation of PRIMECLUSTER to the Non-Global Zone

This procedure varies depending on which environment you use: Solaris 8 Containers environment or Solaris 9 Containers environment. For Solaris 8 Containers environment, refer to Step 1 and for Solaris 9 Containers environment, refer to Step 2.

1. Installation Procedure (for Solaris 8 Containers environment)

   Perform this task in the non-global zone. If you are asked to input some value while executing the pkgadd(1M) and pkgrm(1M) commands, input the default value. If there is no default value, select "y."

   Check that FJSVsnap, FJSVwvcnf, and FJSVwvbs have been installed in the non-global zone.

```
# pkginfo -l FJSVsnap.\*
# pkginfo -l FJSVwvcnf.\*
# pkginfo -l FJSVwvbs.\*
```

   If the version of FJSVsnap is 2.3.1 or earlier, remove its package from the system.

```
# pkgrm FJSVsnap
```

   If the version of FJSVwvcnf is other than 4.11, or the version of FJSVwvbs is other than 4.1.2, remove both packages from the system.

```
# pkgrm FJSVwvcnf
# pkgrm FJSVwvbs
```

   If old version packages were removed in the task above, or packages have not been installed, execute the pkgadd(1M) command to install the target packages. The directory in the non-global zone of PRIMECLUSTER DVD or CD1 mounted or copied in the previous section is <CDROM_DIR>.

```
# pkgadd -d <CDROM_DIR>/Tool/Pkgs FJSVsnap
# pkgadd -d <CDROM_DIR>/Solaris8/products/Base_GUIs/sparc FJSVwvcnf
# pkgadd -d <CDROM_DIR>/Solaris8/products/Base_GUIs/sparc FJSVwvbs
```

   Install the FSUNclnet component. Below is an example when placing FSUNclnet.str to /var/tmp/package in the non-global zone.

```
# cd /var/tmp/package
# pkgadd -d FSUNclnet.str
```

   Install PRIMECLUSTER from PRIMECLUSTER DVD or CD1 mounted or copied in the previous section. The directory in the non-global zone is <CDROM_DIR>.

```
# cd <CDROM_DIR>/Solaris8/products/Base_GUIs/sparc
# pkgadd -d . SMAWcj2re
# pkgadd -d . FJSVwvmpc
# pkgadd -d . SMAWrcadm
# pkgadd -d . SMAWrcaja
# cd <CDROM_DIR>/Solaris8/products/CF/sparc/
# pkgadd -d . FJSVclapi
# pkgadd -d . FJSVcldbm
```

```
# cd <CDROM_DIR>/Tool
# ./cluster_install -p RMS
# ./cluster_install -p WT
```

If the physical interface is multiplexed by GLS in the non-global zone with an exclusive IP zone configuration, install PRIMECLUSTER GLS from PRIMECLUSTER DVD or CD2. The directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris8/CD2.)

```
[Only when using GLS to the non-global zone with an exclusive IP zone configuration]
# mv /usr/sbin/add_drv /usr/sbin/add_drv.save
# touch /usr/sbin/add_drv
# chmod 555 /usr/sbin/add_drv
# pkgadd -d <CDROM_DIR>/Solaris8/products/GLS/sparc FJSVhanet
# mv /usr/sbin/add_drv.save /usr/sbin/add_drv
```

Mount or copy the DVD or SUPPLEMENT CD medium of PRIMECLUSTER on the non-global zone to enable the CD1 medium to access from the non-global zone. In the following example, the directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris8/CD3.)

```
# cd <CDROM_DIR>/Tool
# ./cluster_install -p GUIs
# cd <CDROM_DIR>/Solaris8/products/CF_Add_On/sparc
# pkgadd -d . FJSVclapm
# cd <CDROM_DIR>/Solaris8/products/RMS_Add_On/sparc
# pkgadd -d . FJSVclrms
# pkgadd -d . FJSVclrwz
# pkgadd -d . FJSVclsfw
```

Note

The following messages may be output during installation, but it does not affect the system's behavior.

- Warning: The package <SMAWccbr> has not been installed.
- prtconf: devinfo facility not available

After installing the packages, apply the latest PRIMECLUSTER patches. Apply all the required patches for the non-global zone that are stored in the DVD.

```
<DVD mount point>/Solaris8/patch
```

Below is an example when applying the PRIMECLUSTER patch 901172-32 to the non-global zone.

```
# cd <DVD mount point>/Solaris8/patch
# patchadd 901172-32
```

If you installed PRIMECLUSTER using the medium of PRIMECLUSTER which was used in the migration source, the following PRIMECLUSTER patches need to be applied to the non-global zone.

Solaris 8 Containers PRIMECLUSTER 4.1A30	Solaris 8 Containers PRIMECLUSTER 4.1A40
901167-07 or later	901167-07 or later
901172-34 or later	901172-34 or later
901173-24 or later	901173-24 or later
913855-05 or later	913855-05 or later
914111-03 or later	914111-03 or later
914112-10 or later	914112-10 or later

Solaris 8 Containers PRIMECLUSTER 4.1A30	Solaris 8 Containers PRIMECLUSTER 4.1A40
914120-01 or later	914346-01 or later
914346-01 or later	914351-02 or later
914351-01	914530-01
914530-01	915102-01 or later
915102-01 or later	

Below is an example of applying 901172-32 when placing /var/tmp/patch in the non-global zone.

```
# cd /var/tmp/patch
# zcat 901172-32.tar.Z | tar xf -
# patchadd 901172-32
```

Edit /etc/inet/hosts and add the entry of "host name + RMS in the non-global zone" to any one of IP addresses assigned to the non-global zone.

```
# vi /etc/inet/hosts
ipadress      zonename  zonenameRMS
                        ^^^^^^^^^^^addition
```

ipadress: a given IP address assigned to the non-global zone

zonename: host name in the non-global zone

Set environment variables of java_home for Web-Based Admin View.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam java_home \
 /opt/SMAW/SMAWcj2re/jre
local:java_home         /opt/SMAW/SMAWcj2re/jre
```

2. Installation procedure (for Solaris 9 Containers environment)

   Perform this task in the non-global zone If you are asked to input some value while executing pkgadd(1M) and pkgrm(1M) commands, input the default value. If there is no default value, select "y."

   Check that FJSVsnap, FJSVwvcnf, and FJSVwvbs have been installed in the non-global zone.

```
# pkginfo -l FJSVsnap.\*
# pkginfo -l FJSVwvcnf.\*
# pkginfo -l FJSVwvbs.\*
```

   If the version of FJSVsnap is 2.3.1 or earlier, remove its package from the system.

```
# pkgrm FJSVsnap
```

   If the version of FJSVwvcnf is other than 4.11, or the version of FJSVwvbs is other than 4.1.2, remove both packages from the system.

```
# pkgrm FJSVwvcnf
# pkgrm FJSVwvbs
```

   If old version packages were removed in the task above, or packages have not been installed, execute the pkgadd(1M) command to install the target packages. The directory in the non-global zone of PRIMECLUSTER DVD or CD1 mounted or copied in the previous section is <CDROM_DIR>.

```
# pkgadd -d <CDROM_DIR>/Tool/Pkgs FJSVsnap
# pkgadd -d <CDROM_DIR>/Solaris9/products/Base_GUIs/sparc FJSVwvcnf
# pkgadd -d <CDROM_DIR>/Solaris9/products/Base_GUIs/sparc FJSVwvbs
```

   Install the FSUNclnet component. Below is an example when placing FSUNclnet.str to /var/tmp/package in the non-global zone.

```
# cd /var/tmp/package
# pkgadd -d FSUNclnet.str
```

Install PRIMECLUSTER from PRIMECLUSTER DVD or CD1 mounted or copied in the previous section. The directory in the non-global zone is <CDROM_DIR>.

```
# cd <CDROM_DIR>/Solaris9/products/Base_GUIs/sparc
# pkgadd -d . SMAWcj2re
# pkgadd -d . FJSVwvmpc
# pkgadd -d . SMAWrcadm
# pkgadd -d . SMAWrcaja
# cd <CDROM_DIR>/Solaris9/products/CF/sparc/
# pkgadd -d . FJSVclapi
# pkgadd -d . FJSVcldbm
# cd <CDROM_DIR>/Tool
# ./cluster_install -p RMS
# ./cluster_install -p WT
```

If the physical interface is multiplexed by GLS in the non-global zone with an exclusive IP zone configuration, install PRIMECLUSTER GLS from PRIMECLUSTER DVD or CD2. The directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris9/CD2.)

```
[Only when using GLS to the non-global zone with an exclusive IP zone configuration]
# mv /usr/sbin/add_drv /usr/sbin/add_drv.save
# touch /usr/sbin/add_drv
# chmod 555 /usr/sbin/add_drv
# pkgadd -d <CDROM_DIR>/Solaris9/products/GLS/sparc FJSVhanet
# mv /usr/sbin/add_drv.save /usr/sbin/add_drv
```

Mount or copy the DVD or SUPPLEMENT CD medium of PRIMECLUSTER on the non-global zone to enable the CD1 medium to access from the non-global zone. In the following example, the directory (CD mount point) in the non-global zone is <CDROM_DIR>. (When using the DVD medium of this software, <CDROM_DIR> is <DVD mount point>/Solaris9/CD3.)

```
# cd <CDROM_DIR>/Tool
# ./cluster_install -p GUIs
# cd <CDROM_DIR>/Solaris9/products/CF_Add_On/sparc
# pkgadd -d . FJSVclapm
# cd <CDROM_DIR>/Solaris9/products/RMS_Add_On/sparc
# pkgadd -d . FJSVclrms
# pkgadd -d . FJSVclrwz
# pkgadd -d . FJSVclsfw
```

Note

The following messages may be output during installation, but it does not affect the system's behavior.

- Warning: The package <SMAWccbr> has not been installed.
- prtconf: devinfo facility not available

After installing the packages, apply the latest PRIMECLUSTER patches. Apply all the required patches for the non-global zone that are stored in the DVD.

```
<DVD mount point>/Solaris9/patch
```

Below is an example when applying the PRIMECLUSTER patch 901196-24 to the non-global zone.

```
# cd <DVD mount point>/Solaris9/patch
# patchadd 901196-24
```

If you installed PRIMECLUSTER using the medium of PRIMECLUSTER which was used in the migration source, the following PRIMECLUSTER patches need to be applied to the non-global zone.

Solaris 9 Containers PRIMECLUSTER 4.1A30	Solaris 9 Containers PRIMECLUSTER 4.1A40	Solaris 9 Containers PRIMECLUSTER 4.2
901167-07 or later	901167-07 or later	901196-35 or later
901172-34 or later	901172-34 or later	901215-04 or later
901173-24 or later	901173-24 or later	901217-29 or later
913855-05 or later	913855-05 or later	901254-02 or later
914111-03 or later	914111-03 or later	913855-05 or later
914112-10 or later	914112-10 or later	914111-03 or later
914120-01 or later	914346-01 or later	914112-10 or later
914346-01 or later	914346-01 or later	914346-01 or later
914351-01	914351-02 or later	914351-02 or later
914530-01	914530-01	914530-02 or later
915102-01 or later	915102-01 or later	915102-01 or later

Below is an example of applying 901196-24 (Solaris 9) when placing /var/tmp/patch in the non-global zone.

```
# cd /var/tmp/patch
# zcat 901196-24.tar.Z | tar xf -
# patchadd 901196-24
```

Edit /etc/inet/hosts and add the entry of "host name + RMS in the non-global zone" to any one of IP addresses assigned to the non-global zone.

```
# vi /etc/inet/hosts
ipadress      zonename   zonenameRMS
                         ^^^^^^^^^^^addition
```

ipadress: a given IP address assigned to the non-global zone

zonename: host name in the non-global zone

Set environment variables of java_home for Web-Based Admin View.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam java_home \
 /opt/SMAW/SMAWcj2re/jre
local:java_home         /opt/SMAW/SMAWcj2re/jre
```

16.3.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone)

1. Executing the script to configure the non-global zone

   Execute the hvzonesetup script by specifying the -u option in the global zone.

```
# hvzonesetup -z zone-a -u
```

Information

The following process is performed by the hvzonesetup script.

- Restore the changes of the zonename command implemented by hvzonesetup -z zone-a -c.

2. Changing kernel parameters

   This task is unnecessary if one is not installing PRIMECLUSTER to the non-global zone.

   Add the number of non-global zones to be created and make this the value for the kernel parameter detailed in "Table 16.19 Kernel Parameters Requiring Revision" below to the /etc/system for all of the global zones which comprise the cluster system. Then restart the global zone.

```
# shutdown -y -g0 -i6
```

   Table 16.19 Kernel Parameters Requiring Revision

Kernel Parameters	Attribute	Value	Remarks
shmsys:shminfo_shmmni	Add	30	Value required for resource database per Containers
semsys:seminfo_semmni	Add	20	Value required for resource database per Containers

   Note

   - Do not delete the following definitions written in the **non-global zones'** /etc/system.

     set semsys:seminfo_semmni=30

     set shmsys:shminfo_shmmni=130

     set in_sync=1

   - It will not cause any problems if the following definitions written in the non-global zones' /etc/system are deleted.

     msgsys:msginfo_msgmnb

     msgsys:msginfo_msgmni

     msgsys:msginfo_msgtql

   Check the value of shmsys:shminfo_shmseg set in /etc/system in the non-global zone.

   - If the value is less than 30, or no value has been set

     Set the **non-global zones'** /etc/system to the following.

     set shmsys:shminfo_shmseg = 30

   - If the value is 30 or larger

     No action is required.

   After checking it, add the value of the kernel parameter detailed in table 3.1 to the **non-global zones'** /etc/system. Then, restart the non-global zone.

```
# zlogin zone-a shutdown -y -g0 -i6
```

3. Registering the GDS shared class volume

   This procedure is necessary if attempting to access the GDS shared class volume from the non-global zone.

   Copy the special file of the GDS shared class volume to under /<zonepath>/dev. Execute the following commands from the global zone.
   (If the zonepath is /zone-a-system, the class name is class0001, and the volume name is volume0001)

```
# cd /dev
# tar cvf /var/tmp/dsk.tar sfdsk/class0001/dsk/volume0001
# tar cvf /var/tmp/rdsk.tar sfdsk/class0001/rdsk/volume0001
# cd /zone-a-system/dev
# tar xvf /var/tmp/dsk.tar
# tar xvf /var/tmp/rdsk.tar
```

Note

If the non-global zone is detached and then attached manually, the GDS volume special file copied in the procedure above may be deleted depending on the OS specification. In this case, re-perform this procedure after attaching the non-global zone.

When starting or stopping the non-global zone by starting or stopping the cluster application of PRIMECLUSTER, the GDS special file is not be deleted.

Create the file system to the volume in the non-global zone. Execute the following command from the non-global zone.
(If the class name is class0001, the volume name is volume0001, and the file system is UFS)

```
# newfs /dev/sfdsk/class0001/rdsk/volume0001
```

Note

Perform the creation of the above file system only from the one node first used.

16.3.4.7 Setup of Web-Based Admin View for the Non-Global Zone

Perform this task in the non-global zone.

See "4.2.3 Initial Setup of Web-Based Admin View," and perform the setup and startup for Web-Based Admin View. When doing so, specify the same non-global zone IP addresses as those for both the primary management server and the secondary management server specified with "4.2.3.1 Initial setup of the operation management server." In addition, refer to "7.1 Network address" in "PRIMECLUSTER Web-Based Admin View Operation Guide" to set the same IP addresses to mip and httpip.

(Example: If the non-global zone IP address is 10.20.30.40)

```
# /etc/init.d/fjsvwvcnf stop
# /etc/init.d/fjsvwvbs stop
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam mip 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam httpip 10.20.30.40
# /etc/opt/FJSVwvbs/etc/bin/wvCntl start
# /etc/init.d/fjsvwvcnf start
```

After setup, use the procedure "4.3 Starting the Web-Based Admin View Screen" to confirm that one is able to start up the GUI screen.

16.3.4.8 Initial Setup of the Non-Global Zone Cluster Resource Management Facility

Perform this task in the non-global zone.

After starting up the Web-Based Admin View screen, see "5.1.3 Initial Setup of the Cluster Resource Management Facility" and "5.1.3.1 Initial Configuration Setup" and perform the initial configuration setup for the cluster resource management facility.

It is not necessary to perform CF and CIP setups, shutdown facility setup, or automatic configuration for the non-global zone.

Note

If performing initial configuration setup for the cluster resource management facility, the message below will be output onto the non global zone console, but this will not be a problem for its operation.

/dev/rdsk/*: No such file or directory

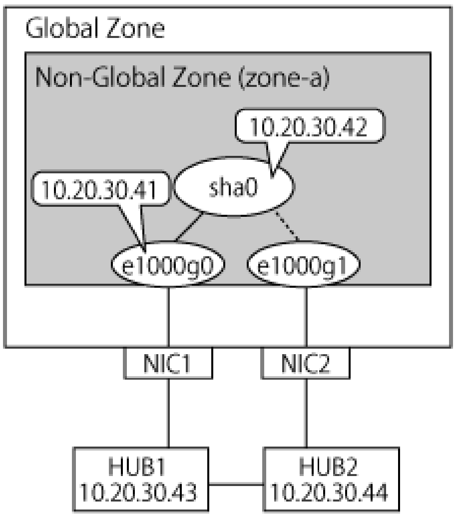
In addition, if initial configuration setup failed, it is possible that the non-global zone kernel parameters were insufficient. See the "A.5 Kernel Parameter Worksheet" and correct the kernel parameter value. After restarting the non-global zone, perform resource database initialization using the clinitreset (1M) command and re-perform the initial configuration setup.

16.3.4.9 Setup of GLS in a Non-Global Zone

This procedure is necessary only if one is using the NIC switching mode with an exclusive IP zone configuration. If setting up GLS on a non global zone, refer to the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function" and perform the setup for multiplexing the physical interface.

Perform this task in all of the non-global zones which are to build the cluster system.

Figure 16.16 Example of an Environment Setup for if Configuring Between the Non-Global Zones with a Warm-standby Configuration



1. System settings

   1-1) Define the IP address to be used and the host name to the /etc/inet/hosts file.

```
10.20.30.42 zone-a0     # zone-a virtual IP(takeover IP)
10.20.30.41 zone-a01    # zone-a physical IP
10.20.30.43 swhub1      # primary monitoring destination HUB IP
10.20.30.44 swhub2      # secondary monitoring destination HUB IP
```

   Note

   Set up the zone-a physical IP address such that it does not overlap with other non-global zone physical IP addresses.

   1-2) Write the host name defined above to the /etc/hostname.e1000g0 file.

   Contents of /etc/hostname.e1000g0

```
zone-a01
```

   1-3) Define the subnet mask to the /etc/inet/netmasks file.

```
10.20.30.0 255.255.255.0
```

2. Rebooting

   Execute the following command and reboot the non-global zone. Perform this command from the global zone. After reboot, execute the ifconfig command to confirm that the e1000g0 is activated.

```
# /usr/sbin/zlogin zone-a shutdown -y -g0 -i6
```

3. Creating the virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha0 -m d -i 10.20.30.42 -e 10.20.30.41 -t
e1000g0,e1000g1
```

Note

Always be sure that the physical IP address defined to the option "-e" matches with the physical IP address set up to the /etc/hostname.e1000g0.

4. Setup of standby patrol function

```
# /opt/FJSVhanet/usr/sbin/hanetconfig create -n sha1 -m p -a 00:00:00:00:00:00 -t sha0
```

Note

- When using the standby patrol function in the non-global zone, you need to apply the PRIMECLUSTER patches (Solaris 8 and Solaris 9: 914111-03 or later) to the function.
- The -a option cannot be omitted in an OSLC environment.
  Specify 00:00:00:00:00:00 to the -a option value and execute the command.

5. Setup of HUB monitoring function

```
# /opt/FJSVhanet/usr/sbin/hanetpoll create -n sha0 -p 10.20.30.43,10.20.30.44 -b off
```

6. Creating the takeover virtual interface

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

Note

This setting is not necessary for single-node cluster operations.

7. Starting HUB monitoring

```
# /opt/FJSVhanet/usr/sbin/hanetpoll on
```

16.3.4.10 Installing Middleware Products to Non-Global Zones

For the installation procedure and points of caution for Wizard for Oracle, see "PRIMECLUSTER Wizard for Oracle Configuration and Administration Guide."

For the installation procedure and points of caution for Wizard for NAS, see "PRIMECLUSTER Wizard for NAS Configuration and Administration Guide."

16.3.4.11 Setup of Non-Global Zone RMS (for single-node cluster operations)

Edit the following lines of "/opt/SMAW/SMAWRrms/bin/hvenv.local" file to enable the automatic startup of RMS in the non-global zone.

<Before changes>

```
export HV_RCSTART=0
```

<After changes>

```
export HV_RCSTART=1
```

16.3.4.12 Setup of Non-Global Zone Cluster Applications

This section explains the procedure for creating cluster applications on the non-global zone.

Perform the following procedure taking into account the cluster resources that are to be set up.

No.	Task Overview	Procedure	Configuration				
			1	2	4	6	7
1	Setup of the Cmdline resource	1	A	A	A	B	A
2	Setup of the Oracle resource	2	A	A	A	B	A
3	Setup of the Netapp resource	3	A	A	A	B	A
4	Setup of the state transition procedure resources	4	A	A	A	B	A
5	Setup of the Gls resource	5	A	B	A	B	B
6	Setup of the Fsystem resource	6	A	A	B	B	A
7	Creation of the cluster applications	7	A	A	A	B	A

A: Perform as required, B: Unrequired

1. Setup of the Cmdline resource

   For the method for setting up the Cmdline resource, refer to "6.7.1.1 Creating Cmdline Resources."

2. Setup of the Oracle resource

   Refer to the "PRIMECLUSTER Wizard for Oracle Configuration and Administration Guide" and perform the setup of the cluster resource.

3. Setup of the NetApp resource

   Refer to the " PRIMECLUSTER Wizard for NAS Configuration and Administration Guide" and perform the setup of the cluster resource.

4. Setup of the state transition procedure resources

   Refer to the middleware manual and set up the state transition procedure resources. As for the availability of middleware products and PRIMECLUSTER in combination in a non-global zone, contact field engineers.

   Refer to "6.7.1.6 Creating Procedure Resources" and perform the setup of the procedure resources.

5. Setup of the Gls resource

   This procedure is necessary only if one is using an exclusive IP zone configuration.

   Refer to "6.7.1.4 Creating Gls Resources" and perform the setup of the Gls resource.

6. Setup of the Fsystem resource

   This procedure is necessary if using a switching file system with a non-global zone.

   Refer to "6.7.1.2 Creating Fsystem Resources" and perform the setup of the Fsystem resource.

   Note that you cannot set ZFS to Fsystem resources in non-global zones. Instead of that, only the UFS file system can be set.

7. Creation of the cluster applications

   Create the cluster applications on non-global zones.

   For the method for creating the cluster applications, follow "6.7.2.1 Creating Standby Cluster Applications." However, there are the following differences in procedure:

   - Cluster application attributes

     - Set No to AutoStartUp, AutoSwitchOver, and HaltFlag. However, when the global zone is operated on a single-node cluster, set Yes to AutoStartUp.

     - If on a warm-standby configuration and wishing to put the cluster applications on the standby system's non-global zone into Standby mode, set the ClearFaultRequest to StandbyTransitions. For all other circumstances, set No to it.

     - Set NONE to Shutdown Priority.

     - Set 0 to Online Priority.

16.3.4.13 Sharing Non-Global Zone Configuration Information

If using cold-standby, stop the non-global zone in the operational system nodes.

```
# zlogin zone-a shutdown -i0 -g0 -y
```

If sharing non-global zone images in cold-standby operation, make it so that one is able to use the information for the non-global zone created thus far from the standby system's nodes as well.

Export the non-global zone configuration information. Execute the following command on the node that the non-global zone was set in the procedure up to this step.

```
# zonecfg -z zone-a export -f /var/tmp/zone-a.exp
```

Copy the output file (/var/tmp/zone-a.exp in the example above) to the standby system nodes.

Import the non-global zone in all other nodes.

```
# zonecfg -z zone-a -f /var/tmp/zone-a.exp
```

Note

When performing import, since it is not necessary to access the non-global zone's file system, do not perform an operation with all other nodes making the cluster application Online. In addition, do not perform an operation which attaches or starts up the non-global zone.

16.3.5 Reconfiguration of Cluster Applications on Global Zone

Build cluster applications corresponding to each non-global zone to the global zone. Perform this task for each non-global zone.

16.3.5.1 Deleting Temporal Cluster Applications

If you created temporary cluster applications in "16.3.2 Creating Cluster Applications on the Global Zone," stop RMS, and then delete them. Do not perform deletion of the resources when deleting the cluster applications.

16.3.5.2 Creating the Cmdline Resource for Non-Global Zone Control

Create the Cmdline resource for controlling non-global zone and also cluster applications configured to the non-global zone from the global zone.

- If performing application monitoring within the non-global zone (except single node cluster operations)

  Create the Cmdline resource.

  Select "Path Input" from "Creation Method" for creating the Cmdline to configure the Start script, Stop script, and Check script respectively as follows:

  - Start script

    ```
    /opt/SMAW/bin/hvzone -c -z <zone_name> -a <app_name> {-s|-n} -t <timeout> -b {Solaris8|Solaris9}
    ```

  - Stop script

    ```
    /opt/SMAW/bin/hvzone -u -z <zone_name> -a <app_name> {-s|-n} -t <timeout> -b {Solaris8|Solaris9}
    ```

  - Check script

    ```
    /opt/SMAW/bin/hvzone -m -z <zone_name> -a <app_name> {-s|-n} -t <timeout> -b {Solaris8|Solaris9}
    ```

  The differences above are only with the -c, -u, and -m options. Specify the zone name to be controlled and the cluster application name configured in the non-global zone to <zone_name> and <app_name> respectively.

Specify in seconds the shutdown process time out value for <timeout>. During Offline processing, this script performs RMS stop and non-global zone stop, but if the processing does not stop even after the time designated here has elapsed, stop the non-global zone using the halt command for zoneadm (zoneadm -z <zone_name> halt).

If sharing non-global zones between cluster nodes, specify the -s option. If not sharing, specify the -n option.

An example is given below. In this example, the configuration is as follows: the zone name is zone1, the cluster application name in the non-global zone is userApp_0, the timeout value is 200 seconds, and the non-global zone is shared between cluster nodes in the global zone.

To control Solaris 8 Containers, add -b Solaris8 option in the end.

To control Solaris 9 Containers, add -b Solaris9 option in the end.

An example is given below. In this example, the configuration is as follows: the zone name is zone1, the cluster application name in the non-global zone is userApp_0, the timeout value is 200 seconds, and the non-global zone is shared between cluster nodes in Solaris 9 Containers.

- Start script

```
/opt/SMAW/bin/hvzone -c -z zone1 -a userApp_0 -s -t 200 -b Solaris9
```

- Stop script

```
/opt/SMAW/bin/hvzone -u -z zone1 -a userApp_0 -s -t 200 -b Solaris9
```

- Check script

```
/opt/SMAW/bin/hvzone -m -z zone1 -a userApp_0 -s -t 200 -b Solaris9
```

After that, set up the script attributes. Click the "Flag" button and set the following values.

Flag	Overview
ALLEXITCODES	Change this to "Yes."
AUTORECOVER	For single-node cluster operations, set "Yes." In this case, do not set the following resources. They are unnecessary for single-node cluster operations because the following resources are used to take over IP addresses among multiple nodes. - Gls resource on the non-global zone - Cmdline resource (shared IP control) on the global zone (For details, see "16.3.5.3 Creating the Cmdline Resource for Shared IP Control.")
STANDBYCAPABLE	If the non-global zone is not shared between cluster nodes, change this to "Yes."
TIMEOUT	The default value is 300 seconds. Set a value large than the total of the following values: the time for the entire startup sequence to finish and the time it takes for the cluster application defined in the non-global zone to enter Online status. It is about 900 seconds.

- If not conducting application monitoring within the non-global zone or single-node cluster is operated:

  1. Placing start_zone.sh script, stop_zone.sh script, and check_zone.sh script

     Create script files for each Cmdline resource. Moreover, create script files for all nodes which use Cmdline resources.

     Below is the example. The contents can be modified according to the elements.

     - Start script: /var/tmp/PCL/rmstools/start_zone.sh
     - Stop script: /var/tmp/PCL/rmstools/stop_zone.sh
     - Check script: /var/tmp/PCL/rmstools/check_zone.sh

     To control Solaris 8 Containers or Solaris 9 Containers, edit the contents of start_zone.sh according to the comments below.

- Create script file

```
# vi /var/tmp/PCL/rmstools/start_zone.sh

(Paste the following the content of Start script file)

# vi /var/tmp/PCL/rmstools/stop_zone.sh

(Paste the following the content of Stop script file)

# vi /var/tmp/PCL/rmstools/check_zone.sh

(Paste the following the content of Check script file)

# chmod +x /var/tmp/PCL/rmstools/start_zone.sh
# chmod +x /var/tmp/PCL/rmstools/stop_zone.sh
# chmod +x /var/tmp/PCL/rmstools/check_zone.sh
```

- Start script

```
#!/bin/sh

MYZONE=$1

zoneadm -z $MYZONE list -p | grep :configured:
if [ $? -eq 0 ]; then
   zoneadm -z $MYZONE attach -F || exit $?
fi

zoneadm -z $MYZONE list -p | grep :running:
if [ $? -eq 0 ]; then
   zoneadm -z $MYZONE reboot
   RET=$?
else
# Remove # of the below line if MYZONE is Solaris 8 container
# /usr/lib/brand/solaris8/s8_p2v $MYZONE
# Remove # of the below line if MYZONE is Solaris 9 container
# /usr/lib/brand/solaris9/s9_p2v $MYZONE
   zoneadm -z $MYZONE boot
   RET=$?
fi

exit $RET
```

- Stop script

```
#!/bin/sh

MYZONE=$1

RET=0
RET2=0

zoneadm -z $MYZONE list -p | grep :running:
if [ $? -eq 0 ]; then
   zoneadm -z $MYZONE halt
   RET=$?
fi

zoneadm -z $MYZONE list -p | grep :installed:
if [ $? -eq 0 ]; then
   zoneadm -z $MYZONE detach
   RET2=$?
fi
```

```
if [ $RET -eq 0 ]; then
   exit $RET2
fi
exit $RET
```

- Check script

```
#!/bin/sh

# Return Offline if zlogin to the NGZ does not end in 30 seconds (Please change if needed)
TIMEOUT=30
MYZONE=$1

zoneadm -z $MYZONE list -p | grep :running: > /dev/null 2>&1
RET=$?

if [ $RET -ne 0 ]; then
   exit $RET
fi

/usr/sbin/zlogin $MYZONE "/usr/bin/ls >/dev/null 2>&1" 2>/dev/null &
PID=$!

i=0
while [ $i -lt $TIMEOUT ]
do
   ps -p $PID > /dev/null 2>&1
   if [ $? -ne 0 ]; then
      wait $PID
      exit $?
   fi
   sleep 1
   i=`expr $i + 1`
done
exit 1
```

2. Creating the Cmdline resource

   Select "Path Input" from "Creation Method" for creating the Cmdline to configure the Start script, Stop script, and Check script respectively as follows (When the zone name is zone1).

   - Start script

```
/var/tmp/PCL/rmstools/start_zone.sh zone1
```

   - Stop script

```
/var/tmp/PCL/rmstools/stop_zone.sh zone1
```

   - Check script

```
/var/tmp/PCL/rmstools/check_zone.sh zone1
```

   After that, Set attributes for scripts. Click the "Flag" button and set the following values.

Flag	Overview
AUTORECOVER	- For single-node cluster operations, set "Yes." - In this case, do not set the following resources. They are unnecessary for single-node cluster operations because the following resources are used to take over IP addresses among multiple nodes.

Flag	Overview
	For the Cmdline resource (shared IP control) on the global zone (For details, see "16.3.5.3 Creating the Cmdline Resource for Shared IP Control."

16.3.5.3 Creating the Cmdline Resource for Shared IP Control

If using the non-global zone in the shared IP zone but not sharing non-global zone images, create a resource for the takeover IP address in the global zone. For details, see "16.2.5.3 Creating the Cmdline Resource for Shared IP Control" of "13.2.5 Reconfiguration of Cluster Applications on Global Zone."

Note

In Solaris 8 Containers environment and Solaris 9 Containers environment, IPv6 addresses cannot be used for takeover IP addresses.

16.3.5.4 Creating Cluster Applications

In addition to the Gds resource, Gls resource, and Fsystem resource previously registered to the global zone, add the Cmdline resource created in "3.5.2 Creating the Cmdline Resource for Non-Global Zone Control" and create the cluster application corresponding to the target non-global zone.

Non-global zone should be stopped to create the cluster application. When the non-global zone is active, take the procedure below to stop the non-global zone that is controlled by the cluster application on all the cluster nodes, and then create the cluster application.

```
# zlogin zone-a shutdown -i0 -g0 -y
```

For the procedure for creating the cluster applications, follow "6.7.2.1 Creating Standby Cluster Applications." However, there is the following difference in procedure.

- Cluster application attributes

  If using warm-standby, be sure to set the "Standby Transitions" to "ClearFaultRequest|StartUp|SwitchRequest."

16.4 Operation

For basic operational procedures, follow "Chapter 7 Operations," but in some cases special procedures and considerations are necessary. This section describes those sorts of special procedures and considerations.

Starting Up and Stopping the Non-Global Zone

When starting up or stopping the cluster applications with the PRIMECLUSTER on the global zone, the non-global zone and the applications running on it will start or stop accordingly. The startup and stop procedures do not change from the normal PRIMECLUSTER operations. For details, refer to "7.2 Operating the PRIMECLUSTER System."

When one performs a stop to the cluster applications, if using warm-standby, the RMS will stop with the non-global zone still started up. If using cold-standby, the non-global zones will enter a detached status in all nodes.

Points of Caution

- During system operation, the RMS and cluster applications on the non-global zone are monitored from the global zone. Because of this, do not perform operations of the PRIMECLUSTER on the non-global zone besides those methods which follow this manual.

  - When performing maintenance for the non-global zone, use the global zone maintenance mode. (For details, refer to "16.5 Maintenance").

- When you make the addition of cluster applications and configuration changes on the global-zone, such as adding non-global zones, stop the RMS and all non-global zones.

- If sharing non-global zone images

  - Things such as software, parameters, and files are stored to the volume for the non-global zone's root file system. Because of this, if these files are destroyed, the failover may fail.

- Since the various types of logs on the non-global zone are stored to the volume for the non-global zone's root file system, they are output to the same log files on the operational system and standby system.

16.5 Maintenance

16.5.1 Maintenance Operations on the Non-Global Zone

The non-global zones are monitored and controlled by the Cmdline resource set up to the global zone. Because of this, if operating the RMS or applications on the non-global zone while the global zone RMS is running, this may be detected as an error on the global zone side. Moreover, because of this events such as forced stop of the non-global zone and switchover may occur.

As such, if performing maintenance tasks which include the following operations, follow the maintenance task procedures explained below.

- Starting up/stopping the non-global zone
- Starting up/stopping the non-global zone RMS
- Starting up/stopping the cluster applications being controlled by the non-global zone RMS

As a basic way of thinking, use the maintenance mode on the global zone for maintenance on the non-global zone.

For the following procedures, it is assumed that the global cluster applications that are to undergo maintenance are in Online status. For troubleshooting in situations of some kind of error occurring on the non-global zone and being unable to put the applications into Online status, refer to "16.5.4 Recovery Operation for When an Error Occurs on the Non-Global Zone."

Note

If the virtual IP addresses of fast switching mode or GS/SURE linkage mode are shared in the shared IP zone, cancel the shared state according to the following procedure before stopping the non-global zone.

Shared logical interfaces will be deleted when stopping the non-global zone. Therefore, you need to cancel the state beforehand so that logical virtual interfaces controlled by GLS are not deleted.

Perform this procedure in the global zone.

1. By using the ifconfig command, check whether the zone setup of the non-global zone name to be stopped (settings of the shared state) is displayed for the logical virtual interfaces (interfaces beginning with "sha") of fast switching mode or GS/SURE linkage mode.

```
# ifconfig -a
sha10:65: flags=1000863<UP,BROADCAST,NOTRAILERS,RUNNING,...
zone zone-a
inet 192.168.100.101 netmask ffffff00 broadcast 192.168.110.255
```

2. Cancel the shared state if it is set.

```
# ifconfig sha10:65 -zone
# ifconfig -a
sha10:65: flags=1000863<UP,BROADCAST,NOTRAILERS,RUNNING,...
inet 192.168.100.101 netmask ffffff00 broadcast 192.168.110.255
```

If GLS logical virtual interfaces are deleted by stopping the non-global zone before performing the procedure by mistake, refer to "Accidentally deleted the virtual interface with ifconfig command" of "Troubleshooting" in the "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function," and then recover GLS logical virtual interfaces as necessary.

The latest release of the following PRIMECLUSTER patches need to be applied to the non-global zone in an OSLC environment.

- Solaris 8
    - When PRIMECLUSTER 4.1A30 was used in the migration source

901167-07 or later
901172-34 or later

901173-24 or later
913855-05 or later
914111-03 or later
914112-10 or later
914120-01 or later
914346-01 or later
914351-01
914530-01
915102-01 or later

- When PRIMECLUSTER 4.1A40 was used in the migration source and using a new PRIMECLUSTER

901167-07 or later
901172-34 or later
901173-24 or later
913855-05 or later
914111-03 or later
914112-10 or later
914346-01 or later
914351-02 or later
914530-01
915102-01 or later

- Solaris 9

    - When PRIMECLUSTER 4.1A30 was used in the migration source

901167-07 or later
901172-34 or later
901173-24 or later
913855-05 or later
914111-03 or later
914112-10 or later
914120-01 or later
914346-01 or later
914351-01
914530-01
915102-01 or later

    - When PRIMECLUSTER 4.1A40 was used in the migration source

901167-07 or later
901172-34 or later
901173-24 or later
913855-05 or later

914111-03 or later
914112-10 or later
914346-01 or later
914351-02 or later
914530-01
915102-01 or later

- When PRIMECLUSTER 4.2 was used in the migration source and using a new PRIMECLUSTER

901196-35 or later
901215-04 or later
901217-29 or later
901254-02 or later
913855-05 or later
914111-03 or later
914112-10 or later
914346-01 or later
914351-02 or later
914530-02 or later
915102-01 or later

If sharing non-global zone images or single-node cluster operations

If sharing non-global zone images and applying a patch to the non-global zones, follow the procedure below. If sharing non-global zone images, patch application by rolling update is not supported.

1. Migration to maintenance mode

   Migrate the corresponding cluster application from the global zone to maintenance mode. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

2. Performing maintenance tasks on the non-global zone

   Perform the necessary maintenance tasks on the non-global zone in the Online state. While the maintenance mode is set on the global zone, it is permitted to use the non-global zone RMS commands (hvcm, hvshut, hvswitch, hvutil, hvdisp, etc.) and to perform the starting up and stopping of the non-global zone as needed.

   For example, if applying a patch to the non-global zone, perform the following procedure.

   1. Stop the non-global zone. From the global zone, login with zlogin and execute the shutdown command or, as shown below, execute a direct shutdown command from the global zone with a zlogin command.

      ```
      # zlogin zone-a shutdown -y -g0 -i0   *1
      ```

      *1: "zone-a" is the zone name (It is the same for below).

   2. Check that the non-global zone has stopped. To be specific, check that STATUS is "installed" using the zoneadm list command.

      ```
      # zoneadm list -vc
        ID NAME         STATUS   PATH                           BRAND  IP
         0 global           running    /                                  native   shared
         - zone-a           installed  /zone-a-system                     native   shared
      ```

3. Start up the non-global zone under single user mode.

```
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)
# zoneadm -z zone-a boot -s
```

4. Apply the patch to the non-global zone.

5. Restart the non-global zone.

```
# zlogin zone-a shutdown -y -g0 -i6
```

Note

- If not sharing non-global zone images, apply the same patch on all of the non-global zones which comprise the cluster system allocated to each global zone.

- Be careful that cluster applications allocated to each global zone are not simultaneously Online.

3. Non-global zone status recovery

   Before canceling maintenance mode, return the non-global zone to the same status it was in at the time it was migrated to maintenance mode. In other words, return it to a status of the non-global zone being started up and the RMS and applications on it being started up. If one manually put the Gds, Gls, Fsystem and other resources into inactive status, also return those resources to active status.

   1. Check that the non-global zone has restarted. This status can be checked by acquiring the console with zlogin -C zone-a.

      For single-node cluster operations, the recovery of the cluster applications is performed automatically to the point of startup. Proceed to the next step, "4. Canceling maintenance mode."

```
# zlogin -C zone-a
```

   2. After login to the non-global zone, start up the RMS by using the procedure below.

```
# hvcm
```

   3. Restore the cluster application on the non-global zone which was in the Online state to the original state.

```
# hvswitch userApp_0   *1
```

      *1: "userApp_0" is the name of the cluster application on the non-global zone.

4. Canceling maintenance mode

   In the global zone, remove the corresponding cluster applications from maintenance mode. It is possible to use GUI and CLI. For the procedures, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

If not sharing non-global zone images

If not sharing non-global zone images, patch application by rolling update is supported.

If applying a patch by rolling update to the non-global zone, perform the following procedure according to the example below.

Example) If applying a patch to the non-global zone on the operational system global zone GZA and standby system global zone GZB using the rolling update.

1. Execute the following command on GZA and switch the cluster application controlling the non-global zone that is to undergo maintenance over to the standby system GZB.

```
# hvswitch userApp_0 GZBRMS   *1
```

   *1: "userApp_0" is the name of the cluster application controlling the non-global zone, whereas "GZBRMS" is the SysNode name.

2. Migrate the cluster application controlling the non-global zone which is to undergo maintenance to maintenance mode. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

3. Check the status of the non-global zone.

```
# zoneadm list -vc
```

4. If STATUS is "configured," attach the non-global zone.

```
# zoneadm -z zone-a attach
```

5. If STATUS is "running," stop the non-global zone. Log in from GZA with zlogin and execute the shutdown command or, as shown below, execute a direct shutdown command from the global zone using the zlogin command.
However, the Cmdline resource that controls the global zones may be in Fault state depending on the settings, continue to perform the steps from now on.

```
# zlogin zone-a shutdown -y -g0 -i0   *2
```

*2: "zone-a" is the zone name (It is the same for below).

6. Check that the non-global zone has stopped. To be specific, execute the zoneadm list command in GZA and check that STATUS is "installed."

```
# zoneadm list -cv
ID NAME STATUS PATH BRAND IP
0 global running / native shared
- zone-a installed /zone-a-system native excl
```

7. From GZA, start up under single user mode the non-global zone which is to undergo maintenance.

```
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)
# zoneadm -z zone-a boot -s
```

8. Apply the patch to the non-global zone.

9. If STATUS is "configured" in Step 3, detach the non-global zone.

```
# zoneadm -z zone-a detach
```

10. If STATUS is "running," restart the non-global zone and start RMS on the non-global zone. If the Cmdline resource which controls the non-global zone becomes Faulted state, wait until the Faulted state of the Cmdline resource is cleared.

```
# zlogin zone-a shutdown -y -g0 -i6   *3
# zlogin zone-a
# hvcm   *4
# hvutil -s userApp_0  *5
```

*3: After this, check that the non-global zone has restarted. This status can be checked by acquiring the console with zlogin -C zone-a.

*4: After login to the non-global zone, start up the RMS and application using the procedure below.

*5: Execute this command when performing the warm standby operation and register resources that can become a standby state to the non-global zone cluster applications. "userApp_0" is the name of a cluster application on the non-global zone.

11. Cancel the maintenance mode for the corresponding cluster applications. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

12. Apply the patch to the non-global zone allocated to GZB as well. Repeat steps 1) through 8). However, replace "GZA" in the procedure with "GZB" and "GZB" in the procedure with "GZA."

Maintenance to Oracle, NetWorker, and NAS device on the Non-Global Zone

When performing maintenance operations for Oracle, Networker, and NAS device on the non-global zone, it is necessary to temporarily suspend monitoring performed by PRIMECLUSTER or Wizard products.

For maintenance procedures for Wizard products and points of caution, refer to each Wizard product manual.

Applying Corrections to Wizard Products on the Non-Global Zone

If applying the PRIMECLUSTER patches to non-global zone Wizard products, put the non-global zone into single user mode and then apply the corrections. At this time, use maintenance mode on the global zone in the same way as for "16.5.1 Maintenance Operations on the Non-Global Zone."

For the procedures and points of caution for applying corrections to Wizard products, refer to each Wizard product manual.

16.5.2 Maintenance Operations on the Global Zone

For the procedure of applying a patch for PRIMECLUSTER to the global zone, follow "Chapter 12 Maintenance of the PRIMECLUSTER System." However, if it is a patch such as an OS patch, and one has applied the patch to the global zone, there is a patch that is also to be applied to the non-global zone. In this instance, the procedure differs depending whether or not the configuration shares non-global zone images. Apply the patch using the respective procedure below.

Procedure for Applying a Patch for a Configuration Sharing Non-Global Zone Images between Cluster Nodes or single-node cluster operations

Perform each procedure in all of the global zones. The situation of performing this with just one of any of the global zones is explained in this procedure.

1. Stop the RMS.

   In any of the global zones, execute the following command.

```
# /opt/SMAW/bin/hvshut -a
```

   Check that the RMS has stopped.

```
# /opt/SMAW/bin/hvdisp -a
hvdisp: RMS is not running
#
```

2. To disable the automatic startup for RMS, edit the file "/opt/SMAW/SMAWRrms/bin/hvenv.local."

   If the file "/opt/SMAW/SMAWRrms/bin/hvenv.local" does not exist, create the file (create the file access privilege in 644), and write in the line shown in [After changes].

   [Before changes]

```
export HV_RCSTART=1
```

   [After changes]

```
export HV_RCSTART=0
```

   After editing the file, check that the "HV_RCSTART" value is at 0.

```
# cat /opt/SMAW/SMAWRrms/bin/hvenv.local
export HV_RCSTART=0
```

3. Start up the system again under single user mode.

```
# shutdown -g0 -i0 -y
```

   Connect to each global zone console. If the ok prompt is displayed on the console, execute the following.

```
ok boot -s
```

4. Login under single user mode and mount the file system.

```
# mountall -l
# zfs mount -a
```

5. Check that all of the non-global zones are detached.

   If the non-global zone STATUS is "configured" in the output results for the "zoneadm list -vc" command, then the non-global zones are detached.

Example

If the non-global zone name is zone-a and the zone path is /zone-a-system

```
# zoneadm list -vc
ID NAME         STATUS       PATH              BRAND    IP
0  global       running      /                 native   shared
-  zone-a       configured   /zone-a-system    native   shared
```

6. Apply the patch.

Example

If an OS patch for Solaris 10

```
# patchadd <patch number>
```

Note

- Make sure that all of the patches being applied to the global zones are the same version level.
- The applying of patches by rolling update is not supported.
- If applying multiple patches, apply all of the patches at this point.

7. Restart under multiuser mode.

```
# shutdown -i6 -g0 -y
```

8. Check that the RMS is stopped.

```
# /opt/SMAW/bin/hvdisp -a
hvdisp: RMS is not running
```

9. From any global zone, start up the GDS volume to which the non-global zone is installed. For a single-node cluster, proceed to Step 11.

```
# sdxvolume -N -c class0001 -v volume0001
```

10. Mount the GDS volume that corresponds to the non-global zone in the same global zone which was used for Step 9.

    It is possible to confirm each non-global zone mount point with /etc/vfstab.pcl.

Example

If the zone path is /zone-a-system

```
# cat /etc/vfstab.pcl
...omitted...
```

```
#RMS#/dev/sfdsk/class0001/dsk/volume0001 /dev/sfdsk/class0001/rdsk/volume0001 /zone-a-system
ufs - no -
```

In the case of this example, perform the mount as follows.

```
# mount -F ufs /dev/sfdsk/class0001/dsk/volume0001 /zone-a-system
```

11. The patch is applied through attaching the non-global zone (example: if the non-global zone name is zone-a) with the same global zone which was used for Step 9.

```
# zoneadm -z zone-a attach -u
```

Note

Since information on applying the patch is displayed on screen, check that there are no errors.

It is possible to check the application log with the non-global zone's "/var/sadm/system/logs/update_log."

12. Check that non-global zone is attached in the same global zone which was used for Step 9.

This can be determined by checking that the non-global zone STATUS is "installed."

Example

If the non-global zone name is zone-a and the zone path is /zone-a-system

```
# zoneadm list -vc
ID NAME            STATUS     PATH              BRAND    IP
0  global          running    /                 native   shared
-  zone-a          installed  /zone-a-system    native   shared
```

If one wishes to check if the patch was applied to the non-global zone, perform the following Steps 13 through 18. If confirmation is not necessary, proceed to Step 19.

13. Start up the non-global zone in the same global zone which was used for Step 9.

```
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)
# zoneadm -z zone-a boot
```

It is possible to check that the non-global zone has started up by using the following command.

```
# zoneadm list -vc
ID NAME            STATUS     PATH              BRAND    IP
0  global          running    /                 native   shared
1  zone-a          running    /zone-a-system    native   shared
```

14. Log in to the non-global zone using the same global zone which was used for Step 9.

```
# zlogin zone-a
```

15. Check that the patch has been applied to the non-global zone.

Example

If applying an OS patch for Solaris 10

```
[zone-a]# showrev -p | grep <patch number>
```

16. Logout from the non-global zone.

```
[zone-a]# exit
```

17. Stop the non-global zone in the same global zone which was used for Step 9.

```
# zlogin zone-a shutdown -i0 -g0 -y
```

18. Detach the non-global zone in the same global zone which was used for Step 9.

```
# zoneadm -z zone-a detach
```

Check that zone-a is detached.

It is possible to determine this by checking if the zone-a STATUS is "configured."

```
#zoneadm list -vc
ID NAME          STATUS        PATH                BRAND    IP
0  global        running       /                   native   shared
-  zone-a        configured    /zone-a-system      native   shared
```

19. If one changed the RMS automatic startup at Step 2, in order to enable the RMS automatic startup, edit the "/opt/SMAW/SMAWRrms/bin/hvenv.local" file.

[Before changes]

```
export HV_RCSTART=0
```

[After changes]

```
export HV_RCSTART=1
```

After editing the file, check that the "HV_RCSTART" value is at "1."

```
# cat /opt/SMAW/SMAWRrms/bin/hvenv.local
export HV_RCSTART=1
```

20. Restart the system.

```
# shutdown -i6 -g0 -y
```

After startup, check that RMS is started up on all global zones.

Example

```
# /opt/SMAW/bin/hvdisp -a
Local System:  GZARMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us
Resource                   Type    HostName            State      StateDetails
-----------------------------------------------------------------------------
GZBRMS                     SysNode                     Online
GZARMS                     SysNode                     Online
userApp_0                  userApp                     Online
Machine001_userApp_0       andOp   GZBRMS
Machine000_userApp_0       andOp   GZARMS              Online
ManageProgram000_Cmdline0 gRes                        Online
MountPoint001_Fsystem0     gRes                        Online
AllDiskClassesOk_Gds0      andOp                       Online
class0001_Gds0             gRes                        Online

#
```

Procedure for Applying a Patch for a Configuration Not Sharing Non-Global Zone Images between Cluster Nodes

Apply the patch as one normally would to the global zone. If performing the application of the patch by rolling update, perform the task by referring to example below. Perform each of these steps to all of the global zones.

Example) If applying a patch performed by rolling update to the operational system global zone GZA and the standby global system GZB.

1. If cluster applications are running, switch all cluster applications over to the standby system.

   On GZA, execute the following command and switch the cluster application over to the standby system GZB.

   ```
   # hvswitch userApp_0 GZBRMS
   ```

   "userApp_0" is the name of the cluster application on the global zone and "GZBRMS" is the SysNode name.

   Check that the user application is Offline in GZA.

   ```
   # /opt/SMAW/bin/hvdisp -a
   Local System:  GZBRMS
   Configuration: /opt/SMAW/SMAWRrms/build/config.us
   Resource                  Type    HostName            State      StateDetails
   -----------------------------------------------------------------------------
   GZARMS                    SysNode                     Online
   GZBRMS                    SysNode                     Online
   userApp_0                 userApp                     Standby
   userApp_0                 userApp GZARMS              Online
   Machine001_userApp_0      andOp   GZBRMS              Offline
   Machine000_userApp_0      andOp   GZARMS
   ManageProgram000_Cmdline0 gRes                        Standby

   #
   ```

2. Start up the GZA again under single user mode.

   ```
   # shutdown -g0 -i0 -y
   ```

   Connect to the console for GZA. If the ok prompt is displayed on the console, execute the following.

   ```
   ok boot -s
   ```

3. Log in to GZA under single user mode and mount the file system.

   ```
   # mountall -l
   # zfs mount -a
   ```

4. If using cold-standby, attach the non-global zone.

   ```
   # zoneadm -z zone-a attach
   ```

5. Check that the non-global zone is attached in GZA.

   If the non-global zone STATUS is given as "installed" in the "zoneadm list -vc" command output result, then the non-global zone is attached.

   Example

   If the non-global zone name is zone-a and the zone path is /zone-a-system

   ```
   # zoneadm list -vc
   ID NAME         STATUS      PATH               BRAND    IP
   0 global        running     /                  native   shared
   - zone-a       installed    /zone-a-system     native   excl
   ```

6. Apply the patch in GZA.

Example

If applying an OS patch for Solaris 10

```
# patchadd <patch number>
```

Note

- Make sure that the patch applied to all of the global zones is the same version level.
- If applying multiple patches, apply all of the patches at this point.

7. If using cold-standby, detach the non-global zone.

```
# zoneadm -z zone-a detach
```

8. Restart GZA under multiuser mode.

```
# shutdown -i6 -g0 -y
```

After startup, check that RMS is started up on all global zones. Also, check that the userApp is in Standby on GZA and that userApp is Online on GZB.

Example

If GZA

```
# /opt/SMAW/bin/hvdisp -a
Local System:  GZARMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us
Resource                  Type    HostName            State        StateDetails
-----------------------------------------------------------------------------
GZBRMS                    SysNode                     Online
GZARMS                    SysNode                     Online
userApp_0                 userApp                     Standby
userApp_0                 userApp GZBRMS              Online
Machine001_userApp_0      andOp   GZARMS              Offline
Machine000_userApp_0      andOp   GZBRMS
ManageProgram000_Cmdline0 gRes
#
```

Example

If GZB

```
# /opt/SMAW/bin/hvdisp -a
Local System:  GZBRMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us
Resource                  Type    HostName            State        StateDetails
-----------------------------------------------------------------------------
GZARMS                    SysNode                     Online
GZBRMS                    SysNode                     Online
userApp_0                 userApp                     Online
Machine001_userApp_0      andOp   GZARMS
Machine000_userApp_0      andOp   GZBRMS              Online
```

```
ManageProgram000_Cmdline0 gRes                          Online
#
```

9. Apply the patch to GZB as well. Repeat Steps 1) through 8). However, replace "GZA" in the procedure with "GZB" and "GZB" with "GZA."

16.5.3 Method for Collecting Troubleshooting Information for the Non-Global Zone

If performing application monitoring, on the installed non-global zone, collect the troubleshooting information from all of the global zones and all of the non-global zones.

For details, refer to "C.1 Collecting Troubleshooting Information" and manuals for each Wizard product and each application.

For the non-global zone, collect the investigation information with the following procedure:

1. Use the zlogin command to log in from the global zone.

```
# zlogin -C zone-a
```

2. Execute the command below to collect PRIMECLUSTER investigation information.

   For details on the the command, see "C.1.1 Executing the fjsnap Command" or "C.1.2 Collecting Information by FJQSS(Information Collection Tool)."

3. Execute the command to collect investigation information for Wizard products.

   For the command to collect investigation information for Wizard for Oracle, see the manual for PRIMECLUSTER Wizard for Oracle.

   For the command to collect investigation information for Wizard for NetWorker, see the manual for PRIMECLUSTER Wizard for NetWorker.

   No command to collect investigation information is provided for Wizard for NAS. Refer to Step 1 and 2, and collect its investigation information using the fjsnap or FJQSS.

16.5.4 Recovery Operation for When an Error Occurs on the Non-Global Zone

If using warm-standby, follow "7.4 Corrective Actions for Resource Failures" and perform the normal recovery operation.

In the following conditions, a non-global stop process will be performed using the Cmdline resource set up to the global zone if an error occurs to a non-global zone, or the RMS or applications running on the non-global zone:

- Configuration of cold-standby operations and sharing non-global zone images
- Single-node cluster operations

When performing a non-global stop process by the Cmdline resource, you cannot log into the non-global zone and perform recovery operations.

The following shows the troubleshooting procedure for such a case.

Note

Before performing this troubleshooting procedure, check that the corresponding cluster application is in a stop status on other nodes.

1. Clearing the Faulted status on the global zone

   Clear the Faulted status for the global zone cluster application. For the procedure using GUI, refer to "7.2.2.4 Bringing Faulted Cluster Application to Available State." If using CLI, execute hvutil -c <cluster application name>.

2. Migrating to maintenance mode

   From the global zone, migrate the corresponding cluster application to maintenance mode. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

3. Activating the Gds resource and Fsystem resource

   Manually perform the mounting of the file system and the startup of GDS volume necessary to the startup of the non-global zone. For the startup procedure for GDS volume, refer to "PRIMECLUSTER Global Disk Services Configuration and Administration Guide." Mount the file system using the mount command.

   This step is not necessary in the case of not sharing non-global zone images or a single-node cluster is operated.

4. Attaching and starting up the non-global zone

   After stopping the cluster application, the non-global zone enters a detached status, so perform the attaching and starting up of the non-global zone using zoneadm attach and zoneadm boot.

```
# zoneadm -z zone-a attach -F   *1
# /usr/lib/brand/solaris8/s8_p2v zone-a  (For Solaris 8 Container)   *2
# /usr/lib/brand/solaris9/s9_p2v zone-a  (For Solaris 9 Container)   *2
# zoneadm -z zone-a boot   *2
```

   *1: Perform this only if sharing a non-global zone between cluster nodes.

   *2: "zone-a" is the zone name.

5. Performing the maintenance operation in the non-global zone

   Perform the necessary maintenance operations on the non-global zone. One is permitted to use RMS commands (hvcm, hvshut, hvswitch, hvutil, hvdisp, etc.) as needed.

   Note

   If something such as a shared disk for data takeover is having maintenance performed using a resource set up to the RMS, be sure that the RMS for the non-global zones allocated to each of the global zones are not simultaneously Online.

6. Stopping and detaching the non-global zone

   Return the non-global zone to a detached status. From the non-global zone, after executing such commands as "shutdown -y -g0 -i0" and stopping the non-global zone, detach the non-global zone using zoneadm detach.

```
# shutdown -y -g0 -i0
# zoneadm -z zone-a detach
```

7. Deactivating the Gds resource and Fsystem resource

   Return the resources activated with the above Step 3 "Activating the Gds resource and Fsystem resource" back to a deactivated status.

   Manually perform an unmount of the file system and a stopping of the GDS volume. Unmount the file system using the unmount command. For the procedure on stopping the GDS volume, refer to "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

   If the Gls resource was also activated, manually perform deactivation.

   This step is not necessary in the case of not sharing non-global zone images or a single-node cluster is operated.

8. Canceling the maintenance mode

   From the global zone, remove the corresponding cluster application from maintenance mode. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

9. Starting up the cluster application

   Start up the cluster application from the global zone as needed.

16.6 Uninstallation Procedure

16.6.1 Uninstalling PRIMECLUSTER from Non-Global Zones

When services to be monitored have been installed in a non-global zone, uninstall PRIMECLUSTER from the non-global zone with the following procedure. If the non-global zone is shared between cluster nodes, uninstall it only in the active node. If not, uninstall it in all nodes.

16.6.1.1 Migrate Applications to Maintenance Mode

Migrate the corresponding cluster application from the global zone to maintenance mode. It is possible to use GUI and CLI. For the procedure, refer to "8.4 Using maintenance mode" of the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

16.6.1.2 Uninstalling Wizard Products

Uninstall Wizard products. For the procedure for uninstalling Wizard products and its consideration, see the respective Wizard product manuals.

16.6.1.3 Uninstalling PRIMECLUSTER on Non-Global Zones

Boot the non-global zone in single-user mode. For the procedure, see the example for applying patches stated in "16.5.1 Maintenance Operations on the Non-Global Zone."

After that, uninstall PRIMECLUSTER with the following procedure.

Log in to the non-global zone and uninstall PRIMECLUSTER in the non-global zone with the following procedure. If you are asked to input some value while executing the pkgadd(1M) command, input the default value. If there is no default value, select "y."

```
# /usr/sbin/removef SMAWRhvto /opt/SMAW/bin/hvksh
# /usr/sbin/removef -f SMAWRhvto
# pkgrm FJSVhanet
# pkgrm FJSVclsfw
# pkgrm FJSVclrwz
# pkgrm FJSVhvgd
# pkgrm FJSVhvgl
# pkgrm SMAWRhvde
# pkgrm SMAWRhvdo
# pkgrm SMAWRhvba
# pkgrm SMAWRhvto
# pkgrm FJSVclrms
# pkgrm SMAWRdfw
# pkgrm SMAWRrms
# pkgrm FJSVclapm
# pkgrm FJSVcldbm
# pkgrm FJSVclapi
# pkgrm FSUNclnet
# pkgrm SMAWcf
# pkgrm FJSVwvsfw
```

```
# pkgrm FJSVwvucw
# pkgrm FJSVwvfrm
# pkgrm SMAWrcaja
# pkgrm SMAWrcadm
# pkgrm SMAWcj2re
# pkgrm FJSVwvmpc
# pkgrm FJSVwvbs
# pkgrm FJSVwvcnf
# pkgrm SMAWskel
```

Delete FJSVsnap if unnecessary.

```
# pkgrm FJSVsnap
```

Next, delete the environment definition file of PRIMECLUSTER according to the following procedure.

```
# cd /opt/SMAW
# rm -r SMAWRrms
# rm -r SMAWcf
# cd /etc
# rm cip.cf
# cd /etc/default
# rm cluster
# rm cluster.config
```

Delete the entry in /etc/inet/hosts added in "16.3.4.5 Installation of PRIMECLUSTER to the Non-Global Zone."

```
# vi /etc/inet/hosts
```

Log out the non-global zone to execute the hvzonesetup script in a global zone.

