FUJITSU

FUJITSU Software PRIMECLUSTER

Installation and Administration Guide 4.4

Linux

J2UL-2101-02ENZ0(05)
May 2020

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. Maintaining the PRIMECLUSTER system	Part 5 Maintenance

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 4.4 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 five parts, appendixes, 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 building, 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 Modifications

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 items related to the maintenance of the PRIMECLUSTER system.

Appendix A PRIMECLUSTER Products

Audience: Users who operate PRIMECLUSTER products on PRIMECLUSTER systems
Contents: This appendix describes the list of products supported by PRIMECLUSTER systems.

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 Registering, Changing, and Deleting State Transition Procedure Resources for PRIMECLUSTER Compatibility

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

Appendix E Configuration Update Service for SA

Audience: All users who use PRIMECLUSTER systems
Contents: This appendix descries Configuration Update Service for SA.

Appendix F Using PRIMECLUSTER in a KVM environment

Audience: All users who use PRIMECLUSTER systems in a KVM environment

Contents: This appendix describes the installation procedures for using the PRIMECLUSTER system in a KVM environment.

Appendix G Using the Migration Function in a KVM Environment

Audience: All users who use the migration function in a KVM Environment
Contents: This appendix describes the procedure for using the migration function in a KVM Environment.

Appendix H Using PRIMECLUSTER in a VMware Environment

Audience: All users who use PRIMECLUSTER systems in a VMware environment
Contents: This appendix describes the installation procedures for using the PRIMECLUSTER system in a VMware environment.

Appendix I Startup Scripts and Startup Daemons, and Port Numbers in PRIMECLUSTER

Audience: System administrators who build PRIMECLUSTER systems
Contents: This appendix provides explanations on scripts and daemons that are started by PRIMECLUSTER, and the port numbers being used.

Appendix J Systemd Service and Startup Daemons, and Port Numbers in PRIMECLUSTER

Audience: System administrators who build PRIMECLUSTER systems
Contents: This appendix provides explanations on systemd services and daemons that are started by PRIMECLUSTER, and the port numbers being used.

Appendix K Using Firewall

Audience: All users who use PRIMECLUSTER systems
Contents: This appendix describes the procedure when using Firewall in the PRIMECLUSTER system.

Appendix L Cloning the Cluster System Environment

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

Appendix M Changes in Each Version

Audience: All users who use PRIMECLUSTER 4.0A20, 4.1A20, 4.1A30, 4.2A00, 4.2A30, 4.3A00, 4.3A10, 4.3A20, 4.3A30, or 4.3A40.
Contents: This appendix describes the changes made to the specifications of PRIMECLUSTER 4.4A00.

Appendix N 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 Installation and Administration Guide FUJITSU Cloud Service K5
- 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 Web-Based Admin View Operation Guide
- PRIMECLUSTER Messages
- PRIMECLUSTER Wizard for Oracle Configuration and Administration Guide
- FJQSS (Information Collection Tool) User's Guide
- Interstage Application Server Reference Manual - Command Edition
- Interstage Application Server Operator's Guide
- PRIMEQUEST 1000 Series User Interface Operating Instructions
- PRIMEQUEST 1000 Series Message Reference
- PRIMEQUEST 1000 Series Tool Reference
- PRIMEQUEST 1000 Series Installation Manual
- PRIMEQUEST 1000 Series Administration Manual
- PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual
- PRIMEQUEST 1000 SERIES HBA blockage function USER'S GUIDE
- PRIMEQUEST 2000 Series Tool Reference
- PRIMEQUEST 2000 Series Installation Manual
- PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual
- PRIMEQUEST 2000 Series HBA blockage function USER'S GUIDE

Note

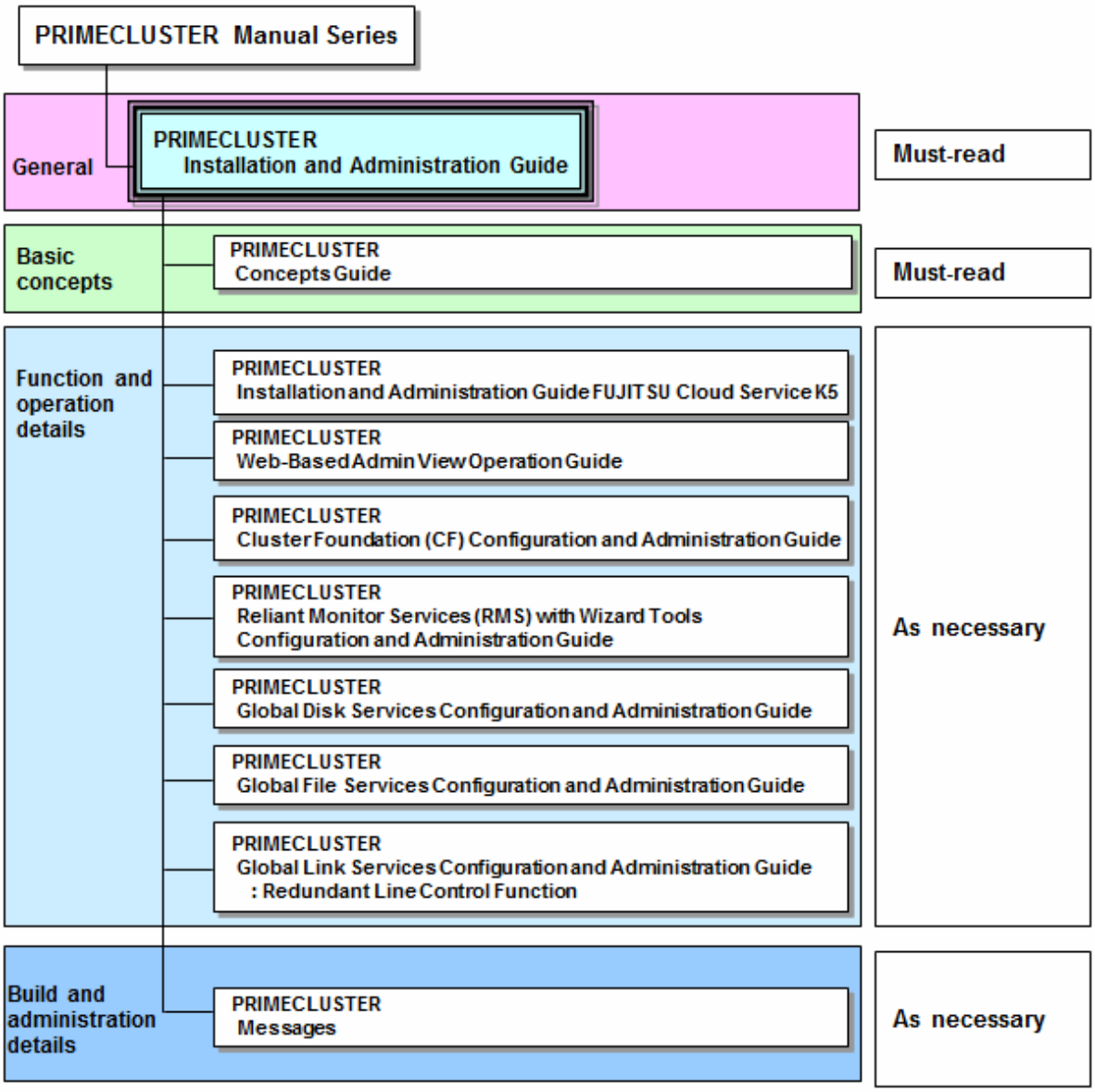
The PRIMECLUSTER documentation includes the following documentation in addition to those listed above:

- PRIMECLUSTER Software Release Guide and Installation Guide

  This Software Release Guide and Installation Guide are provided with each PRIMECLUSTER product package.

  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 Software Release Guide for PRIMECLUSTER that comes with the product.

Adobe Reader is required to read and print this 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 "4.3.1 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 Linux(R) 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:0:0:root:/root:/bin/bash

bin:x:1:1:bin:/bin:/bin/bash

daemon:x:2:2:daemon:/sbin:/bin/bash

lp:x:4:7:lp daemon:/var/spool/lpd:/bin/bash

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

- Red Hat Enterprise Linux is abbreviated as RHEL.
- Red Hat Enterprise Linux AS is abbreviated as RHEL-AS.
- RHEL and RHEL-AS are abbreviated as Linux(R).
- VMware(R) ESXi(TM) is abbreviated as ESXi.
- VMware vSphere(R) is abbreviated as VMware vSphere.
- VMware vSphere(R) High Availability is abbreviated as VMware HA.
- VMware vSphere(R) Fault Tolerance is abbreviated as VMware FT.
- VMware vSphere(R) Distributed Resource Scheduler(TM) is abbreviated as VMware DRS.
- VMware vSphere(R) Distributed Power Management (TM) is abbreviated as VMware DPM.
- VMware vSphere(R) vMotion(R) is abbreviated as VMware vMotion.
- VMware vSphere(R) Storage vMotion(R) is abbreviated as VMware Storage vMotion.
- VMware(R) vCenter(TM) Converter(TM) is abbreviated as VMware vCenter Converter.
- PRIMEQUEST 2000/1000 Series are abbreviated as PRIMEQUEST.

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

Red Hat is a registered trademark of Red Hat, Inc. in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds.

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.

VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions.

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

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

April 2016, First edition
June 2016, Second edition

August 2016, 2.1 edition
June 2017, 2.2 edition
August 2017, 2.3 edition
January 2018, 2.4 edition
May 2020, 2.5 edition

Copyright notice

All Rights Reserved, Copyright (C) FUJITSU LIMITED 2016-2020.

Revision History

Revision	Location	Edition
Added the note about kdump_post in /etc/kdump.conf.	3.3 PRIMECLUSTER Installation	2.1
Added the note when using a bonding device as the cluster interconnect.	5.1.1 Setting Up CF and CIP	
Added the note when executing "sdtool -s" after starting the shutdown facility, and when starting the MMB asynchronous monitoring daemon for the first time.	5.1.2.4.6 Starting the Shutdown Facility	
Added the note when creating the userApplication name.	6.7.2 Setting Up userApplication	
Added the information about the scripts that can be registered to userApplication.	6.7.2.1 Creating Standby Cluster Applications	
Added the note when "none" is set to StopCommands.	6.7.3.1 Setting Up Cmdline Resources	
Added the description about "Timeout of Check Script."	6.11.2.2.4 Timeout of Check Script	
Added the note about the status icon of userApplication in maintenance mode.	7.1.3.1 RMS Tree	
Added the procedure to clear the fault trace of a resource to the corrective action for failed resources.	7.4.1.2 Corrective Action for Failed Resources	
Added the procedure to remove the system board by using Dynamic Reconfiguration.	8.1.2.3 Removing System Board by Dynamic Reconfiguration	
Changed the command execution example in RHEL7 environment.	9.2.1 Changing the IP Address of the Public LAN 9.2.2 Changing the IP Address of the Administrative LAN	
Added the note when all userApplications are deleted.	10.2.2 Deleting a userApplication	
Added the attribute to be used as a resource name to "Contents to be changed" of "When a cluster application that is performing standby operation is to be changed."	10.3 Changing the Cluster Configuration	
Changed the procedure on how to change the startup priority of the state transition procedure.	D.2.2 Changing the Startup Priority of a State Transition Procedure	
Changed the procedure on how to change the registration information of a procedure resource.	D.2.3 Changing registration information of a procedure resource	
Added the note when changing the registration information of the procedure resource.	D.2.3 Changing registration information of a procedure resource	
Added the note about hot swap of the virtual machine software.	H.1 Cluster Systems in a VMware Environment	
Changed the description about the settings of the virtual network when installing VMware vSphere.	H.2.1.1 Installation and Configuration of Related Software	

Revision	Location	Edition
Added the following scripts to /etc/rc3.d. - S05poffinhibit - S07cllkcd - S27SMAWsfex	I.2 Startup Script Lists	
Changed the utilized ports of S13SMAWsf.	I.2 Startup Script Lists	
Changed the utilized ports of smawsf.service.	J.2 systemd Service Lists	
Added poffinhibit.service.	J.2 systemd Service Lists	
Changed the procedure to delete the shared disk resources.	L.3.4 Restoring the GDS Configuration Information	
Changed the description of "Changes" of hvshut command.	Appendix M Changes in Each Version	
Added the description for FUJITSU Cloud Service K5.	2.2 System Design	2.2
Added the description when using mirroring among servers.	3.2.1.1 Host OS setup (before installing the operating system on guest OS) 3.2.1.2 Host OS setup (after installing the operating system on guest OS) 3.2.2.2 Host OS setup (after installing the operating system on guest OS) 3.2.3.1.3 Host OS setup (before installing the operating system on guest OS) 3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)	2.3
Changed the note about synchronizing time.	1.7.1 Common Notes on PRIMERGY/ PRIMEQUEST	2.4
Changed the execution example.	3.1.6.1 PRIMERGY 5.1.2.3.3 Setting up IPMI Shutdown Agent E.3.1 Startup Configuration for the IPMI Service	
Changed the note about the timeout value of CF heartbeat.	3.2.3.3 Building a cluster (Host OS)	
Added the reference of the manual.	4.3.3.2 Confirming Web-Based Admin View Startup	
Changed the description of survival scenarios.	5.1.2.1 Survival Priority	
Added the note when "TestFailed" is displayed as the test status.	5.1.2.3.6 Starting up the Shutdown Facility	
Changed the output example of sfkvmtool -c message.	5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only) G.2.2 Using the Host OS failover function	
Added the description about Check script exit code.	6.6 Setting Up Online/Offline Scripts	
Added the information about the list of attributes settable to userApplication and the details of these attributes, and changed the description of information about the scripts which can be registered to userApplication.	6.7.2.1 Creating Standby Cluster Applications	
Changed the description about the attribute changing procedure.	6.7.3.1 Setting Up Cmdline Resources	
Added the note about the file system on the volume of LVM.	6.7.3.2 Setting Up Fsystem Resources	
Changed the description of how to failover a userApplication in the event of a node failure, resource failure, and RMS stop.	6.10.1 Setting Contents of a Cluster Application	

Revision	Location	Edition
Change the outline of MONITORONLY attribute.	6.11 Notes When Setting Cmdline Resources	
Changed the Gls resource status from "Wait" to "Online processing" in the figure.	6.11.1.5 Flow of the Cmdline Resource Operation	
Added the note about using the maintenance mode.	6.11.2.1.4 Notes When Setting the NULLDETECTOR Flag	
Added the note about Check script exit codes.	6.11.2.2.3 Check Script Exit Code	
Added the note about when using PRIMECLUSTER products and setting the NULLDETECTOR flag.	7.2.2.6 Entering maintenance mode for Cluster Application	
Changed the description of selecting a resource to be deleted.	10.5 Deleting a Resource	
Added the note about the VMware environment where the I/O fencing function is used.	10.5 Deleting a Resource 10.6.2 Changing the Devices of File systems Controlled by the Fsystem Resource	
Deleted hvlogcontrol and config.us.	B.6 RMS	
Deleted the descriptions about hot clone, Data Recovery and backup by VCB.	Appendix H Using PRIMECLUSTER in a VMware Environment	
Changed the description of the I/O fencing function when a cluster interconnect error occurs.	H.1 Cluster Systems in a VMware Environment	
Added the note about when using the forcible stop with the I/O fencing function.	H.1 Cluster Systems in a VMware Environment	
Added the configuration of PRIMECLUSTER Wizard for SAP HANA in comparison with the forcible stop.	H.1 Cluster Systems in a VMware Environment	
Changed the description about forcible stop functions.	H.1 Cluster Systems in a VMware Environment	
Changed the setting of the path policy for the Native Multipathing (NMP)	H.1 Cluster Systems in a VMware Environment H.2 Installation	
Added notes about the cluster application in which Gds resource and Fsystem resource are registered.	H.2 Installation	
Changed the description about settings for shared disks when using the I/O fencing function.	H.2.1.1 Installation and Configuration of Related Software	
Added the description about PreOnline script, OfflineDone script and Fault script.	H.2.4.1 Setting Up I/O Fencing Function	
Added chronyd.service.	J.3 Necessary Services for PRIMECLUSTER to Operate	
Added the setting when using neither GLS nor takeover network.	L.3.7 Changing the Settings of Cluster Application Information	
Added the note when using the ntpdate service in a RHEL7 environment.	3.3 PRIMECLUSTER Installation	2.5
Added the note when setting management servers of Web-Based Admin View.	4.3.3 Initial Setup of Web-Based Admin View	
Added the note about setting IP addresses.	4.3.3.1 Initial setup of the operation management server	
Changed the description of the example of HV_NODENAME.	6.6 Setting Up Online/Offline Scripts 6.11.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts	
Added the note while executing the hvw command.	6.7.1 Starting RMS Wizard	

Revision	Location	Edition
Added the description about the definition of mount points.	6.7.3.2 Setting Up Fsystem Resources	
Added the note when PRIMECLUSTER Wizard for NAS is used.	6.7.5 Registering the Cluster Service of a PRIMECLUSTER-compatible product	
Added the description about a cluster application transiting to the Standby state, to "Information."	6.7.7 Exclusive Relationships Between Cluster Applications	
Changed the description of notes when setting a cluster application.	6.10.2 Notes on Configuration	
Added the description of timeout values of the Check script.	6.11.2.2.4 Timeout of Check Script	
Added the note when creating the Fsystem resource.	6.12.1 Monitoring Fsystem	
Added the note when displaying logs from Cluster Admin.	7.1.3.3 Switchlogs and application logs 7.3.4 Viewing Logs Created by the PRIMECLUSTER System	
Changed the note when starting the maintenance mode.	7.2.2.6 Entering maintenance mode for Cluster Application	
Added the note when starting a system in a RHEL6 environment.	7.5 Notes on Operation	
Changed the procedure when changing the IP address of the public LAN.	9.2.1 Changing the IP Address of the Public LAN	
Changed the procedure when changing the IP address of the administrative LAN and added "See."	9.2.2 Changing the IP Address of the Administrative LAN	
Changed the description about setting RELIANT_SHUT_MIN_WAIT.	11.2.1 Changing Timeout Time during RMS Stop Processing	
Changed the description when changing the host names in the takeover network resources.	L.3.7.2 When Using the Takeover Network	

Contents

Part 1 Planning 1

Chapter 1 Build Flow 2
1.1 Planning 2
1.2 Installation 3
1.3 Development 5
1.4 Test 6
1.5 Operation and Maintenance 9
1.6 Operation Mode Change 9
1.7 Notes When Building a System 10
1.7.1 Common Notes on PRIMERGY/PRIMEQUEST 10
1.7.2 Notes on PRIMERGY 12
1.7.3 Notes on PRIMEQUEST 12
1.7.4 Notes When Building a Cluster System Using a Virtual Machine Function 12

Chapter 2 Site Preparation 14
2.1 PRIMECLUSTER Product Selection 14
2.1.1 Product Selection 14
2.1.2 Function Selection 15
2.2 System Design 15
2.2.1 Virtual Machine Function 16
2.3 Determining the Cluster System Operation Mode 29
2.3.1 Standby Operation 30
2.3.2 Scalable Operation 36
2.3.3 Single-Node Cluster Operation 37
2.4 Determining the Web-Based Admin View Operation Mode 40
2.5 Determining the failover timing of a cluster application 43

Part 2 Installation 44

Chapter 3 Software Installation and Setup 45
3.1 When Not Using the Virtual Machine Function 45
3.1.1 Setting Up the Network 46
3.1.2 Setting Up NTP 46
3.1.3 Setting Up Disk Units 46
3.1.4 Setting Up Hardware Monitoring with ServerView 47
3.1.5 Installing PRIMECLUSTER 47
3.1.6 Setting Up the Cluster High-Speed Failover Function 47
3.1.6.1 PRIMERGY 47
3.1.6.2 PRIMEQUEST 50
3.1.7 Checking and Setting the Kernel Parameters 52
3.1.8 Installing and Setting Up Applications 55
3.2 When Using the Virtual Machine Function 56
3.2.1 When building a cluster system between guest OSes on one host OS 56
3.2.1.1 Host OS setup (before installing the operating system on guest OS) 57
3.2.1.2 Host OS setup (after installing the operating system on guest OS) 57
3.2.1.3 Guest OS setup 62
3.2.1.4 NTP setup (host OS and guest OS) 63
3.2.1.5 Installing PRIMECLUSTER on guest OSes 63
3.2.1.6 Checking and setting the kernel parameters 63
3.2.1.7 Installing and setting up applications 63
3.2.2 When building a cluster system between guest OSes on multiple host OSes without using Host OS failover function 63
3.2.2.1 Host OS setup (before installing the operating system on guest OS) 64
3.2.2.2 Host OS setup (after installing the operating system on guest OS) 65
3.2.2.3 Guest OS setup 69
3.2.2.4 NTP setup (host OS and guest OS) 70

3.2.2.5 Installing PRIMECLUSTER on guest OSes ... 70
3.2.2.6 Checking and setting the kernel parameters ... 70
3.2.2.7 Installing and setting up applications ... 71
3.2.3 When building a cluster system between guest OSes on multiple host OSes using Host OS failover function ... 71
3.2.3.1 Installation and Setup of Software (Host OS) ... 71
3.2.3.1.1 Network setup ... 72
3.2.3.1.2 NTP setup ... 72
3.2.3.1.3 Host OS setup (before installing the operating system on guest OS) ... 72
3.2.3.1.4 Host OS setup (after installing the operating system on guest OS) ... 72
3.2.3.1.5 Installing PRIMECLUSTER on the host OS ... 77
3.2.3.1.6 Setting up the cluster high-speed failover function ... 77
3.2.3.1.7 Checking and setting the kernel parameters ... 77
3.2.3.2 Preparation prior to building a cluster (Host OS) ... 77
3.2.3.3 Building a cluster (Host OS) ... 77
3.2.3.4 Software installation and setup (Guest OS) ... 78
3.2.3.4.1 Guest OS setup ... 78
3.2.3.4.2 NTP setup (Guest OS) ... 79
3.2.3.4.3 Installing PRIMECLUSTER on guest OSes ... 79
3.2.3.4.4 Checking and setting the kernel parameters ... 79
3.2.3.4.5 Installing and setting up applications ... 80
3.2.3.5 Preparation prior to building a cluster (Guest OS) ... 80
3.2.3.6 Building a Cluster (Guest OS) ... 80
3.2.3.7 Building cluster applications (Guest OS) ... 80
3.3 PRIMECLUSTER Installation ... 80
3.4 Installation and Environment Setup of Applications ... 83

Chapter 4 Preparation Prior to Building a Cluster ... 84
4.1 Checking PRIMECLUSTER Designsheets ... 85
4.2 Activating the Cluster Interconnect ... 85
4.3 Preparations for Starting the Web-Based Admin View Screen ... 85
4.3.1 Assigning Users to Manage the Cluster ... 86
4.3.2 Preparing the Client Environment ... 87
4.3.3 Initial Setup of Web-Based Admin View ... 87
4.3.3.1 Initial setup of the operation management server ... 87
4.3.3.2 Confirming Web-Based Admin View Startup ... 88
4.3.3.3 Setting the Web-Based Admin View Language ... 89
4.3.4 Setting Up the Browser ... 90
4.3.5 Setting Up the Java Plug-in ... 90
4.4 Starting the Web-Based Admin View Screen ... 91
4.5 Web-Based Admin View Screen ... 92
4.5.1 Operation Menu Functions ... 93
4.5.2 Global Cluster Services Menu Functions ... 95
4.5.3 Cluster Admin Functions ... 95
4.6 Exiting the Web-Based Admin View Screen ... 96

Chapter 5 Building a Cluster ... 98
5.1 Initial Cluster Setup ... 98
5.1.1 Setting Up CF and CIP ... 99
5.1.2 Setting up the Shutdown Facility ... 101
5.1.2.1 Survival Priority ... 102
5.1.2.2 Setup Flow for Shutdown Facility ... 106
5.1.2.2.1 Setup Flow in PRIMERGY RX/TX Series ... 106
5.1.2.2.2 Setup Flow in PRIMERGY BX Series ... 107
5.1.2.2.3 Setup Flow in PRIMEQUEST ... 107
5.1.2.2.4 Setup Flow in Virtual Machine Environment ... 107
5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY ... 108
5.1.2.3.1 Checking the Shutdown Agent Information ... 108
5.1.2.3.2 Setting up the Shutdown Daemon ... 109

5.1.2.3.3 Setting up IPMI Shutdown Agent....110
5.1.2.3.4 Setting up Blade Shutdown Agent....112
5.1.2.3.5 Setting up kdump Shutdown Agent....114
5.1.2.3.6 Starting up the Shutdown Facility....115
5.1.2.3.7 Test for Forced Shutdown of Cluster Nodes....116
5.1.2.4 Setup Procedure for Shutdown Facility in PRIMEQUEST....116
5.1.2.4.1 Checking the Shutdown Agent Information....116
5.1.2.4.2 Setting up the MMB Shutdown Agent....117
5.1.2.4.3 Setting up the Shutdown Daemon....117
5.1.2.4.4 Starting the MMB Asynchronous Monitoring Daemon....118
5.1.2.4.5 Setting I/O Completion Wait Time....118
5.1.2.4.6 Starting the Shutdown Facility....119
5.1.2.4.7 Test for Forced Shutdown of Cluster Nodes....121
5.1.2.5 Setup Procedure for Shutdown Facility in Virtual Machine Environment....121
5.1.2.5.1 Checking the Shutdown Agent Information....121
5.1.2.5.2 Setting up libvirt Shutdown Agent....122
5.1.2.5.3 Setting Up vmchkhost Shutdown Agent....123
5.1.2.5.4 Setting up the Shutdown Daemon....124
5.1.2.5.5 Starting the Shutdown Facility....125
5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)....126
5.1.2.5.7 Test for Forced Shutdown of Cluster Nodes....128
5.1.3 Initial Setup of the Cluster Resource Management Facility....128
5.1.3.1 Initial Configuration Setup....129
5.1.3.2 Registering Hardware Devices....131
5.2 Setting Up Fault Resource Identification and Operator Intervention Request....135

Chapter 6 Building Cluster Applications....138
6.1 Initial RMS Setup....140
6.2 Initial GLS Setup....140
6.2.1 GLS Setup....140
6.2.2 Setting Up Web-Based Admin View When GLS Is Used....144
6.3 GDS Configuration Setup....145
6.3.1 Setting Up System Disk Mirroring....145
6.3.2 Setting Up Shared Disks....151
6.4 Initial GFS Setup....159
6.4.1 File System Creation....161
6.5 Setting Up the Application Environment....166
6.6 Setting Up Online/Offline Scripts....166
6.7 Setting Up Cluster Applications....171
6.7.1 Starting RMS Wizard....175
6.7.2 Setting Up userApplication....175
6.7.2.1 Creating Standby Cluster Applications....175
6.7.2.2 Creating Scalable Cluster Applications....181
6.7.3 Setting Up Resources....186
6.7.3.1 Setting Up Cmdline Resources....187
6.7.3.2 Setting Up Fsystem Resources....190
6.7.3.3 Preliminary Setup for Gds Resources....194
6.7.3.4 Setting Up Gds Resources....194
6.7.3.5 Setting Up Gls Resources....195
6.7.3.6 Setting Up Takeover Network Resources....197
6.7.3.7 Setting Up Procedure Resources....200
6.7.4 Generate and Activate....202
6.7.5 Registering the Cluster Service of a PRIMECLUSTER-compatible product....203
6.7.6 Attributes....203
6.7.7 Exclusive Relationships Between Cluster Applications....203
6.8 Setting Up the RMS Environment....211
6.9 Checking the Cluster Environment....211

6.10 Setting Contents and Notes on Cluster Application ... 211
6.10.1 Setting Contents of a Cluster Application ... 211
6.10.2 Notes on Configuration ... 220
6.11 Notes When Setting Cmdline Resources ... 221
6.11.1 Scripts and State Transition ... 223
6.11.1.1 Scripts to be Executed in Each Resource State ... 225
6.11.1.2 Script States When Online ... 225
6.11.1.3 Script States When Standby ... 226
6.11.1.4 Script States When Offline ... 227
6.11.1.5 Flow of the Cmdline Resource Operation ... 227
6.11.1.6 Operation for Each Exit Code of the Check Script ... 231
6.11.2 Notes When Creating Scripts ... 234
6.11.2.1 start and stop Scripts ... 235
6.11.2.1.1 Examples of start and stop Scripts ... 235
6.11.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts ... 237
6.11.2.1.3 Exit Code of Start and Stop Scripts ... 238
6.11.2.1.4 Notes When Setting the NULLDETECTOR Flag ... 238
6.11.2.1.5 Timeout of Scripts ... 238
6.11.2.2 Check Script ... 239
6.11.2.2.1 Example of the Check Script ... 239
6.11.2.2.2 Environment Variables that can be referred to within the Check Scripts ... 240
6.11.2.2.3 Check Script Exit Code ... 240
6.11.2.2.4 Timeout of Check Script ... 241
6.11.3 Notes on Scripts ... 241
6.12 Notes When Setting Fsystem Resource ... 243
6.12.1 Monitoring Fsystem ... 243
6.12.2 Fsystem Resource Attribute ... 243
6.12.3 File System on the Shared Disk Device ... 244
6.12.3.1 Corrective Actions for the Forced File System Check ... 244
6.12.3.2 Corrective Actions for delayed allocation ... 245
6.12.4 Other Notes ... 245
6.12.5 Maintaining File Systems Controlled by the Fsystem Resource ... 246

Part 3 Operations ... 248

Chapter 7 Operations ... 249
7.1 Viewing the PRIMECLUSTER System Operation Management Screens ... 249
7.1.1 CF Main Window ... 249
7.1.2 CRM Main Window ... 250
7.1.2.1 Displayed Resource Types ... 251
7.1.2.1.1 Resource Icons ... 251
7.1.2.1.2 Resource States ... 251
7.1.2.1.3 Operations ... 252
7.1.2.2 Detailed Resource Information ... 254
7.1.3 RMS Main Window ... 255
7.1.3.1 RMS Tree ... 255
7.1.3.2 Configuration information or object attributes ... 259
7.1.3.3 Switchlogs and application logs ... 259
7.2 Operating the PRIMECLUSTER System ... 260
7.2.1 RMS Operation ... 260
7.2.1.1 Starting RMS ... 260
7.2.1.2 Stopping RMS ... 260
7.2.2 Cluster Application Operations ... 261
7.2.2.1 Starting a Cluster Application ... 261
7.2.2.2 Stopping a Cluster Application ... 261
7.2.2.3 Switching a Cluster Application ... 262
7.2.2.4 Bringing Faulted Cluster Application to available state ... 262
7.2.2.5 Clearing the Wait State of a Node ... 262

7.2.2.6 Entering maintenance mode for Cluster Application ... 263
7.2.3 Resource Operation ... 263
7.2.3.1 Starting Resources ... 264
7.2.3.2 Stopping Resources ... 264
7.2.3.3 Clearing Fault Traces of Resources ... 265
7.3 Monitoring the PRIMECLUSTER System ... 265
7.3.1 Monitoring the State of a Node ... 265
7.3.2 Monitoring the State of a Cluster Application ... 266
7.3.3 Concurrent Viewing of Node and Cluster Application States ... 267
7.3.4 Viewing Logs Created by the PRIMECLUSTER System ... 268
7.3.4.1 Viewing switchlogs ... 268
7.3.4.2 Viewing application logs ... 268
7.3.5 Viewing Detailed Resource Information ... 269
7.3.6 Displaying environment variables ... 270
7.3.7 Monitoring Cluster Control Messages ... 271
7.4 Corrective Actions for Resource Failures ... 271
7.4.1 Corrective Action in the event of a resource failure ... 271
7.4.1.1 Failure Detection and Cause Identification if a Failure Occurs ... 272
7.4.1.2 Corrective Action for Failed Resources ... 273
7.4.1.3 Recovery of Failed Cluster Interconnect ... 274
7.4.2 Corrective Action in the event of the LEFTCLUSTER state when the virtual machine function is used ... 275
7.4.2.1 When the host OS becomes the panic state ... 275
7.4.2.2 When the host OS hangs up ... 275
7.5 Notes on Operation ... 276
7.5.1 Notes on Switching a Cluster Application Forcibly ... 277
7.6 CF and RMS Heartbeats ... 279
7.7 cron Processing ... 280

Part 4 System Configuration Modification ... 282

Chapter 8 Changing the Cluster System Configuration ... 283
8.1 Adding, Deleting, and Changing Hardware ... 283
8.1.1 Adding Hardware ... 283
8.1.1.1 Adding a shared disk device ... 283
8.1.1.2 Adding a Network Interface Card Used for the Public LAN and the Administrative LAN ... 285
8.1.1.3 Adding Hardware by DR (Dynamic Reconfiguration) ... 286
8.1.2 Deleting Hardware ... 288
8.1.2.1 Deleting a shared disk device ... 288
8.1.2.2 Deleting a network interface card used for the public LAN and the administrative LAN ... 290
8.1.2.3 Removing System Board by Dynamic Reconfiguration ... 291
8.1.3 Changing Hardware ... 294
8.1.3.1 Changing a shared disk device ... 294
8.1.3.2 Changing a network interface card used for the public LAN and the administrative LAN ... 296
8.1.3.3 Changing NIC of CIP ... 297

Chapter 9 Changing the Cluster System Environment ... 299
9.1 Changing the Cluster Configuration information ... 299
9.1.1 Changing a Node Name ... 299
9.1.2 Changing the SF Node Weight ... 300
9.2 Changing the Network Environment ... 300
9.2.1 Changing the IP Address of the Public LAN ... 300
9.2.2 Changing the IP Address of the Administrative LAN ... 302
9.2.3 Changing a CIP Address ... 303
9.2.4 Changing the Subnet Mask of CIP ... 304
9.2.5 Changing the MTU Value of a Network Interface Used for Cluster Interconnects ... 304
9.3 Changing Option Hardware Settings ... 305
9.3.1 Changing MMB Settings ... 305
9.3.1.1 Changing the MMB IP Address ... 305

9.3.1.2 Changing the User Name and Password for Controlling the MMB with RMCP....306
9.3.2 Changing IPMI Settings....306
9.3.2.1 Changing the IPMI IP Address....306
9.3.2.2 Changing the User Name and Password for iRMC....307
9.3.3 Changing Blade Settings....308
9.3.3.1 Changing the IP Address of the Management Blade....308
9.3.3.2 Changing the Slot Number of Server Blades....308
9.4 Changing Virtual Machine Settings....309
9.4.1 Changing Host OS Settings (KVM environment)....309
9.4.1.1 Changing the IP address of the Host OS....309
9.4.1.2 Changing the Password of the Host OS Account (Shutdown Facility)....310
9.4.1.3 Changing the Settings in /etc/sysconfig/libvirt-guests....310

Chapter 10 Configuration change of Cluster Applications....312
10.1 Adding cluster applications....312
10.2 Deleting a Cluster Application....314
10.2.1 Deleting the Hardware Resource....314
10.2.2 Deleting a userApplication....314
10.3 Changing a Cluster Application....316
10.4 Add Resources....325
10.5 Deleting a Resource....327
10.5.1 Settings made when deleting a Gds resource....330
10.6 Changing Resources....330
10.6.1 Changing Host Names and IP Addresses of Takeover Network Resource....330
10.6.2 Changing the Devices of File systems Controlled by the Fsystem Resource....333
10.7 Adding file system to the shared disk by Dynamic Changing Configuration....334

Chapter 11 Changing the Operation Attributes of a Cluster System....338
11.1 Changing the Operation Attributes of a userApplication....338
11.1.1 Changing the Operation Attributes (CUI)....338
11.2 Changing the RMS Environment Variables....343
11.2.1 Changing Timeout Time during RMS Stop Processing....343
11.3 Changing Time to Detect Heartbeat Timeout....344
11.3.1 Changing Time to Detect CF Heartbeat Timeout....344
11.3.2 Changing Time to Detect RMS Heartbeat Timeout....344

Part 5 Maintenance....346

Chapter 12 Maintenance of the PRIMECLUSTER System....347
12.1 Maintenance Types....347
12.2 Maintenance Flow....347
12.2.1 Detaching Resources from Operation....347
12.2.2 Executing Standby Restoration for an Operating Job....348
12.3 Software Maintenance....348
12.3.1 Notes on Applying Corrections to the PRIMECLUSTER System....348
12.3.2 Overview of the Correction Application Procedure....348
12.3.2.1 Procedure for Applying Corrections by Stopping an Entire System....349
12.3.2.2 Procedure for Applying Correction by Rolling Update....350

Appendix A PRIMECLUSTER Products....352

Appendix B Manual Pages....353
B.1 CF....353
B.2 CIP....353
B.3 Operator Intervention....354
B.4 PAS....354
B.5 Cluster Resource Management Facility....354
B.6 RMS....354
B.7 Shutdown Facility (SF)....355

B.8 Tracing Failed Resource ... 355
B.9 SIS ... 355
B.10 Web-Based Admin View ... 356
B.11 Procedure Resource ... 356
B.12 RMS Wizards ... 356

Appendix C Troubleshooting ... 358
C.1 Collecting Troubleshooting Information ... 358
C.1.1 Executing the fjsnap or pclsnap Command ... 358
C.1.2 FJQSS (Information Collection Tool) ... 359
C.1.3 Crash Dump ... 360
C.1.4 PSA Information ... 360
C.1.5 SVmco Information ... 361
C.2 Detecting a Failed Resource ... 361
C.2.1 Failed Resource Message ... 362
C.2.2 Resource Fault History ... 363
C.2.3 Fault Resource List ... 365
C.3 PRIMECLUSTER Log Files ... 366
C.3.1 Output Destination for core Files ... 366
C.3.2 core File Configuration ... 367
C.3.2.1 core Files Output ... 367
C.3.2.2 Setting Output Destination for core Files ... 367
C.3.3 Log Volume When Changing Log Levels ... 368
C.3.4 Rotation and Deletion of RMS Log Files ... 368

Appendix D Registering, Changing, and Deleting State Transition Procedure Resources for PRIMECLUSTER Compatibility ... 369
D.1 Registering a Procedure Resource ... 369
D.2 Changing a Procedure Resource ... 370
D.2.1 Changing a state transition procedure ... 370
D.2.2 Changing the Startup Priority of a State Transition Procedure ... 370
D.2.3 Changing registration information of a procedure resource ... 371
D.3 Deleting a Procedure Resource ... 372

Appendix E Configuration Update Service for SA ... 374
E.1 Feature Description ... 374
E.2 Operation Environment ... 377
E.3 Configuration ... 378
E.3.1 Startup Configuration for the IPMI Service ... 378
E.3.2 Activating Configuration Update Service for SA ... 379
E.3.2.1 Startup Configuration for Update Service for SA ... 379
E.3.2.2 Checking the Configuration ... 379
E.3.2.3 Checking the BMC or iRMC IP Address and the Configuration Information of the Shutdown Agent ... 381
E.4 Operation Check ... 382
E.4.1 Operation Check by Restarting the System ... 382
E.5 Cancellation ... 383
E.5.1 Deactivating Configuration Update Service for SA ... 383
E.5.2 Restoring the Startup Configuration of the IPMI Service ... 383
E.6 Restoration ... 383
E.6.1 Restoration Method When Correct Information is not Distributed to All the Nodes ... 383
E.7 sfsacfgupdate ... 385
E.8 Output Message (Syslog) ... 386

Appendix F Using PRIMECLUSTER in a KVM Environment ... 389
F.1 Controlling and monitoring a guest OS by a cluster application on a host OS ... 389

Appendix G Using the Migration Function in a KVM Environment ... 390
G.1 Design ... 390

G.2 Prerequisites....390
G.2.1 Without using the Host OS failover function....390
G.2.2 Using the Host OS failover function....392
G.3 Operation....394
G.3.1 When performing Live Migration....395
G.3.1.1 When not using the Host OS failover function....395
G.3.1.1.1 Operations before Live Migration....395
G.3.1.1.2 Operations after Live Migration....395
G.3.1.2 When using the Host OS failover function....396
G.3.1.2.1 Operations before Live Migration....396
G.3.1.2.2 Operations after Live Migration....396
G.3.2 When performing Offline Migration....397
G.3.2.1 When not using the Host OS failover function....397
G.3.2.1.1 Operations before Offline Migration....397
G.3.2.1.2 Operations after Offline Migration....397
G.3.2.2 When using the Host OS failover function....398
G.3.2.2.1 Operations before Offline Migration....398
G.3.2.2.2 Operations after Offline Migration....398
G.3.3 When performing Migration by Export/Import....399
G.3.3.1 When not using the Host OS failover function....399
G.3.3.1.1 Operations before Migration by Export/Import....399
G.3.3.1.2 Operations after Migration by Export/Import....399
G.3.3.2 When using the Host OS failover function....399
G.3.3.2.1 Operation before Migration by Export/Import....399
G.3.3.2.2 Operation after Migration by Export/Import....399
G.4 Changing Settings....399
G.4.1 Canceling Prerequisites....399

Appendix H Using PRIMECLUSTER in a VMware Environment....401
H.1 Cluster Systems in a VMware Environment....401
H.2 Installation....406
H.2.1 Software Installation....406
H.2.1.1 Installation and Configuration of Related Software....407
H.2.1.2 Installation and Environment Configuration of Applications....411
H.2.2 Preparation Prior to Building a Cluster....411
H.2.3 Building a Cluster....411
H.2.3.1 Initial Setup of CF and CIP....411
H.2.3.2 Setting Up the Shutdown Facility (when using VMware vCenter Server Functional Cooperation)....411
H.2.3.3 Setting Up the Shutdown Facility (when using I/O fencing function)....414
H.2.3.4 Initial Setup of the Cluster Resource Management Facility....417
H.2.3.5 Setting Up Fault Resource Identification and Operator Intervention Request....417
H.2.4 Building Cluster Applications....417
H.2.4.1 Setting Up I/O Fencing Function....417
H.3 Operations....421
H.4 Changing the Configuration....421
H.5 Maintenance....421

Appendix I Startup Scripts and Startup Daemons, and Port Numbers in PRIMECLUSTER....422
I.1 Explanation Formats....422
I.2 Startup Script Lists....422
I.3 Necessary Daemons for PRIMECLUSTER to Operate....432

Appendix J Systemd Services and Startup Daemons, and Port Numbers in PRIMECLUSTER....433
J.1 Explanation Formats....433
J.2 systemd Service Lists....434
J.3 Necessary Services for PRIMECLUSTER to Operate....452

Appendix K Using Firewall....453

Appendix L Cloning the Cluster System Environment....455
L.1 Preparation....457
L.1.1 Backing up the GFS Configuration Information....457
L.1.2 Backing up the GDS Configuration Information....458
L.1.3 Canceling System Disk Mirroring....458
L.2 Copying System Image Using the Cloning Function....459
L.2.1 Copying Disk Data....459
L.2.2 Setting up System Disk Mirroring....459
L.3 Changing Cluster System Settings....460
L.3.1 Deleting the Setup Information for System Disk Mirroring....460
L.3.2 Setup in Single-User Mode....460
L.3.3 Changing the Settings in Multi-User Mode....467
L.3.4 Restoring the GDS Configuration Information....471
L.3.5 Restoring the GFS Configuration Information....473
L.3.6 Setting Up System Disk Mirroring....474
L.3.7 Changing the Settings of Cluster Application Information....474
L.3.7.1 When Using GLS....474
L.3.7.2 When Using the Takeover Network....476
L.3.7.3 When Using neither GLS nor the Takeover Network....478

Appendix M Changes in Each Version....480
M.1 Changes in PRIMECLUSTER 4.4A00 from 4.0A20....481
M.1.1 clgettree(1) command....482
M.1.2 hvdisp command....483
M.1.3 ciptool command....483
M.1.4 sdtool command....483
M.1.5 hvshut command....484
M.1.6 hvswitch command....484
M.1.7 hvdump command....485
M.1.8 Operator Intervention Request....485
M.1.9 Shutdown Facility....485
M.1.10 Setting Up Cluster Applications....486
M.1.11 Changing the Cluster Application....486
M.1.12 State of the Main CF Table....487
M.1.13 Node state....487
M.1.14 Operation Procedures and Displayed Items for Cluster Application Setup and Modification....488
M.1.15 Setting Up Fsystem Resources....493
M.1.16 Client Environment for Web-Based Admin View....493
M.1.17 Changes of the Behavior of CF Startup....494
M.1.18 HV_CONNECT_TIMEOUT....494
M.1.19 Changes of the ports used by RMS....494
M.1.20 Changes of the port number used by the shutdown facility....495
M.1.21 Changes of the target node to forcibly shut down when a heartbeat failure occurs....495
M.1.22 Display of the resource fault trace....495
M.1.23 Change of /etc/cip.cf file....496
M.1.24 Changes in CF over IP setting window of CF Wizard....496
M.1.25 Changes of the RMS message....496
M.1.26 Changes of the importance of the message in the RMS wizard....497
M.1.27 Changes of RMS console message....497
M.1.28 Changes of the response message for the operator intervention request....498
M.1.28.1 Message: 1421....498
M.1.28.2 Message: 1423....498
M.2 Changes in PRIMECLUSTER 4.4A00 from 4.1A20....499
M.2.1 clgettree(1) command....500
M.2.2 hvdisp command....500
M.2.3 ciptool command....501
M.2.4 sdtool command....501

M.2.5 hvshut command ... 501
M.2.6 hvswitch command ... 502
M.2.7 hvdump command ... 502
M.2.8 Posting Notification of a Resource Failure or Recovery ... 503
M.2.9 Operator Intervention Request ... 503
M.2.10 Node state ... 504
M.2.11 Operation Procedures and Displayed Items for Cluster Application Setup and Modification ... 504
M.2.12 Setting Up Fsystem Resources ... 509
M.2.13 Client Environment for Web-Based Admin View ... 510
M.2.14 Changes of the Behavior of CF Startup ... 510
M.2.15 HV_CONNECT_TIMEOUT ... 510
M.2.16 Changes of the ports used by RMS ... 511
M.2.17 Changes of the port number used by the shutdown facility ... 511
M.2.18 Changes of the target node to forcibly shut down when a heartbeat failure occurs ... 511
M.2.19 Display of the resource fault trace ... 512
M.2.20 Change of /etc/cip.cf file ... 512
M.2.21 Changes in CF over IP setting window of CF Wizard ... 512
M.2.22 Changes of the RMS message ... 513
M.2.23 Changes of the importance of the message in the RMS wizard ... 513
M.2.24 Changes of RMS console message ... 513
M.2.25 Changes of the response message for the operator intervention request ... 514
M.2.25.1 Message: 1421 ... 514
M.2.25.2 Message: 1423 ... 515
M.3 Changes in PRIMECLUSTER 4.4A00 from 4.1A30 ... 516
M.3.1 ciptool command ... 516
M.3.2 sdtool command ... 517
M.3.3 hvshut command ... 517
M.3.4 hvswitch command ... 518
M.3.5 hvdump command ... 518
M.3.6 Posting Notification of a Resource Failure or Recovery ... 518
M.3.7 Operator Intervention Request ... 519
M.3.8 Operation Procedures and Displayed Items for Cluster Application Setup and Modification ... 520
M.3.9 Setting Up Fsystem Resources ... 525
M.3.10 Client Environment for Web-Based Admin View ... 526
M.3.11 Changes of the Behavior of CF Startup ... 526
M.3.12 HV_CONNECT_TIMEOUT ... 526
M.3.13 Changes of the ports used by RMS ... 527
M.3.14 Changes of the port number used by the shutdown facility ... 527
M.3.15 Changes of the target node to forcibly shut down when a heartbeat failure occurs ... 527
M.3.16 Display of the resource fault trace ... 528
M.3.17 Change of /etc/cip.cf file ... 528
M.3.18 Changes in CF over IP setting window of CF Wizard ... 528
M.3.19 Changes of the RMS message ... 529
M.3.20 Changes of the importance of the message in the RMS wizard ... 529
M.3.21 Changes of RMS console message ... 529
M.3.22 Changes of the response message for the operator intervention request ... 530
M.3.22.1 Message: 1421 ... 530
M.3.22.2 Message: 1423 ... 531
M.4 Changes in PRIMECLUSTER 4.4A00 from 4.1A40 ... 531
M.4.1 sdtool command ... 532
M.4.2 hvshut command ... 533
M.4.3 hvswitch command ... 533
M.4.4 hvdump command ... 534
M.4.5 Posting Notification of a Resource Failure or Recovery ... 534
M.4.6 Operator Intervention Request ... 534
M.4.7 Setting Up Fsystem Resources ... 535
M.4.8 Client Environment for Web-Based Admin View ... 536

M.4.9 Changes of the Behavior of CF Startup ... 536
M.4.10 HV_CONNECT_TIMEOUT ... 536
M.4.11 Changes of the ports used by RMS ... 537
M.4.12 Changes of the port number used by the shutdown facility ... 537
M.4.13 Changes of the target node to forcibly shut down when a heartbeat failure occurs ... 537
M.4.14 Display of the resource fault trace ... 538
M.4.15 Change of /etc/cip.cf file ... 538
M.4.16 Changes in CF over IP setting window of CF Wizard ... 538
M.4.17 Changes of the RMS message ... 539
M.4.18 Changes of the importance of the message in the RMS wizard ... 539
M.4.19 Changes of RMS console message ... 539
M.4.20 Changes of the response message for the operator intervention request ... 540
M.4.20.1 Message: 1421 ... 540
M.4.20.2 Message: 1423 ... 541
M.5 Changes in PRIMECLUSTER 4.4A00 from 4.2A00 ... 542
M.5.1 sdtool command ... 542
M.5.2 hvshut command ... 543
M.5.3 hvswitch command ... 543
M.5.4 hvdump command ... 544
M.5.5 Posting Notification of a Resource Failure or Recovery ... 544
M.5.6 Operator Intervention Request ... 545
M.5.7 Setting Up Fsystem Resources ... 545
M.5.8 Client Environment for Web-Based Admin View ... 546
M.5.9 Changes of the Behavior of CF Startup ... 546
M.5.10 HV_CONNECT_TIMEOUT ... 547
M.5.11 Changes of the ports used by RMS ... 547
M.5.12 Configuring the IPMI Shutdown Agent ... 547
M.5.13 Changes of the port number used by the shutdown facility ... 548
M.5.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs ... 548
M.5.15 Display of the resource fault trace ... 548
M.5.16 Change of /etc/cip.cf file ... 549
M.5.17 Changes in CF over IP setting window of CF Wizard ... 549
M.5.18 Changes of the RMS message ... 549
M.5.19 Changes of the importance of the message in the RMS wizard ... 550
M.5.20 Changes of RMS console message ... 550
M.5.21 Changes of the response message for the operator intervention request ... 551
M.5.21.1 Message: 1421 ... 551
M.5.21.2 Message: 1423 ... 551
M.6 Changes in PRIMECLUSTER 4.4A00 from 4.2A30 ... 552
M.6.1 sdtool command ... 553
M.6.2 hvshut command ... 553
M.6.3 hvswitch command ... 554
M.6.4 hvdump command ... 554
M.6.5 Posting Notification of a Resource Failure or Recovery ... 555
M.6.6 Operator Intervention Request ... 555
M.6.7 Setting Up Fsystem Resources ... 556
M.6.8 Client Environment for Web-Based Admin View ... 556
M.6.9 Changes of the Behavior of CF Startup ... 557
M.6.10 HV_CONNECT_TIMEOUT ... 557
M.6.11 Changes of the ports used by RMS ... 557
M.6.12 Configuring the IPMI Shutdown Agent ... 558
M.6.13 Changes of the port number used by the shutdown facility ... 558
M.6.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs ... 559
M.6.15 Display of the resource fault trace ... 559
M.6.16 Change of /etc/cip.cf file ... 559
M.6.17 Changes in CF over IP setting window of CF Wizard ... 560
M.6.18 Changes of the RMS message ... 560

M.6.19 Changes of the importance of the message in the RMS wizard....560
M.6.20 Changes of RMS console message....561
M.6.21 Changes of the response message for the operator intervention request....561
M.6.21.1 Message: 1421....561
M.6.21.2 Message: 1423....562
M.7 Changes in PRIMECLUSTER 4.4A00 from 4.3A00....563
M.7.1 sdtool command....564
M.7.2 hvshut command....564
M.7.3 hvswitch command....564
M.7.4 hvdump command....565
M.7.5 Posting Notification of a Resource Failure or Recovery....565
M.7.6 Operator Intervention Request....566
M.7.7 Setting Up Fsystem Resources....566
M.7.8 Client Environment for Web-Based Admin View....567
M.7.9 Changes of the Behavior of CF Startup....567
M.7.10 HV_CONNECT_TIMEOUT....567
M.7.11 Changes of the ports used by RMS....568
M.7.12 Configuring the IPMI Shutdown Agent....568
M.7.13 Changes of the port number used by the shutdown facility....569
M.7.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs....569
M.7.15 Display of the resource fault trace....569
M.7.16 Change of /etc/cip.cf file....570
M.7.17 Changes in CF over IP setting window of CF Wizard....570
M.7.18 Changes of the RMS message....570
M.7.19 Changes of the importance of the message in the RMS wizard....571
M.7.20 Changes of RMS console message....571
M.7.21 Changes of the response message for the operator intervention request....572
M.7.21.1 Message: 1421....572
M.7.21.2 Message: 1423....572
M.8 Changes in PRIMECLUSTER 4.4A00 from 4.3A10....573
M.8.1 sdtool command....574
M.8.2 hvshut command....574
M.8.3 hvswitch command....575
M.8.4 hvdump command....575
M.8.5 Posting Notification of a Resource Failure or Recovery....576
M.8.6 Operator Intervention Request....576
M.8.7 Setting Up Fsystem Resources....577
M.8.8 Changes of the ports used by RMS....577
M.8.9 Configuring the IPMI Shutdown Agent....577
M.8.10 Changes of the port number used by the shutdown facility....578
M.8.11 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment....578
M.8.12 Changes of the target node to forcibly shut down when a heartbeat failure occurs....579
M.8.13 Display of the resource fault trace....579
M.8.14 Change of /etc/cip.cf file....579
M.8.15 Changes in CF over IP setting window of CF Wizard....580
M.8.16 Changes of RMS console message....580
M.8.17 Changes of the response message for the operator intervention request....581
M.8.17.1 Message: 1421....581
M.8.17.2 Message: 1423....581
M.9 Changes in PRIMECLUSTER 4.4A00 from 4.3A20....582
M.9.1 hvshut command....583
M.9.2 hvswitch command....583
M.9.3 hvdump command....584
M.9.4 Posting Notification of a Resource Failure or Recovery....584
M.9.5 Operator intervention request....585
M.9.6 Setting Up Fsystem Resources....585
M.9.7 Configuring the IPMI Shutdown Agent....586

M.9.8 Changes of the port number used by the shutdown facility....586
M.9.9 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment....586
M.9.10 Changes of the target node to forcibly shut down when a heartbeat failure occurs....587
M.9.11 Display of the resource fault trace....587
M.9.12 Change of /etc/cip.cf file....588
M.9.13 Changes in CF over IP setting window of CF Wizard....588
M.9.14 Changes of RMS console message....588
M.9.15 Changes of the response message for the operator intervention request....589
M.9.15.1 Message: 1421....589
M.9.15.2 Message: 1423....590
M.10 Changes in PRIMECLUSTER 4.4A00 from 4.3A30....590
M.10.1 hvdump command....591
M.10.2 Posting Notification of a Resource Failure or Recovery....591
M.10.3 Operator intervention request....592
M.10.4 Setting Up Fsystem Resources....592
M.10.5 Setting up the Host OS failover function when using it in KVM environment....592
M.10.6 Display of the resource fault trace....593
M.10.7 Change of /etc/cip.cf file....593
M.10.8 Changes in CF over IP setting window of CF Wizard....593
M.11 Changes in PRIMECLUSTER 4.4A00 from 4.3A40....594
M.11.1 Setting up the Host OS failover function when using it in KVM environment....594
M.11.2 Changes in CF over IP setting window of CF Wizard....594
M.11.3 Setting up the migration function when using it in KVM environment....595

Appendix N Release Information....596

Glossary....597

Index....611

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.

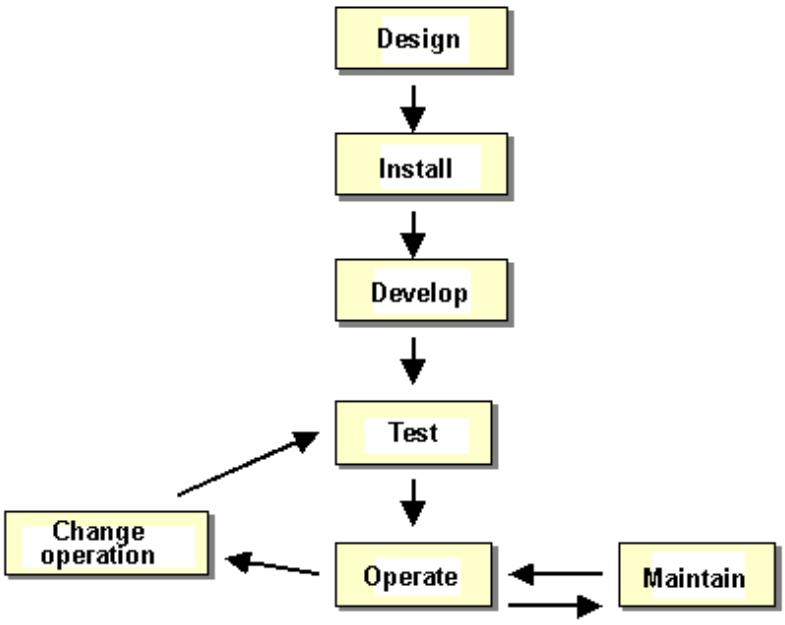
Chapter 1 Build Flow..... 2

Chapter 2 Site Preparation..... 14

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 and whether to use the virtual machine function. This includes selecting the applications to be used and determining the required hardware resources, such as the number of hosts, networks, the number of cluster interconnect paths, and disk size.

   Up to 16 nodes can be added to one cluster system.

   We recommend that you use 2 or more cluster interconnects.

   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."

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) and the takeover address.
   - Determine whether a user-defined RMS configuration script is to be used. Determine whether there are other items to be used as resources.
   - 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 Determining 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 PRIMECLUSTER Designsheets that were created, check that all items on PRIMECLUSTER Designsheets have been entered.

Information

PRIMECLUSTER Designsheets are stored in documents/designsheet in PRIMECLUSTER DVD.

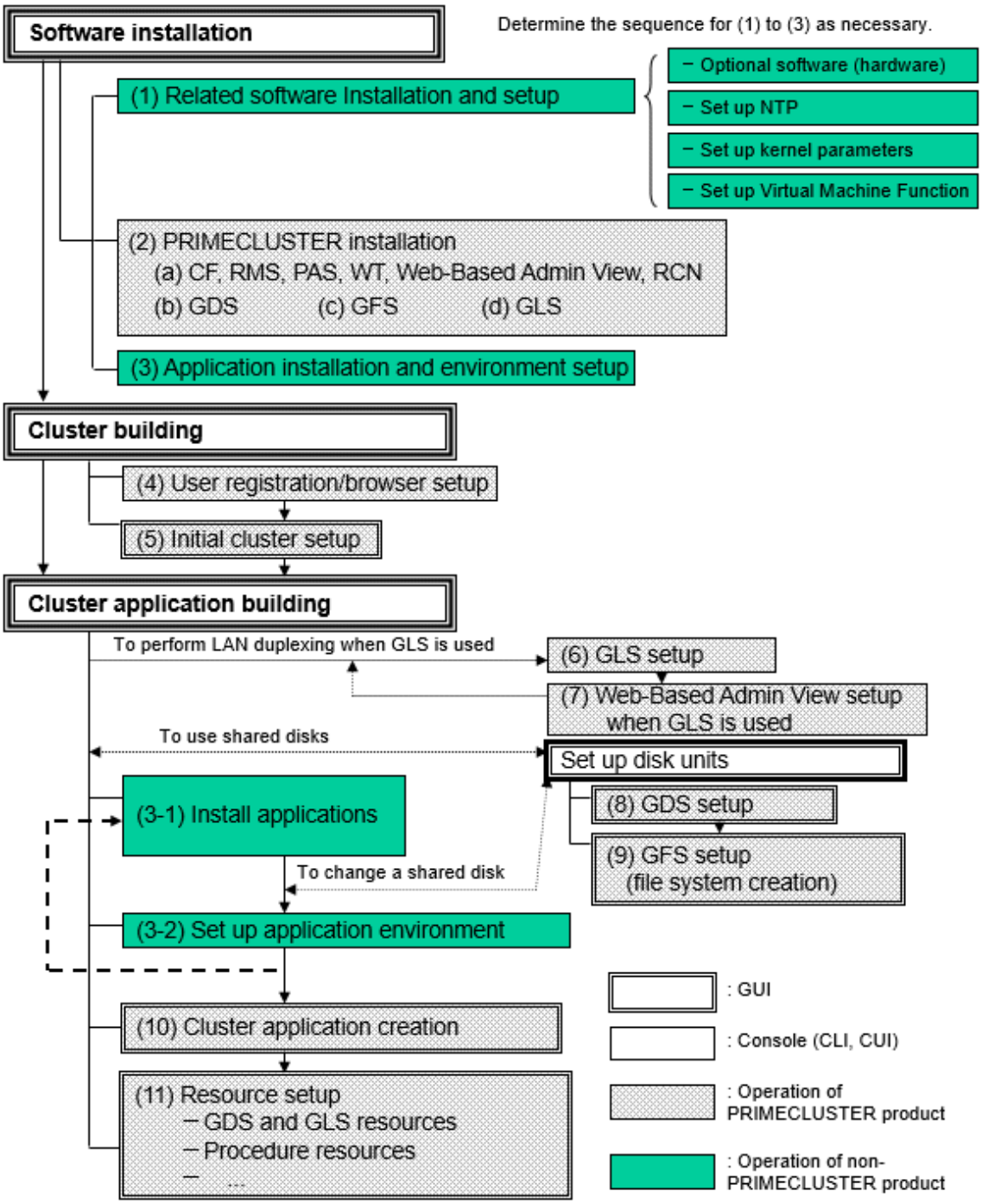
Install the PRIMECLUSTER system by performing the following procedure in sequence from (1).

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

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) Related software Installation and setup	3.1 When Not Using the Virtual Machine Function 3.2 When Using the Virtual Machine Function	Optional
(2) PRIMECLUSTER installation	3.3 PRIMECLUSTER Installation	Required
(3) Application installation and environment setup	3.4 Installation and Environment Setup of Applications	Optional
Cluster building		
(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
(7) Web-Based Admin View setup when GLS is used		Note that it is required when IP address takeover and redundant line control are used.
(8) GDS setup	6.3 GDS Configuration Setup	Optional (required to use GDS)
(9) GFS setup (file system creation)	6.4 Initial GFS Setup	Optional (required to use GFS)
(10) Cluster application creation	6.7 Setting Up Cluster Applications	Required
(11) Resource setup	6.7.3 Setting Up Resources	Optional

GLS: Global Link Services
GDS: Global Disk Services
GFS: Global File 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

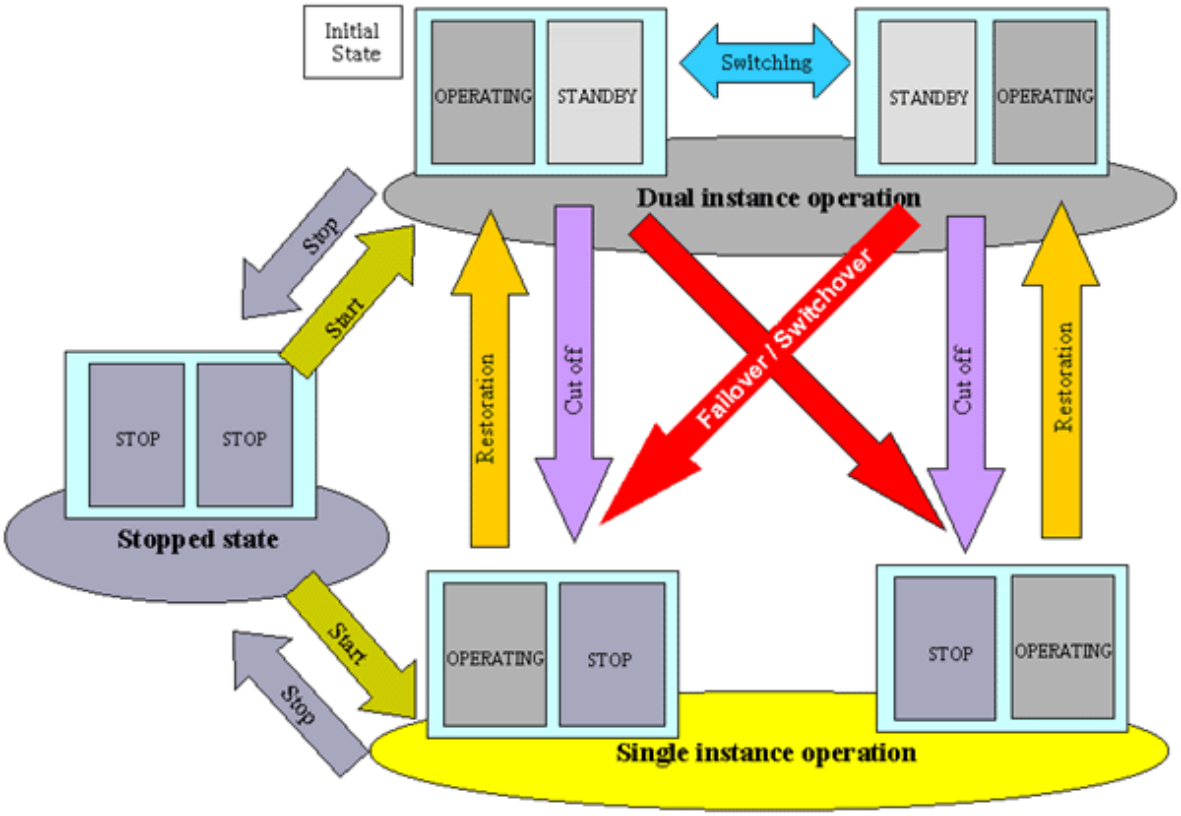
Purpose

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. 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.

State	Description
	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

* It is displayed when referring to the stopped (STOP) cluster application in the status icon of the rms tab in 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 Setting Up userApplication."

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 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(1)" 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 "ip(8)" command or the "ifconfig(8)" command.
- 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(1)" 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 the 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.

Test for forced shutdown of cluster nodes

Check that the settings of the shutdown facility work correctly.

Conduct a test to check that every node in the cluster is shut down at least once with the following viewpoints:

- Induce an OS error to check that the cluster node in which a failure has occurred is forcibly shut down.
- Disconnect the cluster interconnect to check that the cluster node with the lowest priority is forcibly shut down.

Note

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

For example, if the both nodes are connected through a switch instead of being connected directly, 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.

In addition, check that crash dumps for the cluster node which has been forcibly shut down are collected.

See

- For information on the operation procedures for start, clear fault, failover, switchover, and stop, see "7.2 Operating the PRIMECLUSTER System."
- For information on IP address takeover, see "6.7.3.5 Setting Up Gls 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 instance 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 "Chapter 8 Changing the Cluster System Configuration."

1.7 Notes When Building a System

1.7.1 Common Notes on PRIMERGY/PRIMEQUEST

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 the nodes to configure a cluster system

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

For example, if the following messages are output or the OnlinePriority attribute of the cluster application is set, the cluster application may not become Online on the intended node because the last online node cannot be correctly recognized at RMS startup.

(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.

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.

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.

For details on a parameter value, see "Setup (initial configuration)" of PRIMECLUSTER Designsheets.

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.

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.

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 Setting up 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 "1.1.2 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."

Do not use DHCP when configuring CF

A node may be panicked if configuring CF while DHCP is set in the network interface.

Before configuring CF, unset DHCP in all network interfaces on nodes.

Example

When DHCP setting is being set

<Contents of /etc/sysconfig/network-scripts/ifcfg-ethX>

```
DEVICE=ethX
BOOTPROTO=dhcp
HWADDR=xx:xx:xx:xx:xx:xx
ONBOOT=yes
TYPE=Ethernet
DHCP_HOSTNAME=Node1
```

When DHCP setting was unset

<Contents of /etc/sysconfig/network-scripts/ifcfg-ethX>

```
DEVICE=ethX
BOOTPROTO=static
HWADDR=xx:xx:xx:xx:xx:xx
ONBOOT=yes
IPADDR=xxx.xxx.xxx.xxx
NETMASK=xxx.xxx.xxx.x
TYPE=Ethernet
```

To use iptables or ip6tables as Firewall in a cluster node, see "Appendix K Using Firewall."

If Firewall is not set correctly, PRIMECLUSTER may not operate properly.

Do not disable an IPv6 function of the operating system.

Even if you do not use an IPv6 address, set the IPv6 module of the operating system to be loaded.

1.7.2 Notes on PRIMERGY

BMC, iRMC and cluster node must belong to the same network segment.

If they do not belong to the same network segment, the IPMI shutdown agent does not operate properly.

To use the IPMI shutdown agent or the BLADE shutdown agent, also set the kdump shutdown agent.

If the kdump shutdown agent is not set, a node is forcibly stopped without collecting panic dumps.

The kdump shutdown agent is set with the panicinfo_setup command.

LAN interfaces of BMC and iRMC are exclusive to LAN.

LAN interfaces of BMC and iRMC used in the Shutdown Facility are exclusive to LAN. They cannot be used with the administrative LAN or cluster interconnects.

When using the IPMI shutdown agent, assign the iRMC user to the Administrator permission group.

Without the administrator authority, the IPMI shutdown agent will not work correctly.

1.7.3 Notes on PRIMEQUEST

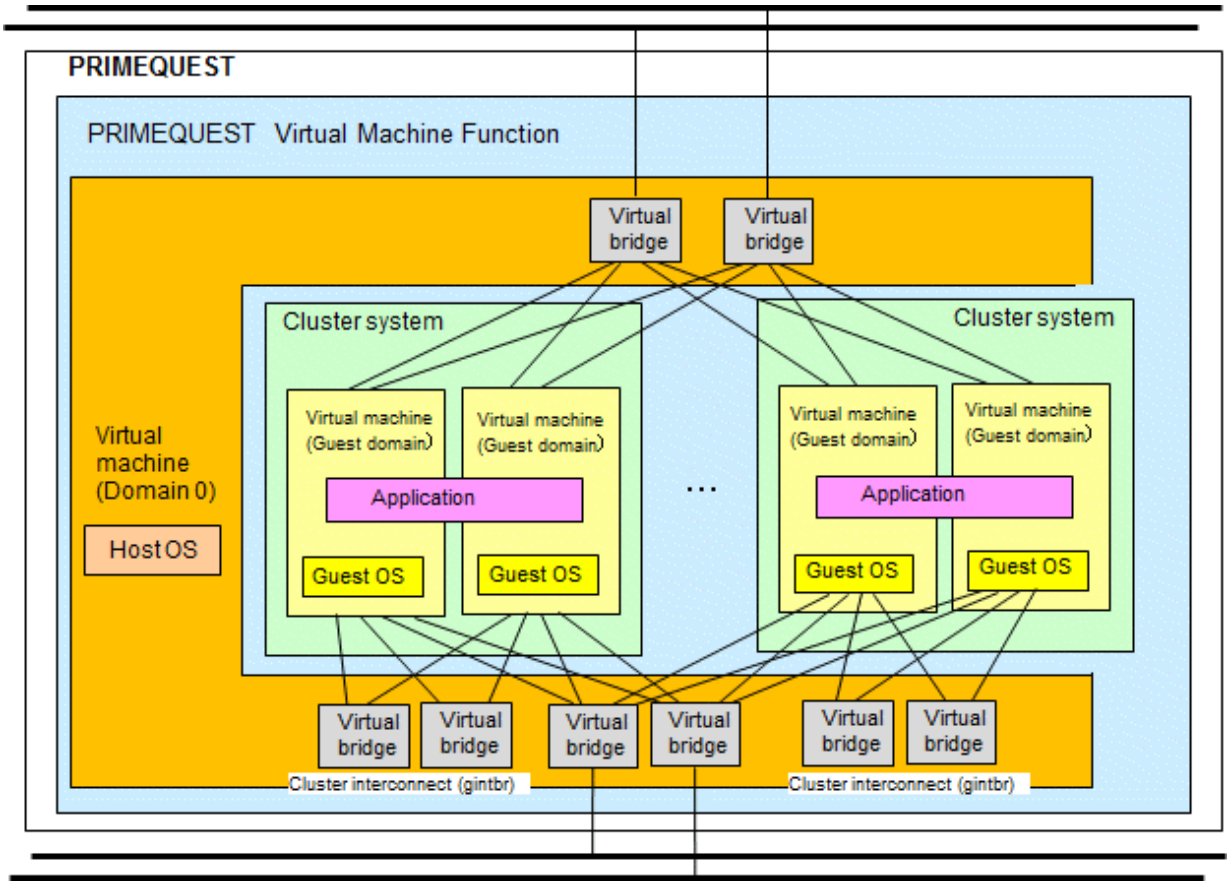
Install PSA/SVmco to each partition

PSA (PRIMEQUEST Server Agent)/SVmco (ServerView Mission Critical Option) needs to be installed to each partition.
If PSA/SVmco has not been installed, a node (partition) cannot be switched to other one properly when a failure occurs.

1.7.4 Notes When Building a Cluster System Using a Virtual Machine Function

To build multiple cluster systems, each cluster system needs its own virtual bridge for cluster interconnects

On the virtual machine function, multiple cluster systems can be built as follows.



To build multiple cluster systems, note the following points:

- For cluster interconnects, use a virtual bridge for each cluster system.
- Use a common virtual bridge for the administrative LAN.

For a virtual bridge used for the administrative LAN, determine whether or not to distinguish cluster systems based on the communication volume used in the operation. The virtual bridge can be distinguished based on the communication volume.

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 "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.

- **PRIMECLUSTER Clustering Base (CB)**

  Cluster foundation software that realizes a switchover-type cluster system.

  This product is only for PRIMERGY.

- **PRIMECLUSTER Lite Pack (LP)**

  Cluster foundation software that runs in a two-node configuration only and on specific models only. It provides superior cost performance, supporting switching (HA) cluster and volume management through the operation mode, 1:1 standby and mutual standby.

  This product is only for PRIMERGY.

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

Components		Products			
Names	Features	EE	HA	CB	LP
PCLsnap	Refers to the function that collects information on a system or cluster that is needed to investigate the failures.	Y	Y	Y	Y
Web-Based Admin View	Refers to the function for realizing PRIMECLUSTER operations and monitoring with the GUI (management view).	Y	Y	Y	Y
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	Y	Y
Reliant Monitor Services (RMS)	Refers to the software monitoring function that is used to realize high-availability (HA) of the application that is to be executed within the cluster.	Y	Y	Y	Y
Wizard Tools	Refers to the function that is used to create an application that is to be controlled with RMS.	Y	Y	Y	Y
RAO	Refers to the function that is used to manage resources that run on PRIMECLUSTER.	Y	Y	Y	Y
SA	Refers to the shutdown agent function for which BMC, iRMC, Blade, and MMB are used.	Y	Y	Y	Y
Global Link Services (hereinafter GLS)	Provides highly reliable transmission routes by setting up redundant network.	Y	Y	-	-
Global File Services (hereinafter 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.	Y	Y	-	-
Global Disk Services (hereinafter GDS)	Refers to the volume management function that is used to improve the availability and manageability of the data stored on the shared disk device.	Y	Y	-	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 components" in "PRIMECLUSTER Concepts Guide."

2.2 System Design

You can use the following configuration of the cluster system. Use PRIMECLUSTER Designsheets 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 created PRIMECLUSTER Designsheets. Make sure to create the designsheet.

See

For details on the operation environment, see "Chapter 2 Operation Environment" in the Installation Guide for PRIMECLUSTER.

Information

- When designing the system without using the virtual machine function, see PRIMECLUSTER Designsheets.
- When using the virtual machine function in a VMware environment, see "Appendix H Using PRIMECLUSTER in a VMware Environment."
- When using PRIMECLUSTER on FUJITSU Cloud Service K5, see "PRIMECLUSTER Installation and Administration Guide 4.4 FUJITSU Cloud Service K5."

Note

- Do not set the name: cipX (X is a number from 0 to 7) for the device name of the network device that exists in the system. Since PRIMECLUSTER creates and uses the name: cipX of the virtual network device, if the name has already existed in the network device, PRIMECLUSTER cannot be set nor operated.
- When the NetworkManager service is enabled, PRIMECLUSTER cannot be set nor operated. Check the NetworkManager service setting. If you change the NetworkManager service setting, see the Linux documentation.

2.2.1 Virtual Machine Function

The virtual machine function is a function for the realization of virtual machine environment.

This function enables you to create multiple independent virtual machines on one physical machine by virtualizing resources such as the CPU, memory, network and disk that are installed on a physical machine.

PRIMECLUSTER can build the cluster system between multiple guest OSes.

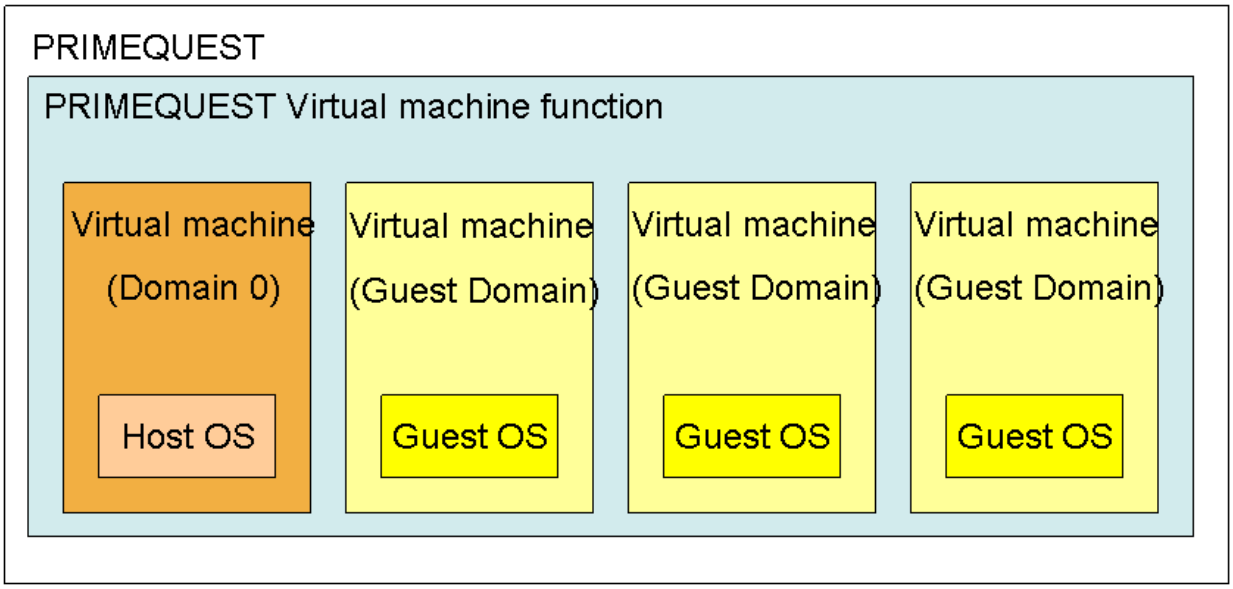
Note

- In a KVM environment, read the "host OS" as "hypervisor," in a VMware environment, read "host OS" as "ESXi host."
- When installing PRIMECLUSTER in a virtual machine environment, do not perform the following procedures:
  - Temporary stopping the Guest OS
  - Restart the Guest OS from a temporary stopped state
  - Restart or stop of the host OS when the guest OS is not stopped

See

- For details on the virtual machine function in a KVM environment, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."
- For details on the virtual machine function in a VMware environment, see the documentation for VMware.

Virtual machine function configuration



Cluster system in the virtual machine function

The virtual machine function provides the following methods to build a cluster system:

- Building a cluster system between guest OSes on one host OS
- Building a cluster system between guest OSes on multiple host OSes without the Host OS failover function
- Building a cluster system between guest OSes on multiple host OSes with the Host OS failover function (only in a KVM environment)

The table below shows uses and notes for each cluster system.

Method	Use	Note

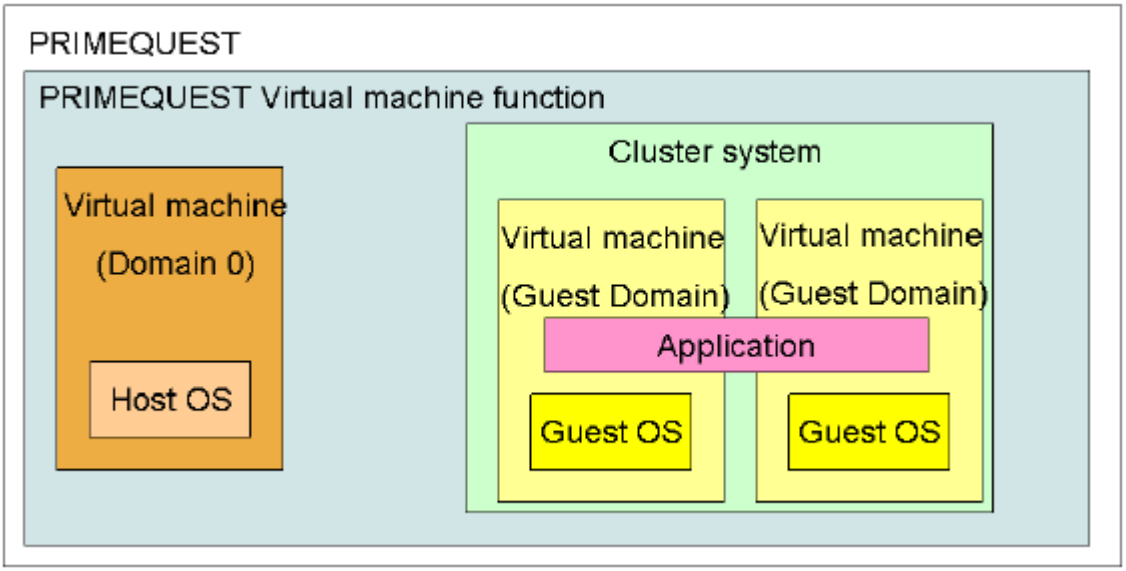
Building a cluster system between guest OSes on one host OS	- In a single physical server, build a cluster environment between guest OSes with the same cluster application configuration as that of the physical environment.	- Since this cluster system is built with a single physical server, all cluster nodes will be shut down and operations will be suspended when the physical server fails. Therefore, this is not suitable to operate services.
		- Do not perform the following procedures.
		  - Temporary stopping the Guest OS.
		  - Restart the Guest OS from a temporary stopped state.
		  - Restart or stop the host OS without stopping the Guest OS.

Building a cluster system between guest OSes on multiple host OSes without the Host OS failover function	- To build a cluster environment between guest OSes with the same cluster application configuration as that of the physical environment and to use it as a development and test environment for cluster applications, or operate services	- Do not install PRIMECLUSTER on the host OS. - If the host OS fails in a KVM environment, the node becomes the LEFTCLUSTER state because guest OS cannot be forcibly shut down. - Do not perform the following procedures. - Temporary stopping the Guest OS.

Method	Use	Note
		- Restart the Guest OS from a temporary stopped state. - Restart or stop the host OS without stopping the Guest OS.
Building a cluster system between guest OSes on multiple host OSes with the Host OS failover function (only in a KVM environment)	- To build a cluster environment between guest OSes with the same cluster application configuration as that of the physical environment and operate services - To switch the cluster application when the host OS fails	- The ShutdownPriority attribute of RMS cannot be set on guest OSes. - The active node may be stopped depending on the settings of survival priority. - When a host OS failure is detected, the host OS is forcibly shut down. Then, all guest OSes on the failed host OS will stop. - Do not perform the following procedures. - Temporary stopping the Guest OS. - Restart the Guest OS from a temporary stopped state. - Restart or stop the host OS without stopping the Guest OS.

When building a cluster system between guest OSes on one host OS

This configuration allows you to run a cluster system on a physical machine. This is effective when you verify that the userApplication runs on PRIMECLUSTER.



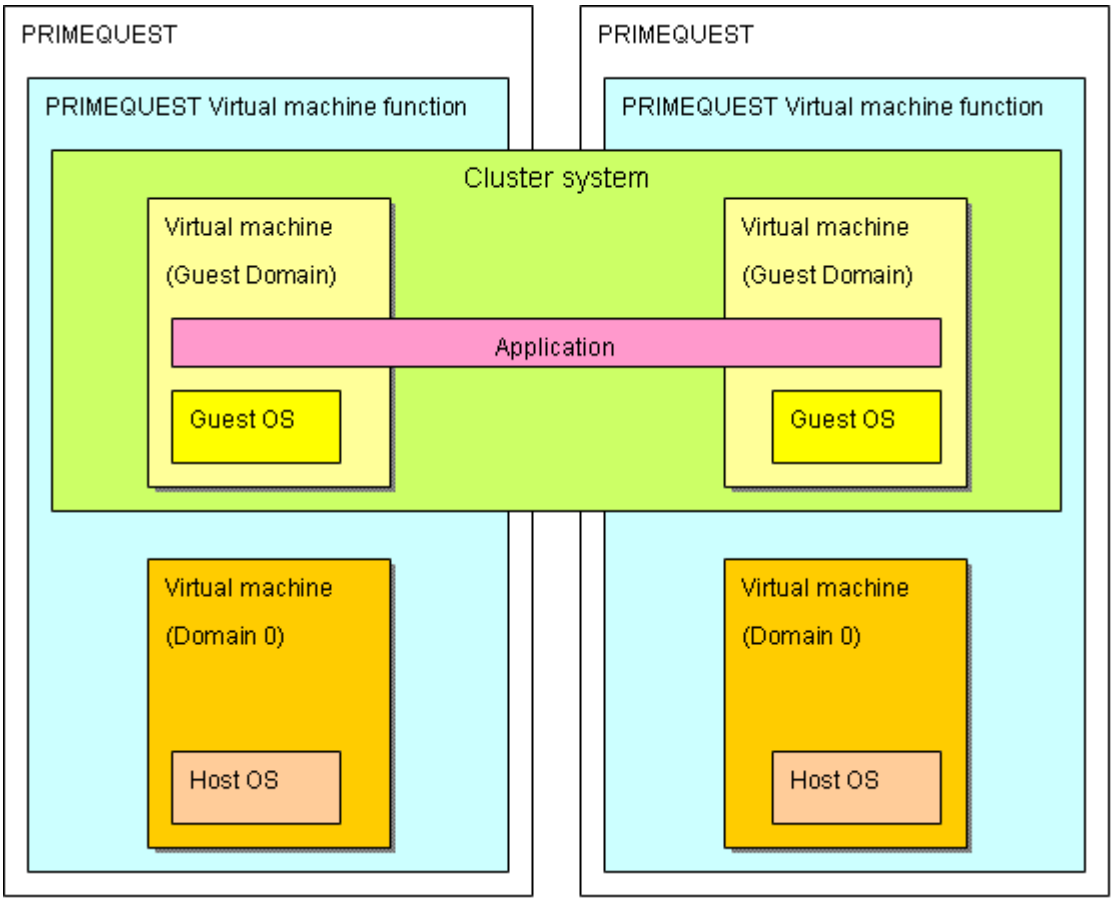
Note

When an error occurs in the guest OS in a VMware environment, the node state becomes LEFTCLUSTER.

For how to recover from LEFTCLUSTER, refer to "5.2 Recovering from LEFTCLUSTER" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide." For the following operations, refer to "7.2 Operating the PRIMECLUSTER System."

When building a cluster system between guest OSes on multiple host OSes

This configuration allows you to continue work processes by a failover even if hardware such as a network or a disk fails.



Note

If the host OS cannot run in a KVM environment, the node may become the LEFTCLUSTER state. For details, see "7.4.2 Corrective Action in the event of the LEFTCLUSTER state when the virtual machine function is used" or "7.2 Operating the PRIMECLUSTER System."

When building a cluster system between guests on multiple host OSes in a KVM environment, you can use a function that automatically perform a failover when the host OS fails (Host OS failover function).

Host OS failover function

When building a cluster between guests in different units on a virtual machine, if an error occurs in the host OS, nodes in the cluster may become the LEFTCLUSTER state. Host OS failover function allows for automatically switching cluster applications on the guest OSes in the case of the following errors in a cluster system between guests in different units in a KVM environment.

- Panic of the host OS

- Hang-up of the host OS (slowdown)

This function is achieved by linking PRIMECLUSTER installed on the host OS with guest OSes.

Note that there are some precautions for operations, for example, setting the priority of RMS is not available by using this function. Then, you should take these precautions into consideration when designing the system.

Note

- When creating a cluster application for a guest OS, do not set the ShutdownPriority attribute of RMS.

- When a host OS failure is detected, the host OS is forcibly shut down. Then, all guest OSes on that host OS with a failure will stop regardless of whether they are clusters or not.
- Do not register resources (except the following resources necessary on the guest OS) in the cluster application.
  - Gls resource which controls the network used on the guest OS
  - Cmdline resource to control the guest OS (see "Appendix F Using PRIMECLUSTER in a KVM Environment")

  If the operation was performed on the host OS and it was overloaded, the host OS is forcibly shut down and it affects the guest OS running on the host OS.

Figure 2.1 Cluster system using the Host OS failover function on the virtual machine

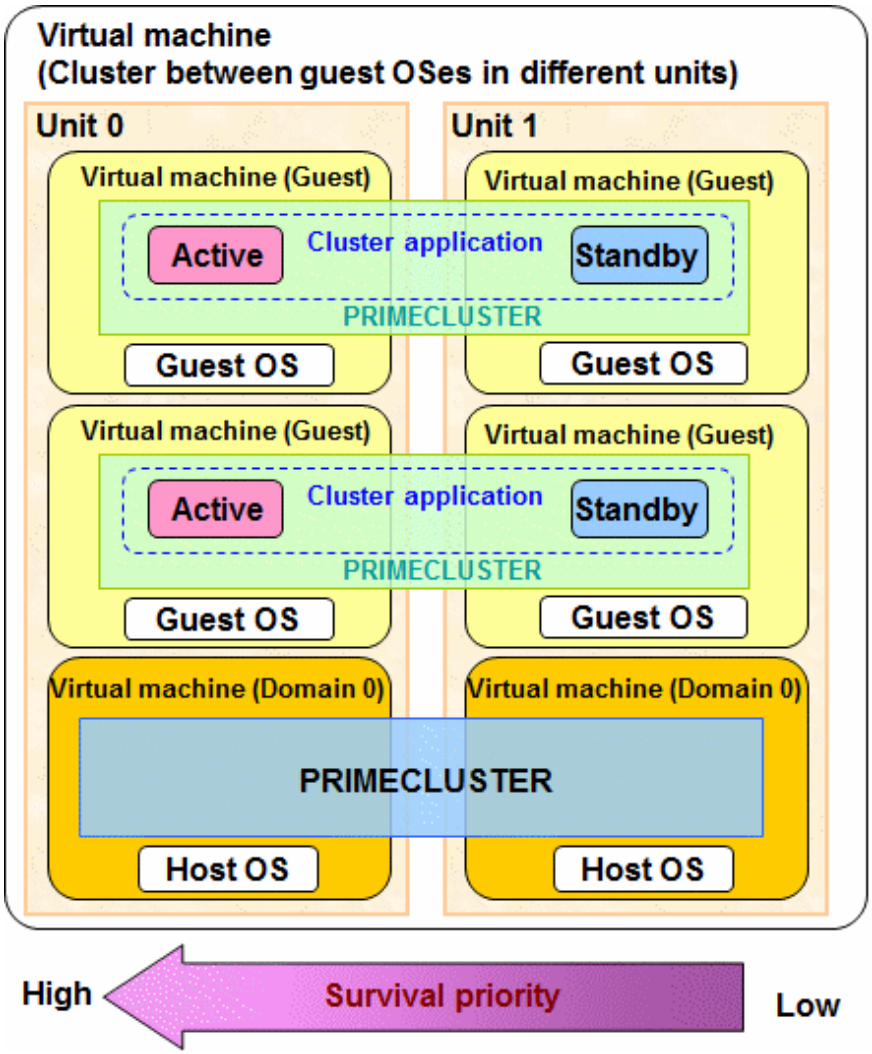
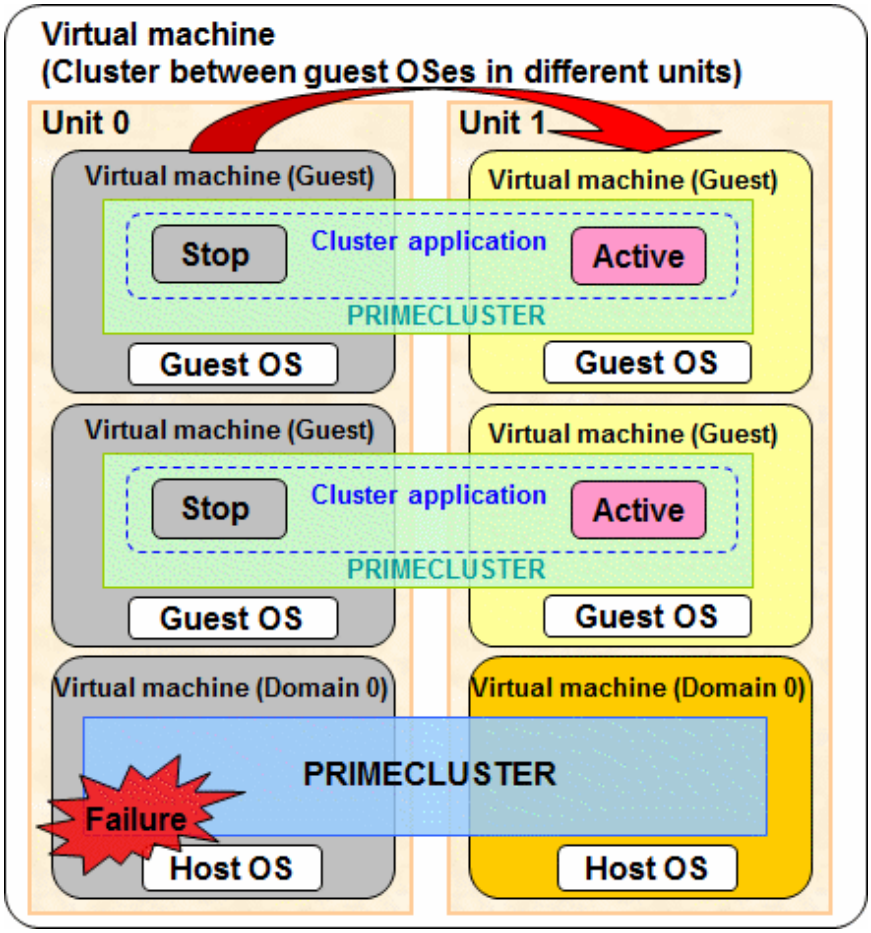


Figure 2.2 Failover image in the case of host OS failure



Moreover, you can replicate the cluster system by doing live migration of guest OSes in which PRIMECLUSTER is installed or by copying the virtual machine image.

Migration for a Cluster System in a KVM Environment

Following three types of the Migration function can be used for a cluster system in a KVM environment:

- Live Migration

  Transferring an active guest OS.

- Offline Migration

  Transferring a suspended guest OS.

- Migration by Export/Import

  Exporting/Importing the XML setup files of stopped guest OSes.

The Migration function in a KVM environment can be used in the following cluster system configurations:

- When building a cluster system between guest OSes on multiple host OSes without using the Host OS failover function
- When building a cluster system between guest OSes on multiple host OSes using the Host OS failover function

- Live Migration

By migrating a guest OS while it is running (Live Migration), you can do server maintenance while maintaining the redundant configuration for active and standby servers.

Figure 2.3 Live Migration for a cluster system

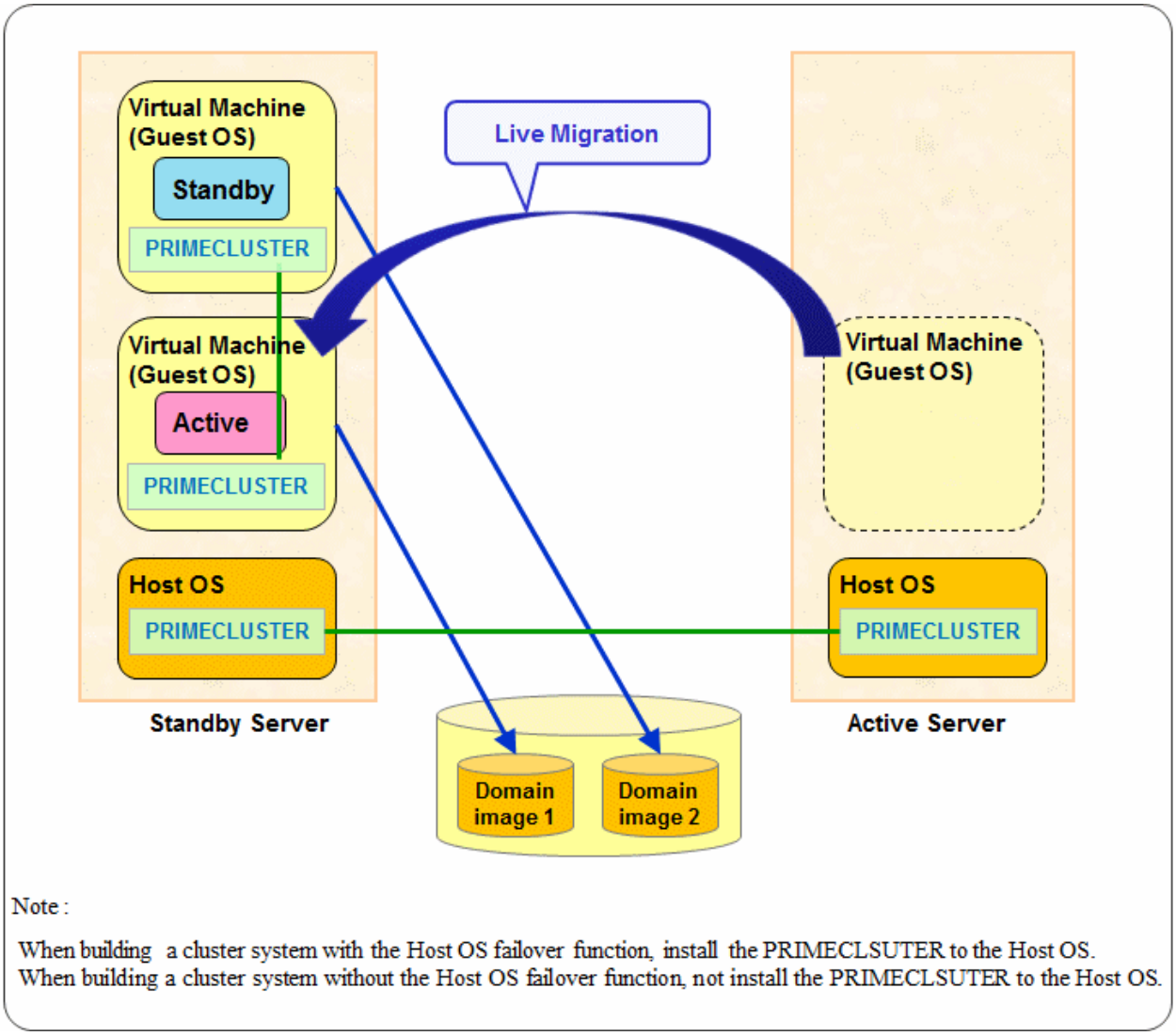
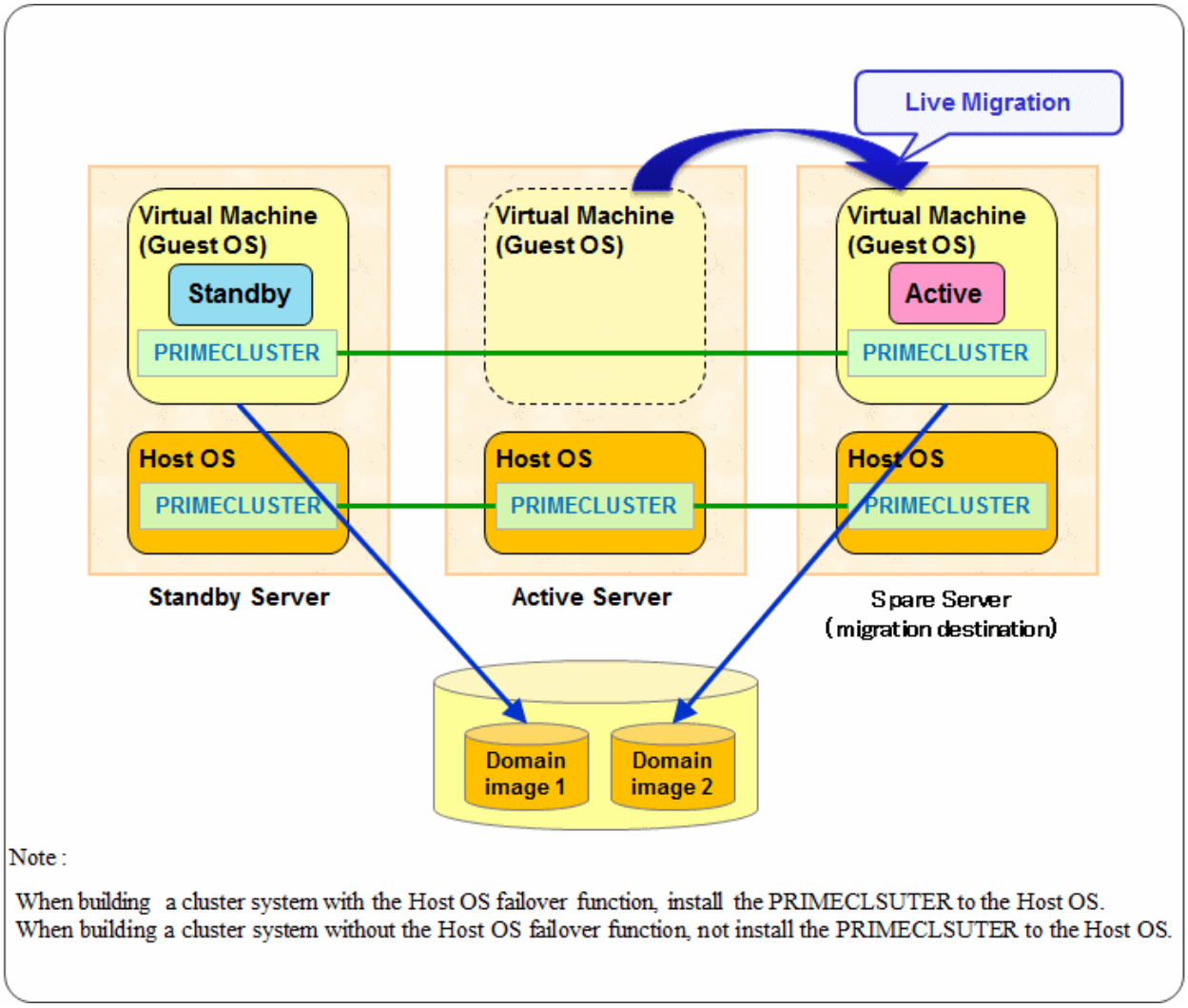


Figure 2.4 Live Migration to a spare server (before performing)



- Offline Migration

  By migrating a suspended guest OS (Offline Migration), you can do standby server maintenance while maintaining the redundant configuration for active and standby servers.

Figure 2.5 Offline Migration to a spare server (before performing)

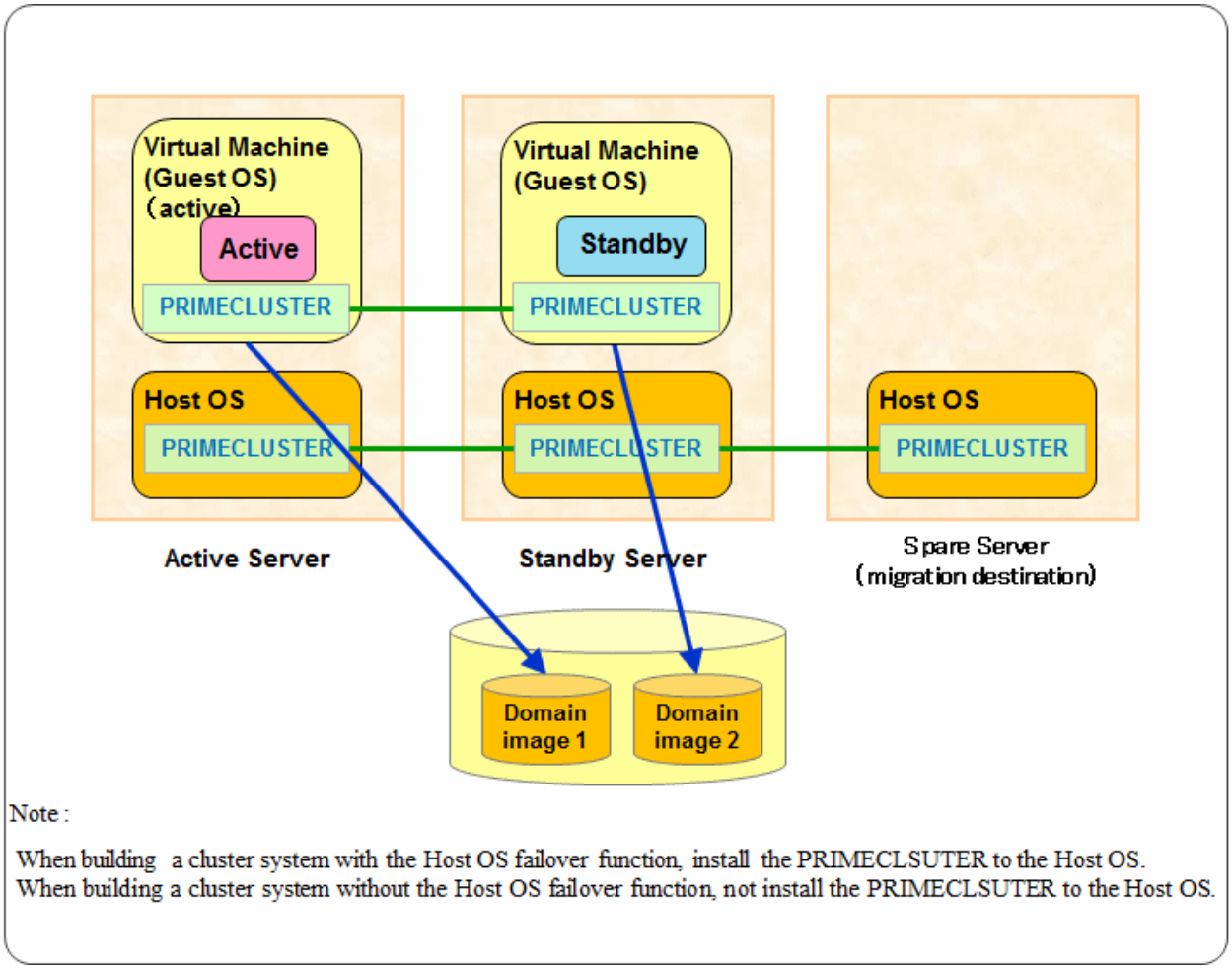
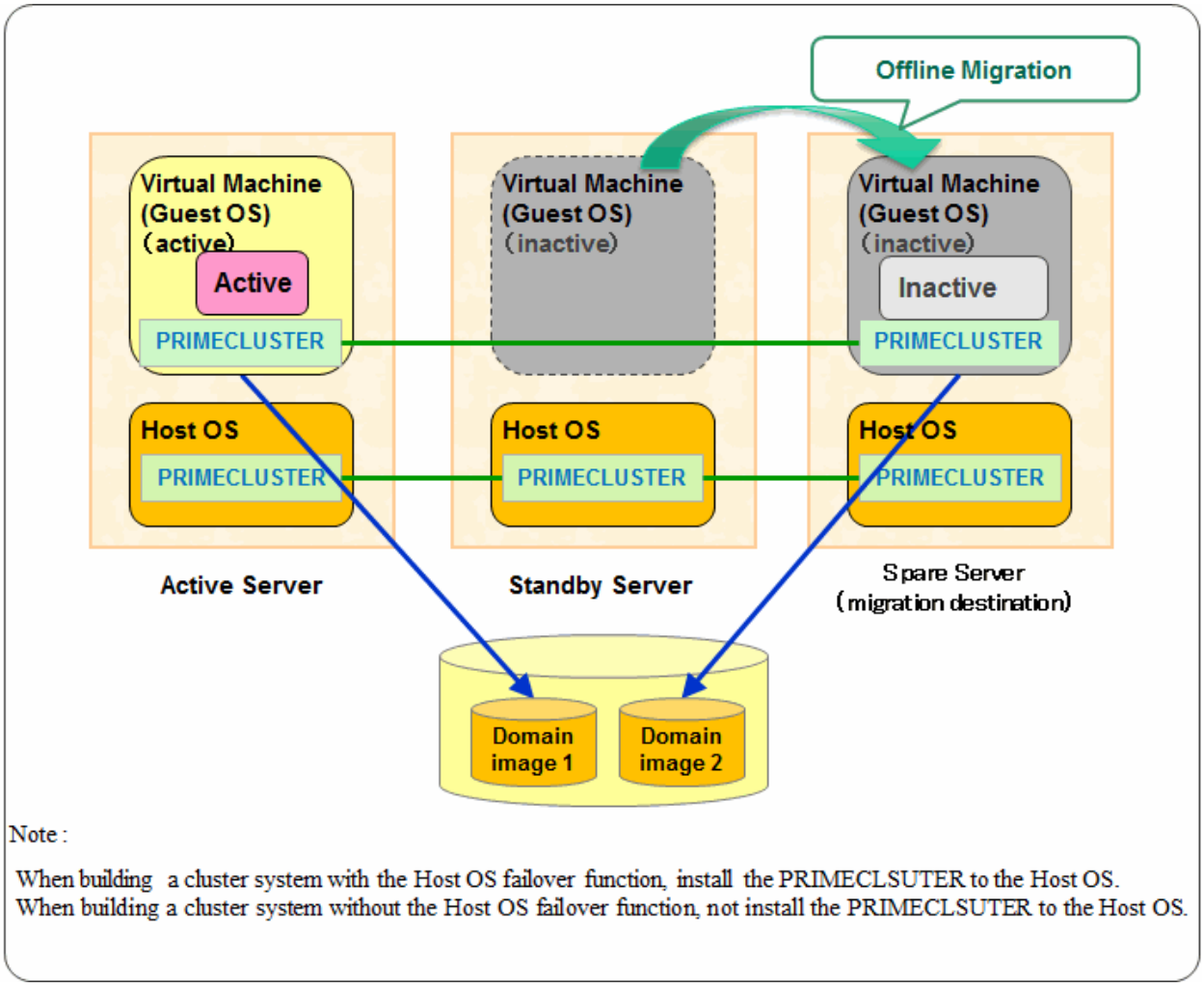


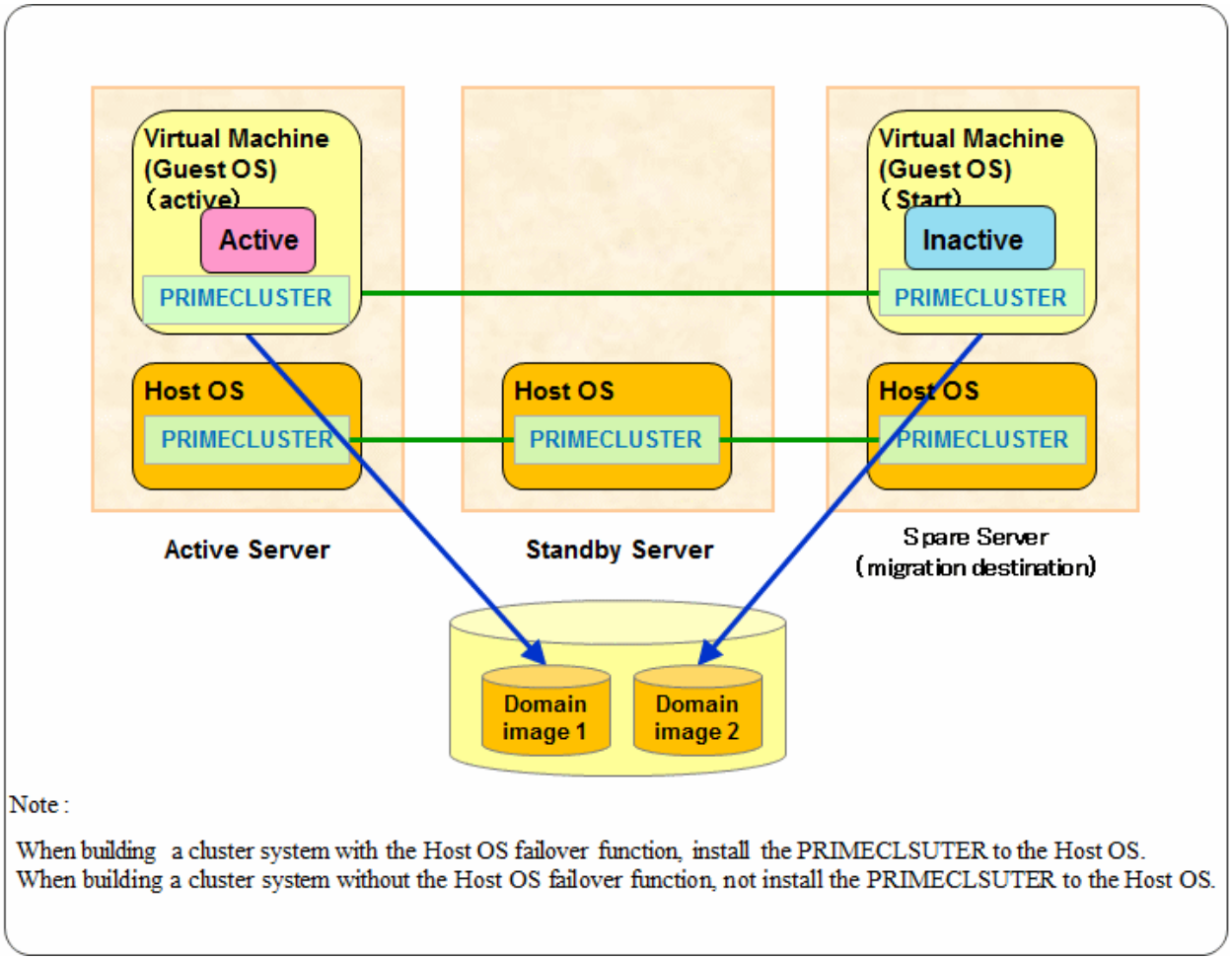
Figure 2.6 Offline Migration to a spare server (in performing)



Note :

When building a cluster system with the Host OS failover function, install the PRIMECLSUTER to the Host OS.
When building a cluster system without the Host OS failover function, not install the PRIMECLSUTER to the Host OS.