```
# hvzonesetup -z zone-a -r
NOTICE: Restoring original setup in zone zone-a.
NOTICE: PCL setup removed in zone zone-a.
```

Information

The following process is performed in this script.

- Deleting PATH settings from /etc/profile

16.6.2 Uninstalling PRIMECLUSTER from the Global Zone

The uninstallation procedure for PRIMECLUSTER from the global zone is the same as in a usual cluster environment.

16.7 Recommended Configuration

This section shows design examples of PRIMECLUSTER system for Oracle Solaris Zones environments. Refer to these examples when designing the PRIMECLUSTER system

System configuration

Figure 16.17 System configuration example (exclusive IP zones)

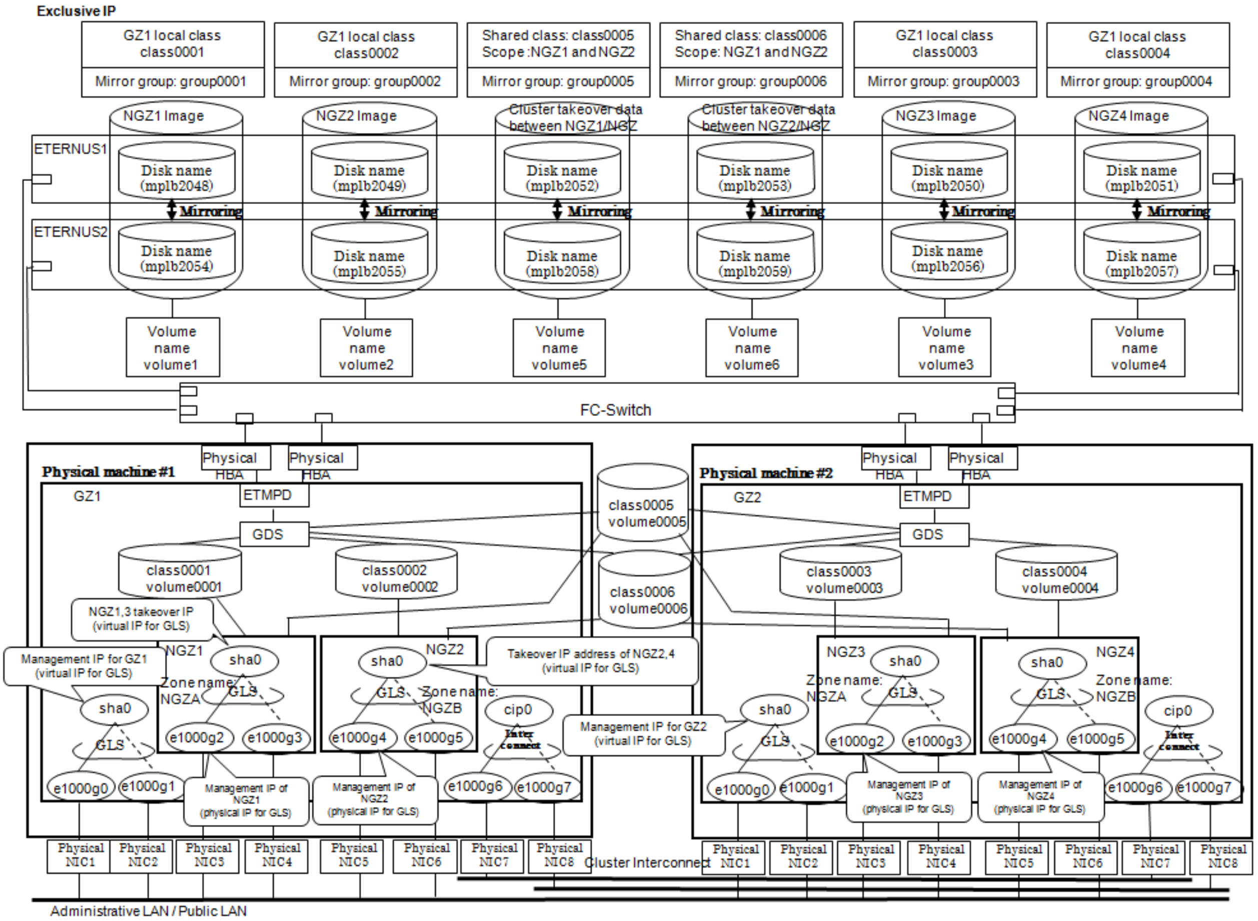
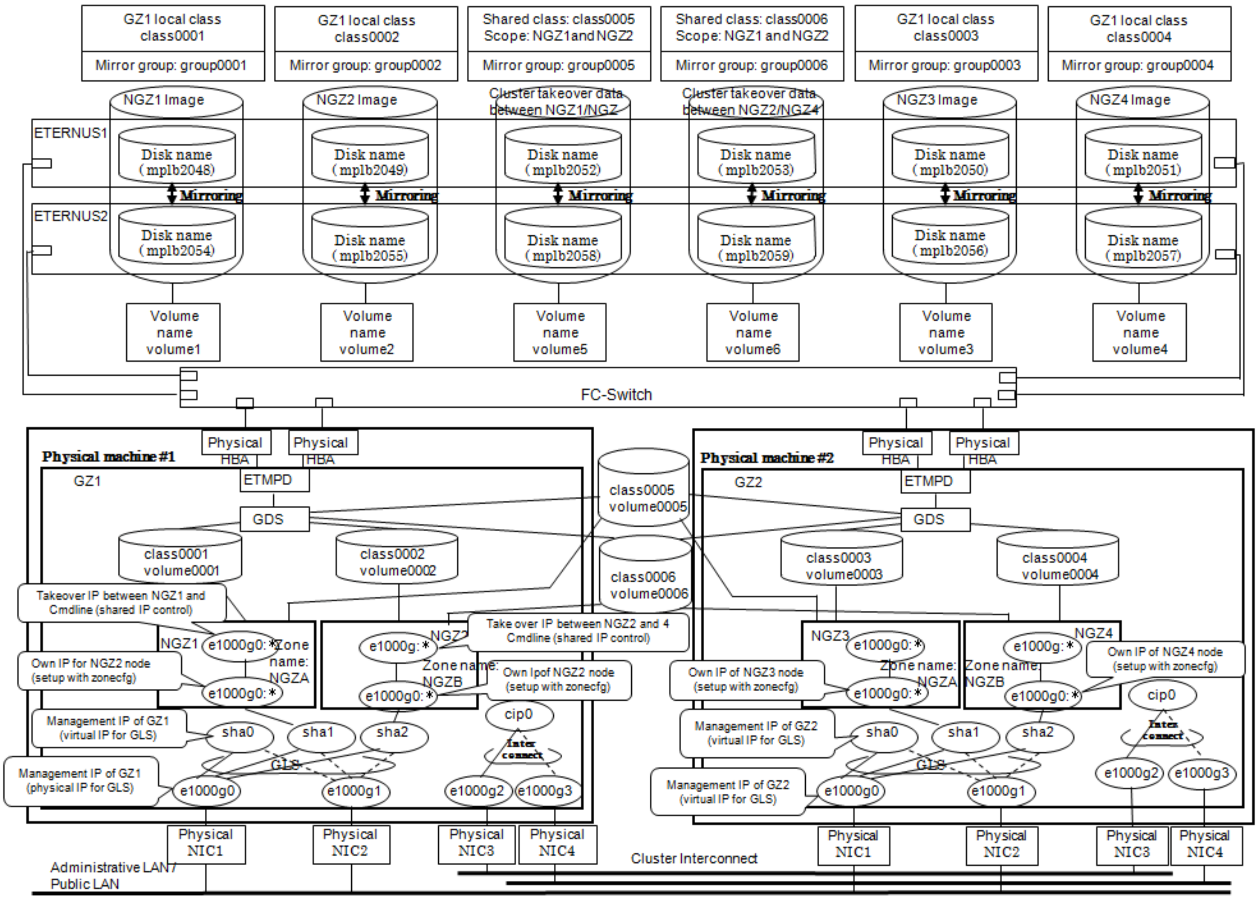


Figure 16.18 System configuration example (shared IP zones)



Used hardware

- SPARC Enterprise M8000

Used software

- Solaris 11
- Enhanced Support Facility 3.2
- PRIMECLUSTER 4.3
- Fujitsu PCI fibre Channel 4.0
- Internet Explorer 8 (installed in the operation management computer)
- J2SE(TM) Runtime Environment 6 (installed in the operation management computer)

16.7.1 Cluster Configuration Worksheet

Shown below is an example of creating the Cluster Configuration Worksheet.

	Item			Setting	Remarks
Shared disk unit: Type 1	Machine model			ETERNUS DX8100 S2	
	Device name			ETERNUSDX8100#002	
	IP address			10.33.66.215	
	Subnet mask			255.255.255.0	
	Interface			fibre channel	
	LUN number			0 - 0F	
	Ports			CM0CA2P0,CM1CA2P0	
	Connection paths			2	
	FC-SW		Use of fibre channel switching hub	■ yes □ no	
			Machine model	ETERNUS SN200 model 630	
			Device name	ETERNUSSN200#630	
			IP address	10.33.66.203	
			Subnet mask	255.255.255.0	
	File system selection (type of file system used in cluster)			UFS	
NTP	Operation mode			□ NTP server (integrated in cluster) ■ NTP client	
	Protocol (only when broadcast is specified)				
	Server	Primary	Host name	pclntp1	
			IP address	10.124.95.11	
			Subnet mask	255.255.255.0	
		Secondary	Host name	pclntp2	
			IP address	10.21.8.3	
			Subnet mask	255.255.255.0	
Web-Based Admin View	Operation mode			□ 1 ■ 2 □ 3 layers	
	Server	Primary	Host name	primecl01	
			IP address	10.34.214.181	
			Subnet mask	255.255.255.0	
			URL	http://10.34.214.181:8081/Plugin.html	
		Secondary	Host name	primecl02	
			IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			URL	http://10.34.214.182:8081/Plugin.html	
	User ID		wvroot group	pclwv	
			clroot group	pclcl	
			cladmin group	pcladm	
			clmon group	pclmon	
			sdxroot group	pclsdx	
Cluster Foundation (CF)	Cluster name			SLV002	
	Cluster nodes			2	
	Cluster node names			primecl01,primecl02	
	Subnets			1	
	Subnet number			192.168.1.0	
	Subnet mask			255.255.255.0	
	Use of CF remote services			■ cfcp (CF file copy) ■ cfsh (CF remote command execution)	
	IP interconnect			◆ [no] ◇ yes (the number of IP interconnects)	
	Use by RMS			◆ [yes] ◇ no	
Shutdown facility (SF)	Shutdown agent			□ RCI □ XSCF (SPARC M series) ■ XSCF (SPARC Enterprise M series) □ ILOM □ ALOM	
Cluster resource management facility (CRM)	Hardware devices stored in resource database			□ Network unit	Required for using takeover of public LAN (IP) and node names
				□ Line switching unit	Required for using line takeover function
				■ Disk unit	Required when using GDS and GFS

Node 1	Device name			DBSV#001	
	Global zone	Domain ID / domain name		-	
		Node name (uname -n)		primecl01	
		CF node name		primecl01	
		Interconnect	IP address	192.168.1.1	
			Subnet mask	255.255.255.0	
			Device name	e1000g6,e1000g7	
		Node in CF quorum set		◆ [yes] ◇ no	Do not change the default setting.
		CF cluster timeout (CLUSTER_TIMEOUT)		"10"	
		Use by RMS		◆ [yes] ◇ no Suffix:primecl01RMS	
		RCI	Address	-	Only when using RCI asynchronous monitoring
		Kernel parameter	scfd:scf_rdctrl_sense_wait	-	
		System console	IP address 1	10.33.66.71	XSCF, ALOM, and ILOM data (The setting of 2 is for SPARC M series)
			Subnet mask 1	255.255.255.0	
			IP address 2		
			Subnet mask 2		
			User ID	root	
			User authority	-	
			Connection type	■ SSH □ telnet	
		Shutdown facility (SF)	Administrative IP addresses	10.33.66.91	
			weight	1	Node weight (priority if cluster partition occurs) used by SF
			User ID	xuser	XSCF, ALOM, and ILOM data
			User authority	platadm	
			Console LAN#1	10.33.66.71	
			Console LAN#2		
			PPAR-ID		SPARC M series
			Domain name		
			Connection type	◆ [SSH] ◇telnet	[ ] Default For XSCF
		Use of shared disk		■ yes □ no	
		scsi-initiator-id (OBP variable #eeprom command)		7	
		SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11 - sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43 - sd74	
		Management LAN 1	IP address	10.33.66.91	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Management LAN 2	IP address	10.33.66.92	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 1	IP address	10.34.214.181	
			Subnet mask	255.255.255.0	
			Device name	e1000g2	
		Public LAN 2	IP address	10.34.215.21	
			Subnet mask	255.255.255.0	
			Device name	e1000g3	

	Non-global zone	Zone name (zonename)		NGZA	
		Node name (uname -n)		guest01	
		Public LAN 1	IP address	10.34.214.186	
			Subnet mask	255.255.255.0	
			Device name	sha0	
	Non-global zone	Zone name (zonename)		NGZB	
		Node name (uname -n)		guest02	
		Public LAN 1	IP address	10.34.214.187	
			Subnet mask	255.255.255.0	
			Device name	sha0	
	Non-global zone	Zone name (zonename)			
		Node name (uname -n)			
		Public LAN 1	IP address		
			Subnet mask		
			Device name		

		Use of shared disk		■ yes □ no	
		scsi-initiator-id (OBP variable #eeprom command)		8	
		SCSI / fibre channel	Circuits (paths)	2	
			Device name 1	fjpfca#0	
			sd device name	sd11 - sd42	
			Device name 2	fjpfca#1	
			sd device name	sd43 - sd74	
		Management LAN 1	IP address	10.33.66.93	
			Subnet mask	255.255.255.0	
			Device name	e1000g0	
		Management LAN 2	IP address	10.33.66.94	
			Subnet mask	255.255.255.0	
			Device name	e1000g1	
		Public LAN 1	IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			Device name	e1000g4	
		Public LAN 2	IP address	10.34.215.22	
			Subnet mask	255.255.255.0	
			Device name	e1000g5	
	Non-global zone	Zone name (zonename)		NGZA	
		Node name (uname -n)		guest03	
		Public LAN 1	IP address	10.34.214.188	
			Subnet mask	255.255.255.0	
			Device name	sha0	
	Non-global zone	Zone name (zonename)		NGZB	
		Node name (uname -n)		guest04	
		Public LAN 1	IP address	10.34.214.189	
			Subnet mask	255.255.255.0	
			Device name	sha0	
	Non-global zone	Zone name (zonename)			
		Node name (uname -n)			
		Public LAN 1	IP address		
			Subnet mask		
			Device name		

	Use of shared disk		■ yes □ no	
	scsi-initiator-id (OBP variable #eeprom command)		8	
	SCSI / fibre channel	Circuits (paths)	2	
		Device name 1	fjpfca#0	
		sd device name	sd11 - sd42	
		Device name 2	fjpfca#1	
		sd device name	sd43 - sd74	
	Management LAN 1	IP address	10.33.66.93	
		Subnet mask	255.255.255.0	
		Device name	e1000g0	
	Management LAN 2	IP address	10.33.66.94	
		Subnet mask	255.255.255.0	
		Device name	e1000g1	
	Public LAN 1	IP address	10.34.214.182	
		Subnet mask	255.255.255.0	
		Device name	e1000g4	
	Public LAN 2	IP address	10.34.215.22	
		Subnet mask	255.255.255.0	
		Device name	e1000g5	
Non-global zone	Zone name (zonename)		NGZA	
	Node name (uname -n)		guest03	
	Public LAN 1	IP address	10.34.214.188	
		Subnet mask	255.255.255.0	
		Device name	sha0	
Non-global zone	Zone name (zonename)		NGZB	
	Node name (uname -n)		guest04	
	Public LAN 1	IP address	10.34.214.189	
		Subnet mask	255.255.255.0	
		Device name	sha0	
Non-global zone	Zone name (zonename)			
	Node name (uname -n)			
	Public LAN 1	IP address		
		Subnet mask		
		Device name		

16.7.2 GLS Setup Worksheet

This worksheet serves to organize necessary input items beforehand for using GLS (Global Link Services) for Oracle Solaris environment.

Shown below is an example of creating the GLS Setup Worksheet for the redundancy method.

- NIC switching mode (exclusive IP zone)

■NIC Switching Mode (Global zone: Logical IP Takeover , Non-global zone: Logical IP Takeover) Worksheet

Item					Setting
GLS setup	Takeover virtual interface name				sha0
		Global zone	Takeover virtual IP address (or host name)		192.168.10.100
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Global zone (GZ1)	Configuration information	Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.10
				Logical IP address (or host name)	192.168.10.100
			Monitoring target information	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)				primecl02
		Global zone (GZ2)	Configuration information	Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.20
				Logical IP address (or host name)	192.168.10.100
			Monitoring destination	Primary monitoring destination IP address (or host	192.168.10.250
				Secondary monitoring destination IP address (or	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
GLS setup	Takeover virtual interface name				sha0
		Non-global zone	Takeover virtual IP address (or host name)		192.168.10.101
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Non-global zone (NGZ1)	Configuration information	Non-global zone name	guest01
				Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g2
				Secondary interface name	e1000g3
				Physical IP address (or host name)	192.168.10.11
				Logical IP address (or host name)	192.168.10.101
			Monitoring destination	Primary monitoring destination IP address (or host	192.168.10.250
				Secondary monitoring destination IP address (or	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)				primecl02
		Non-global zone (NGZ3)	Configuration information	Non-global zone name	guest01
				Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g2
				Secondary interface name	e1000g3
				Physical IP address (or host name)	192.168.10.21
				Logical IP address (or host name)	192.168.10.101
			Monitoring destination	Primary monitoring destination IP address (or host	192.168.10.250
				Secondary monitoring destination IP address (or	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable

GLS setup	Takeover virtual interface name				sha0
		Non-global zone	Takeover virtual IP address (or host name)		192.168.10.102
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Non-global zone (NGZ2)	Configuration information	Non-global zone name	guest02
				Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary physical interface name	e1000g4
				Secondary physical interface name	e1000g5
				Physical IP address (or host name)	192.168.11.12
				Logical IP address (or host name)	192.168.10.102
			Monitoring destination	Primary monitoring destination IP address (or host	192.168.10.250
				Secondary monitoring destination IP address (or	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)				primecl02
		Non-global zone (NGZ4)	Configuration information	Non-global zone name	guest02
				Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary physical interface name	e1000g4
				Secondary physical interface name	e1000g5
				Physical IP address (or host name)	192.168.11.22
				Logical IP address (or host name)	192.168.10.102
			Monitoring destination	Primary monitoring destination IP address (or host	192.168.10.250
				Secondary monitoring destination IP address (or	192.168.10.251
			Optional function	Backup NIC patrol	◇Disable ◆Enable

- NIC switching mode (shared IP zone)

■NIC Switching Mode (Global zone: Logical IP Takeover , Non-global zone: Logical IP Takeover) Worksheet

Item					Setting
GLS setup	Takeover virtual interface name				sha0
		Global zone	Takeover virtual IP address (or host name)		192.168.10.100
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Global zone	Configuration information	Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.10
				Logical IP address (or host name)	192.168.10.100
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
	Node name (2)				primecl02
		Global zone	Configuration information	Virtual interface name	sha0
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.20
				Logical IP address (or host name)	192.168.10.100
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
GLS setup	Takeover virtual interface name				sha1
		Non-global zone	Takeover virtual IP address (Cmdline resource (control of shared IP)) *2		192.168.10.101
			Subnet mask		255.255.255.0
		Global zone	Takeover virtual IP address (or host name) *3		192.168.10.201
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Global zone	Configuration information	Virtual interface name	sha1
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.10
				Logical IP address (or host name) *3	192.168.10.201
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
		Non-global zone	zonecfg information	Non-global zone name	guest01
				IP address of non-global zone (address) *1	192.168.10.11
				Interface of non-global zone (physical)	e1000g0
			Cmdline resource(control of shared IP) information	Non-global zone name (ZONENAME)	guest01
				Takeover IP address (LIP) *2	192.168.10.101
				List of network interface name (IFLIST)	e1000g0 e1000g1
				Non-global zones wait (ZONEWAIT)	◇Disable ◆Enable
	Node name (2)				primecl02
		Global zone	Configuration information	Virtual interface name	sha1
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.20
				Logical IP address (or host name) *3	192.168.10.201
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
		Non-global zone	zonecfg information	Non-global zone name	guest01
				IP address of non-global zone (address) *1	192.168.10.21
				Interface of non-global zone (physical)	e1000g0
			Cmdline resource(control of shared IP) information	Non-global zone name (ZONENAME)	guest01
				Takeover IP address (LIP) *2	192.168.10.101
				List of network interface name (IFLIST)	e1000g0 e1000g1
				Non-global zones wait (ZONEWAIT)	◇Disable ◆Enable

GLS setup	Takeover virtual interface name				sha2
		Non-global zone	Takeover virtual IP address (Cmdline resource (control of shared IP)) *2		192.168.10.102
			Subnet mask		255.255.255.0
		Global zone	Takeover virtual IP address (or host name) *3		192.168.10.202
			Subnet mask		255.255.255.0
	Node name (1)				primecl01
		Global zone	Configuration information	Virtual interface name	sha2
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.10
				Logical IP address (or host name) *3	192.168.10.202
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
		Non-global zone	zonecfg information	Non-global zone name	guest02
				IP address of non-global zone (address) *1	192.168.10.12
				Interface of non-global zone (physical)	e1000g0
			Cmdline resource(control of shared IP) information	Non-global zone name (ZONENAME)	guest02
				Takeover IP address (LIP) *2	192.168.10.102
				List of network interface name (IFLIST)	e1000g0 e1000g1
				Non-global zones wait (ZONEWAIT)	◇Disable ◆Enable
	Node name (2)				primecl02
		Global zone	Configuration information	Virtual interface name	sha2
				Switching mode	◆NIC switching mode (logical IP takeover) ◇NIC switching mode (physical IP takeover)
				Primary interface name	e1000g0
				Secondary interface name	e1000g1
				Physical IP address (or host name)	192.168.10.20
				Logical IP address (or host name) *3	192.168.10.202
			Monitoring destination	Primary monitoring destination IP address (or host name)	192.168.10.250
				Secondary monitoring destination IP address (or host name)	192.168.10.251
			Optional function	Standby NIC patrol	◇Disable ◆Enable
		Non-global zone	zonecfg information	Non-global zone name	guest02
				IP address of non-global zone (address) *1	192.168.10.22
				Interface of non-global zone (physical)	e1000g0
			Cmdline resource(control of shared IP) information	Non-global zone name (ZONENAME)	guest02
				Takeover IP address (LIP) *2	192.168.10.102
				List of network interface name (IFLIST)	e1000g0 e1000g1
				Non-global zones wait (ZONEWAIT)	◇Disable ◆Enable

*1 It is set up using the zonecfg command in the global zone.
*2 The cmdline resource operated in the global zone controls the takeover IP address used in the non-global zone.
*3 When detecting failures in the network, it is used to inform the failures to the user applications. It is not used for the non-global zone transmission.

The item which is not set for the GLS configuration information. It is set using the zoencfg command and Cmdline resource.

16.7.3 GDS Setup Worksheet

Shown below is an example of creating the GDS Setup Worksheet.

Item					Setting
GDS configuration	Class 1	Class name			class0001
		Class type			local
		Class scope (node name)	Node name 1		GZ1
			Node name 2		
		Non-global zone name	Non-global zone name 1		NGZ1
			Non-global zone name 2		
		Disk 1	SDX disk name		diskmplb0001
			Physical disk name		mplb2048
		Disk 2	SDX disk name		diskmplb0007
			Physical disk name		mplb2054
		Disk 3	SDX disk name		
			Physical disk name		
		Disk 4	SDX disk name		
			Physical disk name		
		Group 1	Group name		group0001
			Group type		mirror
			Disk 1		diskmplb0001
			Disk 2		diskmplb0007
			Volume 1	Volume name	volume0001
				Size	8380799
			Volume 2	Volume name	
				Size	
		Group 2	Group name		
			Group type		
			Disk 1		
			Disk 2		
			Volume 1	Volume name	
				Size	
			Volume 2	Volume name	
				Size	
	Class 2	Class name			class0002
		Class type			local
		Class scope (node name)	Node name 1		GZ1
			Node name 2		
		Non-global zone name	Non-global zone name 1		NGZ2
			Non-global zone name 2		
		Disk 1	SDX disk name		diskmplb0002
			Physical disk name		mplb2049
		Disk 2	SDX disk name		diskmplb0008
			Physical disk name		mplb2055
		Group 1	Group name		group0002
			Group type		mirror
			Disk 1		diskmplb0002
			Disk 2		diskmplb0008
			Volume 1	Volume name	volume0002
				Size	8380799

	Class 3	Class name			class0003
		Class type			local
		Class scope (node name)	Node name 1		GZ2
			Node name 2		
		Non-global zone name	Non-global zone name 1		NGZ3
			Non-global zone name 2		
		Disk 1	SDX disk name		diskmplb0003
			Physical disk name		mplb2050
		Disk 2	SDX disk name		diskmplb0009
			Physical disk name		mplb2056
		Group 1	Group name		group0003
			Group type		mirror
			Disk 1		diskmplb0003
			Disk 2		diskmplb0009
			Volume 1	Volume name	volume0003
				Size	8380799
	Class 4	Class name			class0004
		Class type			local
		Class scope (node name)	Node name 1		GZ2
			Node name 2		
		Non-global zone name	Non-global zone name 1		NGZ4
			Non-global zone name 2		
		Disk 1	SDX disk name		diskmplb0004
			Physical disk name		mplb2051
		Disk 2	SDX disk name		diskmplb0010
			Physical disk name		mplb2057
		Group 1	Group name		group0004
			Group type		mirror
			Disk 1		diskmplb0004
			Disk 2		diskmplb0010
			Volume 1	Volume name	volume0004
				Size	8380799

	Class 5	Class name			class0005
		Class type			shared
		Class scope (node name)	Node name 1		GZ1
			Node name 2		GZ2
		Non-global zone name	Non-global zone name 1		NGZ1
			Non-global zone name 2		NGZ3
		Disk 1	SDX disk name		diskmplb0005
			Physical disk name (node1)		mplb2052
			Physical disk name (node2)		mplb2052
		Disk 2	SDX disk name		diskmplb0011
			Physical disk name (node1)		mplb2058
			Physical disk name (node2)		mplb2058
		Group 1	Group name		group0005
			Group type		mirror
			Disk 1		diskmplb0005
			Disk 2		diskmplb0011
			Volume 1	Volume name	volume0005
				Size	8380799
	Class 6	Class name			class0006
		Class type			shared
		Class scope (node name)	Node name 1		GZ1
			Node name 2		GZ2
		Non-global zone name	Non-global zone name 1		NGZ2
			Non-global zone name 2		NGZ4
		Disk 1	SDX disk name		diskmplb0006
			Physical disk name (node1)		mplb2053
			Physical disk name (node2)		mplb2053
		Disk 2	SDX disk name		diskmplb0012
			Physical disk name (node1)		mplb2059
			Physical disk name (node2)		mplb2059
		Group 1	Group name		group0006
			Group type		mirror
			Disk 1		diskmplb0006
			Disk 2		diskmplb0012
			Volume 1	Volume name	volume0006
				Size	8380799

Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment

This chapter describes design, prerequisites and operations when using the Migration function in an Oracle VM Server for SPARC Environment.

Note

- When configuring Oracle Solaris Kernel Zones environment on a guest domain, the Migration cannot be performed.
- In the cluster configuration among the different physical partitions (only between guest domains), only the Live Migration performed on the guest domain is available.

  For prerequisites and operations on the guest domain, see the following sections:

  - 17.2.3 Performing Live Migration of the Cluster on the Guest Domains
  - 17.3.3 Performing Live Migration of the Cluster on a Guest Domain

See

For prerequisites and operations when performing the Migration on a guest domain with ServerView Resource Orchestrator Cloud Edition, see "17.4 Using with ServerView Resource Orchestrator Cloud Edition."

17.1 Design

Following two types of the Migration function can be used for a cluster system in an Oracle VM Server for SPARC Environment:

- Live Migration
  Transferring an active guest domain.
- Cold Migration
  Transferring an inactive guest domain.

For the cluster configuration in which the Migration function of an Oracle VM Server for SPARC can be used, see "2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment."

When performing the following operations on the control domains, all the guest domains that are the targets of Migration must be configured so that the control domains can access the guest domains by SSH connection via the administrative LAN.

- Preconfiguration to the guest domains
- Live Migration operation
- Cold Migration operation

17.2 Prerequisites

This section describes prerequisites when using the Migration function in an Oracle VM Server for SPARC Environment.

17.2.1 Performing Live Migration of the Cluster on the Control Domains

Note

- Before taking the procedure described below, set the shutdown facility of the control domains and the shutdown facility of the guest domains correctly.

- If the host name is specified for the setting of the shutdown facility of the control domain, the host name specified in the control domain and its corresponding IP address must be described in the /etc/inet/hosts file of the guest domain as well.

- Perform steps 1 through 3 on a control domain.

- Perform step 4 on a guest domain.

- Take this procedure only once. It does not have to be taken for each Migration.

  However, if the configuration is changed after this procedure is taken, take the procedure described in "17.2.2 Changing Configuration after Setting Prerequisites."

1. Login to a guest domain (control domain)

   To access a guest domain from the control domain via SSH, you need to complete the user inquiry of the first SSH connection (RSA key generation).

   Log in as the root user to the IP address of a guest domain that is to be set in step 2 on all control domains.

```
# ssh -l root XXX.XXX.XXX.XXX
The authenticity of host 'XXX.XXX.XXX.XXX (XXX.XXX.XXX.XXX)' can't be established.
RSA key fingerprint is xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx.
Are you sure you want to continue connecting (yes/no)? yes <- Enter yes.
```

2. Registering the guest domain information (control domain)

   Execute the following command on one of the cluster nodes of the control domain to register the information of the target guest domain to which a cluster system is migrated with Migration.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest-name guest-clustername guest-cfname
guest_ip guest-user
```

   guest-name

   Guest domain name

   guest-clustername

   Cluster name of the guest domain

   guest-cfname

   CF node name of the guest domain

   guest-ip

   IP address of the guest domain
   Available IP addresses are IPv4 addresses.

   guest-user

   User name to log in to the guest domain
   Specify the root user as the fixed root name.

   Example: When the environment is built with the two-node cluster between guest domains ("Figure 17.1 Cluster configuration example")

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest1 cluster2 cfguest1 10.20.30.51 root
Enter User's Password:
Re-enter User's Password:
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest2 cluster2 cfguest2 10.20.30.52 root
Enter User's Password:
Re-enter User's Password:
#
```

3. Changing the setting of weight connection between shutdown agents (control domain)

   The default weight connection between shutdown agents is the weight connection between the shutdown agent of a control domain and the shutdown agent of a guest domain.

This step 3 is unnecessary if the weight of the shutdown agent of a control domain should be connected to the weight of the shutdown agent of a guest domain.

If the weight connection is unnecessary between the shutdown agent of a control domain and the shutdown agent of a guest domain, execute the following command on one of the cluster nodes of the control domain.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -w off
```

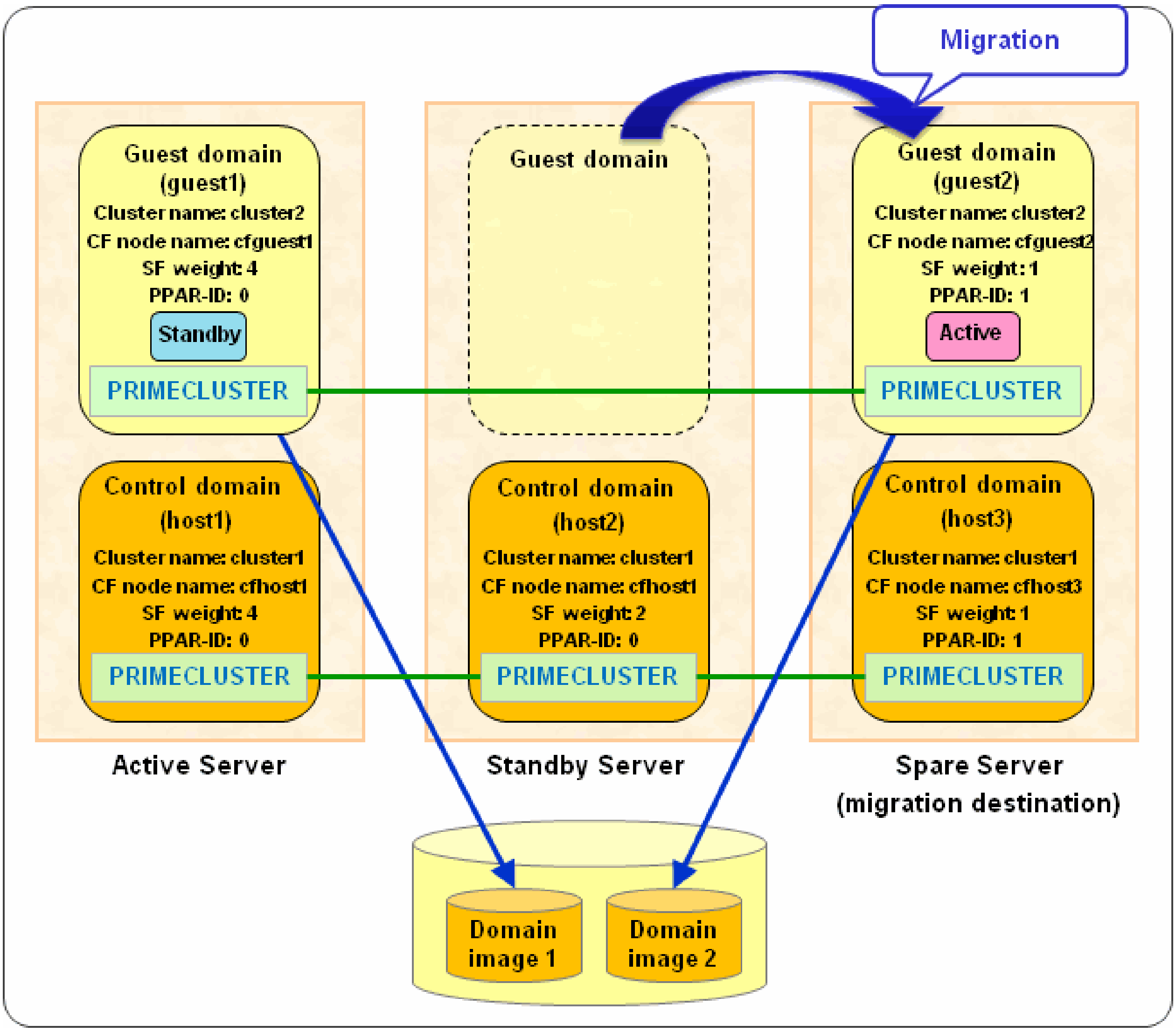
4. Setting /etc/system (guest domain)

   Add the following description to "/etc/system" on the target guest domain to which a cluster system is migrated with Migration. Then, reboot the system.

   This step 4 is unnecessary in the environment where SRU9.5.1 of Solaris 11.1 is applied.

```
forceload: misc/klmmod
```

Figure 17.1 Cluster configuration example



17.2.2 Changing Configuration after Setting Prerequisites

If the configuration is changed after "17.2.1 Performing Live Migration of the Cluster on the Control Domains" has been performed, take the following procedure depending on a configuration change.

Configuration change 1

- IP address of the guest domain
- Password of a user name to log in to the guest domain

Procedure

Execute the following command on any control domain.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -m guest-name -i guest-ip -u guest-user
```

Configuration change 2

- Guest domain name

Procedure

Execute the following commands on any control domain.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -d old-guest-name
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest-name guest-clustername guest-cfname guest_ip
guest-user
```

Configuration change 3

- IP address of the control domain
- Host name registered in the /etc/inet/hosts on the control domain
- SF node weight on the control domain
- PPAR-ID on the control domain
- IP address of XSCF-LAN#0
- IP address of XSCF-LAN#1
- Host name registered in the /etc/inet/hosts of tXSCF-LAN#0
- Host name registered in the /etc/inet/hosts of tXSCF-LAN#1

Procedure

For all the guest domains that are the targets of Migration, execute the following command on any control domain. However, one time execution of this command can change the configuration of only one guest domain.

```
# /etc/opt/FJSVcluster/bin/clovmmigratesetup -a guest-name guest-clustername
guest-cfname guest_ip guest-user
```

17.2.3 Performing Live Migration of the Cluster on the Guest Domains

1. Creating /etc/opt/FJSVcluster/etc/clovmmigrate.conf (guest domain)
   Create the /etc/opt/FJSVcluster/etc/clovmmigrate.conf file on all the nodes, and register the cluster configuration node and the destination server information beforehand.

   Note

   - Keep the clovmmigrate.conf file consistent for all cluster nodes.
   - Take this procedure only once. It does not have to be taken for each Migration.

     However, if the configuration is changed after this procedure is taken, take the procedure described in "17.2.2 Changing Configuration after Setting Prerequisites."

   ```
   primary_IP primary_weight primary_PPAR-ID XSCF-name1 XSCF-name2
                                 :
   ```

   - Create the clovmmigrate.conf file with system administrator authority and set the permission of the file to 600.

- Describe the information of a single node in one line.
- Use a space delimiter between each item.

primary_IP

The host name registered to the IP address or the /etc/inet/hosts file of the control domain

primary_weight

The SF node weight of the control domain

For the cluster configuration among the different physical partitions (only between guest domains), set 1.

primary_PPAR-ID

PPAR-ID of the control domain

Enter the values within 0 to 15.

For SPARC M10-1 environment and SPARC M10-4 environment, specify "0".

For SPARC M10-4S environment, execute the showpparstatus command in XSCF, and check the information of PPAR-ID.

XSCF-name1

The IP address of XSCF-LAN#0 or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

XSCF-name2

The IP address of XSCF-LAN#1 or the host name that is registered in the /etc/inet/hosts file.

Available IP addresses are IPv4 addresses.

For SPARC M10-4S environment, specify the XSCF takeover IP address.

Example: When the environment is built with the two-node cluster between guest domains ("Figure 17.1 Cluster configuration example")

```
host1 4 0 XSCF#1-LAN#0 XSCF#1-LAN#1
host2 2 0 XSCF#2-LAN#0 XSCF#2-LAN#1
host3 1 1 XSCF#3-LAN#0 XSCF#3-LAN#1
```

2. Disabling the link between the node weight of the guest domain and the node weight of the control domain

   Execute the following command on any guest domain.

```
# /etc/opt/FJSVcluster/bin/cldevparam -p WeightMode 0
```

3. Setting /etc/system (guest domain)

   Add the following description to "/etc/system" on the target guest domain to which a cluster system is migrated with Migration. Then, reboot the system.

   This step 3 is unnecessary in the environment where SRU9.5.1 or later of Solaris 11.1 is applied.

```
forceload: misc/klmmod
```

17.3 Operations

This section describes the operation when using the Migration function in an Oracle VM Server for SPARC Environment.

17.3.1 Performing Live Migration of the Cluster on a Control Domain

17.3.1.1 Operation before Performing Live Migration

This section describes the operation before performing the Live Migration in an Oracle VM Server for SPARC Environment.

Note

- Prerequisites in "17.2.1 Performing Live Migration of the Cluster on the Control Domains" are required before performing this operation.
- Make sure that following items are consistent on all control domains of the target clusters: a combination of a user name and password for the XSCF that is registered in the shutdown facility, and the connection method to the XSCF.
- Check that the following settings have been performed for the destination servers:
    - 5.1.2.1.1 Checking XSCF Information
    - 5.1.2.1.2 Setting SNMP
- If you perform this operation, a cluster system will not be switched until "17.3.1.2 Operation after Performing Live Migration" is completed.
- After performing this operation, make sure to perform "17.3.1.2 Operation after Performing Live Migration" even though you cancel the Live Migration.
- After performing this operation, make sure to perform "17.3.1.2 Operation after Performing Live Migration" even though the Live Migration failed.
- Do not perform the Live Migration during a cluster system switchover.

1. Changing the cluster configuration (control domain)

   Change the cluster configuration before performing the Live Migration.

   Execute the following command on the cluster nodes of the source control domain.

   ```
   # /etc/opt/FJSVcluster/bin/clovmmigrate -p source-ldom
   ```

   source-ldom

   The target guest domain name to be migrated

   If you execute this command, the following cluster configurations are changed on all nodes in the cluster between guests that are specified for source-ldom.

   - The timeout value of the CF cluster interconnect (current setting value [default 10 seconds] to 600 seconds)
   - Stopping the shutdown facility

17.3.1.2 Operation after Performing Live Migration

This section describes the operation after performing the Live Migration in an Oracle VM Server for SPARC Environment.

Note

- After performing "17.3.1.1 Operation before Performing Live Migration," perform this operation even though you cancel the Live Migration.
- After performing "17.3.1.1 Operation before Performing Live Migration," perform this operation even though the Live Migration failed.
- A cluster system will not be switched until this operation is completed.
- Perform step 1 on a control domain.

- Perform step 2 on a guest domain.
- Before performing this operation, the configuration information of the saved logical domain needs to be less than six or less. Execute the following command to check it. Also, check that "config_tmp" does not exist in the saved configuration information name of the logical domains.

```
# ldm list-spconfig
```

If the configuration information of the saved logical domain is 7 or more, execute the following command to delete the configuration information to make them 6 or less.

```
# ldm remove-spconfig configuration name
```

If "config_tmp" exists in the configuration name of the logical domain, execute the following command to delete it.

```
# ldm remove-spconfig config_tmp
```

For details on the configuration information of the logical domain, see "Operations and Commands Related to Logical Domain Configurations" in "SPARC M10 Systems Domain Configuration Guide."

1. Changing the cluster configuration (control domain)

   Change the cluster configuration after performing the Live Migration.

   Execute the following command on the cluster nodes of the source control domain:

   ```
   # /etc/opt/FJSVcluster/bin/clovmmigrate -u source-ldom target-host
   ```

   source-ldom

   The guest domain name to be migrated

   target-host

   The host name registered to the IP address or the /etc/inet/hosts file of the destination control domain

   Even if a guest domain is not migrated by such as cancelling the Migration, you need to specify a control domain. To do so, specify the host name registered to the IP address on the source control domain or the /etc/inet/hosts file.

   If you execute this command, the following cluster configurations are changed on all nodes in the cluster between guests that are specified for source-ldom. In addition, the configuration information of the logical domains is saved both by the source control domain and the destination control domain.

   - The timeout value of the CF cluster interconnect (600 seconds to the setting value before Migration [default 10 seconds])
   - Change the setting of the shutdown facility (IP address of XSCF-LAN#0, IP address of XSCF-LAN#1, and the SF weight)
   - Starting the shutdown facility

2. Checking the state of the shutdown facility (guest domain)

   Execute the following commands on each node of a guest domain to check if the cluster is configured correctly after the Live Migration.

   ```
   # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
   # /opt/SMAW/bin/sdtool -s
   ```

   Note

   If TestFailed or InitFailed is displayed, the setting of the shutdown facility may not be changed.
   Go back to step 1 and try again.

   Example: When the Migration is performed for guest 2 in the two-node cluster between guest domains ("Figure 17.1 Cluster configuration example")

```
guest2 # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
device-name cluster-host-name  PPAR-ID  domain-name  IP-address1  IP-address2  user-name
connection-type
-----------------------------------------------------------------------------------------------
--------------------
xscf        cfguest1            0        guest1       10.20.30.71  10.20.40.71  xuser   ssh
xscf        cfguest2            1        guest2       10.20.30.73  10.20.40.73  xuser   ssh
guest2 #                        ^^                    ^^^^^^^^^^^^^^^^^^^^^^^^
guest2 #                                              The target XSCF IP address to be migrated
guest2 #                       The target PPAR-ID to be migrated
guest2 # /opt/SMAW/bin/sdtool -s
Cluster Host      Agent               SA State      Shut State  Test State  Init State
------------      -----               --------      ----------  ----------  ----------
cfguest1          SA_xscfsnmpg0p.so   Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1p.so   Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg0r.so   Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1r.so   Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0p.so   Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg1p.so   Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0r.so   Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg1r.so   Idle          Unknown     TestWorked  InitWorked
guest2 #
```

Note

After performing the Migration, the following message which indicates the time is not synchronized between the cluster nodes may be printed in the switchlog or /var/adm/messages file.

```
(WRP, 34) Cluster host <host> is no longer in time sync with local node. Sane operation of RMS can no
longer be guaranteed.
Further out-of-sync messages will appear in the syslog.
```

If this situation continues, the following message may be periodically printed in the /var/adm/messages file.

```
(WRP, 35) Cluster host <host> is no longer in time sync with local node. Sane operation of RMS can no
longer be guaranteed.
```

This message stops once the time is synchronized. For details on the messages, see "*PRIMECLUSTER Messages*."

17.3.2 Performing Cold Migration of the Cluster on a Control Domain

17.3.2.1 Operation before Performing Cold Migration

This section describes the operation before performing the Cold Migration in an Oracle VM Server for SPARC Environment.

Note

- Prerequisites in "17.2.1 Performing Live Migration of the Cluster on the Control Domains" are required before performing this operation.
- Make sure that following items are consistent on all control domains of the target clusters: a combination of a user name and password for the XSCF that is registered in the shutdown facility, and the connection method to the XSCF.
- Check that the following settings have been performed for the destination servers:
    - 5.1.2.1.1 Checking XSCF Information
    - 5.1.2.1.2 Setting SNMP
- Before performing this operation, make sure that the target guest domain to be migrated is stopped.
  Use the ldm list-domain command to check if STATE is displayed as the bound state.

If the state of STATE is inactive, change its state to bound. The configuration information of the logical domains can be identified by XSCF when the STATE is in the bound state.

- If you cancel the Cold Migration after performing this operation, "17.3.2.2 Operation after Performing Cold Migration" is unnecessary.

1. Changing the cluster configuration (control domain)

   Change the cluster configuration before performing the Cold Migration.

   Execute the following command on the cluster nodes of the source control domain.

   When the guest domain is stopped, check that the command exists with the return value of 1.

```
# /etc/opt/FJSVcluster/bin/clovmmigrate -p source-ldom
```

   source-ldom

      The guest domain name to be migrated.

   By executing this command, you can check if the guest domain specified for source-ldom is stopped.

17.3.2.2 Operation after Performing Cold Migration

This section describes the operation after performing the Cold Migration in an Oracle VM Server for SPARC Environment.

Note

- These steps are unnecessary if you cancel the Cold Migration after performing "17.3.2.1 Operation before Performing Cold Migration."
- Perform steps 1 through 3 on a control domain.
- Perform step 4 on a guest domain.
- Before performing this operation, the configuration information of the saved logical domain needs to be less than six or less. Execute the following command to check it. Also, check that "config_tmp" does not exist in the saved configuration information name of the logical domains.

```
# ldm list-spconfig
```

  If the configuration information of the saved logical domain is 7 or more, execute the following command to delete the configuration information to make them 6 or less.

```
# ldm remove-spconfig configuration name
```

  If "config_tmp" exists in the configuration name of the logical domain, execute the following command to delete it.

```
# ldm remove-spconfig config_tmp
```

  For details on the configuration information of the logical domain, see "Operations and Commands Related to Logical Domain Configurations" in "SPARC M10 Systems Domain Configuration Guide."

1. Changing the cluster configuration (control domain)

   Change the cluster configuration after performing the Cold Migration.

   Execute the following command on the cluster nodes of the source control domain.

   When the guest domain is stopped, the return value of the command is 1.

```
# /etc/opt/FJSVcluster/bin/clovmmigrate -u source-ldom target-host
```

   source-ldom

      The target guest domain name to be migrated target-host

The host name registered to the IP address or the /etc/inet/hosts file of the destination control domain

If you execute this command, information of the logical domains is saved both by the source control domain and the destination control domain.

2. Checking the state of the guest domain and saving the configuration information (control domain)

   Execute the following command on the destination control domain to check the state of the guest domain that was migrated by the Cold Migration.

```
# ldm list-domain
```

   If the migrated guest domain is in the bound state, proceed to the following steps.
   If the migrated guest domain is in the inactive state, change its state to bound. After changing the state of the guest domain to bound, save the configuration information of the logical domains on the source control domain and the destination control domain. The configuration information of the logical domains can be identified by XSCF when the guest domain is in the bound state. For details, see "SPARC M10 Systems System Operation and Administration Guide."

3. Starting the guest domain (control domain)

   Start the stopped guest domain that was migrated by the Cold Migration.
   When the guest domain is started, the following cluster configuration is changed on all nodes of the cluster between guests.

   - Change the setting of the shutdown facility (IP address of XSCF-LAN#0, IP address of XSCF-LAN#1, and the SF weight)

4. Checking the state of the shutdown facility (guest domain)

   Execute the following commands on each node of a guest domain to check if the cluster is configured correctly after the Cold Migration.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -l
# /opt/SMAW/bin/sdtool -s
```

Note

If TestFailed or InitFailed is displayed, the setting of the shutdown facility could have been unchanged.
Go back to step 1 and try again.

Example: When the Migration is performed for guest 2 in the two-node cluster between guest domains ("Figure 17.1 Cluster configuration example")

```
guest2 # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
device-name cluster-host-name  PPAR-ID  domain-name  IP-address1  IP-address2  user-name
connection-type
-------------------------------------------------------------------------------------------
--------------------
xscf        cfguest1           0        guest1       10.20.30.71  10.20.40.71  xuser   ssh
xscf        cfguest2           1        guest2       10.20.30.73  10.20.40.73  xuser   ssh
guest2 #                       ^^                    ^^^^^^^^^^^^^^^^^^^^^^^^
guest2 #                                             The target XSCF IP address to be migrated
guest2 #                     The target PPAR-ID to be migrated
guest2 # /opt/SMAW/bin/sdtool -s
Cluster Host      Agent                SA State      Shut State  Test State  Init State
------------      -----                --------      ----------  ----------  ----------
cfguest1          SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
```

```
cfguest2          SA_xscfsnmpg1r.so    Idle          Unknown      TestWorked  InitWorked
guest2 #
```

17.3.3 Performing Live Migration of the Cluster on a Guest Domain

17.3.3.1 Operation before Performing Live Migration

This section describes the operation before performing the Live Migration in an Oracle VM Server for SPARC Environment.

Note

- Create a /etc/opt/FJSVcluster/etc/clovmmigrate.conf file on all the nodes beforehand. For details, see "17.2.3 Performing Live Migration of the Cluster on the Guest Domains."
- Make sure that following items are consistent on all control domains of the target clusters: a combination of a user name and password for the XSCF that is registered in the shutdown facility, and the connection method to the XSCF.
- Check that the following settings have been performed for the destination servers:
  - 5.1.2.1.1 Checking XSCF Information
  - 5.1.2.1.2 Setting SNMP
- Before performing this operation, make sure that all target cluster nodes between the guest domains work normally.
- If you perform this operation, a cluster system will not be switched until "17.3.3.2 Operation after Performing Live Migration" is completed.
- After performing this operation, make sure to perform "17.3.3.2 Operation after Performing Live Migration" even though you cancel the Live Migration.
- After performing this operation, make sure to perform "17.3.3.2 Operation after Performing Live Migration" even though the Live Migration failed.
- Do not perform the Live Migration during a cluster system switchover.

1. Changing the cluster configuration (guest domain)

   Change the cluster configuration before performing the Live Migration.
   Execute the following command on one of the cluster nodes of the guest domain.

   ```
   # /etc/opt/FJSVcluster/bin/clovmmigrate -p
   ```

   If you execute this command, the following cluster configurations are changed on all the nodes:

   - The timeout value of the CF cluster interconnect (current setting value [default 10 seconds] to 600 seconds)
   - Stopping the shutdown facility

17.3.3.2 Operation after Performing Live Migration

This section describes the operation after performing the Live Migration.

Note

- After performing "17.3.3.1 Operation before Performing Live Migration," perform this operation even though you cancel the Live Migration.
- After performing "17.3.3.1 Operation before Performing Live Migration," perform this operation even though the Live Migration failed.
- A cluster system will not be switched until this operation is completed.
- Perform steps 1 and 2 on a control domain.

- Perform steps 3 and 4 on a guest domain.

1. Saving the logical domains configuration information on the source control domain (control domain)
   On the source control domain, save the logical domains configuration information.
   This operation must be done on the source control domain.

   For details, see "SPARC M10 Systems System Operation and Administration Guide."

2. Saving the logical domains configuration information on the destination control domain (control domain)
   On the destination control domain, save the logical domains configuration information.
   This operation must be done on the source control domain.

   For details, see "SPARC M10 Systems System Operation and Administration Guide."

3. Changing the cluster configuration (guest domain)
   Change the cluster configuration after performing the Live Migration.
   Execute the following command on one of the cluster nodes.

```
# /etc/opt/FJSVcluster/bin/clovmmigrate -u source-ldom target-host
```

   source-ldom

   The target guest domain name to be migrated

   target-host

   The IP address of the destination control domain or the host name registered in the /etc/inet/hosts file
   Even if a guest domain is not migrated by such as cancelling the Migration, you need to specify a control domain. To do so, specify the IP address of the source control domain or the host name registered in the /etc/inet/hosts file.

   If you execute this command, the following cluster configurations are changed on all the nodes:

   - The timeout value of the CF cluster interconnect (600 seconds to the setting value before Migration [default 10 seconds])
   - Change the setting of the shutdown facility (IP address of XSCF-LAN#0, IP address of XSCF-LAN#1, and the SF weight)
   - Starting the shutdown facility

4. Checking the state of the shutdown facility (guest domain)
   Execute the following commands on each node of a guest domain to check if the cluster is configured correctly after the Live Migration.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -l
# /opt/SMAW/bin/sdtool -s
```

   Note

   If TestFailed or InitFailed is displayed, the setting of the shutdown facility could have been unchanged.
   Go back to step 3 and try again.

   Example: When the Migration is performed for guest 2 in the two-node cluster between guest domains ("Figure 17.1 Cluster configuration example")

```
guest2 # /etc/opt/FJSVcluster/bin/clsnmpsetup -l
device-name cluster-host-name  PPAR-ID  domain-name  IP-address1  IP-address2  user-name
connection-type
-----------------------------------------------------------------------------------------------
--------------------
xscf        cfguest1            0        guest1       10.20.30.71  10.20.40.71  xuser   ssh
xscf        cfguest2            1        guest2       10.20.30.73  10.20.40.73  xuser   ssh
guest2 #                        ^^                    ^^^^^^^^^^^^^^^^^^^^^^^^
guest2 #                                              The target XSCF IP address to be migrated
guest2 #                       The target PPAR-ID to be migrated
guest2 # /opt/SMAW/bin/sdtool -s
```

```
Cluster Host      Agent                SA State      Shut State  Test State  Init State
------------      -----                --------      ----------  ----------  ----------
cfguest1          SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
cfguest1          SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0p.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg1p.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg0r.so    Idle          Unknown     TestWorked  InitWorked
cfguest2          SA_xscfsnmpg1r.so    Idle          Unknown     TestWorked  InitWorked
guest2 #
```

Note

After performing the Live Migration, the following message which indicates the time is not synchronized between the cluster nodes may be printed in the switchlog or /var/adm/messages file.

```
(WRP, 34) Cluster host <host> is no longer in time sync with local node. Sane operation of RMS can no
longer be guaranteed.
Further out-of-sync messages will appear in the syslog.
```

If this situation continues, the following message may be periodically printed in the /var/adm/messages file.

```
(WRP, 35) Cluster host <host> is no longer in time sync with local node. Sane operation of RMS can no
longer be guaranteed.
```

This message stops once the time is synchronized. For details on the messages, see "PRIMECLUSTER Messages."

17.4 Using with ServerView Resource Orchestrator Cloud Edition

This section describes prerequisites and operations when using the Migration function in combination with ServerView Resource Orchestrator Cloud Edition in an Oracle VM Server for SPARC Environment.

Note

This function can be used only when PRIMECLUSTER is built also on the control domain in the cluster configuration between guest domains among different physical partitions.

All the guest domains that are the targets of Migration must be accessed from the control domain by SSH connection via the administrative LAN.

17.4.1 Prerequisites

For prerequisites, see "17.2.1 Performing Live Migration of the Cluster on the Control Domains".

17.4.2 Operations

17.4.2.1 Performing Live Migration

Perform the Live Migration from ServerView Resource Orchestrator Cloud Edition.

If this operation is performed, "17.3.1.1 Operation before Performing Live Migration" and "17.3.1.2 Operation after Performing Live Migration" are unnecessary.

Note

- Prerequisites in "17.2.1 Performing Live Migration of the Cluster on the Control Domains" are required before performing this operation.

- Make sure that following items are consistent on all control domains of the target clusters: a combination of a user name and password for the XSCF that is registered in the shutdown facility, and the connection method to the XSCF.
- Check that the following settings have been performed for the destination servers:
  - 5.1.2.1.1 Checking XSCF Information
  - 5.1.2.1.2 Setting SNMP
- If you perform this operation, a cluster system will not be switched until Live Migration is completed.
- Do not perform the Live Migration during a cluster system switchover.
- When performing the Live Migration, the saved configuration information of the logical domains should not be above 6. Execute the following command to check the saved configuration information of the logical domains.

  Check that "config_tmp" does not exist in the saved configuration information name of the logical domains as well.

  ```
  # ldm list-spconfig
  ```

  When the saved configuration information of the logical domains is 7 or more, use the following command to delete the configuration information of the logical domains to 6 or less.

  ```
  # ldm remove-spconfig configuration name
  ```

  If "config_tmp" exists in the configuration information name of the logical domains, execute the following command to delete the configuration information name "config_tmp."

  ```
  # ldm remove-spconfig config_tmp
  ```

  For details on the configuration information of the logical domains, see "Operations and Commands Related to Logical Domain Configurations" in "SPARC M10 Systems Domain Configuration Guide."

17.4.2.2 Performing Cold Migration

Perform the Cold Migration from ServerView Resource Orchestrator Cloud Edition to start the stopped guest domain that was migrated by the Cold Migration.

If this operation is performed, "17.3.2.1 Operation before Performing Cold Migration" and "17.3.2.2 Operation after Performing Cold Migration" are unnecessary.

Note

- Prerequisites in "17.2.1 Performing Live Migration of the Cluster on the Control Domains" are required before performing this operation.
- Make sure that following items are consistent on all control domains of the target clusters: a combination of a user name and password for the XSCF that is registered in the shutdown facility, and the connection method to the XSCF.
- Check that the following settings have been performed for the destination servers:
  - 5.1.2.1.1 Checking XSCF Information
  - 5.1.2.1.2 Setting SNMP
- When performing the Cold Migration, the saved configuration information of the logical domains should not be above 6. Execute the following command to check the saved configuration information of the logical domains.

  Check that "config_tmp" does not exist in the saved configuration information name of the logical domains as well.

  ```
  # ldm list-spconfig
  ```

  When the saved configuration information of the logical domains is 7 or more, use the following command to delete the configuration information of the logical domains to 6 or less.

  ```
  # ldm remove-spconfig configuration name
  ```

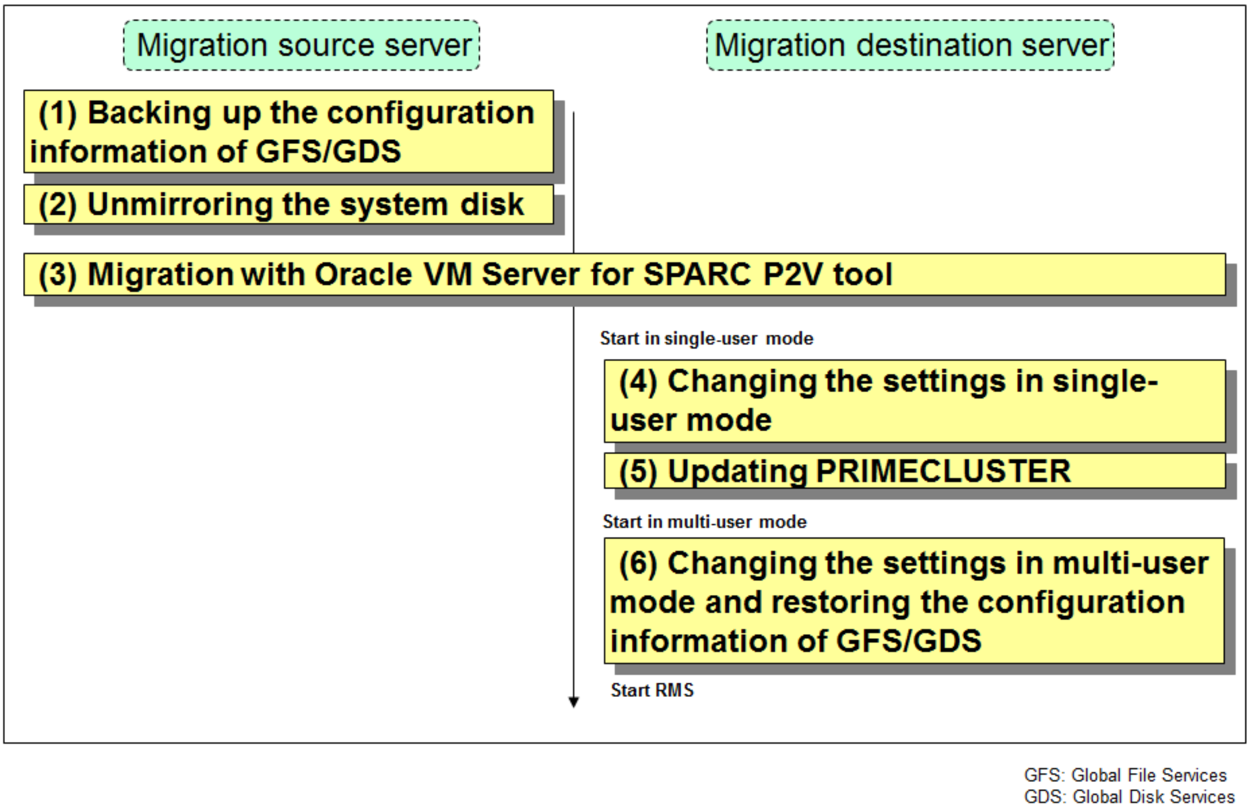
If "config_tmp" exists in the configuration information name of the logical domains, execute the following command to delete the configuration information name "config_tmp."

```
# ldm remove-spconfig config_tmp
```

For details on the configuration information of the logical domains, see "Operations and Commands Related to Logical Domain Configurations" in "SPARC M10 Systems Domain Configuration Guide."

Chapter 18 When Using Oracle VM Server for SPARC P2V Tool to Migrating a Cluster System

This chapter describes the procedure to migrate (P2V) a cluster system in the physical environment that uses PRIMECLUSTER4.2A00 or later to a guest domain (or I/O root domain) in an Oracle VM Server for SPARC Environment.



Note

- The size of the disk to be registered to GDS shared class and GDS local class must be the same in the migration source and migration destination.
- You must migrate user data with ETERNUS storage migration or LUN to LUN such as REC beforehand.
- Use GDS and GLS functions in the cluster system on a guest domain after migration.

Use the same configuration as the migration source after the migration.

See

If a setting has to be changed after completing the cluster system migration, see the following to change the setting:

- "Part 4 System Configuration Modification"
- "*PRIMECLUSTER Global Link Services Configuration and Administration Guide 4.3: Redundant Line Control Function*"
- *"PRIMECLUSTER Global Disk Services Configuration and Administration Guide 4.3"*

You do not need to perform the procedure of unused functions in the cluster systems being used.

In each execution example in the subsequent procedures, a cluster system is constructed in the following configuration.

- Host name : host1, host2
- CF node name: node1, node2
- The root class of GDS: system
- GDS shared classes: class01, class02
- GFS volumes: volume01, volume02, volume03
- The cluster application name: userApp_0
- GLS redundant line control method: NIC switching method

18.1 Backing Up the Configuration Information of GFS/GDS

This section describes the procedure that is performed on the migration source server before migrating in P2V.

1. Back up the management partition information of GFS on the migration source server.

   Execute the following command on any one of the running nodes.

   ```
   # sfcgetconf _backup_file_
   ```

   In the example above, the sfcgetconf(1M) command creates a shell script named "backup_file" to the current directory.

   The content of the shell script is as follows.

   ```
   #!/bin/sh
   # This file is made by:
   #    sfcgetconf _backup_file_
   # Sat Feb 18 09:08:06 2012

   #---- fsid : 1 ----
   # MDS primary   (port) : host1 (sfcfs-1)
   # MDS secondory (port) : host2 (sfcfs-1)
   # MDS other            :
   # AC      : host1,host2
   # options :
   # device  : /dev/sfdsk/class02/rdsk/volume01
   sfcadm -m host1,host2 -g host1,host2 -p sfcfs-1,sfcfs-1 /dev/sfdsk/class02/rdsk/volume01
   #---- fsid : 2 ----
   # MDS primary   (port) : host2 (sfcfs-2)
   # MDS secondory (port) : host1 (sfcfs-2)
   # MDS other            :
   # AC      : host2,host1
   # options :
   # device  : /dev/sfdsk/class02/rdsk/volume02,/dev/sfdsk/class02/rdsk/volume03
   sfcadm -m host2,host1 -g host2,host1 -p sfcfs-2,sfcfs-2 /dev/sfdsk/class02/rdsk/volume02,/dev/
   sfdsk/class02/rdsk/volume03
   # end of backup (_backup_file_)
   ```

2. Backing up the configuration information of GDS on the migration source server.

   Save the results output by the sdxinfo command to a file.

   /var/tmp/Class1.info is used as the path name of the file here.

   Execute the following command on any one of the running nodes. If there are multiple classes, execute the command to all the classes.

   ```
   # sdxinfo -c class01 -e long > /var/tmp/Class1.info
   ```

   Output the configuration of objects within class01 to a file in configuration table format.

   ```
   # sdxconfig Backup -c class01 -o /var/tmp/Class1.conf
   ```

   Confirm that the configuration information is output to the file in the destination.

Note

When using local classes, execute this procedure on the nodes in which local classes are present.

18.2 Unmirroring the System Disk

In a cluster system before migration, if the system disk mirroring using GDS is set, you need to unmirror the system disk before performing the migration.

See

For the procedure to unmirror the system disk, see "*PRIMECLUSTER Global File Services Configuration and Administration Guide 4.3.*"

18.3 Migration with Oracle VM Server for SPARC P2V Tool

Create and expand a file system image by using the Oracle VM Server for SPARC P2V tool.

See

For details on the Oracle VM Server for SPARC P2V tool, see "Oracle VM Server for SPARC Administration Guide" to execute the tool.

Note

- Notes on creating a file system image

  When creating a file system image needed for the migration, stop RMS in the cluster systems on both nodes and then create it.
  For details on how to stop RMS, see "7.2.1.2 Stopping RMS."

- When the IP address that is set in the migration source server and the IP address that is set in the control domain of the migration destination are duplicated, the ldmp2v convert command fails to be executed during the conversion phase. In this case, use the -x skip-ping-test option with the ldmp2v convert command as follows so that checking IP address can be skipped.

```
# ldmp2v convert -i <path to OS image> -d <directory where the files required for P2V are located>
-x skip-ping-test <domain>
```

  <Execution example>

```
# ldmp2v convert -i /var/tmp/ISO/sol_10_113.iso -d /var/tmp/ovm/volumia -x skip-ping-test ldom1
```

  Change the duplicated IP address by taking the procedure of "18.6 Changing the Setting in the Multi-User Mode and Restoration of the Configuration Information of GFS/GDS" after the migration by using P2V tool. When the duplicated IP address is not used for the cluster system, change the IP address when changing the IP address that is used for the cluster system.

18.4 Changing the Setting in the Single-User Mode

To build a cluster system, change the setting in the guest domain after migrating by P2V. Use the following procedure to change the setting.

Although it is not mentioned in the procedure, in general, editor commands such as vi(1) are used for editing files.

1. Start the guest domain in the single-user mode.

```
{0} ok boot -s
```

2. Remove the shared class and the local class forcibly.

   Delete the class definition of both shared and local classes. Execute the following commands on all nodes in the class scope. If there are multiple classes, execute the commands for all the shared and local classes.

```
# cd /etc/opt/FJSVsdx/sysdb.d
# rm class01
# rm class02
```

   Delete the following lines from the class.db file.

```
...
class01     <- Delete this line
class02     <- Delete this line
...
```

   Delete the devname.db file.

```
# cd /etc/opt/FJSVsdx/dev
# rm -f devname.db
```

3. Start inhibition of RMS and SF.

   1. Check the current configuration name of RMS.

```
# /etc/opt/FJSVwvucw/bin/wgcnfclient -g config-name config
```

      The output character string in the above is the currently-used RMS configuration name. If no character string is output, its RMS configuration name is "config".

   2. Start inhibition and delete status of RMS.

      Execute the following command on all nodes.

```
# hvsetenv HV_RCSTART 0
# hvsetenv HV_RCSTART
0    <- Make sure "0" is output
# rm /opt/SMAW/SMAWRrms/.status_info
```

   3. Start inhibition of SF.

      Execute the following command on all nodes.

```
# rm /etc/opt/SMAW/SMAWsf/rcsd.cfg
# ls /etc/opt/SMAW/SMAWsf/rcsd.cfg
/etc/opt/SMAW/SMAWsf/rcsd.cfg: No such file or directory
#
```

4. Set the NIC on the guest domain.

   Change the NIC configuration file name due to the NIC name changes.

   Create the network setting file (/etc/hostname.interface file or /etc/hostname6.interface file) of OS for the primary interface among NICs used in GLS to set the IP address or host name.

   For details, see "3.2.2 Network Configuration" in "*PRIMECLUSTER Global Link Services Configuration and Administration Guide 4.3: Redundant Line Control Function*."

   Delete network configuration files for NICs before migration if these files are left.

18.5 Updating PRIMECLUSTER

If the PRIMECLUSTER version of the migration source is not 4.3A20 or later, upgrade PRIMECLUSTER.

The configuration name of RMS used in here is the RMS configuration name checked in Step 3.1 in "18.4 Changing the Setting in the Single-User Mode."

The following is the update procedure for PRIMECLUSTER Enterprise Edition (EE).

1. Migrate to the OpenBoot environment.

```
# /usr/sbin/shutdown -y -g0 -i0
```

2. Execute the following procedure on all nodes.

    1. Insert the DVD of PRIMECLUSTER 4.3A40 into the DVD drive.

    2. Start the system in the single-user mode.

    ```
    {0} ok boot -s
    ```

    3. Mount the required file system.

    ```
    # /usr/sbin/zfs mount -a
    ```

    If you install it to a directory of the UFS file system, execute the following command.

    ```
    # /usr/sbin/mountall -l
    ```

    4. Mount the DVD.

    ```
    # mkdir -p /cdrom/cdrom0 ***(Note 1)
    # /usr/sbin/mount -F hsfs -o ro /dev/dsk/c0t4d0s2 /cdrom/cdrom0 ***(Note 2)
    ```

    (Note 1) It is required if there is no "/cdrom/cdrom0" directory.

    (Note 2) Check the underlined device name because the name varies depending on the machine environment.

    5. Create a backup directory. "/var/tmp/backup" is used as the backup directory in here.

    ```
    # cd /var/tmp
    # mkdir backup
    ```

    6. Back up the operation environment of PRIMECLUSTER.

    ```
    # cp -p /usr/opt/reliant/etc/hvipalias /var/tmp/backup
    # cp -p /var/opt/FJSVclapm/etc/Tuning_Param /var/tmp/backup
    # cp -p /opt/FJSVcldbm/data/DB/ApiAdjustTbl.CLDB /var/tmp/backup
    # cp -p /etc/opt/FJSVcluster/etc/rmcip.conf /var/tmp/backup
    # /opt/FJSVwvbs/etc/bin/wvSetport fjwv_c > /var/tmp/backup/fjwv_c
    ```

    Check that the backup file has been created.

    ```
    # ls -l /var/tmp/backup
    ```

    7. Back up the operation environment of GLS.

    Back up the operation environment of GLS Redundant Line Control Function.

    Use "hanetYYYYMMDD.bk" as the save file name. YYYYMMDD is the information obtained when the command is executed (YYYY, MM, and DD stand for the year, month and day, respectively).

    ```
    # /opt/FJSVhanet/usr/sbin/hanetbackup -d /var/tmp/backup
    ```

    Back up the operation environment of GLS multipath function.

    ```
    # cd /etc/opt/FJSVmpnet
    # tar cvf - conf | compress > /var/tmp/backup/mpnetfile.tar.Z
    ```

    Check that the backup file has been created.

    ```
    # ls -l /var/tmp/backup
    ```

8. Execute the following script to remove packages.

```
# cd /cdrom/cdrom0/Tool
# ./upgrade_uninstall
Are you sure to remove PRIMECLUSTER software packages from the global zone?(y,n) y
*
*
Uninstallation was successful.
#
```

Check that "Uninstallation was successful." is displayed.

9. For the migration from 4.2A00, check that the following files exist. If so, delete the files with the following procedure.

```
# cd /etc/init.d
# rm fjsvgfs
# rm sfcfsrm
# rm sfcinit
# cd /etc/rc0.d
# rm K00FJSVwvbs
# rm K00FJSVwvcnf
# rm K41sfcfsrm
# rm K62clrmgr
# rm K63clctrl
# rm K68CL20dbm
# rm K71clapi
# rm K72clrms
# cd /etc/rc1.d
# rm K41sfcfsrm
# rm K62clrmgr
# rm K63clctrl
# rm K69CL20dbm
# rm K71clapi
# rm K72clrms
# cd /etc/rc2.d
# rm S07clapi
# rm S08clrms
# rm S65CL10dbm
# rm S66clrmgr
# rm S68clrwz
# rm S71clctrl
# rm S80FJSVwvbs
# rm S81sfcfsrm
# cd /etc/rc3.d
# rm S32clautoconfig
# rm S98clwaitprobe
# rm S99FJSVwvcnf
# rm S99FJSVgfs
# cd /etc/rcS.d
# rm K41sfcfsrm
# rm S76sfcinit
# cd /opt/SMAW/SMAWcf/dep/start.d
# rm S81sfcfs
# cd /opt/SMAW/SMAWcf/dep/stop.d
# rm K41sfcfs
```

10. Stop the system.

```
# /usr/sbin/shutdown -y -g0 -i0
```

Note

Though the following message appears during the shutdown, it does not disrupt ongoing operation.

```
ld.so.1: sfclog: Critical error: libcf.so: Failed to open: No file or directory.
ld.so.1: /etc/opt/FJSVcluster/FJSVcldbm/system/CCMstop: Critical error:
libcf.so: Failed to open: No file or directory.
```

11. Start the system in the single-user mode.

```
{0} ok boot -s
```

12. Mount the required file system.

```
# /usr/sbin/zfs mount -a
```

If you install it to a directory of the UFS file system, execute the following command.

```
# /usr/sbin/mountall -l
```

13. Mount the DVD.

```
# mkdir -p /cdrom/cdrom0  ***(Note 1)
# /usr/sbin/mount -F hsfs -o ro /dev/dsk/c0t4d0s2 /cdrom/cdrom0  ***(Note 2)
```

(Note 1) It is required if there is no "/cdrom/cdrom0" directory.

(Note 2) Check the underlined device name because the name varies depending on the machine environment.

14. Check that FJSVsnap packages have been installed.

Check that the versions of FJSVsnap packages installed are not older than "2.8."

```
# pkginfo -l FJSVsnap.\*
```

If the versions of FJSVsnap packages are old, remove all of them from the system.

```
# pkgrm FJSVsnap
```

If old versions of FJSVsnap packages were removed in the task above, or FJSVsnap packages have not been installed, execute the pkgadd(1M) command to install them.

```
# pkgadd -d /cdrom/cdrom0/Tool/Pkgs FJSVsnap
```

15. Check that FJSViomp 2.3 or later have been installed.

```
# pkginfo -l FJSViomp.\*
```

If the versions of FJSViomp are 2.2 or earlier in the task above, remove all of them from the system.

```
# pkgrm FJSViomp
```

If old versions of FJSViomp packages were removed in the task above, or FJSViomp packages have not been installed, execute the pkgadd(1M) command to install them.

```
# pkgadd -d /cdrom/cdrom0/Tool/Pkgs FJSViomp
```

16. Execute the installation script.

```
# cd /cdrom/cdrom0/Tool
# ./cluster_install -e EE
*
*
```

```
The installation was finished successfully.
#
```

Check that "The installation was finished successfully." is displayed.

17. Unmount the DVD and eject it.

```
# cd /
# /usr/sbin/umount /cdrom/cdrom0
```

18. Restore the operation environment of GLS.

Restore the operation environment used by GLS Redundant Line Control Function.

```
# /opt/FJSVhanet/usr/sbin/hanetrestore -f /var/tmp/backup/<save_file_name>
```

Use "hanetYYYYMMDD.bk" as the save file name.

19. Restore the operation environment of PRIMECLUSTER.

```
# /opt/FJSVwvbs/etc/bin/wvSetport fjwv_c=`cat /var/tmp/backup/fjwv_c`
# cp -p /var/tmp/backup/Tuning_Param /var/opt/FJSVclapm/etc
# cp -p /var/tmp/backup/hvipalias /usr/opt/reliant/etc
# cp -p /var/tmp/backup/ApiAdjustTbl.CLDB /opt/FJSVcldbm/data/DB
# cp -p /var/tmp/backup/rmcip.conf /etc/opt/FJSVcluster/etc/rmcip.conf
```

20. Upadate the autoconf.conf file.

```
# cd /etc/opt/FJSVcluster/etc/
# mv autoconf.conf autoconf.conf.bak
# cp /opt/FJSVclapi/etc/autoconf.conf ./
# chmod 644 autoconf.conf
```

18.6 Changing the Setting in the Multi-User Mode and Restoration of the Configuration Information of GFS/GDS

1. Start the guest domain in the multi-user mode.

```
# shutdown -y -g0 -i6
```

As the NIC, which is set to GLS and the cluster interconnects (CF), does not exist, error messages of GLS and CF appear. Ignore the messages.

Also, the following messages for the cluster resource management facility might appear, but there is no problem.

```
2914 A new disk device(disk ) was found.
6807 Disk device (NodeID NodeID , disk ) cannot be detected.
6836 The disk device (NodeID NodeID , disk ) has changed.
```

2. Change the network address of Web-Based Admin View.

If the IP address to be used in Web-Based Admin View is changed in the environment after the migration, see "7.1 Network address," "7.3 Management server," and "7.5 Multi-network between server and client by classified use" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

3. Change the cluster interconnects.

    1. Change the setting with the cfconfig command (execute it on all nodes that constitute a cluster).

```
# cfconfig -d
# cfconfig -S <CF_node_name> <Cluster_name> <Device_name> [...]
```

After setting CF, execute the following command to check that they have been correctly set.

```
# cfconfig -g
```

2. Start CF on all nodes that constitute a cluster.

   For information on how to start CF, see "5.6 Starting and Stopping CF" in the "*PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide*."

   When executing "Load Driver," at the CF startup, clear the checkboxes of SF and GFS in the "Start CF" pop-up.

3. Check the CF status.

   Check that all CFs are started and cluster connects are in operation in the CF main window.

4. Change an IP address.

   When changing an IP address, change the following settings if necessary:

   - IP address and subnet for CIP

     See "9.2.2 Changing a CIP Address."

     When executing "Load Driver," at the CF startup, clear the checkboxes of SF and GFS in the "Start CF" pop-up.

   - IP address for the public LAN/administrative LAN

     Change the setting with the following procedure:

     1. Edit the /etc/inet/hosts file to change the IP address of each node.

        For Solaris 11, the setting with the ipadm command is also required.

     2. When you need to change the takeover IP address, modify the IP address to be used in the takeover IP address for the /etc/inet/hosts file of each node.