Figure 2.7 Offline Migration to a spare server (after performing)



Note :

When building a cluster system with the Host OS failover function, install the PRIMECLSUTER to the Host OS.
When building a cluster system without the Host OS failover function, not install the PRIMECLSUTER to the Host OS.

- Migration by Export/Import

  By migrating a stopped guest OS by Export/Import, the guest OS can be started in a spare server, and you can do standby server maintenance while maintaining the redundant configuration for active and standby servers.

Figure 2.8 Migration by Export/Import to a spare server (before performing)

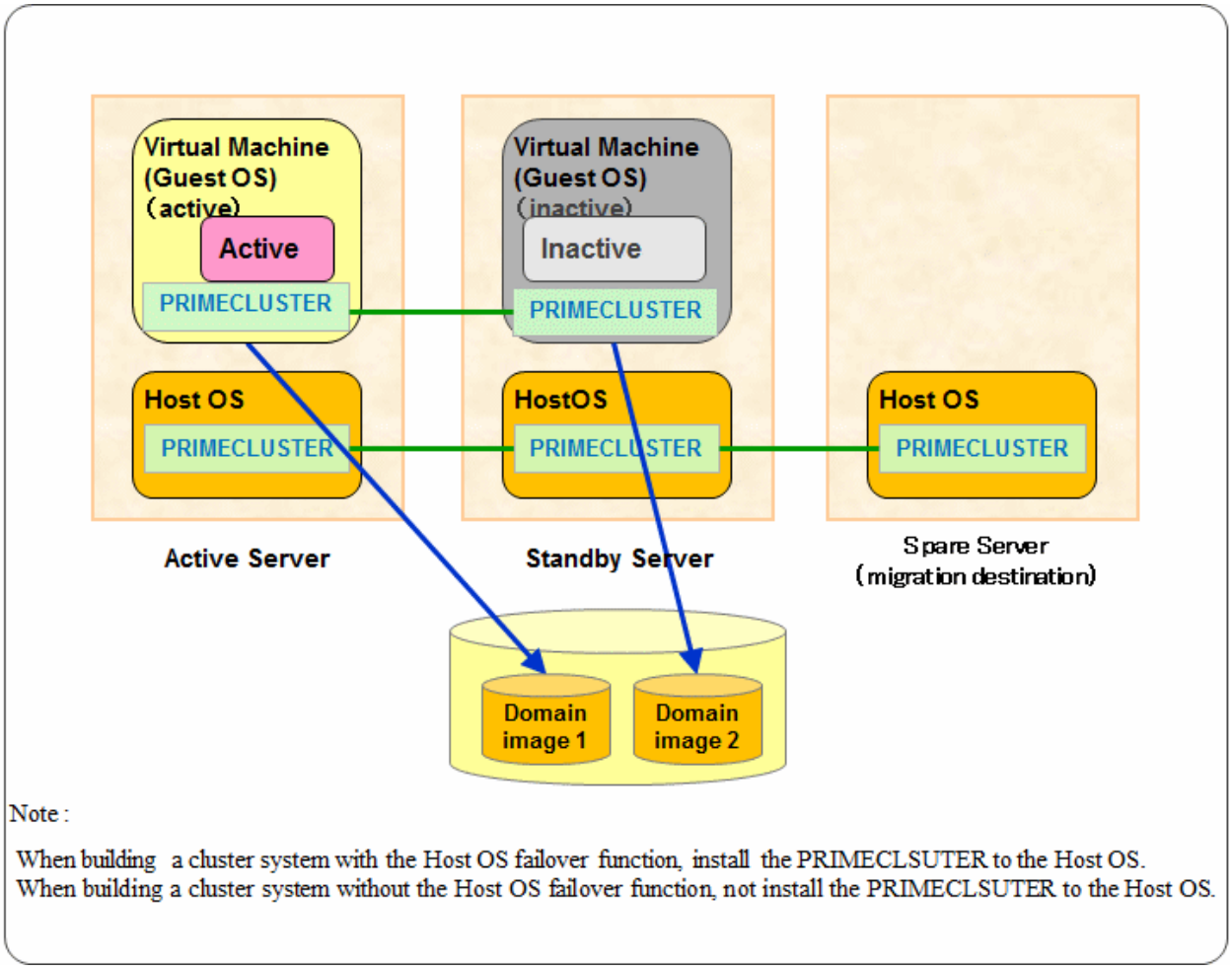
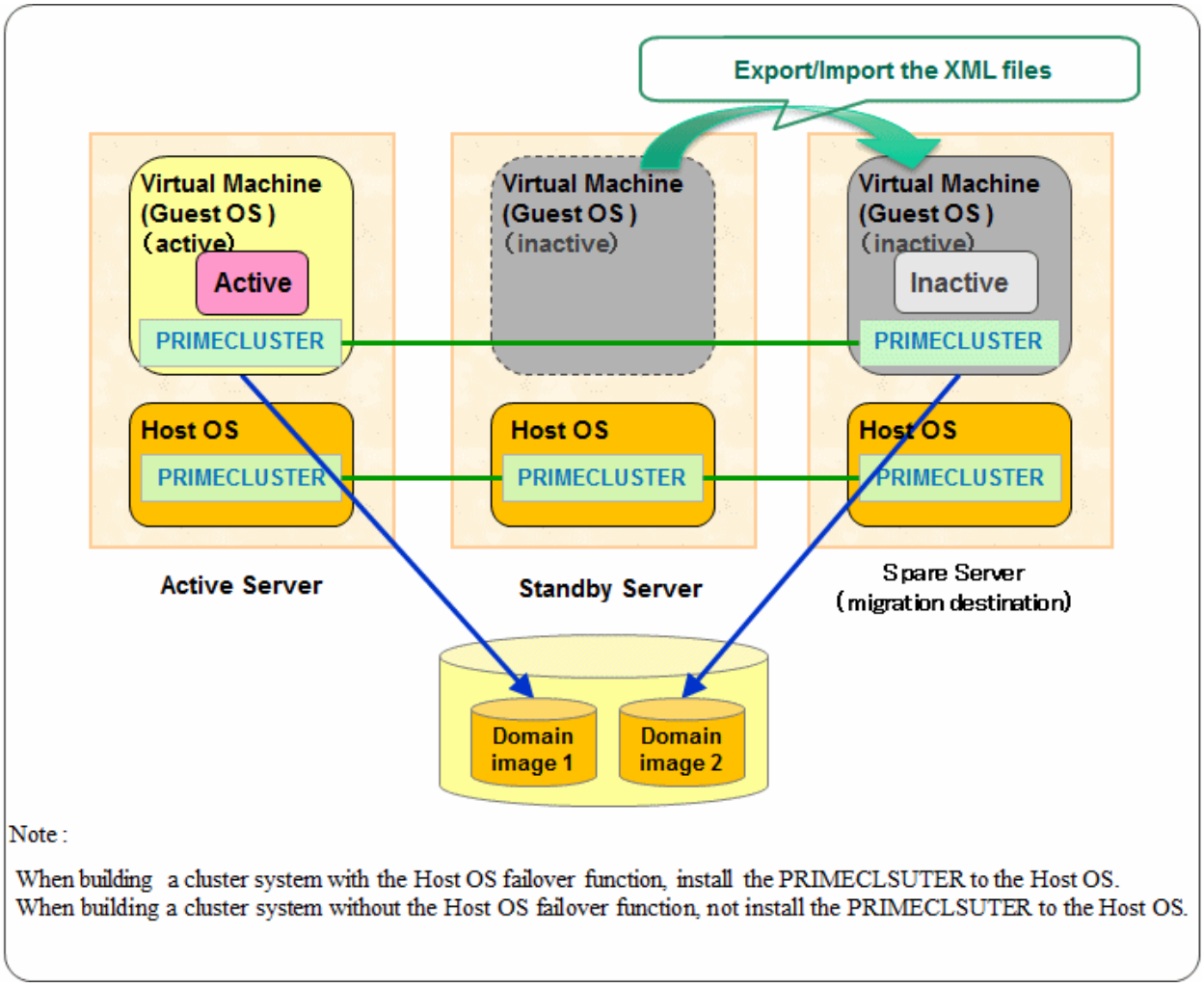


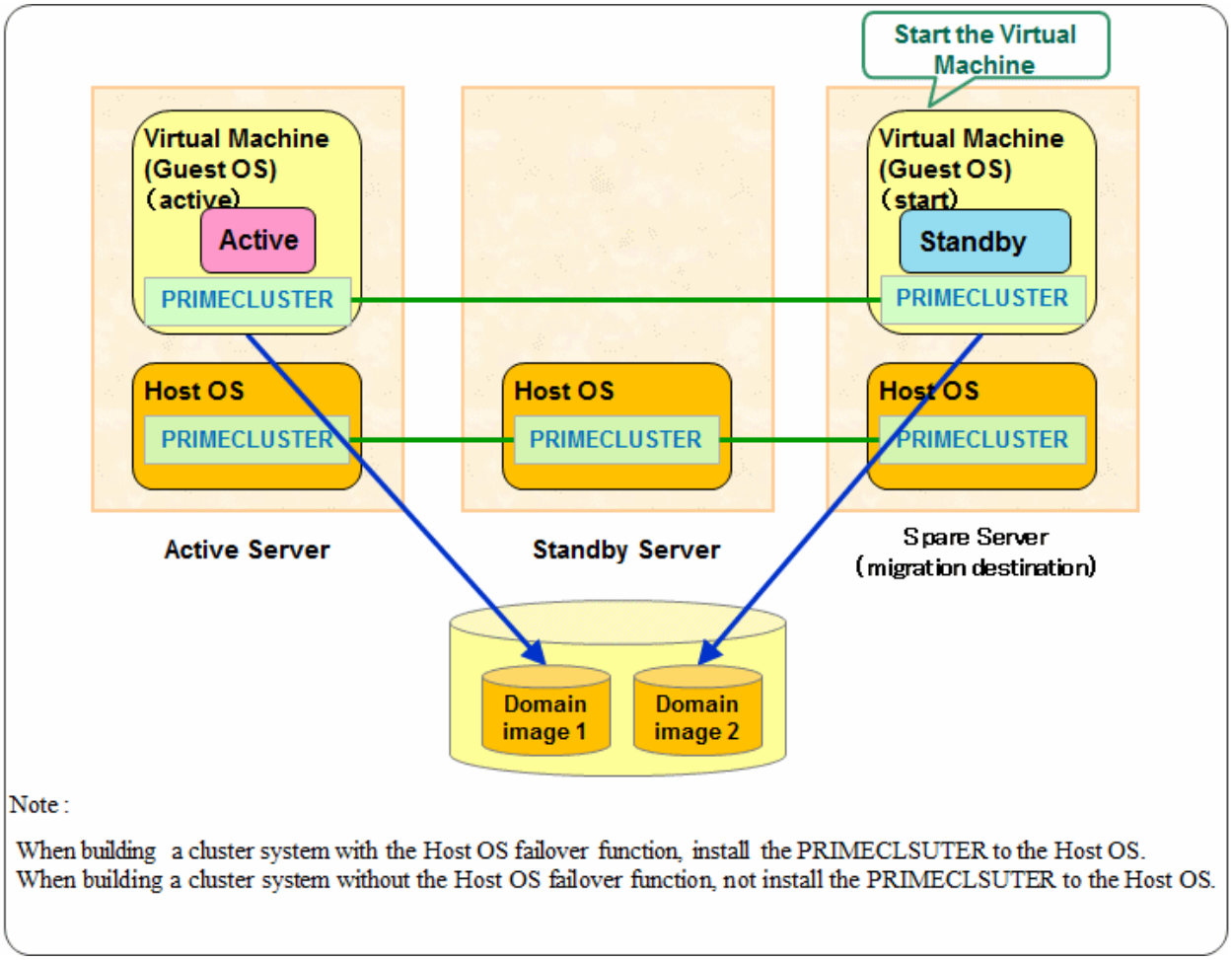
Figure 2.9 Migration by Export/Import to a spare server (in performing)



Note :

When building a cluster system with the Host OS failover function, install the PRIMECLSUTER to the Host OS.
When building a cluster system without the Host OS failover function, not install the PRIMECLSUTER to the Host OS.

Figure 2.10 Migration by Export/Import to a spare server (after performing)



Prerequisites are needed for using the Migration function of KVM in a cluster system. For details, see "Appendix G Using the Migration Function in a KVM Environment."

Note

- A cluster system is not switched during the Migration.
- Do not perform the Migration during a cluster system switchover.

2.3 Determining the Cluster System 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 (number of nodes)	2 to (number of supported nodes)
	N:1 standby	2 to (number of nodes - 1)	3 to (number of supported nodes)

Classification	Operation mode	Number of cluster applications	Number of nodes
	Cascade	1 to (number of nodes)	3 to (number of supported nodes)
	Priority transfer	2 to (number of nodes - 1)	3 to (number of supported nodes)
Scalable operation	Scalable	1 to (number of nodes)	1 to (number of supported nodes)
	High-availability scalable operation	1 to (number of nodes)	2 to (number of supported nodes)
Single-node cluster operation	-	1	1

Note

If an operating node in one side is disconnected abruptly due to a power failure or other power supply problem, failover may not work. Take corrective action as follows:

- Connect all the nodes to UPS.

2.3.1 Standby Operation

The topologies for standby operation are as shown below.

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.

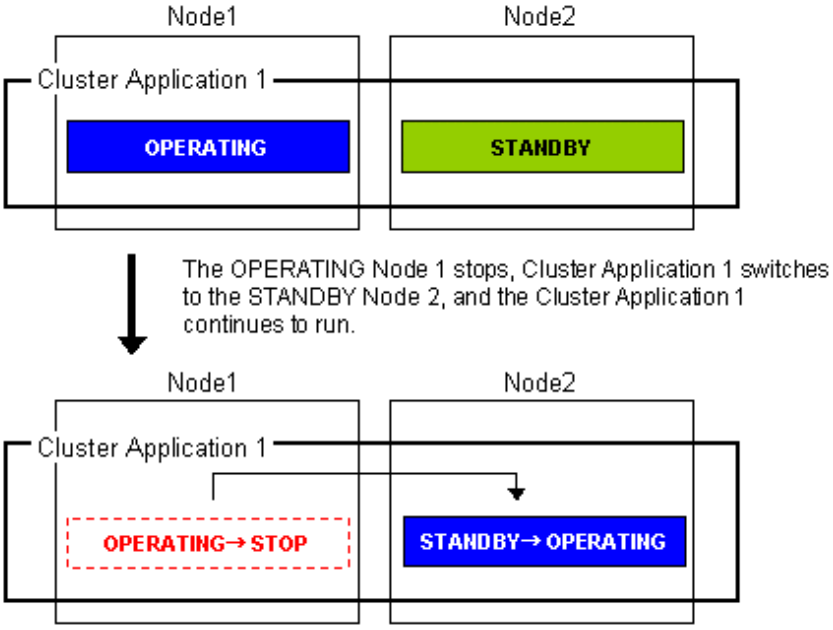
Advantage

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

Note

- Nodes in whole system 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 cluster applications. The operating cluster application has its own standby in each other's node.

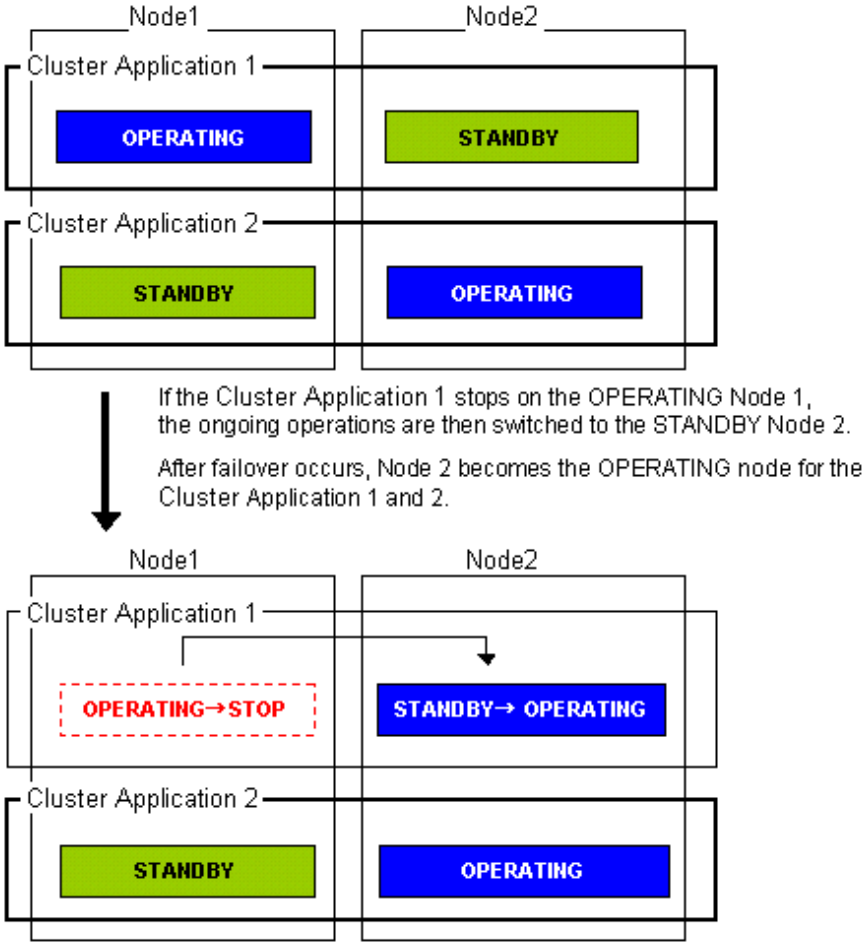
Advantage

- Since all the nodes are operating for cluster application, the nodes in whole system can be used efficiently.

Note

- 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.

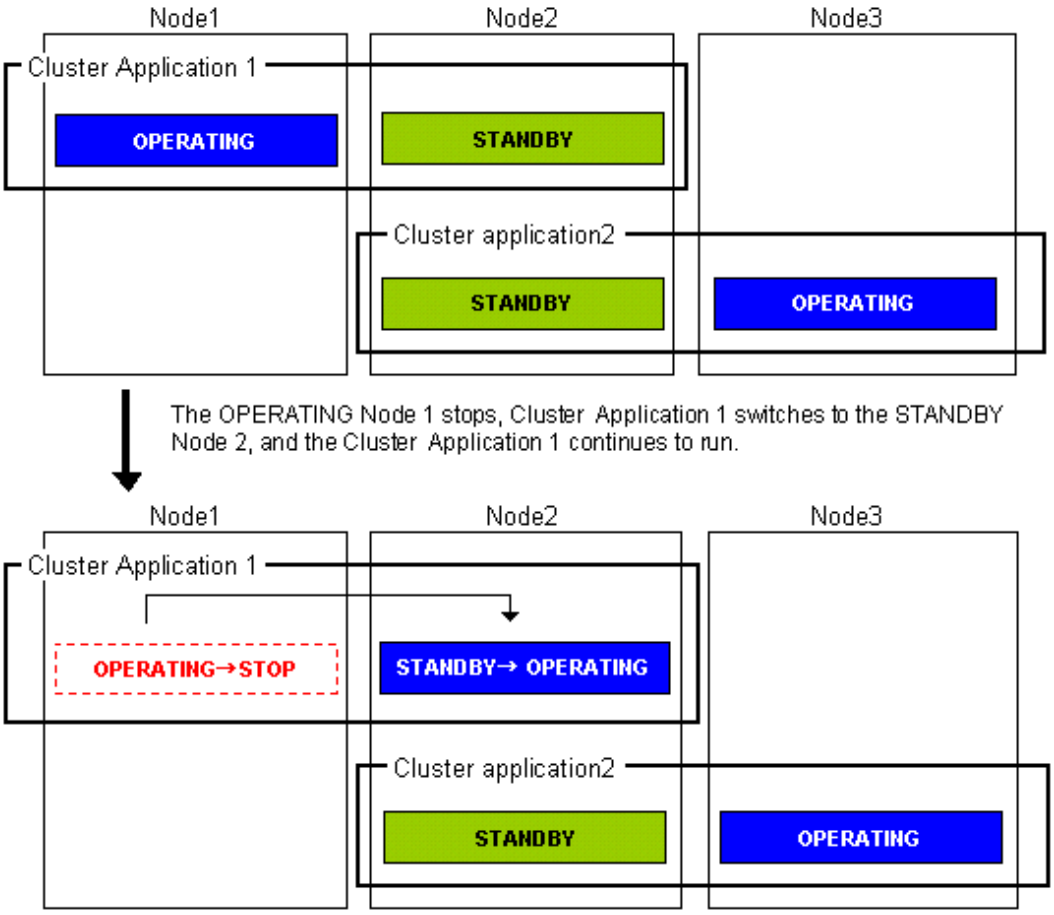
Advantages

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

Note

- If failover occurs for multiple cluster applications, the performance of the cluster applications is reduced because multiple cluster 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 cluster application switches to other standby node.

Advantages

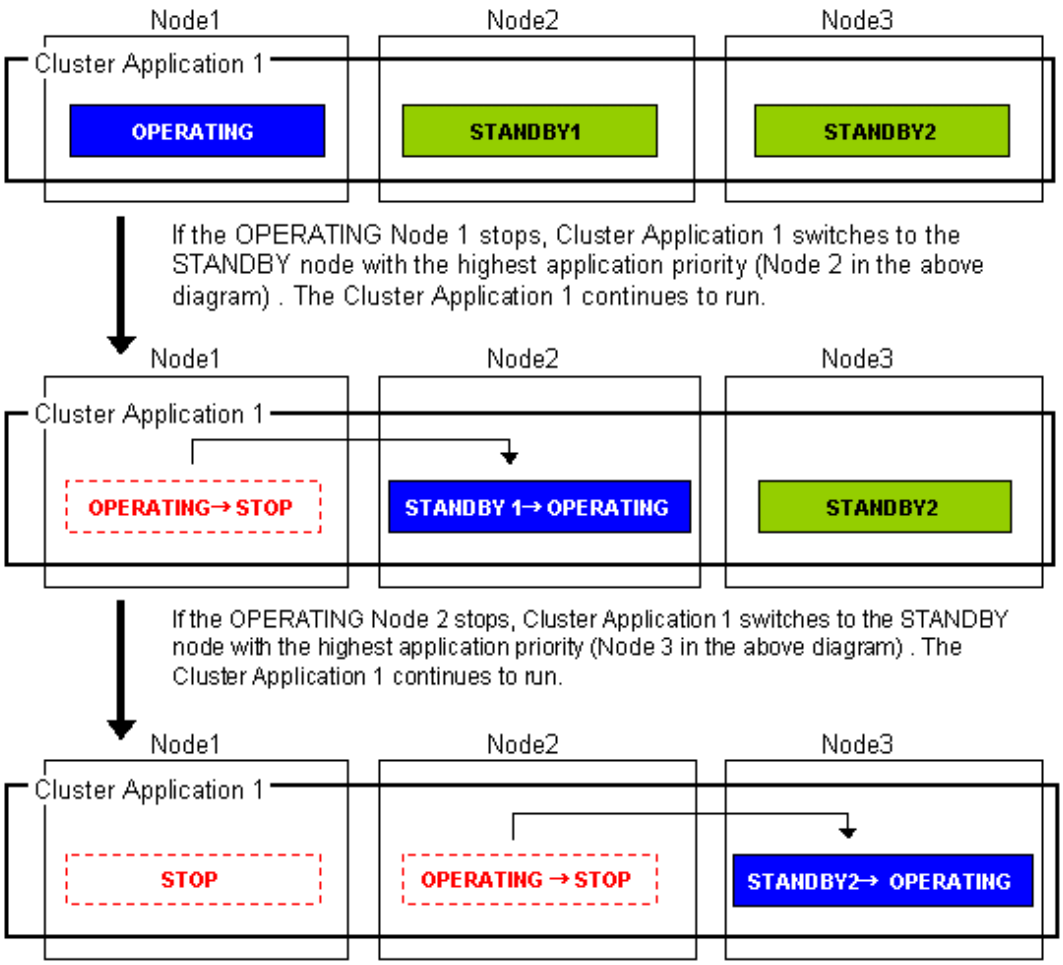
- Even after one node is stopped, the redundant configuration of the cluster application can be maintained by using other nodes. The availability is guaranteed during system maintenance.
- This operation mode ensures the availability of cluster applications even after failover.

Note

- As the system has a redundant configuration, nodes in whole system 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.



See

For information on how to set the cluster application priority, see Step 4 in "6.7.2.1 Creating Standby Cluster Applications."

Priority transfer (application of N:1 standby)

Definition

- One node functions as STANDBY for multiple cluster applications. For the other nodes, one 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

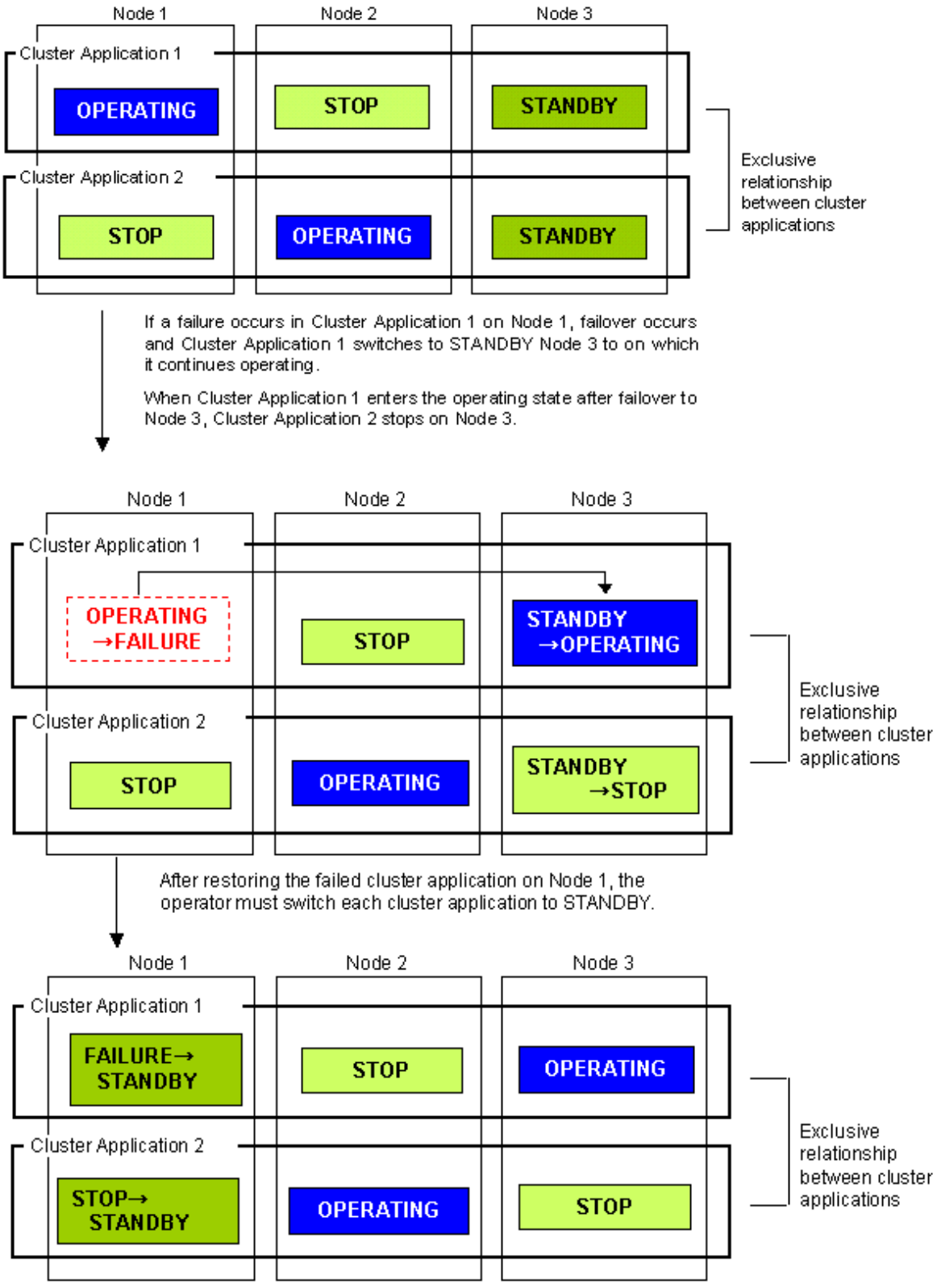
- On that node on which one cluster application is OPERATING, the other cluster applications do not become either OPERATING or STANDBY. Therefore, the throughput of that cluster application is guaranteed even after failover occurs.
- Because failback of the cluster application is not necessary during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

Notes

- Since one 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 one node, the availability decreases because no standby node is available until the completion of the maintenance work.

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.

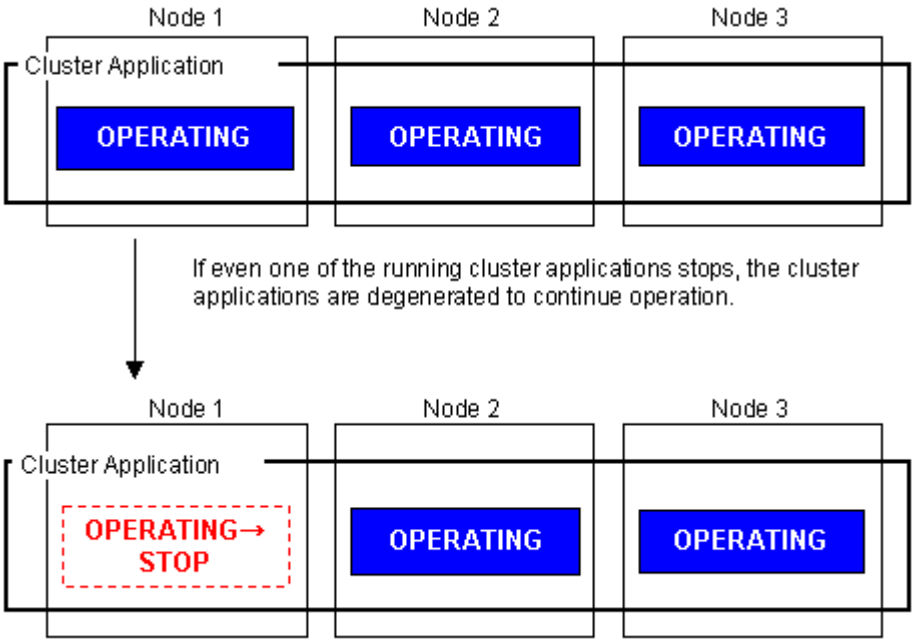
Advantage

- Degenerated operation is possible even if part of the cluster applications stops.

Note

- 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 manuals of PRIMECLUSTER-related products.

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.

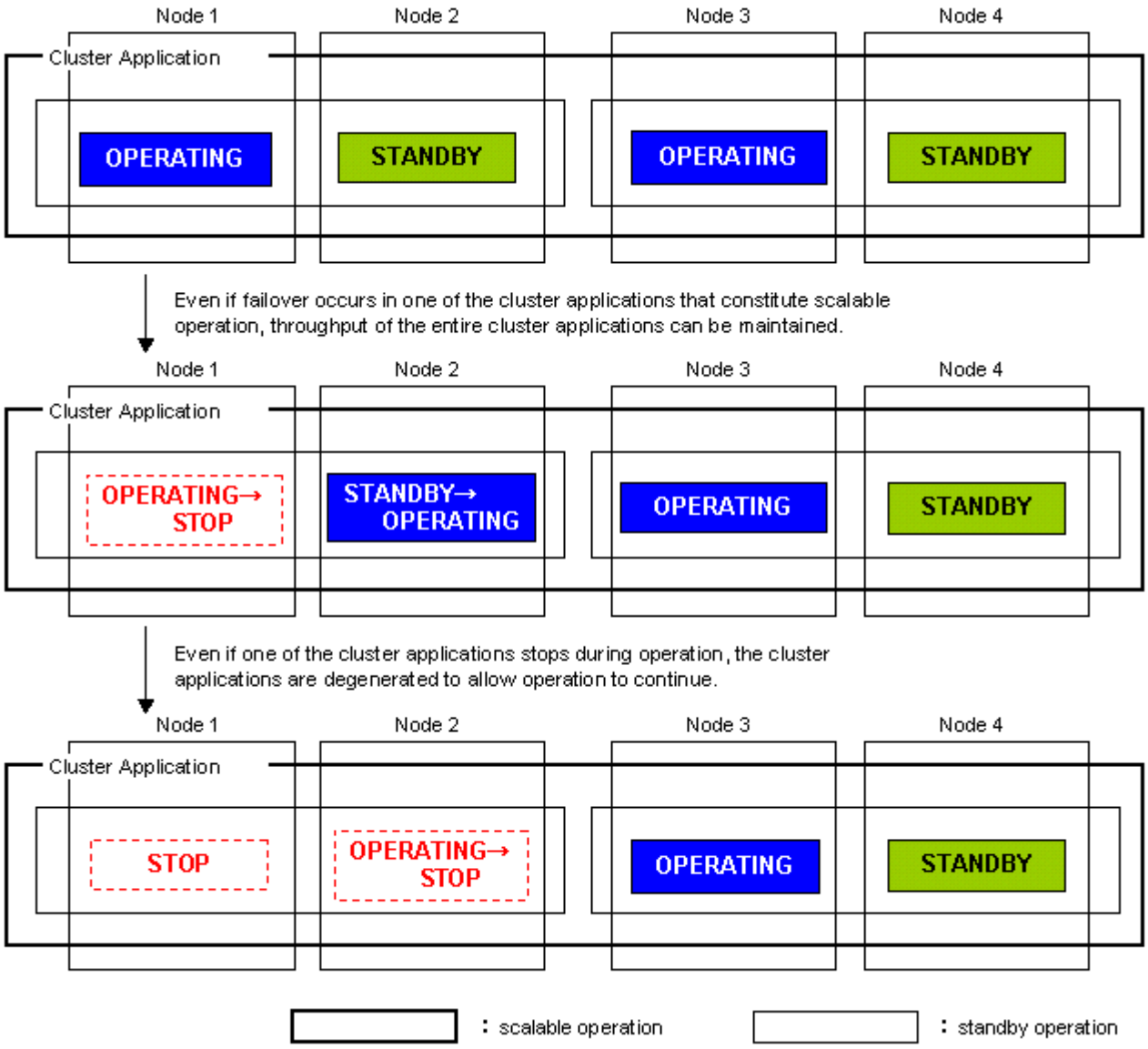
- Degenerated operation is possible even if part of the cluster applications stops.

Note

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

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 manuals of PRIMECLUSTER-related products.

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.

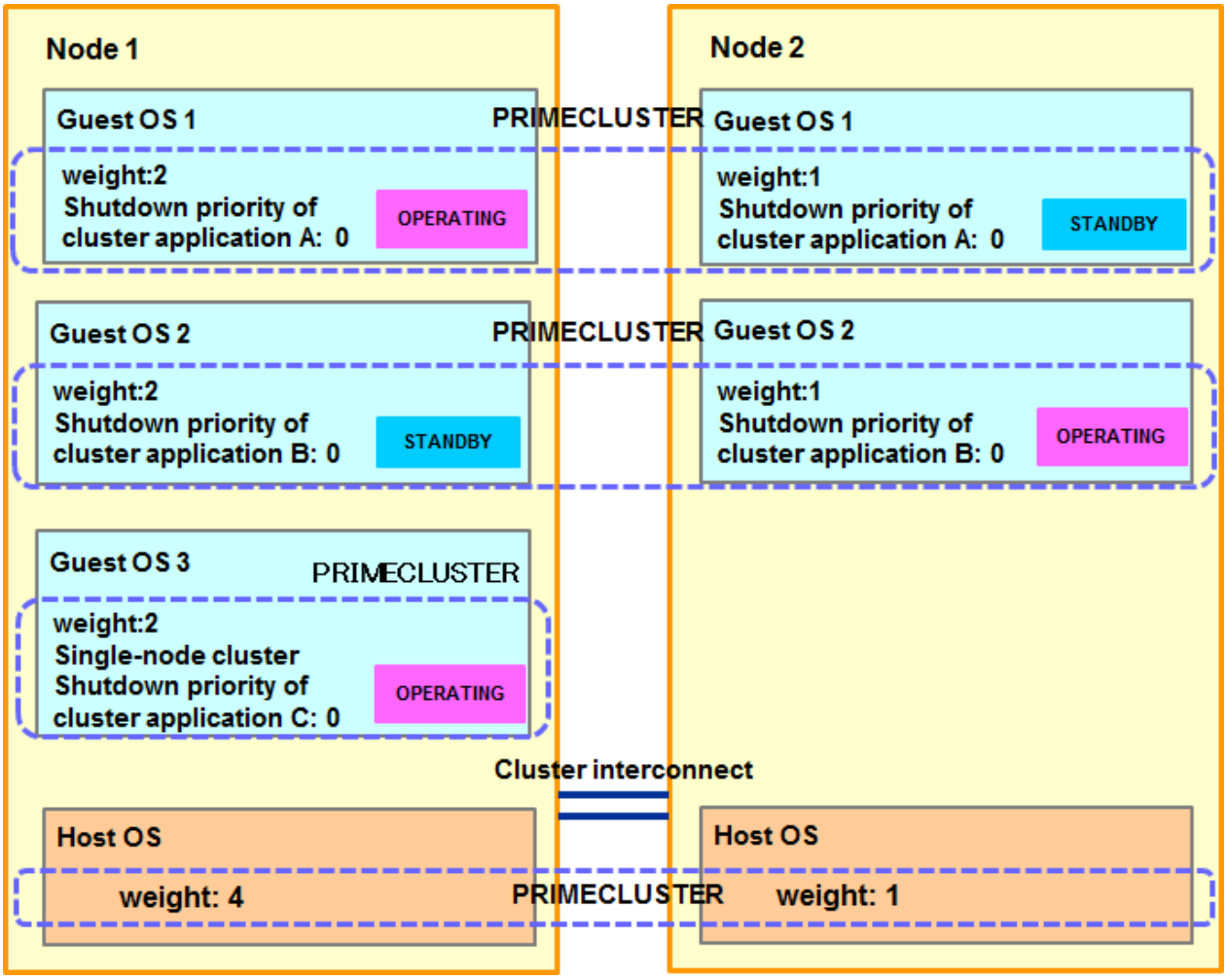
Advantages

- 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.

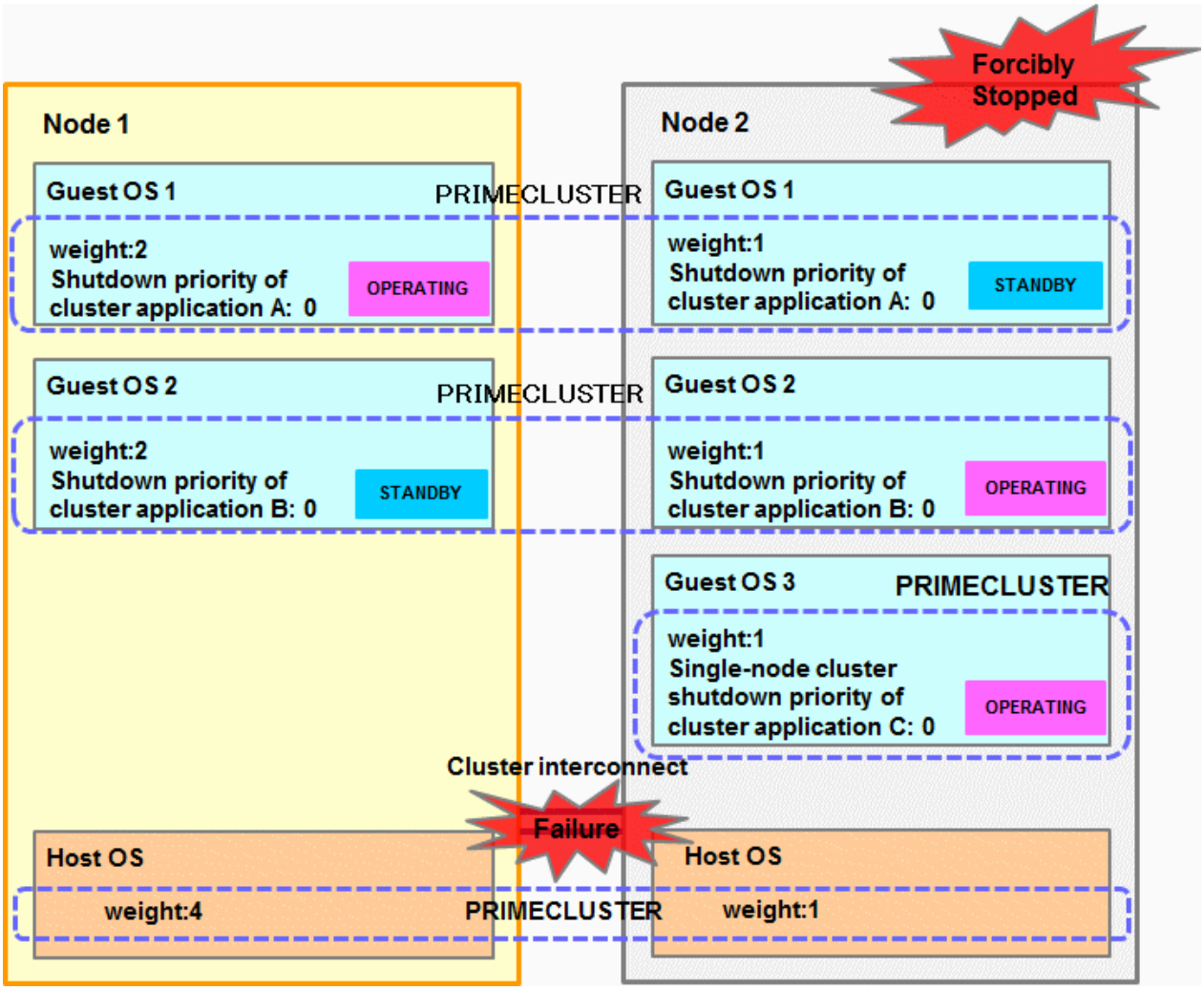
Figure 2.11 Example: Single-node cluster configuration



- In an environment in which the virtual machine environment is used, a guest OS 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

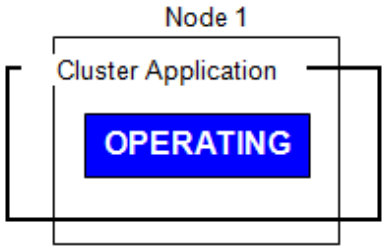
- The node is forcibly shutdown (due to an inter-node communication failure or other causes).

Figure 2.12 Example: Configuration for cluster interconnect failure



Failover image

No failover occurs in the single-node cluster operation.



Automatically restored in the case of resource error.
(When the AUTORECOVER attribute is set.)

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.

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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

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.

Client

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 on the Linux(R) servers.

Cluster nodes

Cluster nodes construct the PRIMECLUSTER system.
Cluster nodes run on the Linux(R) servers.

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

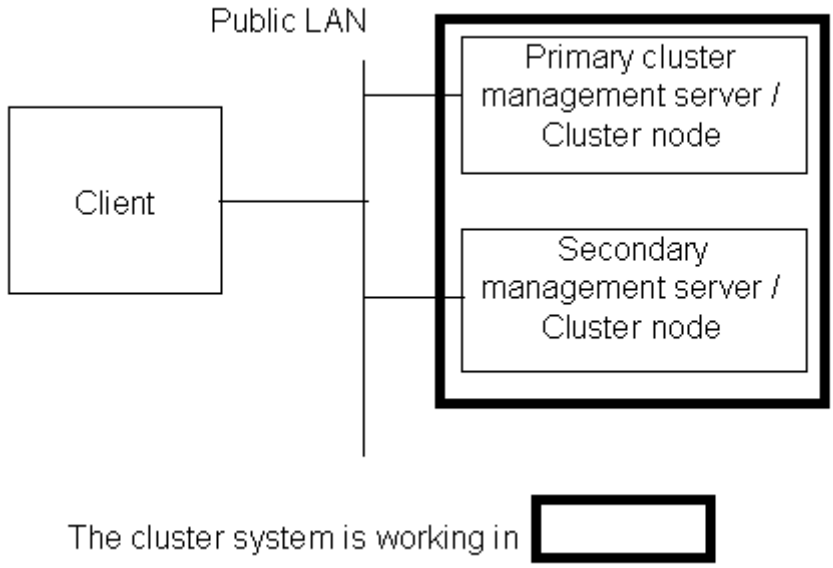
In the 2-tier model, the cluster management server and the cluster node are used together on the same machine, and the client is on a machine other than the nodes and the management servers.

This model supports configurations where the number of nodes is relatively small and which does not require a specific cluster management server.

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

Topology where a network is shared

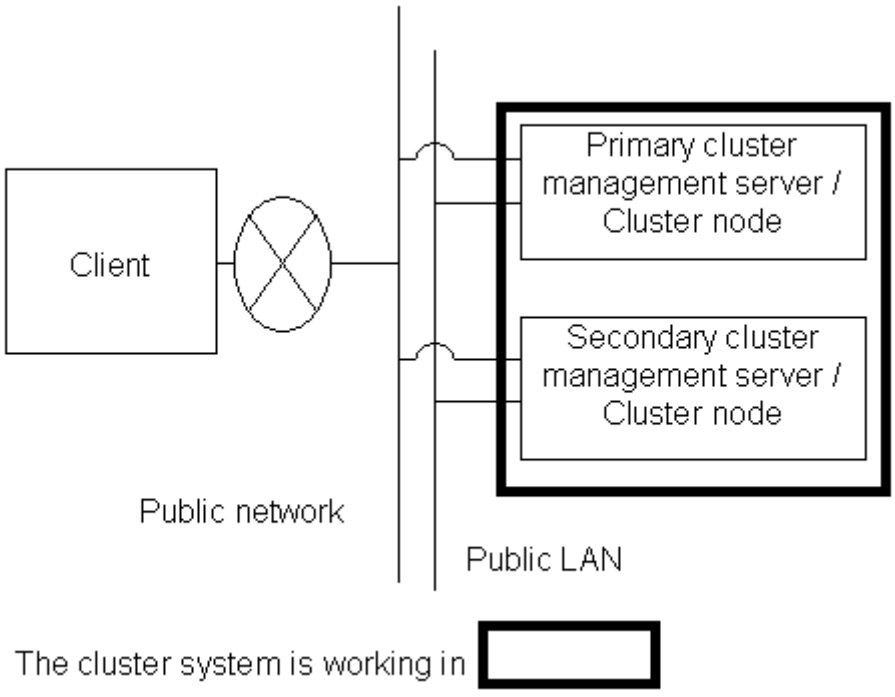
In this topology, the public LAN and the LAN that is connected to the management client are used together. You can adopt this topology if the network users and network range are limited for security. This is the default Web-Based Admin View configuration after PRIMECLUSTER installation.



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 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 (PRIMERGY)

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

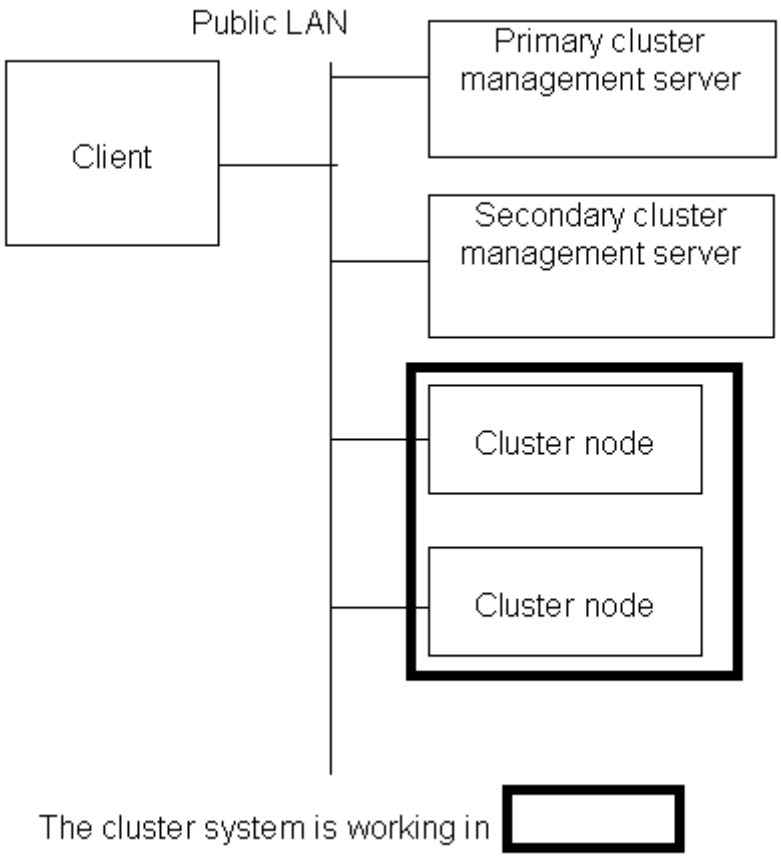
This model is adopted for configurations where the number of nodes is relatively large.

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.

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

Topology where a network is shared

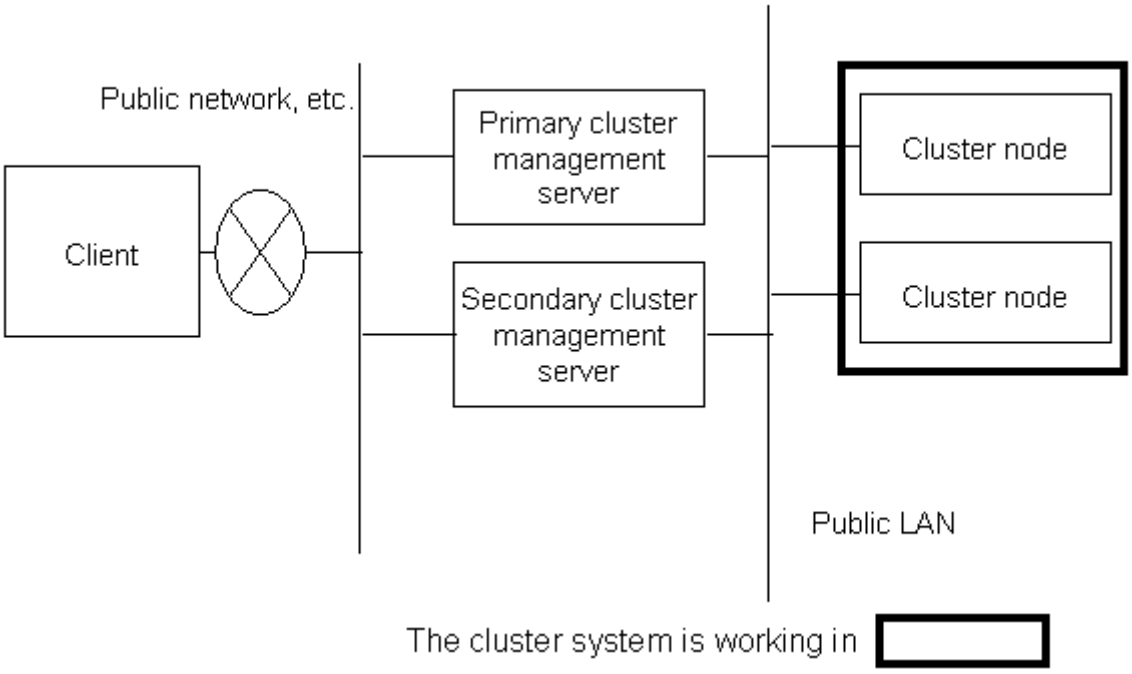
In this topology, the public LAN and the LAN that is connected to the management client are the same. You can adopt this topology if the network users and network range are limited for security. This is the default Web-Based Admin View configuration after PRIMECLUSTER installation.



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



2.5 Determining the failover timing of a cluster application

Determine the failover timing of a cluster application. You can choose from the following:

Multiple choices are possible from 2 to 4.

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

See

The failover timing is set in "6.7.2 Setting Up userApplication."

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 "Chapter 8 Changing the Cluster System Configuration."

Chapter 3 Software Installation and Setup.....45

Chapter 4 Preparation Prior to Building a Cluster.....84

Chapter 5 Building a Cluster.....98

Chapter 6 Building Cluster Applications.....138

Chapter 3 Software Installation and Setup

This chapter describes how to install and set up software products related to PRIMECLUSTER for the following cases:

- When not using the virtual machine function
- When using the virtual machine function

Note

- For the security, set "No Firewall" when a Red Hat Enterprise Linux is installed or when the setup command is executed. If Firewall has already been set for the security, change the setting to "No Firewall." If the "Firewall" setting is left as is, the clsetup (setting of the resource database) command will operate abnormally.
- PRIMECLUSTER guarantees the performance of any required software when the umask value is set to 022. Do not modify the umask value.
- For immediate cluster failover if an I/O device where the system volume is placed fails

  If an I/O device where the system volume is placed fails, a cluster failover does not occur and the system operation may continue based on the data stored on the memory.
  If you want PRIMECLUSTER to trigger a cluster failover by panicking a node in the event that an I/O device where the system volume is placed fails, set the ext3 or the ext4 file system to the system volume and perform the following setting.

  Setting

  Specify "errors=panic" to the mount option of each partition (the ext3 or the ext4 file system) included in the system volume.

  Example: To set it in /etc/fstab (when /, /var, and /home exist in one system volume)

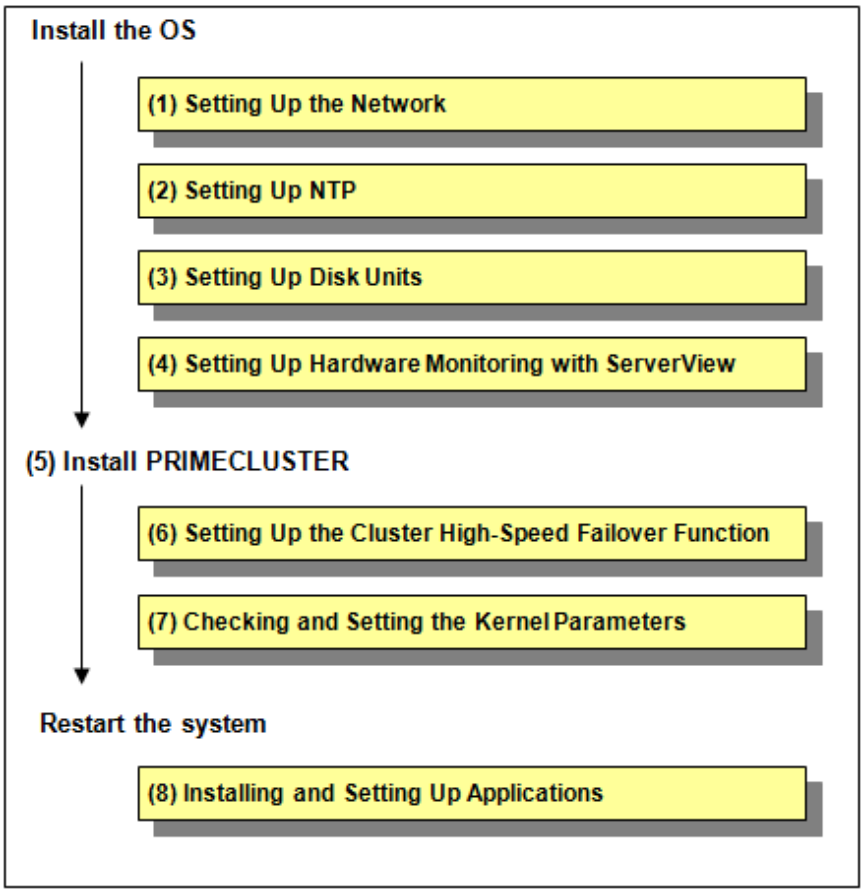
```
LABEL=/     /     ext3 errors=panic 1 1
LABEL=/boot /boot ext3 errors=panic 1 2
LABEL=/var  /var  ext3 errors=panic 1 3
LABEL=/home /home ext3 errors=panic 1 4
```

  However, an immediate cluster failover may not become available due to taking time for an I/O error to reach the file system. The regularly writing to the system volume enhances the detection frequency of I/O error.

3.1 When Not Using the Virtual Machine Function

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.

Perform the steps shown in the figure below as necessary.



3.1.1 Setting Up the Network

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

See

For information on changing the public LAN and administrative LAN that the PRIMECLUSTER system uses, see "9.2 Changing the Network Environment."

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.

3.1.2 Setting Up NTP

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

3.1.3 Setting Up Disk Units

When using a shared disk unit, you need to install and set up the software product.

Installation and Setup of Related Software

Install and set up the software products (ETERNUS Multipath driver) required for using shared disk units. For details on the installation and setup procedure, see "Software Information" for ETERNUS Multipath Driver.

3.1.4 Setting Up Hardware Monitoring with ServerView

Configure software related to hardware monitoring.

ServerView supports monitoring of the fan, temperature sensor, and power supply in addition to watchdog monitoring of the OS.

ServerView allows you to specify the action which will be done in the event of failure for each monitored target. For example, if "Shut down the server immediately" is selected and an error is detected, the failed node will be shut down. The ongoing operations on the failed node are then quickly switched to the standby node.

Note

No failover will be triggered by PRIMECLUSTER even if the operating system hangs up as long as communication with cluster interconnect is performed normally.

This state can be avoided by enabling watchdog timer monitoring.

For information about behavior setup, see the ServerView Operations Manager manual.

3.1.5 Installing PRIMECLUSTER

Install PRIMECLUSTER.

For details, see "3.3 PRIMECLUSTER Installation."

3.1.6 Setting Up the Cluster High-Speed Failover Function

You need to configure software and hardware that enables cluster high-speed failover after installing the OS and PRIMECLUSTER.

3.1.6.1 PRIMERGY

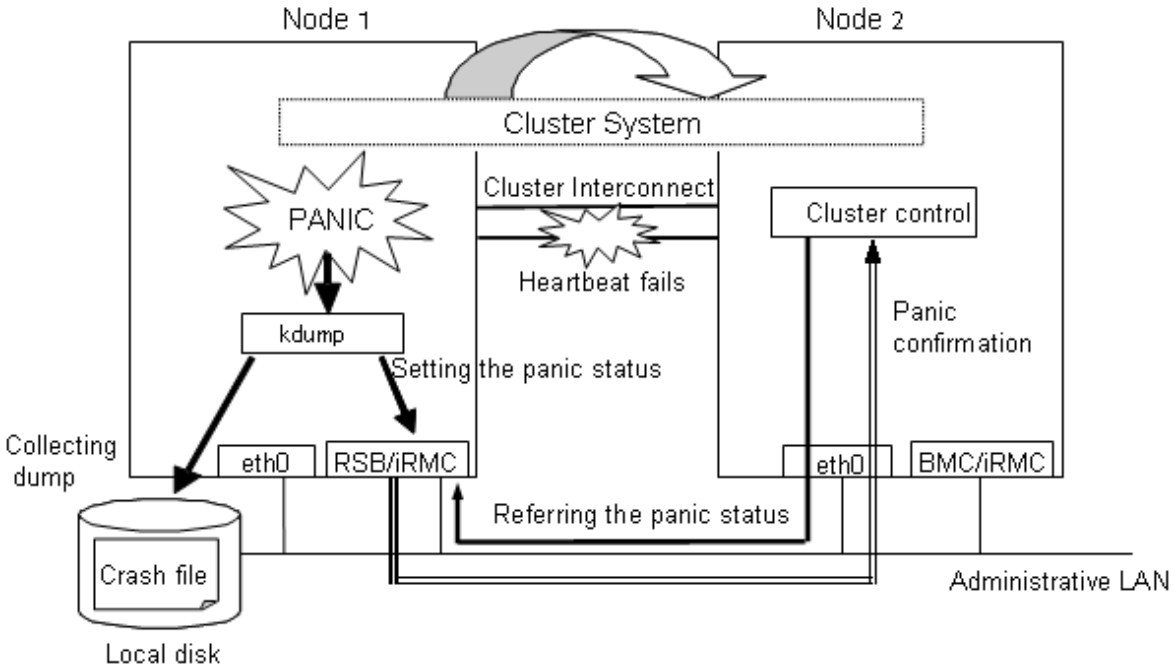
Overview

If one of the nodes that configure a cluster system fails and a heartbeat fails, the PRIMECLUSTER shutdown facility forcibly shuts down the failed node.

If the heartbeat fails due to a panic, the failed node collecting a crash dump is forcibly shut down and then crash dump collection ends in the middle. This means that you may not be able to collect information for troubleshooting.

The cluster high-speed failover function prevents a node from being forcibly shut down during crash dump collection, and at the same time, enables the ongoing operations on the failed node to be quickly moved to another node during crash dump collection.

kdump



As shown in the above figure, the cluster high-speed failover function allows for panic status setting and reference through BMC (Baseboard Management Controller) or iRMC when a heartbeat monitoring failure occurs. The node that detects the failure can consider that the other node is stopped and takes over ongoing operation without forcibly shutting down the node that is collecting a crash dump.

Note

- If you reset the node that is collecting crash dump, collection of the crash dump will fail.
- When the node completes collecting the crash dump after it gets panicked, the behavior of the node follows the setting of kdump.

Required setting for the kdump shutdown agent

1. Configure kdump

   When using kdump, it is necessary to configure the kdump.

   For details on the configuration procedure, see the manual of your OS.

   #### Note

   Configure the kdump again if it is already configured with the installation of Red Hat Enterprise Linux.

2. Check kdump

   [RHEL6]

   Check if the kdump is available. If not, enable the kdump using the "runlevel(8)" and "chkconfig(8)" commands.

   - Check the current runlevel using the "runlevel(8)" command.

     Example:

     ```
     # /sbin/runlevel
     N 3
     ```

     The above example shows that the current runlevel is 3.

- Check if the kdump is available using the "chkconfig(8)" command.

  Example:

```
# /sbin/chkconfig --list kdump
kdump  0:off 1: off 2: off 3: off 4: off 5: off 6: off
```

  The above example shows that the kdump in the current runlevel 3 is off.

- If the kdump is off in the current runlevel, enable the kdump by executing the "chkconfig(8)" command, and then start the kdump by executing the service command.

```
# /sbin/chkconfig kdump on
# /sbin/service kdump start
```

[RHEL7]

Check if the kdump is available. If not, enable the kdump using the "runlevel(8)" and "systemctl(1)" commands.

- Check the current runlevel using the "runlevel(8)" command.

  Example:

```
# /sbin/runlevel
N 3
```

  The above example shows that the current runlevel is 3.

- Check if the kdump is available using the "systemctl(1)" command.

  Example:

```
# /usr/bin/systemctl list-unit-files --type=service | grep kdump.service
kdump.service                               disabled
```

  The above example shows that the kdump in the current runlevel 3 is disabled.

- If the kdump is disabled in the current runlevel, enable the kdump by executing the "systemctl(1)" command, and then start the kdump.

```
# /usr/bin/systemctl enable kdump.service
# /usr/bin/systemctl start kdump.service
```

Prerequisites for the other shutdown agent settings

After you completed configuring the kdump shutdown agent, set the IPMI (Intelligent Platform Management Interface) or BLADE server.

Information

The IPMI shutdown agent is used with the hardware device in which BMC or iRMC is installed.

Prerequisites for the IPMI shutdown agent settings

Set the following for BMC or iRMC.

- IP address
- User for the IPMI shutdown agent (*1)

For details, see "User Guide" provided with the hardware and the ServerView Operations Manager manual.

*1) Assign this user as the administrator. Set the user password with seven-bit ASCII characters except the following characters.
> < " / \ = ! ? ; , &

Prerequisites for the Blade shutdown agent settings

Set the following for the BLADE server:

- Install ServerView
- Set SNMP community for the management blade (*2)
- Set an IP address of the management blade

For details, see the operation manual provided with the hardware and the ServerView Operations Manager manual.

*2) When configuring the cluster across multiple chassis, set the same SNMP community for all the management blades.

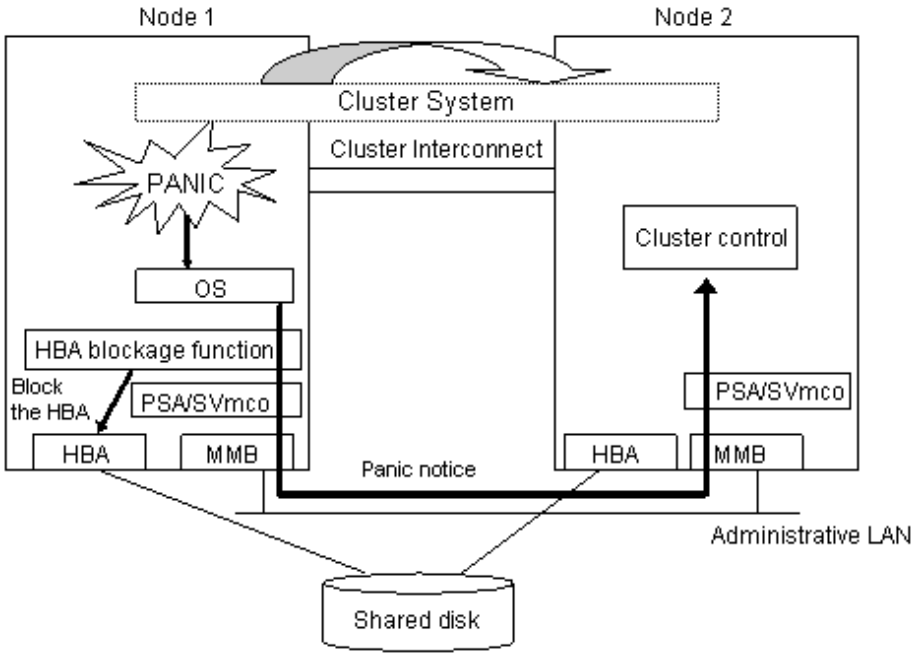
3.1.6.2 PRIMEQUEST

If an error occurs in one of the nodes of the cluster system where PRIMEQUEST is used, the PRIMECLUSTER shutdown facility uses the two methods described below to detect that error. For details, see "2.3.1.7 PRIMECLUSTER SF" in "PRIMECLUSTER Concepts Guide."

(1) Node status change detection through MMB units (asynchronous monitoring)

(2) Heartbeat failure between cluster nodes (NSM: node status monitoring) (cyclic monitoring)

The asynchronous monitoring of (1) allows node errors to be detected immediately, and failover occurs at a higher speed than when detected by the cyclic monitoring of (2).



As shown in the above figure, if a panic occurs, the cluster control facility uses the MMB units to receive the panic notice. This allows the system to detect the node panic status faster than it would be a heartbeat failure.

See

PRIMEQUEST allows you to set the panic environment so that a crash dump is collected if a panic occurs.

For details about the PRIMEQUEST dump function, setup method, and confirmation method, see the following manuals:

- PRIMEQUEST 1000 Series
  - "PRIMEQUEST 1000 Series Installation Manual"
  - "PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual"
- PRIMEQUEST 2000 Series
  - "PRIMEQUEST 2000 Series Installation Manual"
  - "PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual"

To use asynchronous monitoring (1), you must install software that controls the MMB and specify appropriate settings for the driver. This section describes procedures for installing the MMB control software and setting up the driver, which are required for realizing high-speed failover.

1. Installing the HBA blockage function and the PSA/SVmco

   The HBA blockage function and the PSA/SVmco report node status changes through the MMB units to the shutdown facility. Install the HBA blockage function and the PSA/SVmco before setting up the shutdown facility. For installation instructions, see the following manuals:

   - PRIMEQUEST 1000 Series
     - "PRIMEQUEST 1000 SERIES HBA blockage function USER'S GUIDE"
     - "PRIMEQUEST 1000 Series Installation Manual"
     - "PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual"
   - PRIMEQUEST 2000 Series
     - "PRIMEQUEST 2000 Series HBA blockage function USER'S GUIDE"
     - "PRIMEQUEST 2000 Series Installation Manual"
     - "PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual"

2. Setting up the PSA/SVmco and the MMB units

   The PSA/SVmco and MMB must be set up so that node status changes are reported properly to the shutdown facility through the MMB units. Set up the PSA/SVmco units before setting up the shutdown facility. For setup instructions, see the following manuals:

   - PRIMEQUEST 1000 Series
     - "PRIMEQUEST 1000 Series Installation Manual"
     - "PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual"
   - PRIMEQUEST 2000 Series
     - "PRIMEQUEST 2000 Series Installation Manual"
     - "PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual"

   You must create an RMCP user so that PRIMECLUSTER can link with the MMB units.

   In all PRIMEQUEST instances that make up the PRIMECLUSTER system, be sure to create a user who uses RMCP to control the MMB. To create a user who uses RMCP to control the MMB, log in to MMB Web-UI, and create the user from the "Remote Server Management" window of the "Network Configuration" menu. Create the user as shown below.

   - Set [Privilege] to "Admin".
   - Set [Status] to "Enabled".

   Set the user password with seven-bit ASCII characters except the following characters.

   ```
   >  <  "  /  \  =  !  ?  ;  ,  &
   ```

   For details about creating a user who uses RMCP to control the MMB, see the following manuals:

   - PRIMEQUEST 1000 Series

     "PRIMEQUEST 1000 Series Tool Reference"

   - PRIMEQUEST 2000 Series

     "PRIMEQUEST 2000 Series Tool Reference"

   The user name created here and the specified password are used when the shutdown facility is set up. Record the user name and the password.

Note

The MMB units have two types of users:

- User who controls all MMB units
- User who uses RMCP to control the MMB

The user created here is the user who uses RMCP to control the MMB. Be sure to create the correct type of user.

3. Setting up the HBA blockage function

Note

Be sure to carry out this setup when using shared disks.

If a panic occurs, the HBA units that are connected to the shared disks are closed, and I/O processing to the shared disk is terminated. This operation maintains data consistency in the shared disk and enables high-speed failover.

On all the nodes, specify the device paths of the shared disks (GDS device paths if GDS is being used) in the HBA blockage function command, and add the shared disks as targets for which the HBA function is to be stopped. If GDS is being used, perform this setup after completing the GDS setup. For setup instructions, see the following manuals:

- PRIMEQUEST 1000 Series

  "PRIMEQUEST 1000 SERIES HBA blockage function USER'S GUIDE"

- PRIMEQUEST 2000 Series

  "PRIMEQUEST 2000 Series HBA blockage function USER'S GUIDE"

4. Setting the I/O completion wait time

   To maintain consistent I/O processing to the shared disk if a node failure (panic, etc.) occurs and failover takes place, some shared disk units require a fixed I/O completion wait time, which is the duration after a node failure occurs until the new operation node starts operating.

   The initial value of the I/O completion wait time is set to 0 second. However, change the value to an appropriate value if you are using shared disk units that require an I/O completion wait time.

Information

ETERNUS Disk storage systems do not require an I/O completion wait time. Therefore, this setting is not required.

Specify this setting after completing the CF setup. For setting instructions, see "5.1.2.4.5 Setting I/O Completion Wait Time."

Note

If an I/O completion wait time is set, the failover time when a node failure (panic, etc.) occurs increases by that amount of time.

3.1.7 Checking and Setting the Kernel Parameters

To operate the PRIMECLUSTER-related software, you need to edit the values of the kernel parameters based on the environment.

Perform this setup before restarting the installed PRIMECLUSTER.

Target node:

All the nodes in which PRIMECLUSTER is to be installed

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

Check PRIMECLUSTER Designsheets and edit the value if necessary.

Note

To enable modifications, you need to restart the operating system.

Set an appropriate kernel parameter as follows based on the type of "Characteristics" in each table.

- Addition

  Set the total number of the recommended values and specified values for system default values and for each software.

- Maximum value

  Specify the maximum value in the recommended values and specified values for each software.

  However, make sure to use the system default value if the maximum value is less than that.

The kernel parameter values differ depending upon:

- CF Configuration

Kernel parameter	Characteristics	Value	Remarks (parameter description)
SEMMNI value	Addition	20	Maximum number of semaphore ID in whole system
SEMMNS value	Addition	30	Maximum number of semaphore that can be used in entire system
kernel.shmmax	Maximum value	1048576 + value required for resource database (*1)	Maximum size of shared memory segments
kernel.shmmni	Addition	30	Maximum number of shared memory segments

(*1)

Estimate the value required for resource database according to the following equation:

```
Value required for resource database = 2776 x number of resources
```

Estimate the number of resources according to the following equation:

```
Number of resources = Number of disks in shared system devices x (number of shared nodes +1) x 2

Specify the following in "Number of disks in shared system devices":
 - For a disk array unit, specify the number of logical units (LUN).
 - For other than a disk array unit, specify the number of physical units.
Specify the number of nodes connected to the shared disk in "number of shared nodes."
```

Note

For system expansion, if you increase the logical disks, you need to re-estimate the resources and restart each node in the cluster system. If you add disks to the cluster after installation, you must then calculate the resources required for the total number of logical disks after addition.

- 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/sysctl.conf are automatically updated if not defined or defined with smaller value than the following "Value".

Kernel parameter	Characteristics	Value	Remarks (parameter description)
kernel.msgmnb	Maximum value	4194304	Byte size of the message queue

Kernel parameter	Characteristics	Value	Remarks (parameter description)
kernel.msgmax	Maximum value	16384	Maximum size of the message text
kernel.msgmni	Addition	8192	Maximum number of message queues for the entire system

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 above (0x4d2 to 0xcee).
- Even if definitions of the kernel parameters in /etc/sysctl.conf are automatically added/updated, change the value as necessary in consideration of the value required by other software and user applications.

- Using GFS

Kernel parameter	Characteristics	Value	Remarks (parameter description)
SEMMNI value	Addition	2	Maximum value for semaphore identifiers that can be used in the entire system. Add 2 to the current value.
SEMMNS value	Addition	11	Maximum number of semaphore that can be used in the entire system. Add 11 to the current value.

Note

The values used by products and user applications that operate in the PRIMECLUSTER system must also be included in the kernel parameter values.

Described below is the procedure for changing the kernel parameters and setting new values.

1. Check the current values of the kernel parameters.

   To check the current effective values of the kernel parameters, execute the following command:

```
# sysctl -a | grep sem
kernel.sem = 20 90 10 20
```

   The displayed values represent the following kernel parameters:

   kernel.sem = SEMMSL value SEMMNS value SEMOPM value SEMMNI value

```
# sysctl -a | grep shm
kernel.shmmni = 4315
kernel.shmmax = 4000000000

# sysctl -a | grep msg
kernel.msgmnb = 4194304
kernel.msgmni = 8199
kernel.msgmax = 32768
```

2. Determine the kernel parameter values.

   The kernel parameter values are determined by the current effective values that were checked in step 1 and the values in the above table. If the example displayed in step 1 shows the current effective values of the kernel parameters, the edited line becomes the following:

   SEMMSL value: 20

   SEMMNS value: 131

SEMOPM value: 10

SEMMNI value: 42

kernel.shmmni: 4345

kernel.shmmax: 4000000000

kernel.msgmnb: 4194304

kernel.msgmni: 16391

kernel.msgmax: 32768

3. Change the kernel parameters.

   1. Edit the settings file.

      To set the kernel parameters, use an editor and edit the /etc/sysctl.conf file.

      Change the values in the following row to the values that were determined in step 2.

      If the example displayed in step 1 shows the current effective values of the kernel parameters, the edited line becomes the following:

      kernel.sem = 20 131 10 42

      kernel.shmmni = 4345

      kernel.shmmax = 4000000000

      kernel.msgmnb = 4194304

      kernel.msgmni = 16391

      kernel.msgmax = 32768

   2. Apply the settings.

      To change the kernel parameter values to the values in the settings file, execute the following command:

      ```
      # sysctl -p
      ```

4. Check the setting changes in the kernel parameters.

   To check whether the kernel parameter values were changed correctly, execute the following commands and display the current values:

   ```
   # sysctl -a | egrep sem
   kernel.sem = 20 131 10 42
   ```

   ```
   # sysctl -a | grep shm
   kernel.shmmni = 4345
   kernel.shmmax = 4000000000
   ```

   ```
   # sysctl -a | grep msg
   kernel.msgmnb = 4194304
   kernel.msgmni = 16391
   kernel.msgmax = 32768
   ```

   Check that the displayed values are the values that were determined in step 2.

3.1.8 Installing and Setting Up Applications

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

For details, see "3.4 Installation and Environment Setup of Applications."

3.2 When Using the Virtual Machine Function

After installing the PRIMECLUSTER-related software, you need to set up the operating system, hardware, and so on that will be used and administered.

When using PRIMECLUSTER on a virtual machine (KVM environment), setting procedure and contents are different depending on the following the cluster systems:

- When building a cluster system between guest OSes on one host OS

  See "3.2.1 When building a cluster system between guest OSes on one host OS."

- When building a cluster system between guest OSes on multiple host OSes
  - Without using Host OS failover function

    See "3.2.2 When building a cluster system between guest OSes on multiple host OSes without using Host OS failover function."

  - Using Host OS failover function

    See "3.2.3 When building a cluster system between guest OSes on multiple host OSes using Host OS failover function."

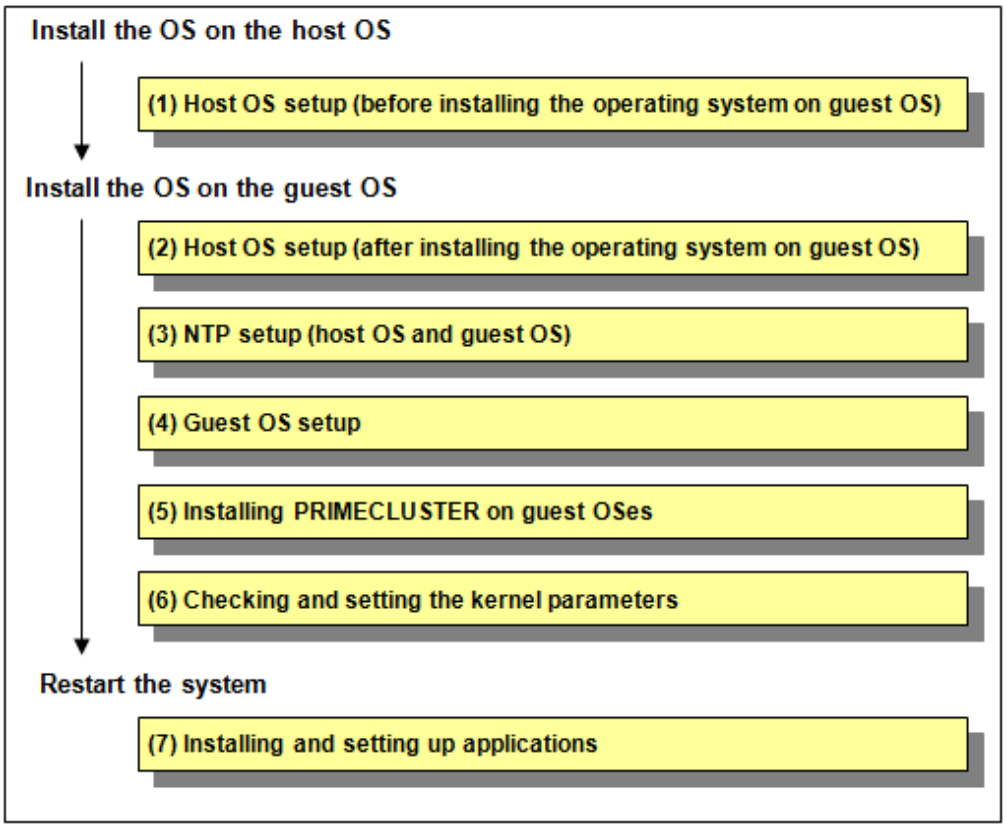
See

When using the virtual machine function in a VMware environment, see "Appendix H Using PRIMECLUSTER in a VMware Environment."

3.2.1 When building a cluster system between guest OSes on one host OS

This section describes how to install and set up related software when building a cluster system between guest OSes on one host OS.

Perform the steps shown in the figure below as necessary.



3.2.1.1 Host OS setup (before installing the operating system on guest OS)

If you plan to operate a guest OS as part of a cluster, set up the required disk devices, virtual bridges, virtual disks, user IDs, and guest OS initializations on the host OS.

Perform the following setup on the host OS after installing the operating system on the host OS and also before installing the operating system on the guest OS.

1. Creating the virtual disk

   When using a shared disk or mirroring among servers on a guest OS, create the virtual disk.

   Create the virtio-SCSI device or the virtio block device. For information on how to create them, see

   "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

   Note

   - For a disk to be added to a guest, specify with the by-id name.
   - Add a non-partitioned disk, not a partition or file, to the guest.

2. Installing and setting up related software

   Install and set up the software product (ETERNUS Multipath Driver) required for using system disk of the guest OS on the host OS. For how to install and set up the related software, see "Software Information" for ETERNUS Multipath Driver.

3. Mirroring the guest OS system disk

   To mirror the guest OS system disk, set up the mirrored volume of the local class or the shared class created on the host OS for the guest OS.

   See

   For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.1.2 Host OS setup (after installing the operating system on guest OS)

Perform the following setup after installing the operating system on guest OS.

1. Setting up the virtual disk

   For using a shared disk or mirroring among servers on a guest OS, you need to set up a virtual disk.

   The following shows the setup procedure for the virtual disk in a KVM environment.

   Using virtio-SCSI device as a shared disk

   1. Stop the guest OS.
   2. Add shareable and cache='none' to the virtio-SCSI device setting that is described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS. Additionally, correct the device attribute to 'lun' if any other value is set.

      ```
      # virsh edit guestname
      ```

      Example before change

      ```
        :
      <disk type='block' device='disk'>
        <driver name='qemu' type='raw'/>
        <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
        <target dev='sdh' bus='scsi'/>
        <address type='drive' controller='0' bus='0' target='0' unit='7'/>
      ```

```
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
    <target dev='sdh' bus='scsi'/>
    <shareable/>
    <address type='drive' controller='0' bus='0' target='0' unit='7'/>
  </disk>
  :
```

3. Start the guest OS.

Using virtio block device as a shared disk

1. Stop the guest OS.
2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar.
3. Click in the toolbar to display the detailed information of hardware.
4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
5. In the [Virtual disk] window, perform the following settings and click [Apply].

   - Select the Shareable check box.

   - Select 'none' for the cache model.
6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <shareable/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <shareable/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.

Using virtio-SCSI device for mirroring among servers

1. Stop the guest OS.
2. If the device attribute other than 'lun' is set in the settings of the virtio-SCSI device described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS, correct the device attribute to 'lun.'

```
# virsh edit guestname
```

Example before change

```
 :
<disk type='block' device='disk'>
  <driver name='qemu' type='raw'/>
  <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
  <target dev='sdh' bus='scsi'/>
  <address type='drive' controller='0' bus='0' target='0' unit='7'/>
</disk>
 :
```

Example after change

```
 :
<disk type='block' device='lun'>
  <driver name='qemu' type='raw'/>
  <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
  <target dev='sdh' bus='scsi'/>
  <address type='drive' controller='0' bus='0' target='0' unit='7'/>
</disk>
 :
```

3. Start the guest OS.

Using virtio block device for mirroring among servers

1. Stop the guest OS.
2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar
3. Click in the toolbar to display the detailed information of hardware.
4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
5. In the [Virtual disk] window, set the serial number on [Serial number] of [Advanced options], and click [Apply].
   The serial number should be a character string of up to 10 characters that does not duplicate in the virtual machine.
6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

Example before change

```
 :
<disk type='block' device='disk'>
  <driver name='qemu' type='raw'/>
  <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
  <target dev='vdb' bus='virtio'/>
  <serial>serial number</serial>
  <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
```

```
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <serial>serial number</serial>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.

9. On the guest OS, make sure that the by-id file of virtual disk exists.

   - Make sure that the by-id files exist in all virtio block devices used for mirroring among servers.

   - Make sure that the serial number set in step 5 is included in the file name of by-id file.

```
# ls -l /dev/disk/by-id
  :
lrwxrwxrwx 1 root root 9 Apr 18 08:44 virtio-disk001 -> ../../vdg
lrwxrwxrwx 1 root root 9 Apr 18 08:43 virtio-disk002 -> ../../vdh
  :                                          serial number
```

2. Setting up the virtual bridge (administrative LAN/public LAN/cluster interconnect)

   For the network interfaces, including the administrative LAN, public LAN and cluster interconnect, that are used by virtual domains, you need to set up virtual bridges for the virtual networks beforehand.

   (1) Setting up a virtual bridge for the administrative LAN

   Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

Note

For HWADDR, set the MAC address of the network interface card you are using.

Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
BOOTPROTO=static
IPADDR=xxx.xxx.xxx.xxx
NETMASK=xxx.xxx.xxx.xxx
ONBOOT=yes
```

Note

For IPADDR and NETMASK, set IP addresses and netmasks to connect to the external network. When IPv6 addresses are required, make the setting so that IPv6 addresses are assigned.

(2) Setting up a virtual bridge for the public LAN

Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

Note

For HWADDR, set the MAC address of the network interface card you are using.

Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
ONBOOT=yes
```

(3) Setting up a virtual bridge for the cluster interconnect

Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
BOOTPROTO=static
ONBOOT=yes
```

3. Setting the guest OS in the host OS (in a KVM environment)

   In a KVM environment, perform the following settings to stop the guest OS normally if the host OS is shut down by mistake while the guest OS running.

   Define the following two values in /etc/sysconfig/libvirt-guests. When values are already defined, change them to the following values:

   - ON_SHUTDOWN=shutdown
   - SHUTDOWN_TIMEOUT=300

   Specify the timeout duration (seconds) for shutdown of the guest OS to SHUTDOWN_TIMEOUT. Estimate the length of time for shutting down the guest OS and set the value. When multiple guest OSes are set, set the time whichever is greater. The above is an example when the time is 300 seconds (5 minutes).

   Note

   - When setting /etc/sysconfig/libvirt-guests, do not describe the setting values and comments on the same line.
   - When changing the settings in /etc/sysconfig/libvirt-guests during operation, make sure to follow the procedure in "9.4.1.3 Changing the Settings in /etc/sysconfig/libvirt-guests."

4. Creating a user ID

   Point

   This user ID will be the one used by the shutdown facility to log in to the host OS to force shut down the nodes. This user ID and password are used for configuring the shutdown facility.

   In a KVM environment, you need to set up a user for the shutdown facility for the guest OS control by PRIMECLUSTER.

   (1) Creating a general user ID (optional)

Create a general user ID (optional) for the shutdown facility in the host OS.

```
# useradd <User ID>
```

(2) Setting up the "sudo" command

You need to 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.

```
<User ID>   ALL=(root) NOPASSWD: ALL
```

Moreover, in order to permit the "sudo" execution without "tty", add "#" to the beginning of the following line to comment it out.

```
Defaults    requiretty
```

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.1.3 Guest OS setup

Perform the following procedure on all guest OSes of a cluster.

1. Setting up the network

   On the guest, you need to set up the network, including IP addresses of the public LAN and administrative LAN.

   This setup should be performed after installing the operating system.

   See

   For information on changing the public LAN and administrative LAN that the PRIMECLUSTER system uses, see "9.2 Changing the Network Environment."

   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.

2. Installing the bundled software on the guest OS

   Install the bundled software on the guest OS.

3. Initial setting

   Initialize the guest OS.

   See

   For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

4. Checking the guest domain name

   Check the guest domain names set on installation of the guest OSes. These names are used when setting up the Shutdown Facility. For information on how to check guest domain names, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.1.4 NTP setup (host OS and guest OS)

This setup is for synchronizing the time on each node comprising the cluster system, which is necessary when creating a cluster.

This setup should be performed on the host OS and guest OS before installing PRIMECLUSTER.

See

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.1.5 Installing PRIMECLUSTER on guest OSes

Install PRIMECLUSTER on guest OSes.

For details, see "3.3 PRIMECLUSTER Installation."

3.2.1.6 Checking and setting the kernel parameters

To operate the PRIMECLUSTER-related software, you need to edit the values of the kernel parameters based on the environment.

Perform this setup before restarting the installed PRIMECLUSTER.

Target node:

All the nodes on which PRIMECLUSTER is to be installed

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

Check "Setup (initial configuration)" of PRIMECLUSTER Designsheets and edit the value if necessary.

See

For information on the kernel parameters, see "3.1.7 Checking and Setting the Kernel Parameters."

Note

To enable modifications, you need to restart the operating system.

3.2.1.7 Installing and setting up applications

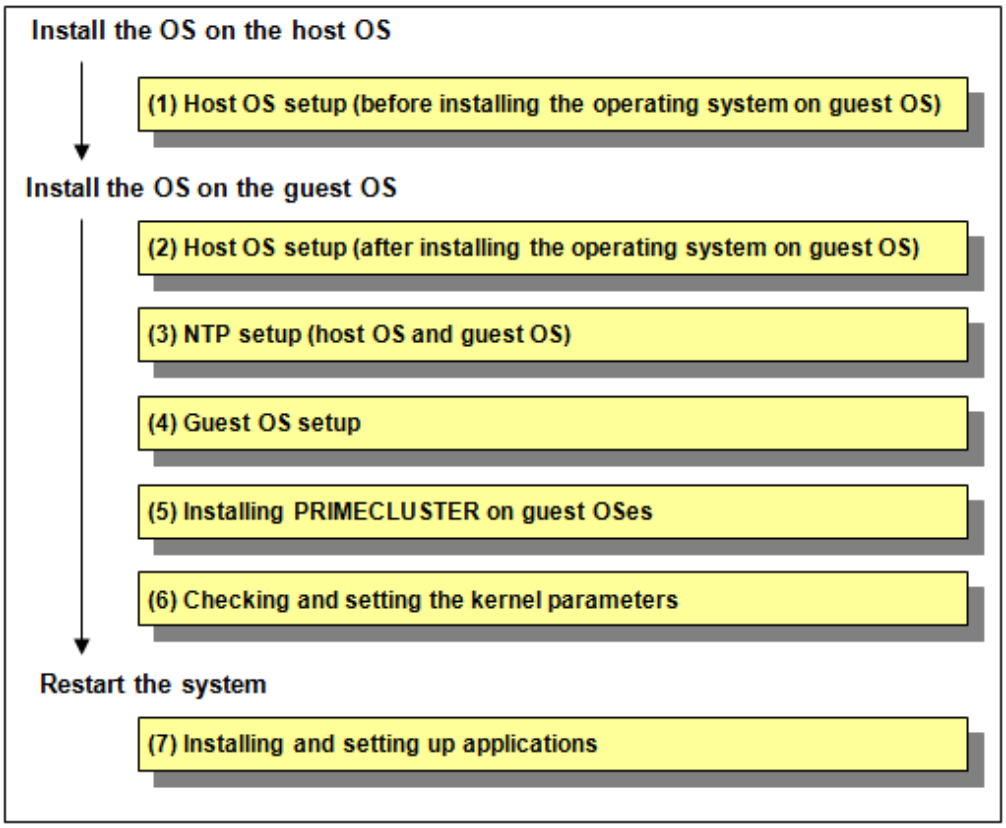
Install software products to be operated on the PRIMECLUSTER system and configure the environment as necessary.

For details, see "3.4 Installation and Environment Setup of Applications."

3.2.2 When building a cluster system between guest OSes on multiple host OSes without using Host OS failover function

This section describes how to install and set up related software when building a cluster system between guest OSes on multiple host OSes without using Host OS failover function.

Perform the steps shown in the figure below as necessary.



3.2.2.1 Host OS setup (before installing the operating system on guest OS)

If you plan to operate a guest OS as part of a cluster, set up the required disk devices, virtual bridges, virtual disks, user IDs, and guest OS initializations on the host OS.

Perform the following setup on the host OS after installing the operating system on the host OS and also before installing the operating system on the guest OS.

1. Creating the virtual disk

   When using a shared disk or mirroring among servers on a guest OS, create the virtual disk.

   Create the virtio-SCSI device or the virtio block device.

   For information on how to create them, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

   Note

   - For a disk to be added to a guest, specify with the by-id name.
   - Add a non-partitioned disk, not a partition or file, to the guest.

2. Installing and setting up related software

   Install and set up the software product (ETERNUS Multipath Driver) required for using system disk of the guest OS on the host OS. For how to install and set up the related software, see "Software Information" for ETERNUS Multipath Driver.

3. Mirroring the guest OS system disk

   To mirror the guest OS system disk, set up the local mirrored volume created on the host OS for the guest OS.

See

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.2.2 Host OS setup (after installing the operating system on guest OS)

Perform the following setup after installing the operating system on guest OS.

1. Setting up the virtual disk

   For using a shared disk or mirroring among servers on a guest OS, you need to set up a virtual disk.

   The following shows the setup procedure for the virtual disk in a KVM environment.

   Using virtio-SCSI device as a shared disk

   1. Stop the guest OS.
   2. Add shareable and cache='none' to the virtio-SCSI device setting that is described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS. Additionally, correct the device attribute to 'lun' if any other value is set.

      ```
      # virsh edit guestname
      ```

      Example before change

      ```
        :
      <disk type='block' device='disk'>
        <driver name='qemu' type='raw'/>
        <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
        <target dev='sdh' bus='scsi'/>
        <address type='drive' controller='0' bus='0' target='0' unit='7'/>
      </disk>
        :
      ```

      Example after change

      ```
        :
      <disk type='block' device='lun'>
        <driver name='qemu' type='raw' cache='none'/>
        <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
        <target dev='sdh' bus='scsi'/>
        <shareable/>
        <address type='drive' controller='0' bus='0' target='0' unit='7'/>
      </disk>
        :
      ```

   3. Start the guest OS.

   Using virtio block device as a shared disk

   1. Stop the guest OS.
   2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar.
   3. Click in the toolbar to display the detailed information of hardware.
   4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
   5. In the [Virtual disk] window, perform the following settings and click [Apply].

      - Select the Shareable check box.

      - Select 'none' for the cache model.

6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <shareable/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <shareable/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.

Using virtio-SCSI device for mirroring among servers

1. Stop the guest OS.
2. If the device attribute other than 'lun' is set in the settings of the virtio-SCSI device described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS, correct the device attribute to 'lun.'

```
# virsh edit guestname
```

Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
    <target dev='sdh' bus='scsi'/>
    <address type='drive' controller='0' bus='0' target='0' unit='7'/>
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
    <target dev='sdh' bus='scsi'/>
    <address type='drive' controller='0' bus='0' target='0' unit='7'/>
```

```
  </disk>
  :
```

3. Start the guest OS.

Using virtio block device for mirroring among servers

1. Stop the guest OS.
2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar
3. Click in the toolbar to display the detailed information of hardware.
4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
5. In the [Virtual disk] window, set the serial number on [Serial number] of [Advanced options], and click [Apply].
   The serial number should be a character string of up to 10 characters that does not duplicate in the virtual machine.
6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

   Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <serial>serial number</serial>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

   Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <serial>serial number</serial>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.
9. On the guest OS, make sure that the by-id file of virtual disk exists.

   - Make sure that the by-id files exist in all virtio block devices used for mirroring among servers.

   - Make sure that the serial number set in step 5 is included in the file name of by-id file.

```
# ls -l /dev/disk/by-id
  :
lrwxrwxrwx 1 root root 9 Apr 18 08:44 virtio-disk001 -> ../../vdg
lrwxrwxrwx 1 root root 9 Apr 18 08:43 virtio-disk002 -> ../../vdh
  :                                        serial number
```

2. Setting up the virtual bridge (administrative LAN/public LAN/cluster interconnect)

   For the network interfaces, including the administrative LAN, public LAN and cluster interconnect, that are used by virtual domains, you need to set up virtual bridges for the virtual networks beforehand.

   (1) Setting up a virtual bridge for the administrative LAN

   Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

   Note

   For HWADDR, set the MAC address of the network interface card you are using.

   Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
BOOTPROTO=static
IPADDR=xxx.xxx.xxx.xxx
NETMASK=xxx.xxx.xxx.xxx
ONBOOT=yes
```

   Note

   For IPADDR and NETMASK, set IP addresses and netmasks to connect to the external network. When IPv6 addresses are required, make the setting so that IPv6 addresses are assigned.

   (2) Setting up virtual bridges for the public LAN and cluster interconnect

   Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

   Note

   For HWADDR, set the MAC address of the network interface card you are using.

   Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
ONBOOT=yes
```

3. Setting the guest OS in the host OS (in a KVM environment)

   In a KVM environment, perform the following settings to stop the guest OS normally if the host OS is shut down by mistake while the guest OS running.

   Define the following two values in /etc/sysconfig/libvirt-guests. When values are already defined, change them to the following values:

- ON_SHUTDOWN=shutdown
- SHUTDOWN_TIMEOUT=300

Specify the timeout duration (seconds) for shutdown of the guest OS to SHUTDOWN_TIMEOUT. Estimate the length of time for shutting down the guest OS and set the value. When multiple guest OSes are set, set the time whichever is greater. The above is an example when the time is 300 seconds (5 minutes).

Note

- When setting /etc/sysconfig/libvirt-guests, do not describe the setting values and comments on the same line.
- When changing the settings in /etc/sysconfig/libvirt-guests during operation, make sure to follow the procedure in "9.4.1.3 Changing the Settings in /etc/sysconfig/libvirt-guests."

4. Creating a user ID

Point

This user ID will be the one used by the shutdown facility to log in to the host OS to force shut down the nodes. This user ID and password are used for configuring the shutdown facility.

In a KVM environment, you need to set up a user for the shutdown facility for the guest OS control by PRIMECLUSTER.

(1) Creating a general user ID (optional)

Create a general user ID (optional) for the shutdown facility in the host OS.

```
# useradd <User ID>
```

(2) Setting up the "sudo" command

You need to 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.

```
<User ID>   ALL=(root) NOPASSWD: ALL
```

Moreover, in order to permit the "sudo" execution without "tty", add "#" to the beginning of the following line to comment it out.

```
Defaults    requiretty
```

3.2.2.3 Guest OS setup

Perform the following procedure on all guest OSes of a cluster.

1. Setting up the network

   On the guest, you need to set up the network, including IP addresses of the public LAN and administrative LAN.

   This setup should be performed after installing the operating system.

See

For information on changing the public LAN and administrative LAN that the PRIMECLUSTER system uses, see "9.2 Changing the Network Environment."

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.

2. Installing the bundled software on the guest OS

   Install the bundled software on the guest OS.

3. Initial setting

   Initialize the guest OS.

See

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

4. Checking the guest domain name

   Check the guest domain names set on installation of the guest OSes. These names are used when setting up the Shutdown Facility. For information on how to check guest domain names, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.2.4 NTP setup (host OS and guest OS)

This setup is for synchronizing the time on each node comprising the cluster system, which is necessary when creating a cluster.

This setup should be performed on the host OS and guest OS before installing PRIMECLUSTER.

See

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.2.5 Installing PRIMECLUSTER on guest OSes

Install PRIMECLUSTER on guest OSes.

For details, see "3.3 PRIMECLUSTER Installation."

3.2.2.6 Checking and setting the kernel parameters

To operate the PRIMECLUSTER-related software, you need to edit the values of the kernel parameters based on the environment.

Perform this setup before restarting the installed PRIMECLUSTER.

Target node:

All the nodes on which PRIMECLUSTER is to be installed

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

Check "Setup (initial configuration)" of PRIMECLUSTER Designsheets and edit the value if necessary.

See

For information on the kernel parameters, see "3.1.7 Checking and Setting the Kernel Parameters."

Note

To enable modifications, you need to restart the operating system.

3.2.2.7 Installing and setting up applications

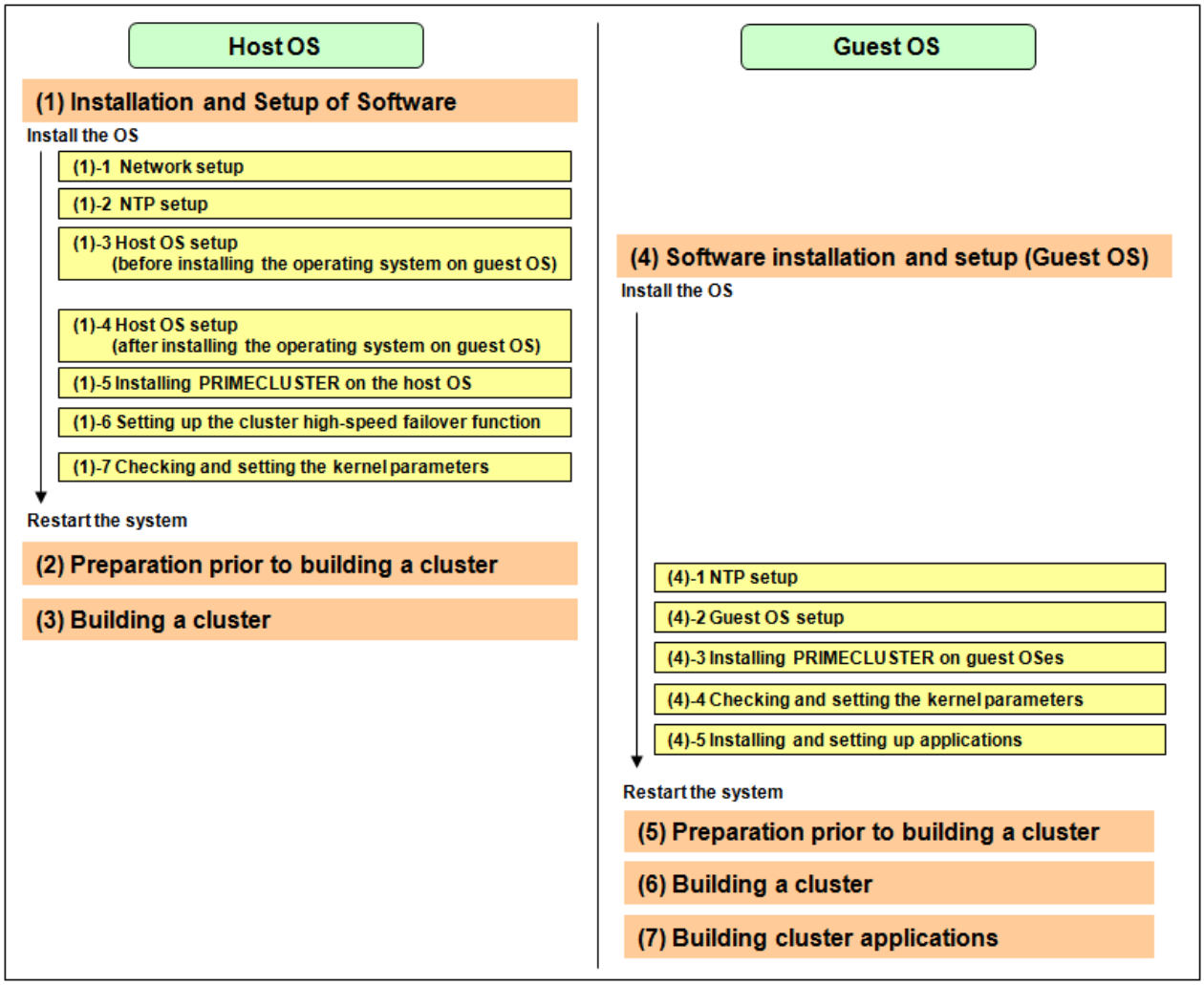
Install software products to be operated on the PRIMECLUSTER system and configure the environment as necessary.

For details, see "3.4 Installation and Environment Setup of Applications."

3.2.3 When building a cluster system between guest OSes on multiple host OSes using Host OS failover function

This section describes how to install and set up related software when building a cluster system between guest OSes on multiple host OSes using Host OS failover function.

Figure 3.1 Flow for building a cluster system when using Host OS failover function



3.2.3.1 Installation and Setup of Software (Host OS)

After installing the PRIMECLUSTER-related software, you need to set up the operating system, hardware, and so on that will be used and administered.

Perform the following as necessary.

3.2.3.1.1 Network setup

In order for the host OS to work as the cluster, network setup is required.

3.2.3.1.2 NTP setup

This setup is for synchronizing the time on each node comprising the cluster system, which is necessary when creating a cluster.

This setup should be performed before installing PRIMECLUSTER.

3.2.3.1.3 Host OS setup (before installing the operating system on guest OS)

If you plan to operate a guest OS as part of a cluster, set up the required disk devices, virtual bridges, virtual disks, user IDs, and guest OS initializations on the host OS.

Perform the following setup on the host OS after installing the operating system on the host OS and also before installing the operating system on the guest OS.

1. Creating the virtual disk

   When using a shared disk or mirroring among servers on a guest OS, create the virtual disk.

   Create the virtio-SCSI device or the virtio block device. For information on how to create them, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

   Note

   - For a disk to be added to a guest, specify with the by-id name.
   - Add a non-partitioned disk, not a partition or file, to the guest.

2. Installing and setting up related software

   Install and set up the software product (ETERNUS Multipath Driver) required for using system disk of the guest OS on the host OS. For how to install and set up the related software, see "Software Information" for ETERNUS Multipath Driver.

3. Mirroring the guest OS system disk

   To mirror the guest OS system disk, set up the mirrored volume of the local class or the shared class created on the host OS for the guest OS.

   See

   For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)

Perform this setup on the host OS according to the following procedure after installing the operating system on the host OS and the guest OSes.

1. Setting up the virtual disk

   For using a shared disk or mirroring among servers on a guest OS, you need to set up a virtual disk.

   The following shows the setup procedure for the virtual disk in a KVM environment.

   Using virtio-SCSI device as a shared disk

   1. Stop the guest OS.
   2. Add shareable and cache='none' to the virtio-SCSI device setting that is described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS. Additionally, correct the device attribute to 'lun' if any other value is set.

      ```
      # virsh edit guestname
      ```

Example before change

```
 :
<disk type='block' device='disk'>
  <driver name='qemu' type='raw'/>
  <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
  <target dev='sdh' bus='scsi'/>
  <address type='drive' controller='0' bus='0' target='0' unit='7'/>
</disk>
 :
```

Example after change

```
 :
<disk type='block' device='lun'>
  <driver name='qemu' type='raw' cache='none'/>
  <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
  <target dev='sdh' bus='scsi'/>
  <shareable/>
  <address type='drive' controller='0' bus='0' target='0' unit='7'/>
</disk>
 :
```

3. Start the guest OS.

Using virtio block device as a shared disk

1. Stop the guest OS.
2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar.
3. Click in the toolbar to display the detailed information of hardware.
4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
5. In the [Virtual disk] window, perform the following settings and click [Apply].

   - Select the Shareable check box.

   - Select 'none' for the cache model.
6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

Example before change

```
 :
<disk type='block' device='disk'>
  <driver name='qemu' type='raw' cache='none'/>
  <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
  <target dev='vdb' bus='virtio'/>
  <shareable/>
  <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
</disk>
 :
```

Example after change

```
 :
<disk type='block' device='lun'>
```

```
    <driver name='qemu' type='raw' cache='none'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <shareable/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.

Using virtio-SCSI device for mirroring among servers

1. Stop the guest OS.
2. If the device attribute other than 'lun' is set in the settings of the virtio-SCSI device described in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS, correct the device attribute to 'lun.'

```
# virsh edit guestname
```

Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
    <target dev='sdh' bus='scsi'/>
    <address type='drive' controller='0' bus='0' target='0' unit='7'/>
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-36000b5d0006a0000006a1296001f0000'/>
    <target dev='sdh' bus='scsi'/>
    <address type='drive' controller='0' bus='0' target='0' unit='7'/>
  </disk>
  :
```

3. Start the guest OS.

Using virtio block device for mirroring among servers

1. Stop the guest OS.
2. Select the stopped guest OS with the Virtual Machine Manager and click the [Open] button in the toolbar
3. Click in the toolbar to display the detailed information of hardware.
4. Select a virtual disk (VirtIO Disk) from the hardware list in the left.
5. In the [Virtual disk] window, set the serial number on [Serial number] of [Advanced options], and click [Apply].
   The serial number should be a character string of up to 10 characters that does not duplicate in the virtual machine.
6. Check the version of the libvirt package on the host OS by using the rpm(8) command.

```
# rpm -qi libvirt
```

7. If the version of the libvirt package is libvirt-0.9.4-23.el6_2.4 or later, change the device attribute from disk to lun, which is set in the guest setting file (/etc/libvirt/qemu/*guestname*.xml) on the host OS.

```
# virsh edit guestname
```

Example before change

```
  :
  <disk type='block' device='disk'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <serial>serial number</serial>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

Example after change

```
  :
  <disk type='block' device='lun'>
    <driver name='qemu' type='raw'/>
    <source dev='/dev/disk/by-id/scsi-1FUJITSU_30000085002B'/>
    <target dev='vdb' bus='virtio'/>
    <serial>serial number</serial>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x06' function='0x0'/>
  </disk>
  :
```

8. Start the guest OS.

9. On the guest OS, make sure that the by-id file of virtual disk exists.

   - Make sure that the by-id files exist in all virtio block devices used for mirroring among servers.

   - Make sure that the serial number set in step 5 is included in the file name of by-id file.

```
# ls -l /dev/disk/by-id
  :
lrwxrwxrwx 1 root root 9 Apr 18 08:44 virtio-disk001 -> ../../vdg
lrwxrwxrwx 1 root root 9 Apr 18 08:43 virtio-disk002 -> ../../vdh
  :                                          serial number
```

2. Setting up the virtual bridge (administrative LAN/public LAN/cluster interconnect)

   For the network interfaces, including the administrative LAN, public LAN and cluster interconnect, that are used by virtual domains, you need to set up virtual bridges for the virtual networks beforehand.

   (1) Setting up a virtual bridge for the administrative LAN

   Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

Note

For HWADDR, set the MAC address of the network interface card you are using.

Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
BOOTPROTO=static
IPADDR=xxx.xxx.xxx.xxx
NETMASK=xxx.xxx.xxx.xxx
ONBOOT=yes
```

Note

For IPADDR and NETMASK, set IP addresses and netmasks to connect to the external network. When IPv6 addresses are required, make the setting so that IPv6 addresses are assigned.

(2) Setting up virtual bridges for the public LAN and cluster interconnect

Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

```
DEVICE=ethX
HWADDR=XX:XX:XX:XX:XX:XX
BOOTPROTO=none
ONBOOT=yes
BRIDGE=brX
```

Note

For HWADDR, set the MAC address of the network interface card you are using.

Create the interface setting file, /etc/sysconfig/network-scripts/ifcfg-brX, for the virtual bridge.

```
DEVICE=brX
TYPE=Bridge
ONBOOT=yes
```

3. Setting the guest OS in the host OS (in a KVM environment)

   In a KVM environment, perform the following settings to stop the guest OS normally if the host OS is shut down by mistake while the guest OS running.

   Define the following two values in /etc/sysconfig/libvirt-guests. When values are already defined, change them to the following values:

   - ON_SHUTDOWN=shutdown
   - SHUTDOWN_TIMEOUT=300

   Specify the timeout duration (seconds) for shutdown of the guest OS to SHUTDOWN_TIMEOUT. Estimate the length of time for shutting down the guest OS and set the value. When multiple guest OSes are set, set the time whichever is greater. The above is an example when the time is 300 seconds (5 minutes).

   ### Note

   - When setting /etc/sysconfig/libvirt-guests, do not describe the setting values and comments on the same line.
   - When changing the settings in /etc/sysconfig/libvirt-guests during operation, make sure to follow the procedure in "9.4.1.3 Changing the Settings in /etc/sysconfig/libvirt-guests."

4. Creating a user ID

   ### Point

   This user ID will be the one used by the shutdown facility to log in to the host OS to force shut down the nodes. This user ID and password are used for configuring the shutdown facility.

   KVM environment

   In a KVM environment, you need to set up a user for the shutdown facility for the guest OS control by PRIMECLUSTER.

   (1) Creating a general user ID (optional)

Create a general user ID (optional) for the shutdown facility in the host OS.

```
# useradd <User ID>
```

(2) Setting up the "sudo" command

You need to 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.

```
<User ID>   ALL=(root) NOPASSWD: ALL
```

Moreover, in order to permit the "sudo" execution without "tty", add "#" to the beginning of the following line to comment it out.