     3. Restart the system on all nodes.

        ```
        # /usr/sbin/shutdown -y -g0 -i6
        ```

     4. When the Ipaddress resource of IPv4 has been set by using the userApplication Configuration Wizard, perform the following procedure to change the takeover IP address in the resource database. When the Ipaddress resource of IPv6 is set, this procedure is unnecessary.

        (1) Identify the resource of the takeover IP address to be changed in the resource database.

        To identify it, use the clgettree(1) command on any one of the cluster node. The resource of the takeover IP address is the resource of the resource class name for the IP address.

        Example: In the following example, the resource of the takeover IP address is the one that the resource IDs are 56 and 57.

        ```
        # /etc/opt/FJSVcluster/bin/clgettree
        Cluster 1 cluster
             Domain 2 RC2
                     Shared 7 SHD_RC2
                             SHD_Host 58 config_Ipaddress0_ip1 UNKNOWN
                                     IP_Address 56 hme0:1 UNKNOWN node1
                                     IP_Address 57 hme0:1 UNKNOWN node2
                     Node 3 node1 ON
                             Ethernet 21 hme0 UNKNOWN
                                     IP_Address 56 hme0:1 UNKNOWN
                             Application 31 proc1 UNKNOWN
                             DISK 19 c0t0d0 UNKNOWN
                     Node 5 node2 ON
                             Ethernet 22 hme0 UNKNOWN
                                     IP_Address 57 hme0:1 UNKNOWN
                             Application 32 proc1 UNKNOWN
                             DISK 20 c0t0d0 UNKNOWN
        ```

        (2) Identify the shared resource of the takeover IP address.

        To identify it, use the clgettree(1) command on any one of the cluster nodes.

The shared resource of the takeover IP address is the resource of the resource class name for SHD_Host.

Example: In the example above, the shared resource of the takeover IP address is the one that the resource ID is 58.

(3) Change the takeover IP address.

Use the clsetrsc(1M) command to change it. The takeover IP address has been defined to the shared resource attribute of the one that was identified in (2) above.

Execute the clsetrsc(1M) command in the following format on each cluster node to change the takeover IP address.

```
# /etc/opt/FJSVcluster/sys/clsetrsc -A ip_addr=<IP_address_after_change>
<Resource_ID_of_shared_resource_for_IP_address>
```

Example: when changing the takeover IP address of the shared resource (resource ID of the resource=58) to 10.10.10.10

```
# /etc/opt/FJSVcluster/sys/clsetrsc -A ip_addr=10.10.10.10 58
```

(4) Check that the takeover IP address has been changed.

Use the clgetrsc(1) command to check it.

Execute the clgetrsc(1) command on any one of the cluster nodes in the following format to check that the takeover IP address has been changed.

```
# /etc/opt/FJSVcluster/sys/clgetrsc -a ip_addr
IP<resource_ID_of_the_shared_resource_for_the_address>
```

Example: when the takeover IP address of the shared resource (resource ID of the resource=58) was changed to 10.10.10.10

```
# /etc/opt/FJSVcluster/sys/clgetrsc -a ip_addr 58
ip_addr 10.10.10.10
```

5. Change the SF setting in the migration destination.

    1. Check the setting of the console asynchronous monitoring.

       Execute the following command on any one of the nodes.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -l
       ```

       <Execution example>

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -l
       device-name cluster-host-name  IP-address host-name user-name connection-type
       -----------------------------------------------------------------------------
       xscf        node1              xscf2      1         xuser     ssh
       xscf        node2              xscf2      1         xuser     ssh
       ```

    2. Cancel the setting of the console asynchronous monitoring.

       Specify the CF node name (cluster-host-name) displayed in 1. on any one of the nodes and execute the following command. In this procedure, all the CF node names displayed in 1. are specified.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -d node1
       # /etc/opt/FJSVcluster/bin/clrccusetup -d node2
       ```

       After specifying all the CF node names on all nodes and executing the command, execute the command in 1. again to check that nothing is displayed.

       ```
       # /etc/opt/FJSVcluster/bin/clrccusetup -l
       #
       ```

    3. Set the SF.

       To set the SF, see "5.1.2.1 For SPARC M10."

The following values can be set here:

- IP address for XSCF
- IP address for the administrative LAN

Note

For the migration from SPARC Enterprise to OVM of SPARC M10, you need to add the following IP addresses:

- Asynchronous monitoring sub-LAN (server side)

- In XSCF, two IP addresses (XSCF-LAN#0/XSCF-LAN#1) and in the redundant configuration for XSCF, six IP addresses (XSCF-LAN#0/XSCF-LAN#1*2, takeover IP*2)

6. Delete the patrol diagnosis of userApplication.

   Delete all the patrol diagnosis of userApplication. If the patrol diagnosis is not registered in the userApplication, this step is unnecessary.

```
# /etc/opt/FJSVcluster/bin/clspconfig -u <userApplication_name> -d
```

   After deleting it, execute the following command to check that the settings have been deleted.

```
# /etc/opt/FJSVcluster/bin/clspconfig -u <userApplication_name>
```

   <Execution example>

```
# /etc/opt/FJSVcluster/bin/clspconfig -u userApp_0 -d
# /etc/opt/FJSVcluster/bin/clspconfig -u userApp_0
FJSVcluster: ERROR: clspconfig: 6690: The specified userApplication or resource is not
monitored. (userApplication:userApp_0)
```

7. Delete resources of the network interface card.

   Use the cldelrsc command to delete resources of the network interface card.

   After deleting the resources of the network interface card, execute the following command to check that no resource of the network interface card are displayed.

```
# /etc/opt/FJSVcluster/bin/clgettree
```

8. Change the setting of GLS.

   Check that the SMF service of GLS has started with the following procedure.

```
# svcs -a | grep fjsvhanet
disabled       HH:MM:SS svc:/network/fjsvhanet:default
disabled       HH:MM:SS svc:/network/fjsvhanet-poll:default
```

   If the SMF service of GLS is in "disable" state, start it.

```
# svcadm enable /network/fjsvhanet
# svcadm enable /network/fjsvhanet-poll
```

```
# svcs -a | grep fjsvhanet
online         HH:MM:SS svc:/network/fjsvhanet:default
online         HH:MM:SS svc:/network/fjsvhanet-poll:default
```

   Check that the SMF service of GLS is in "online" state.

   When the SMF service is started, restart the guest domain.

```
# shutdown -y -g0 -i6
```

   Change the following settings on all nodes:

- NIC name (such as fjgiX ->vnetX)
- MAC address of the standby patrol (00:00:00:00:00:00 in the guest domain)

The following example indicates when changing the NIC used by the virtual interface sha0 from fjgi0 and fjgi2 to vnet0 and vnet1, and changing the MAC address of the standby patrol sha1 to 00:00:00:00:00:00.

How to check the setting

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+----------------------+
 sha0        192.168.0.100    d   192.168.0.1       fjgi0,fjgi2
 sha1        -                p   02:00:00:00:00:01 sha0

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+----------------------+
```

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------+
 sha0:65    192.168.0.100    -
```

How to delete the takeover virtual interface temporarily

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc delete -n sha0:65
hanet: 00000: information: normal end.
```

```
# /opt/FJSVhanet/usr/sbin/dsphanet
[IPv4,Patrol]
 Name       Status   Mode CL  Device
+----------+--------+----+----+------------------------------------------------+
 sha0       Inactive  d   OFF  fjgi0(OFF),fjgi2(OFF)
 sha1       Inactive  p   OFF  sha0(OFF)
[IPv6]
 Name       Status   Mode CL  Device
+----------+--------+----+----+------------------------------------------------+
```

[Supplement]

Use the dsphanet command to check that Status is Inactive.

When the virtual interface, which Status is Active, exists, use the stphanet -n shaX command to inactivate the virtual interface. The virtual interface to be inactivated is the one that Mode is "d," "e," or "c."

How to change the setting

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -t vnet0,vnet1
hanet: 00000: information: normal end.
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha1 -a 00:00:00:00:00:00
hanet: 00000: information: normal end.
```

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
hanet: 00000: information: normal end.
```

How to check the setting

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name        Hostname        Mode MAC Adder/Phys ip Interface List
+-----------+---------------+----+-----------------+----------------------+
 sha0        192.168.0.100    d   192.168.0.1       vnet0,vnet1
```

```
 sha1        -                p   00:00:00:00:00:00 sha0

[IPv6]

 Name        Hostname/prefix                   Mode Interface List
+-----------+---------------------------------+----+-----------------------+
```

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+----------------------------------------------+
 sha0:65    192.168.0.100    -
```

See

For information on how to change IP address, see "*PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function*."

9. Remove the shared class and the local class forcibly.

    1. Delete the class resources.

    ```
    # /etc/opt/FJSVsdx/bin/sdxdcrsc -R -c class01
    # /etc/opt/FJSVsdx/bin/sdxdcrsc -R -c class02
    ```

    If a message indicating that there is no resource, ignore the message and proceed to the next step.

    2. Check that the shared class and the local class have been deleted. Make sure that these classes have been deleted on all nodes in the cluster.

    ```
    # sdxinfo -c class01
    # sdxinfo -c class02
    ```

10. Delete resources of a shared disk device.

    Use the cldeldevice command to delete resources of the registered shared disk device.

    ```
    # /etc/opt/FJSVcluster/bin/cldeldevice -a
    ```

    After deleting resources, execute the following command to check that no resources of the deleted shared disk device are displayed.

    ```
    # /etc/opt/FJSVcluster/bin/clgettree
    ```

11. Register resources again (clautoconfig).

    ```
    # /etc/opt/FJSVcluster/bin/clautoconfig -r
    ```

12. Set the operation for the patrol diagnosis.

    If the patrol diagnosis is not set, this procedure is unnecessary.

    Use the clspconfig(1M) command to set the patrol diagnosis. For procedure to set it, see "6.9 Setting Up Patrol Diagnosis."

    After completing the setting, execute the following command to check that the setting value is correct.

    ```
    # /etc/opt/FJSVcluster/bin/clspconfig -u <userApplication_name>
    ```

    <Execution example>

    ```
    # /etc/opt/FJSVcluster/bin/clspconfig -u userApp_0
    ACTIVE=true
    INTERVAL=360
    TARGET_LAN=vnet2
    TARGET_DISK=c1d0
    TARGET_DISK=c1d1
    ```

```
TARGET_DISK=c1d2
TARGET_DISK=c1d3
```

13. Convert the GDS configuration information.

    If the physical disk names are different between the migration source and migration destination, convert the GDS configuration information with the following procedure. The sdxconfig Convert command ends with the return code "0" if it has been successfully completed. If there are multiple classes, convert the GDS configuration information for all shared and local classes. (Execute the command on any one of the nodes in the class scope.)

    Example

    Change the physical disk c0t4d1 described in the configuration file /var/tmp/Class1.conf to c1d3.

    ```
    # sdxconfig Convert -e replace -c class01 -p c0t4d1=c1d3 -i /var/tmp/Class1.conf -o /var/tmp/
    Class1.conf -e update
    # echo $?
    0
    ```

14. Restore the GDS configuration information.

    If there are multiple classes, restore it for all shared and local classes (on the node where Step 13. was performed).

    ```
    # sdxconfig Restore -c class01 -i /var/tmp/Class1.conf -e chkps
    # shutdown -y -i6 -g0
    ```

    After restarting the system, execute the sdxinfo command to check that the class object in the migration source has been restored as the local class.

15. Change the local class to the shared class.

    Once the configuration information is restored, it is restored as the local class even the backup source is originally in the shared class.

    Use the following command to restore it in a shared class. If there are multiple shared classes in the backup source, restore the configuration information for all the shared classes (on the node where Step 14. was performed).

    ```
    # sdxvolume -F -c class01
    # sdxattr -C -c class01 -a type=shared,scope=node1:node2
    ```

    Use the sdxinfo command to check that types of class01 and scope have been correctly changed.

16. Set Gds resources to be registered in the userApplication.

    This setting is for registering the shared class in the userApplication and using it.

    Execute the following command on any one of the nodes in the class scope for the shared class to be registered in the userApplication.

    ```
    # /opt/SMAW/SMAWRrms/bin/hvgdsetup -a class_name
    ```

17. Restore the management partition information of GFS.

    1. Re-initiate the management partition.

       Execute the following command on any one of the nodes.

       ```
       # sfcsetup -c -f /dev/sfdsk/class01/rdsk/control
       ```

    2. Re-register the node configuration information in the management partition.

       Execute the following command on all nodes.

       ```
       # sfcsetup -a /dev/sfdsk/class01/rdsk/control
       ```

       The path name of the management partition that has been set up can be confirmed by using the sfcsetup(1M) command with the -p option.

```
# sfcsetup -p
/dev/sfdsk/class01/rdsk/control
```

The registered node configuration information can be confirmed by using the sfcsetup(1M) command without any option.

```
# sfcsetup
HOSTID      CIPNAME     MP_PATH
80a4f75b    node1RMS     yes
80960096    node2RMS     yes
```

3. Start the sfcfrmd daemon.

   Execute the following command on all nodes.

```
# sfcfrmstart
```

4. Restore the management partition information.

   Execute the shell script _backup_file_collected which is created by the sfcgetconf(1M) command in "18.1 Backing Up the Configuration Information of GFS/GDS" on any one of the nodes.

```
# sh _backup_file_
get other node information start ... end
get other node information start ... end
```

   Check that the management partition information of GFS has been restored with the sfcinfo(1M) command and the sfcrscinfo(1M) command.

```
# sfcinfo -a
/dev/sfdsk/class02/dsk/volume01:
FSID special                                         size  Type  mount
   1 /dev/sfdsk/class02/dsk/volume01(11500000021)     14422  META  -----
   1 /dev/sfdsk/class02/dsk/volume01(11500000021)      5116  LOG   -----
   1 /dev/sfdsk/class02/dsk/volume01(11500000021)     95112  DATA  -----

/dev/sfdsk/class02/dsk/volume02:
FSID special                                         size  Type  mount
   2 /dev/sfdsk/class02/dsk/volume02(11500000022)     14422  META  -----
   2 /dev/sfdsk/class02/dsk/volume02(11500000022)      5116  LOG   -----
   2 /dev/sfdsk/class02/dsk/volume02(11500000022)     95112  DATA  -----
   2 /dev/sfdsk/class02/dsk/volume03(11500000023)     65512  DATA  -----
```

```
# sfcrscinfo -m -a
/dev/sfdsk/class02/dsk/volume01:
FSID  MDS/AC  STATE  S-STATE   RID-1   RID-2   RID-N  hostname
   1  MDS(P)  stop   -             0       0       0  host1
   1  AC      stop   -             0       0       0  host1
   1  MDS(S)  stop   -             0       0       0  host2
   1  AC      stop   -             0       0       0  host2

/dev/sfdsk/class02/dsk/volume02:
FSID  MDS/AC  STATE  S-STATE   RID-1   RID-2   RID-N  hostname
   2  MDS(P)  stop   -             0       0       0  host2
   2  AC      stop   -             0       0       0  host2
   2  MDS(S)  stop   -             0       0       0  host1
   2  AC      stop   -             0       0       0  host1
```

5. Mount the GFS Shared File System.

   Mount the target file system.

Note

Check the entries for the GFS Shared File System (which is the line that file system type is sfcfs) in the /etc/vfstab file are valid.

If the entry has been commented out (if the line starts with "#") delete the comment of the entry.

See

For details on how to mount the GFS Shared File System, see "11.2.4 Mount" in the "*PRIMECLUSTER Global File Services Configuration and Administration Guide*."

18. Set the system disk mirroring again.

    Re-set the system disk mirroring.

    See

    For details on the procedure, see "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*."

19. Change the setting of the takeover IP address.

    When you use the takeover network resource, follow "10.6.1 Changing the Interface Used by a Resource" and set the takeover network resource again for the network interface of after performing P2V.

20. Set the automatic start for RMS.

    When starting RMS automatically at OS startup, perform the following setting on all nodes.

    How to set

```
# hvsetenv HV_RCSTART 1
```

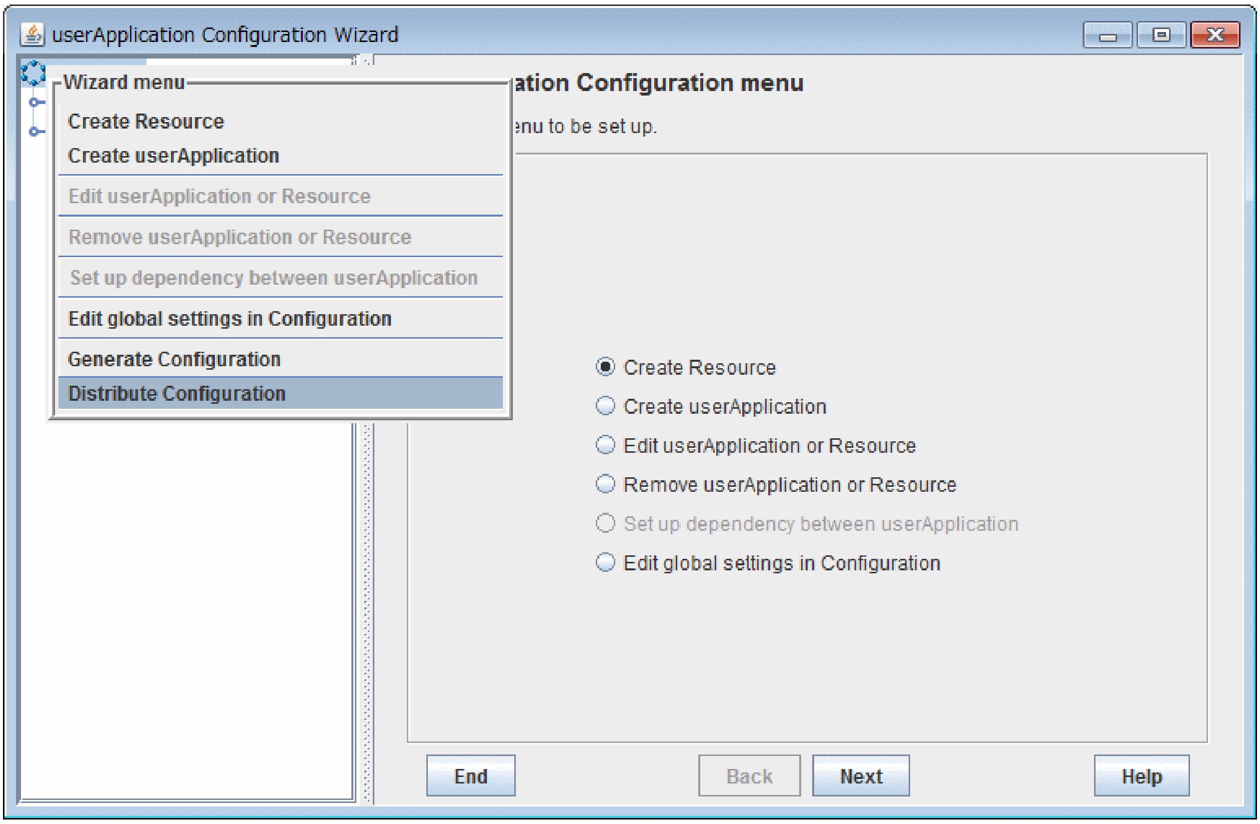
    How to check

```
# hvsetenv HV_RCSTART
1 <- Check that "1" is displayed.
```

21. Enable the RMS setting. Perform this operation on any one of the nodes that constitute a cluster.

    1. Start the userApplication Configuration Wizard.

2. Right-click on the icon which means Configuration on the top of the "tree view" on the left side of the screen.



3. Select and left-click on "Distribute Configuration" displayed in the pull-down menu "Wizard menu."

22. Start RMS.

    Execute the following command on all nodes to start RMS.

```
# hvcm
```

23. Check the RMS startup and the resource state.

    Execute the following command on all nodes to check that RMS has started and the resource state.

```
# hvdisp -a
Local System:  node1RMS
Configuration: /opt/SMAW/SMAWRrms/build/config.us

Resource             Type    HostName            State        StateDetails
-----------------------------------------------------------------------------
node1RMS             SysNode                     Online
node2RMS             SysNode                     Online
userApp_0            userApp                     Online
Machine001_userApp_0 andOp   node2RMS
Machine000_userApp_0 andOp   node1RMS            Online
ManageProgram000_Cmdline1 gRes                   Online
ManageProgram000_Cmdline0 gRes                   Online
ManageProgram000_Cmdline2 gRes                   Online
SDISK.64             gRes                        Online
Ipaddress000_Gls0    gRes                        Online
MountPoint001_Fsystem0 gRes                      Online
MountPoint001_Fsystem1 gRes                      Online
MountPoint001_Fsystem2 gRes                      Online
Zpool000_Fsystem2    gRes                        Online
```

```
AllDiskClassesOk_Gds0 andOp                    Online
class01_Gds0        gRes                       Online
```

Part 7 PRIMECLUSTER Products

This section explains the PRIMECLUSTER products.

- PRIMECLUSTER Product List
- PRIMECLUSTER Wizard for NetWorker
- PRIMECLUSTER Wizard for Oracle
- PRIMECLUSTER Wizard for NAS
- Interstage Application Server Enterprise Edition
- ETERNUS SF AdvancedCopy Manager
- Symfoware Server
- Systemwalker Centric Manager
- Systemwalker Operation Manager
- Systemwalker Service Quality Coordinator

Chapter 19 PRIMECLUSTER Product List

The following lists the version levels of PRIMECLUSTER products and the range of support:

Table 19.1 PRIMECLUSTER product list

Product name	Operation mode	Standby operation					Scalable operation	Single-node cluster operation
	System configuration	Standby			Mutual standby	Cascade		
		1:1		N:1				
	IP address takeover	Yes			Yes	Yes		
	Node name takeover	Yes	No	No	No	No		
Interstage Application Server Enterprise Edition V10.0.0		N	Y	N	Y	N	N	Y
Symfoware Server V10.1.0		Y	Y	Y	N	Y	N	Y
ETERNUS SF AdvancedCopy Manager 14.2		Y	Y	N	Y	N	N	N
Systemwalker Centric Manager Enterprise Edition V13.5.0 Systemwalker Centric Manager Global Enterprise Edition V13.5.0		N	Y	N	N	N	N	N
Systemwalker Operation Manager Enterprise Edition V13.4.1 Systemwalker Operation Manager Global Enterprise Edition V13.4.1		Y	Y	Y	Y	N	N	Y
PRIMECLUSTER Wizard for Oracle 4.3 (*1)		Y	Y	Y	Y	Y	Y	Y
PRIMECLUSTER Wizard for NetWorker 4.3 (*2)		Y	Y	N	N	Y	Y	Y
PRIMECLUSTER Wizard for NAS 4.3		Y	Y	Y	Y	Y	Y	Y
Oracle Real Application Clusters for PRIMECLUSTER 10g (*3)		N	N	N	N	N	Y	N
Systemwalker Service Quality Coordinator Enterprise Edition V13.5.0 (*4)		Y	N	N	N	N	N	N
Netcompo WAN control 1.1.1 or later		Y	Y	Y	Y	N	N	N

(*1) For scalable operation, Oracle Real Application Clusters for PRIMECLUSTER is required.

(*2) Operations of N:1 and Mutual standby between cluster applications which include NetWorker resources are not supported. Operations of N:1 and Mutual standby between cluster applications with NetWorker resources and without NetWorker resources are supported. Scalable operations are supported only for the client and cluster operation.

(*3) PRIMECLUSTER Wizard for Oracle is required.

(*4) For Enterprise Manager and Manager

19.1 PRIMECLUSTER Operation Modes

This section explains the cluster operation mode of PRIMECLUSTER, based on SynfinityCluster.
Note that the "cluster service" of SynfinityCluster is relevant to the "cluster application" of PRIMECLUSTER.

Also, "network takeover" of SynfinityCluster relates to the "public LAN takeover" of PRIMECLUSTER.

For the terminologies, see Appendix "D.1 Terminology".

Cluster operation mode

Cluster operation mode	Description
1:1 standby	A single standby cluster application runs on a system that consists of two nodes. One is an operating node, and the other is a standby node.
N:1 standby	"N" standby cluster applications run on a system consisting of "N+1" nodes. One application runs on each operating node. Two or more applications stand by on one node.
Mutual standby	One cluster application runs, and the other cluster application stands by on each node. Generally, a cluster system consists of 2 nodes. This is referred to as two-node mutual standby.
Cascade	One operating node has two or more standby nodes. Even when one node stops, the cluster application assures its redundancy by using other node. This operation mode is effective for maintaining the availability of a cluster application during maintenance.
Priority transfer	Topology applied from N:1 standby Effective in maintaining availability during maintenance because the restored node becomes a standby node through the transfer of the standby node, even after the standby node stops and the cluster application can adopt a redundant configuration.
Scalable	Topology that is best suited for parallel job execution because one or more cluster applications operate together.
High-availability scalable	Topology in which standby operation is performed for each cluster application that constitutes a scalable configuration.

Cluster operation mode

Cluster operation mode	Description
Standby operation mode	General names for the four operation modes are 1:1 standby, N:1 standby, mutual standby, and cascade.
Scalable operation mode	One scalable cluster application consists of two or more operating nodes. If one node stops in the event of a failure, ongoing operations are continued on other operating nodes. Each operating node can be configured to be in the standby operation mode.

Public LAN takeover function

Public LAN takeover function	Description
IP address takeover	If a failover occurs during standby operation, one or more IP addresses are taken over to other node. In such a case, two or more network interface cards can be used.
Node name takeover	If a failover occurs during standby operation, the node name is taken over to other node. This function enables access to the logical IP address corresponding to the node name.

Note

Physical IP address takeover deactivates a physical IP address at failover. To avoid disrupting ongoing operations of Web-Based Admin View or Shutdown Facility, set up physical IP address takeover for a network interface that is not being used by that Web-Based Admin View or Shutdown Facility.

Chapter 20 PRIMECLUSTER Wizard for NetWorker

20.1 Functional Overview

The PRIMECLUSTER Wizard for NetWorker is a software product which operates NetWorker built on a cluster system in PRIMECLUSTER.

The product performs error monitoring for NetWorker and, when an error is detected, rebooting of NetWorker and switching operational nodes are performed automatically.

It is a required product for backup operations which NetWorker is used under a PRIMECLUSTER environment.

Startup and stop control

In line with the state transition of the cluster application, scripts automatically start or stop NetWorker.

Error monitoring

This function performs process monitoring for NetWorker.

Building the environment

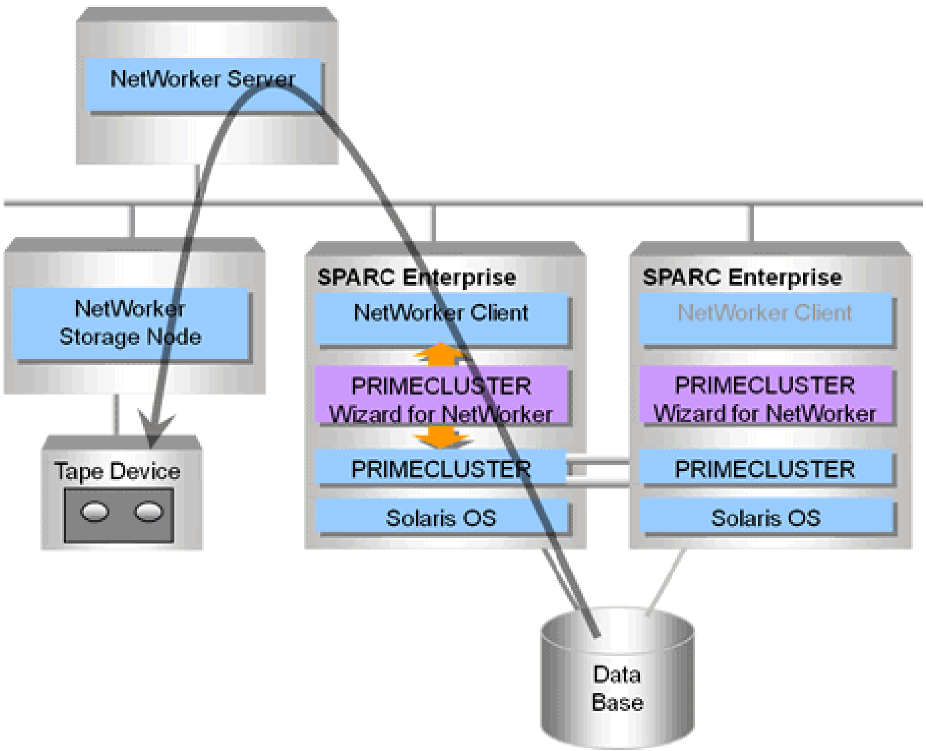
This function provides an environment setup wizard which creates cluster applications and Networker resource settings in PRIMECLUSTER.

20.2 Operation Environment

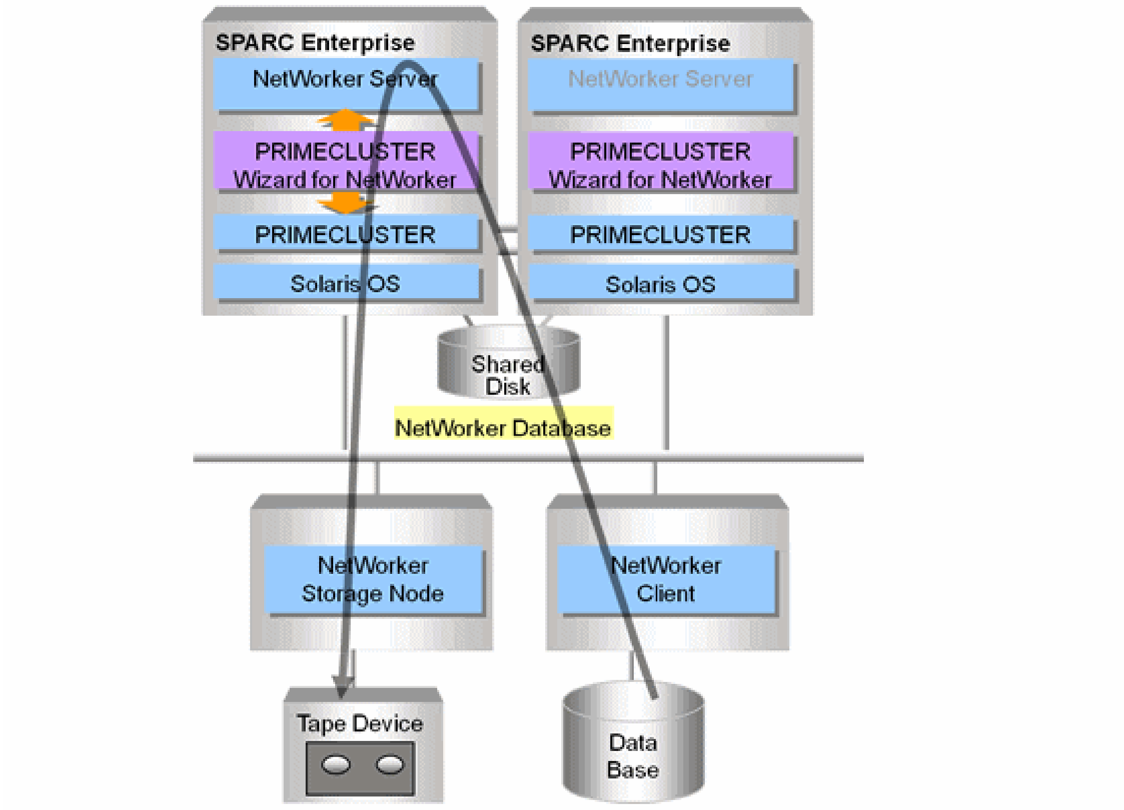
20.2.1 System Configuration

System configuration figures are shown below:

Client cluster operation



Server cluster operation



20.2.2 Supported Modes

For details, see the "Table 19.1 PRIMECLUSTER product list."

20.2.3 Overview of Cluster System Support

The PRIMECLUSTER Wizard for NetWorker provides the environment setup wizard that is used to build a backup system on the cluster system by using NetWorker, startup/stop scripts, and state monitoring detectors, and supports standby operation using NetWorker.

20.3 Notes

For details on the environment setup with the PRIMECLUSTER Wizard for NetWorker, see the "*PRIMECLUSTER Wizard for NetWorker Configuration and Administration Guide,"* which is provided with the PRIMECLUSTER Wizard for NetWorker.

Chapter 21 PRIMECLUSTER Wizard for Oracle

21.1 Functional Overview

PRIMECLUSTER Wizard for Oracle is a software product which operates Oracle Database (abbreviated as Oracle) built on a cluster system in PRIMECLUSTER. PRIMECLUSTER Wizard for Oracle ensures scalable operation with Oracle Real Application Clusters (abbreviated as RAC) and the switching standby operation.

Startup and stop control

In line with the state transition of the cluster system, scripts automatically start or stop the Oracle instance.

- Scalable operation
  - Oracle instance

    Use the Oracle srvctl start/stop instance command to start and stop the Oracle instance.

  - Oracle Clusterware

    Start and stop the Oracle Clusterware.

- Standby operation
  - Oracle instance

    Check the state of the Oracle instance and start while the recovery is performed.

    For example, if the Oracle instance is down during the ONLINE BACKUP of the Oracle instance, execute the end backup automatically to perform the automatic start of the Oracle instance.

    The scripts stop the Oracle instance in the immediate mode.

    When it cannot be stopped normally, it can be switched fast by stopping the Oracle instance with the abort mode.

  - Oracle Listener

    Use the Oracle lsnrctl start/stop command to start and stop the Oracle listener.

    Check the listener process at the same time of start/stop process and confirm that they have been performed correctly.

Monitoring

A detector monitors the Oracle instance and Oracle listener.

- Monitoring of Oracle Instance
  1. Connects itself to Oracle using system user.
  2. Use a dummy table for monitoring to update the data.
  3. Confirming Oracle error code

  A logical failure can be detected as well as the process of the Oracle instance being active or inactive by this monitoring method.

- Monitoring of Oracle listener
  1. Monitor the listener process.
  2. For the standby operation, monitoring by using the Oracle tnsping command is possible. (It is possible when the monitoring items of Wizard are set up).

Environment setup

The environment setup tool provides the Oracle resource setup and Environment setup Wizard to create a cluster application in PRIMECLUSTER.

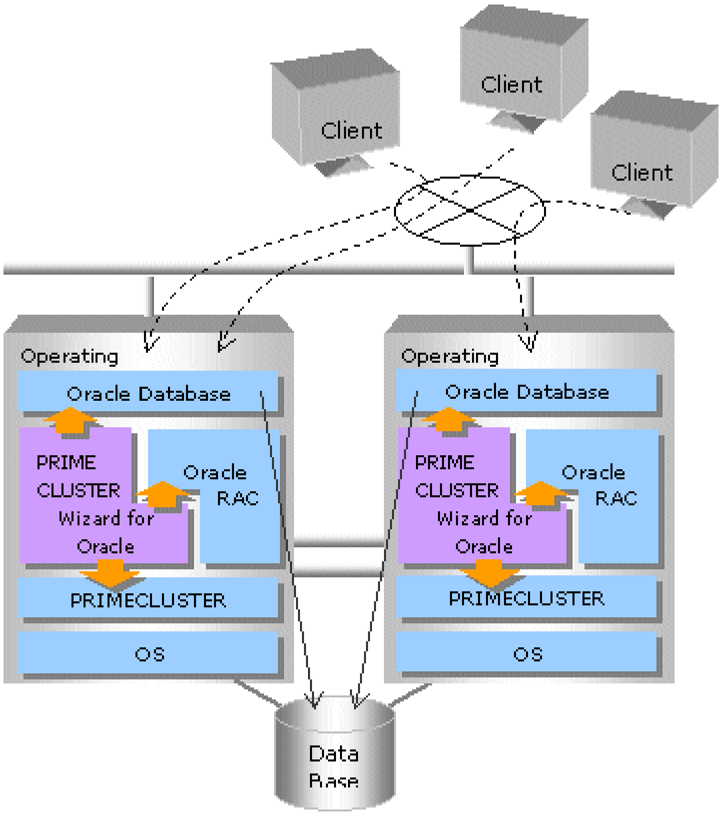
21.2 Operation Environment

21.2.1 System Configuration

System configuration figures are shown below:

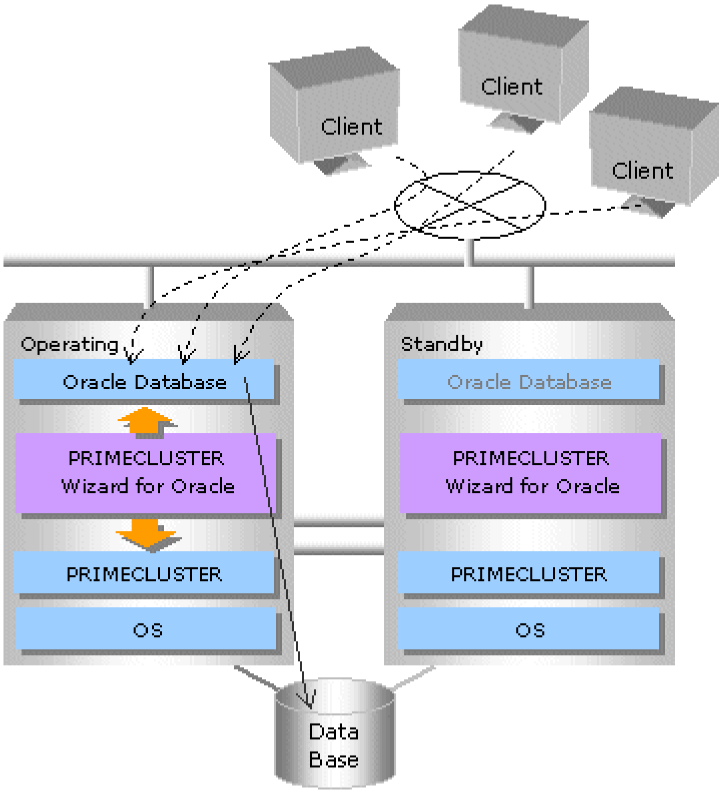
Scalable operation (RAC)

In scalable operation with RAC, Oracle is operated on all nodes. A client can use a database that is connected to either node.



Standby operation (cold-standby)

In standby operation, a cluster system consists of an operational node and standby nodes. On the operational node, Oracle applications, a logical IP address and shared disks are active. On the standby nodes, these resources are inactive. In the event of a failure, production operation is switched to one of the standby nodes, and the resources in this standby node are activated. The client can connect to the running node using the logical IP address without configuration changes.



21.2.2 Supported Modes

For details, see the "Table 19.1 PRIMECLUSTER product list."

21.2.3 Overview of Cluster System Support

The PRIMECLUSTER Wizard for Oracle provides the environment setup wizard that is used to build Oracle on the cluster system, the start/stop script, and the state-monitoring detector. It also supports scalable operation using Oracle9i RAC and Oracle cold-standby operation.

21.3 Notes

For details on the environment setup, etc., see the *"PRIMECLUSTER Wizard for Oracle Configuration and Administration Guide,"* which is attached to the PRIMECLUSTER Wizard for Oracle.

Chapter 22 PRIMECLUSTER Wizard for NAS

22.1 Functional Overview

The PRIMECLUSTER Wizard for NAS is a software product that is for when using an ETERNUS NR1000F series (hereafter referred to as the "NR1000F") as a shared disk device for a PRIMECLUSTER-based cluster system.

Startup and stop control

This function follows the state transition for the userApplication and automatically controls the export/unexport of the NAS device's volume accordingly. Also, during standby operation, this function cancels the file lock that the old operating node was maintaining during startup.

Automatic control in the event of a failure

The PRIMECLUSTER Wizard for NAS does not perform error monitoring for the NAS device, but by means of other resources, error detection and switch control are possible.

- PRIMECLUSTER GLS monitors the access path to the NAS device.
- The RemotedFileSystem resource, which allows remote file system mounting control, monitors accesses to the volume.

Environment setup

The environment setup wizard for running a NAS device as a shared disk on PRIMECLUSTER allows you to configure a cluster environment based on a NAS device easily and safely.

Integrating NR1000F of cluster configuration

- The product supports operation linked with a NR1000F cluster configuration that has made the controller redundant through the use of two NR1000F devices.
- If, between the cluster configuration NR1000F and the nodes, data access cannot be preformed due to something such as a transmission path error, the PRIMECLUSTER Wizard for NAS detects the error and takes over control of the cluster configuration NR1000F. The cluster configuration NR1000F takes over control when there is a malfunction with the controllers between devices, but does not take over for transmission path errors to the NR1000F. Through the PRIMECLUSTER Wizard for NAS linking to the cluster configuration NR1000F, it is possible to continue work even when there is a transmission path error.

22.2 Operating Environment

22.2.1 System Configuration

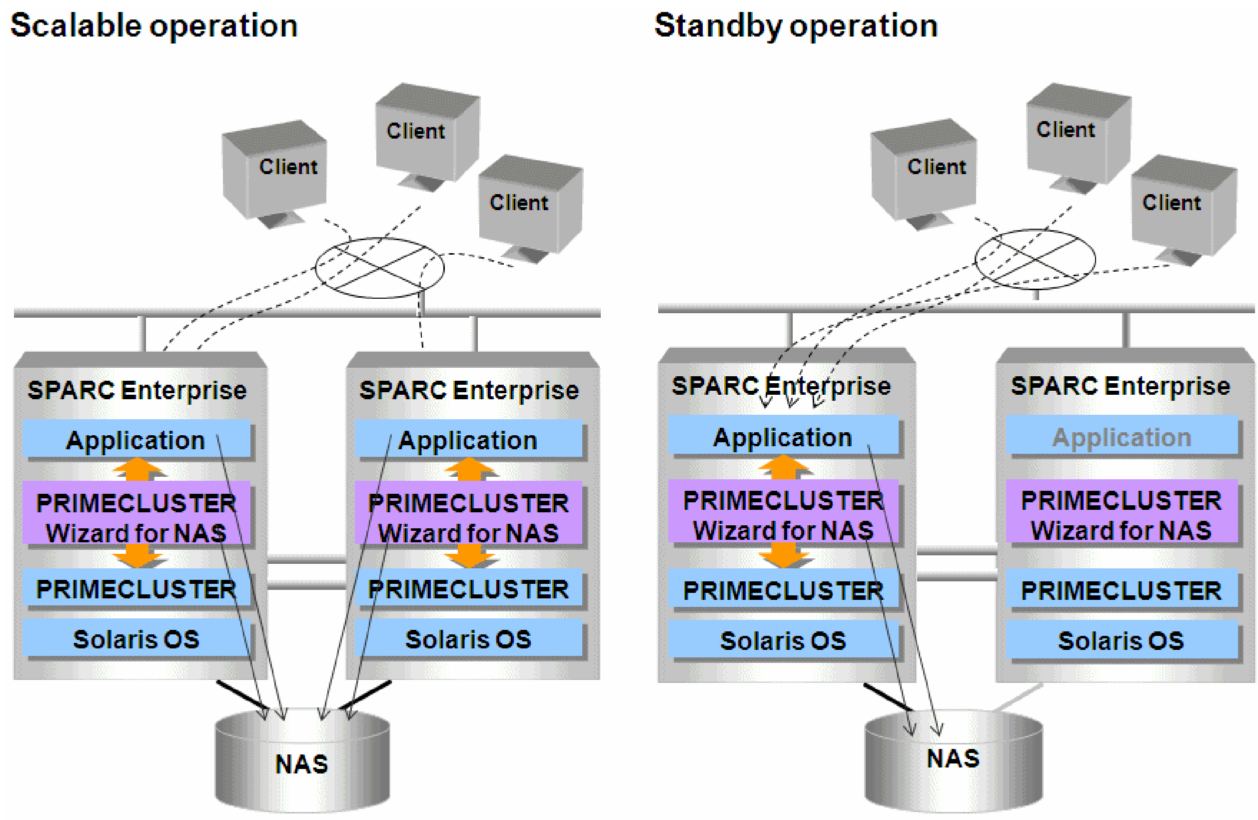
System configuration figures are shown below:

Scalable operation

In scalable operation, a volume of the NAS device is mounted from all nodes. A shared disk can be used from any nodes.

Standby operation

In standby operation, a cluster system consists of an operational node and a standby node. On the operational node, NAS device is only active. In the event of a failure, production operation is switched to one of the standby nodes and the connection is activated after unlocking the operation node in NAS device.



22.2.2 Supported Modes

For details, see the "Table 19.1 PRIMECLUSTER product list."

22.2.3 Overview of Cluster System Support

The PRIMECLUSTER Wizard for NAS provides the environment setup wizard, the start/stop script, the state-monitoring detector and so on to use NR1000F as a shared disk unit of the cluster system.

22.3 Notes

For details on the environment setup and so on, see the "PRIMECLUSTER Wizard for NAS Configuration and Administration Guide" which is attached to the PRIMECLUSTER Wizard for NAS.

Chapter 23 Interstage Application Server Enterprise Edition

23.1 Functional Overview

The Interstage Application Server is an application server product which can simultaneously execute on one server multiple applications developed with different languages (Java, COBOL, and C), and mutually link them together.

The Interstage Application Server Enterprise Edition is the highest-ranking of the editions provided by the Interstage Application Server. Assuming the need to provide round-the-clock service, the Interstage Application Server Enterprise Edition guarantees a stable response even under high load.

23.1.1 Overview of Cluster System Support

The Interstage Application Server Enterprise Edition supports two modes: Namely, 1:1 standby, which is a so-called hot standby system in which the operation is automatically taken over by the standby node if a failure occurs in the OPERATING node; and mutual standby, in which two nodes perform each operation and, if a failure occurs on one node, the other node takes over the operation.

With the Interstage Application Server Enterprise Edition, operating on the cluster system allows high-speed switching to another server to take over the operation even in the event of a failure. The client can continue the transaction through reconnect processing and can minimize downtime.

23.1.2 Supported Modes

For details, see the "Table 19.1 PRIMECLUSTER product list."

23.1.3 Note

For details on building the cluster system, refer to "Chapter 4 Environment Setup Procedure for Cluster Service" of the "Interstage Application Server High Availability System Guide."

Chapter 24 ETERNUS SF AdvancedCopy Manager

24.1 Functional Overview

This product realizes storage management for open systems and consists of the following three solutions:

High-speed backup (round-the-clock support)

Data can be backed up at high speed by using the hardware [advanced copy function (OPC, EC)] of the Fujitsu disk array unit (ETERNUS3000, GR720, GR730, GR740). Since the advanced copy function is a high-speed data copy function that is not dependent on the volume of data, a constant, high-speed backup time can be always maintained even when the amount of data to be backed up increases. Thus, the backup time does not increase even when the data quantity increases.

By linking this solution with the following DBMS, high-speed backup can be done without stopping the job:

- ORACLE
- SymfoWARE

For those jobs that are constructed using a database other than the above or a general file system, the job non-stop time required for the backup is significantly reduced compared with that in the past.

Database backup without stopping the job

Oracle and SymfoWARE database backup can both be done without stopping the job by using ETERNUS SF AdvancedCopy Manager.

When combined with Global Disk Services, ETERNUS SF AdvancedCopy Manager can be used to perform backup without stopping Global Disk Services, even in the mirror volume (SDX object) environment.

High-speed replication

By using the hardware [advanced copy function (OPC, EC)] of the Fujitsu disk array unit (ETERNUS3000, GR720, GR730, GR740), the data can be used for various purposes and a data replication (copy) can be collected to guard against accidental data destruction.

A remote disk array unit in the SAN environment is also available for storing the replication (copy) for the purpose of restoring the data after a catastrophic failure.

24.1.1 Overview of Cluster System Support

With ETERNUS SF AdvancedCopy Manager, units constituting a cluster system that has been constructed by using cluster software are controlled as a single business server. By monitoring failover and the like, recovery from errors can be done quickly.

24.1.2 Supported Modes

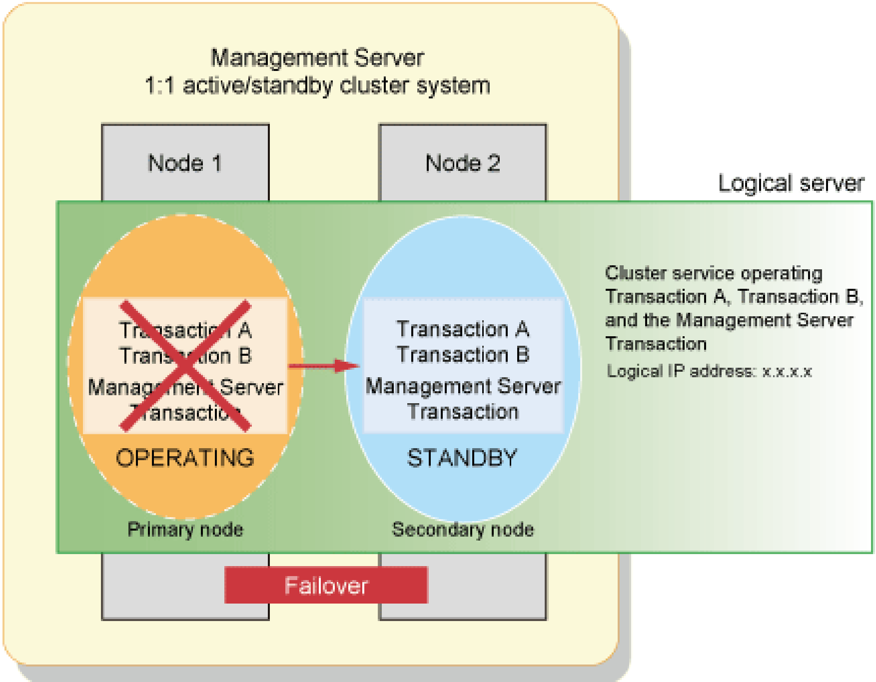
ETERNUS SF AdvancedCopy Manager supports 1:1 standby and mutual standby by means of IP address takeover. For details, see "Table 19.1 PRIMECLUSTER product list."

24.2 Standby Classes

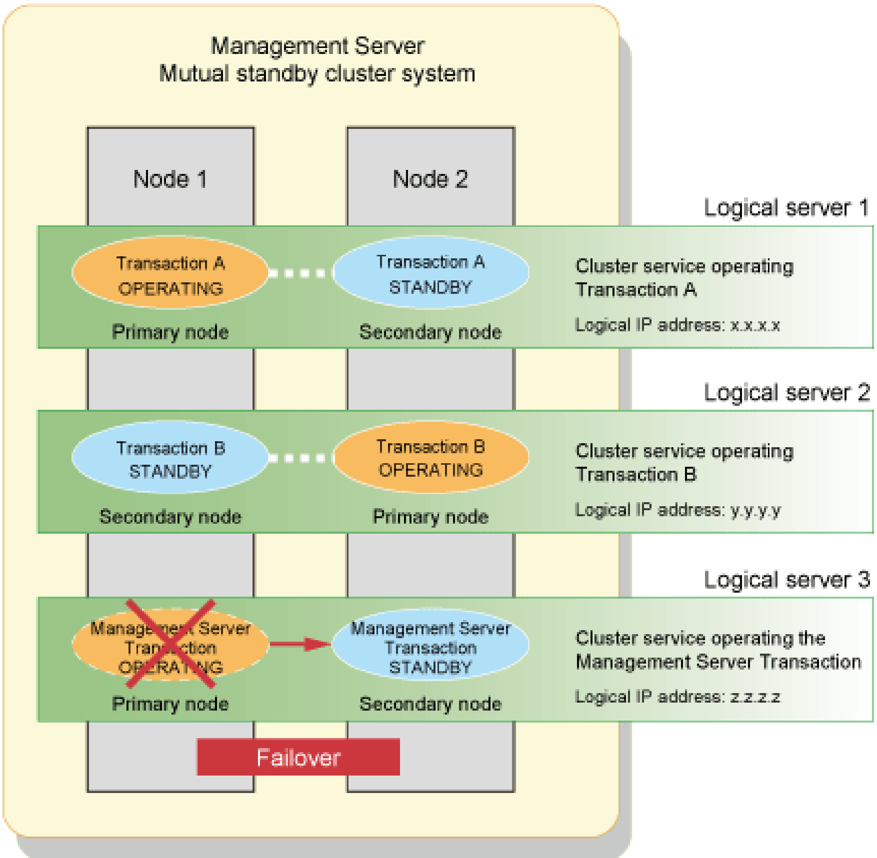
24.2.1 System Configuration

The supported system configurations are shown below:

Standby type cluster system configuration



Mutual standby type cluster system configuration



24.2.2 Setting Up the Environment

For details, see the "ETERNUS SF AdvancedCopy Manager Cluster Application Guide (SynfinityCluster/PRIMECLUSTER)."

Chapter 25 Symfoware Server

This chapter explains the support provided for the Symfoware Server cluster system.

25.1 Functional Overview

Symfoware Server is a database system that has a wide variety of functions applied to the system requirements for mission-critical transactions such as; high-reliability, high-performance, and high-scalability.

25.1.1 Support for Cluster System

Symfoware Server provides failover and load sharing functions to support the cluster system.

You can build a high-available, high-scalable, and high-reliable system by using those functions on the cluster system.

- Failover function

  This is a function to take over the operation running on the node to another node when a failure occurs on one node in the cluster system. There are the standby function and hot-standby function.

- Load sharing function

  This is a function to handle transactions in parallel with multiple nodes on the cluster system.

For details on the failover function and load sharing function for Symfoware Server, see "Symfoware Server Cluster Installation and Operation Guide."

25.1.2 Supported Products

For details, see "Chapter 19 PRIMECLUSTER Product List."

25.1.3 Points of Caution

For details on cluster system configuration, see "Symfoware Server Cluster Installation and Operation Guide."

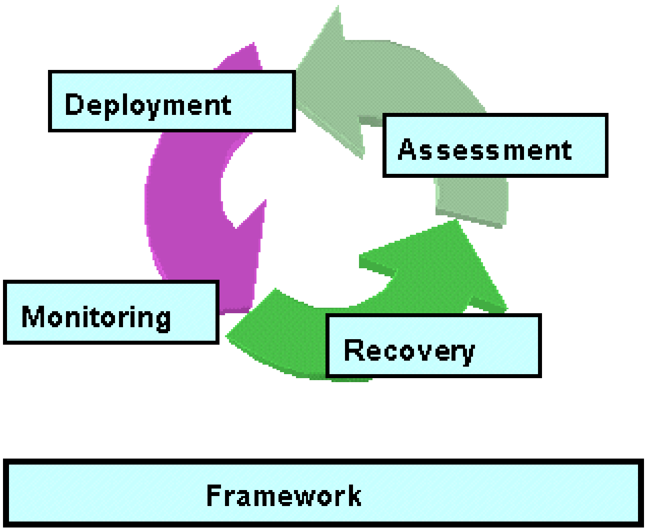
Chapter 26 Systemwalker Centric Manager

26.1 Functional Overview

Systemwalker Centric Manager is a software product that acts as the integrative infrastructure for the management of an information system. It provides life cycle management of IT resources such as networks, hardware, OS, applications, and users that consist of the system according to the policy based management.

Life cycle management

Systemwalker Centric Manager controls the life cycle of IT resources. It provides an effective operation management by categorizing the IT resources into the following 4 phases:



Deployment

Refers to the phase in which IT resources and transactions are installed on the system and set up. The resources to be used for the information system are deployed and applied.

Monitoring

Refers to the phase in which the performance of IT resources and failures are monitored.

Recovery

Refers to the phase in which corrective action is taken in the event of a failure. Troubleshooting and recovery are performed.

Assessment

Refers to the phase in which the stability of operation is evaluated. The service level is evaluated and analyzed from operation information of IT resources. Performance bottlenecks are detected, and capacity planning is done.

Framework

The "framework" refers to the operation management infrastructure that is used to share information between the functions of the life cycle and to enable seamless functional cooperation. The framework collects the configuration information for the entire information system, such as the network, hardware, OS, applications and others, and maintains the integrated management of the configuration information.

26.1.1 Overview of Cluster System Support

Systemwalker Centric Manager enables the integrated management of a cluster system. Systemwalker Centric Manager monitors that the entire cluster system is normally running. Also, Systemwalker Centric Manager eliminates downtime caused by node-down, and provides a highly reliable system by performing quick recovery at failover.

26.1.2 Supported Modes

Systemwalker Centric Manager supports a 1:1 standby operation mode using IP address takeover.

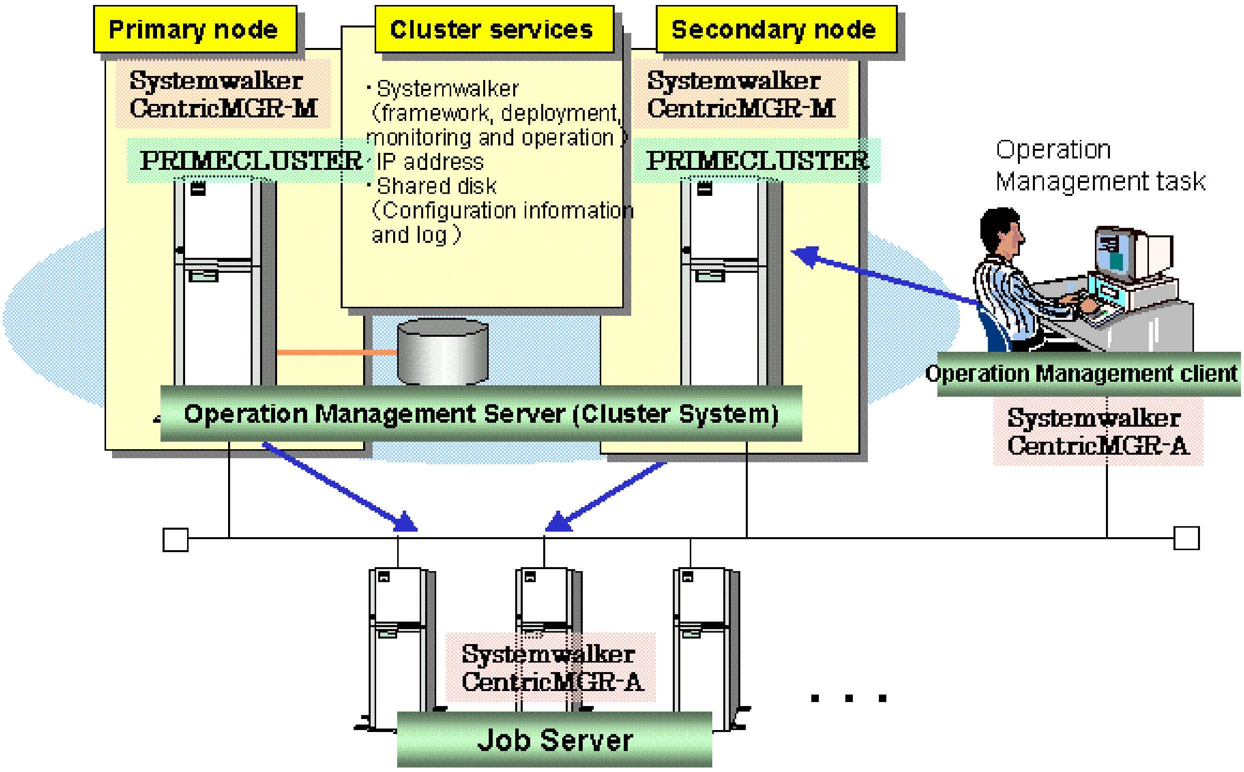
For details, see the "Table 19.1 PRIMECLUSTER product list."

26.2 Standby Class

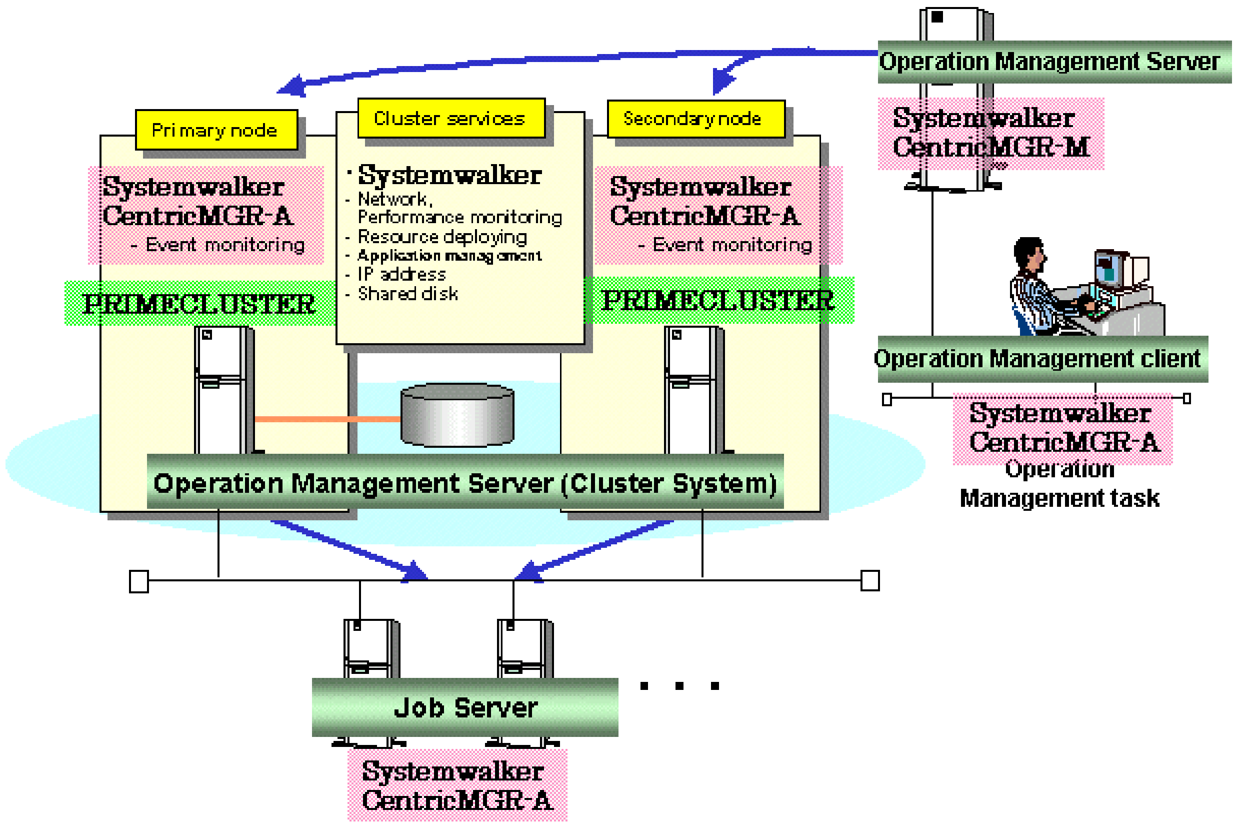
26.2.1 System Configuration

Systemwalker Centric Manager enables a user to use an operation management server, section management server, or job server in a cluster system. The system configuration of using each type of server is illustrated below:

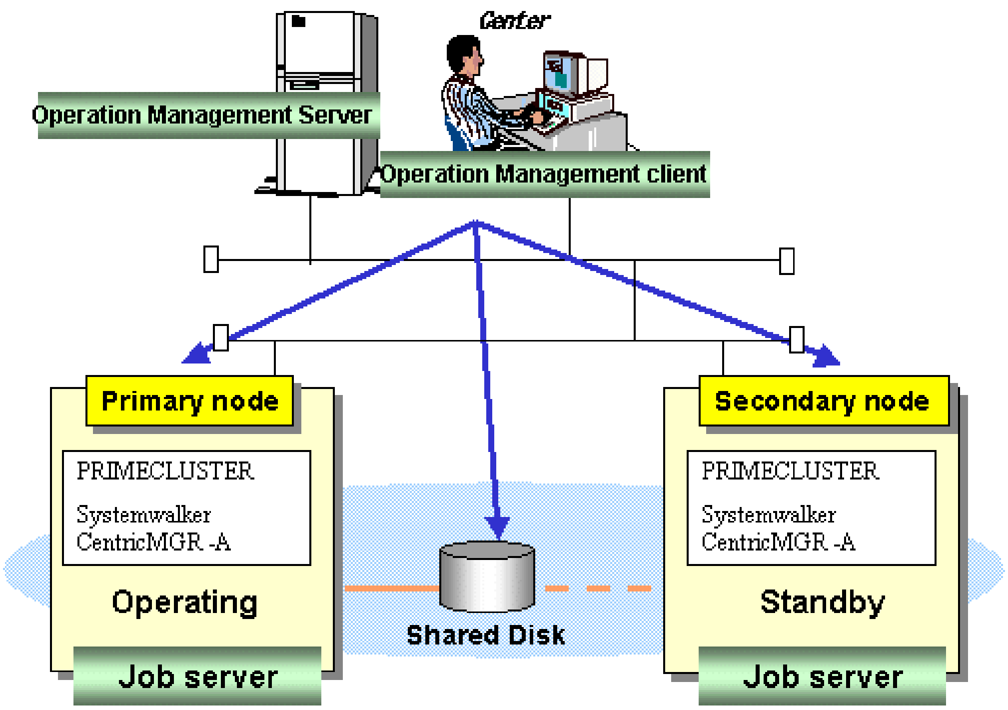
Operation Management Server



Section Management Server



Job server



26.2.2 Environment Setup

For details, see the *"Systemwalker Centric Manager Customization Guide for UNIX."*

Chapter 27 Systemwalker Operation Manager

27.1 Functional Overview

Systemwalker Operation Manager supports the following functions:

Automation of routine system operation and management

By installing Systemwalker Operation Manager, routine system operation and management can be automated as follows:

Automatic control of power on/off

Power to the server can be automatically turned on/off. Power on/off of the server is conducted according to a time schedule that is defined in advance. Also, power to the server can be automatically turned on when a client starts.

Note that this function is enabled only for Solaris servers that are running Systemwalker Operation Manager UNIX version.

Automatic activation of the services/applications

When the server is booted, services and applications are activated in the order in which they are registered. You can change the services and applications on any given day.

Note that this automatic activation of the services can be used only when the connected server is for the Windows version.

Automatic execution of a routine batch job

If you register the times and dates at which a routine batch job is run, the routine batch job is automatically executed according to that schedule. If an error occurs in a routine batch job, it can be corrected automatically without the intervention of the operator.

Scheduling file compression and transfer

The commands used for file compression and transfer are provided as job-associated commands. If you register these files with the job scheduler, you can execute these commands by scheduling.

Monitoring and operating a job on the screen

By installing Systemwalker Operation Manager, you can confirm the job execution status at a glance because the job status is displayed in different colors. You can monitor jobs running not only on a single system but also those running on multiple systems.

Moreover, you can operate a job while monitoring its status.

Control of job execution according to the system operation and status

Systemwalker Operation Manager enables you to control the job execution, as follows:

Control of job execution environment

Jobs can be executed efficiently by adjusting the job execution environments such as the execution order and the number of jobs that can be executed simultaneously.

Job execution with networked servers

Jobs can be executed on any server connected to the network.

Control of networked clients

You can provide clients that are connected over the network with controls such as power-on and file transfer.

27.1.1 Overview of Cluster System Support

Systemwalker Operation Manager provides high availability batch jobs and enables a high-reliability transaction system by running on a cluster system.

27.1.2 Supported Modes

Systemwalker Operation Manager supports the following operation modes:

- 1:1 standby
- N:1 standby
- Two-node mutual standby

For details, see the "Table 19.1 PRIMECLUSTER product list".

27.2 Standby Classes

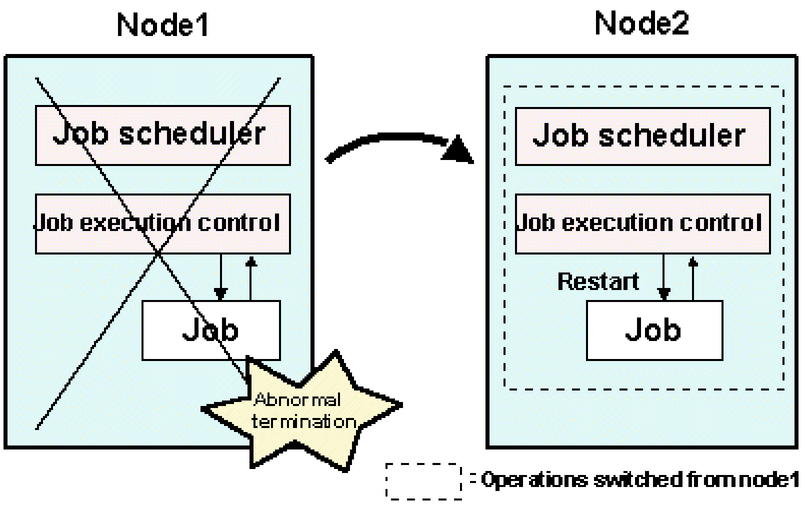
27.2.1 System Configuration

Systemwalker Operation Manager enables the entire servers, schedule server, execution server, and schedule server + execution server to operate on a cluster system in 1:1 standby, N:1 standby, or two-node mutual standby operation mode.

Each cluster system configuration when the entire Systemwalker Operation Manager servers are set up in a cluster configuration is outlined below:

1:1 standby

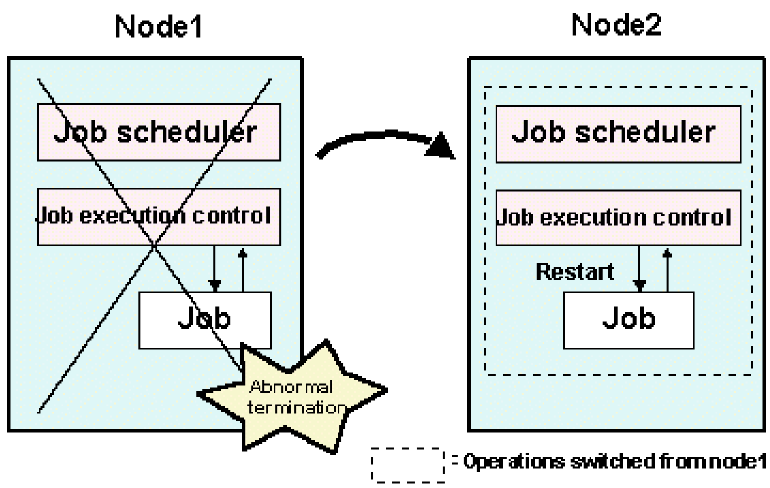
The following figure shows an example in which node 1 operates as an active and node 2 operates as a standby:



If an error occurs in node 1, ongoing operations are switched to node 2.

N:1 standby

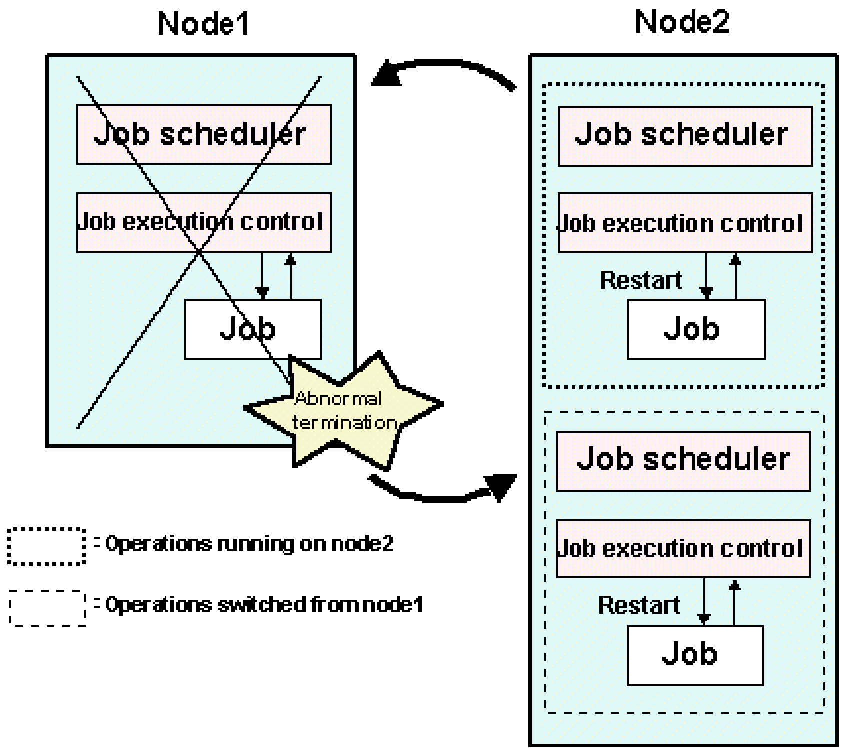
The following figure shows an example in which node 1 and node 3 are active and node 2 operates as a standby. The following shows an example where an error occurs on node 1. Even if an error occurs on node 3, ongoing operations are switched to node 2.



If an error occurs on node 1 or node 3, ongoing operations are switched to node 2. If errors occur on two or more nodes, the transaction running on the node on which the error occurred first is taken over by the standby node. However, the transactions of the nodes on which the errors occurred subsequently will not be taken over.

Two-node mutual standby

The following figure shows an example in which node 1 and node 2 execute their transactions while each operates as a standby for the other. The following shows an example where an error has occurred on node 1. If an error occurs on node 2, ongoing operations are switched to node 1.



For details, see the *"Systemwalker Operation Manager Installation Guide."*

27.2.2 Environment Setup

For details, see the *"Systemwalker Operation Manager Installation Guide."*

Chapter 28 Systemwalker Service Quality Coordinator

28.1 Functional Overview

Systemwalker Service Quality Coordinator is a software product designed to support overall system optimization by enabling the quality of the services provided by an IT system to be represented visually. It will enable the user to conduct the following types of operation management:

- Early detection of any deterioration in performance

  Systemwalker Service Quality Coordinator can collect and accumulate a range of information from the components of an IT system - such as information relating to response times, kernel resource performance, application server performance, and database performance - and use this data to display comparisons and trends from a variety of viewpoints. It can also monitor thresholds to identify signs of deteriorating performance.

- Reduced business recovery time

  Systemwalker Service Quality Coordinator continually collects information needed for troubleshooting. It also provides analytical functions for primary fault isolation. Information to be continually collected can also be provided in the form of templates containing items that have been carefully selected based on Fujitsu's expertise in system operation. This enables the user to automatically collect and accumulate the information needed for performance management, even without advanced performance-related skills.

- Improved availability

  The product allows the user to ascertain the service level of the entire system and the resource balance of individual servers. This makes it possible to:

  - Improve the operating efficiency of each server
  - Employ designs that optimize server resources
  - Employ designs that optimize system resources

  This can contribute to making the operation of the whole system more stable.

28.1.1 Overview of Cluster System Support

The management mechanism provided by a Manager and an Enterprise Manager can be operated in a cluster system. This improves the availability of management operations because if a problem develops on one node, operations can continue on the remaining normal node.

Cluster system operation and corresponding installation types

Enterprise Manager

An Enterprise Manager can be used in cluster configurations.

Manager

A Manager can be used in cluster configurations.

Proxy Manager

The servers making up a cluster system can be monitored individually.

Agent for Business

The servers making up a cluster system can be monitored individually.

Agent for Server

The servers making up a cluster system can be monitored individually.

Agent for Agentless Monitoring

The servers making up a cluster system can be monitored individually. To do so, they must be able to communicate with the physical IP address of each server.

Note

- When collecting the performance information of Managers or Enterprise Managers that make up a cluster system, performance information cannot be collected at the standby node.
- If an Agent for Server, Agent for Business or Proxy Manager is installed on servers making up a cluster system, each of the servers in the cluster system is monitored individually. For this reason, performance information cannot be collected from middleware and other software that is not running on the standby node.

28.1.2 Supported Modes

Systemwalker Service Quality Coordinator supports 1:1 standby by means of IP address takeover. For details, see "Table 19.1 PRIMECLUSTER product list."

28.1.3 Notes

For details on building the cluster system, refer to "3.7 PRIMECLUSTER Cluster System Operation Model" of the "Systemwalker Service Quality Coordinator Installation Guide."

Appendix A PRIMECLUSTER System Design Worksheets

A.1 Worksheet Purpose

The PRIMECLUSTER System Design Worksheets allow you to determine necessary parameters before you configure the PRIMECLUSTER system. Filling out all items in the worksheet will enable you to proceed smoothly with the PRIMECLUSTER installation work.

The worksheets consist of the following:

Overall Design Worksheet

This worksheet is used to design basic items for the PRIMECLUSTER system.

Cluster Installation Environment Setup Worksheet

Use this worksheet to configure the environment before you install Solaris, PRIMECLUSTER, and other required and related software. Also use the worksheet to configure the PRIMECLUSTER system using Web-Based Admin View.

Install each software product on each node using an installation script.

See

For details on each installation method, see the *"PRIMECLUSTER Installation Guide."*

This section identifies the worksheet for preparing and organizing necessary information for each installation method. This worksheet will enable a smooth installation.

Kernel Parameter Worksheet

This worksheet allows you to determine the fundamental PRIMECLUSTER values for estimating the kernel parameters. After installing the Solaris software and PRIMECLUSTER packages, configure the new values for the kernel parameters, as necessary.

Cluster Configuration Worksheet

This worksheet organizes the necessary items for configuring the PRIMECLUSTER system with Web-Based Admin View. Using this worksheet and determining the parameter values beforehand will enable you to smoothly set up operations using Web-Based Admin View.

Cluster Application Configuration Worksheets

This worksheet is used to design how the application should operate when building a cluster application.

Cluster Application Setting Worksheet

This worksheet is used for configuring cluster applications and resources.

A.2 Notes on Worksheet Creation

Note the following points when creating PRIMECLUSTER design worksheets.

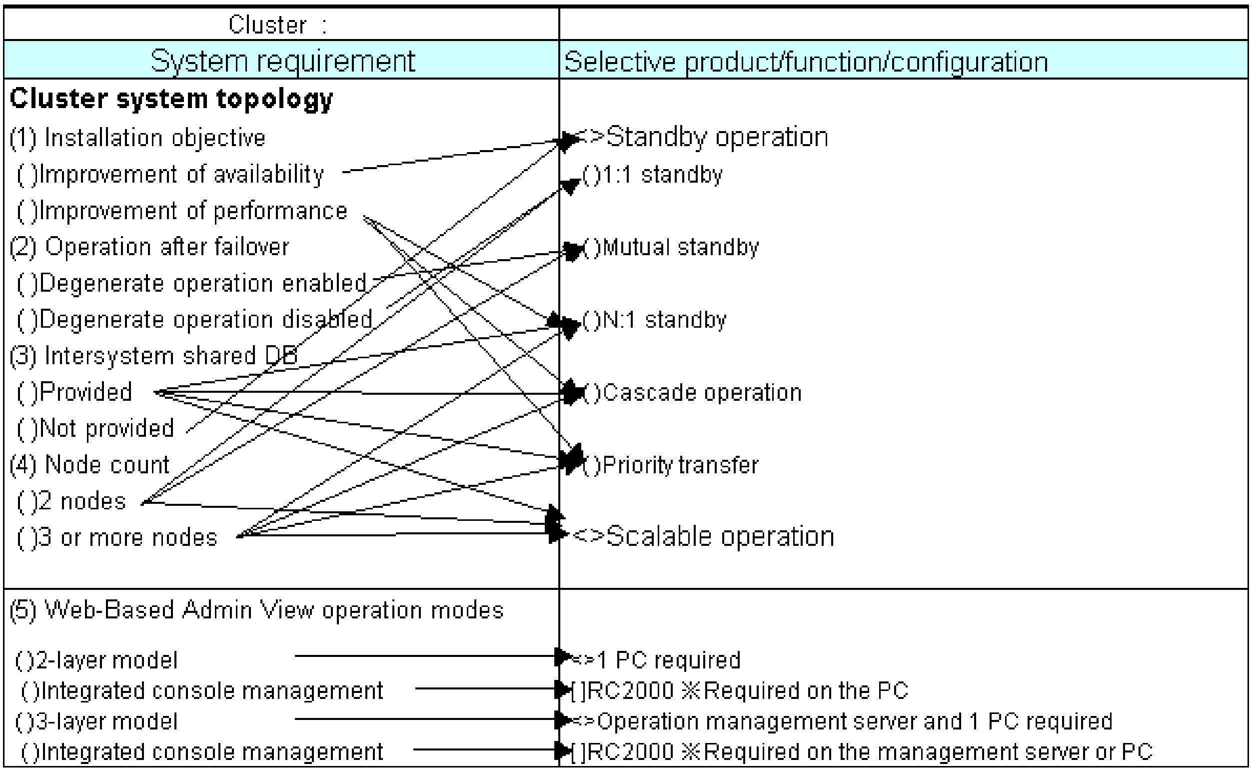
- Specify appropriate values in the blank sections of each design sheet. For instructions on specifying each item, see the point explanations.
- Items marked with "<>"are selection items. Select one of the items written in the box.
- Items marked with "[ ]" are selection items. You can select multiple items written in the box.
- Different values can be specified depending on the model in which PRIMECLUSTER is to be installed. Read the point explanations before specifying values.

A.3 Overall Design Worksheet

Select the operation mode of the cluster system.

See

See "A.10 System Design Examples" when determining values for the setup items in the Overall Design Worksheet.



(6) Select the shared disk units	
1) 2-node configuration	
<>RAID	[ ]Multipath disk control facility
<>Interface	<>MPHD <>ETERNUS multipath driver
<>SCSI	Note: MPHD is supported only by GR710 and higher models
<>Fibre Channel	[ ]Multipath load balance (MPLB)
<>Path	<>MPLB <>ETERNUS multipath driver
<>Single path	[ ]Fujitsu PCI Fibre Channel (PFCA)
<>Path duplex switchover	(hardware attachment)
<>Path duplex load balancing	[ ]GDS (Global Disk Services)
<>Path triplex or higher	
<>Use FDU/FU	
<>Manage disk cabinet	
2) 3 or more nodes (*1)	[ ]Multipath disk control facility
<>RAID	<>MPHD <>ETERNUS multipath driver
<>Interface	Note: MPHD is supported only by GR710 and higher models
<>Fibre Channel	[ ]Multipath load balance
<>Path	<>MPLB <>ETERNUS multipath driver
<>Single path	[ ]Fujitsu PCI Fibre Channel (PFCA)
<>Path duplex switchover	(hardware attachment)
<>Path duplex load balancing	[ ]GDS (Global Disk Services)
<>Path triplex or higher	
<>Manage disk partitions	
3) Switchover file system selection	
<>zfs	
<>ufs	
<>High reliability above ufs	[ ]GFS (Global File Services)
<>Simultaneous sharing	
<>File system	
<>RAW	

(7)NTP		
1)Cluster topology	<>NTP broadcast	
( )No existing NTP network.	<>Place NTP server on the cluster node	
( )Existing NTP network.	<>Synchronous with the NTP server outside the cluster	
2)Protocol (only when broadcast is specified)		
( )Node name takeover	<>multicastclient	
( )No node name takeover	<>broadcast	
3)NTP network setting	Host name	IP address
( )When there is an NTP server outside the cluster system		
( )NTP setting within the cluster system		
Primary NTP server		
Secondary NTP server		

In this chart, ISV products (Veritas, Emulex, EMC PowerPath) should be mentioned.

A.4 Cluster Installation Environment Worksheet

The Cluster Installation Environment Worksheet organizes items related to installation and initial configuration of Solaris software, PRIMECLUSTER, required and related software. Create the Cluster Installation Environment Worksheet and the Solaris Installation Worksheet.

The Cluster Installation Environment Worksheet contains initial configuration items related to the entire PRIMECLUSTER system. Configure a system on each node according to the information specified here.

Cluster system name:		**Cluster Installation Environment Worksheet**
Node name:		

Settings		Notes			
(1) Software to be installed	[ ]Optional software	[ ]RAID Disk Driver (HDDV): Depends on hardware configuration [ ]Fibre Channel driver: Depends on hardware configuration [ ]Multi-path disk control (MPHD) [ ]Multipath load balance (MPLB) [ ]ETERNUS multipath driver			
(2) Version of Solaris	<>10 <>11	Use the same version on all nodes.			
(3) Partition information		Logical path	Size		
	/			Configuration should besynchronized on all nodes	
	swap				
	/usr				
	/var				
	/opt				
	/export/home				
(4) System locale	<> C <> Ja <> Other ( )				
(5) Time zone	<> Japan <> Other ( )				
(6) Node information Node name IP MAC					
(7)Web-Based Admin View					
	Node name	IP	Access I/F	Http port number	RMI port number
Primary management server Secondary management server	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________

Note

Use a single primary management server of Web-Based Admin View to monitor a single cluster system.

Solaris Installation Worksheet

Specify information needed for installing the Solaris software in nodes. Refer to the worksheet created here when installing various software programs to each node individually.

In a cluster system, the items other than the host names and IP addresses of the individual nodes have the same values in all nodes.

See

For details of each item, see the "Solaris X Information Library."

Setting item	Setting value	Remarks
Select a Language	0) English	This item is displayed only for the English version of Solaris. If the environment is not a Japanese language environment, select 0 (English).
Select a Locale	0) USA - English (ASCII only)	This item is displayed only for the English version of Solaris. If the environment is not a Japanese language environment, select 0 (English).
What type of terminal are you using?		Set this item based on the terminal emulator. Select vt52 for using Windows telnet.
Host Name		Specify the node name of each node recorded on the initial setup worksheet.
Network Connectivity	Yes	Specify Yes.
Primary Network Interface	____________________	Specify the network interface connected to the public LAN.
IP address		Specify the node name of each node recorded on the initial setup worksheet.
Name Service	None	Specify None.
Subnets	Yes/No	Specify Yes if the public LAN is part of the subnet.
Netmask	____.____.____.____	If Yes is selected for Subnets, set the net mask of the subnet.
Time Zone		Specify the default time zone.
Data and Time		The initial setup tool matches the date and time of other nodes to the date and time set for the installation server. Set the correct date and time.
Solaris Interactive Installation	Initial	Select Initial.
Allocation Client Services?	Continue	Select Continue.
Select Languages	ja,________________	Select the required language. For a Japanese language environment, you also need to select ja.
Select Software	Entire Distribution	Select Entire Distribution.
Select Disks		Select the disk on which Solaris is installed.
Preserve Data?	Continue	Select Continue.
Automatically Layout File Systems?	Manual Layout	Select Manual Layout. Set so that the disk slice configuration of all nodes is the same.
Customize Disk		Select Manual Layout, so the disk slice configuration is synchronized on all nodes.
Mount Remote File System?	Continue	Select Continue.
Reboot After Installation?	Auto Reboot	Select Auto Reboot to automatically restart after installation is done.

A.5 Kernel Parameter Worksheet

When you build the PRIMECLUSTER system, the default values for the kernel parameters of the Solaris may be inadequate. The Kernel Parameter Worksheet shows the kernel parameter values used for PRIMECLUSTER.

Set up the kernel parameter as follows according to the type of "Characteristics" included in the table.

- Addition

  Specify the total of the system default value and the recommended or specified values for each software product.

- Maximum value

  Specify a maximum for the recommended or specified values for each software product.

  You must, however, use the default value if the maximum value is smaller than the system default value.

The kernel parameter values differ depending upon:

- CF configuration
- RMS configuration
- Using GFS

See

- For details of the kernel parameters and instructions on changing parameter values, see "Solaris System Administration" in *"Solaris X System Administrator Collection Vol. 1."*
- For information on the default values of the Solaris, see the *"Solaris Tunable Parameters Reference Manual."*

Note

- The values used by products and user applications that operate in the PRIMECLUSTER system must also be included in the kernel parameter values.
- To enable values that have been changed, you must restart the system after the installation.
- If a kernel parameter value is already maximized, the change will not be added to the system.

A.5.1 CF Configuration

When the resource database is used

The table below shows the kernel parameter values that are required in the CF configuration when the resource database is used.

The kernel parameter values in /etc/system are automatically changed by the installer. Be sure to return the settings in /etc/system to their original state when you uninstall the package.

Note

For system expansion, if you increase the number of nodes and logical disks, you need to re-estimate the resources and restart each node in the cluster system. If you want to add nodes or logical disks to a cluster system after it is configured, it is necessary to set a kernel parameter in advance considering the number of the nodes and logical disks.

Kernel parameter	Characteristics	Value	Parameter description
semsys:seminfo_semmni	Addition	20	Maximum number of semaphore identifiers.
shmsys:shminfo_shmmax	Maximum value	4194304 *	Maximum size of the System V shared memory segment that can be created.
shmsys:shminfo_shmmni	Addition	30	Maximum number of the shared memory segments that can be created for the entire system.

*

4194304 is the minimum value that is required when the resource database is used.

If you set the smaller value than the default value (8388608) to shmsys:shminfo_shmmax, do not set the smaller value than the minimum value.

Also, You may need to change the value of shmsys:shminfo_shmmax depending on the number of cluster system resources.

Estimate the number of cluster system resources according to the following equation and change the value:

Number of resources = (a) + (b)

(a) Number of disks in shared system devices x (number of shared nodes + 1) x 2

(b) Total number of local disks (number of local disks in all cluster configuration nodes)

Value required for resource database = 1048576 + 2776 x number of resources

- If the value calculated above is larger than the installation default value (8388608):

  shmsys:shminfo_shmmax = Value required for resource database

- If the value calculated above is smaller than the installation default value (8388608):

  You do not need edit shmsys:shminfo_shmmax .

  (The installation default value is used.)

RCI monitoring agent setup

When you set up asynchronous RCI monitoring, you must specify the timeout interval (kernel parameter) in /etc/system for monitoring via SCF/RCI. Kernel parameters vary depending on the server type. Then check your server type so you can set the appropriate timeout interval.

Note

This setting is not required in the following cases:

- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 provided by companies other than Fujitsu in Japan
- SPARC Enterprise M3000, M4000, M5000, M8000, and M9000 with logos of both Fujitsu and Oracle provided in other than Japan

Below table shows the server types that require setting of the monitoring timeout interval.

Server type	Model	Kernel parameter (driver name: scf_rdctrl_sense_wait)
SPARC Enterprise	M3000	scfd:scf_rdctrl_sense_wait
	M4000	
	M5000	
	M8000	
	M9000	
PRIMEPOWER	800	FJSVscf2:scf_rdctrl_sense_wait
	1000	
	2000	
	900	FJSVscf3:scf_rdctrl_sense_wait
	1500	
	2500	

a. Method for Calculating Monitoring Timeout Intervals

   Calculate monitoring timeout intervals as follows:

   - Up to 2 domains: 2 seconds
   - 3 or more domains: 1 second + (0.5 x number of domains)