```
Defaults    requiretty
```

3.2.3.1.5 Installing PRIMECLUSTER on the host OS

Install PRIMECLUSTER on the host OS.

For details, see "3.3 PRIMECLUSTER Installation."

3.2.3.1.6 Setting up the cluster high-speed failover function

You need to configure software and hardware that enables cluster high-speed failover after installing the OS and PRIMECLUSTER.

For details, see "3.1.6 Setting Up the Cluster High-Speed Failover Function."

3.2.3.1.7 Checking and setting the kernel parameters

To operate the PRIMECLUSTER-related software, you need to edit the values of the kernel parameters based on the environment.

Perform this setup before restarting the installed PRIMECLUSTER.

Target node:

All the nodes on which PRIMECLUSTER is to be installed

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

Check "Setup (initial configuration)" of PRIMECLUSTER Designsheets and edit the value if necessary.

See

For information on the kernel parameters, see "3.1.7 Checking and Setting the Kernel Parameters."

Note

To enable modifications, you need to restart the operating system.

3.2.3.2 Preparation prior to building a cluster (Host OS)

Before building a cluster, preparation work is required in the host OS, such as starting up the Web-Based Admin View screen. For details, see "Chapter 4 Preparation Prior to Building a Cluster."

3.2.3.3 Building a cluster (Host OS)

Build a cluster of PRIMECLUSTER on the host OS. For details, see "Chapter 5 Building a Cluster." To build a cluster, perform the procedures described in "5.1.1 Setting Up CF and CIP" and "5.1.2 Setting up the Shutdown Facility." Also, for the shutdown facility, set shutdown agent in the same way as the setting between natives. See "5.1.2 Setting up the Shutdown Facility," and check the hardware model/configuration to set up the appropriate shutdown agent.

Note

- After setting CF, set the timeout value of the cluster system on the host OS to 20 seconds. For details on the setup, refer to "11.3.1 Changing Time to Detect CF Heartbeat Timeout."
- Share the cluster interconnect LAN of the host OS with other guest OSes, and separate networks for each cluster system with Virtual LAN.

3.2.3.4 Software installation and setup (Guest OS)

After building a cluster on the host OS, install the PRIMECLUSTER-related software, and set up the OS and hardware for installing and operating PRIMECLUSTER.

Perform the following as necessary.

3.2.3.4.1 Guest OS setup

Perform the following procedure on all guest OSes of a cluster.

1. Setting up the network

   On the guest, you need to set up the network, including IP addresses of the public LAN and administrative LAN.

   This setup should be performed after installing the operating system.

   #### See

   For information on changing the public LAN and administrative LAN that the PRIMECLUSTER system uses, see "9.2 Changing the Network Environment."

   #### 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.

2. Installing the bundled software on the guest OS

   Install the bundled software on the guest OS.

3. Initial setting

   Initialize the guest OS.

   #### See

   For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

4. Creating an user ID (PRIMEQUEST only)

   Create an user ID used with the host OS failover function of PRIMEQUEST.

   #### Point

   This user ID is used by the host OS failover function to log in to the guest OS. This user ID and password are used for setting the host OS failover function.

1. Creating a general user ID (optional)

   Create a general user ID (optional) for the host OS failover function in the guest OS.

   ```
   # useradd <User ID>
   ```

2. Setting up the "sudo" command

   You need to set up the "sudo" command so that the general user ID (optional) for the host OS failover function 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.

   ```
   <User ID> ALL=(root) NOPASSWD: ALL
   ```

   Moreover, in order to permit the "sudo" execution without "tty", add "#" to the beginning of the following line to comment it out.

   ```
   Defaults requiretty
   ```

5. Checking the guest domain name

   Check the guest domain names set on installation of the guest OSes. These names are used when setting up the Shutdown Facility. For information on how to check guest domain names, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.3.4.2 NTP setup (Guest OS)

This setup is for synchronizing the time on each node comprising the cluster system, which is necessary when creating a cluster.

This setup should be performed on the guest OS before installing PRIMECLUSTER.

See

For details on settings, see "Red Hat Enterprise Linux 6 Virtualization Administration Guide" or "Red Hat Enterprise Linux 7 Virtualization Deployment and Administration Guide."

3.2.3.4.3 Installing PRIMECLUSTER on guest OSes

Install PRIMECLUSTER on guest OSes.

For details, see "3.3 PRIMECLUSTER Installation."

3.2.3.4.4 Checking and setting the kernel parameters

To operate the PRIMECLUSTER-related software, you need to edit the values of the kernel parameters based on the environment.

Perform this setup before restarting the installed PRIMECLUSTER.

Target node:

All the nodes on which PRIMECLUSTER is to be installed

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

Check "Setup (initial configuration)" of PRIMECLUSTER Designsheets and edit the value if necessary.

See

For information on the kernel parameters, see "3.1.7 Checking and Setting the Kernel Parameters."

Note

To enable modifications, you need to restart the operating system.

3.2.3.4.5 Installing and setting up applications

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

For details, see "3.4 Installation and Environment Setup of Applications."

3.2.3.5 Preparation prior to building a cluster (Guest OS)

Before building a cluster, preparation work is required in the host OS, such as starting up the Web-Based Admin View screen. For details, see "Chapter 4 Preparation Prior to Building a Cluster."

3.2.3.6 Building a Cluster (Guest OS)

Build a cluster on the guest OS. For details on each item, see "Chapter 5 Building a Cluster."

Note

- Share the cluster interconnect LAN of the guest OS with other guest OSes and the host OS, and separate networks for each cluster system with Virtual LAN.
- Do not change a timeout value of the guest OS from 10 seconds at the CF setting.
- For setup policy for survival priority, see "Survival scenarios" in "5.1.2 Setting up the Shutdown Facility."

3.2.3.7 Building cluster applications (Guest OS)

Create cluster applications on the guest OS. For details, see "Chapter 6 Building Cluster Applications."

Note

When creating a cluster application for a guest OS, do not set the ShutdownPriority attribute of RMS.

3.3 PRIMECLUSTER Installation

You can install PRIMECLUSTER using an installation script.

The installation script is also called the CLI Installer.

It is used to install PRIMECLUSTER on a system in which Linux(R) software and related Fujitsu software have been installed. It is also used for installation of the cluster management server.

See

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

After PRIMECLUSTER was installed, perform the following settings so that the CF modules and the GDS modules are not incorporated to an initial RAM disk (initramfs) for kdump:

- RHEL6 environment

  1. Add CF modules (cf, symsrv) and GDS modules (sfdsk, sfdsk_lib, sfdsklog, sfdsksys) to the setting of blacklist for /etc/kdump.conf.

Example

- If the setting of blacklist has been described in /etc/kdump.conf

  Before change:

```
blacklist kvm-intel
```

  After change (if GDS is not installed):

```
blacklist kvm-intel cf symsrv
```

  After change (if GDS is installed):

```
blacklist kvm-intel cf symsrv sfdsk sfdsk_lib sfdsklog sfdsksys
```

- If the setting of blacklist has not been described in /etc/kdump.conf

  Add the following line.

  If GDS is not installed:

```
blacklist cf symsrv
```

  If GDS is installed:

```
blacklist cf symsrv sfdsk sfdsk_lib sfdsklog sfdsksys
```

2. Restart the system or the service of kdump.

Note

In physical environment of PRIMERGY, PRIMECLUSTER sets kdump_post in /etc/kdump.conf. Do not set kdump_post in any other environments than PRIMECLUSTER as only one kdump_post is active in /etc/kdump.conf.

PRIMECLUSTER adds the following settings to /etc/kdump.conf when OS is started.

```
### poffinhibit configuration(automatically inserted) ###
kdump_post /etc/poffinhibit/unlock.sh
extra_modules poffinhibit_ipdv poffinhibit
options poffinhibit_ipdv SmBiosAddrParam=0x0
### poffinhibit configuration end ###
```

- RHEL7 environment

1. Add the following description to the line of KDUMP_COMMANDLINE_APPEND in /etc/sysconfig/kdump by the same line.

   If GDS is not installed:

```
rd.driver.blacklist=cf,symsrv
```

   If GDS is installed:

```
rd.driver.blacklist=cf,symsrv,sfdsk,sfdsksys,sfdsklog,sfdsk_lib
```

Example

Before change:

```
KDUMP_COMMANDLINE_APPEND="irqpoll ... acpi_no_memhotplug"
```

After change (if GDS is not installed):

```
KDUMP_COMMANDLINE_APPEND="irqpoll ... acpi_no_memhotplug rd.driver.blacklist=cf,symsrv"
```

After change (if GDS is installed):

```
KDUMP_COMMANDLINE_APPEND="irqpoll ... acpi_no_memhotplug
rd.driver.blacklist=cf,symsrv,sfdsk,sfdsksys,sfdsklog,sfdsk_lib"
```

2. Restart the system or the service of kdump.

See

For details on kdump, see the Linux documentation.

Note

When using the ntpdate service to adjust the time at OS startup in RHEL7, rapid time adjustment may be performed by the ntpdate service after each PRIMECLUSTER service is started. Therefore, considering the startup order of systemd, set the time adjustment by the ntpdate service to be completed before each PRIMECLUSTER service below is started.

- fjsvwvbs.service
- smawcf.service
- fjsvsdx.service (if using GDS)

The operation procedure is as follows.

You can skip these steps when not using the ntpdate service.

Operation Procedure:

Perform the following procedure on all the nodes.

1. Create the directories.

```
# mkdir /etc/systemd/system/fjsvwvbs.service.d
# chmod 755 /etc/systemd/system/fjsvwvbs.service.d
# mkdir /etc/systemd/system/smawcf.service.d
# chmod 755 /etc/systemd/system/smawcf.service.d
```

When using GDS, also create the following directory.

```
# mkdir /etc/systemd/system/fjsvsdx.service.d
# chmod 755 /etc/systemd/system/fjsvsdx.service.d
```

2. Create the configuration files (ntp.conf) in the created directories.

```
# touch /etc/systemd/system/fjsvwvbs.service.d/ntp.conf
# chmod 644 /etc/systemd/system/fjsvwvbs.service.d/ntp.conf
# touch /etc/systemd/system/smawcf.service.d/ntp.conf
# chmod 644 /etc/systemd/system/smawcf.service.d/ntp.conf
```

When using GDS, also create the following configuration file.

```
# touch /etc/systemd/system/fjsvsdx.service.d/ntp.conf
# chmod 644 /etc/systemd/system/fjsvsdx.service.d/ntp.conf
```

3. Add the following setting to each configuration file (ntp.conf) created in step 2.

```
[Unit]
After=time-sync.target
```

4. Reflect the setting of start/stop order of the PRIMECLUSTER services.

```
# systemctl daemon-reload
```

5. Check the setting of start/stop order of the PRIMECLUSTER services. Make sure that time-sync.target is included.

```
# systemctl show fjsvwvbs.service | grep "After="
# systemctl show smawcf.service | grep "After="
```

When using GDS, also check the following service.

```
# systemctl show fjsvsdx.service | grep "After="
```

If the time-sync.target is not included, make sure that settings step 1 to 4 are correctly done.

3.4 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 "Appendix A PRIMECLUSTER Products."
- For details on installing applications, see the manuals, Software Release Guides 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 "Chapter 2 Operation Environment" in the Installation Guide for PRIMECLUSTER.

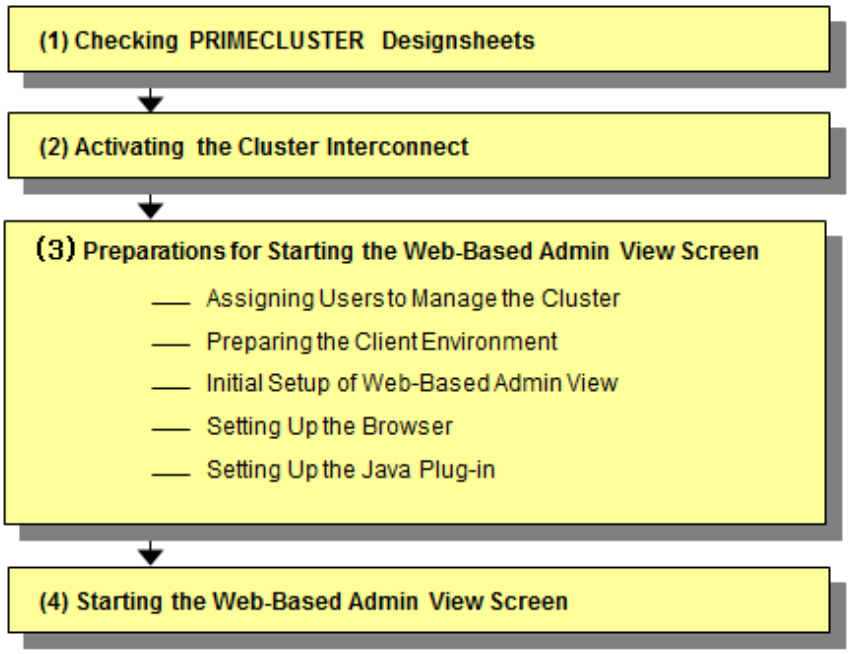


Table 4.1 Operation procedure and manual reference location for starting the Web-Based Admin View screen

	Work item	Execution Node	Required/ Optional	Manual reference location *
(1)	4.1 Checking PRIMECLUSTER Designsheets	-	Required	
(2)	4.2 Activating the Cluster Interconnect	All nodes	Required	
(3)	4.3 Preparations for Starting the Web-Based Admin View Screen			
	4.3.1 Assigning Users to Manage the Cluster	Cluster node	Required	
	4.3.2 Preparing the Client Environment	Client	Required	**WEB** "3.1.2 Prerequisite client environment"
	4.3.3 Initial Setup of Web-Based Admin View	Cluster node	Required	
	4.3.4 Setting Up the Browser	Client	Required	**WEB** "3.1.3.1 Preparing the Web browser"
	4.3.5 Setting Up the Java Plug-in	Client	Required	**WEB** "3.1.3.2 Required for the Web Browser Environment"

	Work item	Execution Node	Required/ Optional	Manual reference location *
(4)	4.4 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 PRIMECLUSTER Designsheets

Make certain of filling PRIMECLUSTER Designsheets. If there is missing information, you should specify values and complete PRIMECLUSTER Designsheets.

4.2 Activating the Cluster Interconnect

As preparation for setting up CF, execute the following operation:

1. Edit the /etc/sysconfig/network-scripts/ifcfg-ethX file.

   Edit "ONBOOT" of the /etc/sysconfig/network-scripts/ifcfg-ethX file as follows:

   ```
   ONBOOT=yes
   ```

   Note

   ethX indicates a network interface that is used for the cluster interconnect.

   A number is specified in X.

2. Confirm the cluster interconnect.

   Confirm the state of the interface for the interconnect with the following command.

   [RHEL6]

   ```
   # ifconfig <relevant interface>
   ```

   If the state flag with the above command is not "UP", execute the following command to confirm if "UP" is set.

   ```
   # ifconfig <relevant interface> up
   ```

   [RHEL7]

   ```
   # ip addr show <relevant interface>
   ```

   If the state flag with the above command is not "UP", execute the following command to confirm if "UP" is set.

   ```
   # ip link set <relevant interface> up
   ```

4.3 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:

   - 4.3.1 Assigning Users to Manage the Cluster
   - 4.3.2 Preparing the Client Environment

- 4.3.3 Initial Setup of Web-Based Admin View

2. Web environment preparation

   You need to set up the browser first.

   - 4.3.4 Setting Up the Browser
   - 4.3.5 Setting Up the Java Plug-in

4.3.1 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	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 allow 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.

The following user groups: wvroot, clroot, cladmin, and clmon are automatically created at the installation of PRIMECLUSTER. Since the sdxroot user group cannot be automatically created, if you want to grant the privileges to users for operating the GDS management view, create it on each primary and secondary management servers. The users must also be assigned to these groups. The Web-Based Admin View group membership should maintain consistency among all management servers associated with a specific cluster system.

To register the above group to a user, you should register the group as a Supplemental Group. To register a group as a Supplemental Group, use the usermod(8) or useradd(8) command.

- To add a user group to a registered user

```
# usermod -G wvroot username
```

- To register a new user

```
# useradd -G wvroot username
```

When you register a new user, use the passwd(8) command to set a password.

```
# passwd username
```

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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

When creating the wvroot user group automatically at installation of PRIMECLUSTER, GID (ID number of the group) is not specified. Even if GID is not changed, it does not affect the behavior of the operation management products running on Web-Based Admin View; however, if you want to specify the same GID between the primary management server and the secondary management server, execute the groupadd(8) command or the groupmod(8) command:

- When specifying GID before installing PRIMECLUSTER and then creating the wvroot user group

```
# groupadd -g <GID> wvroot
```

- When changing GID of the wvroot user group after installing PRIMECLUSTER

```
# groupmod -g <GID> wvroot
```

4.3.2 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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

4.3.3 Initial Setup of Web-Based Admin View

Note

One management server of Web-Based Admin View cannot be set to monitor nodes in multiple cluster systems.

One cluster system must be monitored by one management server of Web-Based Admin View.

4.3.3.1 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 of the primary management server and the secondary management server.

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

   - In the case of multiple-node cluster operation

     Execute the command above on all the management servers and nodes, referring to the example.
     Example: Set "Primary" - "IP address" and "Secondary" - "IP address" found on "Setup (initial configuration)" of PRIMECLUSTER Designsheets.

```
# /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 of the primary management server and the secondary management server, specify the IP address of the own node.

  Example:

```
# /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 addition, no value is displayed in Web-Based Admin View on the secondary management server.

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

   - For RHEL6

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

   - For RHEL7

```
# /etc/init.d/fjsvwvbs start
# /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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

Note

- For making entries to /etc/hosts in Japanese, use EUC encoding and set "ja" for the system requirements variable "lang" for Web-Based Admin View.

  For further details on the Web-Based Admin View language settings, refer to "4.3.3.3 Setting the Web-Based Admin View Language."

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

4.3.3.2 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" 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  3m41s
primaryServer   Sessions: 0
primaryServer   Nodes: 2
        10.20.30.40 node1      Linux-2.4.9-e.8enterprise       3m36s
        10.20.30.41 node2      Linux-2.4.9-e.8enterprise       2m58s
secondaryServer 10.20.30.41 node2  http=10.20.30.41     Run  2m46s
secondaryServer Sessions: 0
```

```
secondaryServer Nodes: 2
        10.20.30.40 node1       Linux-2.4.9-e.8enterprise       2m41s
        10.20.30.41 node2       Linux-2.4.9-e.8enterprise       2m23s
```

Make sure that the information of the nodes connected to each management server is properly displayed. If the information is not properly displayed, check the following points:

- 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 "2.4 Initial Setup of Web-Based Admin View" in "PRIMECLUSTER Web-Based Admin View Operation Guide" and check the parameter settings.
- Communication with the management servers may be blocked by firewall. When using firewalld, iptables, or ip6tables as firewall, permit the communication with the port numbers used by Web-Based Admin View. For details , see "Appendix K Using Firewall."

For details on the "wvstat" command, see the manual page.

4.3.3.3 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 the 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. 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> <setting_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.
   - For RHEL6

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

- For RHEL7

```
# /etc/init.d/fjsvwvbs 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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

4.3.4 Setting Up the Browser

Set up a Web browser on the clients.

See

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

4.3.5 Setting Up the Java Plug-in

To use Web-Based Admin View, install Java Plug-in on the client.

Note

If multiple IP addresses are assigned to the client, the following log display functions of Cluster Admin may not be used:

- CF message (cf tab)
- switchlog (rms tab)
- Application log (rms tab)

When using the log display functions of Cluster Admin, take the following steps to set Java runtime parameter.

1. Select [Start] - [Control Panel] - [Java] - [Java] tab - [View].

   Java Runtime Environment Settings window is displayed.

2. Select [User] tab of Java Runtime Environment Settings window and add the following parameter to the runtime parameter.

   ```
   -Djava.rmi.server.hostname=<IP address>
   ```

   ```
   <IP address>
   Specify the IP address that can communicate with the primary management server and the secondary
   management server.
   ```

   Example: When specifying 10.10.10.10 to the IP address

   ```
   -Djava.rmi.server.hostname=10.10.10.10
   ```

See

For details on the supported Java Plug-in versions, see "3.1.2 Prerequisite client environment" in "PRIMECLUSTER Web-Based Admin View Operation Guide." For instructions on setting up the Java Plug-in, see "3.1.3.2 Required for the Web browser environment" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

4.4 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>

   IP address or 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."

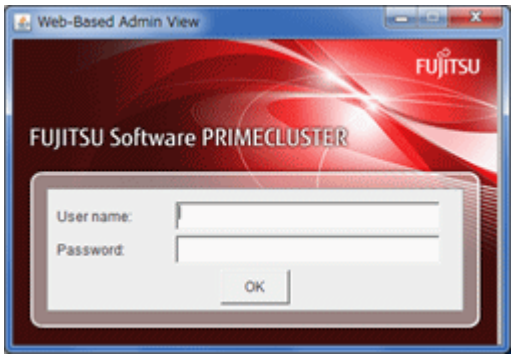
   If the port number has been changed, specify the up-to-date number.
   For instructions on changing the http port number, see "7.2.1 http port number" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

   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 brackets "[ ]".
     (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 "PRIMECLUSTER Web-Based Admin View Operation Guide."
   - For information on the IP address or host name (httpip) used by clients, see "PRIMECLUSTER Web-Based Admin View Operation Guide."
   - When a URL is to be registered for the bookmark of a browser, register "Plugin.cgi."

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> in the upper 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 "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 cluster 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 "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 the corrective action No.3 of "Symptom 17" in "B.1 Corrective action" of "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" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

4.5 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**.

4.5.1 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 categorized into the following two types:

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

b. Management screens and manuals of operation management products that are provided by non-PRIMECLUSTER products (such as Symfoware)

The following operations are possible for the menu of a:

- Operation management product name (PRIMECLUSTER)

  You can operate the screen of the operation management product.

  - Global Cluster Services (CF, CRM, RMS, SIS)
  - Global Disk Services
  - Global File Services

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

- 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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

- Manual

  The PRIMECLUSTER online manual is displayed.

The following operations are possible for the menu of b:

- Operation management product name (non-PRIMECLUSTER)

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

  For example, the following menu category is displayed in the environment where Symfoware has been installed.

  - Symfoware

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

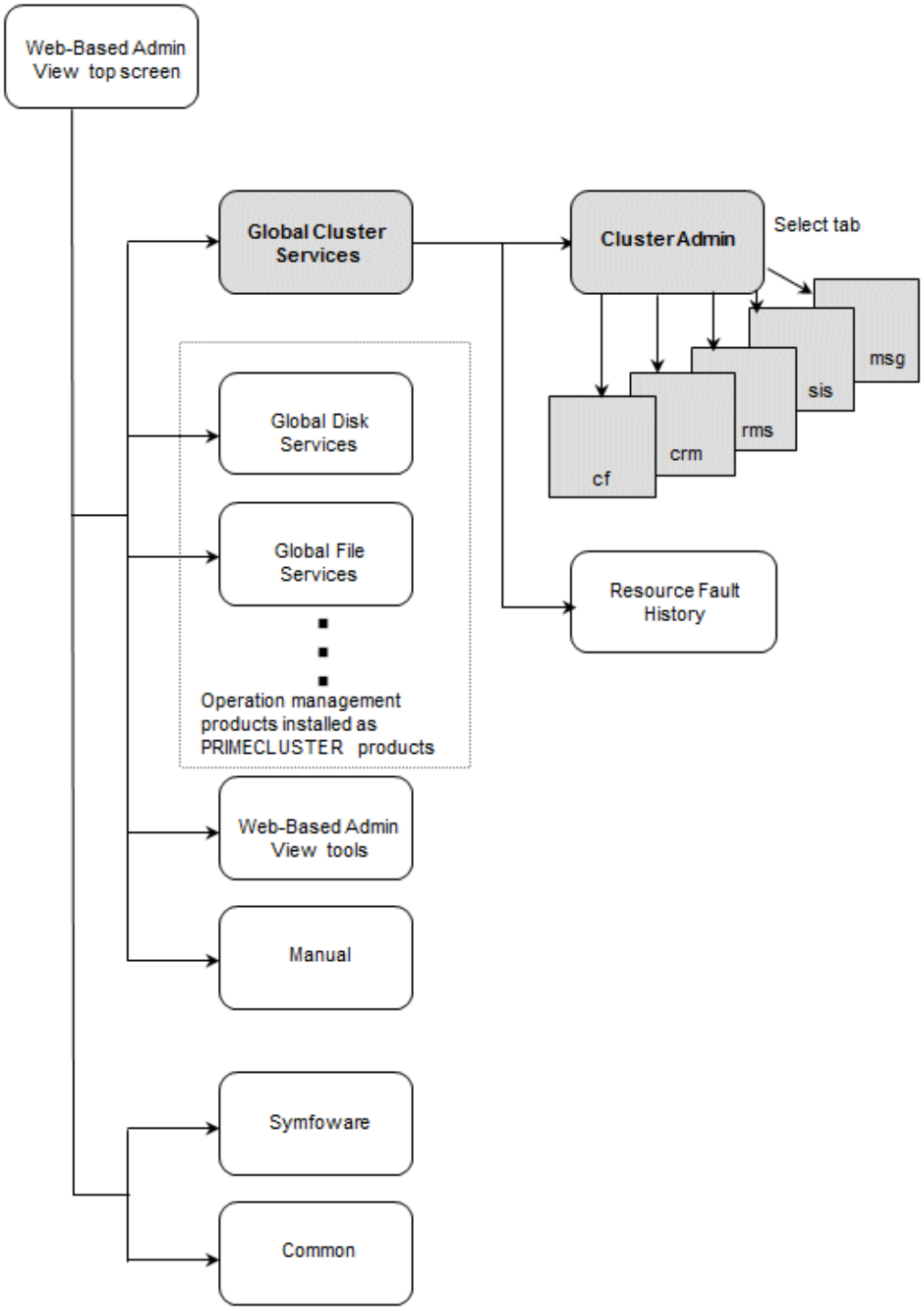
- Common

You can refer to manuals that are available as online manuals.

For details, see "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.



At the Cluster Admin screen, you can switch the window by clicking the following tabs:

- cf: Cluster Foundation

- crm: Cluster Resource Management
- rms: Reliant Monitor Services
- sis: Scalable Internet Services
- msg: Message

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

Note

SIS cannot be used with PRIMECLUSTER 4.4.

4.5.2 Global Cluster Services Menu Functions

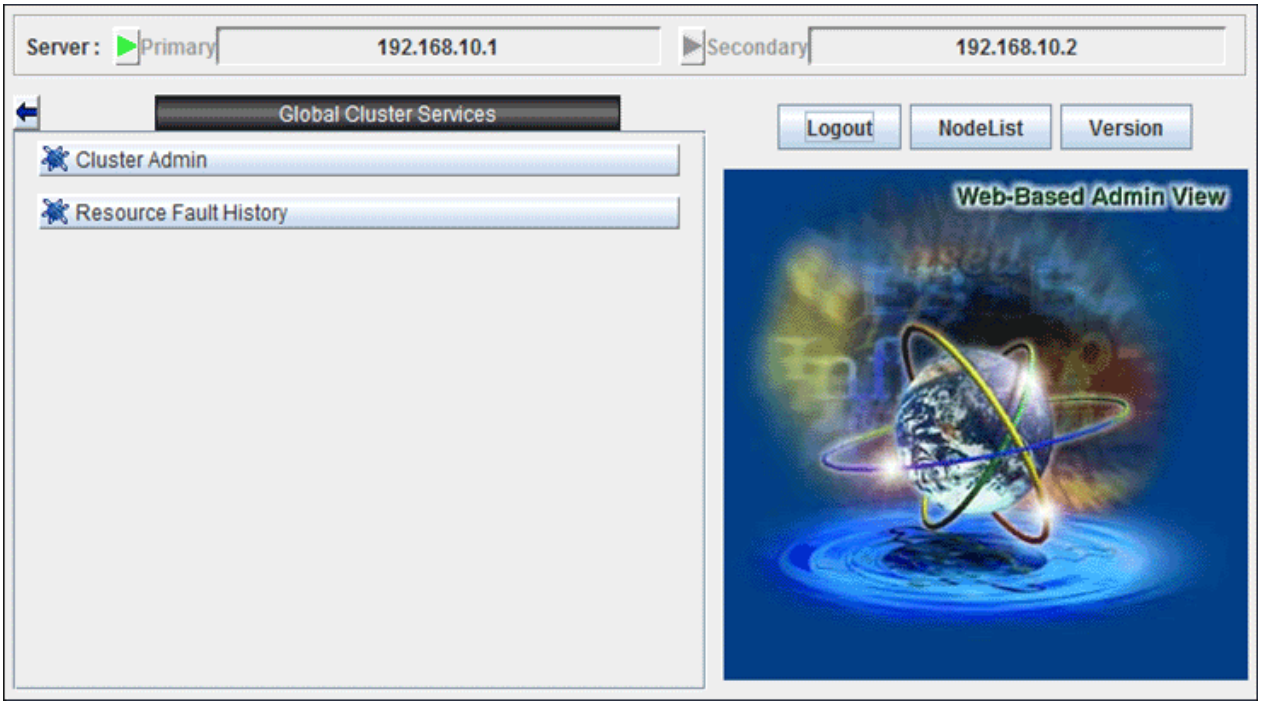
Display procedure

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.

- **Resource Fault History**

  This function allows you to display the resource fault history. For details, see "C.2.2 Resource Fault History."

4.5.3 Cluster Admin Functions

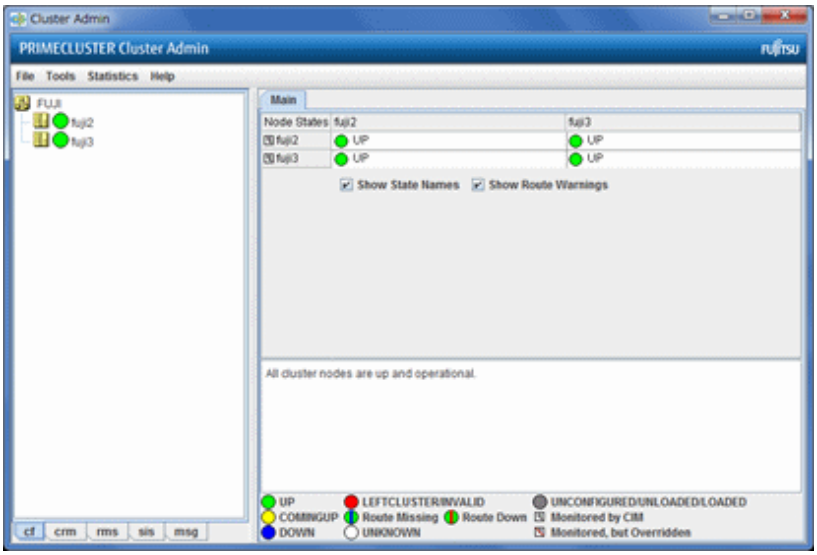
Display procedure

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

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



Cluster Admin supports the functions described below.

The manual reference locations are indicated in "Overview of 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 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"

- **sis (SIS: Scalable Internet Services)**

  In PRIMECLUSTER 4.4, this function is not available.

- **msg (Message)**

  Cluster control messages are displayed.

  Reference location: "Chapter 7 Operations"

4.6 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 <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 <X> in the upper right corner of the screen.
- At the login screen, clicking <X> in the upper 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:

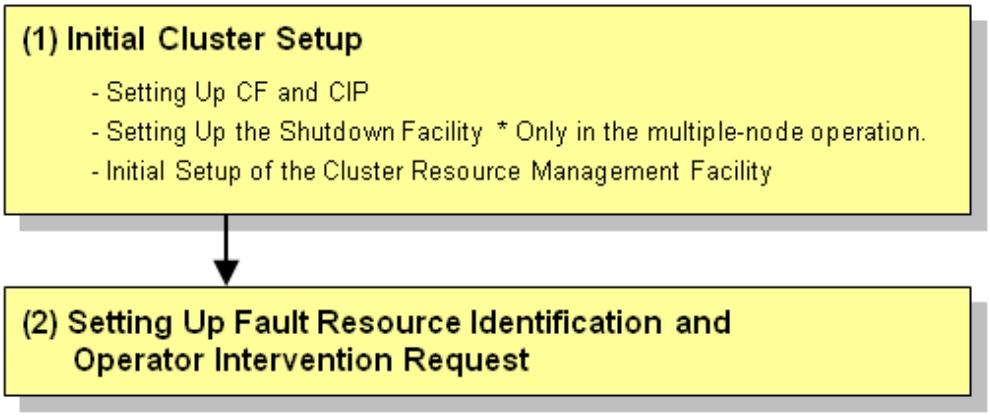


Table 5.1 Cluster building procedure and manual reference locations

	Work item	Execution Node	Required/ Optional	Manual reference location*1
(1)	5.1.1 Setting Up CF and CIP	All nodes	Required	**CF** "1.1 CF, CIP, and CIM configuration"
	5.1.2 Setting up the Shutdown Facility	All nodes	Required***2**	**CF** "7 Shutdown Facility "
	5.1.3 Initial Setup of the Cluster Resource Management Facility	All nodes	Required	**CF** "3.1 Resource Database configuration"
(2)	5.2 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.

Note

- Execute the configuration setting of GDS after initializing the cluster.
- If you deactivate the virtual interfaces such as GLS, tagged VLAN, and virtual bridge, or restart the network service before setting CF and CIP, restart CF first, and then start the settings of CF and CIP.
  For information on how to stop and start CF, see "4.6 Starting and stopping CF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

5.1 Initial Cluster Setup

This section describes the initial cluster setup for PRIMECLUSTER.

If the virtual machine function is used, you need to set up a virtual network first.

For details, refer to "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function" or the Linux documentation.

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

	Contents	Manual reference location*
1	5.1.1 Setting Up CF and CIP (setting up cluster configuration information and IP addresses)	**CF** "1.1 CF, CIP, and CIM configuration"
2	5.1.2 Setting up the Shutdown Facility	**CF** "7 Shutdown Facility"
3	5.1.3 Initial Setup of the Cluster Resource Management Facility	**CF** "3.1 Resource Database configuration"

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

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

When Firewall is enabled, disable it before the initial cluster setup.

When enabling Firewall after completing the installation of the cluster, see "Appendix K Using Firewall."

5.1.1 Setting Up CF and CIP

Set up Cluster Foundation (CF) and CIP by using the CF Wizard of Cluster Admin. The designsheets for the CF and CIP setup are "Setup (initial configuration)" of PRIMECLUSTER Designsheets.

For details on the setting, see "1.1.5 Example of creating a cluster" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

Setup item	Description	Described in designsheets
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"
Cluster nodes	Select the nodes that will construct a cluster system.	"Node name (uname-n)" for "Node 1", "Node 2", "Node 3" and "Node 4"
CF node names	Set the names of the nodes that construct the cluster. The beginning character of each CF node name must be a lower-case alphabet. The length of each node name must be not more than 11 characters which consist of lower-case alphabets, numbers, or symbols (- and _).	"CF node name" for "Node 1", "Node 2", "Node 3" and "Node 4"
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.	"Path 0 NIC name" and "Path 1 NIC name" for "Node 1", "Node 2", "Node 3" and "Node 4"
IP interconnects	Optional. This setup allows you to operate CF over IP.	"IP interconnect setup"
CIP subnets	Set the following items, and set the IP address used by CF: - CIP subnet count - Host suffix - Subnet number	"Number of subnets", "Subnet IP", and "Net mask" for "CIP"

Setup item	Description	Described in designsheets
	- Subnet mask	
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." To use RMS, you need to select at least one function.	"CF remote service usage"
Cluster Integrity Monitor (CIM) configuration	Set the nodes to be monitored by CIM.	"Node in CF quorum set" for "Node 1", "Node 2", "Node 3" and "Node 4"

Note

- Node names of the cluster nodes are automatically input to "CF node names." The CF node name must be within 11characters.
- 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 cluster interconnects are not secured
- Hereinafter, the CF remote services (CFCP and CFSH) must be enabled. To enable the CFCP and CFSH, add the following definition to the "/etc/default/cluster.config" file before configuring the CF.

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

- If the CF and CIP configuration fails, check the following:
  - The cluster interconnect is connected.
  - The network interface that is used for the cluster interconnect is activated.
  - An IP address is configured for the network interface.
- After the CF setup is completed, the "SF Wizard Startup Check" pop-up window is displayed. Click the No button. SF Wizard is not available in PRIMECLUSTER 4.4. Set up the SF according to the instructions in "5.1.2 Setting up the Shutdown Facility."
- To share a NIC with the administrative LAN and the cluster interconnect, see "1.1 CF, CIP, and CIM configuration" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."
- When the bonding device is used for the cluster interconnection, only mode=1(active-backup) can be used.

Note

In the case of the single-node cluster operation

- When constructing multiple single-node clusters, specify a different 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 listed in the above table.
- 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.

- Following messages of the shutdown facility and RMS are output, however, this is no problem, since the setting of the shutdown facility is not performed.
  - Messages of the shutdown facility:

```
fopen of /etc/opt/SMAW/SMAWsf/rcsd.cfg failed, errno 2
Could not correctly read the rcsd.cfg file
```

  - Messages of RMS:

```
(SCR,26): ERROR The sdtool notification script has failed with status 1 after dynamic
modification.
```

See

For information on the corrective action to be applied when the setting of the cluster interconnect fails, see "Chapter 8 Diagnostics and troubleshooting" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

5.1.2 Setting up the Shutdown Facility

This section describes the setup procedure of the shutdown facility for the PRIMERGY, PRIMEQUEST, and virtual machine environment (KVM environment).

The setup procedure for the shutdown facility is different depending on the model/configuration.

The following shows the shutdown agents required for each hardware model/configuration.

Table 5.2 Shutdown agent for PRIMERGY

Server model	IPMI (SA_ipmi)	Blade (SA_blade)	kdump (SA_lkcd)
RX series TX series	Y	-	Y
BX series (For use in combination with ServerView Resource Orchestrator Virtual Edition)	Y (*1)	-	Y
BX series (For not use in combination with ServerView Resource Orchestrator Virtual Edition)	-	Y	Y

(*1) The combination of user and password for BMC or iRMC that is used in the shutdown facility must be the same on all blades.

Table 5.3 Shutdown agent for PRIMEQUEST

Server model		MMB	
		Panic (SA_mmbp)	Reset (SA_mmbr)
PRIMEQUEST	1000/2000 series	Y	Y

Table 5.4 Shutdown agent for virtual machine environment (KVM environment)

Server model	Cluster configuration	libvirt	
		Panic (SA_libvirtgp)	Reset (SA_libvirtgr)
PRIMERGY	Guests in a unit	Y	Y

Server model		Cluster configuration	libvirt	
			Panic (SA_libvirtgp)	Reset (SA_libvirtgr)
		Guests between units	Y	Y
PRIMEQUEST	1000/2000 series	Guests in a unit	Y	Y
		Guests between units	Y	Y

When using the host OS failover function in virtual machine environment (KVM environment), set the following shutdown agents. The shutdown agent that are set on the guest OS are the same as those used in the virtual machine function.

Table 5.5 Shutdown agent for the host OS failover function in virtual machine environment (KVM environment)

Server model	Cluster configuration	Cluster node	Shutdown agent for the server model	libvirt		vmchkhost
				Panic (SA_libvirtgp)	Reset (SA_libvirtgr)	Checking the status (SA_vmchkhost)
PRIMERGY	Guests between units	Host OS	Y (See Table 5.2)	-	-	-
		Guest OS	-	Y	Y	Y
PRIMEQUEST 1000/2000 series	Guests between units	Host OS	Y (See Table 5.3)	-	-	-
		Guest OS	-	Y	Y	Y

See

For details on the shutdown facility, see the following manuals:

1. "2.3.1.7 PRIMECLUSTER SF" in "PRIMECLUSTER Concepts Guide"
2. "Chapter 7 Shutdown Facility" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide"

5.1.2.1 Survival Priority

If the cluster partition occurred due to a fault in the cluster interconnect, all the nodes would still be in the state of accessing the user resources. For details on the cluster partition, see "1.2.2.1 Protecting data integrity" in "PRIMECLUSTER Concepts Guide."

In order to guarantee the data consistency in the user resources, SF must determine the node groups of which nodes remain to survive and which nodes need to be forcibly stopped.

The weight assigned to each node group is referred to as "Survival priority" in 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 the multiple node groups have the same survival priority, the node group that includes the node with the alphabetical earliest node name 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 userApplication."

See

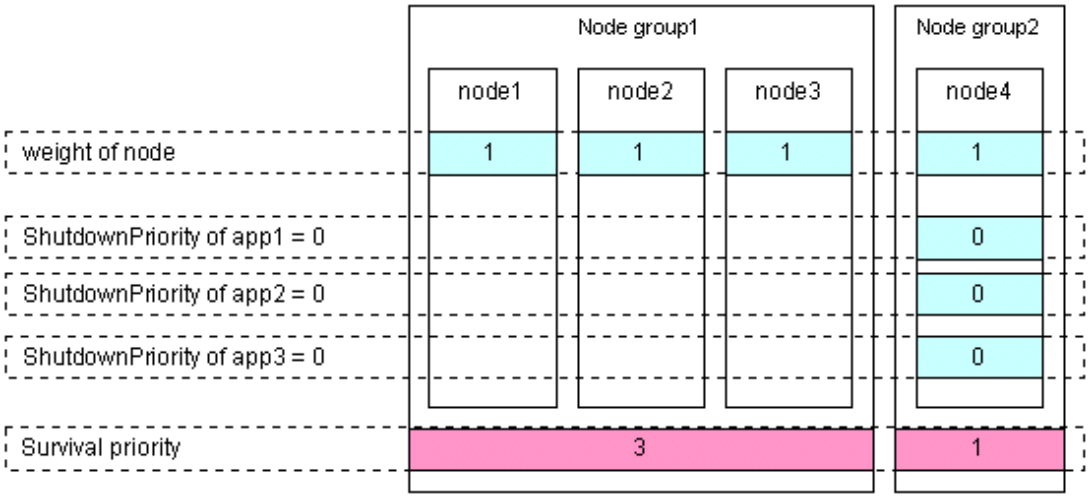
For details on the ShutdownPriority attribute of userApplication, see "12.1 Attributes available to the user" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Survival scenarios

The typical scenarios that are implemented are shown below:

[Largest node group survival]

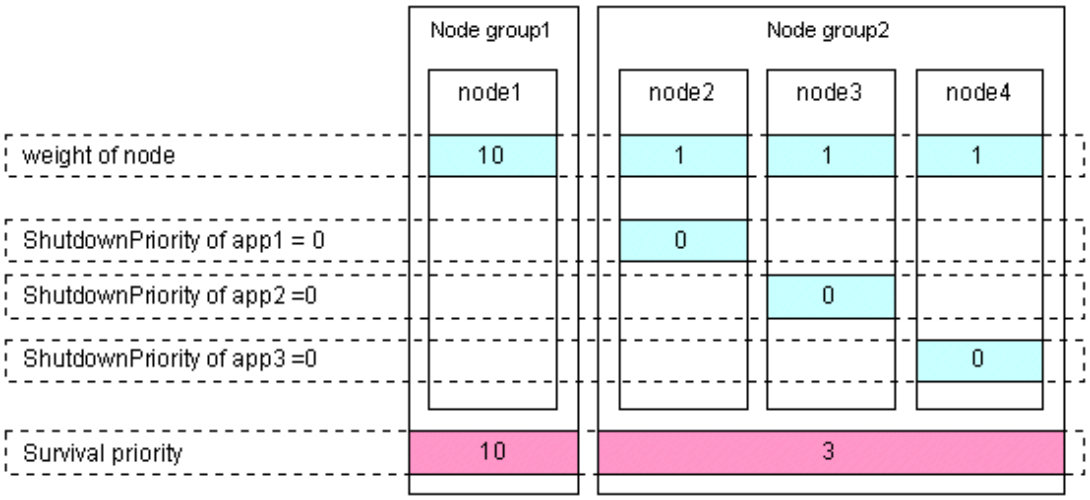
- Set the weight of all the nodes to 1 (default).
- Set the ShutdownPriority attribute of every userApplication to 0 (default).



[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 every userApplication to 0 (default).

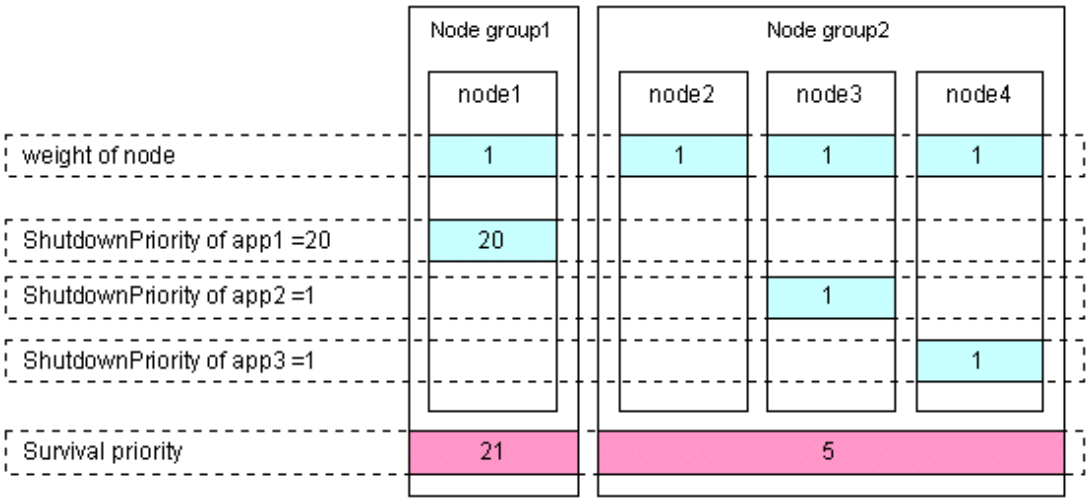
In the following example, node1 is to survive:



[Specific userApplication survival]

- Set the "weight" of all the nodes to 1 (default).
- Set the ShutdownPriority attribute of userApplication 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 the nodes.
- Set the ShutdownPriority attribute within the range of 1 to 20.

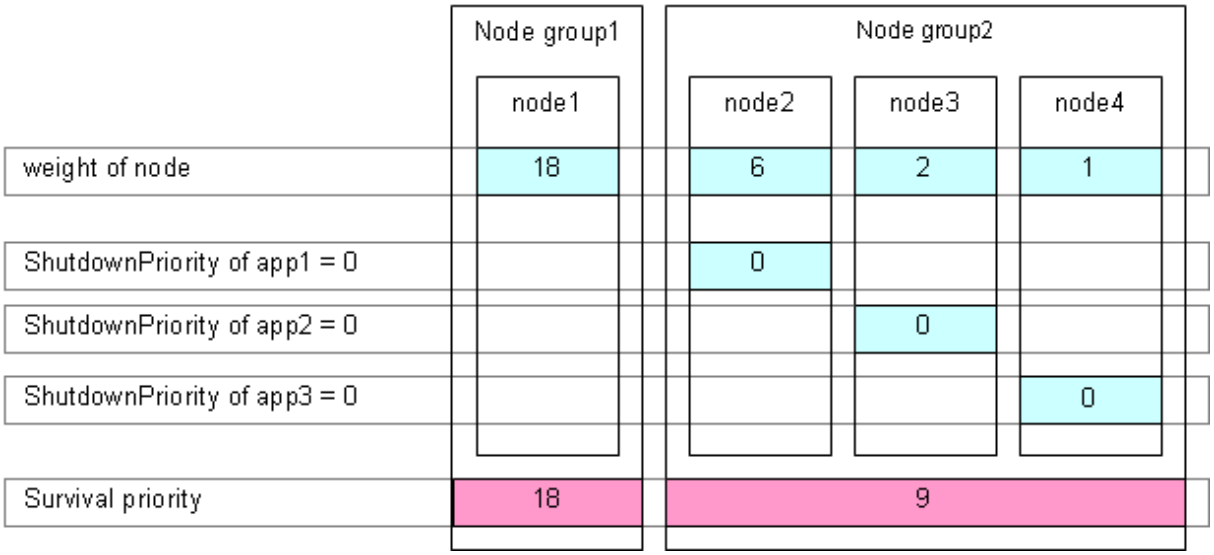
In the following example, the node for which app1 is operating is to survive:



[Node survival in a specific order of node]

- Set the "weight" of the node to survive to a value more than double the total weight of the other nodes which have lower priority.
- Set the ShutdownPriority attribute of every userApplication to 0 (default).

In the following example, node1, node2, node3, and node4 are to survive in this order:

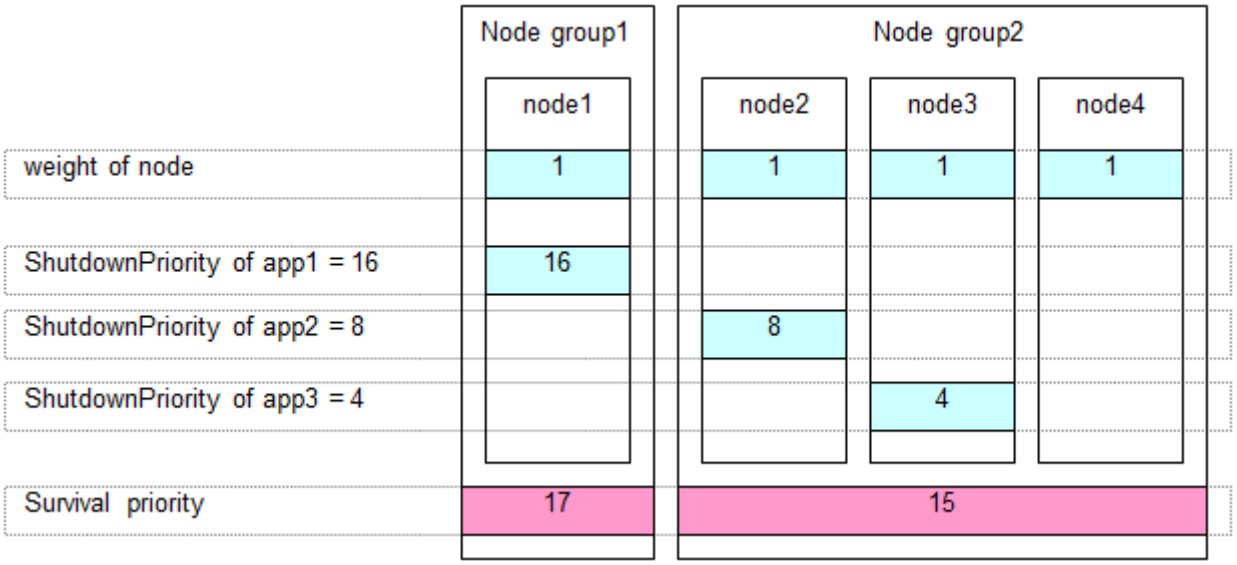


[Node survival in a specific userApplication order]

- Set the "weight" of all the nodes to 1 (default).
- Set the value that is power of 2 (1, 2, 4, 8, 16, ...) to the ShutdownPriority attribute of userApplication if its operation must be continued.
- Calculate the minimum value to be set to the ShutdownPriority attribute using the following formula. The value must be power of 2 (1, 2, 4, 8, 16, ...) and equal to or larger than the calculated value.

  The number of configuration node - 1
  Example: In 2-node configuration, (2 - 1) = 1. The minimum settable value to ShutdownPriority attribute is 1.
  Example: In 3-node configuration, (3 - 1) = 2. The minimum settable value to ShutdownPriority attribute is 2.
  Example: In 4-node configuration, (4 - 1) = 3. The minimum settable value to ShutdownPriority attribute is 4.

The following example shows the survival priority of nodes on which userApplication runs. Sequentially app1, app2, and app3 are prioritized.



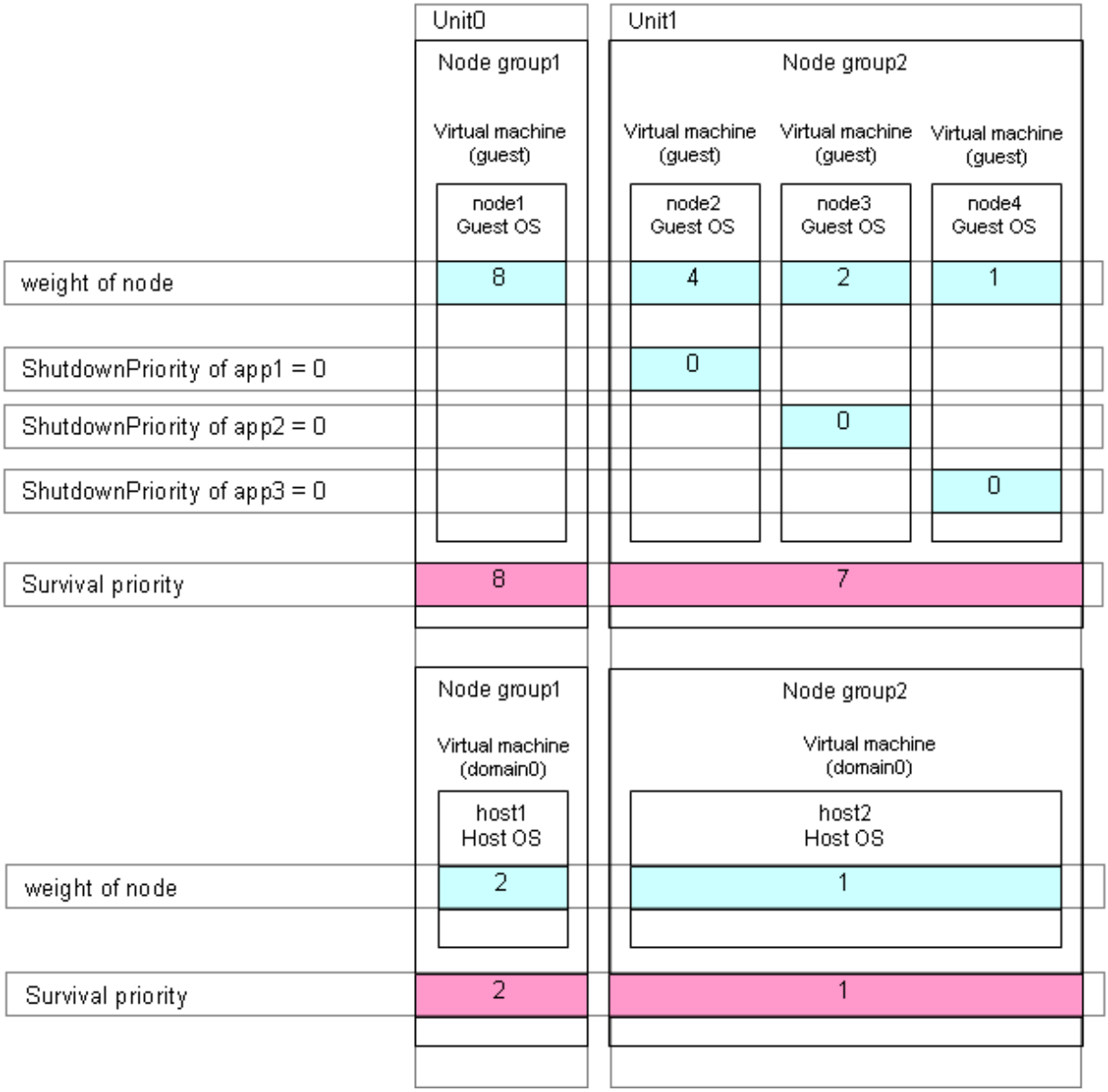
[Host OS failover function]

- Set the "weight" of nodes to a power-of-two value (1,2,4,8,16,...) in ascending order of survival priority in each cluster system.
- The "weight" set to a guest OS should have the same order relation with a corresponding host OS.

  For example, when setting a higher survival priority to host1 than host2 between host OSes, set a higher survival priority to node1 (corresponding to host1) than node2-4 (corresponding to host2) between guest OSes.

- Set the ShutdownPriority attribute of every userApplication to 0 (default).

In the following example, node1, node2, node3, and node4 are to survive in this order:



5.1.2.2 Setup Flow for Shutdown Facility

5.1.2.2.1 Setup Flow in PRIMERGY RX/TX Series

For the setup flow for the shutdown facility in PRIMERGY RX/TX series, take the following steps.

1. Checking the shutdown agent information
2. Setting up the shutdown daemon
3. Configuring the IPMI shutdown agent
4. Configuring the kdump shutdown agent
5. Starting up the shutdown facility
6. Test for forced shutdown of cluster nodes

For the detailed setup procedure, refer to "5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY."

5.1.2.2.2 Setup Flow in PRIMERGY BX Series

When using in combination with ServerView Resource Orchestrator Virtual Edition

When using in combination with ServerView Resource Orchestrator Virtual Edition, for the setup flow for the shutdown facility in PRIMERGY BX series, take the following steps.

1. Checking the shutdown agent information
2. Setting up the shutdown daemon
3. Configuring the IPMI shutdown agent
4. Configuring the kdump shutdown agent
5. Starting up the shutdown facility
6. Test for forced shutdown of cluster nodes

For the detailed setup procedure, refer to "5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY."

When not using in combination with ServerView Resource Orchestrator Virtual Edition

When not using in combination with ServerView Resource Orchestrator Virtual Edition, for the setup flow for the shutdown facility in PRIMERGY BX series, take the following steps.

1. Checking the shutdown agent information
2. Setting up the shutdown daemon
3. Configuring the Blade shutdown agent
4. Configuring the kdump shutdown agent
5. Starting up the shutdown facility
6. Test for forced shutdown of cluster nodes

For the detailed setup procedure, refer to "5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY."

5.1.2.2.3 Setup Flow in PRIMEQUEST

For the setup flow for the shutdown facility in PRIMEQUEST, take the following steps.

1. Checking the shutdown agent information
2. Configuring the MMB shutdown agent
3. Setting up the shutdown daemon
4. Starting the MMB asynchronous monitoring daemon
5. Setting the I/O completion wait time(for using other than ETERNUS disk array as the shared disk)
6. Starting up the shutdown facility
7. Test for forced shutdown of cluster nodes

For the detailed setup procedure, refer to "5.1.2.4 Setup Procedure for Shutdown Facility in PRIMEQUEST."

5.1.2.2.4 Setup Flow in Virtual Machine Environment

When using the host OS failover function

When using the host OS failover function in a virtual machine environment, for the setup flow for the shutdown facility, take the following steps.

1. Setting up the shutdown facility on the host OS in PRIMERGY/PRIMEQUEST
2. Checking the shutdown agent information in a virtual machine environment

3. Configuring the libvirt shutdown agent
4. Configuring the vmchkhost shutdown agent
5. Starting up the shutdown facility
6. Setting up the host OS failover function on the host OS (PRIMEQUEST only)
7. Test for forced shutdown of cluster nodes

For the detailed setup procedure, see the following.

5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY

5.1.2.4 Setup Procedure for Shutdown Facility in PRIMEQUEST

5.1.2.5 Setup Procedure for Shutdown Facility in Virtual Machine Environment

When not using the host OS failover function

When not using the host OS failover function in a virtual machine environment, for setup flow for the shutdown facility, take the following steps.

1. Checking the shutdown agent information in a virtual machine environment
2. Configuring the libvirt shutdown agent
3. Starting up the shutdown facility
4. Test for forced shutdown of cluster nodes

For the detailed setup procedure, refer to "5.1.2.5 Setup Procedure for Shutdown Facility in Virtual Machine Environment."

5.1.2.3 Setup Procedure for Shutdown Facility in PRIMERGY

This section describes the procedure for setting up the shutdown agent in PRIMERGY.

For shutdown facility, set the supported server type to be used.

When creating a redundant administrative LAN used in the shutdown facility by using GLS, set as below.

- For taking over the IP address between nodes

  Configure GLS by using the logical IP address takeover function of the NIC switching mode.

  For shutdown facility, specify a physical IP address instead of a logical IP address.

- For not taking over the IP address between nodes

  Configure GLS by using the physical IP address takeover function of the NIC switching mode.

5.1.2.3.1 Checking the Shutdown Agent Information

RX/TX series

Check the following settings in BMC (Baseboard Management Controller) or iRMC (integrated Remote Management Controller) necessary for setting the IPMI shutdown agent.

- IP address of BMC or iRMC
- User defined in BMC or iRMC
- User password defined in BMC or iRMC

Also, check the following.

- An administrator privilege must be set in the defined user in BMC or iRMC.

- The defined user password in BMC or iRMC must be set with seven-bit ASCII characters other than the following characters.

  > < " / \ = ! ? ; , &

BX series (When using in combination with ServerView Resource Orchestrator Virtual Edition)

Necessary settings are the same as the settings of RX/TX series. Refer to RX/TX series.

BX series (When not using in combination with ServerView Resource Orchestrator Virtual Edition)

Check the following settings for the management blade necessary for setting the Blade shutdown agent.

- IP address of the management blade
- SNMP community name of the management blade
- Slot number of the server blade where a cluster host is operating

5.1.2.3.2 Setting up the Shutdown Daemon

Create /etc/opt/SMAW/SMAWsf/rcsd.cfg on all the nodes as shown below.

Create the rcsd.cfg file by a root user and change the permission to 600.

RX/TX series

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_ipmi,timeout=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_ipmi,timeout=timeout
```

```
CFNameX          : Specify the CF node name of the cluster host.
weight           : Specify the weight of the SF node.
myadmIP          : Specify the IP address of the administrative LAN
                   used in the shutdown facility of the cluster host.
                   It is not the IP address of iRMC or the management blade.
                   Available IP addresses are IPv4 and IPv6 addresses.
                   IPv6 link local addresses are not available.
                   When specifying an IPv6 address, enclose it in brackets "[ ]".
                   (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                   If you specify a host name, make sure it is listed in /etc/hosts.
SA_ipmi          : Specify the IPMI shutdown agent.
timeout          : Specify the timeout duration (seconds) of the IPMI shutdown agent.
                   For the IPMI shutdown agent, specify 25 seconds.
```

Example:

```
node1,weight=1,admIP=10.20.30.100:agent=SA_ipmi,timeout=25
node2,weight=1,admIP=10.20.30.101:agent=SA_ipmi,timeout=25
```

BX series (When using in combination with ServerView Resource Orchestrator Virtual Edition)

Necessary settings are the same as the settings of RX/TX series. Refer to RX/TX series.

BX series (When not using in combination with ServerView Resource Orchestrator Virtual Edition)

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_blade,timeout=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_blade,timeout=timeout
```

```
CFNameX          : Specify the CF node name of the cluster host.
weight           : Specify the weight of the SF node.
myadmIP          : Specify the IP address of the administrative LAN
                   used in the shutdown facility of the cluster host.
                   It is not the IP address of iRMC or the management blade.
                   Available IP addresses are IPv4 and IPv6 addresses.
```

```
                  IPv6 link local addresses are not available.
                  When specifying an IPv6 address, enclose it in brackets "[ ]".
                  (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                  If you specify a host name, make sure it is listed in /etc/hosts.
SA_blade          : Specify the Blade shutdown agent.
timeout           : Specify the timeout duration (seconds) of the Blade shutdown agent.
                    For the Blade shutdown agent, specify 20 seconds.
```

Example:

```
node1,weight=1,admIP=10.20.30.100:agent=SA_blade,timeout=20
node2,weight=1,admIP=10.20.30.101:agent=SA_blade,timeout=20
```

Note

- For using STP (Spanning Tree Protocol) in the administrative LAN used in the shutdown facility, it is necessary to set the timeout value to the current value plus (+) 50 (seconds), taking into account the time STP needs to create the tree and an extra cushion. This setting increases the time required for failover.
- The contents of the rcsd.cfg file must be the same on all the nodes. If different, it does not work.

Information

When the "/etc/opt/SMAW/SMAWsf/rcsd.cfg" file is to be created, the "/etc/opt/SMAW/SMAWsf/rcsd.cfg.template" file can be used as a prototype.

5.1.2.3.3 Setting up IPMI Shutdown Agent

In RX/TX series, or when using in combination with ServerView Resource Orchestrator Virtual Edition in BX series, for the server with the BMC (Baseboard Management Controller) or iRMC (integrated Remote Management Controller) installed, configure the IPMI shutdown agent.

You must configure the IPMI shutdown agent before you configure the kdump shutdown agent.

1. Starting the IPMI service

   [RHEL6]

   Execute the following command on all the nodes to check the startup status of the IPMI service.

   ```
   # /sbin/service ipmi status
   ```

   Execute the following command on all the nodes in which the IPMI service is not activated to start the IPMI service.

   ```
   # /sbin/service ipmi start
   Starting ipmi drivers: [ OK ]
   ```

   [RHEL7]

   Execute the following command on all the nodes to check the startup status of the IPMI service.

   ```
   # /usr/bin/systemctl status ipmi.service
   ipmi.service - IPMI Driver
   Loaded: loaded (/usr/lib/systemd/system/ipmi.service; disabled)
   Active: inactive (dead)
   ```

   Execute the following command on all the nodes in which the "inactive" appeared in the "Active:" field and the IPMI service is not activated to start the IPMI service.

   ```
   # /usr/bin/systemctl start ipmi.service
   ```

2. Setting the run level of the IPMI service

   [RHEL6]

   Execute the following command on all the nodes to read the IPMI service on startup.

```
# /sbin/chkconfig --level 2345 ipmi on
```

   [RHEL7]

   Execute the following command on all the nodes to read the IPMI service on startup.

```
# /usr/bin/systemctl enable ipmi.service
```

3. Encrypting the password

   Execute the sfcipher command to encrypt passwords of a user for the shutdown facility.

   Example: If the password specified when setting up IPMI (BMC and iRMC) is "bmcpwd$"

```
# sfcipher -c
Enter User's Password:  <- enter bmcpwd$
Re-enter User's Password:  <- enter bmcpwd$
/t1hXYb/Wno=
```

   Note: It is not necessary to insert '\' in front of the special characters specified as the password in 4.3A30 or later.

   For information on how to use the sfcipher command, see the "sfcipher" manual page.

For the passwords specified when setting up IPMI (BMC and iRMC), seven-bit ASCII characters are available.
Among them, do not use the following characters as they may cause a problem.

```
>  <  "  /  \  =  !  ?  ;  ,  &
```

4. Setting the shutdown agent

   Create /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg on all the nodes as shown below.

   Create the SA_ipmi.cfg file by a root user and change the permission to 600.

   - For IPv4 address

```
CFName1  ip-address:user:passwd {cycle | leave-off}
CFName2  ip-address:user:passwd {cycle | leave-off}
```

   - For IPv6 address

```
CFName1 [ip-address]:user:passwd {cycle | leave-off}
CFName2 [ip-address]:user:passwd {cycle | leave-off}
```

```
CFNameX          : Specify the CF node name of the cluster host.
ip-address       : Specify the Ip address for IPMI (BMC or iRMC)
                   in the server where a cluster host is operating.
                   Available IP addresses are IPv4 and IPv6 addresses.
                   IPv6 link local addresses are not available.
                   When specifying the IPv6 address, enclose it in brackets "[ ]".
                   (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
user             : Specify the user defined when IPMI (BMC or iRMC) was setup.
passwd           : Password defined when IPMI (BMC or iRMC) was setup.
                   Specify the password encrypted in step 3.
cycle            : Reboot the node after forcibly stopping the node.
leave-off        : Power-off the node after forcibly stopping the node.
```

   Example 1:

When the IP address of iRMC of node1 is 10.20.30.50 and the IP address of iRMC of node2 is 10.20.30.51

```
node1  10.20.30.50:root:/t1hXYb/Wno= cycle
node2  10.20.30.51:root:/t1hXYb/Wno= cycle
```

Example 2:

When the IP address of iRMC of node1 is 1080:2090:30a0:40b0:50c0:60d0:70e0:80f0 and the IP address of iRMC of node2 is 1080:2090:30a0:40b0:50c0:60d0:70e0:80f1

```
node1 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0]:root:/t1hXYb/Wno= cycle
node2 [1080:2090:30a0:40b0:50c0:60d0:70e0:80f1]:root:/t1hXYb/Wno= cycle
```

Information

When the "/etc/opt/SMAW/SMAWsf/SA_ipmi.cfg" file is to be created, the "/etc/opt/SMAW/SMAWsf/SA_ipmi.cfg.template" file can be used as a prototype.

Note

- Check if the setting contents of the /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg file are correct. If there is an error in the setting contents, the shutdown facility cannot be performed normally.
- Check if the IP address (ip-address) of IPMI (BMC or iRMC) corresponding to the cluster host's CF node name (CFNameX) of the /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg file are set. If there is an error in the setting, a different node may be forcibly stopped.
- The contents of the SA_ipmi.cfg file must be the same on all the nodes. If different, it does not work.

5.1.2.3.4 Setting up Blade Shutdown Agent

When not using in combination with ServerView Resource Orchestrator Virtual Edition in BX series, configure the Blade shutdown agent. You must configure the Blade shutdown agent before you configure the kdump shutdown agent.

Create /etc/opt/SMAW/SMAWsf/SA_blade.cfg on all the nodes as shown below.

Create SA_blade.cfg file by a root user and change the permission to 600.

Cluster configuration within a single chassis

```
management-blade-ip IPaddress
community-string SNMPcommunity
CFName1 slot-no {cycle | leave-off}
CFName2 slot-no {cycle | leave-off}
```

```
IPaddress          : Specify the IP address of the management blade.
                     Available IP addresses are IPv4 and IPv6 addresses.
                     IPv6 link local addresses are not available.
                     When specifying the IPv6 address, enclose it in brackets "[ ]".
                     (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
SNMPcommunity      : Specify the SNMP community of the management blade.
CFNameX            : Specify the CF node name of the cluster host.
slot-no            : Specify the slot No. of the server blade where a cluster
                     host is operating.
cycle              : Reboot the node after forcibly stopping the node.
leave-off          : Power-off the node after forcibly stopping the node.
```

Example :

When the IP address of the management blade of node1 and node2 is 10.20.30.50, the slot number of node1 is 1 and the slot number of node2 is 2.

```
management-blade-ip 10.20.30.50
community-string public
node1 1 cycle
node2 2 cycle
```

Cluster configuration across multiple chassis

```
community-string SNMPcommunity
management-blade-ip IPaddress1
CFName1 slot-no {cycle | leave-off}
management-blade-ip IPaddress2
CFName2 slot-no {cycle | leave-off}
```

```
IPaddressX          : Specify the IP address of the management blade
                      in a chassis where a cluster host of CFNameX exists.
                      Available IP addresses are IPv4 and IPv6 addresses.
                      IPv6 link local addresses are not available.
                      When specifying the IPv6 address, enclose it in brackets "[ ]".
                      (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                      Make sure to write it before CFNameX.
SNMPcommunity       : Specify the SNMP community of the management blade.
CFNameX             : Specify the CF node name of the cluster host.
slot-no             : Specify the slot No. of the server blade
                      where a cluster host is operating.
cycle               : Reboot the node after forcibly stopping the node.
leave-off           : Power-off the node after forcibly stopping the node.
```

Note

SNMP community name of the management blade must be same in all the chassis.

Example:

When the IP address of the management blade of node1 is 10.20.30.50, and the slot number of node1 is 1.
Moreover, when the IP address of the management blade of node2 is 10.20.30.51, and the slot number of node2 is 2.

```
community-string public
management-blade-ip  10.20.30.50
node1 1 cycle
management-blade-ip  10.20.30.51
node2 2 cycle
```

Information

When the "/etc/opt/SMAW/SMAWsf/SA_blade.cfg" file is to be created, the "/etc/opt/SMAW/SMAWsf/SA_blade.cfg.template" file can be used as a prototype.

Note

- Check if the setting contents of the /etc/opt/SMAW/SMAWsf/SA_blade.cfg file are correct. If there is an error in the setting contents, the shutdown facility cannot be performed normally.
- Check if the IP address (IPaddress) of the management blade and the slot number (slot-no) of the server blade corresponding to the cluster host's CF node name (CFNameX) of the /etc/opt/SMAW/SMAWsf/SA_blade.cfg file are set. If there is an error in the setting, a different node may be forcibly stopped.
- The contents of SA_blade.cfg file must be same on all the nodes. If different, it does not work.

5.1.2.3.5 Setting up kdump Shutdown Agent

Configure the kdump shutdown agent after configuring IPMI shutdown agent or Blade shutdown agent.

Perform the following procedures.

1. Initializing the configuration file for the kdump

   Execute the following command on any one of the cluster nodes.

   ```
   # /etc/opt/FJSVcllkcd/bin/panicinfo_setup
   ```

   If the following message is output, the setting file (rcsd.cfg) of the shutdown daemon has an error. Correct the file.

   ```
   panicinfo_setup: ERROR: Reading the Shutdown Facility configuration failed.
   ```

   If the following message is output, the setting file (SA_ipmi.cfg or SA_blade.cfg) of the shutdown agent has an error. Correct the file.

   ```
   panicinfo_setup: ERROR: Reading the Shutdown Agent configuration failed.
   ```

   In the environment where panicinfo_setup has already been executed, the following massage is output.

   ```
   panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
   (I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
   ```

   In the case, input "I".

   Note

   To execute the command, CF and CF services (CFSH and CFCP) must be activated. For details, see "5.1.1 Setting Up CF and CIP."

2. Setting crash dump collection

   - In RX/TX series, or when using in combination with ServerView Resource Orchestrator Virtual Edition in BX series

     1. Change /etc/opt/FJSVcllkcd/etc/SA_lkcd.tout as follows on all the nodes.

        Before change

        ```
        PANICINFO_TIMEOUT 5
        RSB_PANIC 0
        ```

        After change

        ```
        PANICINFO_TIMEOUT 10
        RSB_PANIC 3
        ```

     2. Change the timeout value of SA_lkcd in the /etc/opt/SMAW/SMAWsf/rcsd.cfg file as follows on all the nodes.

        Before change

        ```
        agent=SA_lkcd,timeout=20
        ```

        After change

        ```
        agent=SA_lkcd,timeout=25
        ```

   - When not using in combination with ServerView Resource Orchestrator Virtual Edition in BX series

     Change RSB_PANIC of /etc/opt/FJSVcllkcd/etc/SA_lkcd.tout as follows on all the nodes.

     Before change

     ```
     RSB_PANIC 0
     ```

     After change

```
RSB_PANIC 2
```

5.1.2.3.6 Starting up the Shutdown Facility

Start or restart the shutdown facility on all the nodes.

1. Starting the shutdown facility

   Check that the shutdown facility has been started on all the nodes.

   ```
   # sdtool -s
   ```

   If the shutdown facility has already been started, execute the following commands to restart the shutdown facility on all the nodes.

   ```
   # sdtool -e
   # sdtool -b
   ```

   If the shutdown facility has not been started, execute the following command to start the shutdown facility on all the nodes.

   ```
   # sdtool -b
   ```

2. Checking the status of the shutdown facility

   Check the status of the shutdown facility on all the nodes.

   ```
   # sdtool -s
   ```

Information

Display results of the sdtool -s command

- If InitFailed is displayed in Init State, it means that a problem occurred during initialization of that shutdown agent.
- If TestFailed is displayed in Test State, it means that a problem occurred while the agent was testing whether or not the node displayed in the Cluster Host field could be stopped. Some sort of problem probably occurred in the software, hardware, or network resources being used by that agent.
- If Unknown is displayed in Shut State or Init State, it means that SF has not yet executed node stop, path testing, or SA initialization. Unknown is displayed temporarily in Test State and Init State until the actual status can be confirmed.
- If TestFailed or InitFailed is displayed, check /var/log/messages. After the failure-causing problem is resolved and SF is restarted, the status display changes to InitWorked or TestWorked.

Note

If TestFailed is displayed in Test State when "sdtool -s" is executed after the shutdown facility was started, possible causes are as follows:

- The shutdown agent is incorrectly set.
- The IPMI shutdown agent is used without the user password of the shutdown facility encrypted.

Take the following procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

   ```
   # sdtool -e
   ```

2. Review the settings of shutdown facility.
3. Execute the following command on any node to apply changes of the configuration file.

   ```
   # /etc/opt/FJSVcllkcd/bin/panicinfo_setup
   ```

After the following message is displayed, select "I."

```
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
```

4. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

5. Execute the following command on all the nodes and make sure that the shutdown facility operates normally.

```
# sdtool -s
```

5.1.2.3.7 Test for Forced Shutdown of Cluster Nodes

After setting up the shutdown facility, conduct a test for forced shutdown of cluster nodes to check that the correct nodes can be forcibly stopped.

For the detail of the test for forced shutdown of cluster nodes, refer to "1.4 Test."

5.1.2.4 Setup Procedure for Shutdown Facility in PRIMEQUEST

This section describes the setup procedure for the shutdown facility in PRIMEQUEST.

Note

When creating a redundant administrative LAN used in the shutdown facility by using GLS, set as below.

- For taking over the IP address between nodes

  Configure GLS by using the logical IP address takeover function of the NIC switching mode.

  For shutdown facility, specify a physical IP address instead of a logical IP address.

- For not taking over the IP address between nodes

  Configure GLS by using the physical IP address takeover function of the NIC switching mode.

5.1.2.4.1 Checking the Shutdown Agent Information

MMB check items

Check the following settings for MMB blade necessary for setting Blade shutdown agent.

- User name for controlling the MMB with RMCP
- User password for controlling the MMB with RMCP.

Also check that following settings are enabled for the user confirmed above:

- The "Privilege" setting of the user is set to "Admin" so that the user can control the MMB with RMCP.
- The "Status" setting of the user is set to "Enabled" so that the user can control the MMB with RMCP.
- The passwords for controlling MMB with RMCP must be specified seven-bit ASCII characters are available.

  `>  <  "  /  \  =  !  ?  ;  ,  &`

Check the settings for the user who uses RMCP to control the MMB. Log in to MMB Web-UI, and check the settings from the "Remote Server Management" window of the "Network Configuration" menu.

If the above settings have not been set, set up the MMB so that the above settings are set.

Note

The MMB units have two types of users:

- User who controls all MMB units
- User who uses RMCP to control the MMB

The user to be checked here is the user who uses RMCP to control the MMB. Be sure to check the correct type of user.

Checking the time to wait until I/O to the shared disk is completed (when using other than the ETERNUS disk array as the shared disk)

When using any disks other than the ETERNUS disk array as the shared disk, to prevent the data error when the node is down due to a panic or other causes, set the time until I/O to the shared disk is completed.

To set the wait time, panic the node during I/O to the shared disk. After that, check the time until I/O to the shared disk is completed.

5.1.2.4.2 Setting up the MMB Shutdown Agent

Set up the MMB shutdown agent according to the procedure described below.

Take this procedure after taking the procedure described in "5.1.1 Setting Up CF and CIP."

1. Execute the "clmmbsetup -a" command on all the nodes, and register the MMB information.

   For instructions on using the "clmmbsetup" command, see the "clmmbsetup" manual page.

   ```
   # /etc/opt/FJSVcluster/bin/clmmbsetup -a mmb-user
   Enter User's Password:
   Re-enter User's Password:
   ```

   For *mmb-user* and User's Password, enter the following values that were checked in "5.1.2.4.1 Checking the Shutdown Agent Information."

   mmb-user

   User's name for controlling the MMB with RMCP

   User's Password

   User's password for controlling the MMB with RMCP.

   Note

   For the passwords specified when setting MMB, seven-bit ASCII characters are available.

   Among them, do not use the following characters as they may cause a problem.

   ```
   >  <  "  /  \  =  !  ?  ;  ,  &
   ```

2. Execute the "clmmbsetup -l" command on all the nodes, and check the registered MMB information.

   If the registered MMB information was not output on all the nodes in Step 1, start over from Step 1.

   ```
   # /etc/opt/FJSVcluster/bin/clmmbsetup -l
   cluster-host-name  user-name
   ----------------------------------
   node1              mmb-user
   node2              mmb-user
   ```

5.1.2.4.3 Setting up the Shutdown Daemon

On all the nodes, create /etc/opt/SMAW/SMAWsf/rcsd.cfg with the following information.

Create the rcsd.cfg file using root user access privileges and change the permission of the file to 600.

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_mmbp,timeout=timeout:agent=SA_mmbr,timeout=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_mmbp,timeout=timeout:agent=SA_mmbr,timeout=timeout
```

```
CFNameX       :  Specify the CF node name of the cluster host.
weight        :  Specify the weight of the SF node.
myadmIP       :  Specify the IP address of the administrative LAN that is used
                 by the shutdown facility of the cluster host.
                 It is not the IP address of MMB.
                 Available IP addresses are IPv4 and IPv6 addresses.
                 IPv6 link local addresses are not available.
                 When specifying the IPv6 address, enclose it in brackets "[ ]".
                 (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                 If you specify a host name, please make sure it is listed in /etc/hosts.
SA_mmbp       :  Make sure to specify this shutdown agent that panics the node via MMB.
SA_mmbr       :  Make sure to specify this shutdown agent that resets the node via MMB.
timeout       :  Specify the timeout duration (seconds) of the shutdown agent.
                 Specify 20 seconds for "SA_mmbp" and "SA_mmbr".
```

Example:

```
node1,weight=2,admIP=fuji2:agent=SA_mmbp,timeout=20:agent=SA_mmbr,timeout=20
node2,weight=2,admIP=fuji3:agent=SA_mmbp,timeout=20:agent=SA_mmbr,timeout=20
```

Note

- For the shutdown agents to be specified in the rcsd.cfg file, set both the SA_mmbp and SA_mmbr shutdown agents in that order.
- Set the same contents in the rcsd.cfg file on all the nodes. Otherwise, a malfunction may occur.

Information

When creating the /etc/opt/SMAW/SMAWsf/rcsd.cfg file, you can use the /etc/opt/SMAW/SMAWsf/rcsd.cfg.mmb.template file as a template.

5.1.2.4.4 Starting the MMB Asynchronous Monitoring Daemon

Start the MMB asynchronous monitoring daemon.

Check that the MMB asynchronous monitoring daemon has been started on all the nodes.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl
```

If "The devmmbd daemon exists." is displayed, the MMB asynchronous monitoring daemon has been started.

If "The devmmbd daemon does not exist." is displayed, the MMB asynchronous monitoring daemon has not been started. Execute the following command to start the MMB asynchronous monitoring daemon.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl start
```

5.1.2.4.5 Setting I/O Completion Wait Time

When using any disks other than the ETERNUS disk array as the shared disk, to prevent the data error when the node is down due to a panic or other causes, set the time until I/O to the shared disk is completed.

Execute the command in any node that is part of the cluster system, and set the wait time until I/O completion (WaitForIOComp) during failover triggered by a node failure (panic, etc.).

For details about the "cldevparam" command, see the "cldevparam" manual page.

```
# /etc/opt/FJSVcluster/bin/cldevparam -p WaitForIOComp value
```

value　　:　Specify the wait time until I/O completion.
　　Specify the time checked by the procedure described in
　　"5.1.2.4.1 Checking the Shutdown Agent Information."

After setting the wait time, execute the following command to check if the specified value is set.

```
# /etc/opt/FJSVcluster/bin/cldevparam -p WaitForIOComp
value
```

Note

If an I/O completion wait time is set, the failover time when a node failure (panic, etc.) occurs increases by that amount of time.

5.1.2.4.6 Starting the Shutdown Facility

On all the nodes, start or restart the shutdown facility

1. Starting the shutdown facility

   Check that the shutdown facility has been started on all the nodes.

   ```
   # sdtool -s
   ```

   If the shutdown facility has already been started, execute the following commands on all the nodes to restart the shutdown facility.

   ```
   # sdtool -e
   # sdtool -b
   ```

   If the shutdown facility has not been started, execute the following command on all the nodes to start the shutdown facility.

   ```
   # sdtool -b
   ```

2. Checking the status of the shutdown facility

   Check the status of the shutdown facility on all the nodes.

   ```
   # sdtool -s
   ```

Information

Display results of the sdtool -s command

- If InitFailed is displayed in Init State, it means that a problem occurred during initialization of that shutdown agent.
- If TestFailed is displayed in Test State, it means that a problem occurred while the agent was testing whether or not the node displayed in the Cluster Host field could be stopped. Some sort of problem probably occurred in the software, hardware, or network resources being used by that agent.
- If Unknown is displayed in Shut State or Init State, it means that SF has not yet executed node stop, path testing, or SA initialization. Unknown is displayed temporarily in Test State and Init State until the actual status can be confirmed.
- If TestFailed or InitFailed is displayed, check /var/log/messages. After the failure-causing problem is resolved and SF is restarted, the status display changes to InitWorked or TestWorked.

Note

- If TestFailed is displayed in Test State and the message 7210 is output to /var/log/messages at the same time when "sdtool -s" is executed after the shutdown facility was started, possible causes are as follows:

  Make sure each setting is correctly set.

```
7210 An error was detected in MMB. (node:nodename mmb_ipaddress1:mmb_ipaddress1
mmb_ipaddress2:mmb_ipaddress2
node_ipaddress1:node_ipaddress1 node_ipaddress2:node_ipaddress2 status:status detail:detail)
```

- PSA/Svmco is not installed or not set.
- A node is not restarted after installing Svmco manually.
- Incorrect PSA/Svmco settings

  Example: An incorrect IP address (such as MMB IP address) is set to the IP address of the administrative LAN.
- Necessary firewall to activate PSA/Svmco is not set.
- Incorrect MMB settings

  Example 1: An incorrect IP address is set.

  Example 2: Both the virtual IP address of MMB and the physical IP address of MMB are not set.

- If "sdtool -s" is executed immediately after the OS is started, TestFailed may be displayed in Test State for the local node. However, this state is displayed because the snmptrapd daemon is still being activated and does not indicate a malfunction. If "sdtool -s" is executed 10 minutes after the shutdown facility is started, TestWorked is displayed in Test State.

  In the following example, TestFailed is displayed in Test State for the local node (node1).

```
# sdtool -s
Cluster Host    Agent         SA State      Shut State  Test State  Init State
------------    -----         --------      ----------  ----------  ----------
node1           SA_mmbp.so    Idle          Unknown     TestFailed  InitWorked
node1           SA_mmbr.so    Idle          Unknown     TestFailed  InitWorked
node2           SA_mmbp.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_mmbr.so    Idle          Unknown     TestWorked  InitWorked
```

  The following messages may be displayed in the Syslog right after the OS is started by same reason as previously described.

```
3084: Monitoring another node has been stopped.
SA SA_mmbp.so to test host nodename failed
SA SA_mmbr.so to test host nodename failed
```

  These messages are also displayed because the snmptrapd daemon is being activated and does not indicate a malfunction. The following message is displayed in the Syslog 10 minutes after the shutdown facility is started.

```
3083: Monitoring another node has been started.
```

- If "sdtool -s" is executed when MMB asynchronous monitoring daemon is started for the first time, TestFailed may be displayed. This is a normal behavior because the settings are synchronizing between nodes. If "sdtool -s" is executed 10 minutes after the shutdown facility is started, TestWorked is displayed in Test State.
- If nodes are forcibly stopped by the SA_mmbr shutdown agent, the following messages may be displayed in the Syslog. These are displayed because it takes time to stop the nodes and do not indicate a malfunction.

```
Fork SA_mmbp.so(PID pid) to shutdown host nodename
        :
SA SA_mmbp.so to shutdown host nodename failed
        :
Fork SA_mmbr.so(PID pid) to shutdown host nodename
        :
SA SA_mmbr.so to shutdown host nodename failed
        :
MA SA_mmbp.so reported host nodename leftcluster, state MA_paniced_fsnotflushed
        :
MA SA_mmbr.so reported host nodename leftcluster, state MA_paniced_fsnotflushed
        :
Fork SA_mmbp.so(PID pid) to shutdown host nodename
```

```
        :
SA SA_mmbp.so to shutdown host nodename succeeded
```

If "sdtool -s" is executed after the messages above were displayed, KillWorked is displayed in Shut State for SA_mmbp.so. Then, KillFailed is displayed in Shut State for SA_mmbr.so.

The following is the example of "sdtool -s" when the nodes (from node1 to node2) were forcibly stopped and the messages above were displayed.

```
# sdtool -s
Cluster Host    Agent         SA State      Shut State  Test State  Init State
------------    -----         --------      ----------  ----------  ----------
node1           SA_mmbp.so    Idle          Unknown     TestWorked  InitWorked
node1           SA_mmbr.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_mmbp.so    Idle          KillWorked  TestWorked  InitWorked
node2           SA_mmbr.so    Idle          KillFailed  TestWorked  InitWorked
```

To recover KillFailed displayed by "sdtool -s," perform the following procedure.

```
# sdtool -e
# sdtool -b
# sdtool -s
Cluster Host    Agent         SA State      Shut State  Test State  Init State
------------    -----         --------      ----------  ----------  ----------
node1           SA_mmbp.so    Idle          Unknown     TestWorked  InitWorked
node1           SA_mmbr.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_mmbp.so    Idle          Unknown     TestWorked  InitWorked
node2           SA_mmbr.so    Idle          Unknown     TestWorked  InitWorked
```

5.1.2.4.7 Test for Forced Shutdown of Cluster Nodes

After setting up the shutdown facility, conduct a test for forced shutdown of cluster nodes to check that the correct nodes can be forcibly stopped.

For the detail of the test for forced shutdown of cluster nodes, refer to "1.4 Test."

5.1.2.5 Setup Procedure for Shutdown Facility in Virtual Machine Environment

This section describes the setup procedure of the shutdown facility in the virtual machine environment.

Note

When creating a redundant administrative LAN used in the shutdown facility by using GLS, set as below.

- For taking over the IP address between nodes

  Configure GLS by using the logical IP address takeover function of the NIC switching mode.

  For shutdown facility, specify a physical IP address instead of a logical IP address.

- For not taking over the IP address between nodes

  Configure GLS by using the physical IP address takeover function of the NIC switching mode.

5.1.2.5.1 Checking the Shutdown Agent Information

To forcibly stop the domain in the guest OS by using the shutdown facility in KVM environment, log in to the host OS via SSH.

Check in advance the following settings that are necessary for setting the shutdown facility.

- IP address for the host OS
- User for logging in to the host OS
- User password for logging in to the host OS

- Domain name for the gust OS

For information on the user and password for logging in to the host OS, check the following information set up by the procedures described in the following sections:

- When building a cluster system between guest OSes on one host OS, see "3.2.1.2 Host OS setup (after installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes without using the host OS failover function, see "3.2.2.2 Host OS setup (after installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes using host the OS failover function, see "3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)."

Also take the following steps to check that the setting of the sudo command is already completed.

This is necessary for the confirmed user to execute the command as the root user.

1. Execute the visudo command on all the nodes.
2. Check that the following setting is described in the setting file displayed by executing the visudo command.

```
<User ID>    ALL=(root) NOPASSWD: ALL
```

   If this setting information is missing, describe it to the file.

5.1.2.5.2 Setting up libvirt Shutdown Agent

Set up the libvirt shutdown agent.

Take the following steps.

Note

Be sure to perform the following operations from 1. to 3. on all guest OSes (nodes).

1. Encrypt the password.

   Execute the sfcipher command to encrypt the password that was checked in "5.1.2.5.1 Checking the Shutdown Agent Information."

   For details on how to use the sfcipher command, see the manual page of "sfcipher."

```
# sfcipher -c
Enter User's Password:
Re-enter User's Password:
D0860AB04E1B8FA3
```

2. Set up the panicky shutdown agent (SA_libvirtgp) and reset shutdown agent (SA_libvirtgr).

   Set up the panicky shutdown agent (SA_libvirtgp) and reset shutdown agent (SA_libvirtgr).

   Create the /etc/opt/SMAW/SMAWsf/SA_libvirtgp.cfg and the /etc/opt/SMAW/SMAWsf/SA_libvirtgr.cfg as below.

   Create the /etc/opt/SMAW/SMAWsf/SA_libvirtgp.cfg and the /etc/opt/SMAW/SMAWsf/SA_libvirtgr.cfg by using the root user privilege, and change the permission of the file to 600.

```
CFNameX domainX ip-address user passwd
CFNameX domainX ip-address user passwd
```

```
CFNameX    : Specify the CF node name of the cluster host.
domainX    : Specify the guest OS domain name.
             Specify the domain name checked in
             "5.1.2.5.1 Checking the Shutdown Agent Information."
ip-address : Specify the IP address of the host OS.
             Specify the IP address of the host OS checked in
             "5.1.2.5.1 Checking the Shutdown Agent Information."
             Available IP addresses are IPv4 and IPv6 addresses.
```

```
              IPv6 link local addresses are not available.
user        : User to log in to the host OS.
              Specify the user checked in
              "5.1.2.5.1 Checking the Shutdown Agent Information."
passwd      : Password of the user specified by "user".
              Specify the encrypted password that you have checked in 1.
```

Example:

When the guest OS domain name of node1 is domain1, and the IP address of the host OS on which node1 operates is 10.20.30.50. Moreover, when the guest OS domain name of node2 is domain2, and the IP address of the host OS on which node2 operates is 10.20.30.51.

- /etc/opt/SMAW/SMAWsf/SA_libvirtgp.cfg

```
node1 domain1 10.20.30.50 user D0860AB04E1B8FA3
node2 domain2 10.20.30.51 user D0860AB04E1B8FA3
```

- /etc/opt/SMAW/SMAWsf/SA_libvirtgr.cfg

```
node1 domain1 10.20.30.50 user D0860AB04E1B8FA3
node2 domain2 10.20.30.51 user D0860AB04E1B8FA3
```

- Check if the setting contents of the /etc/opt/SMAW/SMAWsf/SA_libvirtgp.cfg file and the /etc/opt/SMAW/SMAWsf/SA_libvirtgr.cfg file are correct. If there is an error in the setting contents, the shutdown facility cannot be performed normally.
- Check if the domain name (domainX) of the guest OS and the IP address (ip-address) of the host OS corresponding to the cluster host's CF node name (CFNameX) of the /etc/opt/SMAW/SMAWsf/SA_libvirtgp.cfg file and the /etc/opt/SMAW/SMAWsf/SA_libvirtgr.cfg file are set. If there is an error in the setting, a different node may be forcibly stopped.
- The contents of the SA_libvirtgp.cfg, SA_libvirtgr.cfg, and rcsd.cfg files of all guest OSes (nodes) should be identical. If not, a malfunction will occur.

3. Log in to the host OS

   The shutdown facility accesses the host OS with SSH. Therefore, you need to authenticate yourself (create the RSA key) in advance, which is required when using SSH for the first time.
   On all guest OSes (nodes), log in to each host OS IP address (ip-address) set in the step 2. using each set user.

   Execute the command as the root user access privilege.

```
# ssh -l user 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."
#
```

5.1.2.5.3 Setting Up vmchkhost Shutdown Agent

When using the host OS failover function, set up the vmchkhost shutdown agent.

Perform this setting after setting up the libvirt shutdown agent.

Be sure to perform the following operations from 2. to 3. on all guest OSes (nodes).

1. Set up the libvirt shutdown agent and check the information of the host OS.

   Check the following information that are set in the libvirt shutdown agent:

- IP address for the host OS
- User for logging in to the host OS
- Encrypted user password for logging in to the host OS

Also check that the following information for the host OS.

- CF node name

2. Set up the vmchkhost shutdown agent.

   Create /etc/opt/SMAW/SMAWsf/SA_vmchkhost.cfg as described in the following.

   Create the SA_vmchkhost.cfg using the root user access privilege and change the permission of the file to 600.

```
guest-cfnameX host-cfnameX ip-address user password
guest-cfnameX host-cfnameX ip-address user password
```

```
guest-cfnameX      : CF node name of the guest OS (cluster node).
host-cfnameX       : CF node name of the host OS.
                     Specify the CF node name checked in step 1.
ip-address         : An IP address of the host OS.
                     Specify the IP address checked in step 1.
user               : User to log in to the host OS.
                     Specify the user checked in step 1.
password           : Password of the user specified by "user".
                     Specify the encrypted password checked in 1.
```

   Example:

   When the CF node name of the host OS on which node1 (CF node name of the guest OS) operates is hostos1, the IP address of the host OS is 10.20.30.50, the CF node name of the host OS on which node2 (CF node name of the guest OS) operates is hostos2, and the IP address of the host OS is 10.20.30.51.

```
node1 hostos1 10.20.30.50 user D0860AB04E1B8FA3
node2 hostos2 10.20.30.51 user D0860AB04E1B8FA3
```

- Check if the setting contents of the /etc/opt/SMAW/SMAWsf/SA_vmchkhost.cfg file are correct. If there is an error in the setting contents, the shutdown facility cannot be performed normally.
- Check if the CF node name of the host OS (host-cfnameX) and the IP address of the host OS (ip-address) corresponding to the CF node name (guest-cfnameX) of the guest OS (clutser host) of the /etc/opt/SMAW/SMAWsf/SA_vmchkhost.cfg file are set. If there is an error in the setting, the shutdown facility cannot be performed normally.
- The contents of the SA_vmchkhost.cfg file of all guest OSes (nodes) should be identical. If not, a malfunction will occur.

3. Log in to the host OS

   The shutdown facility accesses the host OS with SSH. Therefore, you need to authenticate yourself (create the RSA key) in advance, which is required when using SSH for the first time.

   Check that you have already authenticated yourself (created the RSA key) when setting up the libvirt shutdown agent.

5.1.2.5.4 Setting up the Shutdown Daemon

On all the nodes, create /etc/opt/SMAW/SMAWsf/rcsd.cfg with the following information.

Create the rcsd.cfg file using root user access privileges and change the permission of the file to 600.

- When using the host OS failover function

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_libvirtgp,timeout=timeout:agent=SA_libvirtgr,timeout
=timeout:agent=SA_vmchkhost,timeout=timeout
```

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_libvirtgp,timeout=timeout:agent=SA_libvirtgr,timeout
=timeout:agent=SA_vmchkhost,timeout=timeout
```

- When not using the host OS failover function

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_libvirtgp,timeout=timeout:agent=SA_libvirtgr,timeout
=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_libvirtgp,timeout=timeout:agent=SA_libvirtgr,timeout
=timeout
```

```
CFNameX        :  Specify the CF node name of the cluster host.
weight         :  Specify the weight of the SF node.
myadmIP        :  Specify the IP address of the administrative LAN that used by the Shutdown
                  Facility of the cluster host.
                  It is not the IP address of iRMC or the management blade.
                  Available IP addresses are IPv4 and IPv6 addresses.
                  IPv6 link local addresses are not available.
                  When specifying the IPv6 address, enclose it in brackets "[ ]".
                  (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                  If you specify a host name, please make sure it is listed in /etc/hosts.
SA_libvirtgp   :  Make sure to set this shutdown agent that panics the guest OS.
SA_mmbr        :  Make sure to set this shutdown agent that resets the guest OS.
SA_vmchkhost   :  Shutdown agent for the host OS failover function.
timeout        :  Specify the timeout duration (seconds) of the shutdown agent.
                  Specify 35 seconds for SA_libvirtgp, SA_libvirtgr, and SA_vmchkhost.
```

Example1: When using the host OS failover function

```
node1,weight=2,admIP=fuji2:agent=SA_libvirtgp,timeout=35:agent=SA_libvirtgr,timeout=35:agent=SA_vmch
khost,timeout=35
node2,weight=1,admIP=fuji3:agent=SA_libvirtgp,timeout=35:agent=SA_libvirtgr,timeout=35:agent=SA_vmch
khost,timeout=35
```

Example 2: When not using the host OS failover function

```
node1,weight=2,admIP=fuji2:agent=SA_libvirtgp,timeout=35:agent=SA_libvirtgr,timeout=35
node2,weight=1,admIP=fuji3:agent=SA_libvirtgp,timeout=35:agent=SA_libvirtgr,timeout=35
```

Note

- SA_libvirtgp shutdown agent must be set first followed by SA_libvirtgr, and then set SA_vmchkhost as the last of all in the rcsd.cfg file.
- Set the same contents in the rcsd.cfg file on all the nodes. Otherwise, a malfunction may occur.

Information

When creating the /etc/opt/SMAW/SMAWsf/rcsd.cfg file, you can use the /etc/opt/SMAW/SMAWsf/rcsd.cfg.mmb.template file as a template.

5.1.2.5.5 Starting the Shutdown Facility

Start or restart the shutdown facility on all the nodes.

1. Starting the shutdown facility

   Check that the shutdown facility has been started on all the nodes.

   ```
   # sdtool -s
   ```

   If the shutdown facility has already been started, execute the following commands to restart the shutdown facility on all the nodes.

```
# sdtool -e
# sdtool -b
```

If the shutdown facility has not been started, execute the following command to start the shutdown facility on all the nodes.

```
# sdtool -b
```

2. Checking the status of the shutdown facility

   Check the status of the shutdown facility on all the nodes.

```
# sdtool -s
```

Information

Display results of the sdtool -s command

- If InitFailed is displayed in Init State, it means that a problem occurred during initialization of that shutdown agent.
- If TestFailed is displayed in Test State, it means that a problem occurred while the agent was testing whether or not the node displayed in the Cluster Host field could be stopped. Some sort of problem probably occurred in the software, hardware, or network resources being used by that agent.
- If Unknown is displayed in Shut State or Init State, it means that SF has not yet executed node stop, path testing, or SA initialization. Unknown is displayed temporarily in Test State and Init State until the actual status can be confirmed.
- If TestFailed or InitFailed is displayed, check the following files:
  - /var/log/messages
  - /etc/sysconfig/libvirt-guests

  For /etc/sysconfig/libvirt-guests, check whether the following settings are made:

  - When building a cluster system between guest OSes on one host OS, see "3.2.1.2 Host OS setup (after installing the operating system on guest OS)."
  - When building a cluster system between guest OSes on multiple host OSes without using the host OS failover function, see "3.2.2.2 Host OS setup (after installing the operating system on guest OS)."
  - When building a cluster system between guest OSes on multiple host OSes using the host OS failover function, see "3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)."

  After the failure-causing problem is resolved and SF is restarted, the status display changes to InitWorked or TestWorked.

5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)

When using the host OS failover function in PRIMEQUEST, for linking with MMB asynchronous monitoring function, configure the host OS failover function to the host OS.

Set up this setting after setting libvirt shutdown agent and vmchkhost shutdown agent.

Note

Be sure to perform the following operations from 3. to 7. on all the host OSes (nodes).

1. Check the setting information.

   The host OS failover function in PRIMEQUEST, when detecting an host OS error by MMB asynchronous monitoring function, logs in to a guest OS (a cluster node) using SSH and then notifies the shutdown facility of the host OS error.

   For setting the host OS failover function, confirm the following necessary information previously.

   - IP address of the guest OS

- Domain name of the guest OS
- Cluster name of the guest OS
- CF node name of the guest OS

2. Create the user (when logging in to the guest OS not as a root user).

   When the host OS failover function logs in to the guest OS not as a root user, a user for logging in is created. Perform the following procedure on all the guest OS.

   (1) Create the login user.

   Set the user password with seven-bit ASCII characters except the following characters.

   ```
   >  <  "  /  \  =  !  ?  ;  ,  &
   ```

   (2) Set the sudo command so that the created user can execute the command as a root user.

   Execute the visudo command by using the root command. Describe the following setting in the displayed setting file.

   ```
   <User created in (1)>    ALL=(root) NOPASSWD: ALL
   ```

3. Encrypt the password.

   Execute the sfcpher command to encrypt passwords for login to the guest OS as a root user.

   For details on how to use the sfcipher command, see the manual page of "sfcipher."

   ```
   # sfcipher -c
   Enter User's Password:
   Re-enter User's Password:
   D0860AB04E1B8FA3
   ```

4. Create /etc/opt/FJSVcluster/etc/kvmguests.conf.

   Create /etc/opt/FJSVcluster/etc/kvmguests.conf with the following contents.

   Create the kvmguests.conf file using the root user access privilege and change the permission of the file to 600.

   When multiple guest OSes (the cluster nodes) are operating on a host OS that configures the cluster, describe all the guest OSes configured the host OS failover function in this file.

   ```
   guest-name host-cfname guest-clustername guest-cfname guest_IP guest_user guest_passwd
             :
   ```

   - Enter the information of one node in one line.
   - Delimit each item with a single space.
   - The kvmguests.conf file must be the same on all cluster nodes.

   ```
   guest-name          :Specify the domain name of the guest OS.
   host-cfname         :Specify the CF node name of the host OS in which "guest-name" is running.
                        If you execute "cftool -l" on the host OS in which "guest-name" is running,
                        you can confirm the CF node name of the node.
   guest-clustername   :Specify the cluster name of the guest OS.
                        If you execute "cftool -c" on the guest OS, you can confirm the cluster
                        name of the node.
   guest-cfname        :Specify the CF node name of the guest OS.
                        If you execute "cftool -l" on the guest OS, you can confirm the CF node
                        name of the node.
   guest_IP            :Specify the IP address of the guest OS.
                        Available IP address formats are IPv4 and IPv6 addresses.
                        IPv6 link local addresses are not available.
   guest_user          :Specify the user for logging in to the guest OS using SSH.
                        Specify the fixed root or the user created in step 2.
   guest_passwd        :Specify the user password for logging in to the guest OS.
                        Specify the password encrypted in step 3.
   ```

Example: In a two-node configuration between guest OSes, two cluster systems are configured

```
guest11 cfhost1 cluster1 cfguest11  10.20.30.50 root D0860AB04E1B8FA3
guest12 cfhost2 cluster1 cfguest12  10.20.30.51 root D0860AB04E1B8FA3
guest21 cfhost1 cluster2 cfguest21  10.20.30.60 root D0860AB04E1B8FA3
guest22 cfhost2 cluster2 cfguest12  10.20.30.61 root D0860AB04E1B8FA3
```

5. Confirm the log in to the guest OS

   The host OS failover function in PRIMEQUEST accesses the guest OS with SSH. Therefore, you need to authenticate yourself (create the RSA key) in advance, which is required when using SSH for the first time.

   Check that you can connect to all the guest OSes (nodes) which are specified to /etc/opt/FJSVcluster/etc/kvmguests.conf via SSH as a root user.

```
# 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 <- Enter "yes."
```

6. Check the setting in /etc/opt/FJSVcluster/etc/kvmguests.conf.

   Execute the sfkvmtool command on all the host OSes to make sure that the settings in /etc/opt/FJSVcluster/etc/kvmguests.conf are correct.

   If the settings are correct, the following message is output.

```
# /opt/SMAW/SMAWsf/bin/sfkvmtool -c
NOTICE: The check of configuration file succeeded.
```

   If a message other than above is output, review the setting in /etc/opt/FJSVcluster/etc/kvmguests.conf.

7. Start the shutdown facility

   Check that the shutdown facility has already been started on all the nodes.

```
# sdtool -s
```

   If the shutdown facility has already been started, execute the following on all the nodes to restart it.

```
# sdtool -e
# sdtool -b
```

   If the shutdown facility has not been started, execute the following on all the nodes to start it.

```
# sdtool -b
```

5.1.2.5.7 Test for Forced Shutdown of Cluster Nodes

After setting up the shutdown facility, conduct a test for forced shutdown of cluster nodes to check that the correct nodes can be forcibly stopped.

For the detail of the test for forced shutdown of cluster nodes, refer to "1.4 Test."

Note

After shutting down a node (a guest OS) forcibly by SA_libvirtgp, the guest OS may be a temporary stopped state. (For example, when there is no space in /var/crash on the host OS.) In the case, forcibly shutdown the guest OS by the virsh destroy command.

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.

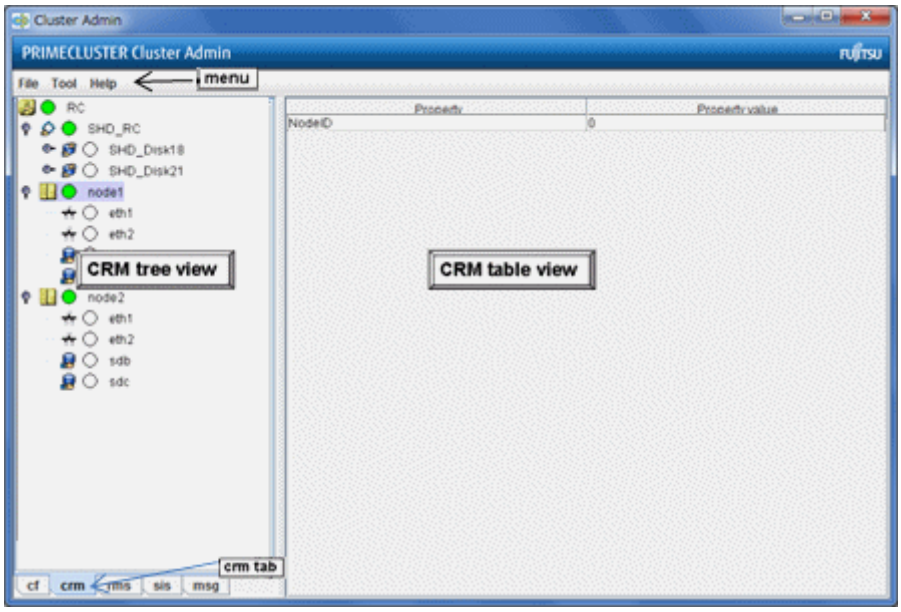
2. Registering Hardware Devices

   Register the connected hardware devices (shared disks and network interface cards) to the resource database that CRM manages.

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

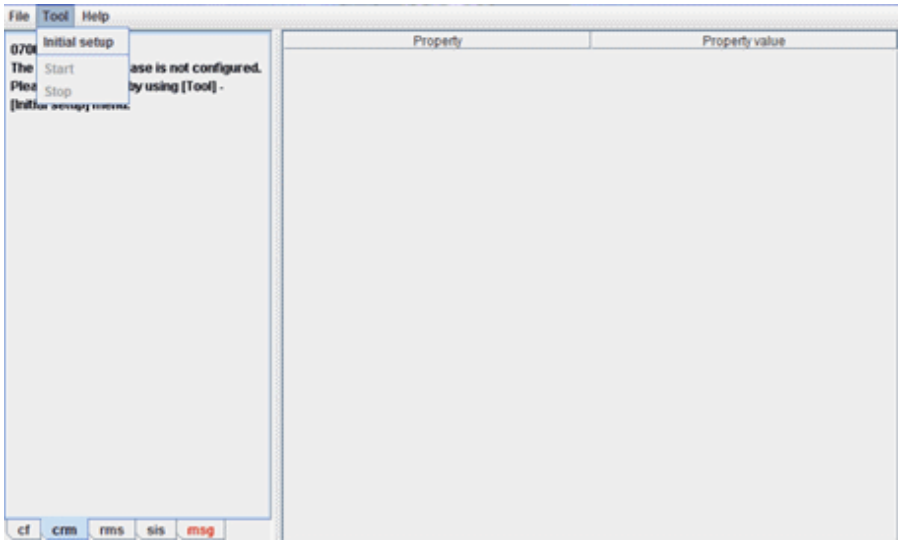
Set up the resource database that CRM manages.

When setting up the initial configuration, make sure that all the 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 5.1 Screen for cluster resource management facility

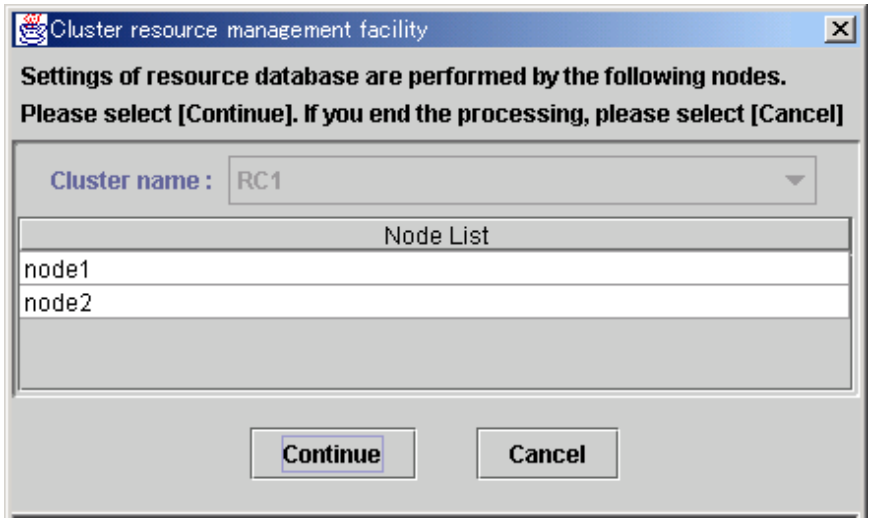


Note

The *Initial setup* can be selected only if the resource database has not been set.

2. The screen for initial setup is displayed.

Figure 5.2 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 the nodes displayed by selecting the *cf* tab in the Cluster Admin screen are Up.
- Whether Web-Based Admin View is operating in all the nodes.

For instructions on checking this, see "4.3.3.2 Confirming Web-Based Admin View Startup."

Continue button

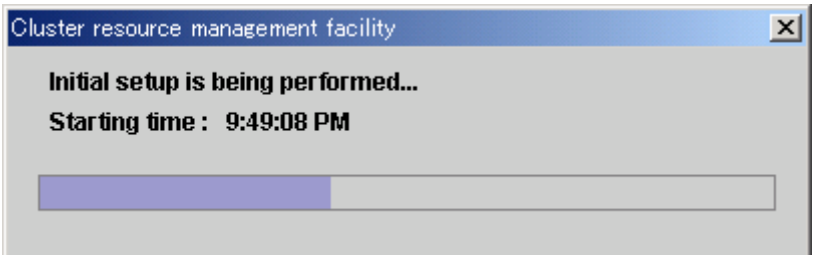
Click this button to set up the resource database for the displayed cluster.

Initial setup is executed on all the nodes displayed in the Node list.