```
Example:
 - 3 domains: 2.5 seconds
 - 4 domains: 3.0 seconds
```

Note

Calculate timeout intervals based on the number of domains in the server that contains the largest number of domains in the RCI network.

b. Method for Setting Timeout Intervals in /etc/system

Before setting up the initial cluster configuration, modify /etc/system for all nodes according to below procedure.

1. Make a backup of /etc/system.

   Example:
   Make a copy of /etc/system and save it under the filename /etc/system.org.

   ```
   # cp /etc/system /etc/system.org
   ```

2. Set the monitoring timeout interval in /etc/system.

   As the monitoring timeout interval is specified in microseconds, you have to multiply the seconds calculated in above item "a." by 1,000,000 for this setting.

   ```
   set driver name: scf_rdctrl_sense_wait = monitoring timeout
   interval [microseconds]
   ```

   Example:
   Setting a 2-second monitoring timeout interval for a SPARC Enterprise server with 2 domains.

   ```
   set scfd:scf_rdctrl_sense_wait = 2000000
   ```

3. Restart the node.

   Example:

   ```
   # /usr/sbin/shutdown -y -g0 -i6
   ```

A.5.2 RMS Configuration

In order to ensure that RMS runs normally, the following kernel parameters need to be set. Therefore, when RMS is installed, the definitions of the parameters in /etc/system are automatically updated if not defined or defined with smaller value than the following "Value".

Kernel parameter	Characteristics	Value	Parameter description
msgsys:msginfo_msgmnb	Maximum value	4194304	Maximum size of the message that can be stored in a single message.
msgsys:msginfo_msgmni	Addition	8192	Maximum number of message queue identifiers that can be used for the entire system.
msgsys:msginfo_msgtql	Maximum value	65535	Maximum number of message headers

Note

- In PRIMECLUSTER, message queues are used for interprocess communication.

  When RMS is running, 2076 message queues are reserved from 0x4d2.

If you are using message queues for any applications, use the range other than the above (0x4d2 to 0xcee).

- Even if definitions of the kernel parameters in /etc/system are automatically added/updated, change the value as necessary in consideration of the value required by other software and user applications.

A.5.3 Using GFS

The kernel parameters required to enable the use of the GFS shared file system are shown below:

Kernel parameter	Characteristics	Value	Parameter description
semsys:seminfo_semmni	Addition	2	Maximum number of semaphore identifiers

A.6 Cluster Configuration Worksheet

The Cluster Configuration Worksheet is required for building a cluster.

Use this worksheet to organize the items before using Web-Based Admin View to set up the PRIMECLUSTER system.

See

For information on the setup items for this worksheet, see "Chapter 5 Building a Cluster." Also see the "A.10.1 Cluster Configuration Worksheet" in the "A.10 System Design Examples."

	Item			Setting	Remarks
Shared disk unit: Type 1	Machine model				
	Device name				
	IP address				
	Subnet mask				
	Interface				
	LUN number				
	Ports				
	Connection paths				
	FC-SW		Use of fibre channel switching hub	☐ yes ☐ no	
			Machine model		
			Device name		
			IP address		
			Subnet mask		
	File system selection (type of file system used in cluster)				
Shared disk unit: Type 2	Machine model				
	Device name				
	IP address				
	Subnet mask				
	Interface				
	LUN number				
	Ports				
	Connection paths				
	FC-SW		Use of fibre channel switching hub	☐ yes ☐ no	
			Machine model		
			Device name		
			IP address		
			Subnet mask		
	File system selection (type of file system used in cluster)				
NTP	Operation mode			☐ NTP server (integrated in cluster) ☐ NTP client	
	Protocol (only when broadcast is specified)				
	Server	Primary	Host name		
			IP address		
			Subnet mask		
		Secondary	Host name		
			IP address		
			Subnet mask		
Web-Based Admin View (Management view)	Operation mode			☐ 1 ☐ 2 ☐ 3 layers	
	Server	Primary	Host name		
			IP address		
			Subnet mask		
			URL		
		Secondary	Host name		
			IP address		
			Subnet mask		
			URL		
	User ID		wvroot group		
			clroot group		
			cladmin group		
			clmon group		
			sdxroot group		

Web-Based Admin View (Management view) For guest domain	Operation mode			☐ 1 ☐ 2 ☐ 3 layers	
	Server	Primary	Host name		
			IP address		
			Subnet mask		
			URL		
		Secondary	Host name		
			IP address		
			Subnet mask		
			URL		
	User ID		wvroot group		
			clroot group		
			cladmin group		
			clmon group		
			sdxroot group		
Cluster Foundation (CF)	Cluster name				
	Cluster nodes				
	Cluster node names				
	Subnets				
	Subnet number				
	Subnet mask				
	Use of CF remote services			☐ cfcp (CF file copy) ☐ cfsh (CF remote command execution)	
	IP interconnect			◇ [no] ◇ yes (the number of IP interconnects)	[] default In a Solaris 11 environment, this function is not available.
	Use by RMS			◇ [yes] ◇ no	[] default
Shutdown facility (SF)	Shutdown agent			☐ RCI ☐ XSCF (SPARC M series) ☐ XSCF (SPARC Enterprise M series) ☐ ILOM ☐ ALOM	
Cluster resource management facility (CRM)	Hardware devices stored in resource database			☐ Network unit	Required for using takeover of public LAN (IP) and node names
				☐ Line switching unit	Required for using line takeover function
				☐ Disk unit	Required when using GDS and GFS

Node 1	Machine model			
	Device name			
	Domain ID / domain name			
	Node name (uname -n)			
	CF node name			
	Interconnect	IP address		
		Subnet mask		
		Device name		
	Node in CF quorum set		◆ [yes] ◇ no	Do not change the default setting.
	CF cluster timeout (CLUSTER_TIMEOUT)			Fixed value in the OVM environment. Do not change the default setting.
	Use by RMS		◇ [yes] ◇ no Suffix:	[] Default
	RCI	Address		Only when using RCI asynchronous monitoring
	Kernel parameter	scfd:scf_rdctrl_sense_wait		
	System console	IP address 1		XSCF, ALOM, and ILOM data (The setting of 2 is for SPARC M series)
		Subnet mask 1		
		IP address 2		
		Subnet mask 2		
		User ID		
		User authority		
		Connection type	■ SSH □ telnet	
	Shutdown facility (SF)	Administrative IP addresses		
		Weight (Node weight)		Node weight (priority if cluster partition occurs) used by SF
		User ID		XSCF, ALOM, and ILOM data
		User authority		
		Console LAN#1		
		Console LAN#2		
		PPAR-ID		SPARC M series
		Domain name		
		Connection type	◆ [SSH] ◇ telnet	[] Default For XSCF
	Use of shared disk		■ yes □ no	
	scsi-initiator-id (OBP variable #eeprom command)			
	SCSI / fibre channel	Circuits (paths)		
		Device name 1		
		sd device name		
		Device name 2		
		sd device name		
	Management LAN	IP address		
		Subnet mask		
		Device name		
	Asynchronous monitoring sub-LAN	IP address		SPARC M series
		Subnet mask		
		Device name		
	Public LAN 1	IP address		
		Subnet mask		
		Device name		
	Public LAN 2	IP address		
		Subnet mask		
		Device name		

Node 2	Machine model			
	Device name			
	Domain ID / domain name			
	Node name (uname -n)			
	CF node name			
	Interconnect	IP address		
		Subnet mask		
		Device name		
	Node in CF quorum set		◆ [yes] ◇ no	Do not change the default setting.
	CF cluster timeout (CLUSTER_TIMEOUT)			Fixed value in the OVM environment. Do not change the default setting.
	Use by RMS		◇ [yes] ◇ no Suffix:	[] Default
	RCI	Address		Only when using RCI asynchronous monitoring
	Kernel parameter	scfd:scf_rdctrl_sense_wait		
	System console	IP address 1		XSCF, ALOM, and ILOM data (The setting of 2 is for SPARC M series)
		Subnet mask 1		
		IP address 2		
		Subnet mask 2		
		User ID		
		User authority		
		Connection type	■ SSH □ telnet	
	Shutdown facility (SF)	Administrative IP addresses		
		Weight (Node weight)		Node weight (priority if cluster partition occurs) used by SF
		User ID		XSCF, ALOM, and ILOM data
		User authority		
		Console LAN#1		
		Console LAN#2		
		PPAR-ID		SPARC M series
		Domain name		
		Connection type	◆ [SSH] ◇ telnet	[] Default For XSCF
	Use of shared disk		■ yes □ no	
	scsi-initiator-id (OBP variable #eeprom command)			
	SCSI / fibre channel	Circuits (paths)		
		Device name 1		
		sd device name		
		Device name 2		
		sd device name		
	Management LAN	IP address		
		Subnet mask		
		Device name		
	Asynchronous monitoring sub-LAN	IP address		SPARC M series
		Subnet mask		
		Device name		
	Public LAN 1	IP address		
		Subnet mask		
		Device name		
	Public LAN 2	IP address		
		Subnet mask		
		Device name		

See

For information on the "setup policy for survival priority if cluster partition occurs" in the Remarks column for weight, see "Survival scenarios" in "5.1.2.2.2 Using the Shutdown Configuration Wizard."

A.7 Cluster Application Configuration Worksheets

Use the Cluster Application Configuration Worksheet to organize the items before building a cluster application.
The following worksheets are required for building a cluster application.

- A.7.1 RMS Setup Worksheet
- A.7.2 GLS Setup Worksheet
- A.8 GDS Setup Worksheets
- A.9 Cluster Application Worksheets

See

For information on the setup items for this worksheet, see "Chapter 6 Building Cluster Applications."

A.7.1 RMS Setup Worksheet

When using RMS, use this worksheet to organize the setting items beforehand.

See

For information on the setup items in this worksheet, see "6.1 Initial RMS Setup." and "6.8 Setting Up the RMS Environment."

Item	Environment variable (global)	Setup value	Remarks
RMS automatic startup	HV_RCSTART	<>No <>[Yes]	(Specified value of CLI) 0 [1]
RMS shutdown wait time	RELIANT_SHUT_MIN_WAIT	(Range: 0 to 2147483647) [2147483647]	

Note: The value in brackets is the default.

[Setup items]

RMS shutdown wait time (RELIANT_SHUT_MIN_WAIT)

When using RMS, you need to change the environment variable to the value corresponding to the configuration setup.

The default value is 2147483647 (seconds) in this version.

If RMS shutdown request ("hvshut" command) is executed, RMS performs shutdown processing after it performs offline processing of the active cluster applications. Add the maximum time required to complete offline processing of cluster applications and the maximum time required to shut down RMS (30 seconds). Then specify the added total time for this value.

See

For details on RELIANT_SHUT_MIN_WAIT, see "RELIANT_SHUT_MIN_WAIT" in "13.2 Global environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

A.7.2 GLS Setup Worksheet

If you will be using Global Link Services (GLS), use the GLS Setup Worksheet to organize the input items beforehand.

This section describes the worksheet settings for the following typical duplexing modes:

- NIC switching mode (logical IP takeover)
- Fast switching mode
- GS/SURE linkage mode

See

- For configuration instructions for other duplexing modes, see "Appendix B Examples of Setting Up" in the *"PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function"* and "Chapter 3 Configuration" in the *"PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function."*
- For information on the items in this sheet, see the explanation at the bottom of the sheet and "6.2 Initial GLS Setup." Also see "A.10.2 GLS Setup Worksheet" in "A.10 System Design Examples."

NIC Switching Mode (Logical IP Takeover)

Item				Setting
GLS settings	Switching mode			<Y>NIC switching mode (logical IP takeover) < >Fast switching mode
	Takeover virtual interface name			
		Takeover virtual IP address (or host name)		
		Subnet mask		
	Node name (1)			
		Configuration information	Virtual interface name	
			Primary physical interface name	
			Secondary physical interface name	
			Physical IP address (or host name)	
			Logical IP address (or host name)	
		Monitoring destination information	IP address (or host name) of primary monitoring destination	
			IP address (or host name) of secondary monitoring destination	
		Optional function	Standby NIC patrol	<>Disable <>Enable
	Node name (2)			
		Configuration information	Virtual interface name	
			Primary physical interface name	
			Secondary physical interface name	
			Physical IP address (or host name)	
			Logical IP address (or host name)	
		Monitoring destination information	IP address (or host name) of primary monitoring destination	
			IP address (or host name) of secondary monitoring destination	
		Optional function	Standby NIC patrol	<>Disable <>Enable

[Setup items]

- Switching mode

  Specify a redundant operation type for GLS; Fast switching mode, NIC (Network Interface Card) switching mode, or RIP mode. Add d (logical IP takeover of the NIC switching mode) to the "-m" option of the "hanetconfig create" command.

- Takeover virtual interface name

  Specify the name of the virtual interface for cluster switchover.
  The virtual interface is automatically created under the name of shaX:NN (X and NN are numbers) as a result of executing the "hanethvrsc create" command.

- Takeover virtual IP address (or host name)

  Specify the IP address (or host name) for cluster switchover.
  The same logical IP address (or host name) that was set in nodes (1) and (2) is set as a result of executing the "hanethvrsc create" command.

- Subnet mask

  Specify the subnet mask value corresponding to the takeover virtual IP address in the "/etc/inet/netmasks" file.
  This value is used for a physical IP and logical IP.
  If subnet is not used, you don't need to specify the value.

- Virtual interface name

  Specify the virtual interface name to be assigned to the physical interface to be duplexed.

  Specify shaX (X is a number) to the "-n" option of the "hanetconfig create" command.
  Set the same name for node (1) and node (2).

- Primary physical interface name

  Specify the name of the interface to be first activated among the duplexed physical interfaces (e.g. hme1, qfe1, etc.)
  Specify the name to the "-t" option of the "hanetconfig create" command.

- Secondary physical interface name

  Among the duplexed physical interfaces, specify the name of the interface to be activated after switchover from the primary physical interface. Specify the name to the "-t" option of the "hanetconfig create" command.

- Physical IP address (or host name)

  Specify the physical IP address (or host name) to be assigned to the physical interface to be duplexed.
  Specify the value to the "-e" option of the "hanetconfig create" command.
  Specify different values for node (1) and node (2).

- Logical IP address (or host name)

  Specify the logical IP address (or host name) to be assigned to the physical interface to be duplexed.
  Specify the value to the "-i" option of the "hanetconfig create" command. Specify the same value for node (1) and node (2).

- Primary monitoring-destination IP address (or host name)

  Specify the IP address (or host name) of the switching hub to be monitored when the primary physical interface is used.
  Specify the value to the "-p" option of the "hanetpoll create" command.

- Secondary monitoring-destination IP address (or host name)

  Specify the IP address (or host name) of the switching hub to be monitored when the secondary physical interface is used.
  Specify the value to the "-p" option of the "hanetpoll create" command.

- Standby NIC patrol

  When monitoring the standby NIC, specify p for standby patrol (automatic switchback in the event of a failure) or q for standby patrol (immediate automatic switchback) to the -m option of the "hanetconfig create" command.

Fast Switching Mode

Item					Setting
GLS settings	Switching mode				◇NIC switching mode (logical IP takeover) ◆Fast switching mode
	Takeover virtual interface name				
		Takeover virtual IP address (or host name)			
		Subnet mask			
	Node name (1)				
		Configuration information	Virtual interface name		
			Physical interface name (1)		
			Physical interface name (2)		
				Virtual IP address (or host name)	
				Subnet mask	
				Physical IP address (or host name) (1)	
				Subnet mask	
				Physical IP address (or host name) (2)	
				Subnet mask	
	Node name (2)				
		Configuration information	Virtual interface name		
			Physical interface name (1)		
			Physical interface name (2)		
				Virtual IP address (or host name)	
				Subnet mask	
				Physical IP address (or host name) (1)	
				Subnet mask	
				Physical IP address (or host name) (2)	
				Subnet mask	

[Setup items]

- Switching mode

  Specify a redundant operation type for GLS; Fast switching mode, NIC (Network Interface Card) switching mode, or RIP mode.
  Add t (fast switching mode) to the "-m" option of the "hanetconfig create" command.

- Takeover virtual interface name

  Specify the name of the virtual interface for cluster switchover.
  The virtual interface is automatically created under the name of shaX:NN (X and NN are numbers) as a result of executing the "hanethvrsc create" command.

- Takeover virtual IP address (or host name)

  Specify the IP address (or host name) for cluster switchover.
  Specify the value to the "-i" option of the "hanethvrsc create" command.

- Subnet mask

  Specify the subnet mask value corresponding to the takeover virtual IP address in the "/etc/inet/netmasks" file.
  This value should be the same as the virtual IP for the subnet mask.
  If subnet is not used, you don't need to specify the value.

- Virtual interface name

  Specify the virtual interface name to be assigned to the physical interface to be duplexed.
  Specify shaX (X is a number) to the "-n" option of the "hanetconfig create" command.
  Set the same name for node (1) and node (2).

- Physical interface name (1) and (2)

  Specify the name of the interface to be first activated among the duplexed physical interfaces (e.g. hme1, qfe1, etc.)
  Specify the name to the "-t" option of the "hanetconfig create" command.

- Virtual IP address (or host name)

  Specify the virtual IP address (or host name) to the "-i" option of the "hanetconfig create" command.
  Specify different values for node (1) and node (2).

- Subnet mask

  Specify the subnet mask value corresponding to the takeover virtual IP address in the "/etc/inet/netmasks" file.
  This value should be the same as the virtual IP for the subnet mask.
  If subnet is not used, you don't need to specify the value.

- Physical IP address (or host name) (1) and (2)

  Specify the physical interface name to be assigned to the physical interface to be duplexed.
  Create the "/etc/hostname. physical interface name" file and specify the IP address (or host name) in the file.
  This value should be different than other IP.

- Subnet mask

  Specify the subnet mask value corresponding to the physical IP address in the "/etc/inet/netmasks" file.
  If subnet is not used, you don't need to specify the value.

GS/SURE linkage mode

Item					Setting
GLS settings	Switching mode				<Y>GS/SURE linkage mode
	Takeover virtual interface name				
		Takeover virtual IP address (or host name)			
		Subnet mask			
	Node name (1)				
		Configuration information	Virtual interface name		
			Physical interface name (1)		
			Physical interface name (2)		
				IP address (or host name)(1)	
				Subnet mask	
				IP address (or host name)(2)	
				Subnet mask	
	Node name (2)				
		Configuration information	Virtual interface name		
			Physical interface name (1)		
			Physical interface name (2)		
				IP address (or host name)(1)	
				Subnet mask	
				IP address (or host name)(2)	
				Subnet mask	
	Remote host name				
		Configuration information	Virtual IP address		
				IP address (1)	
				IP address (2)	
		Monitoring destination information	Remote virtual IP address monitoring		<>Enable <>Disable
		Optional function	RIP packet is sent from remote host		<>Wait <>Not Wait
			TCP relay		Network information of relay destinationNetwork IP address or host name ( )

[Setup items]

- Switching mode

  Specify a redundant operation type for GLS; GS/SURE linkage mode.
  Add n (Physical interface setting) and c (Virtual interface setting) to the "-m" option of the "hanetconfig create" command.

- Takeover virtual interface name

  Specify the name of the virtual interface for cluster switchover.
  The virtual interface is automatically created under the name of shaX:NN (X and NN are numbers) as a result of executing the "hanethvrsc create" command.

- Takeover virtual IP address (or host name)

  Specify the IP address (or host name) for cluster switchover.
  The same logical IP address (or host name) that was set in nodes (1) and (2) is set as a result of executing the "hanethvrsc create" command.

- Subnet mask

  Specify the subnet mask value corresponding to the takeover virtual IP address in the "/etc/inet/netmasks" file.
  This value is used for a physical IP and logical IP.
  If subnet is not used, you don't need to specify the value.

- Virtual interface name

  Specify the virtual interface name to be assigned to the physical interface to be duplexed.
  Specify shaX (X is a number) to the "-n" option of the "hanetconfig create" command.
  Set the same name for node (1) and node (2).

- Primary physical interface name (1) and (2)

  Specify the name of the interface to be first activated among the duplexed physical interfaces (e.g. hme1, qfe1, etc.)
  Specify the interface name followed by "-t" option along with executing "hanetconfig create" command with "-m" option and parameter "n" to set the physical interface.

- IP address (or host name) (1) and (2)

  Specify the interface name followed by "-i" option along with executing "hanetconfig create" command with "-m" option and parameter "n" to set the physical interface.

- Subnet mask (1) and (2)

  Specify the subnet mask value corresponding to the physical IP address in the "/etc/inet/netmasks" file.
  If subnet is not used, you don't need to specify the value.

- Remote host name

  Specify the remote host name.
  Specify the remote host name to the "-n" option of the "hanetobserv create" command. (e.g. GS-1, SURE-1, etc.)

- Virtual IP address

  Specify the virtual IP address of the remote host.
  Specify the virtual IP address to the "-i" option of the "hanetobserv create" command.

- IP address (1) and (2)

  Specify the physical IP address of the remote host.
  Specify the physical IP address to the "-t" option of the "hanetobserv create" command.

- Remote virtual IP address monitoring

  When monitoring the remote virtual IP address, specify "on" to the "-m" option of the "hanetobserv create" command.
  When not monitoring the remote virtual IP address, specify "off" to the "-m" option of the "hanetobserv create" command.

- RIP packet is sent from remote host

  When sending out a report regarding node switch against the remote host, specify whether or not to wait until receiving RIP from the remote host. Specify the parameter "on" using hanetobserv create command with "-r" option to hold off for the report, or "off" for not holding off.

- TCP relay

  Specify the network information of relaying system for TCP relay. Specify client IP address, network address, and subnet mask using hanetobserv create command with "-c" option. (e.g. -c 192.168.72.1,192.168.73.0:255.255.255.0)

A.8 GDS Setup Worksheets

These worksheets serve to organize necessary input items beforehand for setting the GDS (Global Disk Services) configuration.

There are two types of GDS Setup Worksheet, "System Disk Mirror Setup" for setting up system disk mirroring and "GDS Configuration" for setting up shared disks.

- System Disk Mirror Setup Worksheet
- GDS Configuration Worksheet

See

- Before setting up the GDS configuration, you need to know basic information from reading *"PRIMECLUSTER Global Disk Services Configuration and Administration Guide."*
- For information on the items in this worksheet, see "6.3.2 GDS Configuration Setup."

Note

The class name should be unique in the cluster system.

A.8.1 System Disk Mirror Setup Worksheet

Use the System Disk Mirror Setup Worksheet to organize items before mirroring the system disk.

See

- Before setting up the GDS configuration, you need to know basic information from reading *"PRIMECLUSTER Global Disk Services Configuration and Administration Guide."*
- For information on the items in this worksheet, see "6.3.2.1 Setting Up System Disk Mirroring in the ZFS Boot Environment." or "6.3.2.2 Setting Up System Disk Mirroring in the UFS Boot Environment." Also see "A.10.3 System Disk Mirror Setup Worksheet" in "A.10 System Design Examples."

System Disk Mirror Setup Worksheet

Item					Setting
System disk mirror settings	Node name				
		Class name			
			Group name		
				Mount point (for a UFS boot environment) or a ZFS storage pool name (for a ZFS boot environment)	
				Physical disk name	
				Mirror disk name	
				Spare disk name	
	Node name				
		Class name			
			Group name		
				Mount point (for a UFS boot environment) or a ZFS storage pool name (for a ZFS boot environment)	
				Physical disk name	
				Mirror disk name	
				Spare disk name	

A.8.2 Shared Disk Setup Worksheet

If you are setting up shared disks, you must determine the GDS configuration beforehand.
Use the Shared Disk Setup Worksheet to organize the GDS configuration beforehand.

See

- Before setting up the GDS configuration, you need to know basic information from reading *"PRIMECLUSTER Global Disk Services Configuration and Administration Guide."*

- For information on the items in this worksheet, see "6.3.2.3 Setting Up Shared Disks." Also see "A.10.4 GDS Configuration Worksheet" in "A.10 System Design Examples."

GDS Configuration Worksheet

Item					Setting
GDS configuration	Class 1	Class name			
		Class scope (node name)	Node 1		
			Node 2		
		Spare disk 1 (*1)	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
		Single disk 1 (*2)	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
			Single volume 1	Volume name	
				Size	
			Single volume 2	Volume name	
				Size	
		Single disk 2 (*2)	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
			Single volume 1	Volume name	
				Size	
			Single volume 2	Volume name	
				Size	

		Disk 1 to be connected to group (*3)	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
		Disk 2 to be connected to group (*3)	SDX disk name		
			Physical disk name in node 1		
			Physical disk name in node 2		
		Low-order group 1 (*4)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
		Low-order group 2 (*4)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	

		Highest-order group 1 (*5)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
			Volume 1	Volume name	
				Size	
			Volume 2	Volume name	
				Size	
		Highest-order group 2 (*5)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
			Volume 1	Volume name	
				Size	
			Volume 2	Volume name	
				Size	

<table>
<tr><td rowspan="25"></td><td rowspan="25">Class 2</td><td colspan="3">Class name</td><td></td></tr>
<tr><td rowspan="2">Class scope (node name)</td><td colspan="2">Node 1</td><td></td></tr>
<tr><td colspan="2">Node 2</td><td></td></tr>
<tr><td rowspan="3">Spare disk 1 (*1)</td><td colspan="2">SDX disk name</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td></td></tr>
<tr><td rowspan="7">Single disk 1 (*2)</td><td colspan="2">SDX disk name</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td></td></tr>
<tr><td rowspan="2">Single volume 1</td><td>Volume name</td><td></td></tr>
<tr><td>Size</td><td></td></tr>
<tr><td rowspan="2">Single volume 2</td><td>Volume name</td><td></td></tr>
<tr><td>Size</td><td></td></tr>
<tr><td rowspan="7">Single disk 2 (*2)</td><td colspan="2">SDX disk name</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td></td></tr>
<tr><td rowspan="2">Single volume 1</td><td>Volume name</td><td></td></tr>
<tr><td>Size</td><td></td></tr>
<tr><td rowspan="2">Single volume 2</td><td>Volume name</td><td></td></tr>
<tr><td>Size</td><td></td></tr>
<tr><td rowspan="3">Disk 1 to be connected to group (*3)</td><td colspan="2">SDX disk name</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td></td></tr>
<tr><td rowspan="3">Disk 2 to be connected to group (*3)</td><td colspan="2">SDX disk name</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td></td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td></td></tr>
</table>

		Low-order group 1 (*4)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
		Low-order group 2 (*4)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
		Highest-order group 1 (*5)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
			Volume 1	Volume name	
				Size	
			Volume 2	Volume name	
				Size	

		Highest-order group 2 (*5)	Group name		
			Group type		
			Stripe width (*6)		
			Disk /low-order group name	Disk /low-order group 1	
				Disk /low-order group 2	
			Volume 1	Volume name	
				Size	
			Volume 2	Volume name	
				Size	

(*1) This item can be defined only if mirroring is performed.
(*2) This item must be defined only if a single volume is to be created.
(*3) This item must be defined only if mirroring, striping, or concatenation is performed.
(*4) This item must be defined only if the groups are to be arranged into a hierarchy.
(*5) This item must be defined only if mirroring, striping, or concatenation is performed.
(*6) This item must be defined only if the group type is stripe.

A.9 Cluster Application Worksheets

The Cluster Application Worksheets are necessary for setting up cluster applications and resources.

The following worksheets are available:

Resource configuration:

- Cmdline Resource Worksheet
- Fsystem Resource Worksheet

- Gds (Global Disk Services) Resource Worksheet
- Gls (Global Link Services) Resource Worksheet
- Takeover Network Resource Worksheet
- Procedure Resource Worksheet
- Process Monitoring Resource Worksheet
- Line Switching Unit Resource Worksheet (only in Oracle Solaris 10 environment)

Cluster application configuration:

- Cluster Application Setup Worksheet

Exclusive relationship configuration for cluster applications:

- Cluster Application Exclusion Setup Worksheet

Global Settings in Configuration Setup:

- Configuration Global Settings Setup Worksheet

Note

When setting up resources, if the number of objects of any resource exceeds 50, make sure that the numbers of resource objects and detectors do not exceed the maximum number. The maximum numbers of resource objects and detectors are the following:

- Cluster applications: 48
- Resource objects: 500Resource objects: 500
- Detectors: 100

You cannot create the configuration of a cluster application which exceeds the values described above.

Calculate the required number of detectors according to the following procedure and make sure that the number of detectors does not exceed the maximum number.

(1) Calculate the required number of detectors for each resource based on the following table.

Resource name	Number of detectors (round up the decimal point)
Cmdline resource	Number of objects / 2 = (a)
Fsystem resource	Number of objects / 5 = (b)
NFS resource (the number of objects with the SHARE flag among the Fsystem resources)	Number of objects / 10 = (c)
Gds resource	Number of objects / 64 = (d)
Gls resource	Number of objects / 64 = (e)
Takeover network resource	Number of objects / 10 = (f)
Procedure resource	Number of objects / 1 = (g)
Process monitoring resource	Number of objects / 1 = (h)
Line Switching Unit resource	Number of objects / 2 = (i)

- Values such as 2, 5, 10, and 64 in the table indicate the number of the objects which is monitored by a detector of each resource.
- For details on objects corresponding to each resource, refer to setting worksheets of each resource.
- The number of objects indicates the total number of all objects of resources which are registered with all cluster applications. The number of detectors is calculated by using the total number of objects.

(2) Add up the number of detectors required for each resource.

```
                  Number of objects  Number of detectors
Cmdline                 40             (a) 40 / 2 = 20
Fsystem                190             (b) 190 / 5 = 38
NFS                      0             (c) 0 / 10 = 0
Gds                    200             (d) 200 / 64 = 4 (approximately equal to 3.1)
Gls                     50             (e) 50 / 64 = 1 (approximately equal to 0.7)
Takeover network         0             (f) 0 / 10 = 0
Procedure               10             (g) 10 / 1 = 10
Process monitoring      10             (h) 10 / 1 = 10
Line Switching Unit      0             (i) 0 / 2 = 0
                       500
```

In the above example, the total number of detectors is "(a)20 + (b)38 + (c)0 + (d)4 + (e)1 + (f)0 + (g)10 + (h)10 + (i)0 = 83." This configuration is available because the value "83" does not exceed the maximum number of detectors "100."

See

For information on the items in this worksheet, see "6.7 Setting Up Cluster Applications."

A.9.1 Cmdline Resource Worksheet

The Cmdline Resource Setup Worksheet is used for configuring a Cmdline resource.

See

For information on the items in this worksheet, see "6.7.1.1 Creating Cmdline Resources."

Item				Setting value
Cmdline resource	Active node candidate		SysNode	
	Resource name			
	Resource operation	Attribute	InParallel	◇Yes ◇[No]
			NeedAll	◇Yes ◇[No]
		Association	SubApplication	
	Script to be controlled	Object	Start script	
			Stop script	
			Check script	
		Flag	NULLDETECTOR	◇Yes ◇[No]
			ALLEXITCODES	◇Yes ◇[No]
			LIEOFFLINE	◇Yes ◇[No]
			CLUSTEREXCLUSIVE	◇Yes ◇[No]
			AUTORECOVER	◇Yes ◇[No]
			MONITORONLY	◇Yes ◇[No]
			STANDBYCAPABLE	◇Yes ◇[No]
			REALTIME	◇Yes ◇[No]
			TIMEOUT	(Range: 45 to 31535999) [300]

Note: Values enclosed in brackets ([ ]) are default values.

A.9.2 Fsystem Resource Worksheet

The Fsystem Resource Worksheet is used for configuring an Fsystem resource.

See

For information on the setup items in this worksheet, see "6.7.1.2 Creating Fsystem Resources."

Item				Setting value
Fsystem resource	Active node candidate		SysNode	
	Resource name			
	Resource operation	Attribute	NeedAll	◇[Yes] ◇No
			Timeout	(Range: 45 to 3600) [180]
	Mount point to be controlled 1	Object	Mount point	
		Flag	AUTORECOVER	◇Yes ◇[No]
			SHARE	◇Yes ◇[No]
			NFSLOCKFAILOVER	◇Yes ◇[No]
	Mount point to be controlled 2	Object	Mount point	
		Flag	AUTORECOVER	◇Yes ◇[No]
			SHARE	◇Yes ◇[No]
			NFSLOCKFAILOVER	◇Yes ◇[No]
	Mount point to be controlled 3	Object	Mount point	
		Flag	AUTORECOVER	◇Yes ◇[No]
			SHARE	◇Yes ◇[No]
			NFSLOCKFAILOVER	◇Yes ◇[No]

Note: Values enclosed in brackets ([ ]) are default values.

Note

Before you can set the NFSLOCKFAILOVER to Yes, you need to select Yes for NFS Lock Failover and then specify the Lock information directory name for the Global setting in the configuration.

For information on the Global Settings in the Configuration, see "A.9.11 Configuration Global Settings Worksheet."

A.9.3 Gds Resource Worksheet

The Gds Resource Worksheet is used for configuring a Gds (Global Disk Services) resource.

See

For information on the setup items in this worksheet, see "6.7.1.3 Creating Gds Resources."

Item				Setting value
Gds resource	Active node candidate		SysNode	
	Resource name			
	Resource operation	Attribute	Use	Exclusive use ◇[Enable] Hot Standby operation ◇Enable ◇[Disable] ◇Disable
			AutoRecover	◇Yes ◇[No]
			Timeout	(Range: 5 to 3600) [1800]
	Disk class to be controlled 1	Object	Disk Class	
		Flag	MONITORONLY	◇Yes ◇[No]
	Disk class to be controlled 2	Object	Disk Class	
		Flag	MONITORONLY	◇Yes ◇[No]
	Disk class to be controlled 3	Object	Disk Class	
		Flag	MONITORONLY	◇Yes ◇[No]

Note: Values enclosed in brackets ([ ]) are default values.

A.9.4 Gls Resource Worksheet

The Gls Resource Worksheet is used for configuring a Gls (Global Link Services) resource.

See

For information on the setup items in this worksheet, see "6.7.1.4 Creating Gls Resources."

Item				Setting value
GIs resource	Active node candidate		SysNode	
	Resource name			
	Resource operation	Attribute	Timeout	(Range: 5 to 300) [60]
		Association	SubApplication	
	Takeover IP address to be controlled 1	Object	Takeover IP address(IPv4)	
			Takeover IP address(IPv6)	
		Flag	AUTORECOVER	◇Yes ◇[No]
	Takeover IP address to be controlled 2	Object	Takeover IP address(IPv4)	
			Takeover IP address(IPv6)	
		Flag	AUTORECOVER	◇Yes ◇[No]
	Takeover IP address to be controlled 3	Object	Takeover IP address(IPv4)	
			Takeover IP address(IPv6)	
		Flag	AUTORECOVER	◇Yes ◇[No]

Note: Values enclosed in brackets ([ ]) are defaul values.

A.9.5 Takeover Network Resource Worksheet

The Takeover Network Resource Worksheet is used for configuring a takeover network resource.

See

For information on the items in this worksheet, see "6.7.1.5 Creating Takeover Network Resources."

Item				Setting value
Takeover IP address resource	Active node candidate		SysNode	
	Resource name			
	Takeover network type			◇[IP address takeover] ◇Node name takeover + IP address takeover
	Network interfaces	Node 1 ( )		
		Node 2 ( )		
		Node 3 ( )		
		Node 4 ( )		
	Resource operation	Attribute	NeedAll	◇[Yes] ◇No
			Timeout	(Range: 45 to 3600) [60]
			PingHost	
		Association	SubApplication	
	Takeover IP address to be controlled	Object	Host name	
			IP address	
			Net mask	
		Flag	BASE	◇[virtual] ◇base
			AUTORECOVER	◇Yes ◇[No]
			PingHost	

Note: Values enclosed in brackets ([ ]) are default values.

A.9.6 Procedure Resource Worksheet

The Procedure Resource Worksheet is used for configuring a procedure resource.

See

For information on the items in this worksheet, see "6.7.1.6 Creating Procedure Resources."

Item				Setting value
Procedure resource	Active node candicate		SysNode	
	Resource name			
	Resource operation	Attribute	SCRIPTTIMEOUT	(Range: 45 to 31535999) [1800]
		Association	SubApplication	
	Procedure resource to be controlled	Procedure class		◇Application ◇BasicApplication ◇SystemState2 ◇SystemState3 ◇( )
			Procedure resource	

Note: Values enclosed in brackets ([ ]) are default values.

A.9.7 Process Monitoring Resource Worksheet

The Process Monitoring Resource Worksheet is used for configuring a process monitoring resource.

See

For information on the items in this worksheet, see "6.7.1.7 Creating Process Monitoring Resources."

Item				Setting value
Process resource	Active node candidate		SysNode	
	Resource name			
	Resource operation	Attribute	InParallel	<>Yes <>[No]
			NeedAll	<>Yes <>[No]
		Association	SubApplication	
	Command pathto be controlled	Object	Start command path	
			Stop command path	
			Monitored process is in different process group from parent process	If "Yes," check "Process is a daemon."
		Attribute	Restart count for monitored process	(Range: 0 to 99) [3]
			Restart interval for monitored process	(Range: 0 to 3600) [3]

Note: Values enclosed in brackets ([ ]) are default values.

A.9.8 Line Switching Unit Resource Worksheet (Only in Oracle Solaris 10 Environment)

The Line Switching Unit Resource Worksheet is used for configuring a line switching unit resource.

See

For information on the items in this worksheet, see "6.7.1.8 Creating Line Switching Unit Resources."

Item			Setting value
SH_SWLine resource	Active node candidate	SysNode	
	Resource name		
	Control target	Line switching unit resource	

Note: Values enclosed in brackets ([ ]) are default values.

A.9.9 Cluster Application Worksheet

The Cluster Application Worksheet is used for configuring a cluster application.

![See]

See

For information on the items in this worksheet, see "6.7.2 Creating Cluster Applications."

Item					Setting value	Remarks (value specified with GUI/CUI)
Cluster application settings	Configuration name				config	Fixed
	Cluster application name					
	Operating node		SysNode (Priority 1: High)			
			SysNode (Priority 2)			
			SysNode (Priority 3)			
			SysNode (Priority 4: Low)			
	Attribute	Start	Operating node at startup	OnlinePriority	<>[According to cluster application	[0]
					<>Operating node when RMS stopped	1
			Automatic startup	AutoStartUp	<>[Manual]	[No]
					<>Automatic	Yes
		Switchover	Switchover operation	AutoSwitchOver	<>[Manual]	[No]
					<>Automatic (more than one of the following can be selected)	
					[ ]At node failure	HostFailure
					[ ]At resource failure	ResourceFailure
					[ ]When RMS stops	ShutDown
		Standby	Trigger to be Standby	StandbyTransitions	<>[Manual]	[NONE]
					<>Automatic (more than one of the following can be selected)	
					[ ]At startup	StartUp
					[ ]At switchover	SwitchRequest
					[ ]When fault is cleared	ClearFaultRequest
		Operation in the event of a failure	Operation when RMS fails to control 2	HaltFlag	<>[Do not forcibly stop node]	[No]
					<>Forcibly stop node	Yes
			Handling of failed node	PersistentFault	<>[Use if fault is not detected when RMS is restarted]	[0]
					<>Use after operator clears fault	1
			Operation if cluster interconnect fails	ShutdownPriority	<>[NONE]	[NONE]
					<>Priority ( )	Priority (0 to 20) * Higher number indicates higher priority. * See setup policy for survival priority if cluster partition occurs.
	Configuration	Resource				Specified resource name
		Patrol diagnosis			<>Required	Check
					<>[Not required]	[Do not check]

Note: Values enclosed in brackets ([ ]) are default values.

See

For information on the "setup policy for survival priority if cluster partition occurs" in the Remarks column for Shutdown Priority, see "Survival scenarios" in "5.1.2.2.2 Using the Shutdown Configuration Wizard."

A.9.10 Cluster Application Exclusion Worksheet

The Cluster Application Exclusion Worksheet is used for configuring exclusion relationships between cluster applications.

See

For information on the items in this worksheet, see "6.7.3 Setting Up Dependency Relationships Between Cluster Applications."

Item			Setting value
Dependency relationship settings for cluster applications	Exclusion group 1	Group name	Exclusive
		Cluster application name 1	(Priority:)
		Cluster application name 2	(Priority:)
		Cluster application name 3	(Priority:)
		Cluster application name 4	(Priority:)
		Cluster application name 5	(Priority:)
	Exclusion group 2	Group name	Exclusive
		Cluster application name 1	(Priority:)
		Cluster application name 2	(Priority:)
		Cluster application name 3	(Priority:)
		Cluster application name 4	(Priority:)
		Cluster application name 5	(Priority:)
	Exclusion group 3	Group name	Exclusive
		Cluster application name 1	(Priority:)
		Cluster application name 2	(Priority:)
		Cluster application name 3	(Priority:)
		Cluster application name 4	(Priority:)
		Cluster application name 5	(Priority:)
	Exclusion group 4	Group name	Exclusive
		Cluster application name 1	(Priority:)
		Cluster application name 2	(Priority:)
		Cluster application name 3	(Priority:)
		Cluster application name 4	(Priority:)
		Cluster application name 5	(Priority:)

A.9.11 Configuration Global Settings Worksheet

This worksheet is used for configuring the Global Settings in the Configuration.

See

For information on the setup items in this worksheet, see "6.7.4 Editing global settings in Configuration."

Item				Setting value
Global Settings in the Configuration	PreCheck Timeout			(Range: 5 to 2147483647) [300]
	First Available Detector			(Range: 0 to127) [0]
	Last Available Detector			(Range: 0 to127) [127]
	NFS Lock Failover			<>Yes <>[No]
			Lock information directory	*1)Only if Yes is selected above [ ]
	Detector Details	Monitoring Intervals	hvdet_execbin Used for ops, rtp. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_ckhost Used for ops, rtp. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_ckshare Used for ops, rtp.	(Range: 5 to 2147483647) [10]
			hvdet_glbassrt Used for ops. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_gmount Used for Fsystem.	(Range: 5 to 2147483647) [10]
			hvdet_icmp Used for Ipaddress.	(Range: 5 to 2147483647) [10]
			hvdet_locassrt Used for userApplication. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_lvm Used for Lvm. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [18]
			hvdet_nfs Used for Fsystem.	(Range: 5 to 2147483647) [10]
			hvdet_rcfs Used for Rcfs. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [9]
			hvdet_rcvm Used for Rcvm. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [33]
			hvdet_read Used for Rawdisk. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_srdf Used for Srdf. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [59]
			hvdet_stopclnt Used for Fsystem. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [10]
			hvdet_system Used for Cmdline, Fsystem, userApplication.	(Range: 5 to 2147483647) [10]
			hvdet_vxvm Used for Vxvm. (Unused parameter in PRIMECLUSTER for Solaris)	(Range: 5 to 2147483647) [30]
			hvdet_zfs Used for Fsystem.	(Range: 5 to 2147483647) [20]
			ForeignDetectors Used for foreign-code.	(Range: 5 to 2147483647) [30]
		Log level (MemoryLogLevel)		(Range: 0 to 2147483647) [6]
		Detector logging		(Range: 0 to 2147483647) [0]

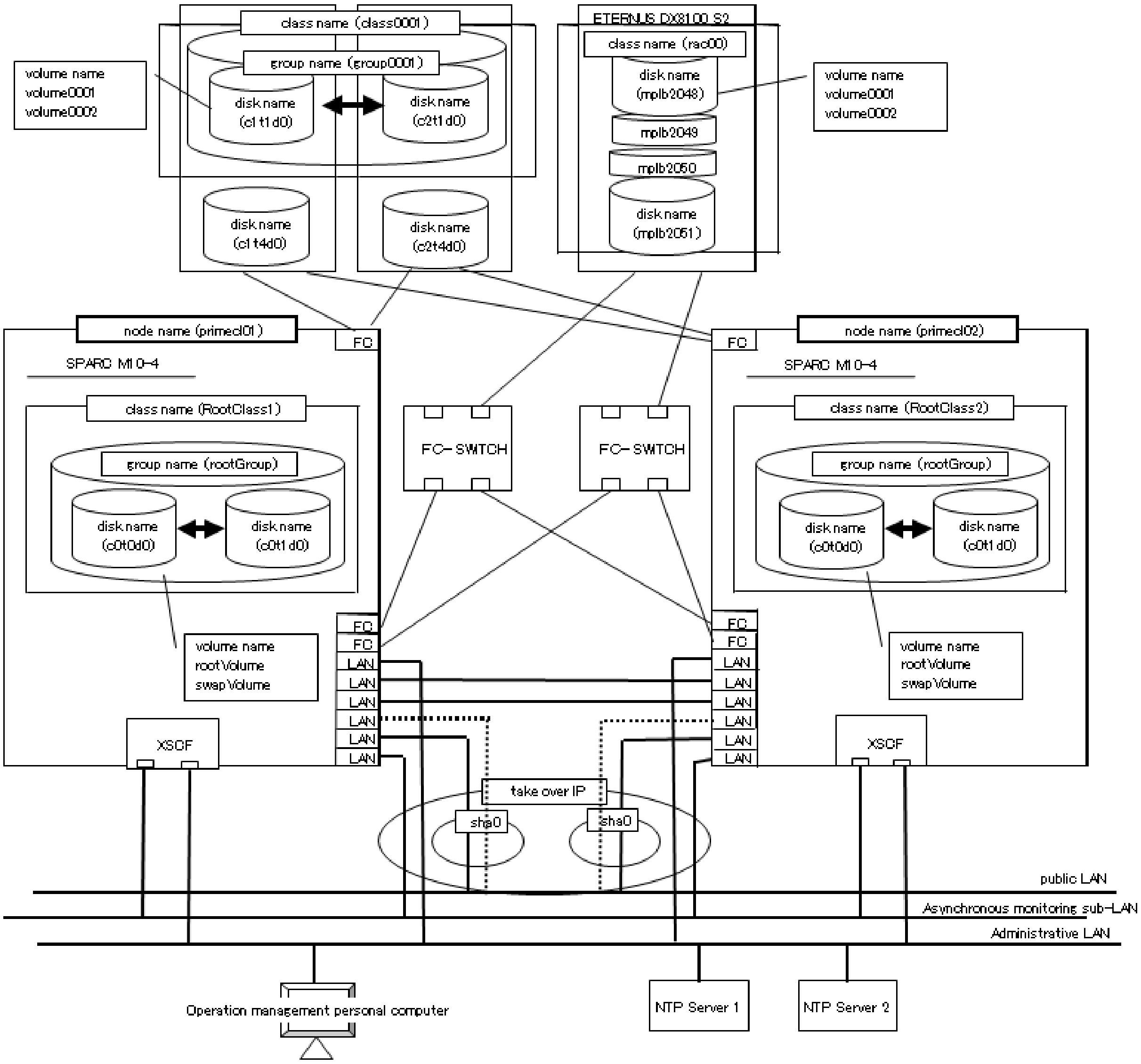
*1) Values enclosed in brackets ([ ]) are default values.
Note: If the settings for items other than NFS Lock Failover need to be changed, please contact Fujitsu systems engineer (SE). However, no changes may be necessary for the normal operation.

A.10 System Design Examples

This section shows design examples of the PRIMECLUSTER system. Refer to these examples when designing the PRIMECLUSTER system.

System configuration

Figure A.1 System configuration example



Note: The NTP server should be set to one of the following operation mode;

- Maintains time that is reliable for special hardware devices such as time units and GPS units.
- Uses another server that maintains time that is reliable for special hardware devices such as time units and GPS units, and uses NTP for time synchronization.

Used hardware

- SPARC Enterprise M10-4 x 2
- Expansion file unit x 2
- Fibre channel switch (SN200) x 2
- Fast Ethernet card x 6
- Fibre channel card x 6
- Disk array unit (ETERNUS DX8100 S2) x 1

- Operation management personal computer x 1

Used software

- Solaris 11
- Enhanced Support Facility 3.2
- PRIMECLUSTER 4.3
- Fujitsu PCI Fibre Channel 2.2
- Microsoft Internet Explorer 9 (Installed in the operation management personal computer)
- Java(TM) 2 Platform Standard Edition Runtime Environment Version 6 (installed in the operation management personal computer)

System disk settings

Node name	Physical disk name	Slice number	ZFS root pool name	Size (MB)
primecl01	c0t0d0	0	rpool	15374
primecl02	c0t0d0	0	rpool	15374

Note

If GDS is used for system disk mirroring, one slice number in the system disk and at least 20 megabytes of free area are used. You need to consider the resources used by GDS when designing the system.

For details, see "*PRIMECLUSTER Global Disk Service Configuration and Administration Guide*."

Shared disk array settings

Device name	Number of LU	Number of RAID groups	RAID level	Size
ETERNUS3000 model 300(FC)	4	1	5	248.2 GB
Expansion file unit (FC) x 2	-	-	-	18.2 GB

Note

To use the disk array units on UNIX, you need to format them by executing the "format (1M)" command.
When formatting the disk array units as part of the cluster system configuration work, you need to estimate the formatting time.

A.10.1 Cluster Configuration Worksheet

Shown below is an example of creating the Cluster Configuration Worksheet.

	Item			Setting	Remarks
Shared disk unit: Type 1	Machine model			ETERNUS DX8100 S2	
	Device name			ETERNUSDX8100#003	
	IP address			10.33.66.215	
	Subnet mask			255.255.255.0	
	Interface			Fibre channel	
	LUN number			0 - 0F	
	Ports			CM0CA2P0,CM1CA2P0	
	Connection paths			2	
	FC-SW		Use of fibre channel switching hub	■ yes □ no	
			Machine model	ETERNUS SN200 Model 630	
			Device name	ETERNUSSN200#003	
			IP address	10.33.66.203	
			Subnet mask	255.255.255.0	
	File system selection (type of file system used in cluster)			UFS	
Shared disk unit: Type 2	Machine model			ETERNUS DX8100 S2	
	Device name			ETERNUSDX8100#004	
	IP address			10.33.66.216	
	Subnet mask			255.255.255.0	
	Interface			Fibre channel	
	LUN number			0 - 0F	
	Ports			CM0CA0P0,CM1CA0P0	
	Connection paths			2	
	FC-SW		Use of fibre channel switching hub	■ yes □ no	
			Machine model	ETERNUS SN200 Model 630	
			Device name	ETERNUSSN200#004	
			IP address	10.33.66.204	
			Subnet mask	255.255.255.0	
	File system selection (type of file system used in cluster)			GFS shared file system	
NTP	Operation mode			□ NTP server (integrated in cluster) ■ NTP client	
	Protocol (only when broadcast is				
	Server	Primary	Host name	pclntp1	
			IP address	10.124.95.11	
			Subnet mask	255.255.255.0	
		Secondary	Host name	pclntp2	
			IP address	10.21.8.3	
			Subnet mask	255.255.255.0	
Web-Based Admin View (Management view) For control domain	Operation mode			□ 1 ■ 2 □ 3 layers	
	Server	Primary	Host name	primecl01	
			IP address	10.34.214.181	
			Subnet mask	255.255.255.0	
			URL	http://10.34.214.181:8081/Plugin.html	
		Secondary	Host name	primecl02	
			IP address	10.34.214.182	
			Subnet mask	255.255.255.0	
			URL	http://10.34.214.182:8081/Plugin.html	
	User ID		wvroot group	pclwv	
			clroot group	pclcl	
			cladmin group	pcladm	
			clmon group	pclmon	
			sdxroot group	pclsdx	

Web-Based Admin View (Management view) for a guest domain of OVM	Operation mode			□ 1 □ 2 □ 3 layers	
	Server	Primary	Host name		
			IP address		
			Subnet mask		
			URL		
		Secondary	Host name		
			IP address		
			Subnet mask		
			URL		
	User ID		wvroot group		
			clroot group		
			cladmin group		
			clmon group		
			sdxroot group		
Cluster Foundation (CF)	Cluster name			SLV002	
	Cluster nodes			2	
	Cluster node names			primecl01,primecl02	
	Subnets			1	
	Subnet number			192.168.1.0	
	Subnet mask			255.255.255.0	
	Use of CF remote services			■ cfcp (CF file copy) ■ cfsh (CF remote command execution)	
	IP interconnect			◆ [no] ◇ yes (the number of IP interconnects)	
	Use by RMS			◆ [yes] ◇ no	[ ] Default
Shutdown facility (SF)	Shutdown agent			□ RCI ■ XSCF (SPARC M series) □ XSCF (SPARC Enterprise M series) □ ILOM □ ALOM	
Cluster resource management facility (CRM)	Hardware devices stored in resource database			■ Network unit	Required for using takeover of public LAN (IP) and node names
				□ Line switching unit	Required for using line takeover function
				■ Disk unit	Required when using GDS and GFS

Node 1	Machine model		SPARC M10-4	
	Device name		DBSV#001	
	Domain ID / domain name		domain0	
	Node name (uname -n)		primecl01	
	CF node name		primecl01	
	Interconnect	IP address	192.168.1.1	
		Subnet mask	255.255.255.0	
		Device name	bge1,bge2	
	Node in CF quorum set		◆ [yes] ◇ no	Do not change the default setting.
	CF cluster timeout (CLUSTER_TIMEOUT)		"10"	Fixed value in the OVM environment. Do not change the default setting.
	Use by RMS		◆ [yes] ◇ no Suffix: primecl01RMS	[ ] Default
	RCI	Address		Only when using RCI asynchronous monitoring
	Kernel parameter	scfd:scf_rdctrl_sense_wait		
	System console	IP address 1	10.33.66.71	XSCF, ALOM, and ILOM data (The setting of 2 is for SPARC M series)
		Subnet mask 1	255.255.255.0	
		IP address 2	10.33.76.71	
		Subnet mask 2	255.255.255.0	
		User ID	cl01sa	
		User authority	platadm	
		Connection type	■ SSH □ telnet	
	Shutdown facility (SF)	Administrative IP addresses	10.33.66.91	
		Weight (Node weight)	1	Node weight (priority if cluster partition occurs) used by SF
		User ID	xuser	XSCF, ALOM, and ILOM data
		User authority	platadm	
		Console LAN#1	10.33.66.71	
		Console LAN#2	10.33.76.71	
		PPAR-ID	0	SPARC M series
		Domain name	primary	
		Connection type	◆ [SSH] ◇ telnet	[ ] Default For XSCF
	Use of shared disk		■ yes □ no	
	scsi-initiator-id (OBP variable #eeprom command)		7	
	SCSI / fibre channel	Circuits (paths)	2	
		Device name 1	fjpfca#0	
		sd device name	sd11 - sd42	
		Device name 2	fjpfca#1	
		sd device name	sd43 - sd74	
	Management LAN	IP address	10.33.66.91	
		Subnet mask	255.255.255.0	
		Device name	bge0	
	Asynchronous monitoring sub-LAN	IP address	10.33.76.91	SPARC M series
		Subnet mask	255.255.255.0	
		Device name	bge0	
	Public LAN 1	IP address	10.34.214.181	
		Subnet mask	255.255.255.0	
		Device name	fjgi0	
	Public LAN 2	IP address	10.34.215.21	
		Subnet mask	255.255.255.0	
		Device name	fjgi1	

Node 2	Machine model		SPARC M10-4	
	Device name		DBSV#002	
	Domain ID / domain name		domain0	
	Node name (uname -n)		primecl02	
	CF node name		primecl02	
	Interconnect	IP address	192.168.1.2	
		Subnet mask	255.255.255.0	
		Device name	bge1,bge2	
	Node in CF quorum set		◆ [yes] ◇ no	[ ] Default Do not change the default setting.
	CF cluster timeout (CLUSTER_TIMEOUT)		"10"	Fixed value in the OVM environment. Do not change the default setting.
	Use by RMS		◆ [yes] ◇ no Suffix: primecl02RMS	[ ] Default
	RCI	Address		Only when using RCI asynchronous monitoring
	Kernel parameter	scfd:scf_rdctrl_sense_wait		
	System console	IP address 1	10.33.66.72	XSCF, ALOM, and ILOM data (The setting of 2 is for SPARC M series)
		Subnet mask 1	255.255.255.0	
		IP address 2	10.33.76.72	
		Subnet mask 2	255.255.255.0	
		User ID	cl02sa	
		User authority	platadm	
		Connection type	■ SSH ☐ telnet	
	Shutdown facility (SF)	Administrative IP addresses	10.33.66.92	
		Weight (Node weight)	2	Node weight (priority if cluster partition occurs) used by SF
		User ID	xuser	XSCF, ALOM, and ILOM data
		User authority	platadm	
		Console LAN#1	10.33.66.72	
		Console LAN#2	10.33.76.72	
		PPAR-ID	1	SPARC M series
		Domain name	primary	
		Connection type	◆ [SSH] ◇ telnet	[ ] Default For XSCF
	Use of shared disk		■ yes ☐ no	
	scsi-initiator-id (OBP variable #eeprom command)		8	
	SCSI / fibre channel	Circuits (paths)	2	
		Device name 1	fjpfca#0	
		sd device name	sd11 - sd42	
		Device name 2	fjpfca#1	
		sd device name	sd43 - sd74	
	Management LAN	IP address	10.33.66.92	
		Subnet mask	255.255.255.0	
		Device name	bge0	
	Asynchronous monitoring sub-LAN	IP address	10.33.76.92	SPARC M series
		Subnet mask	255.255.255.0	
		Device name	bge0	
	Public LAN 1	IP address	10.34.214.182	
		Subnet mask	255.255.255.0	
		Device name	fjgi0	
	Public LAN 2	IP address	10.34.215.22	
		Subnet mask	255.255.255.0	
		Device name	fjgi1	

In each node, specify /etc/inet/ntp.conf as shown below.

```
server 10.34.214.100
server 10.34.214.101
server 127.127.1.0
```

```
fudge 127.127.1.0 stratum 9

enable auth monitor

disable pll
driftfile /etc/ntp/ntp.drift
statsdir /var/ntp/ntpstats/
filegen peerstats file peerstats type day enable
filegen loopstats file loopstats type day enable
filegen clockstats file clockstats type day enable
```

For detailed instructions on specifying /etc/inet/ntp/conf, see the online manual by executing the "xntpd(1M)" command.

A.10.2 GLS Setup Worksheet

Shown below are examples of creating the GLS Setup Worksheet.

NIC Switching Mode (Logical IP Takeover)

Item				Setting
GLS setup	Switching mode			◆NIC switching mode (logical IP takeover) ◇Fast switching mode
	Takeover virtual interface name			sha0:65
		Takeover virtual IP address (or host name)		10.34.214.185
		Subnet mask		255.255.255.0
	Node name (1)			primecl01
		Configuration information	Virtual interface name	sha0
			Primary physical interface name	hme0
			Secondary physical interface name	hme3
			Physical IP address (or host name)	10.34.214.181
			Logical IP address (or host name)	10.34.214.185
		Monitoring destination information	Primary monitoring destination IP address (or host name)	10.34.214.188
			Secondary monitoring destination IP address (or host name)	10.34.214.189
		Optional function	Backup NIC patrol	◇Disable ◆Enable Set to backup NIC. Local MAC address (02:00:00:00:00:01)
	Node name (2)			primecl02
		Configuration information	Virtual interface name	sha0
			Primary physical interface name	hme0
			Secondary physical interface name	hme3
			Physical IP address (or host name)	10.34.214.182
			Logical IP address (or host name)	10.34.214.185
		Monitoring destination information	Primary monitoring-destination IP address (or host name)	10.34.214.188
			Secondary monitoring-destination IP address (or host name)	10.34.214.189
		Optional function	Backup NIC patrol	◇Disable ◆Enable Set to backup NIC Local MAC address (02:00:00:00:00:02)

Fast Switching Mode

Item					Setting
GLS setup	Switching mode				◇NIC switching mode (logical IP takeover) ◆Fast switching mode
	Takeover virtual interface name				sha0:65
		Takeover virtual IP address (or host name)			10.34.214.185
		Subnet mask			255.255.255.0
	Node name (1)				primecl01
		Configuration information	Virtual interface name		sha0
			Physical interface name (1)		hme0
			Physical interface name (2)		hme3
				Virtual IP address (or host name)	10.34.214.181
				Subnet mask	255.255.255.0
				Physical IP address (or host name) (1)	10.34.215.181
				Subnet mask	255.255.255.0
				Physical IP address (or host name) (2)	10.34.216.181
				Subnet mask	255.255.255.0
	Node name (2)				primecl02
		Configuration information	Virtual interface name		sha0
			Physical interface name (1)		hme0
			Physical interface name (2)		hme3
				Virtual IP address (or host name)	10.34.214.182
				Subnet mask	255.255.255.0
				Physical IP address (or host name) (1)	10.34.215.182
				Subnet mask	255.255.255.0
				Physical IP address (or host name) (2)	10.34.216.182
				Subnet mask	255.255.255.0

A.10.3 System Disk Mirror Setup Worksheet

Shown below is an example of creating the System Disk Mirror Setup Worksheet.

Item					Setting
System disk mirroring setup	Node name				primecl01
		Class name			Rootclass1
			Group name		rootGroup
				Mount point (for a UFS boot environment) or a ZFS storage pool name (for a ZFS boot environment)	/, /swap, /usr, /opt
				Physical disk name	c0t0d0
				Mirror disk name	c0t1d0
				Spare disk name	-
	Node name				primecl02
		Class name			Rootclass2
			Group name		rootGroup
				Mount point (for a UFS boot environment) or a ZFS storage pool name (for a ZFS boot environment)	/, /swap, /usr, /opt
				Physical disk name	c0t0d0
				Mirror disk name	c0t1d0
				Spare disk name	-

A.10.4 GDS Configuration Worksheet

Shown below are examples of creating the GDS Configuration Worksheet.

System configuration example

For a system configuration example, set up the worksheet as shown below.

Item					Setting
GDS configuration	Class 1	Class name			rac00
		Class scope (node name)	Node 1		primecl01
			Node 2		primecl02
		Single disk 1	SDX disk name		diskmplb0001
			Physical disk name in node 1		mplb2048
			Physical disk name in node 2		mplb2048
			Single volume 1	Volume name	volume0001
				Size	1024 MB
			Single volume 2	Volume name	volume0002
				Size	1024 MB
		Single disk 2	SDX disk name		diskmplb0002
			Physical disk name in node 1		mplb2049
			Physical disk name in node 2		mplb2049
		Single disk 3	SDX disk name		diskmplb0003
			Physical disk name in node 1		mplb2050
			Physical disk name in node 2		mplb2050
		Single disk 4	SDX disk name		diskmplb0004
			Physical disk name in node 1		mplb2051
			Physical disk name in node 2		mplb2051
	Class 2	Class name			class0001
		Class scope (node name)	Node 1		primecl01
			Node 2		primecl02
		Spare disk 2	SDX disk name		-
			Physical disk name in node 1		-
			Physical disk name in node 2		-
		Disk 1 connected to group	SDX disk name		disk010001
			Physical disk name in node 1		c1t1d0
			Physical disk name in node 2		c1t1d0
		Disk 1 connected to group	SDX disk name		disk020001
			Physical disk name in node 1		c2t1d0
			Physical disk name in node 2		c2t1d0
		Highest-order group 1	Group name		group0001
			Group type		mirror
			Stripe width (*6)		-
			Disk/low-order group name	Disk/low-order group 1	disk010001
				Disk/low-order group 2	disk010002
			Volume 1	Volume name	volume0001
				Size	1024 MB
			Volume 2	Volume name	volume0002
				Size	1024 MB

Examples of creating the GDS Configuration Worksheet are provided below.

Example 1. Managing a disk as a single disk:

In this example, single volumes Volume1 and Volume2 are created in physical disk c3t4d5.

Item				Setting
Class 1	Class name			Class1
	Class scope (node name)	Node 1		Node1
		Node 2		Node2
	Single disk 1	SDX disk name		Disk1
		Physical disk name in node 1		c3t4d5
		Physical disk name in node 2		c2t3d4
		Single volume 1	Volume name	Volume1
			Size	1024 blocks
		Single volume 2	Volume name	Volume2
			Size	1024 blocks

Example 2. If a disk is not mirrored but may be changed to a mirror configuration later:

In this example, mirror group Group1, which consists only of physical disk c3t4d5, is created, and single volumes Volume1 and Volume2 are created in Group1.

Since Group1 consists of only one disk, Volume1 and Volume2 are not mirrored. If another disk is added to Group1, then Volume1 and Volume2 can be changed to a mirror configuration.

Item				Setting
Class 1	Class name			Class1
	Class scope (node name)	Node 1		Node1
		Node 2		Node2
	Disk 1 connected to group	SDX disk name		Disk1
		Physical disk name in node 1		c3t4d5
		Physical disk name in node 2		c2t3d4
	Highest-order group 1	Group name		Group1
		Group type		mirror
		Stripe width		–
		Disk/low-order group name	Disk/low-order group 1	Disk1
		Volume 1	Volume name	Volume1
			Size	1024 blocks
		Volume 2	Volume name	Volume2
			Size	1024 blocks

Example 3. If disk mirroring is to be executed:

In this example, mirror group Group1, which consists of physical disks c3t4d5 and c4t5d6, is created, and mirrored volumes Volume1 and Volume2 are created in Group1.

<table>
<tr><th colspan="4">Item</th><th>Setting</th></tr>
<tr><td rowspan="21">Class 1</td><td colspan="3">Class name</td><td>Class1</td></tr>
<tr><td rowspan="2">Class scope (node name)</td><td colspan="2">Node 1</td><td>Node1</td></tr>
<tr><td colspan="2">Node 2</td><td>Node2</td></tr>
<tr><td rowspan="3">Spare disk 1</td><td colspan="2">SDX disk name</td><td>Disk3</td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td>c1t2d3</td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td>c2t2d2</td></tr>
<tr><td rowspan="3">Disk 1 connected to group</td><td colspan="2">SDX disk name</td><td>Disk1</td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td>c3t4d5</td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td>c2t3d4</td></tr>
<tr><td rowspan="3">Disk 2 connected to group</td><td colspan="2">SDX disk name</td><td>Disk2</td></tr>
<tr><td colspan="2">Physical disk name in node 1</td><td>c4t5d6</td></tr>
<tr><td colspan="2">Physical disk name in node 2</td><td>c3t3d3</td></tr>
<tr><td rowspan="9">Highest-order group 1</td><td colspan="2">Group name</td><td>Group1</td></tr>
<tr><td>Group type</td><td></td><td>mirror</td></tr>
<tr><td>Stripe width</td><td></td><td>–</td></tr>
<tr><td rowspan="2">Disk/low-order group name</td><td>Disk/low-order group 1</td><td>Disk1</td></tr>
<tr><td>Disk/low-order group 2</td><td>Disk2</td></tr>
<tr><td rowspan="2">Volume 1</td><td>Volume name</td><td>Volume1</td></tr>
<tr><td>Size</td><td>1024 blocks</td></tr>
<tr><td rowspan="2">Volume 2</td><td>Volume name</td><td>Volume2</td></tr>
<tr><td>Size</td><td>1024 blocks</td></tr>
</table>

Example 4. If disk striping is to be executed:

In this example, striping group Group1, which consists of physical disks c3t4d5 and c4t5d6, is created, and striped volumes Volume1 and Volume2 are created in Group1.

Item				Setting
Class 1	Class name			Class1
	Class scope (node name)	Node 1		Node1
		Node 2		Node2
	Disk 1 connected to group	SDX disk name		Disk1
		Physical disk name in node 1		c3t4d5
		Physical disk name in node 2		c2t3d4
	Disk 2 connected to group	SDX disk name		Disk2
		Physical disk name in node 1		c4t5d6
		Physical disk name in node 2		c3t3d3
	Highest-rder group 1	Group name		Group1
		Group type		stripe
		Stripe width		32
		Disk/low-order group name	Disk/low-order group 1	Disk1
			Disk/low-order group 2	Disk2
		Volume 1	Volume name	Volume1
			Size	1024 blocks
		Volume 2	Volume name	Volume2
			Size	1024 blocks

Example 5. If disk concatenation is to be executed:

In this example, concatenation group Group1, which consists of physical disks c3t4d5 and c4t5d6, is connected, and volumes Volume1 and Volume2 are created in Group1.

Item				Setting
Class 1	Class name			Class1
	Class scope (node name)	Node 1		Node1
		Node 2		Node2
	Disk 1 connected to group	SDX disk name		Disk1
		Physical disk name in node 1		c3t4d5
		Physical disk name in node 2		c2t3d4
	Disk 2 connected to group	SDX disk name		Disk2
		Physical disk name in node 1		c4t5d6
		Physical disk name in node 2		c3t3d3
	Highest-order group 1	Group name		Group1
		Group type		concat
		Stripe width		-
		Disk/low-order group name	Disk/low-order group 1	Disk1
			Disk/low-order group 2	Disk2
		Volume 1	Volume name	Volume1
			Size	1024 blocks
		Volume 2	Volume name	Volume2
			Size	1024 blocks

Example 6. If disk striping and mirroring are to be executed:

In this example, mirror group Group1, which consists of striping group Group2 and striping group Group3, is created, and volumes Volume1 and Volume2 are created in Group1. Striping group Group2 consists of physical disks c3t4d5 and c4t5d6, and striping group Group3 consists of physical disks c3t4d6 and c4t5d7.

Item				Setting
Class 1	Class name			Class1
	Class scope (node name)	Node 1		Node1
		Node 2		Node2
	Spare disk 1	SDX disk name		Disk5
		Physical disk name in node 1		c1t2d3
		Physical disk name in node 2		c2t2d2
	Disk 1 connected to group	SDX disk name		Disk1
		Physical disk name in node 1		c3t4d5
		Physical disk name in node 2		c2t3d4
	Disk 2 connected to group	SDX disk name		Disk2
		Physical disk name in node 1		c4t5d6
		Physical disk name in node 2		c3t3d3
	Disk 3 connected to group	SDX disk name		Disk3
		Physical disk name in node 1		c3t4d6
		Physical disk name in node 2		c2t3d5
	Disk 4 connected to group	SDX disk name		Disk4
		Physical disk name in node 1		c4t5d7
		Physical disk name in node 2		c3t3d4
	Low-order group 1	Group name		Group2
		Group type		stripe
		Stripe width		32
		Disk/low-order group name	Disk/low-order group 1	Disk1
			Disk/low-order group 2	Disk2
	Low-order group 2	Group name		Group3
		Group type		stripe
		Stripe width		32
		Disk/low-order group name	Disk/low-order group 1	Disk3
			Disk/low-order group 2	Disk4
	Highest-order group 1	Group name		Group1
		Group type		mirror
		Stripe width		–
		Disk/low-order group name	Disk/low-order group 1	Group2
			Disk/low-order group 2	Group3
		Volume 1	Volume name	Volume1
			Size	1024 blocks
		Volume 2	Volume name	Volume2
			Size	1024 blocks

Appendix B Manual Pages

This appendix provides online manual page lists for CCBR, CF, CIP, operator intervention, PAS, the cluster resource management facility, RMS, shutdown facility (SF), SIS, tracing failed resource, Web-Based Admin View, procedure resource, process monitoring feature, and the RMS wizards.

To view a manual page, enter the following command:

$ **man** *man_page_name*

Note:

To view these manual pages, you must set the MANPATH environment variable so that /etc/opt/FJSVcluster/man is included.

To print a hard copy of a manual page, enter the following command:

% **man** *man_page_name* **|col-b |lpr**

B.1 CCBR

System administrator

Command	Function
cfbackup(1M)	Creates a backup copy of all the cluster configuration information related to the PRIMECLUSTER nodes.
cfrestore(1M)	Restores the cluster configuration information stored on the PRIMECLUSTER node.

B.2 CF

System administrator

Command	Function
cfconfig(1M)	Configures or unconfigures a node for a PRIMECLUSTER cluster.
cfset(1M)	Applies or modifies /etc/default/cluster.config entries into the CF module.
cftool(1M)	Prints the node communications state of a node or the cluster.

B.3 CIP

System administrator

Command	Function
cipconfig(1M)	Starts or stops CIP 2.0.
ciptool(1M)	Retrieves CIP information about local and remote nodes in the cluster.

File format

File	Format
cip.cf(4)	CIP configuration file format

B.4 Operator Intervention

System administrator

Command	Function
clreply(1M)	Responds to an operator intervention request message.

B.5 PAS

System administrator

Command	Function
mipcstat(1M)	MIPC statistics
clmstat(1M)	CLM statistics

B.6 Cluster Resource Management Facility

System administrator

Command	Function
clautoconfig(1M)	Executes automatic resource registration.
clbackuprdb(1M)	Saves the resource database.
clexec(1M)	Executes a remote command.
cldeldevice(1M)	Deletes a resource that was registered during automatic resource registration.
cldelrsc(1M)	Deletes a network interface card resource that was created with the automatic resource registration facility.
clinitscript(1M)	Reports the connection confirmation results for shared disk units.
clrestorerdb(1M)	Restores the resource database.
clsetacfparam(1M)	Checks the connections of shared disk units and sets up the operation for automatic resource registration.
clsetparam(1M)	Displays and changes the operation environment for a resource database, failed resource identification, or operation intervention request.
clsetrsc(1M)	Changes the takeover IP address.
clsetup(1M)	Sets up the resource database.
clspconfig(1M)	Sets up the operation of patrol diagnosis.
clsptl(1M)	Executes patrol diagnosis.
clstartrsc(1M)	Activates a resource (GDS only).
clstoprsc(1M)	Deactivates a resource (GDS only).
clsyncfile(1M)	Distributes a file between cluster nodes.

GDS: Global Disk Services

User command

Point

There is also a clgettree command in the Web-Based System Administration tool WSA.

Command	Function
clgetrsc(1)	Acquires takeover IP address.

Command	Function
clgettree(1)	Outputs tree information for the resource database.

B.7 RMS

System administrator

Command	Function
hvassert(1M)	Asserts (tests for) an RMS resource state.
hvcm(1M)	Starts the RMS configuration monitor.
hvconfig(1M)	Displays or saves the RMS configuration file.
hvdisp(1M)	Displays RMS resource information.
hvdispall(1M)	Displays RMS resource information on all nodes.
hvdump(1M)	Collects debugging information about RMS.
hvlogclean(1M)	Cleans the RMS log files.
hvshut(1M)	Shuts down RMS.
hvswitch(1M)	Switches control of an RMS user application or resource to another host.
hvutil(1M)	Manipulates the availability of an RMS resource.

File format

File	Format
hvenv.local(4)	RMS local environment variables file

B.8 Shutdown Facility (SF)

System administrator

Command	Function
sdtool(1M)	Interface tool for shutdown daemon
rcsd(1M)	Shutdown daemon for shutdown manager
clrcimonctl(1M)	Displays, starts, stops, and restarts the operation of the RCI asynchronous monitoring daemon.
clrccumonctl(1M)	Displays, starts, stops, and restarts the operation of the console asynchronous monitoring daemon.
clrccusetup(1M)	Registers, changes, deletes, or displays console information

File format

File	Format
rcsd.cfg(4M)	Configuration file for shutdown daemon
SA_wtinps.cfg(4M)	Configuration file for WTI NPS shutdown agent
SA_sunF.cfg(4M)	Configuration file for sunF system controller shutdown agent
SA_sspint.cfg(4M)	Configuration file for Sun E10000 shutdown agent
SA_rps.cfg(4M)	Configuration file for shutdown agent that uses remote power switch (RPS)
SA_rccu.cfg(4M)	Configuration file for XSFC shutdown agent

B.9 SIS

System administrator

Command	Function
dtcpadmin(1)	Starts the SIS administration utility.
dtcpd(1)	Starts the SIS daemon for configuring VIPs.
dtcpdbg(1)	Displays SIS debugging information.
dtcpstat(1)	Displays state information on SIS.

B.10 Tracing Failed Resource

System administrator

Command	Function
cldispfaultrsc(1M)	Outputs a list of the current failed resources.

B.11 Web-Based Admin View

System administrator

Command	Function
fjsvwvbs(1M)	Stops Web-Based Admin View.
fjsvwvcnf(1M)	Starts, stops, or restarts the web server for Web-Based Admin View.
wgcnfclient(1M)	Sets and displays the RMS configuration name.
wvCntl(1M)	Starts, stops, or gets debugging information for Web-Based Admin View.
wvGetparam(1M)	Displays the Web-Based Admin View environment variables.
wvSetparam(1M)	Sets the Web-Based Admin View environment variables.
wvstat(1M)	Displays the operating state of Web-Based Admin View.

B.12 Procedure Resource

System administrator

Command	Function
claddprocrsc(1M)	Registers an application resource that uses a state transition procedure.
cldelproc(1M)	Deletes a state transition procedure.
cldelprocrsc(1M)	Deletes an application resource that uses state transition procedure.
clgetproc(1M)	Gets a state transition procedure.
clsetproc(1M)	Registers a state transition procedure.
clsetprocrsc(1M)	Changes the registered information of an application resource that uses a state transition procedure.

User command

Command	Function
cldspproc(1)	Outputs information on the resource that uses the state transition procedure.

B.13 Process Monitoring Feature

System administrator

Command	Function
clmonproc(1M)	Requests the process monitoring facility to monitor a process.

B.14 RMS Wizards

RMS Wizards and RMS Application Wizard

The RMS Wizard manual will be saved in the following directory when the SMAWRhvdo package is installed.

```
/usr/opt/reliant/htdocs.solaris/wizards.en
```

Appendix C Troubleshooting

This appendix explains how to collect troubleshooting information if an error occurs in the PRIMECLUSTER system.

C.1 Collecting Troubleshooting Information

If an error occurs in the PRIMECLUSTER system, collect the information required for the error investigation from all nodes that construct the cluster and the cluster management servers. Then, contact your customer support representative.

1. PRIMECLUSTER investigation information

    - Use fjsnap or FJQSS to collect information.

    - Retrieve the system dump.

      If a hardware error, an operating system error, a panic, or a login failure occurs, collect the following:

        - System dump

        - XSCF log (only in SPARC M10)

    - Collect the Java Console on the clients.

      See "Appendix B.2.2 Java console" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

    - Collect a hard copy of the client screens.

      See "Appendix B.2.3 Screen hard copy" in the "*PRIMECLUSTER Web-Based Admin View Operation Guide*."

2. Investigation information for the failed application

3. Error reproduction procedure description if the error can be reproduced

Information

- When reporting a problem, collect the information required for an error investigation. If you do not provide information for problem checking and error reproduction execution, it may take a long time to reproduce and diagnose the problem or it may become impossible to do so.

- Collect investigation material promptly from all nodes of the PRIMECLUSTER system. Necessary information may become lost if a long time elapses after the error occurs. This applies especially to information collected by fjsnap or FJQSS.

- If a node was forcibly terminated, execute sync in OBP mode, and collect a system dump.

C.1.1 Executing the fjsnap Command

The "fjsnap" command is a system information tool of the Solaris, and it comes with the FJSVsnap Expanded Support Function Package. This tool enables you to collect information for troubleshooting if a failure occurs in the PRIMECLUSTER system.

The procedure for executing the fjsnap command is described below.

1. Log in with system administrator authority.

2. Execute the "fjsnap" command.

   /opt/FJSVsnap/bin/fjsnap -h *output*

   or

   /opt/FJSVsnap/bin/fjsnap -a *output*

    - If -a is specified, the amount of data becomes large because all detailed information is collected. If -h is specified, only cluster control information is collected.

    - Specify a special file name or an output file name (ex: /dev/rmt/0) to output information. The specified name is the file to which collected information is output when the "fjsnap" command is executed.

- When executing the fjsnap command while some nodes which constitute a cluster has been stopped, the following message may be output to the switchlog and the /var/adm/messages file. However, no corrective action is required.
  (BM, 8) Failed sending message <message> to object <object> on host <host>.
  (WRP, 11) Message send failed, queue id <queueid>, process <process>, <name>, to host <node>.

See

For details on the "fjsnap" command, see the "README" file, which is the README file included in the "FJSVsnap" package.

Information

Execution timings for the fjsnap command

- For problems that occur during operation, for example, if an error message is output, execute the "fjsnap" command immediately after the problem occurs.
- If the "fjsnap" command cannot be executed because the system hangs, collect a system dump. Then start the system in single user mode, and execute the "fjsnap" command.
  To collect the system dump, input the abort key sequence (for example, Break signal) to forcibly stop the node to OBP mode, and then execute "sync." For detailed instructions on forcibly stopping the node to OBP mode, see the "*System Administration Guide*" of the Solaris.
- After an error occurs, if a node restarts automatically (the node could not be started in single-user mode) or if the node is mistakenly started in multi-user mode, execute the "fjsnap" command.
- If investigation information cannot be collected because the "fjsnap" command results in an error or the "fjsnap" command does not return, then collect a system dump.

C.1.2 Collecting Information by FJQSS(Information Collection Tool)

The procedure is as follows.

1. Execute the following command.

```
# /opt/FJSVqstl/fjqss_collect
```

2. The product selection menu appears. Input the number of the product of which you want to collect the information, then press the [Y] key.

   Select from the following product numbers:

   - PRIMECLUSTER Enterprise Edition
   - PRIMECLUSTER HA Server
   - PRIMECLUSTER Clustering Base
   - PRIMECLUSTER Lite Pack

   If GDS and GLS are installed, and the above selection is performed, the information for investigation of PRIMECLUSTER including those products will be collected at a time.

3. Press the [Y] key according to the instruction in the prompt.
4. After the FJQSS has completed the collection, the name of the output directory of the collected information appears. Verify that the information have been collected in the directory.
5. The following file is created in the output directory of the collected information. Please send it to field engineers.

   result*YYYYMMDDHHMMSS*.tar.gz

   (*YYYYMMDDHHMMSS*: time (year, month, day, hour, minute, and second) that the collection started)

See

About FJQSS (Information Collection Tool) and its usage

You can collect the information necessary for the trouble investigation with FJQSS (Information Collection Tool). See the FJQSS User's Guide bundled to the installation medium of the product.

When you see the FJQSS User's Guide, open the following file in the installation medium of the product by the browser.

Documentation/fjqss-manual_sollnx/index_en.html

C.1.3 System Dump

If a system dump is collected when a node is in a panic, obtain the system dump as investigation information. The system dump is written into the dump file during node startup. For details on a system dump, see the "*System Administration Guide*" of the Solaris.

C.1.4 XSCF Log

Collect an XSCF log if conditions like those listed below occur in the shutdown facility of the SPARC M10 environment.

- 7240 and 7241

How to collect an XSCF log

For how to collect an XSCF log, see "SPARC M10 Systems System Operation and Administration Guide."

C.1.5 Core Dump of a Kernel Zone

In Kernel Zones environment, a forcible stop is performed by both the KZONE Panic shutdown agent and the KZONE Reset shutdown agent. A crush dump is output by only the KZONE Panic shutdown agent.

However, by setting a core dump output as necessary, a core dump of a Kernel Zone can be collected by the KZONE Reset shutdown agent.

This section explains how to set a core dump output.

Note

- Setting a core dump output is optional.
- If the core dump output is set, the KZONE Reset shutdown agent waits until the collection of the core dump, and then it stops the Kernel Zone forcibly. Therefore, the required time for a failover that causes a node error and a double fault of a resource may be long.
- The required time for outputting a core dump and for executing the halt subcommand of the zoneadm command depend on the I/O performance of a disk and memory usage.
  Therefore, calculate the time after completing a construction of the cluster application, as well as letting the cluster application operated.

For how to set the core dump output in the Kernel Zone, perform the following procedure on all nodes which constitute a cluster.

1. Stopping the shutdown facility

   Execute the following command to stop the shutdown facility.

   ```
   # /opt/SMAW/SMAWsf/bin/sdtool -e
   ```

2. Changing the timeout value for the KZONE Reset shutdown agent (SA_kzoner)

   Edit the rcsd.cfg file using an editor such as vi.

   ```
   # vi /etc/opt/SMAW/SMAWsf/rcsd.cfg
   ```

   ```
   CFNameX,weight=weight,admIP=myadmIP:
   agent=SA_kzonep,timeout=p_timeout:agent=SA_kzoner,timeout=r_timeout:agent=SA_kzchkhost,timeout=
   c_timeout
   ```

```
CFNameX,weight=weight,admIP=myadmIP:
agent=SA_kzonep,timeout=p_timeout:agent=SA_kzoner,timeout=r_timeout:agent=SA_kzchkhost,timeout=
c_timeout
```

r_timeout: Specifies a timeout period (in seconds) for the KZONE Reset shutdown agent.

Calculate the value using the following calculation formula.

```
r_timeout = Required time for creating a file with the memory size set in the Kernel Zone +
            Required time for executing the halt subcommand of the zoneadm command + 20 seconds
```

* The minimum value of the r_timeout value is default 70 seconds. If the calculation result for the r_timeout value is less than 70 seconds, use the default value of 70 seconds.

Example

When changing the timeout period for the KZONE Reset shutdown agent from 70 seconds to 100 seconds

- Before edit

```
CFNameX,weight=weight,admIP=myadmIP:
agent=SA_kzonep,timeout=45:agent=SA_kzoner,timeout=70:agent=SA_kzchkhost,timeout=20
CFNameX,weight=weight,admIP=myadmIP:
agent=SA_kzonep,timeout=45:agent=SA_kzoner,timeout=70:agent=SA_kzchkhost,timeout=20
```

- After edit

```
CFNameX,weight=weight,admIP=myadmIP:
agent=SA_kzonep,timeout=45:agent=SA_kzoner,timeout=100:agent=SA_kzchkhost,timeout=20
CFNameX,weight=weight,admIP=myadmIP:
agent=SA_kzonep,timeout=45:agent=SA_kzoner,timeout=100:agent=SA_kzchkhost,timeout=20
```

Calculate each processing time on the global zone host where the Kernel Zone operates is as follows.

In the following example, the timeout value is 70 + 10 + 20 = 100 seconds.

Example: The required time for creating a file with the memory size set in the Kernel Zone and for executing the halt subcommand of the zoneadm command when the memory size of the Kernel Zone kzone1 is 8 GB.

```
# /usr/bin/time mkfile 8G /var/crash/tmpfile
real       70.0
user        0.0
sys         0.0
# /usr/bin/time zoneadm -z kzone1 halt
real       10.0
user        0.0
sys         0.0
#
```

The memory size for the Kernel Zone can be checked with the zonecfg command as follows.

```
# zonecfg -z kzone1 info
zonename: kzone1
brand: solaris-kz
...
capped-memory:
     physical: 8G
...
#
```

3. Creating a file for enabling the core dump collection function

   Execute the following command to enable the core dump collection function.

```
# touch /etc/opt/SMAW/SMAWsf/enablekzcore.cfg
# chmod 600 /etc/opt/SMAW/SMAWsf/enablekzcore.cfg
```

* The core dump is output to the following directory on the global zone host.

```
/var/crash/<Kernel Zone name>.<Execution time for the savecore subcommand(YYYYMMDDHHMMSS
format)>.kzcore
```

Example: When the core dump of kzone1 was output at 12:34:56 January 1st, 2015

```
/var/crash/kzone1.20150101123456.kzcore
```

4. Starting the shutdown facility

   Execute the following command to start the shutdown facility.

```
# /opt/SMAW/SMAWsf/bin/sdtool -b
```

C.2 Troubleshooting

C.2.1 GUI in General

This section explains how to take corrective actions for problems that may occur while you are using the GUI.

If you find no relevant descriptions in this chapter, see also "B.1 Corrective action" in the "*PRIMECLUSTER Web-Based Admin Operation Guide*."

Symptom 1: When disk class creation, deletion, or name change is executed on the GDS screen, the message, "An error occurred in the log monitoring facility in node XXXX. Monitoring stops. After confirming that the node is active, update to the latest information." or "0007 The connection to the management server is disconnected. Attempt to connect to the active management server?" is displayed. Web-Based Admin View is disconnected.

Corrective action

This error might be due to a JavaVM related problem. Close the browser, and then display the screen again.

If you are using the Cluster Admin screen, close the screen first, and create or delete a disk class or change the disk class name to avoid this error. If the same error occurs frequently even after closing the Cluster Admin screen, contact field engineers.

Symptom 2: If you operate the userApplication Configuration Wizard while the registration of cluster application and resource configuration is in progress (while 0805 or 0813 dialog box is displayed), the message, "0880 A non-classified error occurred." appears.

Corrective action

If this phenomenon occurs, respond to the message, exit the userApplication Configuration Wizard and then restart it. After restarting, re-execute the operation that was being made before the 0880 error message was displayed.

C.3 Detecting a Failed Resource

If a failure occurs in a resource, you can specify the resource by referring to the following:

- The message displayed if a failure occurs in the resource
- Resource Fault History
- Fault Resource List

Note

To use the history function of the failed resource, the resource database must be set up correctly. Also, the "AutoStartUp" and "PersistentFault" attributes of userApplication must be set to Yes(1). For information on the resource database settings, see the "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

To use the detection function of the failed resources, you must enable an operator intervention request. For information on the use of the operator intervention request, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

The operator intervention function and the failed resource history function are both dependent on the "clwatchlogd" daemon. This daemon can be started automatically with the "rc" script in multi-user mode. The "clwatchlogd" daemon uses the "RELIANT_LOG_PATH" environment variable of RMS. The value of this variable is set when the "rc" script starts up for the first time.

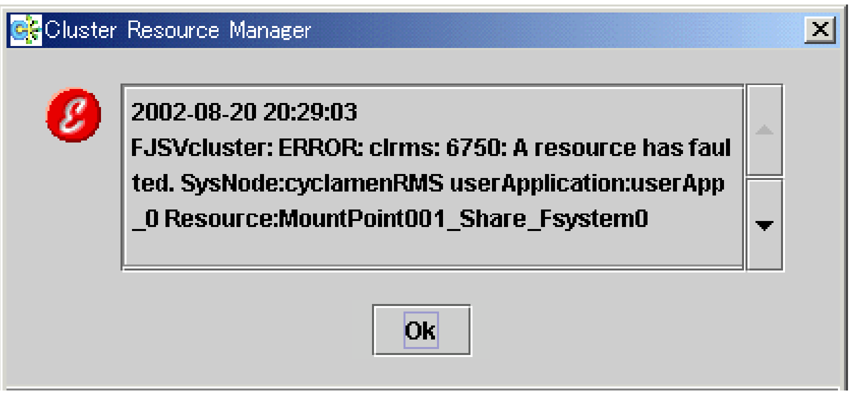
When this value is changed, you need to send the "SIGHUP" signal to clwatchlogd. When clwatchlogd receives this signal, clwatchlogd acquires the latest value of RELIANT_LOG_PATH. After completing the above processing, start RMS.

This manual is installed in the /etc/opt/FJSVcluster/man directory.

Before executing the "man (1)" command, add this directory to the beginning of MANPATH. Usually, a directory name is added to the line beginning with "setenv MANPATH" within the ".cshrc" file or the line beginning with "export MANPATH" within the ".profile" file.

C.3.1 Failed Resource Message

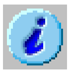
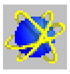
If a resource failure or recovery is detected, a message screen pops up as the Cluster Admin popup screen. An example of failed resource message screen is shown below:



Note

If a message frame title says "Cluster resource management facility," see "3.2 CRM View Messages" and "Chapter 4 FJSVcluster Format Messages" in the "PRIMECLUSTER Messages."

The severity icon is defined as follows:

Icon	Meaning
	Notice
	Warning
	Error
	Other

Respond to the failed resource message screen as follows:

Procedure

1. Click on the *OK* button to respond to the message.
2. Click the up arrow mark or down arrow mark to go to the previous or next message. Then, a message appears to remind you that you have not yet entered a response or confirmed the displayed message.

If you subsequently enter a response, the message is cleared and the next message appears. If the next message does not appear and the message prior to that for which a response was entered is still available, the previous message will appear. If there is any message for which confirmation or a response has not yet been entered, the message screen closes. For information on the message contents, refer to "Failed Resource and Operator Intervention Messages" and for information on how to display previous messages, refer to "Resource Fault History."

Note

If you close Web-Based Admin View or Cluster Admin after this message is displayed, a fault resource message with the same contents will not be displayed. Therefore, you are recommended to confirm the message contents if a fault resource message is displayed for the first time. After you have closed the message, refer to the fault history on the "Resource Fault History" screen. For information on the message display language, refer to "4.2.3.3 Setting the Web-Based Admin View Language."

If the Cluster Admin screen is not displayed on the client PC when the fault resource message is displayed, the message is transmitted only to the client to which the management server was first connected.

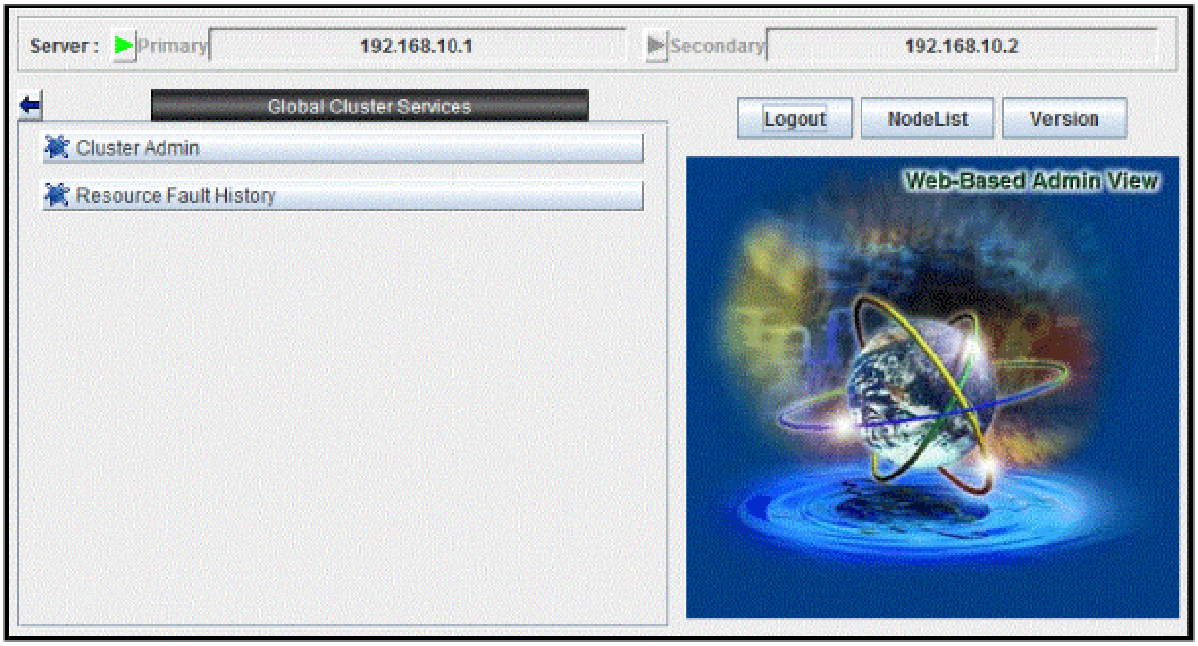
Each management server administers its fault resource messages. If you change the management server after confirming the message, the same message will be displayed again. To delete these messages, select *Cluster Admin* by using the GUI of *Web-Based Admin View* after closing *Cluster Admin*, and then open *Cluster Admin* again.

C.3.2 Resource Fault History

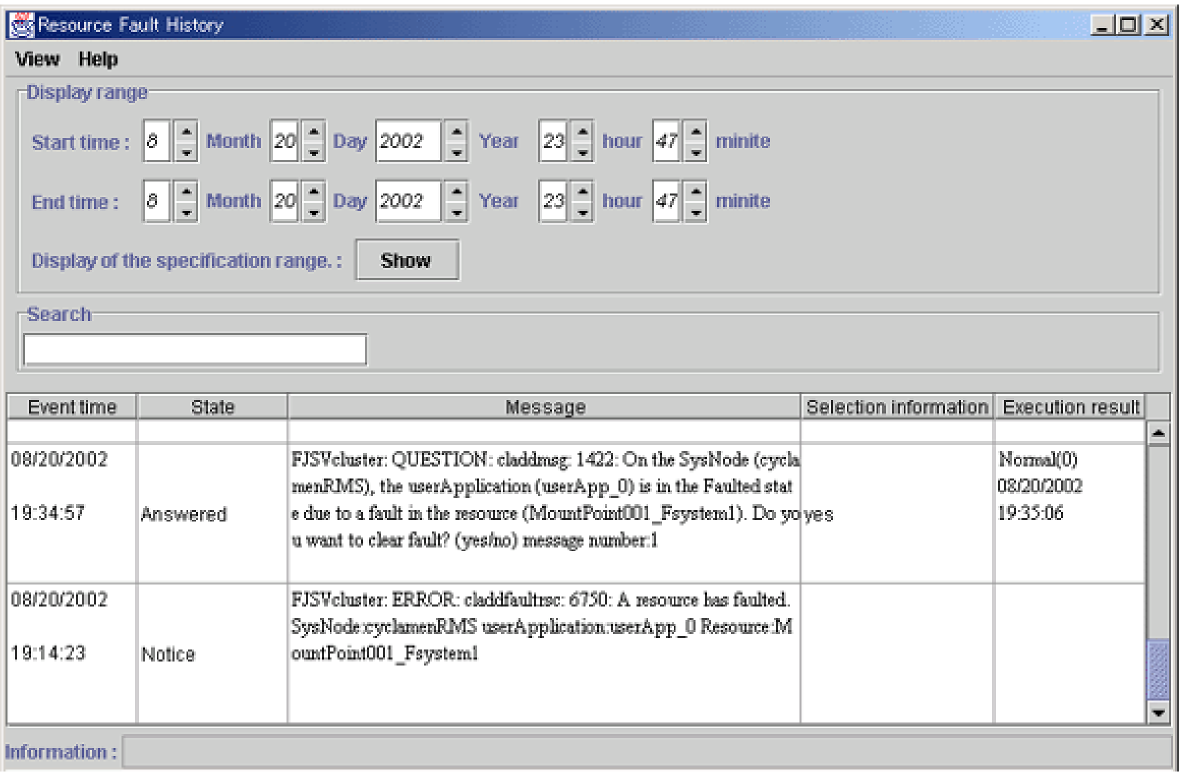
Display the "Resource Fault History" screen, in which the resource fault history is displayed, in the following procedure.

Procedure

1. Open the "Web-Based Admin View" screen and then select *Global Cluster Services*.
2. Choose *Resource Fault History*.



The "Resource Fault History" will be displayed.



Note

The "Resource Fault History" cannot be displayed automatically. To display the latest history information, select *View* -> *Update* menu.

Menu of the fault resource list screen

The "Resource Fault History" screen contains the following menu items:

Menu	Function
View -> Update latest information	The duration is initialized to the present time and date. A maximum of 100 of the latest history resources are displayed.
View -> Fault Resource List	A list of resources in which failures are present is displayed (see "C.3.3 Fault Resource List").
View -> Exit	The "Resource Fault History" screen is cleared.
Help -> Help	The GUI help screen is displayed.

Setting the range of time

A fault resource history listing can be displayed by specifying a date and time.

- *Start time* - A start time is set up.
- *End time* - An end time is set up.

If you click the *View* button after setting up the required values, a maximum of 100 of the most recently failed resources within the specifiable range can be displayed.

Search with a keyword

The fault resource history list can be narrowed by specifying "*Keyword*".

If a duration is set, the history of up to the 100 latest failed resources that satisfy both conditions can be displayed.

How to read the list

The following information is displayed on the "Resource Fault History" screen.

- Event time - The time at which the RMS detected a resource failure is displayed.
- State - One of the following statuses is indicated.
  - Responded - The operator has already responded the message.
  - Not responded - The operator has not responded to the message for which a response is required.
  - Responding - The operator is currently responding to the message.
  - Confirm - Notification message for which no response is required.
- Message - The message is displayed.
- Selection information - Operator intervention message information from the client that is connected to the management server is displayed. If the message is canceled or if a response to the message is entered by executing the "clreply(1M)" command, nothing will be displayed.
- Execution result - The result and time of the response processing are displayed.

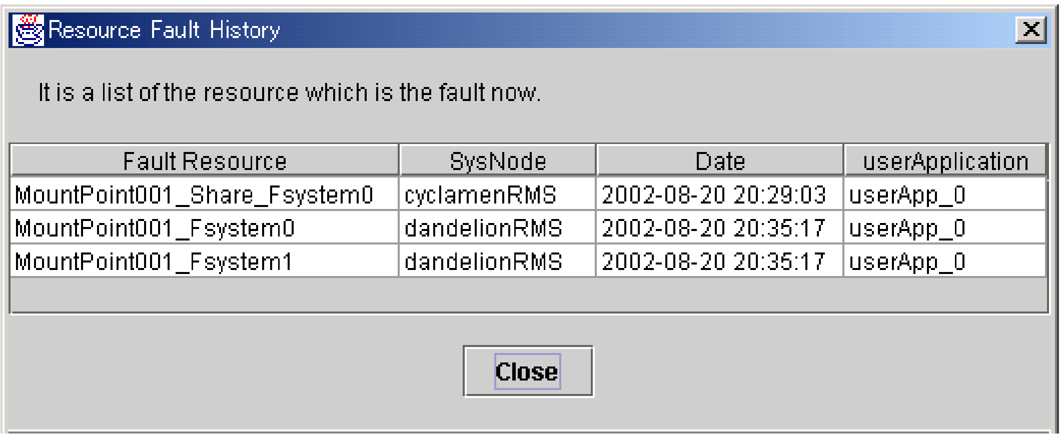
Information field

The information related to error detection during the acquisition or read-in of the history files is displayed. The following items will be displayed:

- Processing - History data is being collected from the management server.
- An error is included in the specified duration. - The specified duration is incorrect. Correct it and then click the *View* button.
- Part of the data acquisition failed. - Parts of the history files could be damaged. This will not disrupt ongoing operation, but the corrupted data will not be displayed.

C.3.3 Fault Resource List

If you select *View* -> *Fault Resource List* on the "Resource Fault History" screen, the fault resource list is displayed as follows:



The following information is displayed on the fault resource list screen:

- Fault Resource - The name of the fault resource is displayed.
- SysNode - The SysNode to which the fault resource belongs is displayed.
- Date - The date and time at which the RMS detected a resource failure are displayed.

- userApplication - The name of userApplication to which the failed resource belongs is displayed.

If you click *Close*, the failed resource list screen is closed.

The list of resources that are currently in the failed state can be displayed by executing the "cldispfaultrsc(1M)" command.

C.4 PRIMECLUSTER Log Files

This appendix describes PRIMECLUSTER log files.

C.4.1 Output Destination for core Files

When each process of PRIMECLUSTER is finished abnormally, core is output in the following directories.

The output destination cannot be changed.

/

/var/opt/FJSVcluster/cores/FJSVcldev/devsnmpd

/var/opt/FJSVcluster/cores/FJSVcldev/devrccud

/var/opt/FJSVcluster/cores/FJSVcldev/devrccuxd

/var/opt/FJSVcluster/cores/FJSVcldev/devscfd

/var/opt/FJSVcluster/cores/FJSVcldev/devscfmond

/var/opt/FJSVcluster/cores/dcmevmd

/var/opt/FJSVwvbs/logs/node

/var/opt/FJSVwvbs/logs/server

/var/opt/FJSVwvcnf

/var/opt/SMAWsf/log

/opt/SMAW/SMAWRrms

Current directory (command)

The core of the following processes are output to each directory:

Under /

prmd
clrmd
rcsd_monitor
devmalogd
cfregd

Under /var/opt/FJSVcluster/cores/FJSVcldev/devsnmpd

devsnmpd

Under /var/opt/FJSVcluster/cores/FJSVcldev/devrccud

devrccud

Under /var/opt/FJSVcluster/cores/FJSVcldev/devrccuxd

devrccuxd

Under /var/opt/FJSVcluster/cores/FJSVcldev/devscfd

devscfd

Under /var/opt/FJSVcluster/cores/FJSVcldev/devscfmond

devscfmond

Under /var/opt/FJSVcluster/cores/dcmevmd

dcmevmd
dcmmond
dcmmstd
dcmevmd
dcmfcpd
dcmsynd
dcmprcd
dcmcfmd
dcmdbud
dcmcomd
dcmdbcd
dcmlckd
clwatchlogd

Under /var/opt/FJSVwvbs/logs/node

wvAgent

Under /var/opt/FJSVwvbs/logs/server

java

Under /var/opt/FJSVwvcnf

wvcnfd

Under /var/opt/SMAWsf/log

rcsd

Under /opt/SMAW/SMAWRrms

bm
hvdet_system
hvdet_gmount
hvdet_icmp
hvdet_prmd
hvdet_execproc

C.4.2 core File Configuration

C.4.2.1 core Files Output

Core files are not output due to errors of applications, daemons, and commands.

To identify the cause when an error occurs, be sure to set core files to be output.

To output core files, change /etc/profile as follows:

</etc/profile>

[Before change]

```
ulimit -S -c 0 > /dev/null 2>&1
```

[After change]

```
ulimit -S -c unlimited > /dev/null 2>&1
```

C.4.2.2 Setting Output Destination for core Files

The default value of the current directories started with the OnlineScript of PRIMECLUSTER is /opt/SMAW/SMAWRrms (the default value of an environment variable RELIANT_PATH).

In a system environment where core files are set to be output, if an error of the application started via the OnlineScript occurs, the core files to be output are written under /opt.

If large number of core files are output under /opt, it weighs on the /opt file system. As a result, a double operation may not be performed because the necessary information for operating PRIMECLUSTER cannot be written, or PRIMECLUSTER may not be started or switched. To avoid this, change the current directory to an appropriate directory with one of the following methods:

- Changing the current directory in the head of the OnlineScript
- Changing the current directory within an application

Check files under /opt periodically and if core files exist, mode them to other directory not to weigh on the /opt file system.

C.4.3 Log Volume When Changing Log Levels

Changing log levels allows RMS to investigate details of an error.

When log levels are changed, the volume of dynamic disk resources required for PRIMECLUSTER is increased.

When changing log levels (maximum value of the log level 0), the log volume increased per day is as follows:

Calculation formula for increased log volume per day

(number of nodes x 80) + (number of registered resources x 25) + 25 = log volume increased per day (unit MB)

Note

Increased log volume varies depending on the system operation state. It is an approximated value.

For the actual increased system volume, check the increased movement of log volume under RELIANT_LOG_PATH.

C.4.4 Rotation and Deletion of RMS Log Files

RMS follows the following RMS environment variables, rotate and delete RMS log files:

- RELIANT_LOG_LIFE
- HV_LOG_ACTION_THRESHOLD
- HV_LOG_WARN_THRESHOLD
- HV_LOG_ACTION

For the value of this environment variable, you can change it corresponding to the system requirement. For the meaning of each RMS environment variable, see "PRIMECLUSTER Reliant Monitor Services with Wizard Tools Configuration and Administration Guide."

Note

1. RMS log files are deleted by the setting of the RELIANT_LOG_LIFE. This process is executed by hvlogcron, which is activated by a cron.
   For notes and contents of hvlogcron, see "7.7 cron Processing."
2. When deleting RMS log files with RELIANT_LOG_LIFE setting, the log files that RMS is outputting are not deleted. In the operation that RMS is operated one day or more continuously and also in the operation to dispatch old log information, which had been created before the RELIANT_LOG_LIFE was created, from RMS log files and delete them, set the hvlogclean command to be executed once a day to the cron configuration.

Appendix D Using SynfinityCluster Products in PRIMECLUSTER

Target users:

Users who use SynfinityCluster products

This appendix describes the required information for applying a SynfinityCluster-related topic in a reference manual to PRIMECLUSTER. The information focuses on the following points:

- Terminology
- Manual series
- Building procedure

D.1 Terminology

The terms of SynfinityCluster and PRIMECLUSTER are as follows:

SynfinityCluster	PRIMECLUSTER
Cluster services	Cluster applications, userApplication
SynfinityCluster agent products	PRIMECLUSTER products
State Transition Procedure	Online/Offline scripts
Cluster Service Instance	None (Expressions such as operating node and standby node are used instead.)
Cluster Domain	None
Node	Node or Host
Failover	Switch Over (Switchover or Failover)
Private LAN	Interconnect
Standby Patrol	Patrol diagnosis
Public LAN Takeover	Network Takeover
Monitoring Function	RMS Monitoring Function
SynfinityLink	GLS (Global Link Services)
SynfinityDisk, SynfinityDisk/Global	GDS (Global Disk Services)
SynfinityFile, SynfinityFile/Global	GFS (Global File Services)

D.2 Manual Series

The manual series for SynfinityCluster and PRIMECLUSTER are shown below.

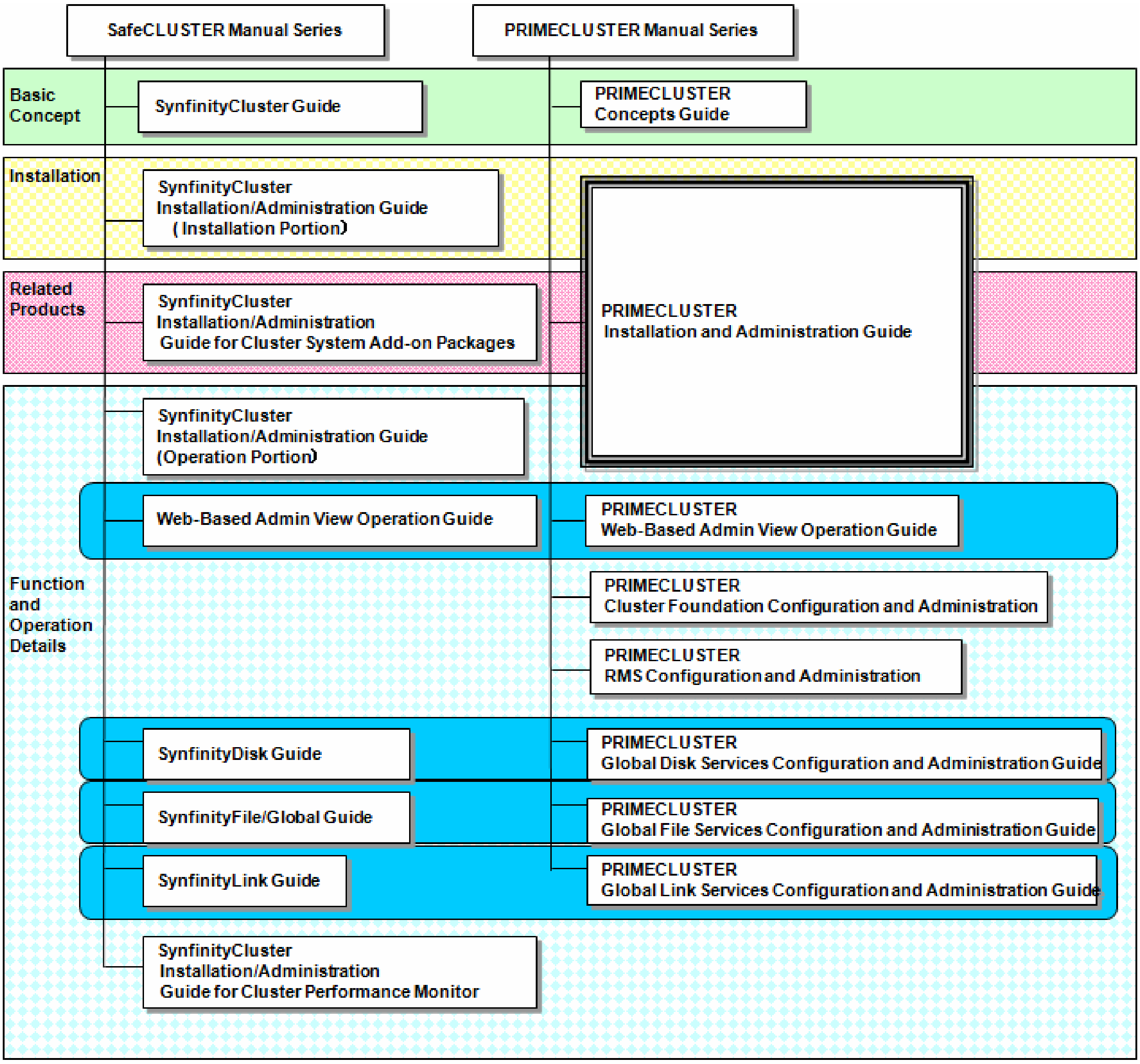


Table D.1 Manual list and manual uses

PRIMECLUSTER manual	Use	Reference
PRIMECLUSTER Installation/ Administration Guide (This Manual)	This manual is a general guide for all PRIMECLUSTER operations (installation, design, and maintenance). The manual also explains differences between PRIMECLUSTER and SynfinityCluster, the predecessor product, and information on PRIMECLUSTER products.	Installation - Design to maintenance
PRIMECLUSTER Concepts Guide	This manual allows readers to learn basic knowledge (technology, architecture, functions, etc.) related to each cluster system.	Design
PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide	This manual describes theory to setup information related to the Cluster Foundation (CF) functions.	Installation to operations
PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide	This manual describes theory to setup information related to the RMS functions.	Installation to operations

PRIMECLUSTER manual	Use	Reference
PRIMECLUSTER Web-Based Admin View Operation Guide	This manual describes theory to setup information related to Web-Based Admin View.	Installation to operations
PRIMECLUSTER Global Disk Services Configuration and Administration Guide	This manual describes information on Global Disk Services (GDS).	Installation to operations
PRIMECLUSTER Global File Services Configuration and Administration Guide	This manual describes information on Global File Services (GFS).	Installation to operations
PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function	This manual describes information related to the transmission path duplexing function of Global Link Services (GLS).	Installation to operations
PRIMECLUSTER Global Link Services Configuration and Administration Guide: Multipath Function	This manual describes information related to the multipath function of Global Link Services (GLS).	Installation to operations

D.3 Building Procedure

This section compares the PRIMECLUSTER operation procedures and the SynfinityCluster operation procedures.

D.3.1 Software Installation

The installation methods for the cluster products and related software are shown below.

Table D.2 Installation method

Operation	SynfinityCluster	PRIMECLUSTER
Installation using a custom jumpstart	Installation using custom Jumpstart	
Installation in each node	Installation in node units	CLI Installer (installation using an installation script)

D.3.2 Initial Cluster Setup

The initial setup procedures for the cluster products are shown below.

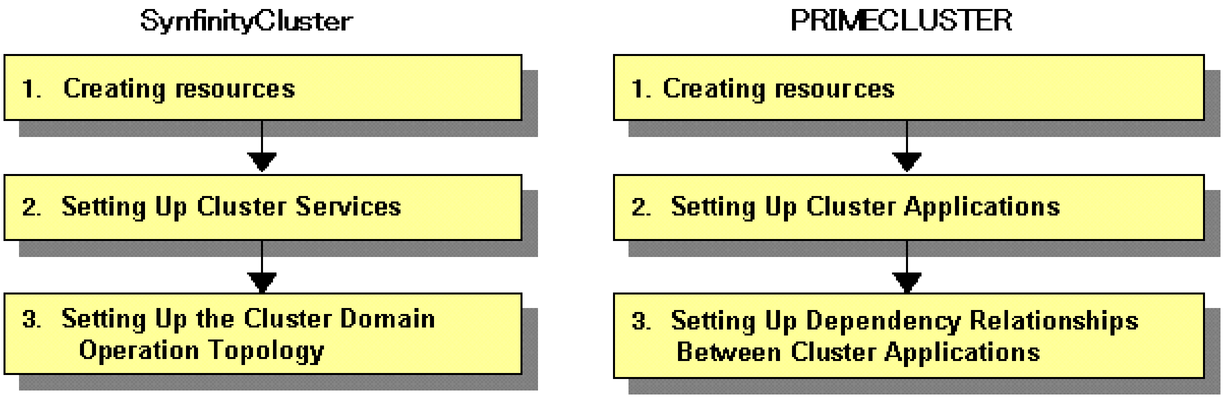
Table D.3 Initial setup procedure of the cluster

Operation	SynfinityCluster	PRIMECLUSTER
Initial setup of the cluster system	Setting up cluster configuration information and the resource database Cluster operation management view (GUI) *SynfinityCluster initial setup*	5.1 Initial Cluster Setup 5.1.3 Initial Setup of the Cluster Resource Management Facility
	Setting up cluster configuration information Setting up a private LAN	5.1.1 Setting Up CF and CIP (CF Wizard used)
	Setting up console asynchronous monitoring	5.1.2 Configuring the Shutdown Facility
	Setting up optional hardware resources	5.2 Setting Up Power Supply Linkage

Operation	SynfinityCluster	PRIMECLUSTER
	Setting up shared disk units	5.3 Setting Up Shared Disk Connection Confirmation
	Changing the operation environment for standby patrol	6.9 Setting Up Patrol Diagnosis

D.3.3 Application Setup

The setup procedures for operating applications with each product are shown below.



Setting up resources

Create the elements used by cluster services (cluster application in PRIMECLUSTER) as resources.

Operation	SynfinityCluster	PRIMECLUSTER
Setting up the takeover network	Set Cluster Operation GUI Setting up a takeover network	6.7.1.5 Creating Takeover Network Resources
Setting up state transition processing (script) of an application	Set Cluster Operation GUI Set State Transition Procedure of Set Application Resource	E.1 Registering a Procedure Resource 6.7.1.6 Creating Procedure Resources Note: This function is for making a SynfinityCluster product compatible for use in PRIMECLUSTER.
Using scripts and commands as resources	-	6.7.1.1 Creating Cmdline Resourcesing Cmdline Resources
Mount information for shared disks	Manual	6.7.1.2 Creating Fsystem Resources
Setting up shared disks	Set Cluster Operation GUI Shared resource units	6.7.1.3 Creating Gds Resources
Setting up high availability resources	Manual operation of SynfinityLink (CLI)	6.7.1.4 Creating Gls Resources

Setting up services

Services that operate in the cluster are provided for switchover units. These services are called "cluster services" in SynfinityCluster and "cluster applications" in PRIMECLUSTER.

Operation	SynfinityCluster	PRIMECLUSTER
Setting the application startup priority	Set Cluster Operation GUI Set Cluster Service - Set Application Start/Stop Priority menu items	Resource association
Setting a disk group	Set Cluster Operation GUI Set Cluster Service-Set Group menu items	GDS GUI
Setting the type of shared disk device	Set Cluster Operation GUI Set Cluster Service-Set Shared Disk Unit by Use menu items	6.7.1.3 Creating Gds Resources

GDS: Global Disk Services

Setting up Inter-service relationships

Operation	SynfinityCluster	PRIMECLUSTER
Setting a launch priority	Set Cluster Domain Operation Mode-Set Cluster Service Startup Priority menu items	-
Setting a low job priority	Set Cluster Domain Operation Mode-Set Cluster Service Application Priority/Exclusive Relationship menu items	-
Setting an exclusive relationship	Set Cluster Domain Operation Mode-Set Cluster Service Application Priority/Exclusive Relationship menu items	"userApplication Configuration Wizard" menu-Set Dependency Relationship Between userApplications-Set userApplication exclusion 6.7.3.1 Setting Exclusive Relationships Between Cluster Applications

D.3.4 Operations

The operations of each cluster product are shown below.

System state monitoring

Operation	SynfinityCluster	PRIMECLUSTER
Start the operation management screen	From the top menu of the cluster operation management view, select *SynfinityCluster* and then select *Cluster Operation Management*.	From the top menu of Web-Based Admin View, select *Global Cluster Services* and then select *Cluster Admin*.
Configuring the operation management screen	- Node view Display a hierarchical list of the resources (system resources) as seen from a node. - Service view Display a hierarchical configuration of the cluster resources.	- CF GUI - CRM main window (displays only resources related to hardware) - RMS main window - MSG main window

When the resource state changes

Operation	SynfinityCluster	PRIMECLUSTER
Checking resource errors	- Message - Node View - Service View	- Message - Resource Fault History screen - Failed resource list
Patrol diagnosis	1. Detection by standby patrol 2. Confirmation with messages	1. Detection by patrol diagnosis 2. Confirmation with messages and CRM main window

Appendix E Registering, Changing, and Deleting State Transition Procedure Resources for SynfinityCluster Compatibility

Operation target:

The procedures described here are necessary only if you will be using products that have been migrated from SynfinityCluster.

To use a procedure resource in a cluster application, you must register the procedure resource before setting up the cluster application.

This appendix explains how to register, change, and delete procedure resources.

E.1 Registering a Procedure Resource

This section explains how to register a procedure resource.

Take the following steps to register a procedure resource on all nodes where the procedure resource needs to be registered.

Operation Procedure:

1. Log in with the system administrator authority to the node in which the procedure resource is to be registered.
2. Execute the "clsetproc(1M)" command to register the state transition procedure.

   See

   For details on the "clsetproc(1M)" command, see the manual page.

   Example

   To register the "/var/tmp/program" state transition procedure as program (file name) to the BasicApplication class

   ```
   # /etc/opt/FJSVcluster/bin/clsetproc -c BasicApplication -m program /var/tmp/program
   ```

   Point

   To overwrite a state transition procedure that is already registered, specify the -o option.

3. Registering the procedure resource

   Execute the "claddprocrsc(1M)" command to register the procedure resource.

   See

   For details on the claddprocrsc(1M) command, see the manual page for claddprocrsc (1M).

   Example

   When registering a procedure resource, this procedure resource has to meet the following conditions;

   - The resource key of the procedure resource is SDISK,
   - The procedure name is program,
   - This procedure resource uses the state transition procedure registered in the BasicApplication class,

- The procedure resource is registered in the node (NODE1), and
- The state transition procedure receives the requests of START RUN AFTER and STOP RUN BEFORE

```
# /etc/opt/FJSVcluster/bin/claddprocrsc -k SDISK -m program -c BasicApplication -s NODE1 -K
AFTER -S BEFORE
```

E.2 Changing a Procedure Resource

This section explains how to change a procedure resource.

E.2.1 Changing a State Transition Procedure

Take the following steps to change the state transition procedure on all nodes where it needs to be changed.

Operation Procedure:

1. Log in with the system administrator authority to the node in which the state transition procedure is to be changed.
2. Execute the "clgetproc(1M)" command to retrieve the state transition procedure.

   See

   For details on the "clgetproc(1M)" command, see the manual page.

   Example

   When retrieving a state transition procedure, this procedure resource has to meet the following conditions;

   - The state transition procedure is retrieved to the "/var/tmp" directory.
   - The file name registered in the BasicApplication class is program

   ```
   # /etc/opt/FJSVcluster/bin/clgetproc -c BasicApplication -f /var/tmp/program program
   ```

3. Modifying the state transition procedure

   Using a text editor such as vi(1), modify the state transition procedure that was retrieved in Step 2.

4. Registering the state transition procedure

   For instruction on registering the procedure, see Step 2 of "E.1 Registering a Procedure Resource."

E.2.2 Changing the Startup Priority of a State Transition Procedure

Take the following steps to change the startup priority of a state transition procedure on all nodes where it needs to be changed.

Note

To change the startup priority of a state transition procedure, you need to delete a procedure resource with the procedure for changing a cluster application configuration and create the procedure resource again.

For more details, see "10.3.1 Changing the Cluster Application Configuration."

Operation Procedure:

1. Log in with the system administrator authority to the node in which the state transition procedure of the procedure resource is to be changed.

2. Execute the "clsetprocrsc(1M)" command to change the startup priority of the procedure resource.

See

For details on the "clsetprocrsc(1M)" command, see the manual page.

Example

When changing the startup priority of the procedure resource to 10000, this procedure resource has to meet the following conditions;

- The resource class registered in the node (NODE1) is the BasicApplication class.
- The resource name is SDISK.