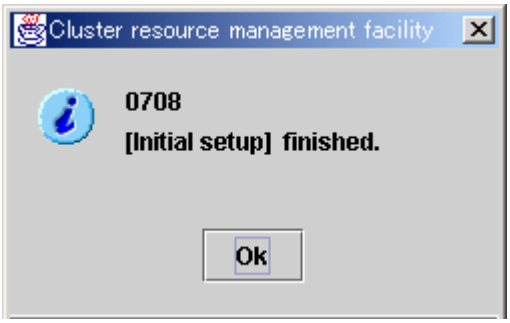
Cancel button

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.



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 "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 Registering Hardware Devices

Register the connected hardware devices (shared disks and network interface cards) to the resource database that CRM manages.

Operation Procedure:

1. Registering the network interface card
   1. Confirm that all the nodes have been started in multi-user mode.
   2. Perform the following procedure on any node in the cluster system.
      1. Log in the node using system administrator access privileges.

2. Execute the "clautoconfig" command.

```
# /etc/opt/FJSVcluster/bin/clautoconfig -r -n
```

For details on this command, see the manual pages of "clautoconfig".

Note

- Do not execute the "clautoconfig" command on the node in which the "clautoconfig" command is being executed or on any other node while the "clautoconfig" command is being executed. If you execute it, a shared disk device cannot be registered correctly. If you have executed it, execute the following operation on all the nodes that constitute the cluster system to re-execute "5.1.3 Initial Setup of the Cluster Resource Management Facility" described in this chapter:

  1. Reset the resource database using the "clinitreset" command. For details on this command, see the manual pages of "clinitreset".
  2. Restart the node.

- Only an active network interface card is automatically detected. Confirm the state of the network interface card using the "ip(8)" command in RHEL7 or later, or the "ifconfig(8)" command in RHEL6 or earlier.
  Execute the following command to activate the network interface:

  [RHEL6 or earlier]

  ```
  # ifconfig network interface card up
  ```

  (Example) Enable the network interface card eth1

  ```
  # ifconfig eth1 up
  ```

  [RHEL7 or later]

  ```
  # ip link set dev network interface card up
  ```

  (Example) Enable the network interface card eth1

  ```
  # ip link set dev eth1 up
  ```

2. Registering a shared disk

When you use GDS, register a shared disk in the resource database using the following steps on any one of the nodes of the cluster system. These steps are required also when performing the mirroring among servers.
For details on the procedure, see "Shared Disk Resource Registration" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

1. Log in any one of the nodes of the cluster system using system administrator access privileges.
2. Set the disk for performing the mirroring among servers.

   For performing the mirroring among servers, set the local disk device to be accessed from each node as an iSCSI device.

   For details, see "Disk Setting for Performing Mirroring among Servers" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

   By this setting, the target disk device can be used from each node as the shared disk device is used. For the procedure below, describe the iSCSI device in the shared disk definition file.
3. Create a shared disk configuration file in the following format.

   The configuration file defines settings of a shared disk connected to all the nodes of the cluster system.

   Create a shared disk definition file with an arbitrary name.

```
<Resource key name> <device name> <node identifier>
<Resource key name> <device name> <node identifier>
                    :
```

- Define "resource key name device name node identifier" for each shared disk in one row.
- "resource key name", "device name", and "node identifier" are delimited by a single space.
- Set up resource key name, device name and node identifier as follows;

Resource key name

Specify a resource key name that indicates the sharing relationship for each shared disk. You need to specify the same name for each disk shared between nodes. The resource key name should be specified in the "shd number" format. "shd" is a fixed string. For "number", you can specify any four-digit numbers. If multiple shared disks are used, specify unique numbers for each shared disk.

(Example) When /dev/sdb and /dev/sdc are shared between nodes

```
Resource key name of /dev/sdb: shd0001
Resource key name of /dev/sdc: shd0002
```

Device name

Specify a device name by the full device path of the shared disk.

(Example) In the case of /dev/sdb

```
/dev/sdb
```

Note

- When using DM-MP
  - Describe a device name with /dev/mapper/mpath*X* format.
  - Do not describe a device name with /dev/dm-*X* format.
  - Do not describe a native device (sd device) which composes mpath devices.
- For a guest in the virtual environment

  Describe a device name for a guest.

  For example, for the virtio block device of the KVM guest, describe the device name for the KVM guest /dev/vd*X*, not the device name for the host OS /dev/sd*X*.

Node identifier

Specify a node identifier for which a shared disk device is available. Confirm the node identifier by executing the "clgettree" command. For details on this command, see the manual pages of "clgettree".

(Example) node1 and node2 are node identifiers in the following case:

```
# /etc/opt/FJSVcluster/bin/clgettree
Cluster 1 cluster
    Domain 2 PRIME
        Shared 7 SHD_PRIME
        Node 3 node1 ON
        Node 5 node2 ON
```

The following example shows the configuration file of the shared disk when shared disks /dev/sdb and /dev/sdc are shared between node1 and node2.

```
shd0001 /dev/sdb node1
shd0001 /dev/sdb node2
shd0002 /dev/sdc node1
shd0002 /dev/sdc node2
```

- When adding a shared disk device and registering the added shared disk device on the resource database, define only the information of the added shared disk device.

  Example: When registering the added disk device /dev/sdd (*1) on the resource database after shd0001 and shd0002 are already registered on the resource database:

```
shd0003 /dev/sdd node1
shd0003 /dev/sdd node2
```

  (*1) Note

  The device name of the added shared disk device may not follow the device name of the registered device in alphabetical order. Make sure to check the device name of the added shared disk device before defining the information of the added disk device.

4. Execute the "clautoconfig" command to register the settings of the shared disk device that is stored in the configuration file in the resource database.

   Specify the "clautoconfig" command in the following format:

   (Format)

   /etc/opt/FJSVcluster/bin/clautoconfig -f [full path of the shared disk definition file]

   (Example)

```
# /etc/opt/FJSVcluster/bin/clautoconfig -f /var/tmp/diskfile
```

Note

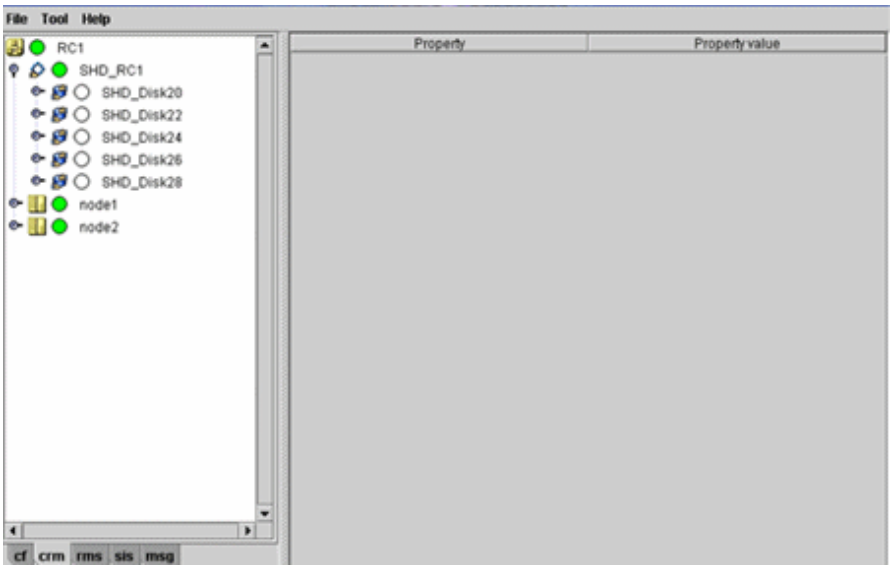
- If the "clautoconfig" command ends abnormally, take corrective action according to the error message. For details on the messages of this command, see "PRIMECLUSTER Messages."
- This command does not check whether the shared disk defined in the configuration file is physically connected.
- If the device name of the shared disk device varies depending on a node, execute the "clautoconfig" command on the nodes in which all the device files written in the shared disk configuration file exist. If a device file written in the shared disk configuration file does not exist on the node in which the "clautoconfig" command is executed, the resource registration fails and the following message is displayed.

  FJSVcluster: ERROR: clautoconfig: 6900: Automatic resource registration processing terminated abnormally.
  (detail: /dev/*device_name*)

  For details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."
- If you found an error in the shared disk configuration file after executing the "clautoconfig" command, reset the resource database by executing the "clinitreset" command and restart the node.

3. Registration completed

When the registration of the hardware device is completed, the following screen appears.



4. Checking registered resource

Go to the CRM main window and confirm that the resource registration is completed successfully by checking the following. In the CRM main window of Initial Setup, check that the selected device resources are registered correctly.

- Whether the disk configuration is different among the nodes.
- Whether the number of disks in each node differs from the number of shared disk units.
- 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.

If the actual device configuration and the resources do not match each other as described above, the following may be the causes:

- There is a connection path failure (such as cable disconnection) 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.

If the resources are not registered correctly, first review the above causes, and then register the resources again.

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 "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.

5.2 Setting Up Fault Resource Identification and Operator Intervention Request

The fault resource identification is a function that outputs a message to syslogd(8) 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, specify the settings for enabling fault resource identification and operator intervention request. An example of a message displayed by fault resource identification is shown below.

```
6750: A resource has faulted. 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(8) 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.: number
```

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 "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

Note

- To view the manual pages of each command, add "/etc/opt/FJSVcluster/man" to the MANPATH variable.

Preparation prior to displaying fault resource identification and operator intervention request messages

The fault resource identification and operator intervention request messages are displayed by using syslogd(8)/rsyslogd(8). daemon.err is specified to determine the priority (facility.level) of the fault resource identification and operator intervention request messages.

For details on the priority, see the manual page describing syslog.conf(5)/rsyslogd.conf(5).

If the fault resource identification and operator intervention request messages need to be output to the console, execute the following procedure on all the nodes.

Procedure:

1. Log in the node using system administrator access privileges.
2. Check the setting of syslogd/rsyslogd and change it so that daemon.err will be displayed on the console.
   - RHEL6
     1. Check the setting of rsyslogd in /etc/rsyslog.conf to see that daemon.err is set to be displayed on the console.

        (Example) Daemon.err is set to be displayed on the console.

        ```
        daemon.err        /dev/console
        ```

        For further details on /etc/rsyslog.conf, see the manual pages of rsyslog.conf(5).

     2. If daemon.err is not set to be displayed on the console, change the setting of rsyslogd in /etc/rsyslog.conf.

        To enable this change, restart the system log daemon by executing the following command.

        ```
        # /etc/init.d/rsyslog restart
        ```

   - RHEL7
     1. Check the setting of rsyslogd in /etc/rsyslog.conf to see that daemon.err is set to be displayed on the console.

        (Example) Daemon.err is set to be displayed on the console.

```
daemon.err        /dev/console
```

For further details on /etc/rsyslog.conf, see the manual pages of rsyslog.conf(5).

2. If daemon.err is not set to be displayed on the console, change the setting of rsyslogd in /etc/rsyslog.conf.

   To enable this change, restart the system log daemon by executing the following command.

```
# systemctl restart rsyslog.service
```

3. Starting the console.

   If you are using the graphical environment, execute the following command to start the console. For the textual environment or the remote environment using SSH or Telnet, this step is not required.

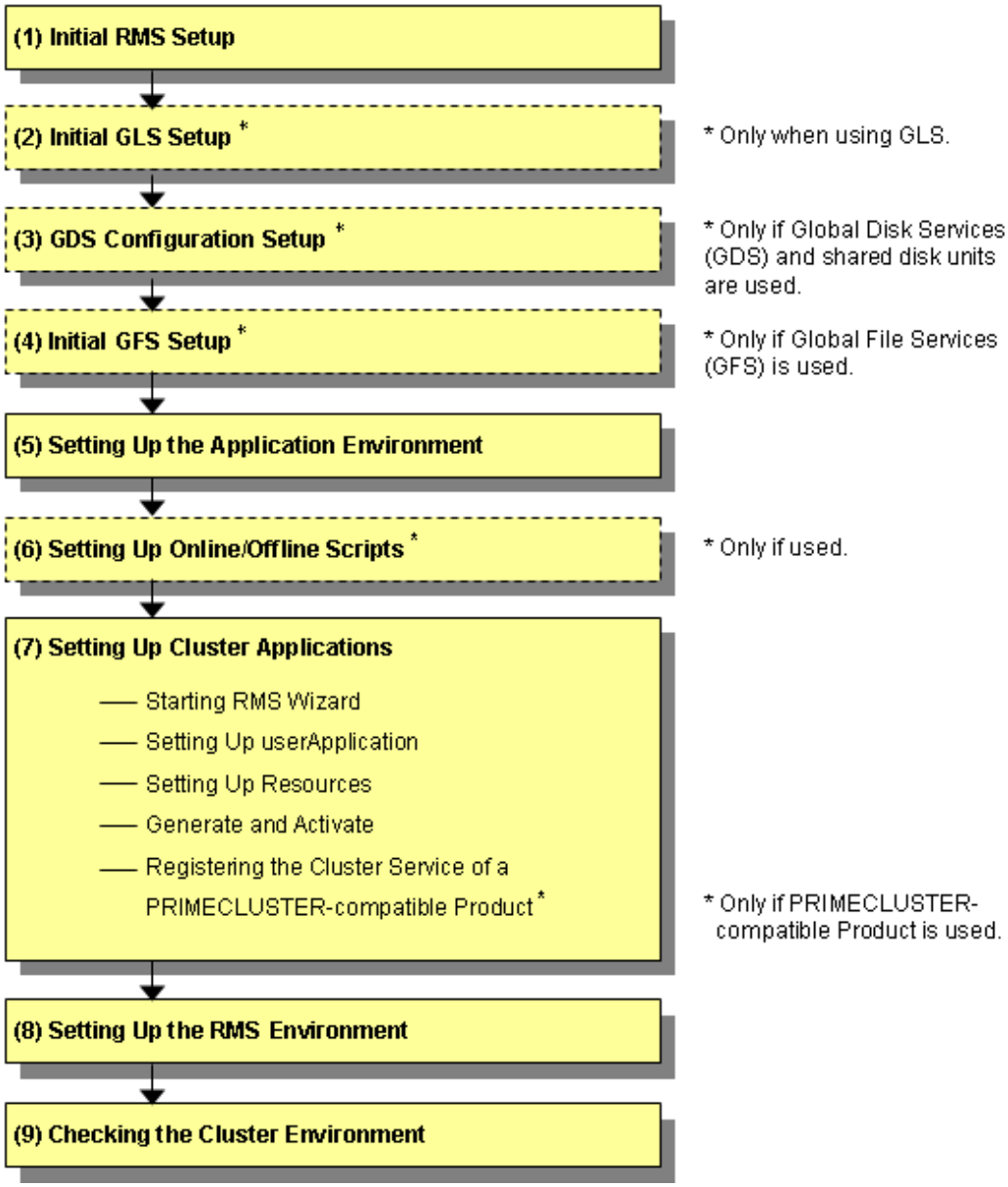
```
# xterm -C
```

Identifying the fault resource and changing the operation setting of operator intervention request

Use the clsetparam command to change the setting. For details, see the manual page of 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	GLSR

Work item		Execution Nodes	Required/ optional	Manual reference location*
			when the GLS redundant line control function is used)	
(3)	6.3 GDS Configuration Setup	All nodes	Optional (required when GDS is used)	GDSG "Chapter 5 Operation"
(4)	6.4 Initial GFS Setup	All nodes	Optional (required when GFS is used)	GFSG
(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," "12 Appendix - Environment variables"
(7)	6.7.1 Starting RMS Wizard	All nodes	Required	-
	6.7.2 Setting Up userApplication			
	6.7.3 Setting Up Resources			
	6.7.4 Generate and Activate			
	6.7.5 Registering the Cluster Service of a PRIMECLUSTER-compatible product	All nodes	Optional (required when a PRIMECLUSTER-compatible product is used)	
(8)	6.8 Setting Up the RMS Environment	All nodes	Required	RMS "2.9 Environment variables," "12 Appendix - Environment variables"
(9)	6.9 Checking the Cluster Environment	All nodes	Required	-

GLS: Global Link Services
GDS: Global Disk Services
GFS: Global File Services

***** The names of the reference PRIMECLUSTER manuals are abbreviated as follows:

- **RMS:** PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide
- **GDSG:** PRIMECLUSTER Global Disk Services Configuration and Administration Guide
- **GFSG:** PRIMECLUSTER Global File Services Configuration and Administration Guide
- **GLSR:** PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function

6.1 Initial RMS Setup

When RMS is to be used, you must first check "Setup (initial configuration)" of PRIMECLUSTER Designsheets and change the following environment variable as required:

- Automatic startup of RMS (HV_RCSTART)

In this version, "Start up automatically" is set as default.

If you want to set RMS not to be started automatically along with node startup, select "Does not start up automatically."

When RMS is to be used, we recommend that you set "Start up automatically."

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 GLS.

6.2.1 GLS Setup

For information on the initial GLS setup, see "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

This section describes how to set up "Example of the Cluster system" of "Example of configuring NIC switching mode (IPv4)" that GLS (redundant line control function) provides. This procedure is described in the example below.

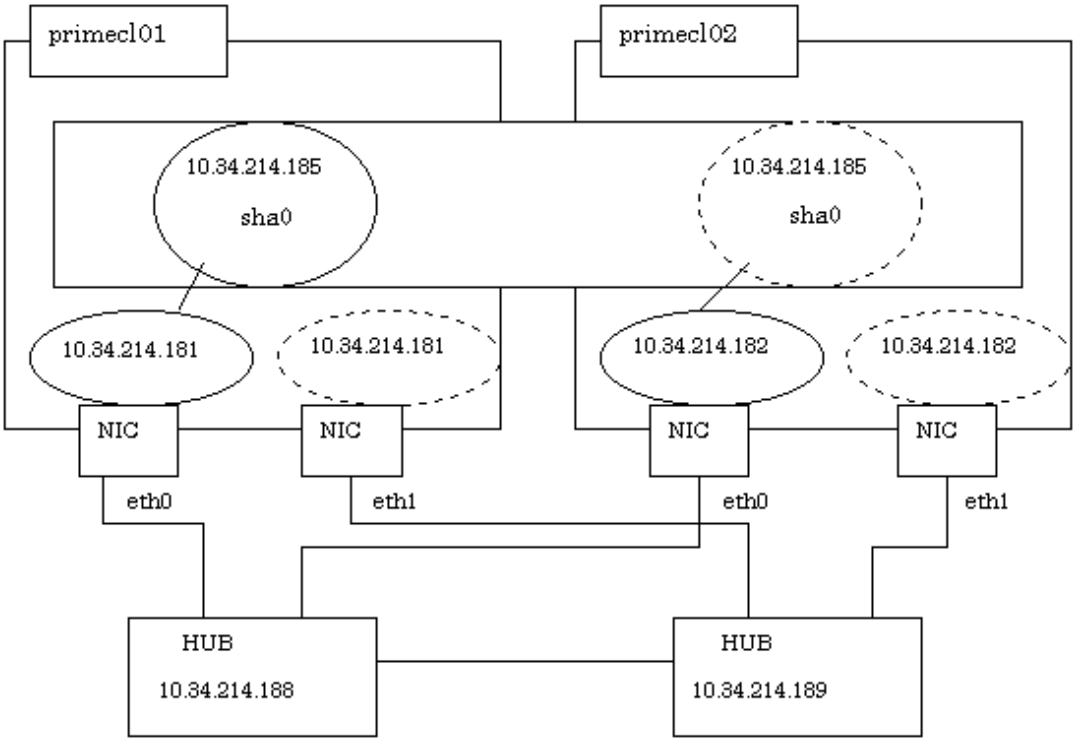
Note

Although it is possible to have "takeover network" for PRIMECLUSTER and "IP address takeover" for GLS on the same cluster system, you must not configure them on the same interface. If you do so, the communication through "takeover IP address" will be disabled.

For example, when you select 'eth1' for the interface when you set "takeover network" for PRIMECLUSTER, do not use 'eth1' for GLS environment settings (do not specify 'eth1' by using the '-t' option for "hanetconfig create" command).

When you need to duplex the interface for a takeover network, use "IP address takeover" for GLS. You cannot set "takeover network" for the bonding interface.

The setup values correspond to the values in "Setup (GLS_Monitoring Parameter)", "Setup (GLS_Virtual Interface)", "Setup (GLS_GS Linkage Mode Monitoring Destination Information)", and "Setup (GLS_Common Parameter)" of PRIMECLUSTER Designsheets.



Operation Procedure:

If the OPERATING node is [HOST-primecl01]

1. Setting up the system

    1. Define the IP address and Host name in /etc/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. Specify the IP address defined in Step 1-1 above to the /etc/sysconfig/network-scripts/ifcfg-ethX (X is either 0 or 1) file.

        - Contents of /etc/sysconfig/network-scripts/ifcfg-eth0

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
IPADDR=10.34.214.181
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

        - Contents of /etc/sysconfig/network-scripts/ifcfg-eth1

```
DEVICE=eth1
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
ONBOOT=yes
TYPE=Ethernet
```

Note

If you specify "HWADDR=XX:XX:XX:XX:XX:XX" in the settings for the physical interfaces bundled by GLS (the /etc/sysconfig/network-scripts/ifcfg-ethX file), add "HOTPLUG=no."

Information

Setting of "HOTPLUG=no" does not disable the PCI hot plug function.

You can perform hot maintenance for NIC (PCI card) to the physical interfaces with "HOTPLUG=no."

3. When the system is RHEL, on the /etc/sysconfig/network file, define a statement which enables the network configuration.

```
NETWORKING=yes
```

2. Restarting

   Run the following command and restart OS. After restarting OS, verify eth0 is enabled using the "ip(8)" or the "ifconfig(8)" command.

```
# /sbin/shutdown -r now
```

3. Setting a subnet mask

   For the underlined parameter, specify the network address and the subnet mask of the virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanetmask create -i 10.34.214.0 -m 255.255.255.0
```

   Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetmask print
```

Note

For details on the subnet mask value, see "hanetmask Command" in "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

4. 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
eth0,eth1
```

   Check that the virtual interface has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
```

5. 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
```

6. 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
```

7. Creating of the takeover IP address (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
```

If the STANDBY node is [HOST-primecl02]

1. Setting up the system
    1. Define the IP address and Host name in /etc/hosts file. Defined content is same as HOST-primecl01.
    2. Specify the IP address defined in Step 1-1 above to the /etc/sysconfig/network-scripts/ifcfg-ethX (X is either 0 or 1) file.
        - Contents of /etc/sysconfig/network-scripts/ifcfg-eth0

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
IPADDR=10.34.214.182
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

        - Contents of /etc/sysconfig/network-scripts/ifcfg-eth1

```
DEVICE=eth1
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
ONBOOT=yes
TYPE=Ethernet
```

    3. When the system is RHEL, on the /etc/sysconfig/network file, define a statement which enables the network configuration.

```
NETWORKING=yes
```

2. Restarting

    Run the following command and restart OS. After restarting OS, verify eth0 is enabled using the "ip(8)" or the "ifconfig(8)" command.

```
# /sbin/shutdown -r now
```

3. Setting a subnet mask

    For the underlined parameter, specify the network address and the subnet mask of the virtual interface.

```
# /opt/FJSVhanet/usr/sbin/hanetmask create -i 10.34.214.0 -m 255.255.255.0
```

    Check that the facility has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetmask print
```

Note

For details on the subnet mask value, see "hanetmask Command" in "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

4. 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
eth0,eth1
```

   Check that the virtual interface has been set up correctly.

```
# /opt/FJSVhanet/usr/sbin/hanetconfig print
```

5. Setting up the HUB monitoring function

   In 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
```

6. 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
```

7. Creating of the takeover IP address (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
```

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 "6.7.3.5 Setting Up Gls Resources" and "6.7 Setting Up Cluster Applications."

Then, start RMS and check the RMS tree to confirm whether the Gls resources are displayed correctly. For details, see "7.1.3.1 RMS Tree."

The Gls resource name is displayed as GlsX (X is integer).

See

For information on GLS (redundant line control function) and other operation modes, see "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 Virtual NIC mode.

See

For setup details, see "2.3 Setup with GLS" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

6.3 GDS Configuration Setup

The GDS setup operations are described below.

- "6.3.1 Setting Up System Disk Mirroring"

  Set up system disk mirroring in both nodes, and then restart both nodes.

- "6.3.2 Setting Up Shared Disks"

  Set up the shared disk volumes.

Note

- If you are using a shared disk unit, you must use GDS to manage that unit.
- Execute the configuration setting of GDS after initializing the cluster.
- To use EC or REC function of the ETERNUS Disk storage systems without using PRIMECLUSTER GDS Snapshot, do not add a GDS class that includes a copy destination disk of EC or REC to a cluster application.
  When EC or REC is either the synchronous processing in process or equivalency maintain status, a program running on the server may fail to access the destination disk with error. Therefore, if the class that includes the copy destination disk is added to a cluster application, the program running on the server may fail to access the destination disk. This may lead to a failover of the cluster application.

6.3.1 Setting Up System Disk Mirroring

Take the following setup procedures to enable system disk mirroring.

The setup values correspond to the values in "Setup (GDS System Disk Mirror)" of PRIMECLUSTER Designsheets. In the operation procedure, "Setup (GDS System Disk Mirror)" of PRIMECLUSTER Designsheets is abbreviated as "designsheet".

See

For setup details, see "System Disk Mirroring Settings [EFI]" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Note

To mirror the system disk of a guest OS by using GDS in KVM environment, you need to configure a mirror volume of a local class or a shared class, which is created on the host OS, for the guest OS. For information on how to set up the host OS, see the following:

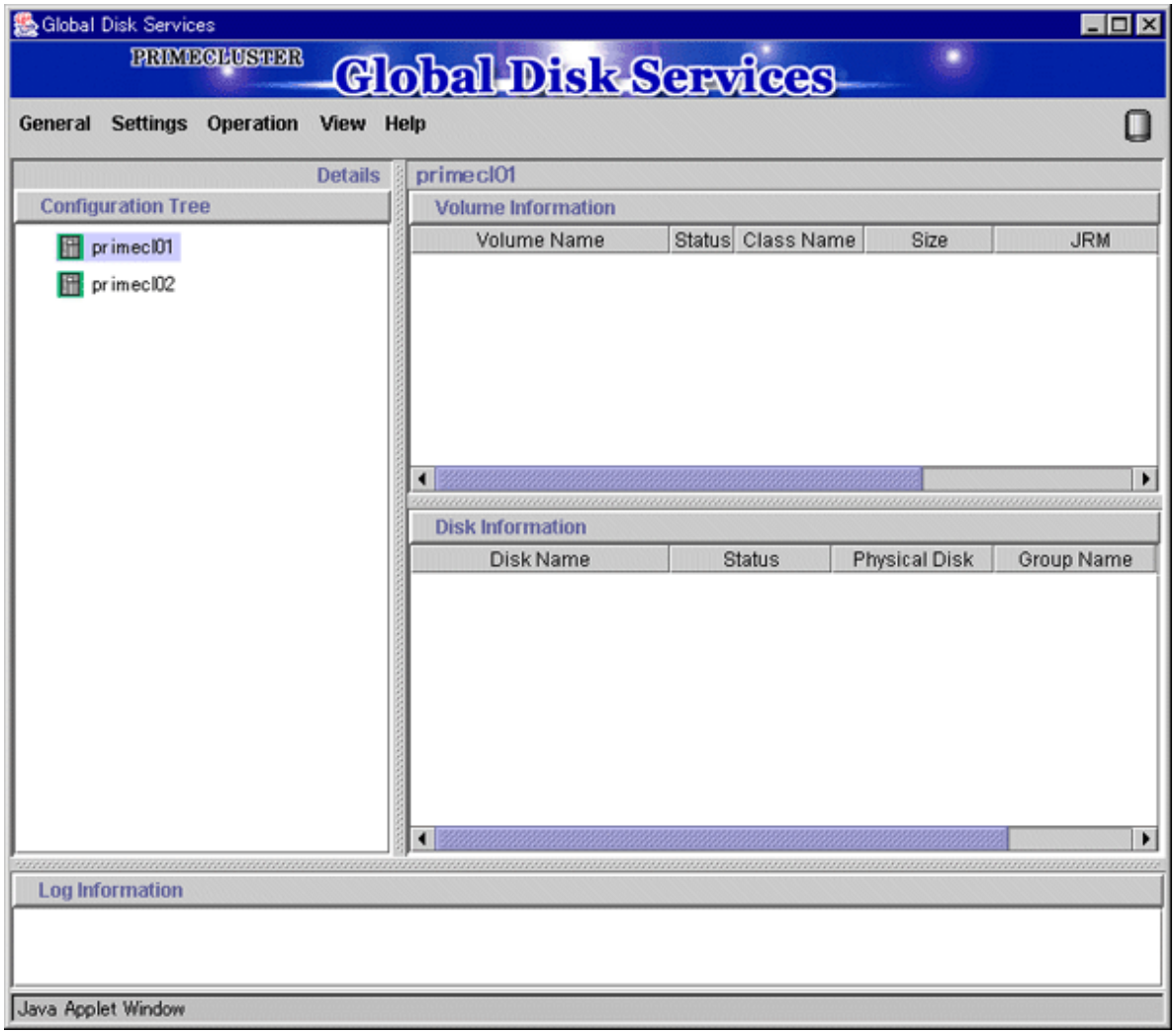
- When building a cluster system between guest OSes on one host OS, see "1. Setting up disks and related devices" in "3.2.1.1 Host OS setup (before installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes without using Host OS failover function, see "1. Setting up disks and related devices" in "3.2.2.1 Host OS setup (before installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes using Host OS failover function, see "1. Setting up disks and related devices" in "3.2.3.1.3 Host OS setup (before installing the operating system on guest OS)."

For details on settings, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Operation Procedure:

1. Select Global Disk Services at the Web-Based Admin View top screen.

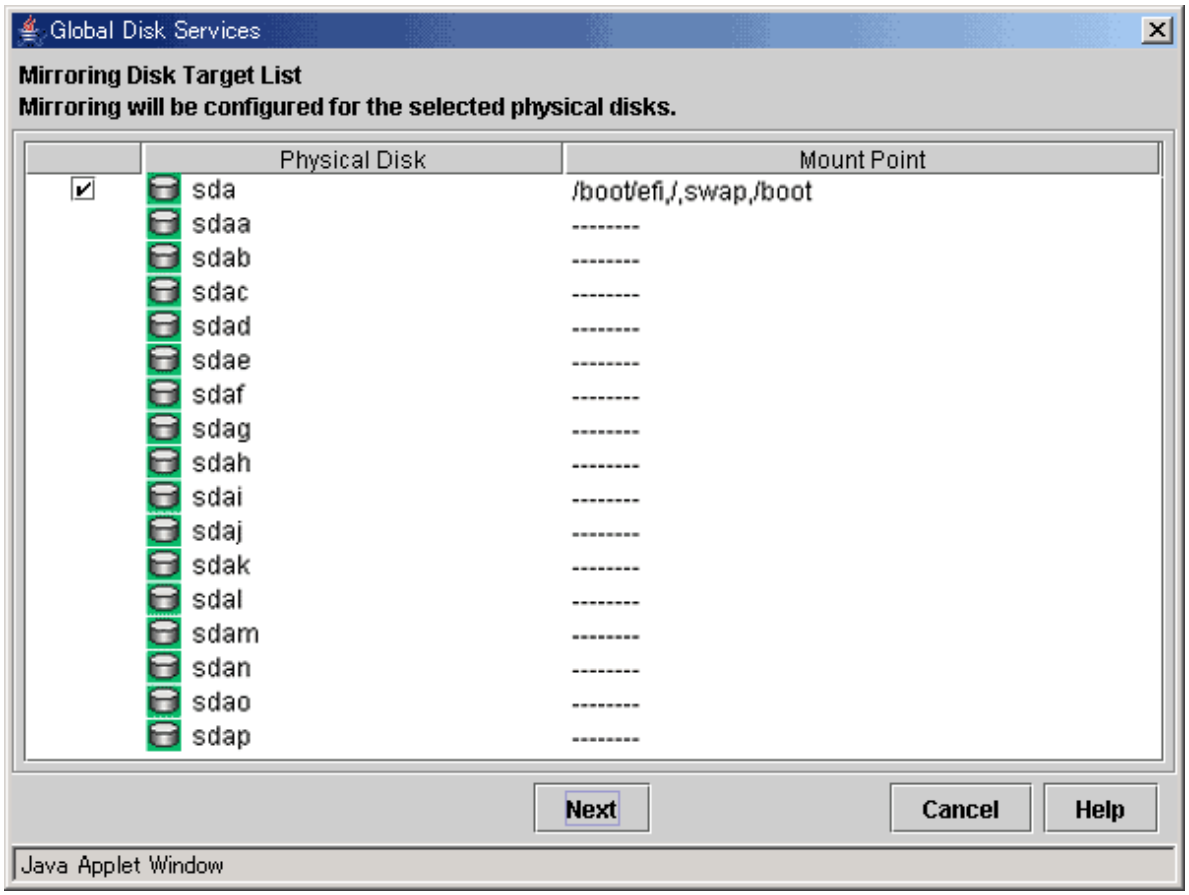
   The GDS Management screen (hereinafter main screen) is displayed.



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.

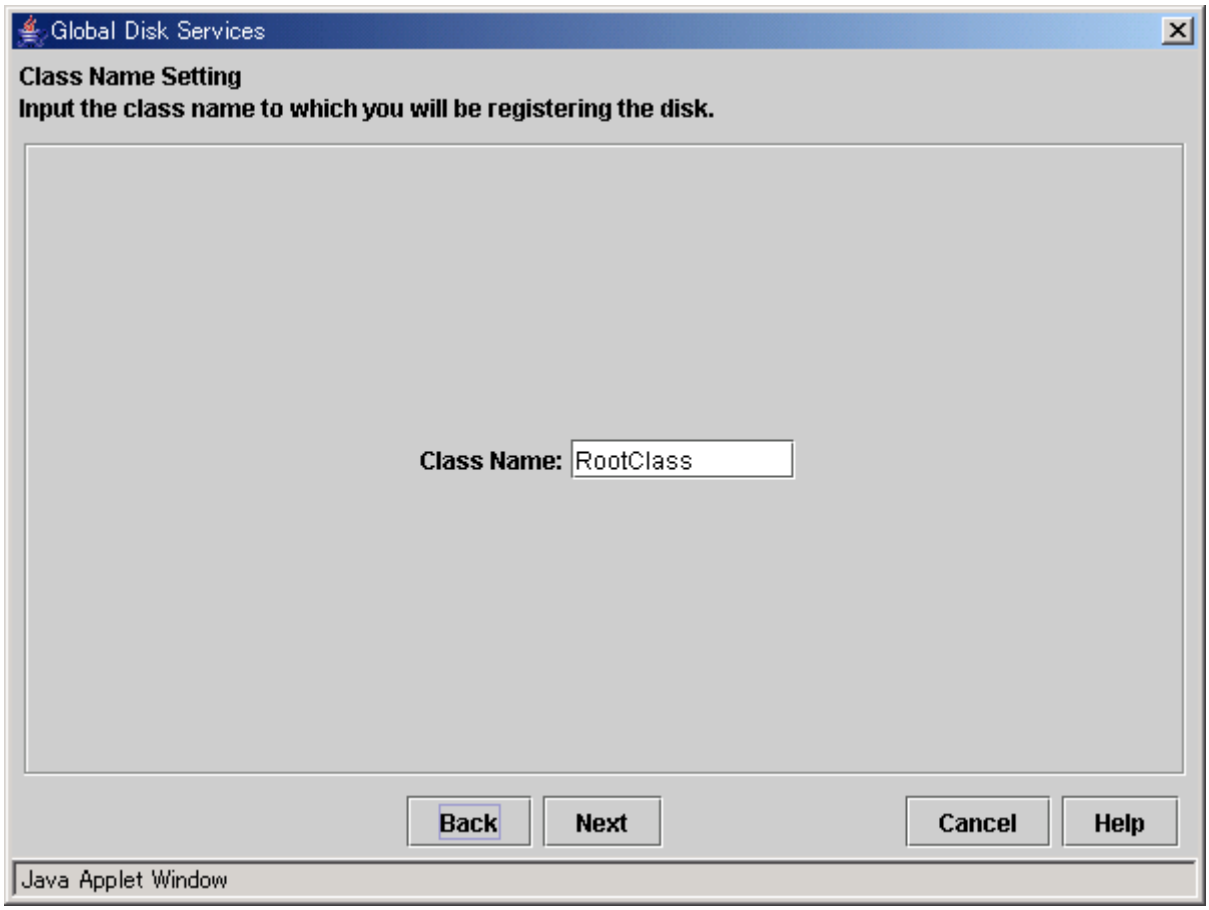
   A list of disks that can be used for mirrored disks for the selected node is displayed.

Select the system disk ("Physical disk name" on the designsheet), and click Next.



3. Specify class name of the root class.

   Enter the class name ("Class name" on the designsheet) of the root class, and click Next.



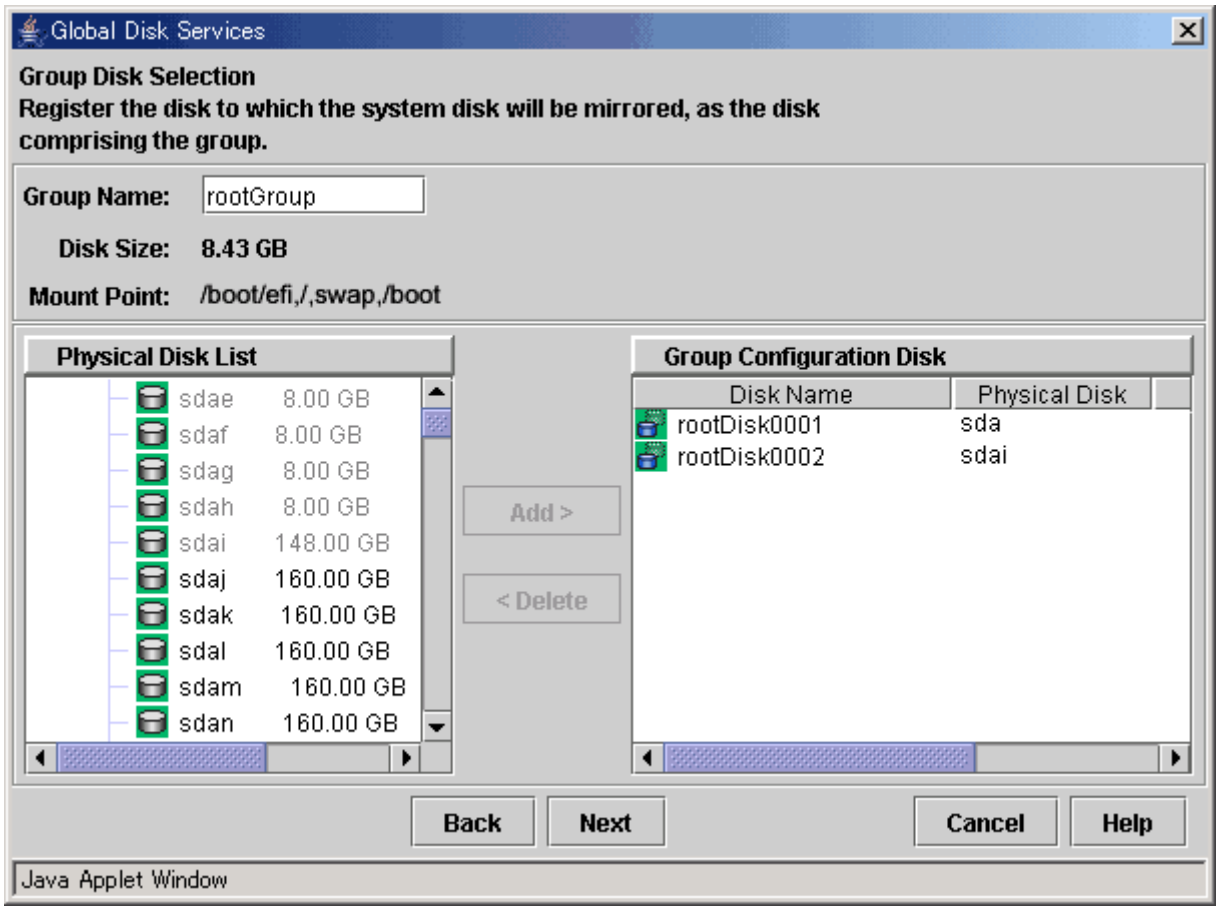
Note

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 designsheet), then from the "Physical Disk List," select the "mirror disk" ("Mirror disk name" on the designsheet) for the system disk, and click Add.

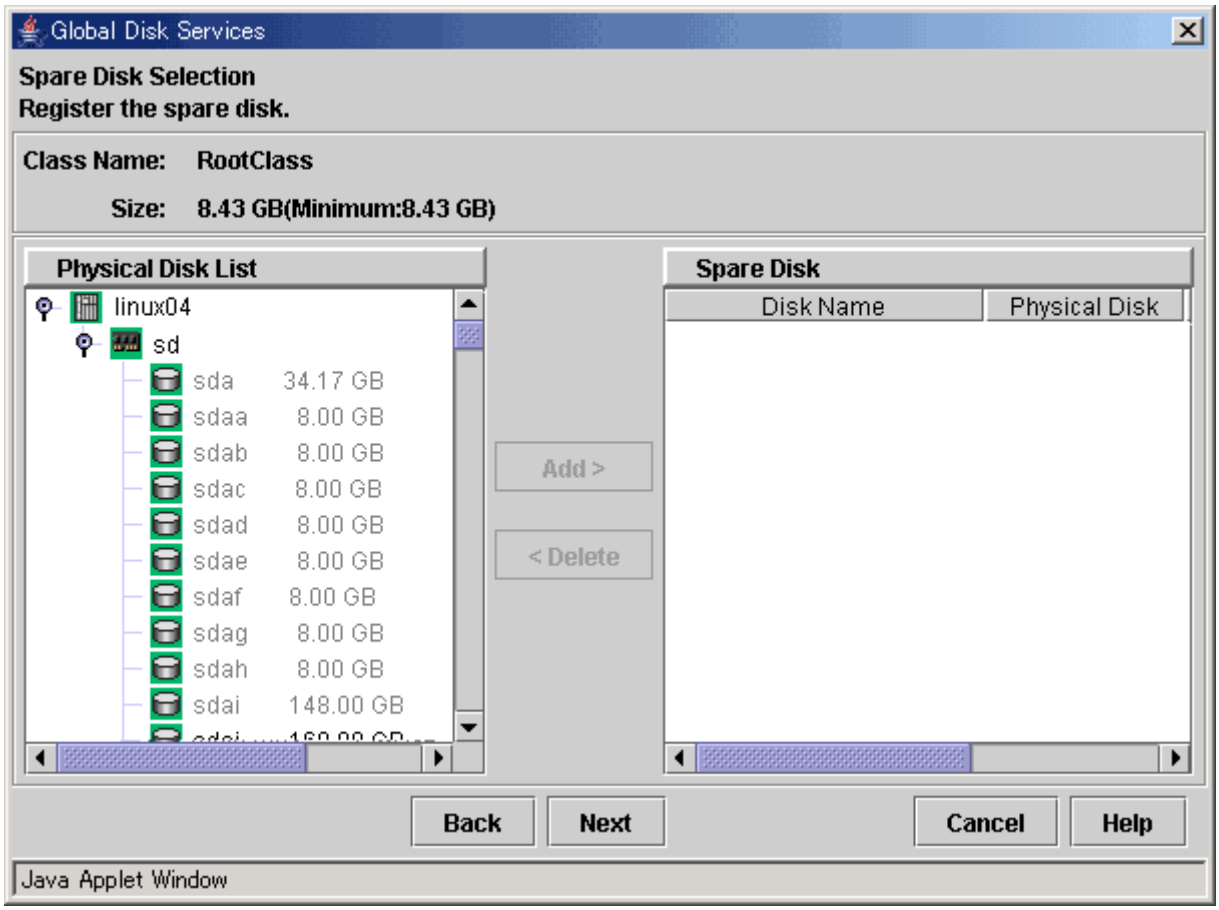
   Check that the "mirror disk" that was selected is displayed in "Group Configuration Disk," and then click Next.



5. Select a Spare Disk ("Spare disk name" on the designsheet) from the "Physical Disk List," and click Add.

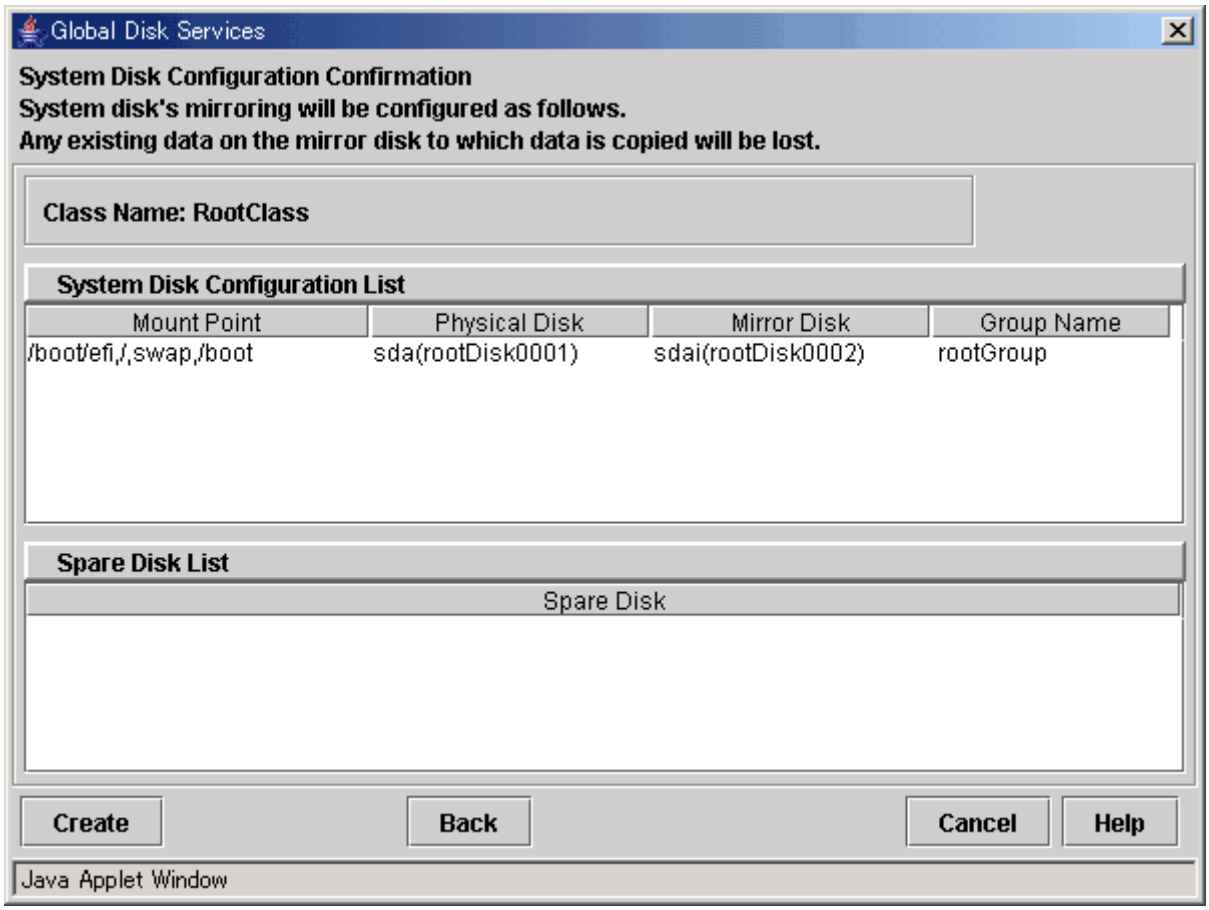
   Check that the spare disk that was selected is registered to "Spare Disk," and then click Next.

   If a spare disk is unnecessary, go to Step 6.



6. Check the system disk configuration.

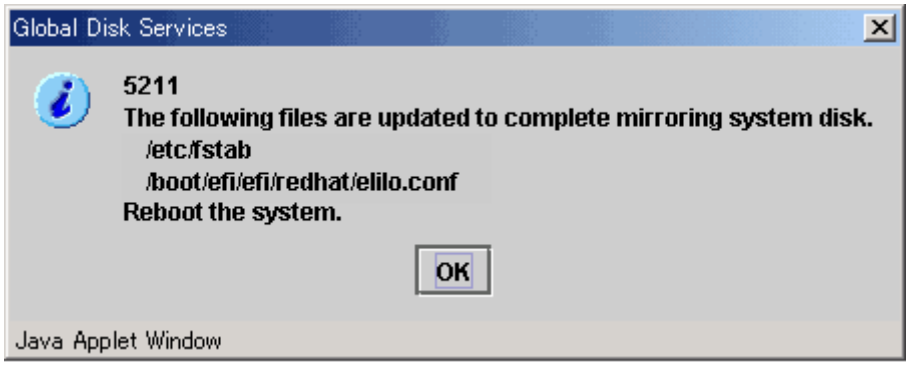
   Check the physical disk name and the mirror disk name, and then click Create.



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 Setting Up Shared Disks

Set up the shared disk as described below:

When mirroring is used among servers, the setting procedures are the same; select the netmirror under Type on the Group Attributes Definition screen.

- **When the GFS Shared File System is to be used**

  1. Execute "Volume setup."
  2. Execute "6.4 Initial GFS Setup."

- **When the ext3 file system is to be used**

  1. Execute "Volume setup."
  2. Execute "File system setup."
  3. Create a Gds resource and register it to a cluster application.
     For details, see "6.7.3.4 Setting Up Gds Resources" and "6.7 Setting Up Cluster Applications."

- **When the file system is not to be used**

  1. Execute "Volume setup."
  2. For use as a RAW device, make the settings according to the manual pages related to the "raw(8)" command and rawdevices.
  3. Create a Gds resource and register it to a cluster application.
     For details, see "6.7.3.4 Setting Up Gds Resources" and "6.7 Setting Up Cluster Applications."

- "When the GFS Shared File System is to be used" and "When the file system is not to be used," "File system setup" is not necessary.
- The setup procedures for "When the ext3 file system is to be used" and "When the file system is not to be used" must be carried out before the Gds resources are set up. For details, see "6.7.3.3 Preliminary Setup for Gds Resources."
- "When the GFS Shared File System is to be used," "6.7.3.4 Setting Up Gds Resources" must not be carried out.
- The local class disks or shared class disks used by GDS on the guest OS should be configured as the following virtual disks if they are used in the virtual machine environment.
  - KVM environment

    virtio-SCSI devices or virtio block devices

Volume setup

There are five types of volumes:

a. Single volume
b. Mirror volume
c. Stripe volume
d. Volume created in a concatenation group
e. Netmirror volume

This section separately describes the volume setup procedures for a single volume (a) and for other volumes (b, c, d, e). For details, see "Class, Group, Volume Settings" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

The values to be set for the individual items correspond to the values in "Setup (GDS Local Class)" and "Setup (GDS Shared Class)" of PRIMECLUSTER Designsheets.

- 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.
- When neither the system nor the GDS Management screen are reactivated after "2. Registering a shared disk" of "5.1.3.2 Registering Hardware Devices," the registered shared disk might not be correctly recognized to GDS. In this case, setup the volume after updating physical disk information. Physical disk information can be updated by selecting *Update Physical Disk Information* from *Operation* menu of the main screen.

Single volume setup

If you are not using a single volume, this setup is unnecessary.

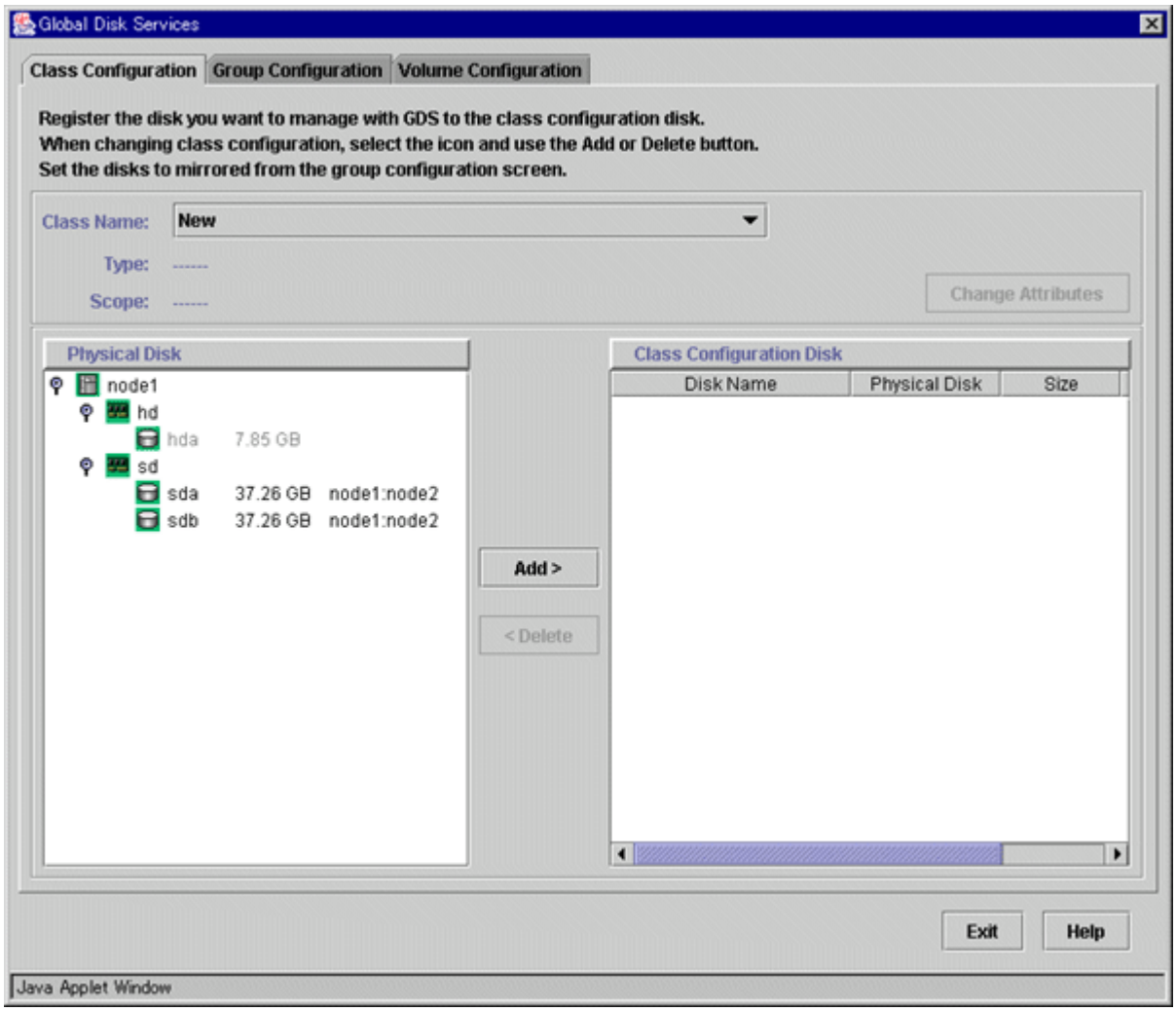
Operation Procedure:

1. Start the main screen.

   Choose *Global Disk Services* on the Web-Based Admin screen.