```
# /etc/opt/FJSVcluster/bin/clsetprocrsc -n SDISK -c BasicApplication -s NODE1 -p 10000
```

E.2.3 Changing Registration Information of a Procedure Resource

Take the following steps to change the registration information of the procedure resource on all nodes where the state transition procedure needs to be changed.

Operation Procedure:

1. Log in with the system administrator authority to the node in which the state transition procedure of the procedure resource is to be changed.
2. Execute the "clsetprocrsc(1M)" command to change the registration information of the procedure resource.

See

For details on the "clsetprocrsc(1M)" command, see the manual page.

Example

When adding the requests of START RUN BEFORE state transition to the procedure resource that meet the following conditions;

- The resource key of the procedure resource is SDISK,
- This procedure resource uses the state transition procedure registered in the BasicApplication class,
- The procedure resource is registered in the node (NODE1), and
- The state transition procedure receives the requests of START RUN AFTER and STOP RUN BEFORE

```
# /etc/opt/FJSVcluster/bin/clsetprocrsc -n SDISK -c BasicApplication -s NODE1 -K BEFORE,AFTER
-S BEFORE
```

E.3 Deleting a Procedure Resource

Take the following steps to delete a procedure resource on all nodes where the procedure resource needs to be changed.

Operation Procedure:

1. Log in with the system administrator authority to the node from which the procedure resource is to be deleted.
2. Execute the "cldelprocrsc(1M)" command to delete the procedure resource.

See

For details on the "cldelprocrsc(1M)" command, see the manual page.

Example

When deleting a procedure resource, the procedure resource needs to meet the following conditions:

- The resource key of the procedure resource is SDISK,
- This procedure resource uses the state transition procedure registered in the BasicApplication class, and
- The node identification name is NODE1

```
# /etc/opt/FJSVcluster/bin/cldelprocrsc -n SDISK -c BasicApplication -s NODE1
```

3. Deleting the state transition procedure

   If a state transition procedure becomes unnecessary after all procedure resources that use that state transition procedure have been deleted, execute the cldelproc(1M) command to delete the state transition procedure.

See

For details on the "cldelproc(1M)" command, see the manual page.

Example

When deleting a procedure resource, the procedure resource needs to meet the following conditions;

- The procedure name is program, and
- This procedure resource uses the state transition procedure registered in the BasicApplication class

```
# /etc/opt/FJSVcluster/bin/cldelproc -c BasicApplication program
```

Appendix F SafeCLUSTER Compatibility Function Settings

This chapter explains the setup for when acquiring the case of status migration for cluster service that has used the clgettrigger command.

F.1 Setup Procedures for When Building Cluster Applications

1. Use the userApplication Configuration Wizard and create the cluster applications.
2. Login to one optional node using system administrator privileges.
3. Stop the RMS.

```
# hvshut -a
```

4. Start up the hvw.

   *For this procedure, the configuration definition file is assumed to be "config."

```
# hvw -n config
```

5. Select "4) Application-Edit."

```
banana: Main configuration menu, current configuration: config
No RMS active in the cluster
1) HELP                                  10) Configuration-Remove
2) QUIT                                  11) Configuration-Freeze
3) Application-Create                    12) Configuration-Thaw
4) Application-Edit                      13) Configuration-Edit-Global-Settings
5) Application-Remove                    14) Configuration-Consistency-Report
6) Application-Clone                     15) Configuration-ScriptExecution
7) Configuration-Generate                16) RMS-CreateMachine
8) Configuration-Activate                17) RMS-RemoveMachine
9) Configuration-Copy
Choose an action: 4
```

6. Select "4) OPTIONS."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
Application Name: 4
```

7. Select "5) ShowAllAvailableWizards."

```
Edit: selection criteria:
1) HELP
2) RETURN
3) NONE
4) ShowTurnkeyWizardsOnly
5) ShowAllAvailableWizards
6) ApplicationsOnly
7) SearchPattern
8) ShowSubApplications
9) FlagApplications
Set selection criteria: 5
```

8. Select the number for userApplication.

   Here the userApplication is set as "userApp_0," so select "7)."

```
Edit: Application selection menu:
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) Procedure0
6) Procedure1
7) userApp_0
Application Name: 7
```

9. Select the number for (PreCheckScript=).

   Here select "13" for "(PreCheckScript=)."

```
Settings of application type "generic" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=userApp_0
 5) AdditionalMachine
 6) AdditionalConsole
 7) AdditionalSubApplication
 8) Machines[0]=bananaRMS
 9) Machines[1]=grapeRMS
10) SubApplications[0]=Procedure0
11) SubApplications[1]=Procedure1
12) (HostSpecificSubApplication=no)
13) (PreCheckScript=)
14) (PreOnlineScript=)
15) (PostOnlineScript=)
16) (PreOfflineScript=)
17) (PostOfflineScript=)
18) (OfflineDoneScript=)
19) (FaultScript=)
20) (AutoStartUp=yes)
21) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
22) (PreserveState=no)
23) (PersistentFault=1)
24) (ShutdownPriority=)
25) (OnlinePriority=0)
26) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
27) (LicenseToKill=no)
28) (AutoBreak=yes)
29) (AutoBreakMaintMode=no)
30) (HaltFlag=yes)
31) (PartialCluster=0)
32) (ScriptTimeout=)
Choose the setting to process: 13
```

10. Select "4) FREECHOICE."

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming ONLINE: 4
```

11. On the bottom of the same screen ">>" is displayed, and one has become able to input the PreCheckScript script's file path.

    Specify /etc/opt/FJSVcluster/sys/clprechktrigger for the PreCheckScript script's file path.

```
1) HELP
2) RETURN
3) NONE
```

```
4) FREECHOICE
Enter the command line to start prior to the application becoming ONLINE: 4
       >> /etc/opt/FJSVcluster/sys/clprechktrigger
```

12. Select the number for (PostOnlineScript=).

    Here, select "15)" for "(PostOnlineScript=)."

```
Settings of application type "generic" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=userApp_0
 5) AdditionalMachine
 6) AdditionalConsole
 7) AdditionalSubApplication
 8) Machines[0]=bananaRMS
 9) Machines[1]=grapeRMS
10) SubApplications[0]=Procedure0
11) SubApplications[1]=Procedure1
12) (HostSpecificSubApplication=no)
13) (PreCheckScript='/etc/opt/FJSVcluster/sys/clprechktrigger')
14) (PreOnlineScript=)
15) (PostOnlineScript=)
16) (PreOfflineScript=)
17) (PostOfflineScript=)
18) (OfflineDoneScript=)
19) (FaultScript=)
20) (AutoStartUp=yes)
21) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
22) (PreserveState=no)
23) (PersistentFault=1)
24) (ShutdownPriority=)
25) (OnlinePriority=0)
26) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
27) (LicenseToKill=no)
28) (AutoBreak=yes)
29) (AutoBreakMaintMode=no)
30) (HaltFlag=yes)
31) (PartialCluster=0)
32) (ScriptTimeout=)
Choose the setting to process: 15
```

13. Select "4) FREECHOICE."

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start after the application became ONLINE: 4
```

14. At the bottom of the same screen, ">>" is displayed, and one has become able to input the PostOnlineScript script's file path.

    Specify /etc/opt/FJSVcluster/sys/clpostontrigger for the PostOnlineScript script's file path.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start after the application became ONLINE: 4
        >> /etc/opt/FJSVcluster/sys/clpostontrigger
```

15. Select the number for (PreOfflineScript=).

    Here, select "16)" for "PreOfflineScript=)."

```
Settings of application type "generic" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=userApp_0
 5) AdditionalMachine
 6) AdditionalConsole
 7) AdditionalSubApplication
 8) Machines[0]=bananaRMS
 9) Machines[1]=grapeRMS
10) SubApplications[0]=Procedure0
11) SubApplications[1]=Procedure1
12) (HostSpecificSubApplication=no)
13) (PreCheckScript='/etc/opt/FJSVcluster/sys/clprechktrigger')
14) (PreOnlineScript=)
15) (PostOnlineScript='/etc/opt/FJSVcluster/sys/clpostontrigger')
16) (PreOfflineScript=)
17) (PostOfflineScript=)
18) (OfflineDoneScript=)
19) (FaultScript=)
20) (AutoStartUp=yes)
21) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
22) (PreserveState=no)
23) (PersistentFault=1)
24) (ShutdownPriority=)
25) (OnlinePriority=0)
26) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
27) (LicenseToKill=no)
28) (AutoBreak=yes)
29) (AutoBreakMaintMode=no)
30) (HaltFlag=yes)
31) (PartialCluster=0)
32) (ScriptTimeout=)
Choose the setting to process: 16
```

16. Select "4) FREECHOICE."

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming OFFLINE: 4
```

17. At the bottom of the same screen, ">>" is displayed, and one has become able to input the PreOfflineScript script's file path.

    Specify /etc/opt/FJSVcluster/sys/clpreofftrigger for the PreOfflineScript script's file path.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming OFFLINE: 4
        >> /etc/opt/FJSVcluster/sys/clpreofftrigger
```

18. Select the number for (OfflineDoneScript=).

    Here, select "18)" for "(OfflineDoneScript=)."

```
Settings of application type "generic" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
```

```
 3) SAVE+EXIT
 4) ApplicationName=userApp_0
 5) AdditionalMachine
 6) AdditionalConsole
 7) AdditionalSubApplication
 8) Machines[0]=bananaRMS
 9) Machines[1]=grapeRMS
10) SubApplications[0]=Procedure0
11) SubApplications[1]=Procedure1
12) (HostSpecificSubApplication=no)
13) (PreCheckScript='/etc/opt/FJSVcluster/sys/clprechktrigger')
14) (PreOnlineScript=)
15) (PostOnlineScript='/etc/opt/FJSVcluster/sys/clpostontrigger')
16) (PreOfflineScript='/etc/opt/FJSVcluster/sys/clpreofftrigger')
17) (PostOfflineScript=)
18) (OfflineDoneScript=)
19) (FaultScript=)
20) (AutoStartUp=yes)
21) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
22) (PreserveState=no)
23) (PersistentFault=1)
24) (ShutdownPriority=)
25) (OnlinePriority=0)
26) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
27) (LicenseToKill=no)
28) (AutoBreak=yes)
29) (AutoBreakMaintMode=no)
30) (HaltFlag=yes)
31) (PartialCluster=0)
32) (ScriptTimeout=)
Choose the setting to process: 18
```

19. Select "4) FREECHOICE."

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start after the application became OFFLINE: 4
```

20. At the bottom of the same screen, ">>" is displayed, and one has become able to input the OfficeDoneScript script's file path.

    Specify /etc/opt/FJSVcluster/sys/clpostofftrigger for the OfflineDoneScript script's file path.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start after the application became OFFLINE: 4
      >> /etc/opt/FJSVcluster/sys/clpostofftrigger
```

21. Select "3) SAVE+EXIT."

```
Settings of application type "generic" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=userApp_0
 5) AdditionalMachine
 6) AdditionalConsole
 7) AdditionalSubApplication
 8) Machines[0]=bananaRMS
 9) Machines[1]=grapeRMS
10) SubApplications[0]=Procedure0
```

```
11) SubApplications[1]=Procedure1
12) (HostSpecificSubApplication=no)
13) (PreCheckScript='/etc/opt/FJSVcluster/sys/clprechktrigger')
14) (PreOnlineScript=)
15) (PostOnlineScript='/etc/opt/FJSVcluster/sys/clpostontrigger')
16) (PreOfflineScript='/etc/opt/FJSVcluster/sys/clpreofftrigger')
17) (PostOfflineScript=)
18) (OfflineDoneScript='/etc/opt/FJSVcluster/sys/clpostofftrigger')
19) (FaultScript=)
20) (AutoStartUp=yes)
21) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
22) (PreserveState=no)
23) (PersistentFault=1)
24) (ShutdownPriority=)
25) (OnlinePriority=0)
26) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
27) (LicenseToKill=no)
28) (AutoBreak=yes)
29) (AutoBreakMaintMode=no)
30) (HaltFlag=yes)
31) (PartialCluster=0)
32) (ScriptTimeout=)
Choose the setting to process: 3
```

22. Select "3) RETURN."

```
Edit: Application selection menu:
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) Procedure0
6) Procedure1
7) userApp_0
Application Name: 3
```

23. Select "2) QUIT" and end hvw.

```
banana: Main configuration menu, current configuration: config
No RMS active in the cluster
 1) HELP                                     10) Configuration-Remove
 2) QUIT                                     11) Configuration-Freeze
 3) Application-Create                       12) Configuration-Thaw
 4) Application-Edit                         13) Configuration-Edit-Global-Settings
 5) Application-Remove                       14) Configuration-Consistency-Report
 6) Application-Clone                        15) Configuration-ScriptExecution
 7) Configuration-Generate                   16) RMS-CreateMachine
 8) Configuration-Activate                   17) RMS-RemoveMachine
 9) Configuration-Copy
Choose an action: 2
```

24. Execute the following command on the node where hvw was terminated.

```
# /opt/SMAW/SMAWRrms/bin/hvw -xj -F Configuration-Activate
```

25. Write to the /etc/opt/FJSVcluster/etc/sc.d/clgettrigger.conf for all nodes the information for procedure resources using this function in the following format.

```
RID"RID of procedure resource"="userApplication name"
RID"RID of procedure resource"="userApplication name"      :
```

In the setup example provided in this procedure, RID-29,30 is registered to the appl's userApplication,whereas the RID=36,37 is registered to the app2's userApplication.

```
RID29=app1
RID30=app1
RID36=app2
RID37=app2
      :
```

26. Restart all nodes.

```
# /usr/sbin/shutdown -g0 -i6 -y
```

F.2 Setup Procedure for When Changing Cluster Application Configurations

If the configuration for a cluster application has been changed using the userApplication Configuration Wizard, re-perform the procedures for steps 2 to 26 of the "F.1 Setup Procedures for When Building Cluster Applications."

Note that if the RID for the procedure resource is not being changed, it is not necessary to perform step 26.

F.3 Setup Procedure for When Deleting Cluster Applications

If the cluster applications have been deleted using the userApplication Configuration Wizard, delete the information for the deleted userApplication from the /etc/opt/FJSVcluster/etc/sc.d/clgettrigger.conf for all nodes.

In the following setting example, deletion of the appl's userApplication is assumed.

```
RID36=app2
RID37=app2
      :
```

Appendix G Startup Scripts and Startup Daemons in PRIMECLUSTER

This appendix provides explanations on scripts and daemons that are started by PRIMECLUSTER.

G.1 Explanation Formats

Scripts and daemons are explained with the following formats:

Startup script

Name of startup script.

Function

Function of startup script and daemon.

Effect if stopped

Effect if startup script and daemon are stopped.

Startup daemon

Daemon started by startup script.

G.2 Startup Script Lists

/etc/inittab

inittab

Function

Basic part of GDS.

Effect if stopped

If GDS daemon terminates with an error, it will not be restarted automatically.

Startup daemon

/usr/sbin/sdxmond

/etc/rc2.d

S06sfdsk2

Function

Startup of GDS local classes in which iscsi devices are registered.

Effect if stopped

GDS local classes in which iscsi devices are registered cannot be used.

Startup daemon

None.

S99SMAWsfclean

Function

Deleting an existing lock file of the shutdown facility.

Effect if stopped

The shutdown facility cannot be started.

Startup daemon

None.

Appendix H SMF Services and Port Numbers in PRIMECLUSTER

This appendix provides explanations on SMF services and startup daemons for PRIMECLUSTER, and the port numbers being used.

H.1 Explanation Formats

SMF services and port numbers are explained with the following formats:

Name of service

Name of SMF service.

Function

Function of service.

Effect if stopped

Effect if service is stopped.

Dependency relationship with other services

dependency

Remote service required for starting own service.

dependent

Remote service requiring own service.

Startup daemon

Daemons started by service.

Utilized port

Port

Port number.

Protocol

Protocol - TCP or UDP.

Send/Receive

"s" if port sends data, "r" if it receives data "s, r" for both.

Network

Utilized network - any of Cluster interconnect, administrative LAN, or public LAN.

Target

Node that uses the port.

Communication target

Port

Port number of communication target.

Target

Node or device that uses the port of the communication target.

Remarks

Remarks.

H.2 SMF Service Lists

/network/fjsvmpnet (*1)

Function

mpnet instance activation.

Effect if stopped

Creation of LAN redundancy using the Multipath function is not available. (*2)

Dependency relationship with other services

dependency (*3)

/system/filesystem/local

dependent

None.

Startup daemon

/opt/FJSVmpnet/daemon/mpnetd or /opt/FJSVmpnet/daemon/mpnetpolld (*4)

Utilized port

None.

Remarks

(*1) This SMF service exists only in Solaris 10.

(*2) For using the Multipath function, it is necessary to enable the following services in addition to this service:
/network/fjsvmpnet-inet
/network/fjsvmpnet-inet6
/network/fjsvmpnet-stop

(*3) If disabled, the service cannot start.

(*4) One line monitoring daemon is started for each mpnet instance.

/network/fjsvmpnet-inet (*1)

Function

IPV4 interface activation.

Effect if stopped

Creation of LAN redundancy using the Multipath function is not available.

Dependency relationship with other services

dependency

/network/fjsvmpnet

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

(*1) This SMF service exists only in Solaris 10.

/network/fjsvmpnet-inet6 (*1)

Function

IPV6 interface activation.

Effect if stopped

Creation of LAN redundancy using the Multipath function is not available.

Dependency relationship with other services

dependency

/network/fjsvmpnet-inet

dependent

/system/sysidtool:net
/network/initial

Startup daemon

None.

Utilized port

None.

Remarks

(*1) This SMF service exists only in Solaris 10.

/network/fjsvmpnet-stop (*1)

Function

Stopping of daemons for Multipath function when system is stopped.

Effect if stopped

Daemons for Multipath function will not be stopped when system is stopped.

Dependency relationship with other services

dependency

None.

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

(*1) This SMF service exists only in Solaris 10.

/network/fjsvhanet

Function

Startup of daemons and activation of virtual interfaces.

Effect if stopped

Creation of LAN redundancy using the Redundant Line Control function is not available. (*1)

Dependency relationship with other services

dependency (*2)

/milestone/network
/system/filesystem/local (*3)

dependent

/network/initial
/network/routing-setup
/network/fjsvhanet-poll

Startup daemon

/opt/FJSVhanet/etc/sbin/hanetctld
/opt/FJSVhanet/etc/sbin/hanetptld (*4)

Utilized port (*5)

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
1807	UDP	s, r	Public LAN	Cluster node	1807	Remote node (GS or SURE)

Remarks

(*1) For using the Redundant Line Control function, it is necessary to enable the following service in addition to this service: /network/fjsvhanet-poll

(*2) If disabled, the service cannot start.

(*3) This dependency relationship does not exist in the following environments:

- Solaris 11 environment

- Solaris 10 environment where the PRIMECLUSTER patch T011644SP-01 or later is applied

(*4) Exists only when using the standby patrol function in exclusive-IP zones. Availability of startup and the number of processes rely on the configuration. Also, this may be suspended according to the monitoring status.

(*5) Port only used by GS/SURE linkage modes. It does not associate the service name.

/network/fjsvhanet-poll

Function

Startup of line monitoring.

Effect if stopped

Creation of LAN redundancy using the Redundant Line Control function is not available.

Dependency relationship with other services

dependency

/network/fjsvhanet (*1)
/network/service
/network/routing-setup
/system/zones

dependent

None.

Startup daemon

/opt/FJSVhanet/etc/sbin/hanetmond (*2)
/opt/FJSVhanet/etc/sbin/hanetselect (*3)

Utilized port

None.

Remarks

(*1) If disabled, the service cannot start.

(*2) Starts only when the self-checking function is used.

(*3) Availability of startup and the number of processes rely on the configuration. Also, this may be suspended according to the monitoring status.

/milestone/fjsvsdx

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Dependency relationship with other services

dependency

/milestone/devices
/milestone/network

dependent

/milestone/single-user

Startup daemon

/usr/sbin/sdxservd
/usr/sbin/sdxlogd
/usr/sbin/sdxexd

Utilized port

None.

Remarks

None.

/milestone/fjsvclapi

Function

Beginning of online tracing of the Cluster Resource Management facility (1).

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvclrms

Function

Beginning of online tracing of the Cluster Resource Management facility (2).

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvcldbm

Function

Startup of cluster configuration management facility.

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclapi
/milestone/fjsvclrms
/milestone/smawcf

dependent

None.

Startup daemon

/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmmond
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmmstd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmevmd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmfcpd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmsynd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmprcd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmcfmd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmdbud
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmcomd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmdbcd
/etc/opt/FJSVcluster/FJSVcldbm/daemons/dcmlckd
/etc/opt/FJSVcluster/FJSVclrms/daemons/clwatchlogd

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9331 (*1)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9379 (*2)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9378 (*3)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9377 (*4)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9376 (*5)	TCP	s, r	Interconnect	Cluster node	ANY	Local cluster node
9375 (*6)	TCP	s, r	Interconnect	Cluster node	ANY	Local cluster node

Remarks

When using the IP filtering function of Oracle Solaris, allow communications between all nodes.

(*1) No.9331 is set to support the service name "dcmcom."

(*2) No.9379 is set to support the service name "dcmsync."

(*3) No.9378 is set to support the service name "dcmlck."

(*4) No.9377 is set to support the service name "dcmfcp."

(*5) No.9376 is set to support the service name "dcmevm."

(*6) No.9375 is set to support the service name "dcmmst."

/milestone/fjsvclrmgr

Function

Startup of Cluster Resource Management facility (1).

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvcldbm

dependent

None.

Startup daemon

/etc/opt/FJSVcluster/FJSVcldbm/daemons/clrmd

Utilized port

None.

Remarks

None.

/milestone/fjsvclctrl

Function

Waiting for completion of startup of the Cluster Resource Management facility; basic part of GDS.

Effect if stopped

Clusters and functions of GDS cannot be used.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclrmgr

dependent

None.

Startup daemon

/usr/sbin/sdxcld
/usr/sbin/sdxclc
/usr/sbin/sdxcle

Utilized port

None.

Remarks

None.

/milestone/fjsvclrwz

Function

Setting of cluster applications.

Effect if stopped

Cluster applications cannot be configured correctly, or will not work correctly.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclctrl

dependent

/milestone/smawrrms

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvclprmd

Function

Startup of process monitoring facility.

Effect if stopped

Applications using the process monitoring functions will not work.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclctrl

dependent

/milestone/smawrrms

Startup daemon

/etc/opt/FJSVcluster/FJSVclapm/daemons/prmd

Utilized port

None.

Remarks

None.

/milestone/fjsvclrmgr2

Function

Startup of Cluster Resource Management facility (2).

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclctrl

dependent

/milestone/smawrrms

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvwvbs

Function

Startup of daemons on Web-Based Admin View management server or monitoring nodes.

Effect if stopped

Settings and monitoring via the GUI provided by Web-Based Admin View will not be available.

Dependency relationship with other services

dependency

/milestone/multi-user
/network/initial

dependent

None.

Startup daemon

[For the node working as the primary or the secondary management server]
/opt/SMAW/SMAWcj2re/jre/bin/java
wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0-2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0-2 processes)

[For nodes other than those described above]
wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0-2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0-2 processes)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9399 (*1)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9398 (*2)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9397 (*3)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote cluster nodes
9396 (*4)	UDP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote cluster nodes

Remarks

When using the IP filtering function of Oracle Solaris, allow communications between all nodes.

(*1) No.9399 is set to support the service name "fjwv_c."

(*2) No.9398 is set to support the service name "fjwv_s."

(*3) No.9397 is set to support the service name "fjwv_n."

(*4) No.9396 is set to support the service name "fjwv_g."

(*5) Including concurrent use with cluster nodes.

(*6) PC or Solaris workstation.

/milestone/fjsvclautoconfig

Function

Beginning of shared disk connection confirmation.

Effect if stopped

Connections of shared disks cannot be checked.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclctrl

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvclwaitprobe

Function

Waiting for shared disk connection confirmation.

Effect if stopped

Connections of shared disks cannot be checked.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvclautoconfig

dependent

/milestone/smawrrms

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvwvcnf

Function

WWW server for sending Java applets, Java classes, and HTML contents to clients.

Effect if stopped

Settings and monitoring via the GUI provided by Web-Based Admin View will not be available.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/fjsvwvbs

dependent

None.

Startup daemon

/opt/FJSVwvcnf/bin/wvcnfd

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
8081 (*1)	TCP	s, r	Administrative LAN	Administrative server (*2)	ANY	WebView client (*3)

Remarks

(*1) No.8081 is set to support the service name "fjwv-h."

(*2) Including concurrent use with cluster nodes.

(*3) PC or Solaris workstation.

Note: For wvcnfd, there is an additional child process of the same name while processing a request from a client.

This process, however, terminates immediately after processing the request.

/milestone/fjsvgfsinit

Function

Initialization processing for monitoring facility of GFS shared file system.

Effect if stopped

Functions of GFS shared file system cannot be used.

Dependency relationship with other services

dependency

/system/filesystem/minimal

dependent

/milestone/single-user

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvgfsfsrm

Function

Startup control for monitoring facility of GFS shared file system, mount control for GFS shared file system.

Effect if stopped

Functions of GFS shared file system cannot be used.

Dependency relationship with other services

dependency

/milestone/multi-user
/milestone/smawcf
/milestone/fjsvclctrl
/milestone/fjsvgfsinit
/milestone/fjsvclrmgr2

dependent

/milestone/smawrrms

Startup daemon

/usr/lib/fs/sfcfs/sfcpncd
/usr/lib/fs/sfcfs/sfcprmd
/usr/lib/fs/sfcfs/sfchnsd
/usr/lib/fs/sfcfs/sfcfrmd

/usr/lib/fs/sfcfs/sfcfsd
/usr/lib/fs/sfcfs/sfcfsmg

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9200 (*1)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node
9100-9163 (*2)	TCP	s, r	Interconnect and Administrative LAN	Cluster node	ANY	Local and remote cluster nodes

Remarks

When using the IP filtering function of Oracle Solaris, allow communications between all nodes.

(*1) No. 9200 is set to support the service name "sfcfsrm."

(*2) From No. 9100 to No. 9163 are set to support the service names from sfcfs-1 to sfcfs-64.

/milestone/fjsvgfs

Function

Size check of internal log files of GFS shared file system, and switching of logs as necessary.

Effect if stopped

Functions of GFS shared file system cannot be used.

Dependency relationship with other services

dependency

/milestone/multi-user-server
/milestone/fjsvgfsfsrm

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/fjsvgfs-zones

Function

Automatic startup of a non-global zone that shares the GFS Shared File System.

Effect if stopped

When starting a non-global zone that shares the GFS Shared File System automatically, the GFS shared file system cannot be used.

Dependency relationship with other services

dependency

/milestone/fjsvgfsfsrm

dependent

/system/zones

Startup daemon

None.

Utilized port

None.

Remarks

Enable this service only when starting a non-global zone that shares the GFS Shared File System automatically.

/milestone/smawrhvto

Function

Initialization processing of RMS.

Effect if stopped

RMS functions cannot be used.

Dependency relationship with other services

dependency

/milestone/multi-user

dependent

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

/milestone/smawrrms

Function

Startup of RMS.

Effect if stopped

Even if HV_RCSTART=1 is set, RMS will not start automatically at node startup.

Dependency relationship with other services

dependency

/milestone/smawcf
/milestone/multi-user-server

dependent

None.

Startup daemon

/opt/SMAW/SMAWRrms/bin/bm

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9786 (*1)	TCP	s, r	Interconnect	Cluster node	9786	Remote cluster node

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
8000	UDP	s, r	Interconnect	Cluster node	8000	Remote cluster node

Remarks

(*1) No.9786 is set to support the service name "rmshb."

In order to avoid duplicate use of the same port number by other applications, modify the port number in the other applications.

/milestone/fjsvcldev

Function

Startup of asynchronous monitoring.

Effect if stopped

As the asynchronous monitoring cannot function, detection of any nodes being down will be delayed.

Also, when a node is hanging, the LEFTCLUSTER status cannot be resolved without intervention by an operator. (*1)

Dependency relationship with other services

dependency

/milestone/smawcf

dependent

None.

Startup daemon

/etc/opt/FJSVcluster/sys/devscfd (*2)
/etc/opt/FJSVcluster/sys/devscfmond (*2)
/etc/opt/FJSVcluster/sys/devrccud (*2)
/etc/opt/FJSVcluster/sys/devrccuxd (*2)
/etc/opt/FJSVcluster/sys/devrcircculogd
/etc/opt/FJSVcluster/sys/devsnmpd (*3)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
ANY	TCP	s, r	Administrative LAN	Cluster node	22 (*4)	XSCF (*5), ILOM (*6)
ANY	TCP	s, r	Administrative LAN	Cluster node	23 (*4)	XSCF (*7)

Remarks

(*1) If you are using the asynchronous monitoring in order to automatically stop a node on which an error occurred, it is necessary to enable the following service in addition to this service:
/milestone/smawsf

(*2) Not available on SPARC M10.

(*3) Available only on SPARC M10.

(*4) ssh port number 22 and telnet port number 23 in communication targets cannot be modified.

(*5) When connecting to XSCF from SPARC M10, SPARC Enterprise M3000, M4000, M5000, M8000, M9000 via SSH.

(*6) When connecting to ILOM from SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series via SSH.

(*7) When connecting to XSCF from SPARC M10, SPARC Enterprise M3000, M4000, M5000, M8000, M9000 via telnet.

/milestone/fjsvcldevtrap

Function

Monitoring the SNMP trap receiving daemon used for the shutdown facility.

Effect if stopped

As the asynchronous monitoring cannot function, detection of any nodes being down will be delayed.

Dependency relationship with other services

dependency

None.

dependent

None.

Startup daemon

/usr/sfw/sbin/snmptrapd (*1)

/usr/sbin/snmptrapd (*2)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9385(*3)	UDP	r	Administrative LAN	Cluster node	ANY	XSCF (*4)

Remarks

This SMF service is avilable only on SPARC M10.

(*1) Available only on SPARC M10 in the Solaris 10 environment.

(*2) Available only on SPARC M10 in the Solaris 11 environment.

(*3) No. 9385 is set to support the service name "sftrap."

(*4) For SPARC M10.

/milestone/smawsf

Function

Startup of Shutdown Facility.

Effect if stopped

Nodes on which an error occurs cannot be stopped automatically.

Dependency relationship with other services

dependency

/milestone/fjsvcldev

dependent

/milestone/smawrrms

Startup daemon

/opt/SMAW/SMAWsf/bin/rcsd

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
2316 (*1)	UDP	s, r	Administrative LAN	Cluster node	2316	Remote cluster node
23 (*2)	TCP	s, r	Administrative LAN	Cluster node	23 (*2)	ALOM

Remarks

(*1) No. 2316 is set to support the service name "sfadv."

(*2) The telnet port number 23 in communication targets cannot be modified.

/milestone/smawcf

Function

Loading of CF and CIP drivers.

Effect if stopped

The cluster cannot be started.

Dependency relationship with other services

dependency

/milestone/multi-user

dependent

None.

Startup daemon

/opt/SMAW/SMAWcf/bin/cfregd

/opt/SMAW/SMAWcf/bin/genqm

Utilized port

None.

Remarks

None.

/network/fjsvrvnet (*1)

Function

Startup of daemons and activation of virtual interfaces.

Effect if stopped

Creation of the LAN redundancy using the Redundant Line Control Function for Virtual NIC Mode is not available.

Dependency relationship with other services

dependency (*2)

/network/install

dependent

/network/physical

Startup daemon

/opt/FJSVrvnet/bin/rvnetmd

/opt/FJSVrvnet/bin/rvnetpathmd (*3)

Utilized port

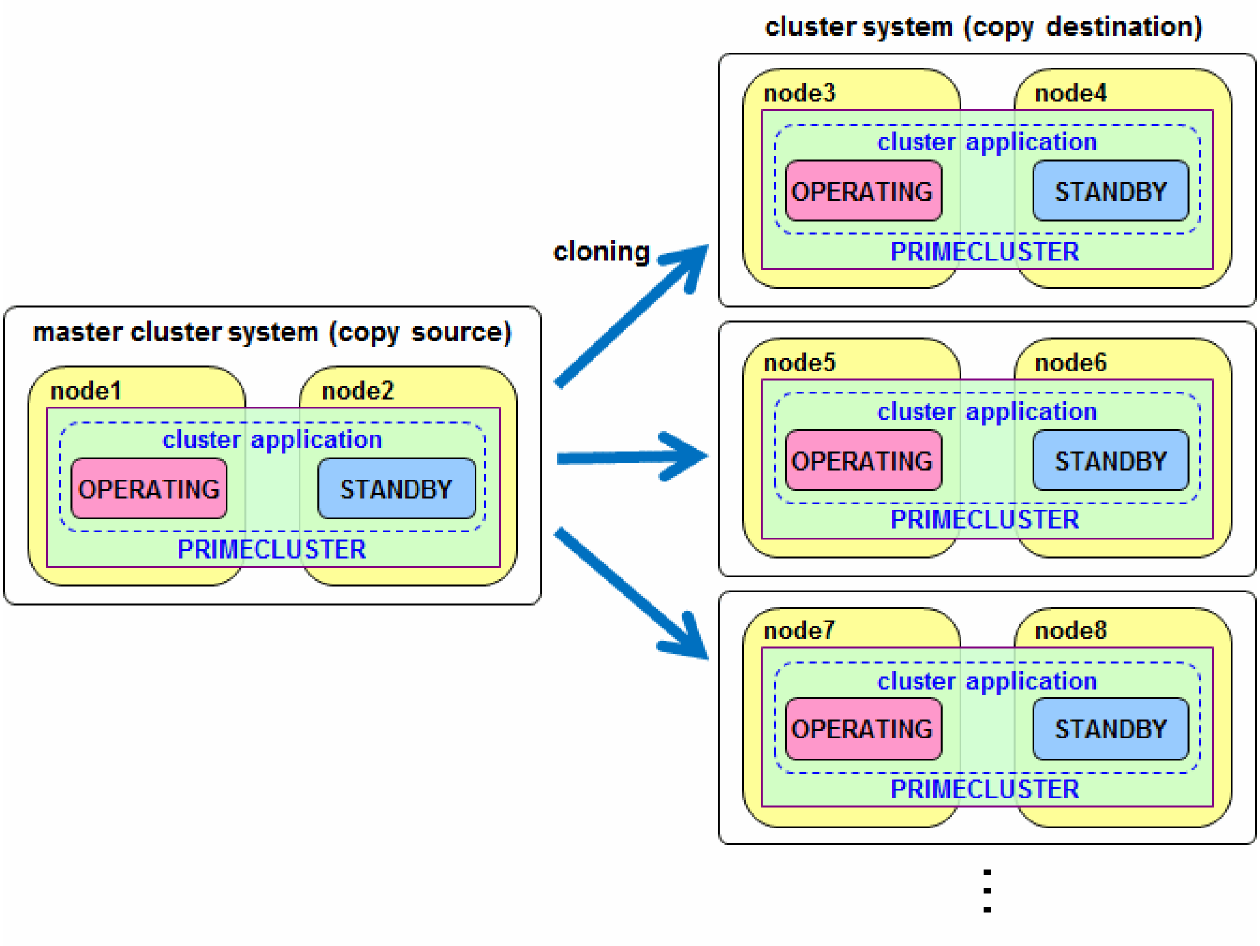
None.

Remarks

(*1) This SMF service exists only in Solaris 11.
(*2) If disabled, the service cannot start.
(*3) Whether this daemon starts or not, and the number of processes depend on the configuration of the virtual NIC mode. Also, this daemon may be suspended according to the monitoring status.

Appendix I Cloning the Cluster System Environment

PRIMECLUSTER allows you to configure a new cluster system by cloning an already configured cluster system.



The cluster system can be cloned in the following environments if the same OS version is used in both the copy source and the copy destination.

Table I.1 Solaris 10

		Copy destination	
		Physical environment	Guest domain environment in Oracle VM Server for SPARC
Copy source	Physical environment	Y	Y
	Guest domain environment in Oracle VM Server for SPARC	-	Y

Y: Supported, -: Not supported

Table I.2 Solaris 11

		Copy destination		
		Physical environment	Guest domain environment in Oracle VM Server for SPARC	Oracle Solaris Kernel Zones environment
Copy source	Physical environment	Y	Y	Y

	Guest domain environment in Oracle VM Server for SPARC	-	Y	-
	Oracle Solaris Kernel Zones environment	-	-	Y

Y: Supported, -: Not supported

Note

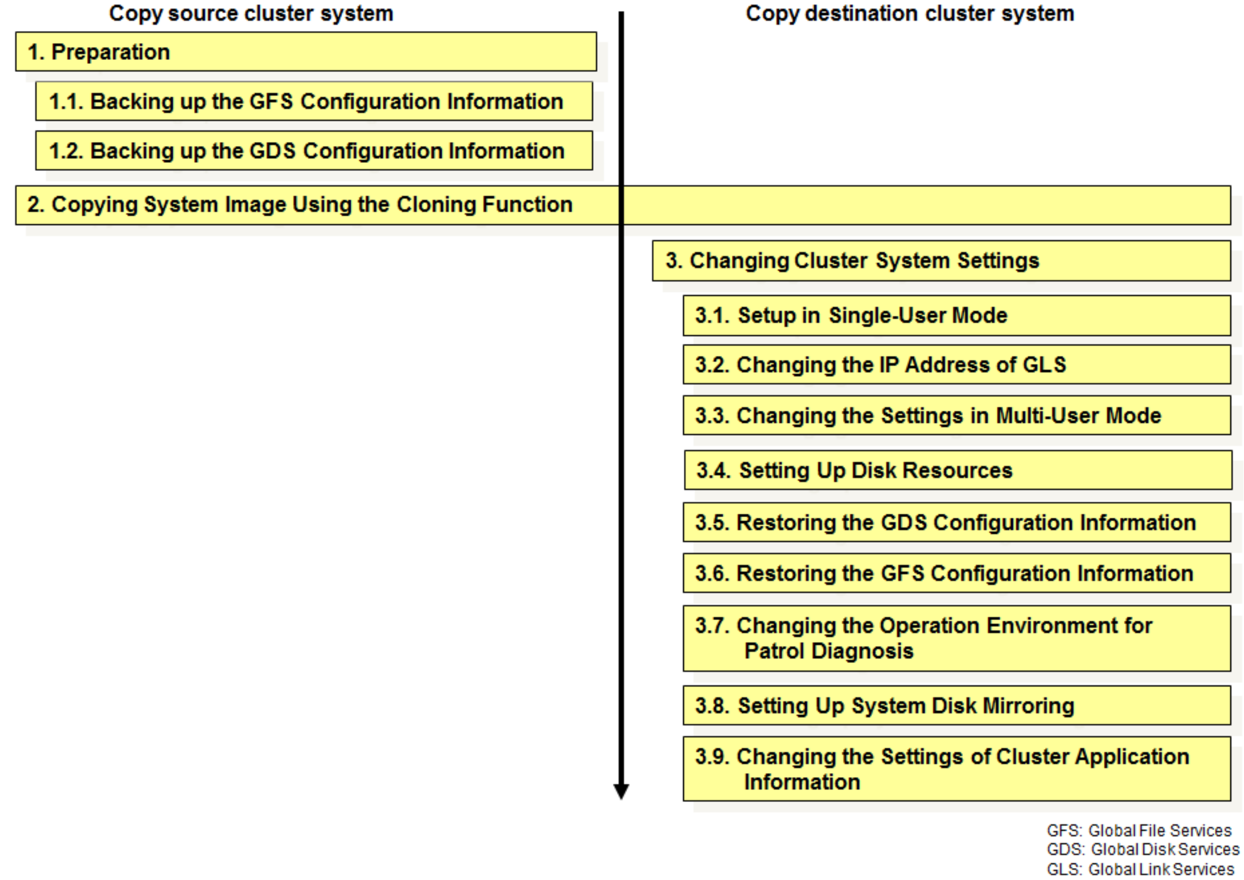
- The following items are not included in the supported range:
    - Building a single-node cluster from multiple clusters (cases of which can be seen with Disaster Recovery and so on)
    - After building a single-node cluster, copying it to multiple nodes to build multiple-node clusters
    - Building multiple-node clusters by copying a node within the multiple-node clusters to the multiple nodes
- Make sure that the sizes of disks managed by GDS are the same at both copy source and copy destination.
- Before starting up the copy destination system, make sure that the NIC cables are disconnected or the copy source is stopped, or connect from the copy source system to an isolated network, taking care that there are no IP addresses in duplicate with the copy source system.
- When you carry out cloning, you should follow the conditions of the cloning software/function (a flash archive for Solaris 10 and Unified Archives for Solaris 11) to be used.

Here, the cloning procedure is explained with the cases of cloning a cluster system of standby operation and a two-node cluster in the physical environment.

Procedure for Configuration by Cloning

The procedure for configuration by cloning in PRIMECLUSTER is as follows.

Figure I.1 Procedure for Configuration by Cloning



Note

If mirroring a system disk in the system of the copy source, the system disk mirroring in the copy destination is automatically canceled due to the functions of a flash archive and Unified Archives. Therefore, you must set it manually again.

The description of the steps in the following execution example, is given for building a cluster system with the following configuration.

	Copy source		Copy destination	
Cluster Name	PRIMECLUSTER1		PRIMECLUSTER2	
CF Node Name	fuji2	fuji3	fuji4	fuji5
CIP/SysNode Name	fuji2RMS	fuji3RMS	fuji4RMS	fuji5RMS
IP address on Administrative LAN	10.20.30.100	10.20.30.101	10.20.30.102	10.20.30.103
IP address for cluster interconnect	192.168.0.1	192.168.0.2	192.168.0.3	192.168.0.4
Physical IP address/hostname for GLS	10.34.214.181/ primecl01	10.34.214.182/ primecl02	10.34.214.191/ primecl03	10.34.214.192/ primecl04
Virtual IP address/hostname for GLS	10.34.214.185/takeoverIP		10.34.214.195/takeoverIP2	

GLS: Global Link Services

I.1 Preparation

This part describes the preliminary operation executed before cloning is applied.

I.1.1 Backing up the GFS Configuration Information

This section describes the items executed before cloning is applied while GFS Shared File System is used on the copy source server.

1. Back up the management partition information of the GFS Shared File System from the copy source server.

   Execute the following command on any running node.

   ```
   # sfcgetconf _backup_file_
   ```

   In the above example, sfcgetconf(1M) generates a shell script named _backup_file_ in the current directory.

   Note

   Execute the above procedure if you are going to copy data from a shared disk.

2. Edit _backup_file_ you retrieved in Step 1.

   Change the names of the nodes written in the execution procedure of the "sfcadm" command contained in _backup_file_ to the node names on the destination server.

   Example: The node names on the copy source server are host2 and host3, and, the node names on the destination server are host4 and host5.

   [Before Change]

   ```
   #!/bin/sh
   # This file is made by:
   # sfcgetconf _backup_file_
   # Thu May 26 09:23:04 2014
   #---- fsid : 1 ----
   # MDS primary (port) : host2 (sfcfs-1)
   # MDS secondory (port) : host3 (sfcfs-1)
   # MDS other :
   # AC : host2, host3
   # options :
   # device : /dev/sfdsk/gfs01/dsk/volume01
   sfcadm -m host2,host3 -g host2,host3 -p sfcfs-1,sfcfs-1 /dev/sfdsk/gfs01/dsk/volume01
   ...
   ```

   [After Change]

   ```
   #!/bin/sh
   # This file is made by:
   # sfcgetconf _backup_file_
   # Thu May 26 09:23:04 2014
   #---- fsid : 1 ----
   # MDS primary (port) : host4 (sfcfs-1)
   # MDS secondory (port) : host5 (sfcfs-1)
   # MDS other :
   # AC : host4, host5
   # options :
   # device : /dev/sfdsk/gfs01/dsk/volume01
   sfcadm -m host4,host5 -g host4,host5 -p sfcfs-1,sfcfs-1 /dev/sfdsk/gfs01/dsk/volume01
   ...
   ```

   Note

   If there are multiple file systems, there also are multiple lines in the execution procedure of the "sfcadm" command. Modify the node names in all lines.

3. Check the setup of the startup procedure of the sfcfrmd daemon.

```
# sfcsetup -m
wait_bg
```

Record the output value.

This value is used when restoring the GFS configuration information on the source destination server.

I.1.2 Backing up the GDS Configuration Information

1. When using PRIEMCLUSTER GDS Snapshot, remove the association between Master and Proxy beforehand.

   See

   For procedure for removing it, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

2. Back up the local class and shared class object configurations for GDS on the copy source server.

   Execute the following procedure on any node of the copy source server. If there are multiple classes, perform this operation for all classes.

   Example: The object configuration data of class Class1 is output to file /var/tmp/Class1.conf.

```
# sdxconfig Backup -c Class1 -o /var/tmp/Class1.conf
```

3. Save the GDS configuration data in a file on the copy source server. Output the class configuration data of all GDS classes to files.

   Example: The data of class Class1 is output to the /var/tmp/Class1.info file

```
# sdxinfo -c Class1 -e long > /var/tmp/Class1.info
```

I.2 Copying System Image Using the Cloning Function

This section describes the procedure for copying system images using the cloning function.

Note

- Before starting up the copy destination system, make sure that the NIC cables are disconnected or the copy source is stopped, or connect from the copy source system to an isolated network, taking care that there are no IP addresses in duplicate with the copy source system.
- The MAC addresses of the copy source system and destination system NICs are different. Depending on the cloning software/function you are using, update the MAC addresses either by initializing the NIC settings when cloning, or by modifying the NIC settings manually after cloning.

1. Copy the system Disk

   Copy the system disk image to the destination system.

   Copy the system disk image with the following functions of the OS:

   - For Solaris 10: Flash archive
   - For Solaris 11 or later: Unified Archives

   Note

   - For Solaris 11 or later, apply the PRIMECLUSTER patch T012012SP-01 or later to the copy source environment first before creating the archive by the archiveadm create command. If the archive is already created without applying the PRIMECLUSTER patch, apply it to the copy destination environment.

- For Solaris 11 or later, when creating an archive by the archiveadm create command, specify the --root-only option to exclude storage pools other than the ZFS root pool from the archive.

It is not necessary to copy mirrored disk when the system disk mirroring is set and also the boot disk in the alternative boot environment when using PRIMECLUSTER GDS Snapshot.

After copying the system disk image, change the settings of the OS and other MW referring to the manuals for each product.

2. Copy the disks that are registered in a local class or a shared class of GDS.

   The disks registered in local or shared classes of GDS can be copied by one of the following methods:

   a. Copy the whole data of the disk including the GDS private slice.

   b. Copy the data of the GDS private slice only.

   c. Copy the data of the volume area only.

   d. Do not copy any of the disk data.

   Determine the copy range by the specifications of the cloning software or function you use for data copying (data of which area can be copied) and by the need of copying the data from the volume area.

I.3 Changing Cluster System Settings

This section explains the procedure for changing cluster system settings required on the system image copy destination.

I.3.1 Setup in Single-User Mode

Execute the following procedure on all nodes of the copy destination.

1. Start the system in single-user mode.

   Note

   When using the GFS Shared File System in the copy source server, check that the following message has been output to the /var/adm/messages file before migrating the node to single-user mode.

   WARNING: sfcfsrm:5001: Starting the sfcfrmd daemon was suspended because quorum does not exist

   If this message is displayed, the node cannot be migrated to single-user mode with shutdown(1M) because the system startup is suspended. In this case, migrate the node to single-user mode after it is stopped forcibly in OBP mode.

2. Add or change the host name.

   Refer to the system in the copy source to add or change the host name to /etc/inet/hosts according to the environment in the copy destination.

3. Change the primary management sever, secondary management server, httpip, and mip in the Web-Based Admin View.

   1. Set the IP addresses or host name of the primary management server and secondary management server.

   ```
   # /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server <IP-address-or-host-name-of-primary-management-server>
   # /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server <IP-address-or-host-name-of-secondary-management-server>
   ```

   2. Set httpip.

   ```
   # /etc/opt/FJSVwvbs/etc/bin/wvSetparam httpip <IP-address-or-host-name-used-for-a-client>
   ```

3. Set mip.

```
# /etc/opt/FJSVwvbs/etc/bin/wvSetparam mip <IP-address-or-host-name-for-identifying-own-host>
```

4. Change the CF node name, CIP/SysNode name, cluster name, and device names of the interconnect.

Note

For the naming conventions (cluster name and CF node name), see "5.1.1 Setting Up CF and CIP".

1. Change the string of the CF node name within the CF node name and the CIP/SysNode name that are described in /etc/cip.cf.

[Before change]

```
fuji2       fuji2RMS:netmask:255.255.255.0
fuji3       fuji3RMS:netmask:255.255.255.0
```

[After change]

```
fuji4       fuji4RMS:netmask:255.255.255.0
fuji5       fuji5RMS:netmask:255.255.255.0
```

2. Refer to the system in the copy source to correct the string of the CF node name within the CIP/SysNode name that are described in /etc/inet/hosts according to the environment in the copy destination.

[Copy source]

```
192.168.0.1     fuji2RMS
192.168.0.2     fuji3RMS
```

[After change]

```
192.168.0.3     fuji4RMS
192.168.0.4     fuji5RMS
```

3. Change the CF node name, cluster name, and device names of the interconnect described in /etc/default/cluster.

[Before change]

```
nodename fuji2
clustername PRIMECLUSTER1
device /dev/hme2
device /dev/hme3
```

[After change]

```
nodename fuji4
clustername PRIMECLUSTER2
device /dev/vnet2
device /dev/vnet3
```

5. Restore the symbolic links to the device file.
For Solaris 11 or later, restore the symbolic links to the device file that was removed due to cloning.

```
# /opt/SMAW/SMAWcf/bin/cfrecoverdev
```

6. Cancel the SF settings.

Delete the /etc/opt/SMAW/SMAWsf/rcsd.cfg file.

```
# rm /etc/opt/SMAW/SMAWsf/rcsd.cfg
```

7. Delete the file of the Cluster Resource Management Facility.

   If the /etc/opt/FJSVcluster/FJSVcldbm/config/shmno file exists, delete the file.

```
# rm /etc/opt/FJSVcluster/FJSVcldbm/config/shmno
```

8. Change the node of the Cluster Resource Management Facility.

   Note

   This procedure is unnecessary when the Cluster Resource Management Facility.

   Execute the following command to change the node name of the Cluster Resource Management Facility.

```
# /etc/opt/FJSVcluster/bin/clchgnodename
```

9. Delete the information in the management partition of GFS.

   Note

   This procedure is unnecessary when the GFS Shared File System is not being used.

   Delete the information in the management partition of the GFS Shared File System. Execute the following command on all nodes.

```
# rm /var/opt/FJSVgfs/sfcfsrm.conf
```

10. Change the GDS settings.

    Delete the local class and shared class definitions of GDS. Perform the following procedure on all nodes.

    1. Delete the settings of class name in the "/etc/opt/FJSVsdx/sysdb.d/class.db" file.

```
...
Class1          <- Delete all the lines of cluster name
...
```

    2. Delete all files named with the class name in the "/etc/opt/FJSVsdx/sysdb.d" directory.

```
# cd /etc/opt/FJSVsdx/sysdb.d
# rm Class1
```

    3. Correct the sfdsk.conf file.

       When cloning a cluster system from other environment other than a Kernel Zone to the Kernel Zone, add the line: SDX_KZONE_SUPPORT=on to the /kernel/drv/sfdsk.conf file.

       When cloning a cluster system from the Kernel Zone to the environment other than the Kernel Zone, delete the line: SDX_KZONE_SUPPORT=on from the /kernel/drv/sfdsk.conf file.

    4. Delete the devname.db file.

```
# cd /etc/opt/FJSVsdx/dev
# rm -f devname.db
```

11. Prepare for changing the GLS settings.

    If you are using GLS, perform the following:

    1. Refer to the system in the copy source to correct the /etc/inet/hosts file according to the environment in the copy destination.

       [Copy source]

```
10.34.214.185 takeoverIP # Virtual IP
10.34.214.181 primecl01  # primecl01 physical IP
```

```
10.34.214.182 primecl02  # primecl02 physical IP
10.34.214.188 swhub1     # primary HUB IP
10.34.214.189 swhub2     # secondary HUB IP
```

[After change]

```
10.34.214.195 takeoverIP2 # Virtual IP
10.34.214.191 primecl03   # primecl03 physical IP
10.34.214.192 primecl04   # primecl04 physical IP
10.34.214.188 swhub1      # primary HUB IP
10.34.214.189 swhub2      # secondary HUB IP
```

2. Isolate the copy destination system so that there are no IP addresses in duplicate with the copy source system.

   Beforehand, make sure that the NIC cables are disconnected or the copy source is stopped, or connect from the copy source system to an isolated network, taking care that there are no IP addresses in duplicate with the copy source system.

12. Restriction of the automatic startup of RMS

    Restrict the automatic startup of RMS

    Check the settings of the current automatic startup of RMS and execute the following command according to the settings.

```
# hvsetenv HV_RCSTART
1    <- Check this value
```

- If "0" is set, the automatic startup of RMS has been restricted. Go to Step 13.
- If "1" is set, execute the following commands to restrict the automatic startup of RMS.

```
# hvsetenv HV_RCSTART 0
# hvsetenv HV_RCSTART
0    <- Check "0" is output
```

13. After completing above procedure on all nodes of the copy destination, start up all nodes in multi-user mode.

I.3.2 Changing IP Address, MAC Address, and Interface Name of GLS

Change the settings for the IP address, the MAC address, and the interface name of GLS on all nodes.

See

For details on the settings, see "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

Note

The procedure depends on the data communication mode. The following procedure is for changing the IP address within the same network as the configuration using the NIC switching mode. In the procedure below, sha0 indicates the virtual interface name and sha1 indicates the interface name of the standby patrol.

1. Delete all settings for the takeover virtual Interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc delete -n all
```

2. Stop the HUB monitoring function and the patrol monitoring function.

```
# /opt/FJSVhanet/usr/sbin/hanetpoll off
# /opt/FJSVhanet/usr/sbin/stpptl -n sha1
```

3. Disable the interface status monitoring function.

```
# touch /var/opt/FJSVhanet/tmp/disable_watchif
```

4. Modify the IP address of the primary physical interface.

   The following is the example of changing the IP address of the primary interface net1 in the Solaris 11 environment. IP address is changed by using the ipadm(1M) command.

   For [primecl03]

```
# /usr/sbin/ipadm delete-ip net1
# /usr/sbin/ipadm create-ip net1
# /usr/sbin/ipadm create-addr -T static -a 10.34.214.191/24 net1/v4
# /usr/sbin/ipadm show-addr net1/v4  <- Check that the IP address has been changed.
```

   For [primecl04]

```
# /usr/sbin/ipadm delete-ip net1
# /usr/sbin/ipadm create-ip net1
# /usr/sbin/ipadm create-addr -T static -a 10.34.214.192/24 net1/v4
# /usr/sbin/ipadm show-addr net1/v4  <- Check that the IP address has been changed.
```

   Note

   When bundling interfaces that are different in the copy source and copy destination, use the ipadm(1M) command to delete the IP address in the copy source, and then assign the IP address for the interface used in the copy destination.

5. Change the IP address of the virtual interface.

   For [primecl03]

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -i 10.34.214.195 -e 10.34.214.191
# /opt/FJSVhanet/usr/sbin/hanetconfig print -n sha0  <- Check that the IP address has been
changed.
```

   For [primecl04]

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -i 10.34.214.195 -e 10.34.214.192
# /opt/FJSVhanet/usr/sbin/hanetconfig print -n sha0  <- Check that the IP address has been
changed.
```

6. Change the MAC address that is set in the standby patrol function.

   Take the following procedure on all the nodes in the copy destination. These are the examples of [primecl03] and [primecl04].

   1. Check the MAC address that is set in the standby patrol function.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
[IPv4,Patrol]

 Name       Hostname        Mode MAC Adder/Phys ip    Interface List
+----------+--------------+----+-------------------+---------------------------+
 sha0       10.34.214.185   d    10.34.214.181       net1,net2
 sha1       -               p    02:03:04:05:06:07   sha0
                                 ^^^^^^^^^^^^^^^^^ <- Check this value.
```

   2. Take the following procedure depending on the current value.

      - If the current value is "00:00:00:00:00:00", this procedure is not necessary. Go to step 7.

      - If the current value is not "00:00:00:00:00:00", change the MAC address by executing the following command.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha1 -a 0:0:0:0:0:0
```

7. Change the interfaces bundled by GLS as necessary.

   When bundling the interfaces that are different in the copy source and copy destination, change the interfaces bundled by GLS. If the interface names have not been changed, skip this step. The following is the procedure when the interfaces bundled by the virtual interface sha0 of GLS need to be changed. Change the primary interface to net3 and the secondary interface to net4.

   For [primecl03]

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -t net3,net4
```

   For [primecl04]

```
# /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -t net3,net4
```

8. Create the takeover virtual interface again.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
```

9. Check the changed configuration definition.

   The following is the example of [primecl03].

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc print
 ifname     takeover-ipv4    takeover-ipv6
+----------+----------------+--------------------------------------------------+
 sha0:65    10.34.214.195    -            <- Check that the takeover virtual interface is
                                             set correctly.
# /opt/FJSVhanet/usr/sbin/hanetconfig print -n sha1
[IPv4,Patrol]

 Name      Hostname       Mode MAC Adder/Phys ip   Interface List
+---------+-------------+----+-------------------+---------------------------+
 sha0      10.34.214.195   d   10.34.214.191       net3,net4 <- Check IP addresses and
                                                              interface names.
 sha1      -               p   00:00:00:00:00:00   sha0      <- Check the MAC address.
```

10. To enable the setting change of IP address, you must restart OS on all nodes. If the NIC cable has been disconnected in the preliminary operation, stop OS first and then connect the NIC cable again. After that, restart OS.

See

For changing the IP address to a different network, the subnet mask of the virtual interface and the monitoring IP address of the HUB monitoring function need to be changed. For details, see "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

I.3.3 Changing the Settings in Multi-User Mode

This work is to be performed after completing the settings on all nodes in the cluster system of the copy destination in single-user mode.

1. Start all nodes in multi-user mode.

   At this time, the following error message may be output to the console. However, no corrective action is necessary.

```
SDX:sdxservd: ERROR: class_name: failed to start shared volumes, class closed down,
node=node_name
```

2. Set up the class Cluster Integrity Monitor (CIM).

   Delete the CF node names that were used in the copy source, and set the CF node names to be used in the copy destination.

   Perform the settings on any node that configures the cluster system.

   Example: The CF node names used in the copy source are fuji2 and fuji3, and those used in the copy destination are fuji4 and fuji5.

```
# rcqconfig -d fuji2 fuji3
# rcqconfig -a fuji4 fuji5
```

3. Check the CF setting item.

   Check if the changed CF node name, CIP/SysNode name, cluster name, and device names of the interconnect are correct.

   a. Checking the CF node name, cluster name, and device names of the interconnect.

      Execute the cfconfig -g command on each node to check if the set CF node name, cluster name, and device names of the interconnect are correct.

      Example: When the CF node name is fuji4, the cluster name is PRIMECLUSTER2, and the device names of the interconnect are /dev/vnet2 and /dev/vnet3 in the copy destination.

      ```
      # cfconfig -g
      fuji4 PRIMECLUSTER2 /dev/vnet2 /dev/vnet3
      ```

   b. Checking the CIP/Sysnode name

      Check that all the CIP/SysNode names set in the remote host are enabled to communicate. Check the communication status on all nodes.

      Example: When the SysNode name set in the remote host is fuji5RMS

      ```
      # ping fuji5RMS
      ```

   If an error occurs in the above step a or b, check if the CF node name, CIP/SysNode name, cluster name, and device names of the interconnect that are set in /etc/cip.cf, /etc/default/cluster or /etc/inet/hosts are correct.

   If an error occurs, take the procedure below:

   1. Start the system in single-user mode.

   2. Perform "4. Change the CF node name, CIP/SysNode name, cluster name, and device names of the interconnect." of "I.3.1 Setup in Single-User Mode" again, and then restart the node.

   3. Perform "I.3.3 Changing the Settings in Multi-User Mode" again.

4. Check the existence of the symbolic links to the device file.
   For Solaris11 or later, check that the symbolic links to the device file are correctly restored.

   ```
   # pkgchk SMAWcf
   ```

   If "pathname does not exist" is displayed, take the following procedure.

   1. Start the system in single-user mode.

   2. Perform "5. Restore the symbolic links to the device file" of "I.3.1 Setup in Single-User Mode", and then restart the node.

   3. Perform "I.3.3 Changing the Settings in Multi-User Mode" again.

5. Change the cluster name of the Cluster Resource Management Facility.

   Change the cluster name of the Cluster Resource Management Facility.

   Perform the settings on any node that configures the cluster system.

   Example: The new cluster name of the copy destination is "PRIMECLUSTER 2".

   ```
   # /etc/opt/FJSVcluster/bin/clsetrsc -n PRIMECLUSTER2 1
   # /etc/opt/FJSVcluster/bin/clsetrsc -n PRIMECLUSTER2 2
   ```

6. Delete the network interface resources.

   Delete the network interfaces registered in the resource database.

   Perform the settings on any node that configures the cluster system.

1. Check the resource IDs of the registered network interfaces.
   The resource IDs of the network interfaces are the underlined values displayed on the "Ethernet" line in the following command output result.

```
# /etc/opt/FJSVcluster/bin/clgettree
...
  Ethernet 27 vnet0 UNKNOWN
  Ethernet 28 vnet1 UNKNOWN
...
```

2. Delete all the network interface resources that were checked in Step 1.

   Example: The resource IDs of the registered network interfaces are 27 and 28.

```
# /etc/opt/FJSVcluster/bin/cldelrsc -r 27
# /etc/opt/FJSVcluster/bin/cldelrsc -r 28
```

7. Register the network interface resources again.

   Register the network interface resources in the resource database again.

   For the procedure, see "8.1.1.2 Adding a Network Interface Card Used for the Public LAN and the Administrative LAN."

8. Change the SF settings.

   1. Delete the information of the asynchronous monitoring that was used in the copy source.

      For SPARC M10

      Execute the following command on any node to check the information of the SNMP asynchronous monitoring that was used in the copy source.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -l
```

      Execute the following commands on any node to delete the information of the SNMP asynchronous monitoring that was used in the copy source.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -d fuji2
# /etc/opt/FJSVcluster/bin/clsnmpsetup -d fuji3
```

      Execute the following command on any node to check that information of the SNMP asynchronous monitoring is not displayed.

```
# /etc/opt/FJSVcluster/bin/clsnmpsetup -l
```

      For SPARC Enterprise M3000, M4000, M5000, M8000, M9000, SPARC Enterprise T5120, T5220, T5140, T5240, T5440, SPARC T3, T4, T5, T7, S7 series

      Execute the following command on any node to check the information of the console asynchronous monitoring that was used in the copy source.

```
# /etc/opt/FJSVcluster/bin/clrccusetup -l
```

      Execute the following command on any node to delete the information of the console asynchronous monitoring that was used in the copy source.

```
# /etc/opt/FJSVcluster/bin/clrccusetup -d fuji2
# /etc/opt/FJSVcluster/bin/clrccusetup -d fuji3
```

      Execute the following command on any node to check that the information of the console asynchronous monitoring is not displayed.

```
# /etc/opt/FJSVcluster/bin/clrccusetup -l
```

      For SPARC Enterprise T1000 and T2000

      Delete the /etc/opt/SMAW/SMAWsf/SA_sunF.cfg file.

```
# rm /etc/opt/SMAW/SMAWsf/SA_sunF.cfg
```

2. See "5.1.2 Configuring the Shutdown Facility" to set the shutdown facility again.

I.3.4 Setting Up Disk Resources

Perform the following procedure in the copy destination cluster system.

1. Deletion of disk resources

   Perform this setting on any node configuring a cluster system.

   1. Delete all GDS class resources

      Example: Deleting class resource Class1

      ```
      #/etc/opt/FJSVsdx/bin/sdxdcrsc -R -c Class1
      ```

   2. Confirm the resource IDs of the registered disk resources.

      The resource IDs of the disk resources are the underlined portions of the entries for "SHD_DISK" and "DISK" in the following command output results.

      ```
      # /etc/opt/FJSVcluster/bin/clgettree
      Cluster 1 cluster
        Domain 2 CLUSTER
          Shared 7 SHD_CLUSTER
           SHD_DISK 35 SHD_Disk35 UNKNOWN
             DISK 37 c1t4d0 ON fuji4
             DISK 153 c1t4d1 ON fuji5
      ...
        Node 3 fuji4 ON
          DISK 19 c0t0d0 UNKNOWN
          DISK 37 c1t4d0 ON
      ...
      ```

   3. Delete all the disk resources that were checked in Step 2.

      Example: The resource IDs of the registered disk resources are "19", "35", "37", and "153".

      ```
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 19
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 35
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 37
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 153
      ```

2. Re-registration of disk resources

   If disk resources are used in the copy source, re-register the disks in the resource database.

   For details, see "6.3.1.1 Executing Automatic Configuration."

   Example:

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -r
   ```

I.3.5 Restoring the GDS Configuration Information

Restore the GDS configuration information in the copy destination cluster system.

1. Change the physical disk name in GDS configuration files.

   If the physical disk names registered in the GDS class are different in the copy source and destination systems, use the "sdxconfig Convert" command to change the physical disk names in the configuration file to the physical disk names in the copy destination system. This setting is performed on the node where the configuration file was backed up.

   Example: Changing the physical disk described in the "/var/tmp/Class1.conf" configuration file from "c1t4d0" to "c1t4d2"

```
# sdxconfig Convert -e replace -c Class1 -p c1t4d0=c1t4d2 -i /var/tmp/Class1.conf -o /var/tmp/
Class1.conf -e update
```

2. Change of physical disk names in the Excluded List of GDS

   In environments using the Excluded List, if the physical disk names entered in the Excluded List are different in the copy source and destination systems, change the physical disk names to those entered in the Excluded List for the copy destination system. Perform this task on all nodes.

   For details on the Excluded List, see "*PRIMECLUSTER Global Disk Services Configuration and Administration Guide*".

3. Restoring the object configuration information of local and shared classes

   Execute the following command on the node where the configuration file was backed up.

   - If private slice data were copied

   ```
   # sdxconfig Restore -c Class1 -i /var/tmp/Class1.conf -e chkps
   ```

   - If private slice data were not copied

   ```
   # sdxconfig Restore -c Class1 -i /var/tmp/Class1.conf
   ```

   Note

   After restoring with the "sdxconfig Restore" command, shared classes become local classes.
   If the following message is displayed, take corrective measures with reference to the "PRIMECLUSTER Global Disk Services Configuration and Administration Guide":

   ERROR: *device*: disk label is not matched with class *class*

4. Restart OS on all nodes.

5. Change of the class attribute

   If the Class is a shared class, change the restored class from local class to shared class.

   Perform the following operation on the nodes on which you restored the class object configuration in Step 3.

   1. Stop the GDS volume.

   ```
   # sdxvolume -F -c Class1
   ```

   2. Change class attribute to shared class.

   ```
   # sdxattr -C -c Class1 -a type=shared,scope=fuji4:fuji5
   ```

6. Preliminary setup for Gds resources used in RMS

   Perform the following operation on any node.

```
# /opt/SMAW/SMAWRrms/bin/hvgdsetup -a Class1
```

7. Associate Master with Proxy.

   If PRIMECLUSTER GDS Snapshot is used, associate Master with Proxy.

   See

   For how to associate them, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

8. Start the GDS volume.

   If the GDS volume stopped in Step 5. includes any GDS shared class volumes which are not registered to RMS (e.g. classes used by GFS), start the volume on all the nodes in the class scope by executing the sdxvolume -N command, since it will not start automatically at the time of starting the RMS.

   Example: Starting classes used by GFS (gfs and gfs01)

   Execute the following commands on all the nodes in the class scope.

   ```
   # sdxvolume -N -c gfs
   # sdxvolume -N -c gfs01
   ```

I.3.6 Restoring the GFS Configuration Information

Restore the GFS configuration information to the copy destination servers.

Note

This procedure is required when using a GFS Shared File System on the copy source servers.

1. Reinitialize the management partition on the one node of the copy destination servers.

   Example: Initializing the /dev/sfdsk/gfs/rdsk/control file as the management partition.

   ```
   # sfcsetup -cf /dev/sfdsk/gfs/rdsk/control
   ```

2. Reregister the information of the configuration node on each node.

   ```
   # sfcsetup -a /dev/sfdsk/gfs/rdsk/control
   ```

3. On the one node of the copy destination servers, redo the settings for the startup method of the sfcfrmd daemon as recorded in "I.1.1 Backing up the GFS Configuration Information" in Step 3.

   Example: For setting the startup method of sfcfrmd daemon to wait_bg

   ```
   # sfcsetup -m wait_bg
   ```

   Note

   This procedure is required when changing the startup method of the sfcfrmd daemon from the default value wait.

4. Confirm that the management partition is reinitialized.

   The path name of the management partition for which the settings were made can be confirmed by executing the "sfcsetup(1M)" command with the -p option. .

   ```
   # sfcsetup -p
   /dev/sfdsk/gfs/rdsk/control
   ```

   The registered node information can be confirmed by executing the "sfcsetup(1M)" command without any option.

   ```
   # sfcsetup
   HOSTID          CIPNAME         MP_PATH
   80380000        fuji4RMS        yes
   80380001        fuji5RMS        yes
   ```

   The startup method of the sfcfrmd daemon can be confirmed by executing the"sfcsetup(1M)" command with the -m option.

   ```
   # sfcsetup -m
   wait_bg
   ```

5. Start the sfcfrmd daemon by executing the following command on all nodes.

```
# sfcfrmstart
```

6. Set the GFS Shared File System.

   - Without copying the data on the shared disk

     Create the GFS Shared File System.

     See

     For details on how to create a GFS Shared File System, see "Creating a file system" and "Selecting a communication protocol" of "Creation" of "File System Operations (Command)" in "PRIMECLUSTER Global File Services Configuration and Administration Guide."

   - Copying the data on the shared disk

     Restore the information of the management partition.

     Execute the shell script edited in "I.1.1 Backing up the GFS Configuration Information" on any one node.

```
# sh _backup_file_
get other node information start ... end
```

     Confirm that restoration of the management partition of GFS was successful by running the "sfcinfo(1M)" command and the "sfcrscinfo(1M)" command.

```
# sfcinfo -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID special                                        size  Type  mount
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)  14422  META  -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)   5116  LOG   -----
   1 /dev/sfdsk/gfs01/dsk/volume01(15000000021)  95112  DATA  -----
```

```
# sfcrscinfo -m -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID  MDS/AC  STATE  S-STATE   RID-1   RID-2   RID-N  hostname
   1  MDS(P)  stop   -             0       0       0  host4
   1  AC      stop   -             0       0       0  host4
   1  MDS(S)  stop   -             0       0       0  host5
   1  AC      stop   -             0       0       0  host5
```

7. Mount the GFS Shared File System on all nodes.

```
# sfcmntgl <mount point>
```

I.3.7 Changing the Operation Environment for Patrol Diagnosis

If the patrol diagnosis facility is used, change the LAN and shared disk to be monitored to the device in the cloning destination with the clspconfig(1M) command.

For details on the "clspconfig(1M)" command, see the manual page for "clspconfig(1M)."

I.3.8 Setting Up System Disk Mirroring

To enable system disk mirroring, it is required to set up system disk mirroring on all the target nodes. To create an alternative environment, create it by using the PRIMECLUSTER GDS Snapshot here.

See

For details on the setting procedure, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide".

I.3.9 Setting a File System on a Shared Disk

When using a file system on a shared disk, create the file system on the target node (If a file system has been copied by restoring GDS, this step is not necessary).

For how to create a file system, see "6.4 Initial File System Setup."

In addition, copy data from the file system on the copy source node as necessary.

I.3.10 Changing the Settings of Cluster Application Information

This section describes how to change the SysNode name that constitutes a cluster application and the host name of the takeover network resource as well as the CF node name such as the procedure resource.

1. Obtain the RMS Configuration name. Follow the procedure below to output the file content. The character string after -c will be the RMS Configuration name. Perform this operation on any one of the nodes constituting a cluster.

   Example

   If the RMS Configuration name is 'config'

   ```
   # cat /opt/SMAW/SMAWRrms/etc/CONFIG.rms
   hvcm -c config <- RMS Configuration name
   ```

2. Collect the backup of the configuration file. Perform this operation on all cluster nodes.

   Execute the following command.

   ```
   cd /opt/SMAW/SMAWRrms/build/wizard.d
   cp -rp <RMS Configuration name> <RMS Configuration name>.backup
   ```

   When the takeover network resource is used, execute the following command on all cluster nodes.

   ```
   cp -p /opt/SMAW/SMAWRrms/etc/hvipalias /opt/SMAW/SMAWRrms/etc/hvipalias.backup
   ```

   Example

   If <RMS Configuration name> is config

   ```
   # cd /opt/SMAW/SMAWRrms/build/wizard.d
   # cp -rp config config.backup
   # cp -p /opt/SMAW/SMAWRrms/etc/hvipalias /opt/SMAW/SMAWRrms/etc/hvipalias.backup
   ```

3. Modify the SysNode name of the configuration file. Perform this operation on the same node where the step 1 was performed.

   1. Move it to the directory where the configuration file is stored.

      ```
      # cd /opt/SMAW/SMAWRrms/build/wizard.d/<RMS Configuration name>
      ```

      Example

      If <RMS Configuration name> is config

      ```
      # cd /opt/SMAW/SMAWRrms/build/wizard.d/config
      ```

2. Search the target file with the following command.

```
grep -l HvpMachine *
```

Example

```
# grep -l HvpMachine *
userApp_0.m
userApp_1.m
    :
```

The displayed ".m" file is the target file.

3. Search and change the line starting "HvpMachine" in the displayed file. Open the file with the vi command.

```
# vi userApp_0.m
HvpApplication=userApp_0
HvpAutoBreak=yes
      :
    (Omitted)
      :
HvpMachine000=fuji2RMS  <- Target line to be changed
HvpMachine001=fuji3RMS  <- Target line to be changed
      :
   (Omitted)
```

Change the SysNode name that is set in the right side of "-" to the new SysNode name.

The SysNode name is set by adding RMS (capital letters) to the CF node name.

The following example indicates when changing the CF node name from fuji2 and fuji3 to fuji4 and fuji5.

[Before change]

```
HvpMachine000=fuji2RMS
HvpMachine001=fuji3RMS
```

[After change]

```
HvpMachine000=fuji4RMS
HvpMachine001=fuji5RMS
```

Perform this procedure on all the ".m" files obtained in step 3-2.

4. Check that the old SysNode does not exist.

Example

```
# grep HvpMachine *.m
userApp_0.m:HvpMachine000=fuji4RMS
userApp_0.m:HvpMachine001=fuji5RMS
userApp_1.m:HvpMachine000=fuji4RMS
userApp_1.m:HvpMachine001=fuji5RMS
```

5. Rename the <SysNode name>.s file under the directory where the configuration file is stored with the mv command. Change the file with the old SysNode name to the new SysNode name.

Example

```
# mv fuji2RMS.s fuji4RMS.s
# mv fuji3RMS.s fuji5RMS.s
```

4. Modify the CF node name of the procedure resource, line switching unit resource, and patrol diagnosis. This operation is only performed on the node where the step 3 was performed.

   This procedure is unnecessary if the procedure resource, line switching unit resource, and patrol diagnosis are not being used.

   1. Move it to the directory where the configuration file is stored.

      ```
      # cd /opt/SMAW/SMAWRrms/build/wizard.d/<RMS Configuration name>
      ```

      Example

      If <RMS Configuration name> is config

      ```
      # cd /opt/SMAW/SMAWRrms/build/wizard.d/config
      ```

   2. Search the target file with the following command.

      ```
      grep -l HvpCrmScopeFilter *
      ```

      Example

      ```
      # grep -l HvpCrmScopeFilter *
      Procedure0.m
      Procedure1.m
         :
      ```

      The displayed ".m" file is the target file.

   3. Search and change the line starting "HvpCrmScopeFilter" in the displayed file. Open the file with the vi command.

      ```
      # vi Procedure0.m
      HvpApplType=RESOURCE
      HvpApplication=Procedure0
      HvpClassNameFilter=BasicApplication
      HvpConsistent=consistent
      HvpCrmFlags000=OT1800
      HvpCrmResourceId000=33
      HvpCrmResourceName000=SDISK
      HvpCrmScopeFilter=fuji2:fuji3  <- Target line to be changed
      HvpPlugin=BasicApplication
      HvpPreCheckHeritageIn=''
         :
      (Omitted)
      ```

      Change the CF node name that is set in the right side of "=" to the new CF node name.

      The following example indicates when changing the CF node name from fuji2 and fuji3 to fuji4 and fuji5.

      [Before change]

      ```
      HvpCrmScopeFilter=fuji2:fuji3
      ```

      [After change]

      ```
      HvpCrmScopeFilter=fuji4:fuji5
      ```

Perform this procedure on all the ".m" files displayed in step 4-2.

Note

Only one CF node name may be set to HvpCrmScopeFilter due to the resource type.

4. Check that the old CF node name does not exist.

```
grep HvpCrmScopeFilter *.m
```

Example

```
# grep HvpCrmScopeFilter *.m
Procedure0.m:HvpCrmScopeFilter=fuji4:fuji5
Procedure1.m:HvpCrmScopeFilter=fuji4:fuji5
    :
```

5. Modify the host name of the takeover network. This procedure is only performed on the same node as step 3.
If the takeover network resource is not used, this step and the following steps 6 and 7 are not necessary. Skip those steps and go to step 8.

1. Move it to the directory where the configuration file is stored.

```
# cd /opt/SMAW/SMAWRrms/build/wizard.d/<RMS Configuration name>
```

Example

If <RMS Configuration name> is config

```
# cd /opt/SMAW/SMAWRrms/build/wizard.d/config
```

2. Search the target file with the following command.

```
grep -l HvpPlugin=Ipaddress *
```

Example

```
# grep -l HvpPlugin=Ipaddress *
Ipaddress0.m
Ipaddress1.m
    :
```

The displayed ".m" file is the target file.

3. In the displayed file, search and change the line starting "HvpInterface". Open the file with the vi command.

```
# vi Ipaddress0.m
HvpApplType=IPADDR
HvpApplication=Ipaddress0
HvpConsistent=consistent
HvpInterface000=V:oldname  <- Target line to be changed
HvpIntfFilter=''
HvpNeedAll=yes
      :
   (Omitted)
```

Change the host name set in the right side of (":") to the new host name.

The following example indicates when changing the host name (oldname) corresponding to the IP address of the takeover network to newname.

[Before change]

```
HvpInterface000=V:oldname
```

[After change]

```
HvpInterface000=V:newname
```

Perform this procedure on all the ".m" files obtained in step 4-2.

4. Check that the old host name of the takeover network does not exist.

Example

```
# grep HvpInterface *.m
Ipaddress0.m: HvpInterface000=V:newname
```

6. Change the CF node name and the takeover network in the hvipalias file.

This file is stored in /opt/SMAW/SMAWRrms/etc/hvipalias. For each line, the CF node name is set in the first field, the host name of the takeover network is set in the second field, and the network interface to be used for the takeover network is set in the third field. Modify each name to the new name. Perform this operation on all the nodes constituting a cluster.

Example

When the CF node names are changed from fuji2 and fuji3 to fuji4 and fuji5, the host names of the takeover network are changed from oldname to newname, and the network interface to be used for the takeover network are changed from net0 to net1.

[Before change]

```
fuji2  oldname    net0  0xffffff00       # sh_rid=34 rid=32 192.168.100.233
fuji3  oldname    net0  0xffffff00       # sh_rid=34 rid=33 192.168.100.233
```

[After change]

```
fuji4  newname    net1  0xffffff00       # sh_rid=34 rid=32 192.168.100.233
fuji5  newname    net1  0xffffff00       # sh_rid=34 rid=33 192.168.100.233
```

7. Check that the SysNode name and takeover network in the /etc/inet/hosts file have been changed.

8. Execute the following command on the same node where the step 1 was performed.

```
hvw -F Configuration-Activate -xj -n <RMS Configuration name>
```

Example

```
# hvw -F Configuration-Activate -xj -n config
```

In the command output result, check the following contents of (1) and (2).

```
Testing for RMS to be up somewhere in the cluster ... done.

Arranging sub applications topologically ... done.

Check for all applications being consistent ... done.

Running overall consistency check ... done.
```

```
Generating pseudo code [one dot per (sub) application]: ..... done.

Generating RMS resources [one dot per
resource]: ............................................................ done


hvbuild using /usr/opt/reliant/build/wizard.d/config/config.us
About to distribute the new configuration data to hosts: fuji4RMS,fuji5RMS (1)

The new configuration was distributed successfully.

About to put the new configuration in effect ... done.

The activation has finished successfully. (2)
#
```

In (1), check that the new SysNode name is output.

In (2), check that "The activation has finished successfully." is displayed.

If the output results are different, check the following and take a necessary action:

- Check if RMS is in operation on any node that constitutes a cluster. If it is, stop RMS.
- Check that all nodes that constitute a cluster run in multi-user mode.
- Check that the host name in the /etc/hosts file or the CF node name has been change on all nodes.

After taking the necessary action, perform the operation in step 8 again.

9. In "I.3.2 Changing IP Address, MAC Address, and Interface Name of GLS," if you change the IP address of GLS, delete the takeover IP address of Gls resource before cloning, and set it again. For how to set it, see "10.3.1 Changing the Cluster Application Configuration", "6.7.1.4 Creating Gls Resources."

10. If the RMS environment variables was changed in "I.3.1 Setup in Single-User Mode," return it with the following procedure. Perform this operation on all nodes where RMS is used.

```
# hvsetenv HV_RCSTART 1
# hvsetenv HV_RCSTART
1   <- Check that the returned value is output
```

11. Execute the following command on any one node to start RMS.

```
# hvcm -a
```

Appendix J Using Remote Power Distribution Unit

This appendix provides explanations on using the Remote Power Distribution Unit:

- Design
- Setup procedure
- Test

J.1 Design

PRIMECLUSTER forcibly stops a failed node by the shutdown facility and prevents the conflict of user resources.

If stopping other nodes forcibly, use the option hardware such as RCI or XSCF.

Therefore, if an error occurs in some machines, which the option hardware is not made redundant, the failed node cannot be forcibly stopped and the service may not be switched.

Even in this situation, by using the Remote Power Distribution Unit, the node's power and redundant power are cut and the node can be forcibly stopped for sure. Then, the cluster system is switched.

Using the Remote Power Distribution Unit is optional.

When using the Remote Power Distribution Unit, control it by the RPDU shutdown agent of the shutdown facility to check the status of the power supply to the node and cut the power.

J.1.1 Operation Environment

This section explains the supported configuration on using the Remote Power Distribution Unit:

- Remote Power Distribution Unit
    - SP0X9RP1F Remote Power Distribution Unit (100V)
    - SP0X9RP2F Remote Power Distribution Unit (200V)

    * You must connect the LAN for the Remote Power Distribution Unit to the administrative LAN.

    In addition, connect the Remote Power Distribution Unit only to the cluster node. If you want to control peripherals such as a shared disk from the Remote Power Distribution Unit, prepare it separately for controlling peripherals.
- 100V-200V Linked switch
    - PDU-224 by ISA (Recommended)
- Server type
    - SPARC Enterprise M3000, M4000, M5000
    - SPARC M10-1, M10-4
- Virtualized environment

    Virtualized environment is unsupported.

    When using PRIMECLUSTER in Oracle VM Server for SPARC environment and Oracle Solaris Kernel Zones environment, the Remote Power Distribution Unit cannot be used.

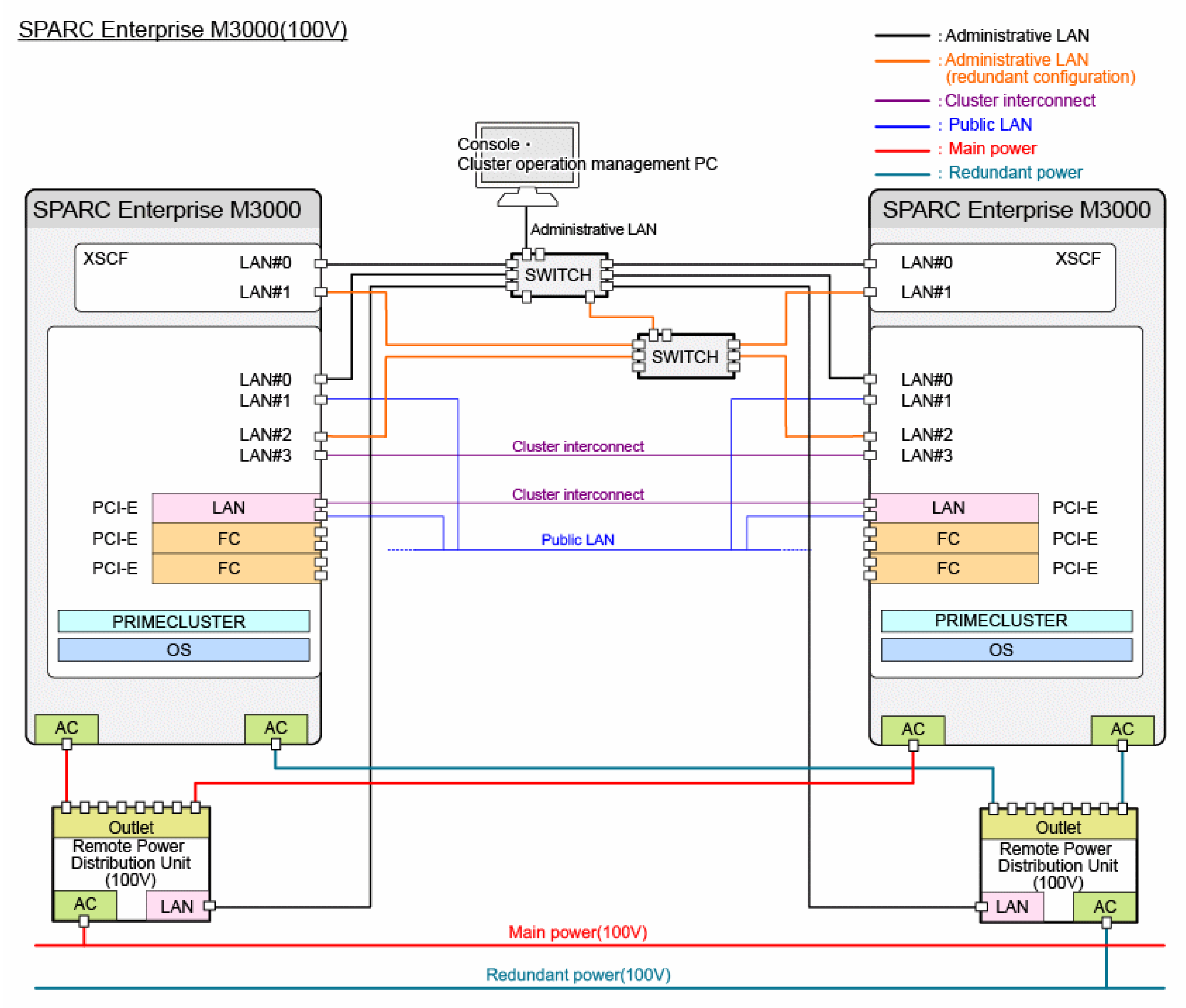
J.1.2 Hardware Configuration

This section explains the hardware configuration on using the Remote Power Distribution Unit:

- SPARC Enterprise M3000(100V)

  Two for up to two Remote Power Distribution Units (100V) SP0X9RP1F SPARC Enterprise M3000

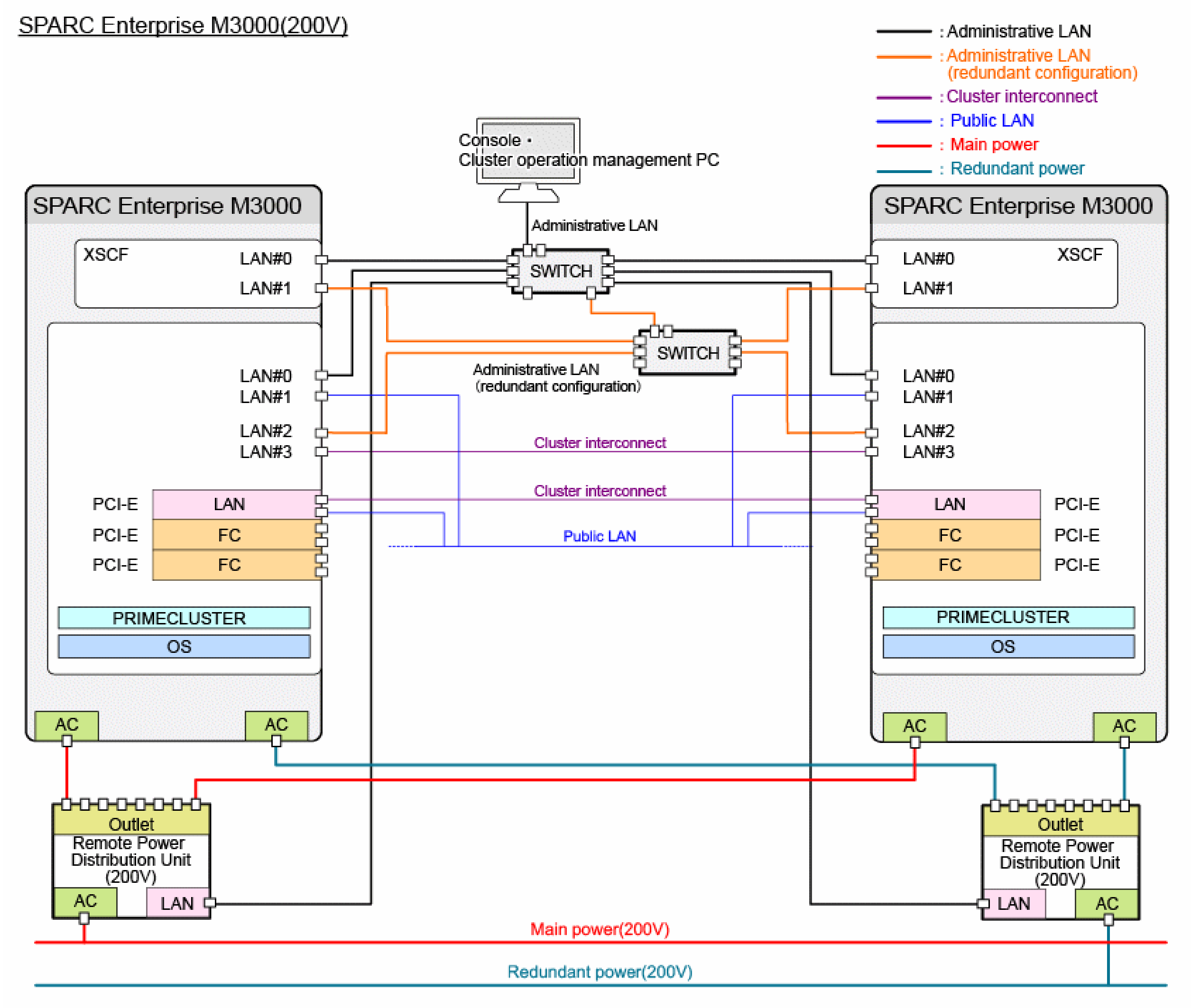
Figure J.1 Connection diagram for SPARC Enterprise M3000(100V)



- SPARC Enterprise M3000(200V)

  Two for up to seven Remote Power Distribution Units (200V) SP0X9RP2F SPARC Enterprise M3000

Figure J.2 Connection diagram for SPARC Enterprise M3000(200V)

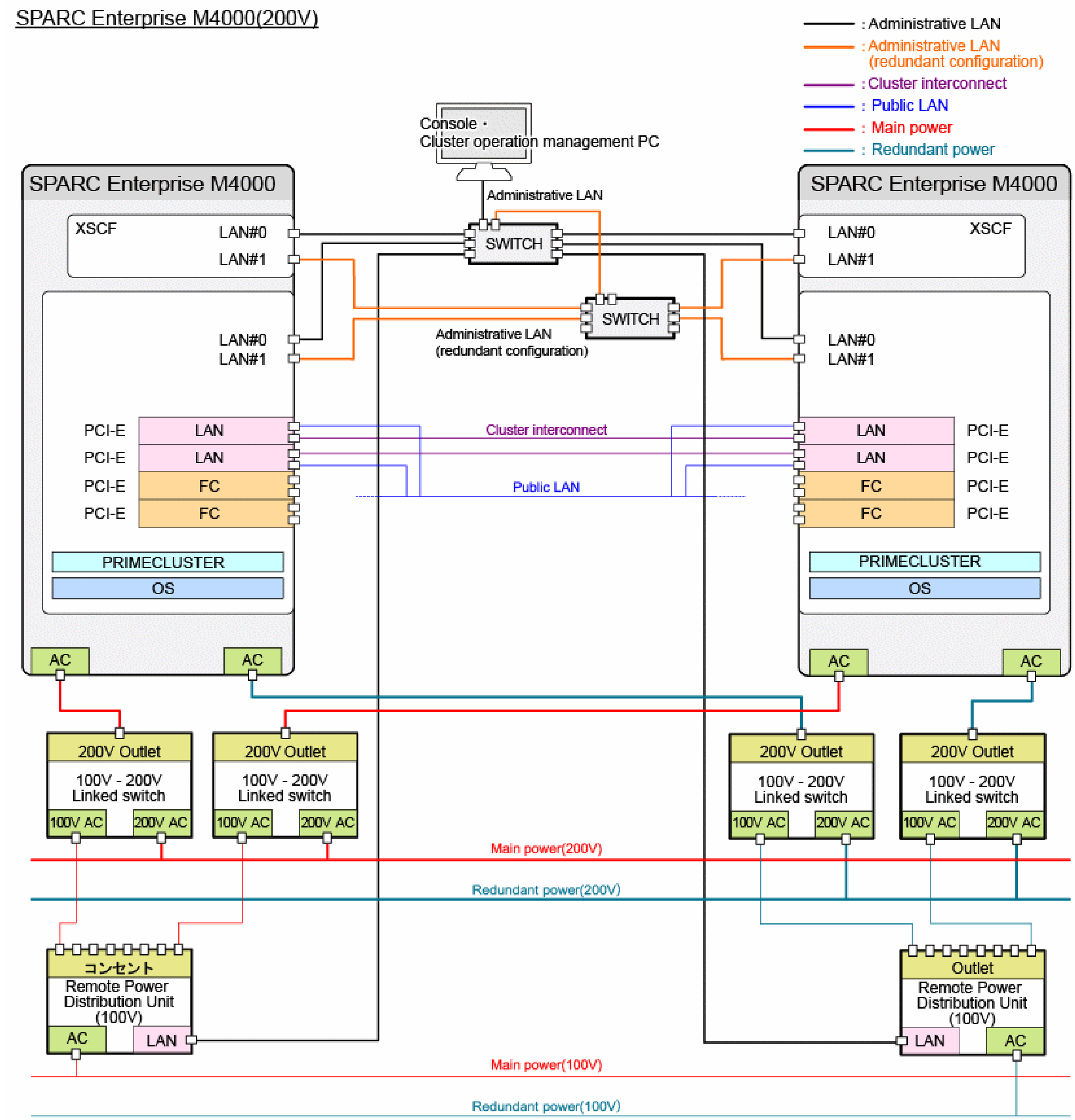


- SPARC Enterprise M4000(200V)
  - Two for up to eight Remote Power Distribution Units (100V) SP0X9RP1F SPARC Enterprise M4000
  - Two for one 100V-200V linked switch (PDU-224 by ISA (Recommended)) SPARC Enterprise M4000

Note

In addition to 200V power supply, 100V power supply for control is required in this configuration.
If the 100V power supply is cut, the 200V outlet, linked with the power supply, will also be OFF.

Figure J.3 Connection diagram for SPARC Enterprise M4000(200V)

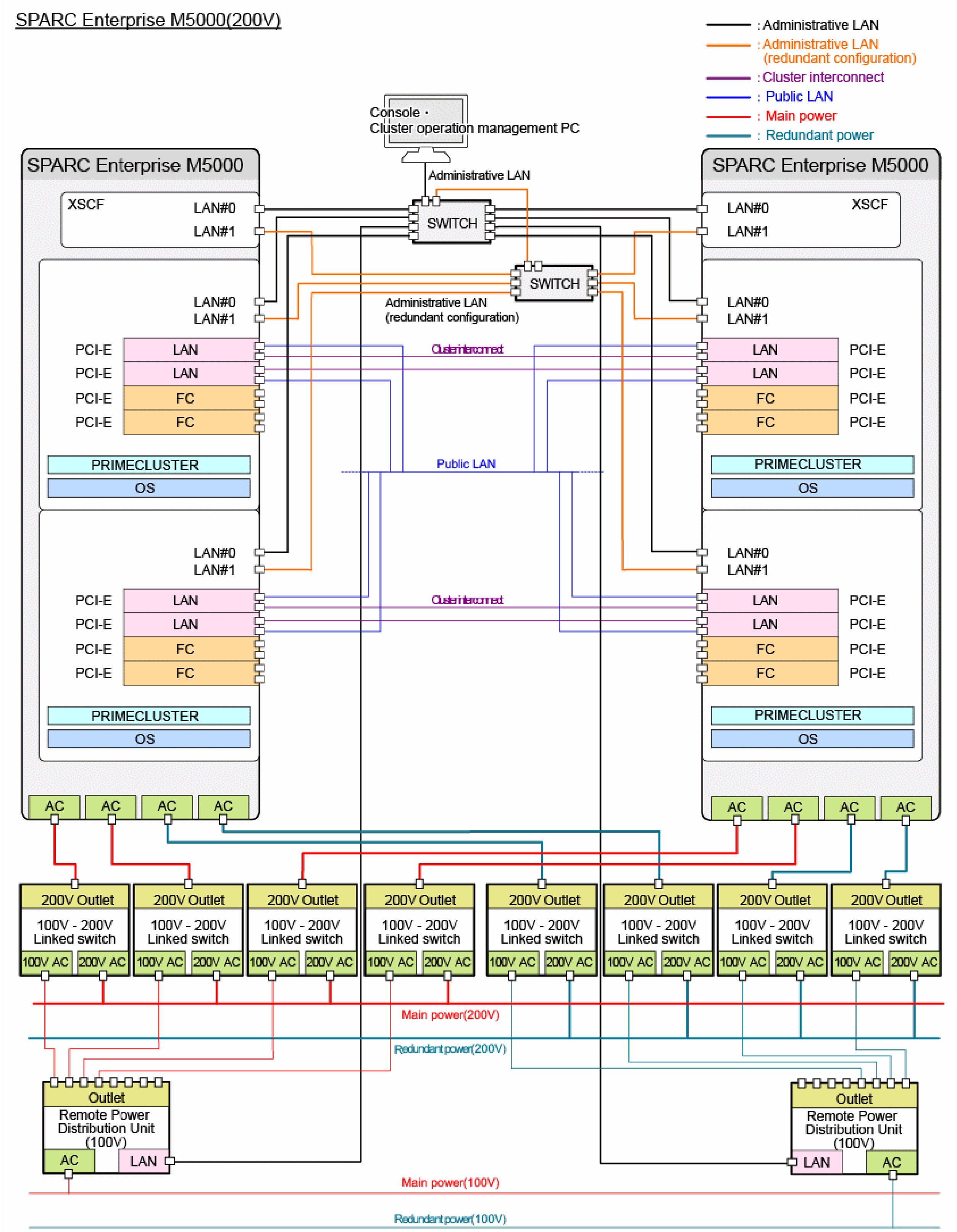


- SPARC Enterprise M5000(200V)
    - Two for up to four Remote Power Distribution Units(100V) SP0X9RP1F SPARC Enterprise M5000
    - Four for one 100V-200V linked switch (PDU-224 by ISA (Recommended)) SPARC Enterprise M5000

Note

In addition to 200V power supply, 100V power supply for control is required in this configuration.
If the 100V power supply is cut, the 200V outlet, linked with the power supply, will also be OFF.

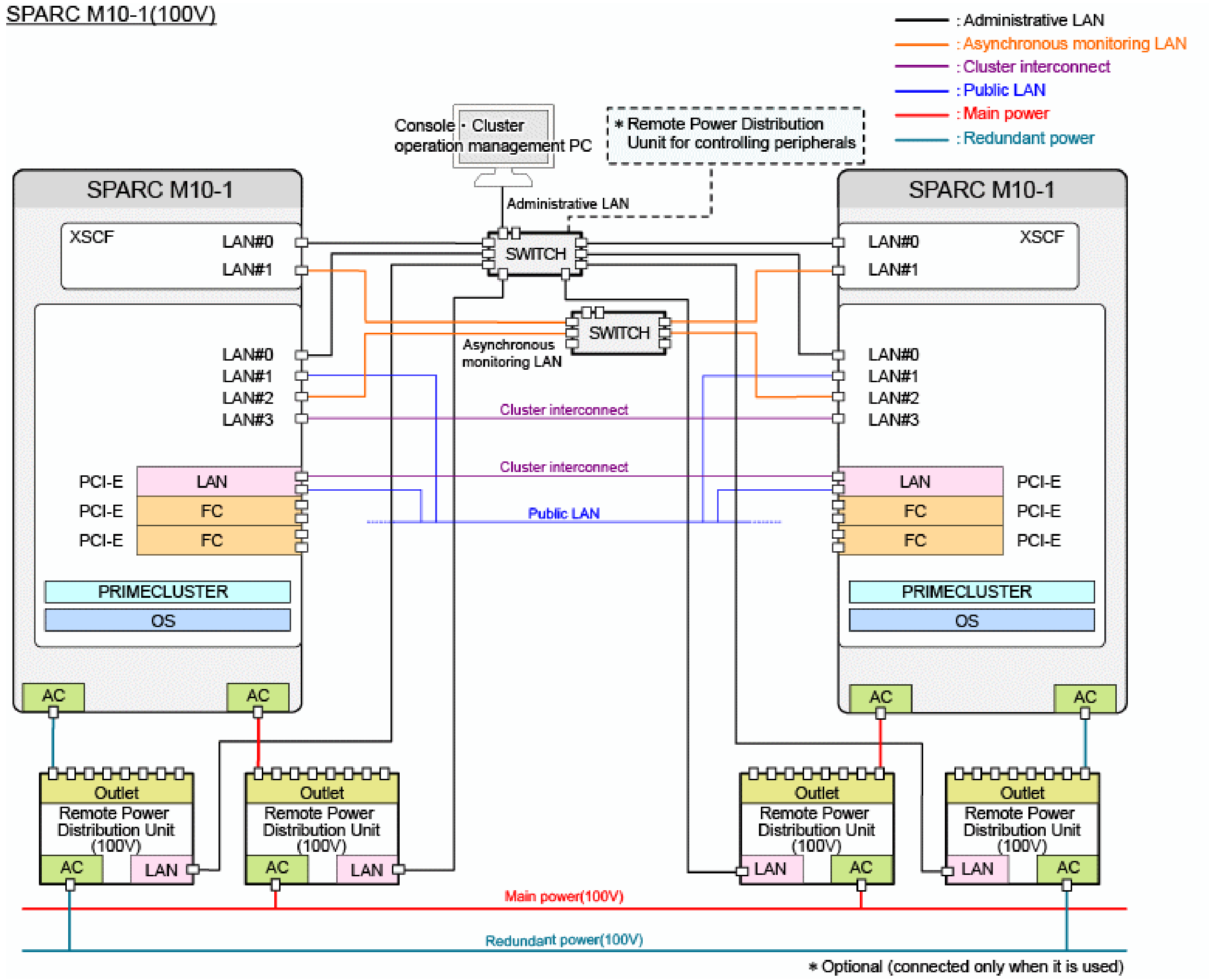
Figure J.4 Connection diagram for SPARC Enterprise M5000(200V)



- SPARC M10-1(100V)

Two for one Remote Power Distribution Unit (100V) SP0X9RP1F SPARC M10-1

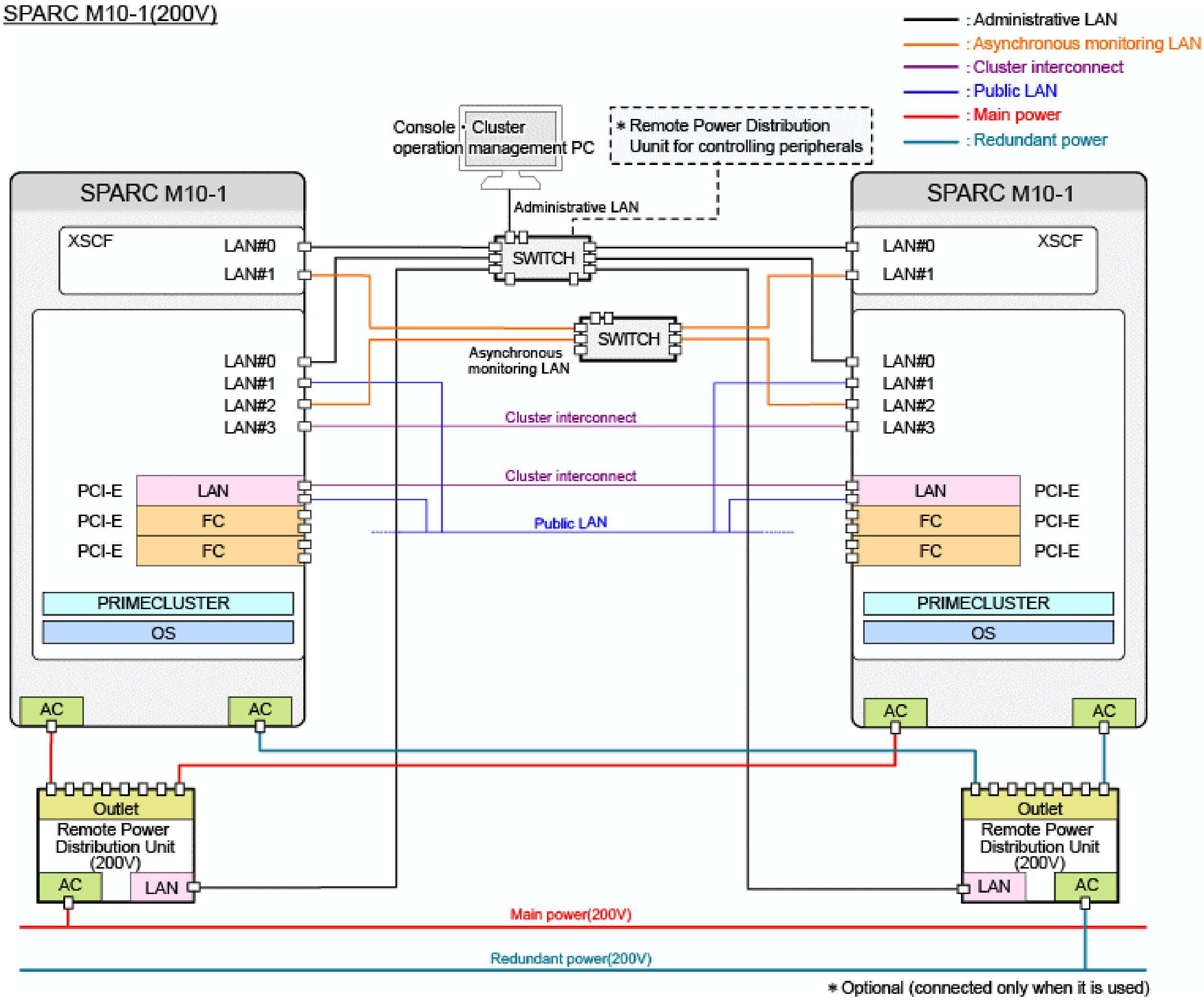
Figure J.5 Connection diagram for SPARC M10-1(100V)



- SPARC M10-1(200V)

  Two for up to five Remote Power Distribution Units (200V) SP0X9RP2F SPARC M10-1

Figure J.6 Connection diagram for SPARC M10-1(200V)



- SPARC M10-4 (200V)

  - Two for up to eight Remote Power Distribution Units (100V) SP0X9RP1F SPARC M10-4

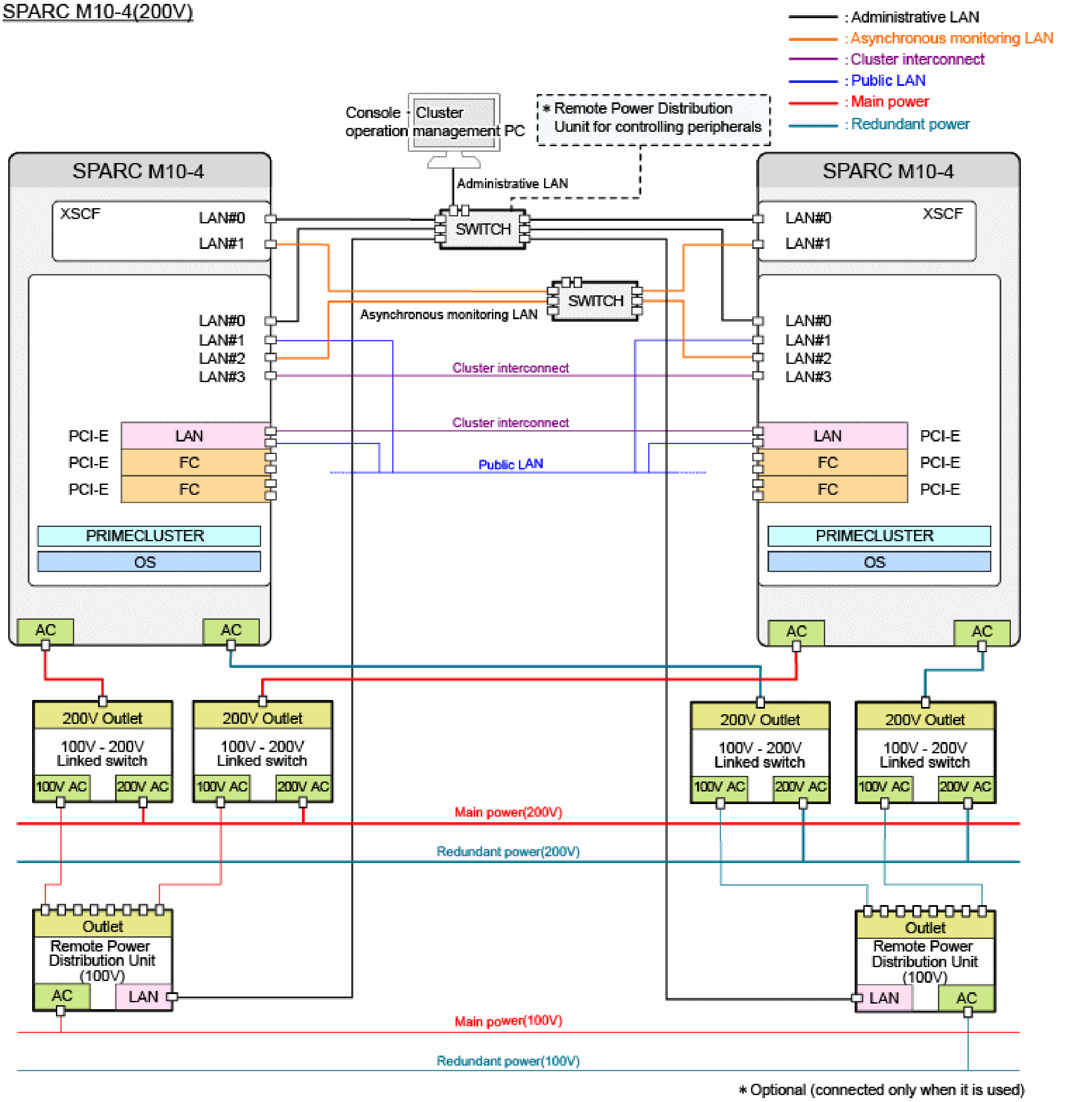
  - Two for one 100V-200V linked switch (PDU-224 by ISA (Recommended)) SPARC M10-4

Note

In addition to 200V power supply, 100V power supply for control is required in this configuration.
If the 100V power supply is cut, the 200V outlet, linked with the power supply, will also be OFF.

-

Figure J.7 Connection diagram for SPARC M10-4(200V)



J.1.3 Notes

This section explains notes on using the Remote Power Distribution Unit:

- SPARC M10-4S is unsupported.
- Connect the Remote Power Distribution Unit only to the cluster node.

  When connecting it to peripherals such as a shared disk, prepare the Remote Power Distribution Unit for peripherals separately.
- Connect the LAN for the Remote Power Distribution Unit to the administrative LAN.
- This function is only performed when the node could not be shut down by the option hardware facilities such as RCI or XSCF.
- When using it for SPARC Enterprise M4000 and M5000, all partitions of the node to be forcibly stopped are forcibly stopped. Shutdown is not performed at each partition.

- Virtualized environment is unsupported.

  When using PRIMECLUSTER in Oracle VM Server for SPARC environment and Oracle Solaris Kernel Zones environment, the Remote Power Distribution Unit cannot be used.

- If an error occurs in one Remote Power Distribution Unit connected among two of them that is connected to the node, the node's power cannot be cut and the stop of the node may not guaranteed. In this case, to prevent the conflict of user resources, the node is manually switched, not forcibly switched for this case.

J.2 Setup Procedure

This section explains how to setup the cluster system using the Remote Power Distribution Unit.

J.2.1 Installing Remote Power Distribution Unit

Stop all cluster nodes and connect the Remote Power Distribution Unit to the cluster nodes.

For the connection between cluster nodes and the Remote Power Distribution Unit, see "J.1.2 Hardware Configuration."

J.2.2 Setting up Remote Power Distribution Unit

Connect the LAN for the Remote Power Distribution Unit to the administrative LAN. When you set the LAN port for the Remote Power Distribution Unit, set the network of the same subnet as the administrative LAN.

J.2.3 Creating a User for Shutdown Facility

Since PRIMECLUSTER controls the Remote Power Distribution Unit, you must create a user for the shutdown facility.

Moreover, you must set the system and outlet permissions for target shutdown facility.

Perform the following:

Creating a user for Remote Power Distribution Unit

1. Enter the following URL from a browser such as Microsoft Internet Explorer or Mozilla Firefox to connect to the Remote Power Distribution Unit Web interface.

   http(s)://<IP address>

   *<IP address> is the IP address specified for the initial setting for the Remote Power Distribution Unit.

2. If a security warning message is displayed, click **OK** or **Yes**.

3. The Login page is displayed. Enter a username and password for login.

4. From the Remote Power Distribution Unit Web interface menu, select User Management > Users & Groups to display the Users/Group Management page.

5. On the Group Management panel in the Users/Group Management page, enter the following information, and then click **Create**.

Item	Setting value
New Group Name	pcl

   Check that the following message indicating the group creation has been completed, appears at the top of the page.

   Group created successfully.

6. On the User Management panel in the Users/Group Management page, enter the following information, and then click **Create**.

Item	Setting value
New user name	pcl
Full Name	pcl

Item	Setting value
Password Confirm Password	Optional *Only 4 - 32 characters can be specified. *Only one-byte characters and symbols The available characters can be used for the password, not a space.
User Group	pcl

Check that the following message indicating the user creation has been completed, appears at the top of the page.

User created successfully.

Setting up system permissions

1. From the Remote Power Distribution Unit Web interface menu, select User Management > User/Group System Permissions to display the User/Group System Permissions page.
2. From the User/Group System Permissions page, select the pcl group from the Group drop list. The permissions that can be applied for the pcl group are displayed.

   Enter the following information and click Apply.

Item	Setting value
IPMI Privilege Level	OEM
Log View	Yes
Server Status via IPMI	Yes
Unit & Outlet Configuration	Yes

   Check that the following message indicating that system permissions have been set up, appears at the top of the page.

   Operation completed successfully.

Setting up outlet permissions

1. From the Remote Power Distribution Unit Web interface menu, select User Management > User/Group Outlet Permissions to display the User/Group Outlet Permissions page.
2. From the User/Group Outlet Permissions page, select the pcl group from the Group drop list. The permissions that can be applied for the pcl group are displayed.

   Set "Yes" for the outlet number to connect the node, and then click Apply.

   Check that the following message indicating the outlet permission has been completed, appears at the top of the page.

   Operation completed successfully.

J.2.4 Starting a Cluster Node

Start a cluster node.

1. From the Remote Power Distribution Unit Web interface, display the Home page.
2. The power status for each outlet is displayed in the Outlets list.

   Click ON for the outlet number to which the cluster node is connected.
3. A dialog box for checking the operation is displayed. Click OK, and then the outlet power turns ON.
4. From the Remote Power Distribution Unit Web interface menu, select the Home page to update the screen. Check that the outlet number of the outlet, which was turned ON in step 2 and 3, is now ON.
5. Perform the step 2 to 4 in all outlets to which the cluster node is connected and turn the power ON.

6. Turn the power switch of the cluster node ON for startup.

J.2.5 Setting the Shutdown Facility

The following table indicates shutdown agents necessary for using the Remote Power Distribution Unit.

Table J.1 Shutdown Agents necessary for using the Remote Power Distribution Unit (only in physical environment)

<table>
<tr><th colspan="4" rowspan="2">Server machine type name</th><th rowspan="2">RPDU</th><th colspan="3">XSCF SNMP</th><th colspan="2">RCI</th><th colspan="3">XSCF</th></tr>
<tr><th>Panic</th><th>Reset</th><th>PPAR Reset</th><th>Panic</th><th>Reset</th><th>Panic</th><th>Reset</th><th>Break</th></tr>
<tr><td rowspan="2">SPARC Servers</td><td colspan="3">M10-1</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="3">M10-4</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td rowspan="3">SPARC Enterprise</td><td rowspan="3">M3000
M4000
M5000</td><td rowspan="2">Japan</td><td>Fujitsu</td><td>Y</td><td>-</td><td>-</td><td>-</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td></tr>
<tr><td>Other than Fujitsu</td><td>Y</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>Y</td><td>Y</td><td>Y</td></tr>
<tr><td colspan="2">Other than Japan</td><td>Y</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>Y</td><td>Y</td><td>Y</td></tr>
</table>

Check the hardware type and see "5.1.2.1 For SPARC M10" or "5.1.2.2 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000."

J.2.6 Checking the Operation Performance

After setting up the RPDU shutdown agent, check the operation performance to check that the target node can be forcibly stopped.

For how to check the operation performance of the RPDU shutdown agent, see "J.3 Test."

J.2.7 Restoration Method

This section explains how to restore the cluster node if it was forcibly stopped by the RPDU shutdown agent. There are the following two restoration procedures. Use one of the methods to turn the outlet power to which the forcibly stooped cluster node ON.

- Restoration by the sfrpdupoweron command
- Restoration from the Remote Power Distribution Unit Web interface

Restoration by the sfrpdupoweron command

The following explains the procedure for restoring the forcibly stopped cluster node by the sfrpdupoweron command.

For the sfrpdupoweron command, see the manual page of the sfrpdupoweron(1M) command.

1. Log in with one of the cluster nodes and perform the following with the system administrator privilege.
2. Execute the following command to check the outlet statuses to which the cluster node that its power to be restored is connected.

```
# /opt/SMAW/SMAWsf/bin/sfrpdupoweron -l CF-node-name
```

3. Execute the following command to turn the outlet power to which the cluster node is connected ON.

```
# /opt/SMAW/SMAWsf/bin/sfrpdupoweron -p CF-node-name
```

4. Execute the following command to check that the outlet number that its power was turned ON in step 3 is now ON.

```
# /opt/SMAW/SMAWsf/bin/sfrpdupoweron -l CF-node-name
```

Restoration from the Remote Power Distribution Unit Web interface

The following explains the procedure for restoring the forcibly stopped cluster node from the Remote Power Distribution Unit Web interface.

1. Enter the following URL from a browser such as Microsoft Internet Explorer or Mozilla Firefox to connect to the Remote Power Distribution Unit Web interface to which the cluster node that its power is to be restored is connected.

   http(s)://<IP address>

   *<IP address> is the IP address specified for the initial setting for the Remote Power Distribution Unit.

2. If a security warning message is displayed, click **OK** or **Yes**.
3. The Login page is displayed. Enter a username and password for login.
4. From the Remote Power Distribution Unit Web interface menu, display the Home page.
5. The power status for each outlet is displayed in the Outlets list.

   Click **ON** for the number of the outlet that recovers the power.

6. A dialog box for checking the operation is displayed. Click **OK**, and then the outlet power turns ON.
7. From the Remote Power Distribution Unit Web interface menu, select the Home page to update the screen. Check that the outlet number of the outlet, which was turned ON in step 5 and 6, is now ON.

J.3 Test

When using the Remote Power Distribution Unit, check that the RPDU shutdown agent is correctly set by using the following procedure:

1. Shut down a standby node. For SPARC Enterprise M4000 and M5000, shut down all partitions. For SPARC M10, shut down all domains.
2. In the Remote Power Distribution Unit Web interface, connected to the main power and redundant power of the standby node, turn the outlet number to which the standby node is connected OFF.
   1. Enter the following URL from a browser such as Microsoft Internet Explorer or Mozilla Firefox to connect to the Remote Power Distribution Unit Web interface.

      http(s)://<IP address>

      *<IP address> is the IP address specified for the initial setting for the Remote Power Distribution Unit.

   2. If a security warning message is displayed, click **OK** or **Yes**.
   3. The Login page is displayed. Enter the created pcl user and password for login.
   4. From the Remote Power Distribution Unit Web interface menu, display the Home page.
   5. The power status for each outlet is displayed in the Outlets list. Click **OFF** for the outlet number to which the cluster node is connected.

      If you cannot click **OFF**, the settings of "System Permission" and "Outlet Permission" for the pcl user may be insufficient.

      See "J.2.3 Creating a User for Shutdown Facility" to review the permission setting for the pcl user.

   6. A dialog box for checking the operation is displayed. Click **OK**, and then the outlet power turns OFF.
   7. From the Remote Power Distribution Unit Web interface menu, select the Home page to update the screen. Check that the outlet number of the outlet, which was turned OFF in step 5 and 6, is now OFF.
3. Execute the following command in the operation node.

```
# /opt/SMAW/SMAWsf/bin/sfrpdupoweron -p CF-node-name-of-the-standby-node
```

   If the following message was output, the settings of "System Permission" and "Outlet Permission" for the pcl user may be insufficient.

   See "J.2.3 Creating a User for Shutdown Facility" to review the permission setting for the pcl user.

   outlet power-on failed.

4. Check that the outlet number on which its power was turned OFF in step 2, is now ON. If not, review the setting for the shutdown facility.

5. For SPARC Enterprise M4000 and M5000, perform step 2 to 4 on all partitions used the RPDU shutdown agent.

6. Start the standby node.

7. After starting the standby node, shut down the operation node. For SPARC Enterprise M4000 and M5000, shut down all partitions. For SPARC M10, shut down all domains.

8. Turn the power of the outlet number to which the operation node is connected OFF in the Remote Power Distribution Unit Web interface.

9. Execute the following command on the standby node.

   ```
   # /opt/SMAW/SMAWsf/bin/sfrpdupoweron -p CF-node-name-of-the-operation-node
   ```

   If the following message was output, the settings of "System Permission" and "Outlet Permission" for the pcl user may be insufficient.

   See "J.2.3 Creating a User for Shutdown Facility" to review the permission setting for the pcl user.

   outlet power-on failed.

10. Check that the outlet number to which its power was turned ON in step 8 in the Remote Power Distribution Unit Web interface. If not, review the setting for the shutdown facility.

11. For SPARC Enterprise M4000 and M5000, perform step 8 to 10 in all the partitions where the RPDU shutdown agent is used.

12. Start the operation node.

Appendix K Changes in Each Version

This chapter explains the changes made to the specifications of PRIMECLUSTER 4.3A40.

The changes are listed in the following table.

Table K.1 List of changes

Category	Item	Version
Incompatible commands	hvshut command	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	hvswitch command	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	sdtool command	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.3A40
	clrccusetup command	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.3A40
	hvdump command	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.3A40
Incompatible functions	MAC address takeover	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Operator intervention request	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Behavior of CF startup	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	IP Interconnect	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Setting Up Fsystem Resources	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Shutdown Facility	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Procedure resources	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	HV_CONNECT_TIMEOUT	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Shutdown Configuration Wizard	(Before change) PRIMECLUSTER 4.2A00 or earlier, or PRIMECLUSTER 4.3A10 (After change) PRIMECLUSTER 4.3A40
	Shutdown agent selection screen for the shutdown configuration wizard	(Before change) PRIMECLUSTER 4.2A00 or earlier, PRIMECLUSTER 4.3A10, or PRIMECLUSTER 4.3A20 (After change) PRIMECLUSTER 4.3A40
	CF wizard	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.3A40

Category	Item	Version
	Posting Notification of a Resource Failure or Recovery	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.3A40
	Changes of the port numbers for SNMP	(Before change) PRIMECLUSTER 4.3A20 (After change) PRIMECLUSTER 4.3A40
	Display of the resource fault trace	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.3A40
	Changes of the target node to forcibly shut down when a heartbeat failure occurs	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.3A40
Incompatible messages	Changes of RMS console message	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Changes of the response message for the operator intervention request	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Changes of the RMS message	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Changes of the importance of the message in the RMS wizard	(Before change) PRIMECLUSTER 4.2A00 or earlier (After change) PRIMECLUSTER 4.3A40
	Messages of the shutdown configuration wizard	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.3A40
	Method to display the messages of the shutdown configuration wizard	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.3A40

K.1 Changes in PRIMECLUSTER 4.3A40 from 4.2A00

Incompatible commands

The following commands of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.2A00.

- K.1.1 hvshut command
- K.1.2 hvswitch command
- K.1.3 sdtool command
- K.1.4 clrccusetup command
- K.1.5 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.2A00.

- K.1.6 MAC address takeover
- K.1.7 Operator intervention request
- K.1.8 Behavior of CF startup
- K.1.9 IP Interconnect
- K.1.10 Setting Up Fsystem Resources
- K.1.11 Shutdown Facility

- K.1.12 Procedure Resources
- K.1.13 HV_CONNECT_TIMEOUT
- K.1.14 Shutdown Configuration Wizard
- K.1.15 Shutdown agent selection screen for the shutdown configuration wizard
- K.1.16 CF wizard
- K.1.17 Posting Notification of a Resource Failure or Recovery
- K.1.18 Display of the resource fault trace
- K.1.19 Changes of the target node to forcibly shut down when a heartbeat failure occurs

Incompatible messages

The following messages of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.2A00.

- K.1.20 Changes of RMS console message
- K.1.21 Changes of the response message for the operator intervention request
- K.1.22 Changes of the RMS message
- K.1.23 Changes of the importance of the message in the RMS wizard
- K.1.24 Messages of the shutdown configuration wizard
- K.1.25 Method to display the messages of the shutdown configuration wizard

K.1.1 hvshut command

Details on incompatibilities

The default value of the environment variable RELIANT_SHUT_MIN_WAIT, which sets the timeout duration of the hvshut command, is changed from 900 (seconds) to 2147483647 (seconds). With this change, even if you leave the environment variable to default, the command will not timeout.

Point

A resource in a cluster application does not stop and may remain running because the RMS ends abnormally when the hvshut command times out.

In this situation, data corruption may occur when RMS and cluster application with the resource is forcibly started on another node, if shared disk is controlled by the resource. This is because the resource is started on multiple nodes at the same time.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

If the resource offline processing invoked by the hvshut command is not completed within 900 seconds in an environment where the RELIANT_SHUT_MIN_WAIT environment variable is set to its default value, the command times out and then RMS terminates abnormally. In this case, the resource remains running without stopped.

After upgrading [PRIMECLUSTER 4.3A40]

In an environment where the RELIANT_SHUT_MIN_WAIT environment variable is set to its default value, the command will not time out even if stop process using the hvshut command is not completed.

Note

When using RMS, you need to change the environment variable to the value corresponding to the configuration setup.

K.1.2 hvswitch command

Details on incompatibilities

In the forced startup (when using -f option) of a cluster application is issued, data corruption may occur if you start cluster applications when nodes where RMS is not running exist in the cluster. Therefore, to deal with this issue, the function is added. This function forcibly shuts down the nodes where RMS is not running before forced startup of cluster applications.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

When using -f option, RMS performs forced startup of cluster applications even if nodes where RMS is not running exist in the cluster and it may lead to data corruption.

After upgrading [PRIMECLUSTER 4.3A40]

In the use of -f option, when nodes where RMS is not running exist in the cluster, RMS performs the forced startup cluster applications after forcibly shutting down the nodes for reducing the risk of data corruption. However, if RMS failed to the forced shutdown, the forced startup of cluster applications are not performed.

Note

When using -f option, confirm "7.5.1 Notes on Switching a Cluster Application Forcibly " and then execute the command.

K.1.3 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The number of characters displayed by "Agent" of "sdtool -s" is 14 characters (including spaces).

The number of characters displayed by "Admin IP" of "sdtool -C" is 16 characters (including spaces).

After upgrading [PRIMECLUSTER 4.3A40]

The number of characters displayed by "Agent" of "sdtool -s" is 21 characters (including spaces).

When an IPv6 address is used for the administrative LAN of the shutdown facility, the number of characters displayed by "Admin IP" of "sdtool -C" is 40 characters (including spaces). When an IPv4 address is used, the number of characters is not changed.

Note

None

K.1.4 clrccusetup command

Details on incompatibilities

The number of characters displayed by "clrccusetup -l" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The number of characters displayed by "IP-address" of "clrccusetup -l" is 16 characters (including spaces).

After upgrading [PRIMECLUSTER 4.3A40]

When an IPv6 address is used for the IP address of the console, the number of characters displayed by "IP-address" of "clrccusetup -l" is 40 characters (including spaces). When an IPv4 address is used, the number of characters is not changed.

Note

None

K.1.5 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.3A40]

The default work directory is /var/tmp.

Note

None.

K.1.6 MAC address takeover

Details on incompatibilities

MAC address takeover is not supported.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

MAC address takeover is supported.

After upgrading [PRIMECLUSTER 4.3A40]

MAC address takeover is not supported.

Note

None.

K.1.7 Operator intervention request

Details on incompatibilities 1

In the forced startup of a cluster application is issued, data corruption may occur if you start cluster applications when nodes without running RMS exist in the cluster.

Therefore, to deal with issue, the function is added. This function forcibly shuts down the nodes without running RMS before forced start the cluster application.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

In the forced startup of a cluster application is issued, even if the nodes without running RMS exist in the cluster and it may cause the data corruption, forcibly starts the cluster application according to the user's operation.

After upgrading [PRIMECLUSTER 4.3A40]

For reducing the risk of data corruption in the forced startup of a cluster application is issued, forcibly starts the cluster application after forcibly shuts down the nodes without running RMS.

Note

Read through the items in "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

Details on incompatibilities 2

With the default settings made when the cluster was installed, the operator intervention request is always enabled. For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The operator intervention request will not work with the default setting at installation. The default value of AppWatch set when the cluster was installed is set to OFF, and the operator intervention request will not work with this default value.

After upgrading [PRIMECLUSTER 4.3A40]

The operator intervention request will work with the default setting at installation. The operator intervention request, is disabled only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.1.8 Behavior of CF startup

Details on incompatibilities

CF starts even if some of the network interfaces for the cluster interconnects are not recognized.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.3A40]

CF starts if at least one of the network interfaces for the cluster interconnects is recognized.

Note

If there are any network interfaces that are not recognized on CF startup, the following message appears:

CF: <NIC>: device not found.

<NIC> will be the name of the network interface (e.g.hme0).

This message is also available in 4.2A00.

K.1.9 IP Interconnect

Details on incompatibilities

IP interconnect has not been supported in a Solaris 11 environment.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

IP interconnect is supported.

After upgrading [PRIMECLUSTER 4.3A40]

IP interconnect is not supported in a Solaris 11 environment.

Note

None.

K.1.10 Setting Up Fsystem Resources

Details on incompatibilities

The file which defines the mount point of the file system used as Fsystem resource and the entry of NFS has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The mount point was defined in /etc/vfstab.

The entry of NFS was defined in /etc/dfs/dfstab.

After upgrading [PRIMECLUSTER 4.3A40]

It is necessary to define the mount point in /etc/vfstab.pcl.

It is necessary to define the entry of NFS in/etc/dfs/dfstab.pcl.

For details, see "6.7.1.2 Creating Fsystem Resources."

Note

None.

K.1.11 Shutdown Facility

Details on incompatibilities

Even if "TestFailed" is displayed by the shutdown agent test, a test is conducted 10 minutes afterward.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

If "TestFailed" is displayed, a test is not conducted 10 minutes afterward.

After upgrading [PRIMECLUSTER 4.3A40]

Even if "TestFailed" is displayed, a test is conducted 10 minutes afterward.

Note

None.

K.1.12 Procedure Resources

Details on incompatibilities

If a process is started without process monitoring from a state transition procedure, the hard- and software limits of the file descriptor (fd) for the process are 1024 or more.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The hard- and software limits of the file descriptor (fd) for the process that is started without process monitoring from a state transition procedure are set to 1024.

After upgrading [PRIMECLUSTER 4.3A40]

If a process is started without process monitoring from a state transition procedure, the hard- and software limits of the file descriptor (fd) for the process are as follows:

- If the hard- and software limits of the file descriptor for the system are less than 1024, the limits are set to 1024.
- If the hard- and software limits of the file descriptor for the system are 1024 or more, the limits are not changed.

Note

In the process started from a state transition procedure, if you need to set the value less than 1024 for the hard- and software limits of the file descriptor, add a "ulimit" definition at the beginning of processing for the state transition procedure, and specify the hard- and software limits of the file descriptor.

K.1.13 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds).

After upgrading [PRIMECLUSTER 4.3A40]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in the "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide".

K.1.14 Shutdown Configuration Wizard

There are incompatibilities of a setup of the shutdown facility and the shutdown configuration wizard in the following models:

- K.1.14.1 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000
- K.1.14.2 For SPARC Enterprise T1000, T2000
- K.1.14.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series

K.1.14.1 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000

Details on incompatibilities

Changed the screen of the shutdown configuration wizard to set up the shutdown facility for SPARC Enterprise M3000, M4000, M5000, M8000, or M9000.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

- There were selection items of *SCON* from the selection screen of a shutdown agent.
- Select *RCI Panic*, *XSCF Panic*, *Console Break*, *RCI Reset*, and *XSCF Reset* from the selection screen of a shutdown agent and make settings.
- In the setup screen of Wait for PROM, the check box of *Wait for PROM* has checked by default.
- Set a timeout value of a shutdown agent in the setup screen of the timeout value.
- One IP address of XSCF can be set.

After upgrading [PRIMECLUSTER 4.3A40]

- Deleted *SCON* from the selection screen of the shutdown agent.

- Added *XSCF* (*SPARC M10*) to the selection screen of the shutdown agent.
- Changed to *XSCF* (*SPARC Enterprise M-series*) from *XSCF* on the selection screen of the shutdown agent.
- When *XSCF* (*SPARC M10*) is selected from the selection screen of a shutdown agent, *XSCF Domain Panic*, *XSCF Domain Reset*, and *XSCF PPAR Reset* are automatically set according to the hardware state.
- When *XSCF* (*SPARC Enterprise M-series*) is selected from the selection screen of a shutdown agent, *XSCF Panic*, *XSCF Break*, and *XSCF Reset* are automatically set.
  When *XSCF* (*SPARC Enterprise M-series*) and *Use RCI* are selected from the selection screen of a shutdown agent, *RCI Panic*, *XSCF Panic*, *XSCF Break*, *RCI Reset*, and *XSCF Reset* are automatically set.
- In the setup screen of Wait for PROM, the checkbox of *Wait for PROM* is not checked by default.
- Timeout value of a shutdown agent is automatically set.
- Up to two IP addresses of XSCF can be set.

Note

The *XSCF* supported in 4.2A00 is changed to *XSCF* (*SPARC Enterprise M-series*).

K.1.14.2 For SPARC Enterprise T1000, T2000

Details on incompatibilities

The shutdown configuration wizard is available to set up the shutdown facility for SPARC Enterprise T1000 or T2000.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

Setup of the shutdown facility for SPARC Enterprise T1000 or T2000 is performed by CLI.

After upgrading [PRIMECLUSTER 4.3A40]

Setup of the shutdown facility for SPARC Enterprise T1000 or T2000 is performed by the shutdown configuration wizard.

Note

None.

K.1.14.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series

Details on incompatibilities

The shutdown configuration wizard is available to set up the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

Setup of the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series is performed by CLI.

After upgrading [PRIMECLUSTER 4.3A40]

Setup of the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series is performed by the shutdown configuration wizard.

Note

None.

K.1.15 Shutdown agent selection screen for the shutdown configuration wizard

Details on incompatibilities

The shutdown agent selection screen for the shutdown configuration wizard has been changed to support SPARC M10 and Oracle Solaris Kernel Zones.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

You can select the following SA on the selection screen of the shutdown agent:

- XSCF

After upgrading [PRIMECLUSTER 4.3A40]

You can select the following SAs on the selection screen of the shutdown agent:

- XSCF (SPARC M10)
- XSCF (SPARC Enterprise M-series)
- ILOM
- ALOM
- KZONE(SPARC M10)
- KZONE(SPARC T4)

Note

None.

K.1.16 CF wizard

Details on incompatibilities

The only lowercase characters are allowed to set the beginning of the CF node names on the screen to edit CF node names of the CF wizard.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

If a CF node name begins with a letter other than lowercase characters, you could enter a numerical value or a symbol at the beginning of the CF node name, even though RMS Wizard Tools becomes unavailable to operate.

After upgrading [PRIMECLUSTER 4.3A40]

If a CF node name begins with a letter other than lowercase characters, the process to check the input value is added, because RMS Wizard Tools becomes unavailable to operate.

If you enter a letter other than lowercase characters at the beginning of the CF node name, the following message is displayed:

message .2978

The first letter of the CF node name "{0}" is a letter other than lowercase characters.

RMS Wizard Tools cannot operate with this setting.

Specify a name beginning with a lowercase character.

Note

If the error message of 2978 is displayed, specify a lowercase character at the beginning of the CF node name.

K.1.17 Posting Notification of a Resource Failure or Recovery

Details on incompatibilities

The default setting at installation is that notification of a resource failure or recovery is posted with PRIMECLUSTER 4.3A40. For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Message No.	Message overview
2700	Recovering from a resource failure
2701	Recovering from a node failure
6750	Resource failure
6751	Node failure

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

Notification of a resource failure or recovery will be not posted in the default setting of cluster installation.

The default value of AppWatch at cluster installation is OFF and notification of the resource failure or recovery will not be posted.

After upgrading [PRIMECLUSTER 4.3A40]

Notification of a resource failure or recovery will be posted in the default setting of cluster installation.

A resource failure or recovery will not be posted only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.1.18 Display of the resource fault trace

Details on incompatibilities

When the resource is failed, the display of StateDetails of the failed resource object is changed. As a result, it can be able to distinguish the failed resource.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The StateDetails information is not displayed.

After upgrading [PRIMECLUSTER 4.3A40]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

K.1.19 Changes of the target node to forcibly shut down when a heartbeat failure occurs

Details on incompatibilities

The selecting method of the target node, which is forcibly shut down when a heartbeat failure occurs by temporary causes such as the overloaded, is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility determines the node to forcibly shut down according to the setup policy for survival priority.

After upgrading [PRIMECLUSTER 4.3A40]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility forcibly stops the node on which CF cannot perform regardless of the setup policy for survival priority.

Note

None.

K.1.20 Changes of RMS console message

Details on incompatibilities

Because of the change mentioned in "K.1.2 hvswitch command", hvswitch -f outputs different console message from that of previous versions.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The use of the -f (force) flag could cause your data to be corrupted and could cause your node to be killed. Do not continue if the result of this forced command is not clear.

The use of force flag of hvswitch overrides the RMS internal security mechanism. In particular RMS does no longer prevent resources, which have been marked as "ClusterExclusive", from coming Online on more than one host in the cluster. It is recommended to double check the state of all affected resources before continuing.

Do you wish to proceed? (default: no) [yes, no]:

After upgrading [PRIMECLUSTER 4.3A40]

The use of the -f (force) flag could cause your data to be corrupted and could cause your node to be killed. Do not continue if the result of this forced command is not clear.

The use of force flag of hvswitch overrides the RMS internal security mechanism. In particular RMS does no longer prevent resources, which have been marked as "ClusterExclusive", from coming Online on more than one host in the cluster. It is recommended to double check the state of all affected resources before continuing.

IMPORTANT: This command may kill nodes on which RMS is not running in order to reduce the risk of data corruption!

Ensure that RMS is running on all other nodes. Or shut down OS of the node on which RMS is not running.

Do you wish to proceed ? (default: no) [yes, no]:

Note

None.

K.1.21 Changes of the response message for the operator intervention request

K.1.21.1 message 1421

Details on incompatibilities

The response message of the operator intervention request 1421 has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

1421 The userApplication "*userApplication*" did not start automatically because not all of the nodes where it can run are online.
Do you want to force the userApplication online on the SysNode "*SysNode*"?
Message No.: *number*
Do you want to do something? (yes/no)
Warning: Forcing a userApplication online ignores potential error conditions. Used improperly, it can result in data corruption. You should not use it unless you are certain that the userApplication is not running anywhere in the cluster.

After upgrading [PRIMECLUSTER 4.3A40]

1421 The userApplication "*userApplication*" did not start automatically because not all of the nodes where it can run are online.
Forcing the userApplication online on the SysNode "*SysNode*" is possible.
Warning: When performing a forced online, confirm that RMS is started on all nodes in the cluster, manually shutdown any nodes where it is not started and then perform it.For a forced online, there is a risk of data corruption due to simultaneous access from several nodes.In order to reduce the risk, nodes where RMS is not started maybe forcibly stopped.
Are you sure wish to force online? (no/yes)
Message No.: *number*

Note

Read through the items for the applicable messages in "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

K.1.21.2 message 1423

Details on incompatibilities

The response message of the operator intervention request 1423 has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

1423 On the SysNode "*SysNode*", the userApplication "*userApplication*" has the faulted resource "*resource*". The userApplication "*userApplication*" did not start automatically because not all of the nodes where it can run are online.
Do you want to force the userApplication online on the SysNode "*SysNode*"?
Message No.: *number*
Do you want to do something? (yes/no)
Warning: Forcing a userApplication online ignores potential error conditions. Used improperly, it can result in data corruption. You should not use it unless you are certain that the userApplication is not running anywhere in the cluster.

After upgrading [PRIMECLUSTER 4.3A40]

1423 On the SysNode "*SysNode*", the userApplication "*userApplication*" has the faulted resource "*resource*". The userApplication "*userApplication*" did not start automatically because not all of the nodes where it can run are online.
Forcing the userApplication online on the SysNode "*SysNode*" is possible.
Warning: When performing a forced online, confirm that RMS is started on all nodes in the cluster, manually shutdown any nodes where it is not started and then perform it. For a forced online, there is a risk of data corruption due to simultaneous access from several nodes. In order to reduce the risk, nodes where RMS is not started maybe forcibly stopped.
Are you sure wish to force online? (no/yes)
Message No.: *number*

Note

Read through the items for the applicable messages in "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

K.1.22 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

(SYS, 8): ERROR: RMS failed to shut down the host <host> via a Shutdown Facility, no further kill functionality is available.

The cluster is now hung.

After upgrading [PRIMECLUSTER 4.3A40]

(SYS, 8): ERROR: RMS failed to shut down the host <host> via a Shutdown Facility, no further kill functionality is available.

The cluster is now hung. An operator intervention is required.

Note

None.

K.1.23 Changes of the importance of the message in the RMS wizard

Details on incompatibilities

The importance of the following message in the RMS wizard has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.3A40]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

K.1.24 Messages of the shutdown configuration wizard

Details on incompatibilities

By defining the CF node name used on the shutdown configuration wizard as "CF node," messages displayed on nodes are changed.

Changes

The following messages are changed:

- No.2586
- No.2587
- No.2588
- No.2591
- No.2941
- No.2942
- No.2943
- No.2944
- No.2950
- No.2967
- No.2968

Note

None.

K.1.24.1 Message 2586

Before upgrading [PRIMECLUSTER 4.2A00]

Invalid CF name for node. Lowercase a-z, 0-9, _ and - are allowed.

After upgrading [PRIMECLUSTER 4.3A40]

Invalid CF node name for node. Lowercase a-z, 0-9, _ and - are allowed.

K.1.24.2 Message 2587

Before upgrading [PRIMECLUSTER 4.2A00]

The CF name for node1 and node2 are the same.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node1 and node2 are the same.

K.1.24.3 Message 2588

Before upgrading [PRIMECLUSTER 4.2A00]

The CF name for node is empty.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node is empty.

K.1.24.4 Message 2591

Before upgrading [PRIMECLUSTER 4.2A00]

The CF name for node1 is the same as the public name of node2.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node1 is the same as the public name of node2.

K.1.24.5 Message 2941

Before upgrading [PRIMECLUSTER 4.2A00]

You must enter weight for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must enter weight for each of the CF nodes.

K.1.24.6 Message 2942

Before upgrading [PRIMECLUSTER 4.2A00]

Invalid host weight entered.

The entered weight of the host is invalid.

After upgrading [PRIMECLUSTER 4.3A40]

Invalid CF node weight entered.

The entered weight of the CF node is invalid.

K.1.24.7 Message 2943

Before upgrading [PRIMECLUSTER 4.2A00]

You must enter admin IP for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must enter admin IP for each of the CF nodes.

K.1.24.8 Message 2944

Before upgrading [PRIMECLUSTER 4.2A00]

Host weight must be between 1 and 1000000.

After upgrading [PRIMECLUSTER 4.3A40]

CF node weight must be between 1 and 1000000.

K.1.24.9 Message 2950

Before upgrading [PRIMECLUSTER 4.2A00]

You must specify XSCF-Name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify XSCF-Name and User-Name for each of the CF nodes.

K.1.24.10 Message 2967

Before upgrading [PRIMECLUSTER 4.2A00]

You must specify ILOM-name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify ILOM-name and User-Name for each of the CF nodes.

K.1.24.11 Message 2968

Before upgrading [PRIMECLUSTER 4.2A00]

You must specify ALOM-name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify ALOM-name and User-Name for each of the CF nodes.

K.1.25 Method to display the messages of the shutdown configuration wizard

Details on incompatibilities

On the screen to set information of each shutdown agent, the message is displayed in the pop-up window when you enter different values for the password and confirmation.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

On the screen to set information of each shutdown agent, the message is displayed in the bottom right of the SF wizard screen when you enter different values for the password and confirmation.

After upgrading [PRIMECLUSTER 4.3A40]

On the screen to set information of each shutdown agent, the message is displayed in the pop-up window when you enter different values for the password and confirmation.

Note

None.

K.2 Changes in PRIMECLUSTER 4.3A40 from 4.3A10

Incompatible commands

The following commands of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.3A10.

- K.2.1 sdtool command
- K.2.2 clrccusetup command
- K.2.3 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.3A10.

- K.2.4 Shutdown Configuration Wizard
- K.2.5 Shutdown agent selection screen for the shutdown configuration wizard
- K.2.6 CF wizard
- K.2.7 Posting Notification of a Resource Failure or Recovery
- K.2.8 Operator Intervention Request
- K.2.9 Display of the resource fault trace
- K.2.10 Changes of the target node to forcibly shut down when a heartbeat failure occurs

Incompatible messages

The following messages of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.3A10.

- K.2.11 Messages of the shutdown configuration wizard
- K.2.12 Method to display the messages of the shutdown configuration wizard

K.2.1 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The number of characters displayed by "Agent" of "sdtool -s" is 14 characters (including spaces).

The number of characters displayed by "Admin IP" of "sdtool -C" is 16 characters (including spaces).

After upgrading [PRIMECLUSTER 4.3A40]

The number of characters displayed by "Agent" of "sdtool -s" is 21 characters (including spaces).

When an IPv6 address is used for the administrative LAN of the shutdown facility, the number of characters displayed by "Admin IP" of "sdtool -C" is 40 characters (including spaces). When an IPv4 address is used, the number of characters is not changed.

Note

None.

K.2.2 clrccusetup command

Details on incompatibilities

The number of characters displayed by "clrccusetup -l" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The number of characters displayed by "IP-address" of "clrccusetup -l" is 16 characters (including spaces).

After upgrading [PRIMECLUSTER 4.3A40]

When an IPv6 address is used for the IP address of the console information, the number of characters displayed by "IP-address" of "clrccusetup -l" is 40 characters (including spaces). When an IPv4 address is used, the number of characters is not changed.

Note

None.

K.2.3 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.3A40]

The default work directory is /var/tmp.

Note

None.

K.2.4 Shutdown Configuration Wizard

There are incompatibilities of a setup of the shutdown facility and the shutdown configuration wizard in the following models:

- K.2.4.1 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000
- K.2.4.2 For SPARC Enterprise T1000, T2000
- K.2.4.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series

For SPARC Enterprise M3000, M4000, M5000, M8000, and M9000, details on the incompatibilities are different depending on whether or not T006771SP-01 or later is applied to PRIMECLUSTER 4.3A10.

For SPARC Enterprise T1000, T2000, SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series, there are incompatibilities with environments where T006771SP-01 or later is applied to PRIMECLUSTER 4.3A10.

K.2.4.1 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000

Details on incompatibilities

Changed the screen of the shutdown configuration wizard to set up the shutdown facility for SPARC Enterprise M3000, M4000, M5000, M8000, or M9000.

Changes 1

Before upgrading [PRIMECLUSTER 4.3A10]

- There were selection items of *SCON* from the selection screen of a shutdown agent.
- Select *RCI Panic*, *XSCF Panic*, *Console Break*, *RCI Reset*, and *XSCF Reset* from the selection screen of a shutdown agent and make settings.
- In the setup screen of Wait for PROM, the check box of *Wait for PROM* has checked by default.
- Set a timeout value of a shutdown agent in the setup screen of the timeout value.
- One IP address of XSCF can be set.

After upgrading [PRIMECLUSTER 4.3A40]

- Deleted *SCON* from the selection screen of the shutdown agent.
- Changed to *XSCF* (*SPARC Enterprise M-series*) from *XSCF* on the selection screen of the shutdown agent.
- Added *XSCF* (*SPARC M10*) to the selection screen of the shutdown agent.
- When *XSCF* (*SPARC Enterprise M-series*) is selected from the selection screen of a shutdown agent, *XSCF Panic*, *XSCF Break*, and *XSCF Reset* are automatically set.
- When *XSCF* (*SPARC Enterprise M-series*) and *Use RCI* are selected from the selection screen of a shutdown agent, *RCI Panic*, *XSCF Panic*, *XSCF Break*, *RCI Reset*, and *XSCF Reset* are automatically set.
- In the setup screen of Wait for PROM, the checkbox of *Wait for PROM* is not checked by default.
- Timeout value of a shutdown agent is automatically set.
- Up to two IP addresses of XSCF can be set.

Changes 2

Before upgrading [PRIMECLUSTER 4.3A10 (T006771SP-01 or later is applied)]

- When *XSCF* is selected from the selection screen of a shutdown agent, *XSCF Panic*, *XSCF Break*, and *XSCF Reset* are automatically set.
- When *XSCF* and *Use RCI* are selected from the selection screen of a shutdown agent, *RCI Panic*, *XSCF Panic*, *XSCF Break*, *RCI Reset*, and *XSCF Reset* are automatically set.

After upgrading [PRIMECLUSTER 4.3A40]

- Changed to *XSCF* (*SPARC Enterprise M-series*) from *XSCF* on the selection screen of the shutdown agent.
- Added *XSCF* (*SPARC M10*) to the selection screen of the shutdown agent.
- When *XSCF* (*SPARC Enterprise M-series*) is selected from the selection screen of a shutdown agent, *XSCF Panic*, *XSCF Break*, and *XSCF Reset* are automatically set.
- When *XSCF* (*SPARC Enterprise M-series*) and *Use RCI* are selected from the selection screen of a shutdown agent, *RCI Panic*, *XSCF Panic*, *XSCF Break*, *RCI Reset*, and *XSCF Reset* are automatically set.

Note

XSCF supported in 4.3A10 is changed to *XSCF* (*SPARC Enterprise M-series*).

K.2.4.2 For SPARC Enterprise T1000, T2000

Details on incompatibilities

The shutdown configuration wizard is available to set up the shutdown facility for SPARC Enterprise T1000 or T2000.

Changes

Before upgrading [PRIMECLUSTER 4.3A10 (T006771SP-01 or later is not applied)]

Setup of the shutdown facility for SPARC Enterprise T1000 or T2000 is performed by CLI.

After upgrading [PRIMECLUSTER 4.3A40]

Setup of the shutdown facility for SPARC Enterprise T1000 or T2000 is performed by the shutdown configuration wizard.

Note

None.

K.2.4.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series

Details on incompatibilities

The shutdown configuration wizard is available to set up the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series.

Changes

Before upgrading [PRIMECLUSTER 4.3A10 (T006771SP-01 or later is not applied)]

Setup of the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series is performed by CLI.

After upgrading [PRIMECLUSTER 4.3A40]

Setup of the shutdown facility for SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3 series is performed by the shutdown configuration wizard.

Note

None.

K.2.5 Shutdown agent selection screen for the shutdown configuration wizard

Details on incompatibilities

The shutdown agent selection screen for the shutdown configuration wizard has been changed to support SPARC M10 and Oracle Solaris Kernel Zones.

Changes 1

Before upgrading [PRIMECLUSTER 4.3A10]

You can select the following SA on the selection screen of the shutdown agent:

- XSCF

After upgrading [PRIMECLUSTER 4.3A40]

You can select the following SAs on the selection screen of the shutdown agent:

- XSCF (SPARC M10)
- XSCF (SPARC Enterprise M-series)
- ILOM
- ALOM
- KZONE(SPARC M10)
- KZONE(SPARC T4)

Changes 2

Before upgrading [PRIMECLUSTER 4.3A10 (T006771SP-01 or later is applied)]

You can select the following SAs on the selection screen of the shutdown agent:

- XSCF
- ILOM
- ALOM

After upgrading [PRIMECLUSTER 4.3A40]

You can select the following SAs on the selection screen of the shutdown agent:

- XSCF(SPARC M10)
- XSCF(SPARC Enterprise M-series)
- ILOM
- ALOM
- KZONE(SPARC M10)
- KZONE(SPARC T4)

Note

None.

K.2.6 CF wizard

Details on incompatibilities

The only lowercase characters are allowed to set the beginning of the CF node names on the screen to edit CF node names of the CF wizard.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

If a CF node name begins with a letter other than lowercase characters, you could enter a numerical value or a symbol at the beginning of the CF node name, even though RMS Wizard Tools becomes unavailable to operate.

After upgrading [PRIMECLUSTER 4.3A40]

If a CF node name begins with a letter other than lowercase characters, the process to check the input value is added, because RMS Wizard Tools becomes unavailable to operate.

If you enter a letter other than lowercase characters at the beginning of the CF node name, the following message is displayed:

message 2978

The first letter of the CF node name "{0}" is a letter other than lowercase characters.

RMS Wizard Tools cannot operate with this setting.

Specify a name beginning with a lowercase character.

Note

If the error message of 2978 is displayed, specify a lowercase character at the beginning of the CF node name.

K.2.7 Posting Notification of a Resource Failure or Recovery

Details on incompatibilities

The default setting at installation is that notification of a resource failure or recovery is posted with PRIMECLUSTER 4.3A40. For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Message No.	Message overview
2700	Recovering from a resource failure
2701	Recovering from a node failure
6750	Resource failure
6751	Node failure

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

Notification of a resource failure or recovery will be not posted in the default setting of cluster installation.

The default value of AppWatch at cluster installation is OFF and notification of the resource failure or recovery will not be posted.

After upgrading [PRIMECLUSTER 4.3A40]

Notification of a resource failure or recovery will be posted in the default setting of cluster installation.

A resource failure or recovery will not be posted only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.2.8 Operator Intervention Request

Details on incompatibilities

With the default settings made when the cluster was installed, the operator intervention request is always enabled.

For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The operator intervention request will not work with the default setting at installation.

The default value of AppWatch set when the cluster was installed is set to OFF, and the operator intervention request will not work with this default value.

After upgrading [PRIMECLUSTER 4.3A40]

The operator intervention request will work with the default setting at installation.

The operator intervention request, is disabled only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.2.9 Display of the resource fault trace

Details on incompatibilities

When the resource is failed, the display of StateDetails of the failed resource object is changed. As a result, it can be able to distinguish the failed resource.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The StateDetails information is not displayed.

After upgrading [PRIMECLUSTER 4.3A40]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

K.2.10 Changes of the target node to forcibly shut down when a heartbeat failure occurs

Details on incompatibilities

The selecting method of the target node, which is forcibly shut down when a heartbeat failure occurs by temporary causes such as the overloaded, is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility determines the node to forcibly shut down according to the setup policy for survival priority.

After upgrading [PRIMECLUSTER 4.3A40]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility forcibly stops the node on which CF cannot perform regardless of the setup policy for survival priority.

Note

None.

K.2.11 Messages of the shutdown configuration wizard

Details on incompatibilities

By defining the CF node name used on the shutdown configuration wizard as "CF node," messages displayed on nodes are changed.

Changes

The following messages are changed:

- No.2586
- No.2587
- No.2588
- No.2591
- No.2941
- No.2942
- No.2943
- No.2944
- No.2950

- No.2967
- No.2968

Note

None.

K.2.11.1 Message 2586

Before upgrading [PRIMECLUSTER 4.3A10]

Invalid CF name for node. Lowercase a-z, 0-9, _ and - are allowed.

After upgrading [PRIMECLUSTER 4.3A40]

Invalid CF node name for node. Lowercase a-z, 0-9, _ and - are allowed.

K.2.11.2 Message 2587

Before upgrading [PRIMECLUSTER 4.3A10]

The CF name for node1 and node2 are the same.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node1 and node2 are the same.

K.2.11.3 Message 2588

Before upgrading [PRIMECLUSTER 4.3A10]

The CF name for node is empty.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node is empty.

K.2.11.4 Message 2591

Before upgrading [PRIMECLUSTER 4.3A10]

The CF name for node1 is the same as the public name of node2.

After upgrading [PRIMECLUSTER 4.3A40]

The CF node name for node1 is the same as the public name of node2.

K.2.11.5 Message 2941

Before upgrading [PRIMECLUSTER 4.3A10]

You must enter weight for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must enter weight for each of the CF nodes.

K.2.11.6 Message 2942

Before upgrading [PRIMECLUSTER 4.3A10]

Invalid host weight entered.

The entered weight of the host is invalid.

After upgrading [PRIMECLUSTER 4.3A40]

Invalid CF node weight entered.

The entered weight of the CF node is invalid.

K.2.11.7 Message 2943

Before upgrading [PRIMECLUSTER 4.3A10]

You must enter admin IP for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must enter admin IP for each of the CF nodes.

K.2.11.8 Message 2944

Before upgrading [PRIMECLUSTER 4.3A10]

Host weight must be between 1 and 1000000.

After upgrading [PRIMECLUSTER 4.3A40]

CF node weight must be between 1 and 1000000.

K.2.11.9 Message 2950

Before upgrading [PRIMECLUSTER 4.3A10]

You must specify XSCF-Name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify XSCF-Name and User-Name for each of the CF nodes.

K.2.11.10 Message 2967

Before upgrading [PRIMECLUSTER 4.3A10]

You must specify ILOM-name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify ILOM-name and User-Name for each of the CF nodes.

K.2.11.11 Message 2968

Before upgrading [PRIMECLUSTER 4.3A10]

You must specify ALOM-name and User-Name for each of the hosts.

After upgrading [PRIMECLUSTER 4.3A40]

You must specify ALOM-name and User-Name for each of the CF nodes.

K.2.12 Method to display the messages of the shutdown configuration wizard

Details on incompatibilities

On the screen to set information of each shutdown agent, the message is displayed in the pop-up window when you enter different values for the password and confirmation.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

On the screen to set information of each shutdown agent, the message is displayed in the bottom right of the SF wizard screen when you enter different values for the password and confirmation.

After upgrading [PRIMECLUSTER 4.3A40]

On the screen to set information of each shutdown agent, the message is displayed in the pop-up window when you enter different values for the password and confirmation.

Note

None.

K.3 Changes in PRIMECLUSTER 4.3A40 from 4.3A20

Incompatible commands

The following command of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.3A20.

- K.3.1 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.3A40 are incompatible with PRIMECLUSTER 4.3A20.

- K.3.2 Shutdown agent selection screen for the shutdown configuration wizard
- K.3.3 Posting Notification of a Resource Failure or Recovery
- K.3.4 Operator Intervention Request
- K.3.5 Changes of the port numbers for SNMP
- K.3.6 Display of the resource fault trace
- K.3.7 Changes of the target node to forcibly shut down when a heartbeat failure occurs

K.3.1 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.3A40]

The default work directory is /var/tmp.

Note

None.

K.3.2 Shutdown agent selection screen for the shutdown configuration wizard

Details on incompatibilities

The shutdown agent selection screen for the shutdown configuration wizard has been changed to support Oracle Solaris Kernel Zones.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

You can select the following SAs on the selection screen of the shutdown agent:

- XSCF (SPARC M10)
- XSCF (SPARC Enterprise M series)

- ILOM
- ALOM

After upgrading [PRIMECLUSTER 4.3A40]

You can select the following SA on the selection screen of the shutdown agent:

- XSCF (SPARC M10)
- XSCF (SPARC Enterprise M-series)
- ILOM
- ALOM
- KZONE(SPARC M10)
- KZONE(SPARC T4)

Note

None.

K.3.3 Posting Notification of a Resource Failure or Recovery

Details on incompatibilities

The default setting at installation is that notification of a resource failure or recovery is posted with PRIMECLUSTER 4.3A40. For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Message No.	Message overview
2700	Recovering from a resource failure
2701	Recovering from a node failure
6750	Resource failure
6751	Node failure

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

Notification of a resource failure or recovery will be not posted in the default setting of cluster installation.

The default value of AppWatch at cluster installation is OFF and notification of the resource failure or recovery will not be posted.

After upgrading [PRIMECLUSTER 4.3A40]

Notification of a resource failure or recovery will be posted in the default setting of cluster installation.

A resource failure or recovery will not be posted only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.3.4 Operator Intervention Request

Details on incompatibilities

With the default settings made when the cluster was installed, the operator intervention request is always enabled.

For details, see "5.4 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

The operator intervention request will not work with the default setting at installation.

The default value of AppWatch set when the cluster was installed is set to OFF, and the operator intervention request will not work with this default value.

After upgrading [PRIMECLUSTER 4.3A40]

The operator intervention request will work with the default setting at installation.

The operator intervention request, is disabled only when the AppWatch parameter is set to OFF with clsetparam.

Note

After you have changed the AppWatch parameter with clsetparam, you have to restart all the nodes to validate the setting.

K.3.5 Changes of the port numbers for SNMP

Details on incompatibilities

The port number of the SNMP trap receiving daemon (snmptrapd) has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

The standard port number (162) is used.

After upgrading [PRIMECLUSTER 4.3A40]

The port number of the shutdown facility (9385) is used.

Note

None.

K.3.6 Display of the resource fault trace

Details on incompatibilities

When the resource is failed, the display of StateDetails of the failed resource object is changed. As a result, it can be able to distinguish the failed resource.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

The StateDetails information is not displayed.

After upgrading [PRIMECLUSTER 4.3A40]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

K.3.7 Changes of the target node to forcibly shut down when a heartbeat failure occurs

Details on incompatibilities

The selecting method of the target node, which is forcibly shut down when a heartbeat failure occurs by temporary causes such as the overloaded, is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility determines the node to forcibly shut down according to the setup policy for survival priority.

After upgrading [PRIMECLUSTER 4.3A40]

If CF becomes temporarily disabled by the overloaded or other causes, and then a heartbeat failure occurs, the shutdown facility forcibly stops the node on which CF cannot perform regardless of the setup policy for survival priority.

Note

None.

Appendix L Release Information

This appendix describes the main contents for change of this manual.

No.	Version / Level	Change Location	Details
1	Second edition	2.1.2 Function Selection	Added Patrol diagnosis to the Other feature.
2	Second edition	2.2 System Design 2.2.1 Virtual Machine Function 2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment 2.2.1.3 Cluster System Operating in Oracle Solaris Zones Environment Chapter 3 Software Installation	Added the description regarding the virtual machine function.
3	Second edition	2.3 Determining the Cluster System Operation Mode	Added the supported configurations and notes when using the clusters between control domains in an Oracle VM Server for SPARC Environment.
4	Second edition	Chapter 3 Software Installation	Added notes on the installation of the operation system when mirroring the system disk using GDS in a ZFS boot environment.
5	Second edition	Chapter 3 Software Installation	Added notes on umask.
6	Second edition	3.2.1 Setting Up the Network	Deleted the description regarding IP address takeover.
7	Second edition	4.3 Starting the Web-Based Admin View Screen 5.1.2.2.2 Using the Shutdown Configuration Wizard 5.1.2.3.2 Using the Shutdown Configuration Wizard 5.1.2.4.2 Using the Shutdown Configuration Wizard 6.7.1.5 Creating Takeover Network Resources 7.1.2.2 Detailed Resource Information 9.2.1 Changing an IP Address on the Public LAN 9.2.2 Changing a CIP Address	Added the description of IPv6.
8	Second edition	4.4.1 Operation Menu Functions	Added notes on SIS.
9	Second edition	5.1.1 Setting Up CF and CIP	Added the description to IP interconnects.
10	Second edition	5.1.2 Configuring the Shutdown Facility 5.1.2.1 For SPARC M10 5.1.2.2.2 Using the Shutdown Configuration Wizard	Added SPARC M10 to the supported server models.
11	Second edition	5.1.2 Configuring the Shutdown Facility	Added XSCF SNMP to the shutdown agents required to be set up.

No.	Version / Level	Change Location	Details
12	Second edition	5.1.2 Configuring the Shutdown Facility	Added a table that shows the shutdown agents necessary for virtualized environments.
13	Second edition	5.1.2.2 For SPARC Enterprise M3000, M4000, M5000, M8000, or M9000	Deleted "Specifying the Timeout Value."
14	Second edition	5.1.2.2.2 Using the Shutdown Configuration Wizard	Changed the description of setting up the shutdown agent.
15	Second edition	5.1.2.2.3 Setting of the connection method to the XSCF	Added the confirmation procedure after changing the connection method.
16	Second edition	5.1.2.3 For SPARC Enterprise T5120, T5220, T5140, T5240, T5440, or SPARC T3, T4, T5, T7, S7 series 5.1.2.4 For SPARC Enterprise T1000, T2000 9.2.1 Changing an IP Address on the Public LAN 9.4.1 Changing the User Name and Password to Control the Console	Changed the operation method of setting and changing the shutdown facility from CLI to the shutdown configuration wizard.
17	Second edition	5.1.3.2 Automatic Configure 6.9 Setting Up Patrol Diagnosis	Added the description to Note when specifying shared disk units by the patrol diagnosis.
18	Second edition	6.2.1 GLS Setup	Added the procedure to set up the subnet masks.
19	Second edition	6.6 Setting Up Online/Offline Scripts	Modified the description of HV_AUTORECOVER.
20	Second edition	6.7.1.2 Creating Fsystem Resources	Changed the description of "Notes on using the file lock facility."
21	Second edition	6.7.1.2.1 Prerequisites	Changed the following: - The description regarding the location of the directories where NFS file lock information is stored in "Carry out preparations for enabling NFS Lock Failover" - The description and notes regarding the file systems to be shared in NFS in "Set up the NFS entry of PRIMECLUSTER"
22	Second edition	6.7.1.5 Creating Takeover Network Resources	Added notes on network interfaces for the case of non-global zones.
23	Second edition	6.7.1.5.1 Setup Method	Added the description for the case of Solaris 11 in Note under "Network type selection."
24	Second edition	7.1.3.1 RMS Tree 7.3.1 Monitoring the State of a Node 7.3.2 Monitoring the State of a Cluster Application	Added the explanation of icons that shows the node status in the GUI windows.
25	Second edition	8.1.1.1 Adding a Shared Disk Device	Added Information to the procedure to add a shared disk device.

No.	Version / Level	Change Location	Details
26	Second edition	12.3.2.2 Procedure for Applying Correction by Rolling Update	Changed the description of switching a cluster application.
27	Second edition	Part 6 Virtualized Environments Chapter 14 Using PRIMECLUSTER in Oracle VM Server for SPARC Environment Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment	- Added a new Part. - Added new sections by using the following appendices: - Appendix G Using PRIMECLUSTER in Oracle VM Server for SPARC Environments - Appendix H Using PRIMECLUSTER in Oracle Solaris Zones Environments
28	Second edition	Chapter 25 Symfoware Server	Changed the description.
29	Second edition	A.5.2 RMS Configuration	Added the note for message queues.
30	Second edition	A.6 Cluster Configuration Worksheet	Changed worksheets.
31	Second edition	Appendix G Startup Scripts and Startup Daemons in PRIMECLUSTER	Added Appendix of "Startup Scripts and Startup Daemons."
32	Second edition	Appendix H SMF Services and Port Numbers in PRIMECLUSTER	Added Appendix of "SMF Services and Port Numbers."
33	Second edition	Appendix K Changes in Each Version	Added changes made to the specifications from 4.3A10 to the table of "List of changes."
34	Second edition	Appendix K Changes in Each Version K.1 Changes in PRIMECLUSTER 4.3A40 from 4.2A00 K.1.9 IP Interconnect K.1.14 Shutdown Configuration Wizard K.1.15 Shutdown agent selection screen for the shutdown configuration wizard K.1.16 CF wizard	Added incompatible functions.
35	Second edition	Appendix K Changes in Each Version K.1 Changes in PRIMECLUSTER 4.3A40 from 4.2A00 K.1.24 Messages of the shutdown configuration wizard K.1.25 Method to display the messages of the shutdown configuration wizard	Added incompatible messages.
36	Second edition	K.2 Changes in PRIMECLUSTER 4.3A40 from 4.3A10	Added sections.
37	Second edition	Glossary	Deleted "Console Break agent."
38	Third edition	2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment	- Added the instruction for how to specify a timeout option of a virtual disk. - Deleted the description "Migration function is not available" from Note. - Added the description of the Migration in an Oracle VM Server for SPARC Environment.

No.	Version / Level	Change Location	Details
39	Third edition	4.2.3.1 Initial setup of the operation management server 4.4.1 Operation Menu Functions 4.4.2 Global Cluster Services Menu Functions 6.7.1.5.1 Setup Method 10.6.2 Changing the Attributes Used by a Resource or a Resource Interface	Added the description of IPv6.
40	Third edition	5.1.2 Configuring the Shutdown Facility 14.1.1.1 Cluster System Between Guest Domains Within a Same Physical Partitions 14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions	Added the description of using the Migration function of Oracle VM Server for SPARC to Note.
41	Third edition	7.2.2.5 Clearing the Wait State of a Node	Change the description of clearing the Wait state of a node.
42	Third edition	Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	Added a new Chapter.
43	Third edition	C.1.3 System Dump	Change the description of collecting a system dump.
44	Fourth edition	2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment Operation before Performing Live Migration	Added the note when performing the Live Migration.
45	Fourth edition	2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment Chapter 18 When Using Oracle VM Server for SPARC P2V Tool to Migrating a Cluster System	Added the description when migrating a cluster system in the physical environment to a guest domain in Oracle VM Server for SPARC Environment.
46	Fourth edition	2.2.1.3 Cluster System Operating in Oracle Solaris Zones Environment Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment	Added the description when using Oracle Solaris Legacy Containers (OSLC).
47	Fourth edition	Part 2 Installation	Added Note when Oracle Solaris Zones of an exclusive IP zone exist on the system.
48	Fourth edition	5.1.2 Configuring the Shutdown Facility	Added Note when making the administrative LAN, used in the shutdown facility, redundant by GLS.
49	Fourth edition	5.1.2 Configuring the Shutdown Facility	Added SPARC T4 series.
50	Fourth edition	Chapter 7 Operations	Added notes on operation.
51	Fourth edition	7.5 Notes on Operation	Added the note on the operation.
52	Fourth edition	9.2.1 Changing an IP Address on the Public LAN	Added Note when the connection to the XSCF is SSH connection.
53	Fourth edition	12.3.2.2 Procedure for Applying Correction by Rolling Update	Added the note on the operation.

No.	Version / Level	Change Location	Details
		14.3.1.2 Maintenance of Guest Domains	
54	Fourth edition	14.5.1 Cluster Configuration Worksheet 16.7.1 Cluster Configuration Worksheet A.6 Cluster Configuration Worksheet A.6 Cluster Configuration Worksheet	Described how to reboot a node.
55	Fourth edition	Chapter 16 Using PRIMECLUSTER in Oracle Solaris Zones Environment	Added the description about installing a new PRIMECLUSTER in OSLC environments.
56	Fourth edition	16.2.4 Creating Non-Global Zones	Changed the description about installing middleware products to non-global zones.
57	Fourth edition	Setting Prerequisites on a Guest Domain	Added the procedure that should be done before performing Migration.
58	Fourth edition	Appendix H SMF Services and Port Numbers in PRIMECLUSTER	Added the utilized port of /milestone/ fjsvcldev.
59	Fourth edition	K.2 Changes in PRIMECLUSTER 4.3A40 from 4.3A10	Added the following incompatible command: J.2.2 clrccusetup command
60	Fifth edition	1.7 Notes When Building a System 2.2 System Design 5.1.2 Configuring the Shutdown Facility 9.2.1 Changing an IP Address on the Public LAN Chapter 15 Using PRIMECLUSTER in Oracle Solaris Kernel Zones Environment Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment C.1 Collecting Troubleshooting Information	Added the description of the Oracle Solaris Kernel Zones.
61	Fifth edition	1.7 Notes When Building a System 6.11 Setting Contents and Notes on Cluster Application 6.12 Notes When Setting Cmdline Resources 6.13 Notes When Setting Fsystem Resource 7.6 CF and RMS Heartbeats 7.7 cron Processing 8.1.3.3 Changing a network interface card used for CIP 9.2.3 Changing the Subnet Mask of CIP 11.2.1 Changing Timeout Period during RMS Stop Processing 11.3.1 Changing Time to Detect CF Heartbeat Timeout 11.3.2 Changing Time to Detect RMS Heartbeat Timeout	Newly added the descriptions.

No.	Version / Level	Change Location	Details
		C.4 PRIMECLUSTER Log Files	
62	Fifth edition	Chapter 2 Site Preparation 5.1.2.1.1 Checking XSCF Information 5.1.2.1.3 Using the Shutdown Configuration Wizard	Added the description of XSCF configuration when using SPARC M10.
63	Fifth edition	2.2.1 Virtual Machine Function	Added the note on the virtual machine function.
64	Fifth edition	2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment 5.1.1 Setting Up CF and CIP	Added the note on the cluster interconnect.
65	Fifth edition	2.2.1.1 Cluster Systems in Oracle VM Server for SPARC Environment 14.1.1.1 Cluster System Between Guest Domains Within a Same Physical Partitions 14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions 14.1.1.3 Cluster System Between Control Domains	Added the description and note when building the cluster application on the control domain.
66	Fifth edition	2.2.1.1.1 Cluster System Configuration in Oracle VM Server for SPARC Environment 14.1.1.2 Cluster System Between Guest Domains Among Different Physical Partitions	Changed the description of the cluster system between guest domains among different physical partitions.
67	Fifth edition	2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment	Described two types of migration functions (Live Migration and Cold Migration).
68	Fifth edition	2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	Changed the note when using the Migration function.
69	Fifth edition	2.2.1.1.2 Migration for a Cluster System in Oracle VM Server for SPARC Environment Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	Added the description when using the Migration function in combination with ServerView Resource Orchestrator Cloud Edition.
70	Fifth edition	2.2.1.1.3 When Migrating a Cluster System in the Physical Environment to a Guest Domain in Oracle VM Server for SPARC Environment (Physical to Virtual) 5.2 Setting Up Power Supply Linkage 18.3 Migration with Oracle VM Server for SPARC P2V Tool	Changed the reference manual.
71	Fifth edition	2.5 Setting the Failover Timing of a Cluster Application	Changed the setting procedure for the failover timing of a cluster application.
72	Fifth edition	Chapter 3 Software Installation	Added the note on a message output when the shutdown facility is set.

No.	Version / Level	Change Location	Details
73	Fifth edition	4.2.3.1 Initial setup of the operation management server	Added the note.
74	Fifth edition	5.1.1 Setting Up CF and CIP	Added the note for a cluster interconnect in Oracle VM Server for SPARC environment.
75	Fifth edition	5.1.1 Setting Up CF and CIP	Changed the description of the CF node name.
76	Fifth edition	5.1.2.1.1 Checking XSCF Information	Added the required XSCF information to be checked before setting the shutdown facility.
77	Fifth edition	5.1.2.1.2 Setting SNMP 9.2.4 Changing Port Numbers for SNMP	Added the description when changing the port number for SNMP. Added the description of how to confirm the received SNMP.
78	Fifth edition	5.1.2.1.3 Using the Shutdown Configuration Wizard	Added the note on displaying the configuration status of the shutdown facility.
79	Fifth edition	5.1.2.1.3 Using the Shutdown Configuration Wizard	Added the description of "Specific node survival" to "Survival scenarios."
80	Fifth edition	5.1.2.1.3 Using the Shutdown Configuration Wizard	Changed the description of PPAR-ID in "Configuring XSCF."
81	Fifth edition	5.1.2.1.3 Using the Shutdown Configuration Wizard 5.1.2.1.4 Setting of the connection method to the XSCF	Added the note on the setting of XSCF when using the Migration function.
82	Fifth edition	5.1.2.1.3 Using the Shutdown Configuration Wizard 5.1 Initial Cluster Setup Appendix J Using Remote Power Distribution Unit	Added the description on using the Remote Power Distribution Unit.
83	Fifth edition	5.4 Setting Up Fault Resource Identification and Operator Intervention Request	Changed the default setting of the fault resource identification and the operator intervention request.
84	Fifth edition	6.3.1 Automatic Configuration of Shared Disks	Changed the note on the automatic configuration of shared disks.
85	Fifth edition	6.3.2.3 Setting Up Shared Disks	Changed the setting procedure of the file system.
86	Fifth edition	6.4.1.1 Design for using ZFS with a Cluster System	Changed the description of supported configurations and notes.
87	Fifth edition	6.4.1.3 Notes on the Operation	Added the note on the non-legacy zfs file system.
88	Fifth edition	6.6 Setting Up Online/Offline Scripts	Added the description of environment variables that can be referred by the Check script.
89	Fifth edition	6.6 Setting Up Online/Offline Scripts	Added the reference manual when setting the Online/Offline script.
90	Fifth edition	6.7.1.1 Creating Cmdline Resources	Changed the note on characters that cannot be used in the file path and the arguments.
91	Fifth edition	6.7.1.2 Creating Fsystem Resources	Added the note when using the mountpoint that shares NFS.

No.	Version / Level	Change Location	Details
92	Fifth edition	6.7.1.2.1 Prerequisites	Changed the example when the mountpoint is not set to the ZFS storage pool.
93	Fifth edition	6.7.1.2.1 Prerequisites	Changed the preparation steps to share a file system in a network (NFS).
94	Fifth edition	6.7.1.2.2 Setup Method	Changed the note when selecting mount points.
95	Fifth edition	6.7.1.4 Creating Gls Resources	Modified the description of checking registration information for Gls resources.
96	Fifth edition	6.7.1.5 Creating Takeover Network Resources	Deleted the unnecessary description about taking over the node name.
97	Fifth edition	6.7.2.1 Creating Standby Cluster Applications	Added the note on a resource name of the patrol diagnosis facility.
98	Fifth edition	6.7.3.1 Setting Exclusive Relationships Between Cluster Applications	Deleted the description of the hvreset command.
99	Fifth edition	6.7.3.1 Setting Exclusive Relationships Between Cluster Applications	Added the note on the cluster applications with an exclusive relationship.
100	Fifth edition	7.1.3.1 RMS Tree	Added the note on displaying the other object state.
101	Fifth edition	7.2.2.6 Entering maintenance mode for Cluster Application	Added the description when changing the resource state of a cluster application while in maintenance mode.
102	Fifth edition	7.2.3 Resource Operation	Changed the operation procedure for a resource and added "Information."
103	Fifth edition	7.5.1 Notes on Switching a Cluster Application Forcibly	Added the description about a resource to the note on the forced switch request.
104	Fifth edition	7.7 cron Processing	Added/Changed the table and note of the cron processing.
105	Fifth edition	Chapter 8 Changing the Cluster System Configuration Chapter 9 Changing the Cluster System Environment Chapter 10 Changing the Cluster Application Configuration Chapter 11 Changing the Operation Attributes of a Cluster System	Changed the chapter organization.
106	Fifth edition	Chapter 8 Changing the Cluster System Configuration Chapter 9 Changing the Cluster System Environment	Added the description to check the PRIMECLUSTER environment.
107	Fifth edition	8.1 Adding, Deleting, and Changing Hardware	Added "Replacing XSCF."
108	Fifth edition	Chapter 9 Changing the Cluster System Environment	Added "Changing a CF Node Name."
109	Fifth edition	9.2.1 Changing an IP Address on the Public LAN Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	Provided the consistent description for performing the Live Migration on the control domain.

No.	Version / Level	Change Location	Details
110	Fifth edition	10.1 Adding a Cluster Application	Added "Adding a Cluster Application."
111	Fifth edition	10.4 Adding a Resource	Added "Adding a Resource."
112	Fifth edition	11.2 Changing the RMS Environment Variables	Changed the description of changing the RMS environment variables.
113	Fifth edition	12.3.1 Notes on Applying Corrections to the PRIMECLUSTER System	Added the note when applying SRU.
114	Fifth edition	13.2 Restoring the PRIMECLUSTER Operation Environment	Changed the procedure to restore the PRIMECLUSTER Operation Environment
115	Fifth edition	13.2 Restoring the PRIMECLUSTER Operation Environment 16.2.3.1 Creating an Archive on the Migration Source Node 16.2.4.3 OS Installation to the Non-Global Zone 16.2.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone) 16.2.4.13 Sharing Non-Global Zone Configuration Information 16.3.3.3 Creating a Flash Archive from the Migration Source Node 16.3.4.2 Creating Containers from a Flash Archive 16.3.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone) 16.3.4.13 Sharing Non-Global Zone Configuration Information SCF Dump E.1 Registering a Procedure Resource E.2.1 Changing a State Transition Procedure	Changed the work are from /tmp to /var/tmp.
116	Fifth edition	14.4 Collection Troubleshooting Information in Oracle VM Server for SPARC Environment 16.5.3 Method for Collecting Troubleshooting Information for the Non-Global Zone C.1 Collecting Troubleshooting Information C.1.2 Collecting Information by FJQSS(Information Collection Tool)	Added FJQSS as a method for collecting the investigation material.
117	Fifth edition	14.1.1.1 Cluster System Between Guest Domains Within a Same Physical Partitions	Changed the descriptions on setting the control domain and building the guest domain.
118	Fifth edition	16.1 Design	Described the migratable PRIMECLUSTER versions in the migration from an

No.	Version / Level	Change Location	Details
			environment with PRIMECLUSTER operation to OSLC environment.
119	Fifth edition	16.1 Design	Added the reference when using GFS.
120	Fifth edition	16.2.4.3 OS Installation to the Non-Global Zone	Changed the procedure when newly installing OS.
121	Fifth edition	16.2.4.6 Global Zone Environment Setup (After Installation of PRIMECLUSTER to the Non-Global Zone)	Modified the execution example for the environment setting of the global zone in the cluster building procedure in Oracle Solaris Zone environment.
122	Fifth edition	16.2.5.4 Creating Cluster Applications 16.3.5.4 Creating Cluster Applications	Added the description of creating the cluster application in the cluster building procedure in Oracle Solaris Zone environment.
123	Fifth edition	16.3.4 Creating Non-Global Zones	Added the note on creating the container from a flash archive when creating the non-global zone in building procedure of the cluster to use OSLC.
124	Fifth edition	16.3.4.2 Creating Containers from a Flash Archive 16.3.4.3 Non-Global Zone Startup and OS Setup 16.5.1 Maintenance Operations on the Non-Global Zone 16.5.2 Maintenance Operations on the Global Zone 16.5.4 Recovery Operation for When an Error Occurs on the Non-Global Zone	Modified the execution example of the procedure to restart the non-global zone.
125	Fifth edition	Chapter 17 When Using the Migration Function in Oracle VM Server for SPARC Environment	Changed the procedure and description when using the Migration function.
126	Fifth edition	18.3 Migration with Oracle VM Server for SPARC P2V Tool 18.6 Changing the Setting in the Multi-User Mode and Restoration of the Configuration Information of GFS/GDS	Modified the description and added the note for the Migration with P2V Tool.
127	Fifth edition	A.5.1 CF Configuration	Changed the description of the required value for the resource database.
128	Fifth edition	A.9 Cluster Application Worksheets	Changed the number of supported cluster application.
129	Fifth edition	A.9.11 Configuration Global Settings Worksheet	Added hvdet_ckshare and hvdet_zfs.
130	Fifth edition	A.9.11 Configuration Global Settings Worksheet	Changed the following items: - hvdet_ckhost - hvdet_execbin - hvdet_glbassrt - hvdet_locassrt - hvdet_lvm - hvdet_rcfs

No.	Version / Level	Change Location	Details
			- hvdet_rcvm - hvdet_read - hvdet_srdf - hvdet_stopclnt - hvdet_vxvm
131	Fifth edition	B.7 RMS	Changed the description of "hvswitch."
132	Fifth edition	B.7 RMS	Added "hvdispall."
133	Fifth edition	Appendix C Troubleshooting	Added the XSCF log to "PRIMECLUSTER investigation information."
134	Fifth edition	C.1 Collecting Troubleshooting Information	Deleted the SCF dump.
135	Fifth edition	C.4.1 Output Destination for core Files	Changed the core file.
136	Fifth edition	G.2 Startup Script Lists	Added S99SMAWsfclean.
137	Fifth edition	Appendix H SMF Services and Port Numbers in PRIMECLUSTER	Added a service name corresponding to each port number to "Remarks."
138	Fifth edition	H.2 SMF Service Lists	Added the descriptions and remarks for SMF services, startup daemons, and utilized ports.
139	Fifth edition	H.2 SMF Service Lists	Added /milestone/fjsvgfs-zones and /milestone/fjsvcldevtrap services.
140	Fifth edition	Appendix I Cloning the Cluster System Environment	Added the appendix.
141	Fifth edition	Appendix K Changes in Each Version	Changed the incompatible information.

Glossary

AC

See *Access Client*.

Access Client

GFS kernel module on each node that communicates with the Meta Data Server and provides simultaneous access to a shared file system.

See also *Meta Data Server*.

Administrative LAN

In PRIMECLUSTER configurations, an Administrative LAN is a private local area network (LAN) on which machines such as the System Console and Cluster Console reside. Because normal users do not have access to the Administrative LAN, it provides an extra level of security. The use of an Administrative LAN is optional.

See also *public LAN*.

API

See *Application Program Interface*.

application (RMS)

A resource categorized as userApplication used to group resources into a logical collection.

Application Program Interface

A shared boundary between a service provider and the application that uses that service.

application template (RMS)

A predefined group of object definition value choices used by RMS Wizard kit to create object definitions for a specific type of application.

attribute (RMS)

The part of an object definition that specifies how the base monitor acts and reacts for a particular object type during normal operations.

automatic configuration (GDS)

Function that automatically recognizes the physical connection configuration of shared disk units and registers the units to the resource database.

automatic power control

This function is provided by the Enhanced Support Facility (ESF), and it automatically switches the power of the server on and off.

automatic switchover (RMS)

The procedure by which RMS automatically switches control of userApplication over to another host after specified conditions are detected.

See also *directed switchover*, *failover*, *switchover*, and *symmetrical switchover*.

availability

Availability describes the need of most enterprises to operate applications via the Internet 24 hours a day, 7 days a week. The relationship of the actual to the planned usage time determines the availability of a system.

base cluster foundation (CF)

This PRIMECLUSTER module resides on top of the basic OS and provides internal interfaces for the CF (Cluster Foundation) functions that the PRIMECLUSTER services use in the layer above.

See also *Cluster Foundation*.

BM (base monitor) (RMS)

The RMS module that maintains the availability of resources. The base monitor is supported by daemons and detectors. Each host being monitored has its own copy of the base monitor

Cache Fusion

The improved interprocess communication interface in Oracle 9i that allows logical disk blocks (buffers) to be cached in the local memory of each node. Thus, instead of having to flush a block to disk when an update is required, the block can be copied to another node by passing a message on the interconnect, thereby removing the physical I/O overhead.

CCBR

See *Cluster Configuration Backup and Restore.*

CB

Clustering Base

ccbr.conf

The environment configuration file that is used for backup and restore operations, and is placed in the "/opt/SMAW/ccbr" directory. This file is used in the "$CCBRHOME" variable setting. For details, see the manual pages for the "cfbackup(1M)" and "cfrestore(1M)" commands and the comments in the "ccbr.conf" file.

ccbr.gen

The file that stores the generation number and is placed in the "/opt/SMAW/ccbr" directory. A value of 0 or higher is stored in this file. For details, see the manual pages for the "cfbackup(1M)" and "cfrestore(1M)" commands.

CCBRHOME variable

The variable that identifies the directory in which backup data is stored. The initial value is the "/var/spool/pcl4.1/ccbr" directory. This variable can be set only in the "ccbr.conf" file.

CF (Cluster Foundation or Cluster Framework)

See Cluster Foundation.

child (RMS)

A resource defined in the configuration file that has at least one parent. A child can have multiple parents, and can either have children itself (making it also a parent) or no children (making it a leaf object).

See also *resource*, *object*, *parent*, and *leaf object*.

CIM

Cluster Integrity Monitor

CIP

Cluster Interconnect Protocol

CLI

command-line interface

CLM

Cluster Manager

class (GDS)

See *disk class*.

cluster

A set of computers that work together as a single computing source. Specifically, a cluster performs a distributed form of parallel computing.

See also *RMS configuration*.

Cluster Configuration Backup and Restore (CF)

CCBR provides a simple method to save the current PRIMECLUSTER configuration information of a cluster node. It also provides a method to restore the configuration information.

CB

Cluster BaseCluster Foundation

The set of PRIMECLUSTER modules that provides basic clustering communication services.

See also *base cluster foundation*.

cluster interconnect (CF)

The set of private network connections used exclusively for PRIMECLUSTER communications.

Cluster Join Services (CF)

This PRIMECLUSTER module handles the forming of a new cluster and the addition of nodes.

Cluster Resource Management facility

Facility that manages hardware units that are shared among multiple nodes.

cold-standby

The operation which does not allow the preliminary operation needed to establish the operating state immediately on the standby node.

concatenated virtual disk

Concatenated virtual disks consist of two or more pieces on one or more disk drives. They correspond to the sum of their parts. Unlike simple virtual disks where the disk is subdivided into small pieces, the individual disks or partitions are combined to form a single large logical disk. (Applies to transitioning users of existing Fujitsu Technology Solutions products only.)

See also *mirror virtual disk, simple virtual disk, striped virtual disk, virtual disk*.

concatenation (GDS)

The linking of multiple physical disks. This setup allows multiple disks to be used as one virtual disk that has a large capacity.

configuration file (RMS)

The RMS configuration file that defines the monitored resources and establishes the interdependencies between them. The default name of this file is config.us.

Console Break agent

The Console Break agent is used for the Shutdown Facility to eliminate a node by sending a break signal from RCCU.

control domain

Domain in which the Oracle VM Server for SPARC is installed. All platforms that are using Oracle VM Server for SPARC must contain a control domain. By using the ldm command within this domain, other domains can be created and controlled.

control domain cluster

Cluster that is configured between different control domains

CRM

Cluster Resource Management

custom detector (RMS)

See *detector*.

custom type (RMS)

See *generic type*.

daemon

A continuous process that performs a specific function repeatedly.

detector (RMS)

A process that monitors the state of a specific object type and reports a change in the resource state to the base monitor.

directed switchover (RMS)

The RMS procedure by which an administrator switches control of userApplication over to another host.

See also *automatic switchover*, *failover*, *switchover*, and *symmetrical switchover*.

disk class (GDS)

Collection of SDX objects. The shared type disk class is also a resource unit that can be used by the PRIMECLUSTER system. A disk class is sometimes simply called a "class."

disk group (GDS)

A collection of disks or low-order groups that become the unit for mirroring, striping, or concatenation. Disk and low-order groups that belong to the same disk group are mutually mirrored, striped, or concatenated according to the type attribute (mirror, stripe, or concatenation) of that disk group.

A disk group is sometimes simply called a "group."

DLPI

Data Link Provider Interface

Domain

A set of one or more system boards that function as an independent system. While the server is shared, an operating system can be installed in each domain to enable each domain to operate as an independent system.

Each domain consists of a logical system board assigned to it. Each domain is electrically insulated by each hardware partition. Therefore, if one domain fails, it does not affect the other domains in the server.

DOWN (CF)

A node state that indicates that the node is unavailable (marked as down). A LEFTCLUSTER node must be marked as DOWN before it can rejoin a cluster.

See also *UP*, *LEFTCLUSTER*, *node state*.

EE

Enterprise Edition

ENS (CF)

See *Event Notification Services*.

environment variables (RMS)

Variables or parameters that are defined globally.

error detection (RMS)

The process of detecting an error. For RMS, this includes initiating a log entry, sending a message to a log file, or making an appropriate recovery response.

Ethernet

LAN standard that is standardized by IEEE 802.3. Currently, except for special uses, nearly all LANs are Ethernets. Originally the expression Ethernet was a LAN standard name for a 10 megabyte per second type LAN, but now it is also used as a general term that includes high-speed Ethernets and gigabyte Ethernets.

Event Notification Services (CF)

This PRIMECLUSTER module provides an atomic-broadcast facility for events.

Fast switching mode

One of the LAN duplexing modes presented by GLS.

This mode uses a multiplexed LAN simultaneously to provide enhanced communication scalability between Solaris servers and high-speed switchover if a LAN failure occurs.

fault tolerant network

A network with the ability to withstand faults (fault tolerant). Fault tolerant is the ability to maintain and continue normal operation even if a fault occurs in part of the computer system. A fault tolerant network is therefore a network that can continue normal communication even if a flat occurs in part of the network system.

GDS

See *Global Disk Services*.

generation number

Data generation management is enabled in the PRIMECLUSTER backup and restore operations. The current generation number is added as part of the backup and restore data name. Integers of 0 or higher are used as generation numbers, and the generation number is incremented each time backup is successful. The generation number is stored in the "ccbr.gen" file and can be specified as an optional argument in the "cfbackup(1M)" and "cfrestore(1M) " commands.
For details, see the manual pages for the "cfbackup(1M)" and "cfrestore(1M)" commands.

generic type (RMS)

An object type which has generic properties. A generic type is used to customize RMS for monitoring resources that cannot be assigned to one of the supplied object types.

See also *object type*.

GFS

See *Global File Services*.

GFS shared file system

A shared file system that allows simultaneous access from multiple Solaris systems that are connected to shared disk units, while maintaining data consistency, and allows processing performed by a node to be continued by other nodes even if the first node fails.

A GFS shared file system can be mounted and used concurrently from multiple nodes.

Global Disk Services

This optional product provides volume management that improves the availability and manageability of information stored on the disk unit of the Storage Area Network (SAN).

Global File Services

This optional product provides direct, simultaneous accessing of the file system on the shared storage unit from two or more nodes within a cluster.

Global Link Services

This PRIMECLUSTER optional module provides network high availability solutions by multiplying a network route.

GLS

See *Global Link Services*.

graph (RMS)

See *system graph*.

graphical user interface

A computer interface with windows, icons, toolbars, and pull-down menus that is designed to be simpler to use than the command-line interface.

group (GDS)

See *disk group*.

GS/SURE linkage mode

One of the LAN duplexing modes presented by GLS.

This mode uses a duplexed LAN simultaneously and high reliance communication with Global server or SURE system is realized.

GUI

See *graphical user interface*.

guest domain

Virtualized hardware environment in which an independent operating system is running. It can be started and stopped without any influence on other domains.

HA

high availability

high availability

This concept applies to the use of redundant resources to avoid single points of failure.

highest-order group (GDS)

Group that does not belong to another group. A volume can be created in the highest-order group.

hot-standby

The operation which enables preliminary operation so that the operating state can be established immediately on the standby node.

hub

Star-type wiring device used for LAN or fibre channels.

ICF

Internode Communication Facility

I/F

Interface

installation server

System having a Solaris CD image on the disk or CD-ROM drive to distribute the Solaris CD image to other systems over the network.

interconnect (CF)

See *cluster interconnect*.

Internet Protocol address

A numeric address that can be assigned to computers or applications.

See also *IP aliasing*.

internode communication facility

Communication function between cluster nodes that are used by PRIMECLUSTER CF. Since this facility is designed especially for communication between cluster nodes, the overhead is less than that of TCP/IP, and datagram communication services that also guarantee the message arrival sequence can be carried out.

I/O

input/output

I/O domain

A domain in an Oracle VM Server for SPARC Environment that is allocated only the PCle end point device, which is managed by the control domain through the Direct I/O function.

I/O root domain

The logical domain which holds a physical I/O device in Oracle VM Server for SPARC Environments. This holds one or more root complex. (I/O root domains exceeding the number of root complex within a partition cannot be created.)

IP address

See *Internet Protocol address*.

IP aliasing

This enables several IP addresses (aliases) to be allocated to one physical network interface. With IP aliasing, the user can continue communicating with the same IP address, even though the application is now running on another host.

See also *Internet Protocol address*.

JOIN (cluster join services module) (CF)

See *Cluster Join Services*.

keyword (reserved words)

A word that has special meaning in a programming language. For example, in the configuration file, the keyword node identifies the kind of definition that follows.

LAN

local area network

latency (RMS)

Time interval from when a data transmission request is issued until the actual response is received.

leaf object (RMS)

A bottom object in a system graph. In the configuration file, this object definition is at the beginning of the file. A leaf object does not have children.

LEFTCLUSTER (CF)

A node state that indicates that the node cannot communicate with other nodes in the cluster. That is, the node has left the cluster. The purpose for the intermediate LEFTCLUSTER state is to avoid the network partition problem.

See also *UP*, *DOWN*, *network partition*, *node state*.

line switching unit (only in Oracle Solaris 10 environment)

This device connects external lines to more than one node and switches the connected nodes by the RCI.

link (RMS)

Designates a child or parent relationship between specific resources.

local area network

See *public LAN*.

local host

The host from which a command or process is initiated.

See also *remote host*.

local MAC address

MAC address that the system administrator of a local area network (LAN) system guarantees to be unique within that system.

log file

The file that contains a record of significant system events or messages. The base monitor, wizards, and detectors can have their own log files.

logical volume (GDS)

General term for a virtual disk device that the user can access directly. The user can access a logical volume in the same way as accessing a physical disk slice (partition). A logical volume is sometimes simply called a "volume."

low-order group (GDS)

Group that belongs to another group. A volume cannot be created in a low-order group.

MA

Monitoring Agents

MAC address

Address that identifies the office or node that is used by the MAC sublayer of a local area network (LAN).

MDS

See Meta Data Server.

message

A set of data transmitted from one software process to another process, device, or file.

message queue

A designated memory area which acts as a holding place for messages.

Meta Data Server (GFS)

GFS daemon that centrally manages the control information of a file system (meta-data).

MIB

Management Information Base

MIPC

Mesh Interprocessor Communication

mirrored volume (GDS)

A volume that is created in a mirror group. Data redundancy is created by mirroring.

mirror group (GDS)

A disk group of the mirror type. This a collection of mutually mirrored disks or low-order groups.

mirroring (GDS)

A setup that maintains redundancy by writing the same data to multiple slices. Even if an error occurs in some of the slices, this setup allows access to the volume to continue as long as a normal slice remains.

mirror virtual disk (VM)

Mirror virtual disks consist of two or more physical devices, and all output operations are performed simultaneously on all of the devices. (Applies to transitioning users of existing Fujitsu Technology Solutions products only.)

See also *concatenated virtual disk, simple virtual disk, striped virtual disk,* and *virtual disk*.

mixed model cluster

A cluster system that is built from different SPARC Enterprise models. For example, one node is a SPARC Enterprise M3000 machine, and another node is a SPARC Enterprise M4000 machine.

The models are divided into several groups, which are represented by the SPARC M10-1/M10-4/M10-4S machines, SPARC S7-2/S7-2L machines, SPARC T7-1/T7-2/T7-4 machines, SPARC T5-2/T5-4/T5-8 machines, SPARC T4-1/T4-2/T4-4 machines, SPARC T3-1/T3-2/T3-4, SPARC Enterprise T1000/T2000, SPARC Enterprise T5120/T5220/T5140/T5240/T5440, and the SPARC Enterprise M3000/M4000/M5000/M8000/M9000 machines.

monitoring agent

Component that monitors the state of a remote cluster node and immediately detects if that node goes down. This component is separate from the SA function.

mount point

The point in the directory tree where a file system is attached.

multihosting

Same disk via multiple controllers. (Applies to transitioning users of existing Fujitsu Technology Solutions products only.)

native operating system

The part of an operating system that is always active and translates system calls into activities.

network adapter

A LAN network adapter.

network interface card

See *network adapter*.

network partition (CF)

This condition exists when two or more nodes in a cluster cannot communicate over the interconnect; however, with applications still running, the nodes can continue to read and write to a shared device, compromising data integrity.

NIC

network interface card

NIC switching mode

One of the LAN duplexing modes presented by GLS. The duplexed NIC is used exclusively, and LAN monitoring between the Solaris server and the switching HUB, and switchover if an error is detected are implemented.

node

A host which is a member of a cluster. A computer node is the same as a computer.

node state (CF)

Every node in a cluster maintains a local state for every other node in that cluster. The node state of every node in the cluster must be either UP, DOWN, or LEFTCLUSTER.

See also *UP*, *DOWN*, *LEFTCLUSTER*.

NSM

Node State Monitor

object (RMS)

In the configuration file or a system graph, this is a representation of a physical or virtual resource.

See also *leaf object*, *object definition*, *node state*, *object type*.

object definition (RMS)

An entry in the configuration file that identifies a resource to be monitored by RMS. Attributes included in the definition specify properties of the corresponding resource. The keyword associated with an object definition is object.

See also *attribute*, *object type*.

object type (RMS)

A category of similar resources monitored as a group, such as disk drives. Each object type has specific properties, or attributes, which limit or define what monitoring or action can occur. When a resource is associated with a particular object type, attributes associated with that object type are applied to the resource.

See also *generic type*.

online maintenance

The capability of adding, removing, replacing, or recovering devices without shutting or powering off the host.

operating system dependent (CF)

This module provides an interface between the native operating system and the abstract, OS-independent interface that all PRIMECLUSTER modules depend upon.

OPS

See *Oracle Parallel Server*.

Oracle Parallel Server

Oracle Parallel Server allows access to all data in the database to users and applications in a clustered or MPP (massively parallel processing) platform.

Oracle VM Server for SPARC

Virtualization function using Hypervisor, which is provided as part of the firmware.

OSD (CF)

See *operating system dependent*.

OSLC (Oracle Solaris Legacy Containers)

A virtualization function to migrate Oracle Solaris 8/9 environment to hardware on which Solaris 10 is installed.

parent (RMS)

An object in the configuration file or system graph that has at least one child.

See also *child*, *configuration file*, and *system graph*.

PAS

Parallel Application Services

patrol diagnosis

A function that periodically diagnoses hardware faults.

physical IP address

IP address that is assigned directory to the interface (for example, hme0) of a network interface card. See also *logical IP address*. For information about the logical interface, see the explanation of logical interface in ifconfig(1M).

primary host (RMS)

The default host on which a user application comes online when RMS is started. This is always the hostname of the first child listed in the userApplication object definition.

PRIMECLUSTER services (CF)

Service modules that provide services and internal interfaces for clustered applications.

private network address

Private network addresses are a reserved range of IP addresses specified by RFC1918. They may be used internally by any organization but, because different organizations can use the same addresses, they should never be made visible to the public internet.

private resource (RMS)

A resource accessible only by a single host and not accessible to other RMS hosts.

See also *resource*, *shared resource*.

PS

Parallel Server

public LAN

The local area network (LAN) by which normal users access a machine.

See also *Administrative LAN*.

queue

See *message queue*.

quorum

State in which integrity is maintained among the nodes that configure the cluster system. Specifically, the CF state in all nodes that configure the cluster system is either UP or DOWN (there is no LEFTCLUSTER node).

RAO

RMS-Add on

RCI

Remote Cabinet Interface

redundancy

This is the capability of one object to assume the resource load of any other object in a cluster, and the capability of RAID hardware and/or RAID software to replicate data stored on secondary storage devices.

Reliant Monitor Services (RMS)

The package that maintains high availability of user-specified resources by providing monitoring and switchover capabilities.

remote console connection unit

Device that converts an RS232C interface and a LAN interface. This device allows another device (personal computer) that is connected to the LAN to use the TTY console functions through the Telnet function.

remote host

A host that is accessed through a telecommunications line or LAN.

See also *local host*.

remote node

See *remote host*.

reporting message (RMS)

A message that a detector uses to report the state of a particular resource to the base monitor.

resource (RMS)

A hardware or software element (private or shared) that provides a function, such as a mirrored disk, mirrored disk pieces, or a database server. A local resource is monitored only by the local host.

See also *private resource*, *shared resource*.

resource database (CF)

Database that manages information on hardware units that are shared among multiple nodes.

The resource database is managed by the cluster resource management facility.

resource definition (RMS)

See *object definition*.

resource label (RMS)

The name of the resource displayed in a system graph.

resource state (RMS)

Current state of a resource.

RMS

See *Reliant Monitor Services*.

RMS command

Commands that enable RMS resources to be administered from the command line.

RMS configuration

A configuration in which two or more nodes are connected to shared resources. Each node has its own copy of operating system and RMS software, as well as its own applications.

RMS Wizard kit

Each component of the RMS Wizard Kit adds new menu items to the RMS Wizard Tools for a specific application.

See also *RMS Wizard Tools*, *Reliant Monitor Services (RMS)*.

RMS Wizard Tools

A software package composed of various configuration and administration tools used to create and manage applications in an RMS configuration.

See also *RMS Wizard kit*, *Reliant Monitor Services*.

route

In the PRIMECLUSTER Concepts Guide, this term refers to the individual network paths of the redundant cluster interfaces that connect the nodes to each other.

Rolling update

Update method used to fix an application or maintenance within the cluster system. Fix application is enabled by applying fixes to each node sequentially without stopping jobs.

SA

Shutdown Agent

SAN (Storage Area Network)

See *Storage Area Network*.

SC

Scalability Cluster

scalability

The ability of a computing system to dynamically handle any increase in work load. Scalability is especially important for Internet-based applications where growth caused by Internet usage presents a scalable challenge.

scope (GDS)

The range of nodes that can share objects in the shared type disk class.

script (RMS)

A shell program executed by the base monitor in response to a state transition in a resource. The script may cause the state of a resource to change.

SD

Shutdown Daemon

SDX disk (GDS)

General term for disks that GDS manages. Depending on its use, an SDX disk may be called a single disk, a keep disk, a spare disk, or an undefined disk. An SDX disk is sometimes simply called a "disk."

SDX object (GDS)

General term for resources that GDS manages. The resources include classes, groups, SDX disks, and volumes.

SF

Shutdown Facility

shared disk connection confirmation

Function that checks whether that all shared disk units are turned on and all cable connections are correct when a node is started.

shared resource

A resource, such as a disk drive, that is accessible to more than one node.

See also private resource, resource.

Shutdown Facility

A facility that forcibly stops a node in which a failure has occurred. When PRIMECLUSTER decides that system has reached a state in which the quorum is not maintained, it uses the Shutdown Facility (SF) to return the cluster system to the quorum state.

shutdown request

Instruction that forcibly stops the specified node so that the quorum is restored.

simple virtual disk

Simple virtual disks define either an area within a physical disk partition or an entire partition.

See also *concatenated virtual disk*, *striped virtual disk*, and *virtual disk*.

single disk (GDS)

SDX disk that does not belong to a group and can be used to create a single volume.

single-node cluster

An operation mode of a cluster system consisting of one node.

single volume (GDS)

A volume that is created in a single disk that does not belong to a group. There is no data redundancy.

spare disk (GDS)

A spare disk for restoring the mirroring state in place of a failed disk.

state

See *resource state*.

state transition procedure

The state transition procedure receives a state transition instruction from the cluster control and controls activation and deactivation of the resource (start and stop of the application).

Storage Area Network

The high-speed network that connects multiple, external storage units and storage units with multiple computers. The connections are generally fiber channels.

striped group (GDS)

A disk group of the stripe type. This is a collection of disks or low-order groups that become striping units.

striped virtual disk

Striped virtual disks consist of two or more pieces. These can be physical partitions or further virtual disks (typically a mirror disk). Sequential I/O operations on the virtual disk can be converted to I/O operations on two or more physical disks. This corresponds to RAID Level 0 (RAID0).

See also *concatenated virtual disk*, *mirror virtual disk*, *simple virtual disk*, *virtual disk*.

striped volume (GDS)

A volume that is created in a striped group. Striping allows the I/O load to be distributed among multiple disks. There is no data redundancy.

stripe width (GDS)

The size is which data is divided when striping takes place.

striping (GDS)

Dividing data into fixed-size segments, and cyclically distributing and writing the data segments to multiple slices. This method distributes I/O data to multiple physical disks and issues I/O data at the same time.

switching mode

LAN duplexing mode presented by GLS.

There is a total of five switching mode types: fast switching mode, NIC switching mode, GS/SURE linkage mode, multipath mode, and multilink Ethernet mode:

switchover

The process by which a user application transfers processes and data inherited from an operating node to a standby node, based on a user request.

switchover (RMS)

The process by which RMS switches control of userApplication over from one monitored host to another.

See also *automatic switchover*, *directed switchover*, *failover*, and *symmetrical switchover*.

symmetrical switchover (RMS)

This means that every RMS host is able to take on resources from any other RMS host.

See also *automatic switchover*, *directed switchover*, *failover*, and *switchover.*

synchronized power control

When the power of one node is turned in the cluster system, this function turns on all other powered-off nodes and disk array unit that are connected to nodes through RCI cables.

system disk (GDS)

The disk in which the operating Solaris is installed. This term refers to the entire disk, including slices that are currently operating as one of the following file systems or swap area:

/, /usr, /var, or swap area

system graph (RMS)

A visual representation (a map) of monitored resources used to develop or interpret the configuration file.

See also *configuration file*.

template

See *application template*.

type

See *object type*.

UP (CF)

A node state that indicates that the node can communicate with other nodes in the cluster.

See also *DOWN*, *LEFTCLUSTER*, *node state*.

user group

A group that limits the environment setup, operation management, and other operations presented by Web-Based Admin View and the Cluster Admin GUI. There are four user groups: wvroot, clroot, cladmin, and clmon. Each user ID is registered in an appropriate user group by the operation system administrator of the management server.

VIP

Virtual Interface Provider

virtual disk

With virtual disks, a pseudo device driver is inserted between the highest level of the Solaris logical Input/Output (I/O) system and the physical device driver. This pseudo device driver then maps all logical I/O requests on physical disks.

See also *concatenated virtual disk*, *mirror virtual disk*, *simple virtual disk*, *striped virtual disk*.

volume (GDS)

See *logical volume (GDS)*.

warm-standby

In Oracle Solaris Zones environments, with the non-global zones started up on both the operating server and standby server as is, this operation switches over only the applications operating within the non-global zone, and takes over services. Since the standby system's non-global zone OS enters a startup status, a faster switchover than the cold-standby is possible.

watchdog timer monitoring

Timer value that measures operating system hangs and boot failures.

Web-Based Admin View

This is a common base enabling use of the Graphic User Interface of PRIMECLUSTER. This interface is in Java.

Wizard (RMS)

An interactive software tool that creates a specific type of application using pretested object definitions. An enabler is a type of wizard.

WK

Wizard Kit

WT

Wizard Tools

XSCF

Abbreviation for eXtended System Control Facility. XSCF is a system monitoring facility that consists of dedicated processors that are independent from a main CPU. XSCF performs integrated management of the cooling system (FAN unit), power supply unit, system monitoring, and power on/off and monitoring system of peripherals. This is enabled from remote places, providing functions to monitor a main unit, notify a system administrator of a system failure, and perform console input/output from remote places via serial port or Ethernet port.

Zone name

The name to be defined by the -z of the zonecfg and the zoneadm.

Index

[T]

[U]

[V]