2. Disk class creation and physical disk registration

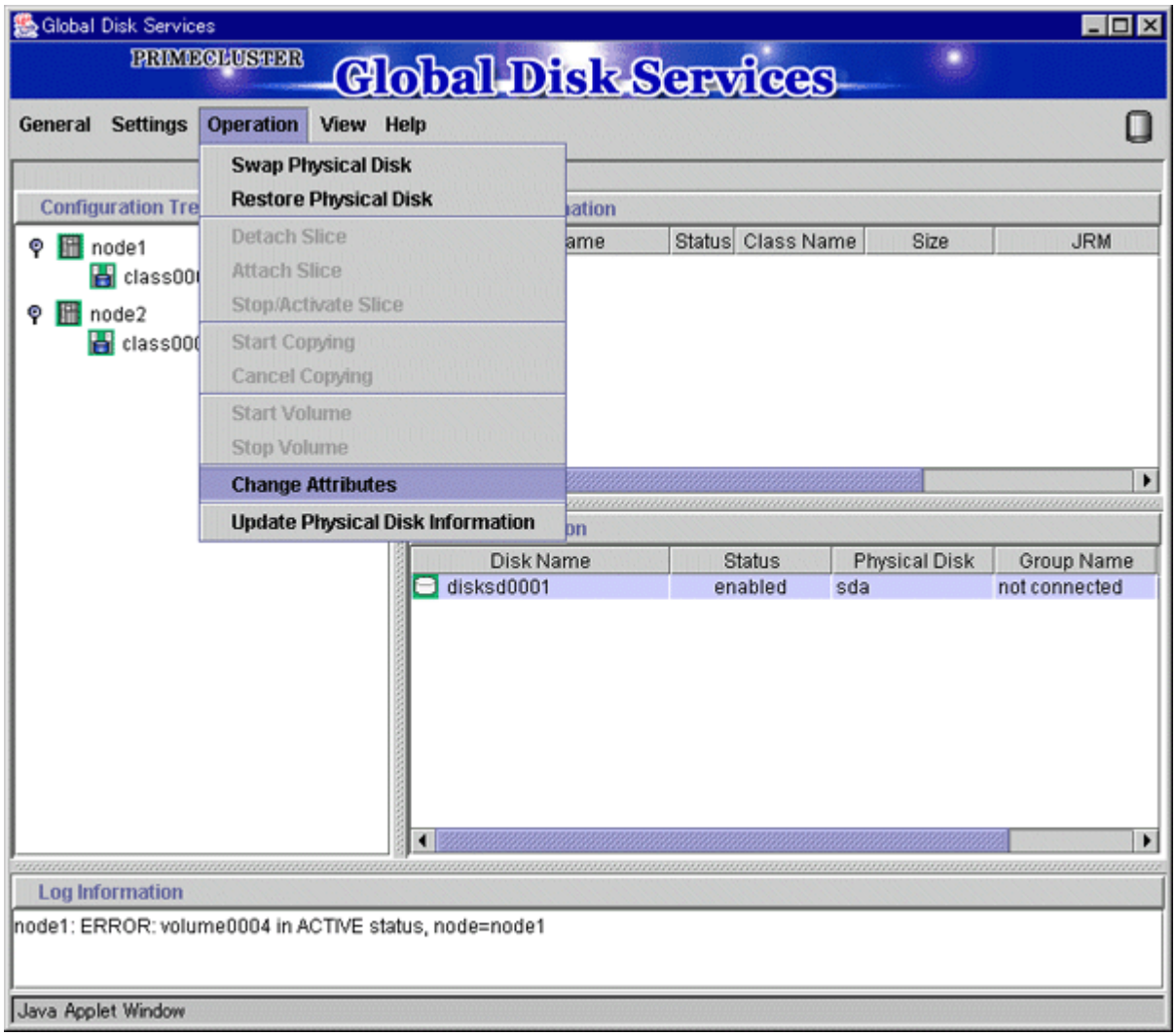
   At the main screen, select *Class Configuration* from the *Settings* menu.



   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 Attributes Definition screen opens. Enter the *Class Name* but do not change the *Type* value (leave the value as "shared").

3. Disk type attribute setup

   At the main screen, select the disk that was registered in Step 2 from the Disk Information field, and select *Operation -> Change Attributes* from the menu bar.



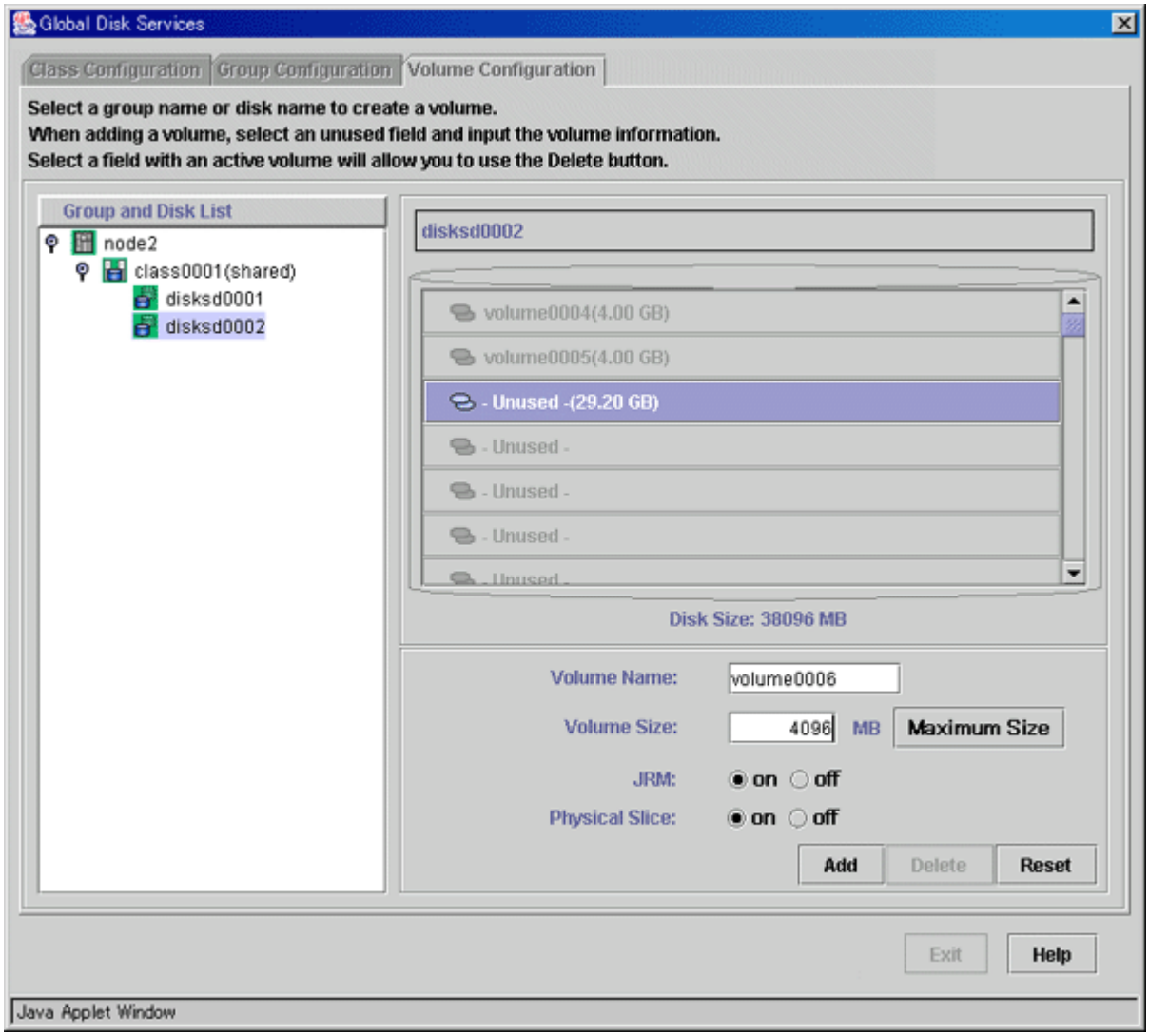
Set *Disk Type* to "single," and then click *OK*.

4. Volume creation

   Select *Settings* -> *Volume Configuration*, and then select the disk that was registered in Step 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 settings, and then click *Exit*.



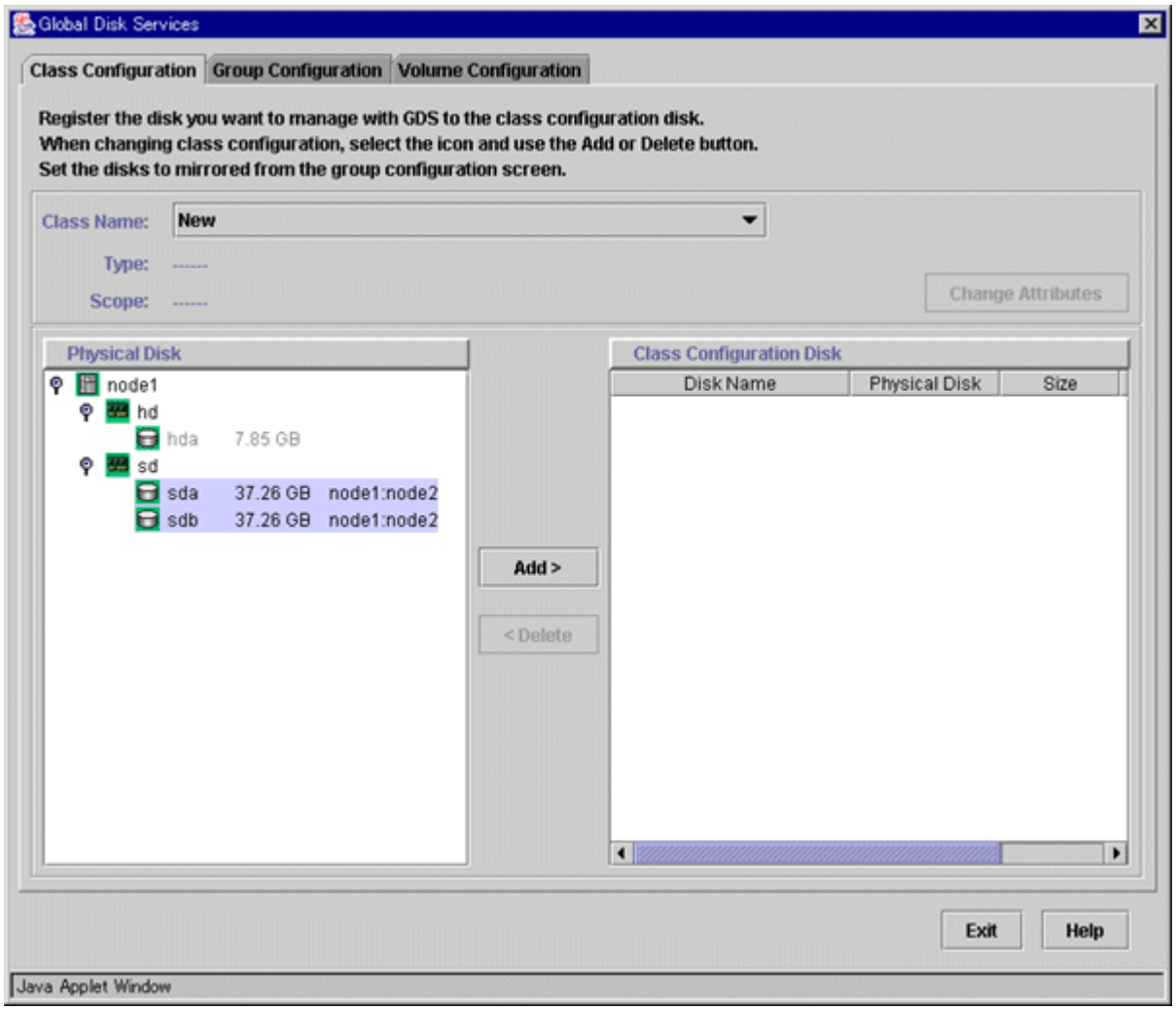
Setup for other volumes

If you are using only "single" volumes, this setup is unnecessary.

Operation Procedure:

1. Creating a disk class and registering physical disks

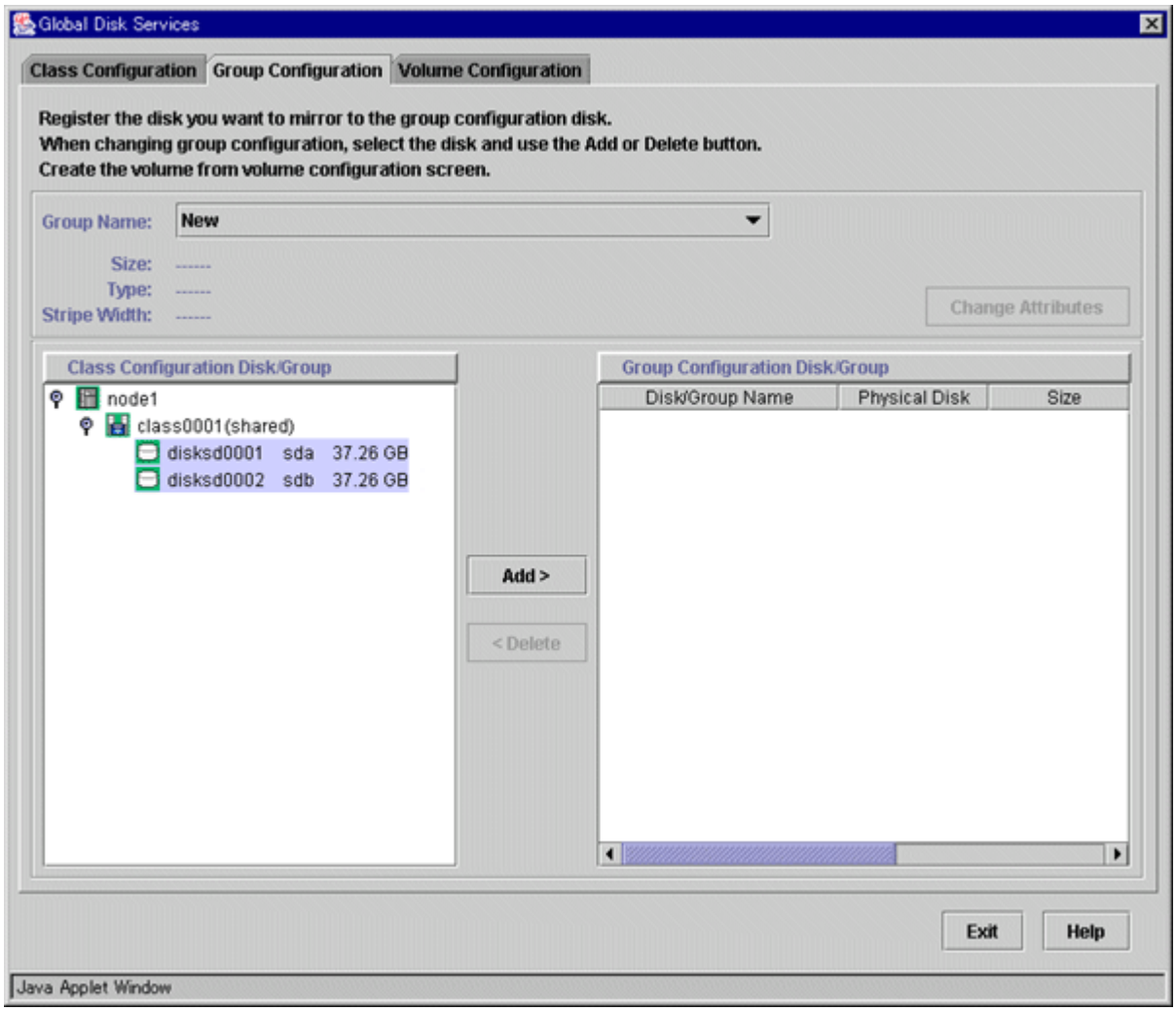
   At the main screen, select "Class Configuration″from "Settings″menu.



At the above screen, select the physical disks to be registered from "Physical Disk″list, and then click "Add″. When "Add" is clicked, the Class Attributes Definition screen opens. Enter "Class Name″but do not change "Type″value (leave the value as "shared"). Then click "Exit″.

2. Setting up the disk group configuration

   Click the Group Configuration tab.

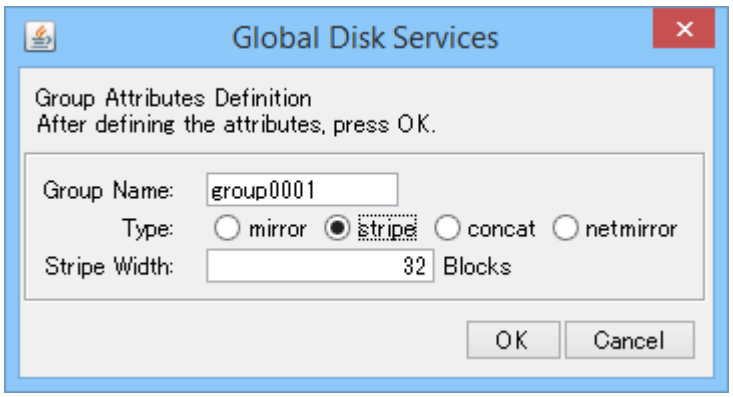


At the above screen, select the disks to be added to the group from "Class Configuration Disk/Group ″list, and then click "Add".

Enter "Group Name", "Type", and "Stripe width" in the Group Attributes Definition screen, and then click "OK ″.

For the mirroring among servers, select "netmirror" for "Type".

Enter "Stripe width" only when selecting "stripe" for "Type".

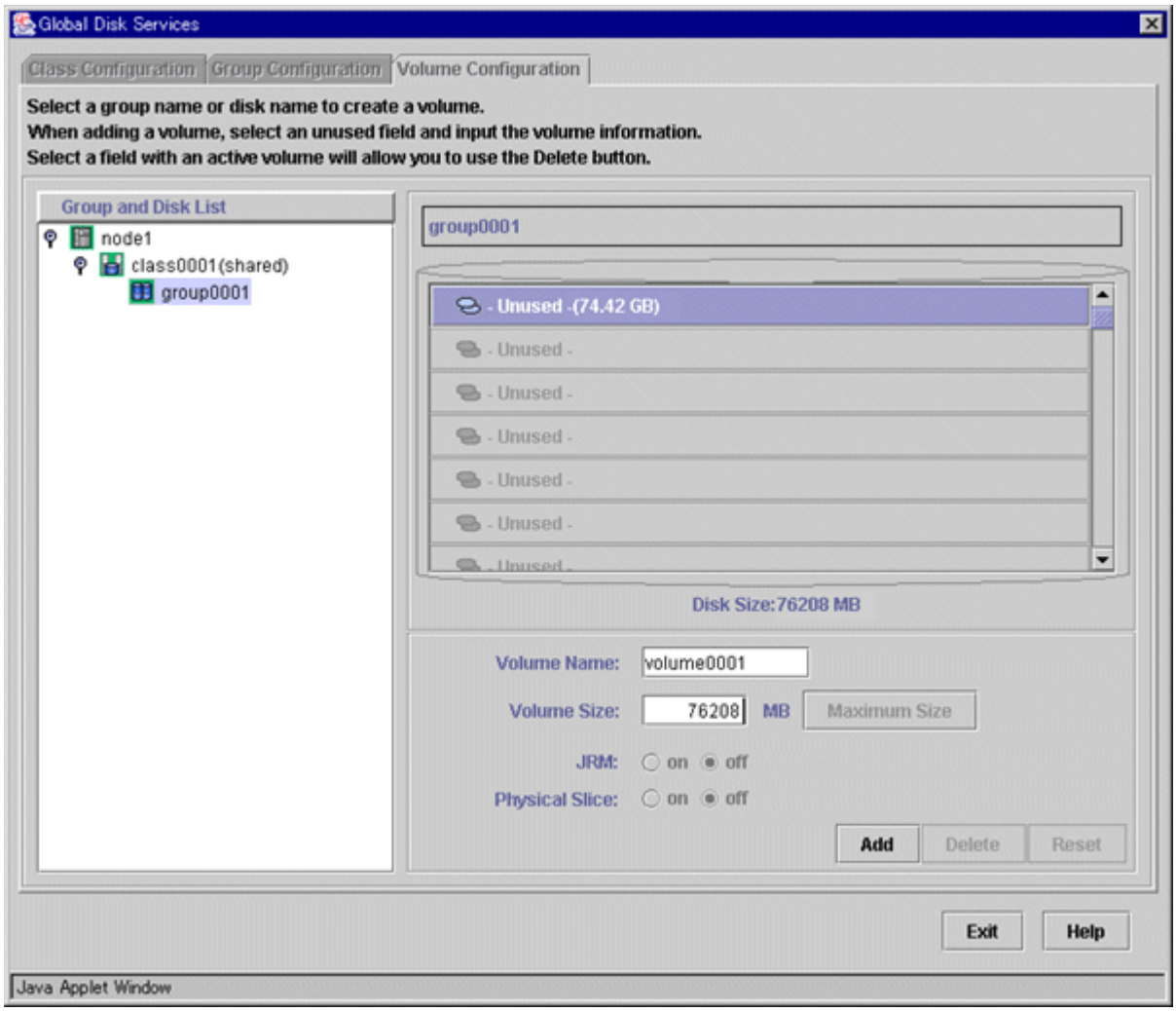


3. Creating a volume

   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.

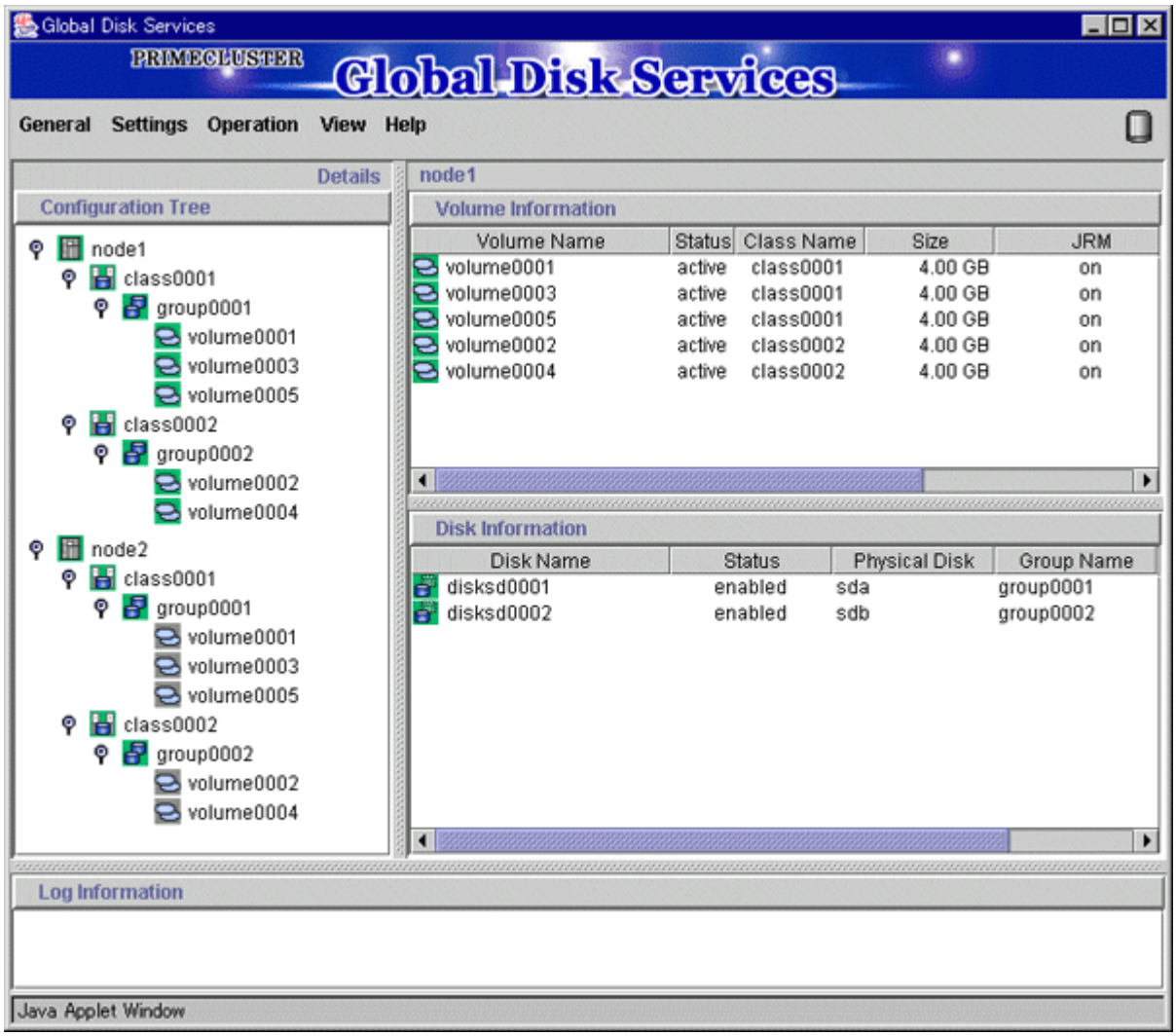
   Click *Add* to enable the settings.

   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 each created volume.

Example: class name = Class1, volume name = Volume1, and file system type = ext3

```
# mkfs -t ext3 /dev/sfdsk/Class1/dsk/Volume1
```

See

For how to create file system, see the file system manual.

6.4 Initial GFS Setup

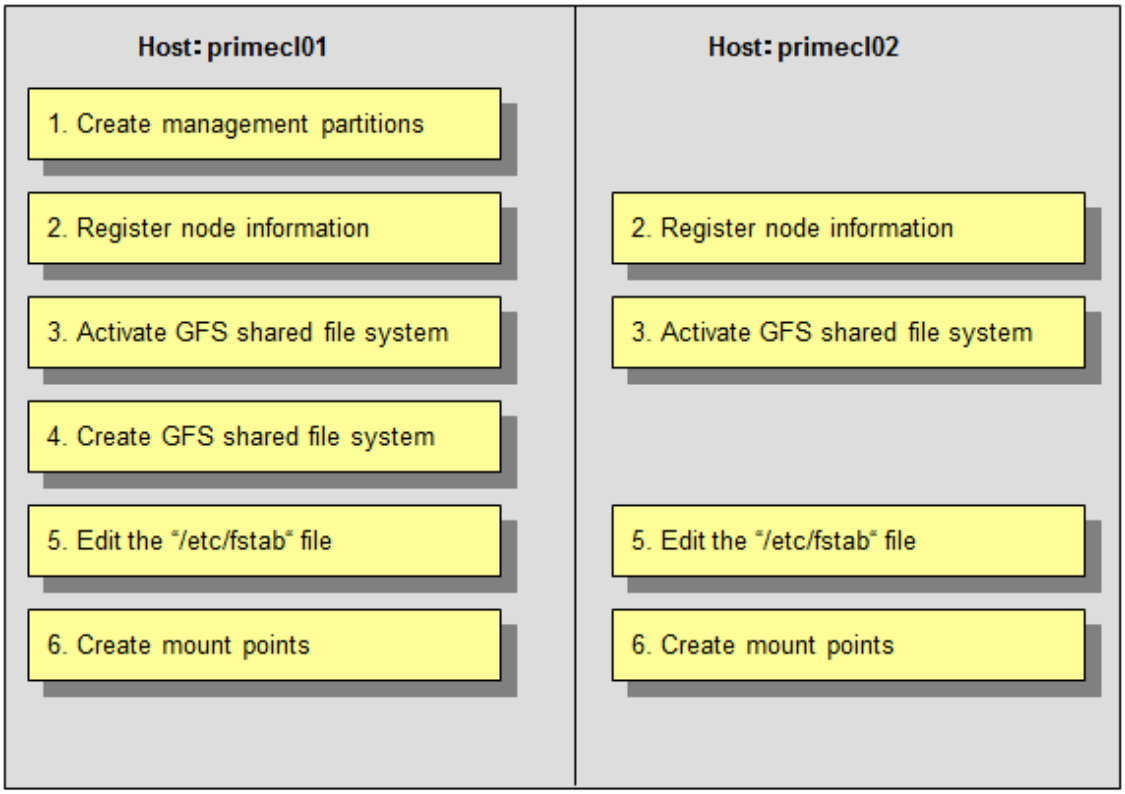
For the GFS Shared File System to be created on GDS volume, there must be a GDS volume and that volume must be active. If the volume is stopped, start the volume.

See

The volume is started by the [Start Volume] of [Operation] menu of GDS management view or the "sdxvolume -N" command.

For details, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

To use the GFS shared file system in RMS cluster operation, you need to set up GFS according to the flow below:



GFS: Global File Services

The device name and mount points that are specified here correspond to the values in "Setup (GFS Shared File System)" and "Setup (GFS Shared File System 2)"of PRIMECLUSTER Designsheets.

Note

- You need to prepare a management partition that is exclusive to the GFS shared file system. The GDS volume disk class is used for a switching file system and non-switching file system. For the management partition, non-switching file system must be allocated.
- If you are using a GFS shared file system, you must not carry out "6.7.3.4 Setting Up Gds Resources."

Operation Procedure:

1. Create a management partition for the GFS shared file system on any one of the nodes.

```
# sfcsetup -c /dev/sfdsk/class0001/dsk/GFSctl
```

2. Register the node information in the management partition on each node.

```
primecl01# sfcsetup -a /dev/sfdsk/class0001/dsk/GFSctl
primecl02# sfcsetup -a /dev/sfdsk/class0001/dsk/GFSctl
```

3. Activate the GFS shared file system on each node.

```
primecl01# sfcfrmstart
primecl02# sfcfrmstart
```

Note

If sfcfrmstart ends abnormally, confirm that sfcprmd is started with the "ps" command. If sfcprmd has not been started, execute the following command on the node on which sfcprmd is not started:

- For RHEL6

```
# service sfcfsrm start
```

- For RHEL7

```
# systemctl stop fjsvgfsfsrm.service
# systemctl start fjsvgfsfsrm.service
```

4. Create a GFS shared file system on any one of the nodes.

```
# sfcmkfs -o node=primecl01,primecl02 /dev/sfdsk/class0002/dsk/volume0001
```

5. Add the mount information of the GFS shared file system to /etc/fstab on each node. Specify "noauto" in the "mount options" filed of the mount information. Do not specify "noatrc" in the same field.

```
/dev/sfdsk/class0002/dsk/volume0001 /sfcfs1 sfcfs rw,noauto 0 0
```

6. Create a mount point on each node.

```
primecl01# mkdir /sfcfs1
```

```
primecl02# mkdir /sfcfs1
```

See

The operations described in procedures 4, 5, and 6 can be set up by using the GUI management view. For details, see "6.4.1 File System Creation."

6.4.1 File System Creation

This section explains how to create a file system.

Operation Procedure:

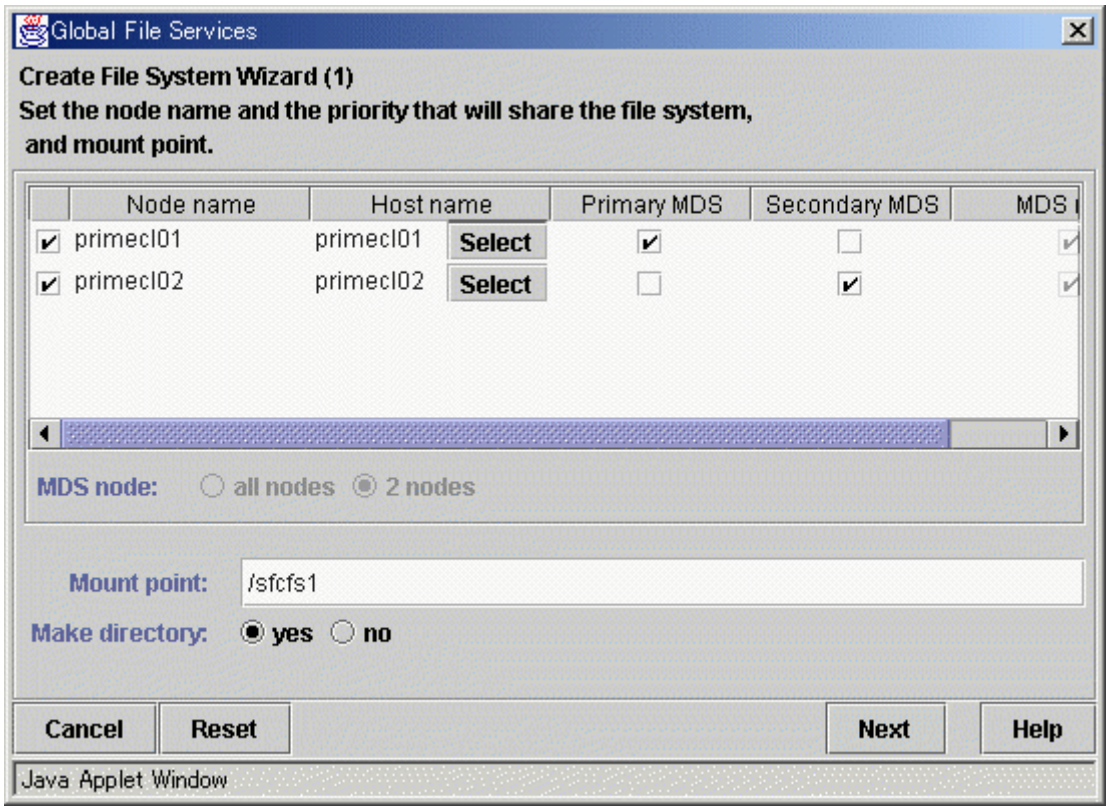
1. Start the GFS management view.

   Choose Global File Services on the Web-Based Admin screen, select a node from the node list, and then display the main screen of Global File Services.

2. Set up shared information.

   Select *Operation* -> *Create* and then open the "Create File System Wizard (1)."

   Set up the node information that is to be shared within the file system and then mount the information using the "Create File System Wizard (1)."



Selecting "Node name"

Select the node names to be shared with "Node Names." You must select two nodes.
Note that the selection of the local node (displayed node) cannot be canceled.

Selecting a "Host name"

To select a host name other than that which is currently displayed, click the *Select* button and specify the host name of the LAN to be used on each node. Note that two or more host names cannot be specified.

Setting the "Primary MDS" and "Secondary MDS"

Specify the nodes that boot the management server of the shared file system in "Primary MDS" and "Secondary MDS."

Setting the "Mount point" and "Make directory"

Specify the full path for the "Mount point." If you select "yes" from "Make directory," creates a directory with the following attributes:

- Owner: root
- Group: sys
- Access authority: 775

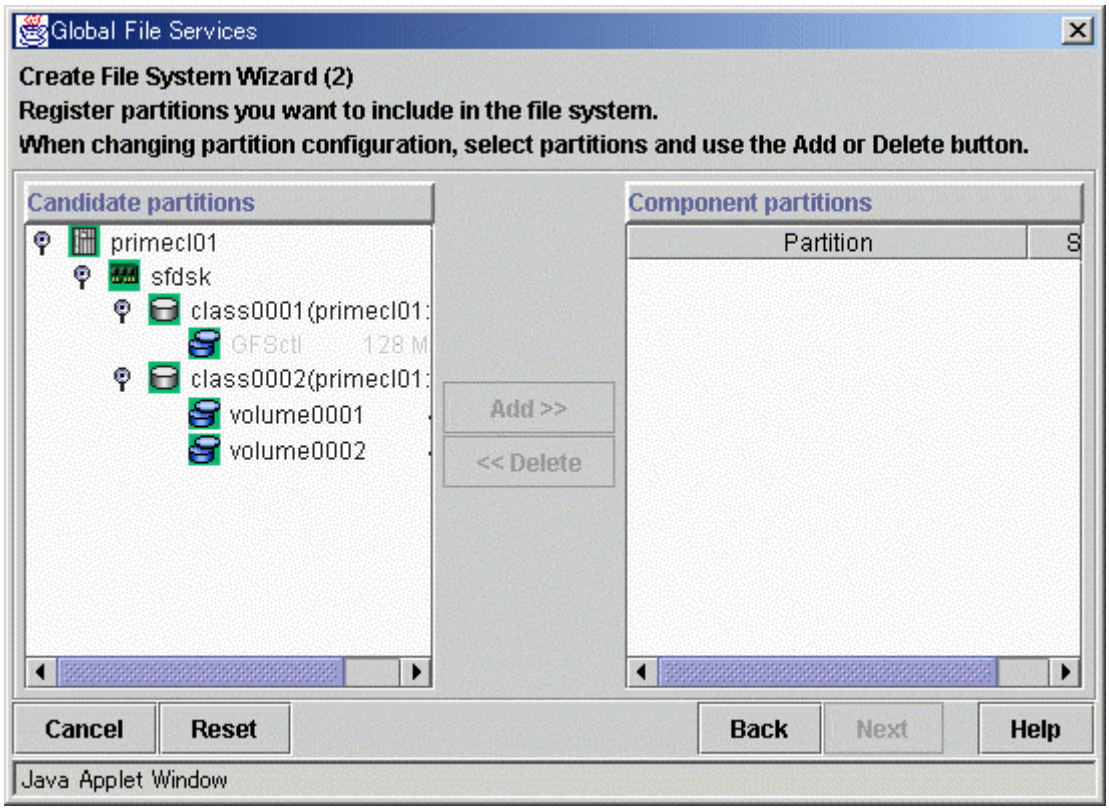
After setting or changing this information, click the *Next* button to open the "Create File System Wizard (2)."

To return each setup item to its default value, click the *Reset* button.

To stop the processing of the file system creation, click the *Cancel* button.

3. Select the configuration partition.

   Using the "Create File System Wizard (2)," register the partition that is to be used as the file system.



Select the partition to be used from the [*Candidate partitions*] list and then click the *Add* button.

Only one partition can be selected at a time. A partition that is already being used as a file system or as a management partition cannot be selected.

After the partition has been selected, click the *Next* button to open the "Create File System Wizard (3)."

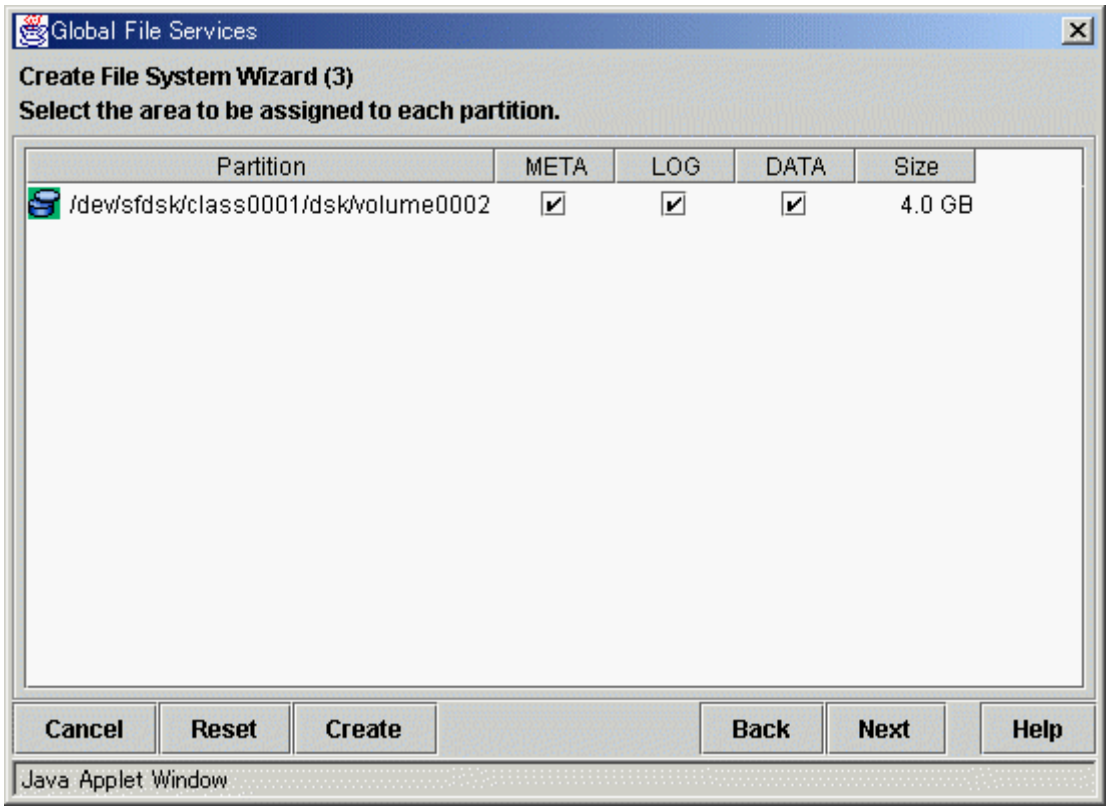
To return to the "Create File System Wizard (1)," click the *Back* button.

To abandon file system creation, click the *Cancel* button.

4. Set up the partition information.

   Using the "Create File System Wizard (3)," select the meta, log, and data areas that are to be allocated to each of the partitions selected with the "Create File System Wizard (2)."

   The partition to which the meta data area is allocated is used as the representative partition.



After setting the above information, click the *Next* button to open the "Create File System Wizard (4)."

No information can be set with the "Create File System Wizard (4)." Go to the "Create File System Wizard (5)."

To return each setup item to its default value, click the *Reset* button.

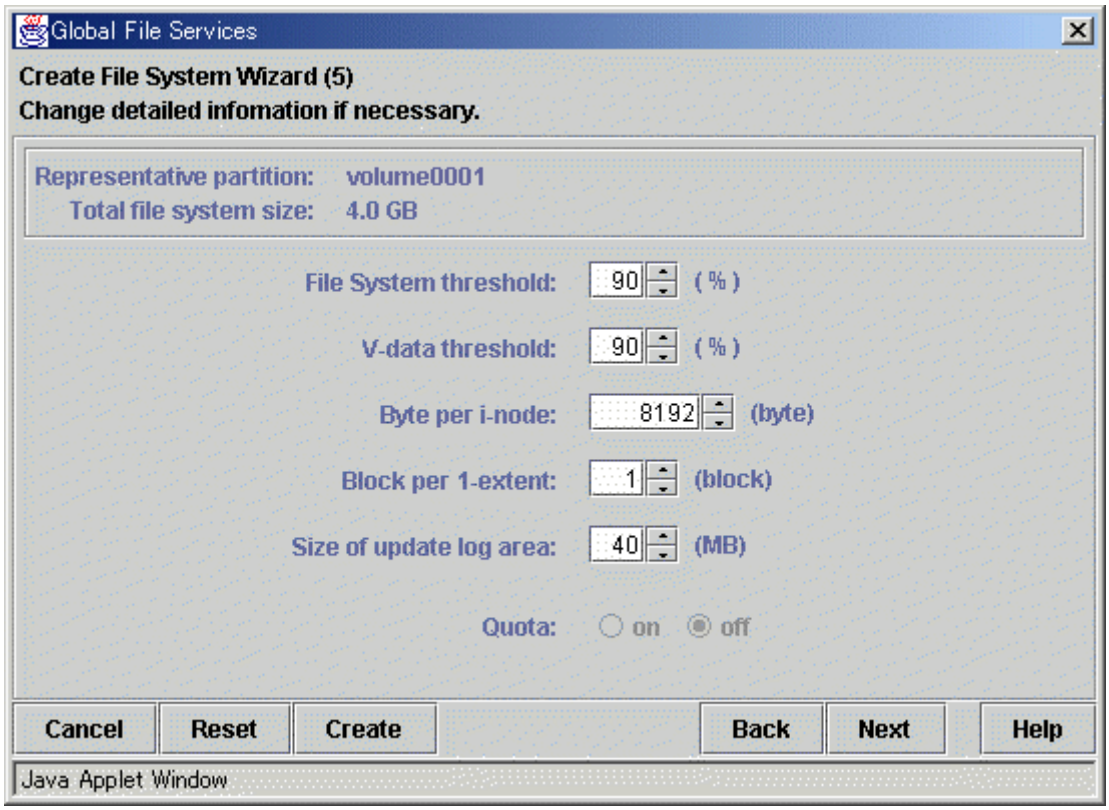
To return to "Create File System Wizard (2)," click the *Back* button.

To abandon file system creation, click the *Cancel* button.

To create the file system while leaving the default settings of the extended, detailed, and mount information as is, click the *Create* button.

5. Set up the detailed information.

   Set up the "Detailed information" by using the "Create File System Wizard (5)."



After setting the above information, click the *Next* button to open the "Create File System Wizard (6)."

To return each setup item to its default value, click the *Reset* button.

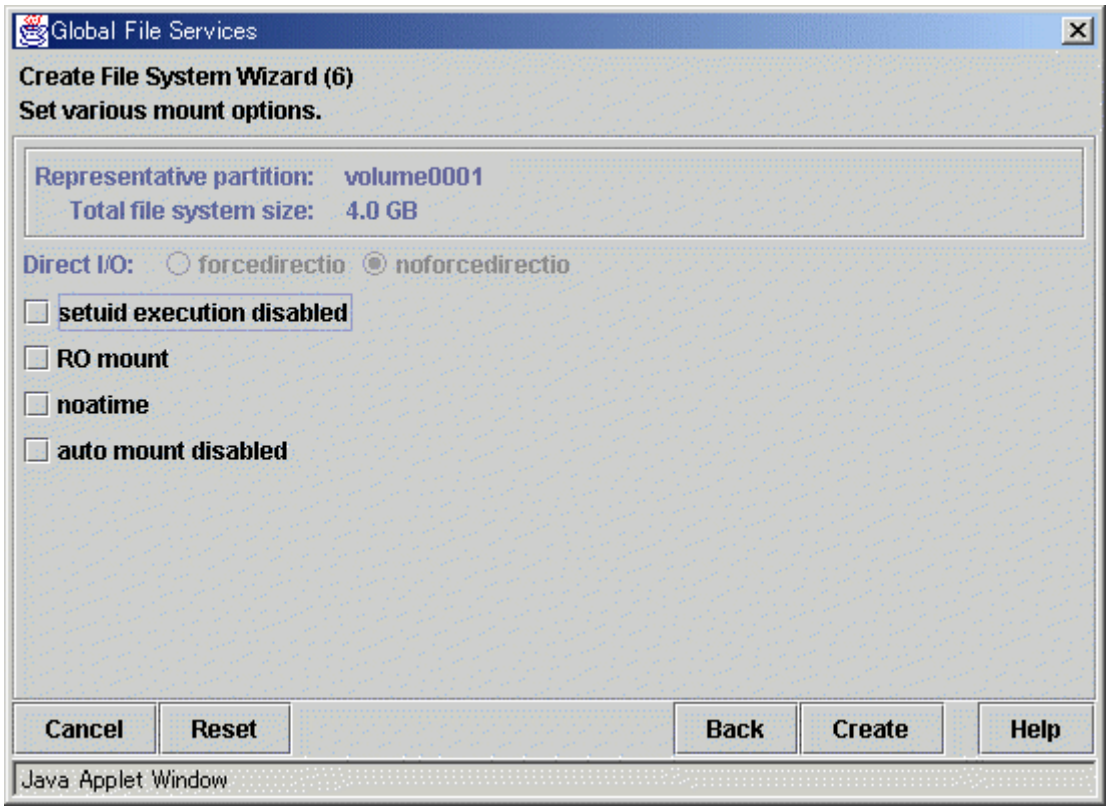
To return to the "Create File System Wizard (4)," click the *Back* button.

To abandon file system creation, click the *Cancel* button.

To create the file system while leaving the default setting of the mount information as is, click the *Create* button.

6. Set up the mount information.

   Set up the "Mount information" by using the "Create File System Wizard (6)."



After setting the above information, click the *Create* button to create the file system. To return each setup item to its default value, click the *Reset* button.

To return to the "Create File System Wizard (5)," click the *Back* button.

To abandon file system creation, click the *Cancel* button.

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 "6.7.3.1 Setting Up 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 ("/etc/profile" or "etc/bashrc", for example) 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.

Start script/Stop script

```
#!/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
      # Start your application
elif [[ $1 = "-u" ]]; then
      # Stop your application
else
      # Default operation
      exit 1 # Error
fi
exit 0
```

The above script sample covers both the Start script and the Stop script.
An example of Check script is shown below:

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 "6.7.3.1 Setting Up 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.

   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.

   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 "6.7.3.1 Setting Up 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 Fault state.

3: Indicates the Unknown state.

4: Indicates the Standby state.

5: Indicates the Onlinewarning state.

6: Indicates the Offlinefaulted state.

* The exit codes 3, 5, 6 indicate the special status. Use these codes only when the instructions from PRIMECLUSTER products are received. Do not use any exit codes other than the described above.

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 "Change the attribute" of "6.7.3.1 Setting Up 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 "10.3 Changing a Cluster Application."

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 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

RMS also has other environment variables.

See

- For details 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."

6.7 Setting Up Cluster Applications

This section explains how to set up a cluster application.

You can use any one of the nodes of the cluster system for the settings described in this section.

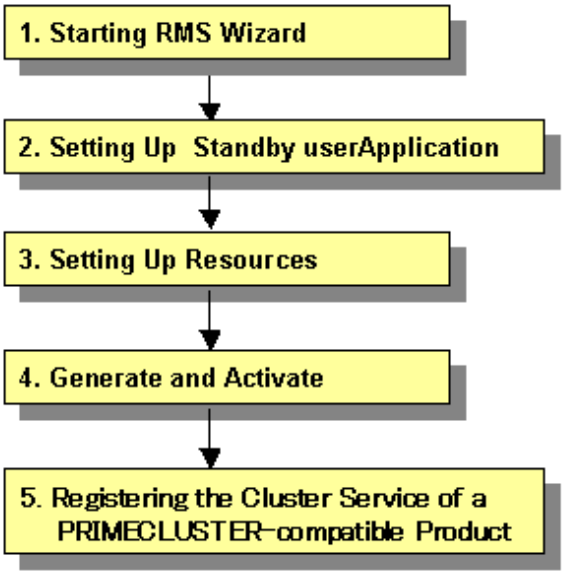
Cluster application setup flow

The setup flow for a cluster application is explained for each topology, below.

For information on these topologies, see "2.3 Determining the Cluster System Operation Mode."

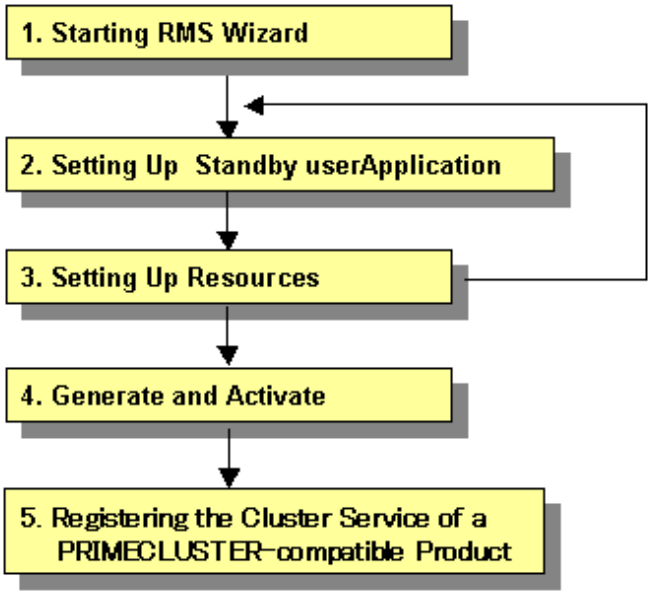
1) 1:1 standby for standby operation

Set up 1:1 standby for standby operation as follows.



2) Mutual standby for standby operation

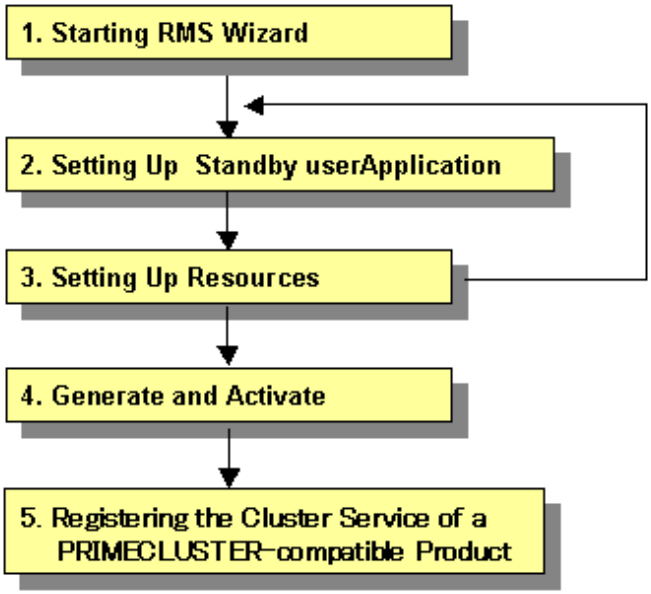
Set up mutual standby for standby operation as follows.



- To create two cluster applications, repeat steps 2. to 3.

3) N:1 standby for standby operation

Set up N:1 standby for standby operation as follows.

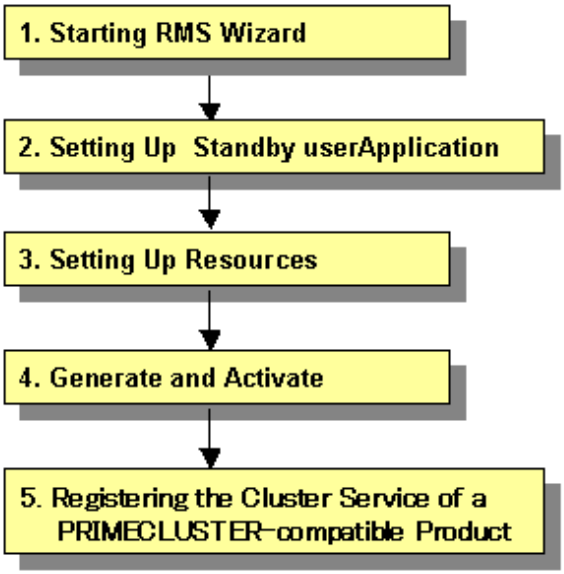


- To create two or more cluster applications, repeat steps 2. and 3.

  Example) For 2:1 standby, repeat steps 2. and 3. two times, to create two cluster applications.

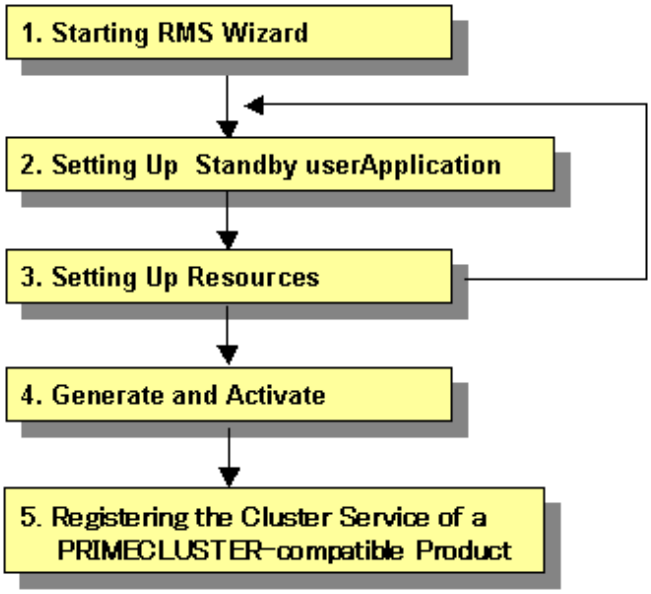
4) Cascaded standby operation

Set up cascaded standby operation as follows.



5) Priority transfer of standby operation

Set up the priority transfer of standby operation as follows.



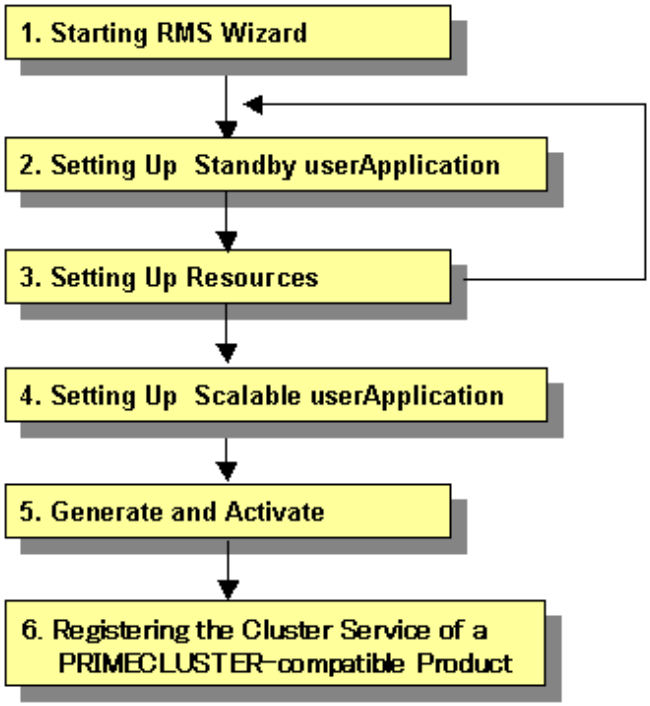
- To create two or more cascade cluster applications, repeat steps 2. and 3.
- Establish an exclusive relationship between the cluster applications.

  For details on exclusive relationships, see "6.7.7 Exclusive Relationships Between Cluster Applications."

  Example) For priority transfer with two cascade cluster applications, repeat steps 2. and 3. two times to create two cascade cluster applications.

6) Scalable operation

Set up a scalable operation as follows.



- Before you create cluster applications as part of scalable operation, create cluster applications in standby operation that act as the constituent factors of the cluster applications in scalable operation. To create cluster applications in standby operation, repeat steps 2. to 3.

  Example 1) For scalable operation with three nodes, repeat steps 2. and 3. three times to create three cluster applications of standby operation.

  Example 2) For high-availability scalable 1:1 standby (standby operation), repeat steps 2. and 3. once to create 1:1 standby cluster applications.

See

- 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 of a cluster application depending on the operation, and notes on its setting, see "6.10 Setting Contents and Notes on Cluster Application."

Note

- Set up the cluster application and resources based on the cluster application and resource information in "Setup (cluster application)"of PRIMECLUSTER Designsheets that was created in the design stage, and the sheet corresponding to each resource. If you need to change the cluster application after it is created, the designsheets are helpful. Make sure to create the designsheets before performing necessary operation.
- You cannot share one resource with multiple userApplication.
- Generate and Activate process fail if RMS is running. Using Cluster Admin or hvdisp, you need to confirm that RMS has not started before creating a cluster application. If RMS has already started, stop RMS from Cluster Admin or execute the "hvshut" command to stop RMS on all the nodes of the cluster system. For details on "hvdisp" and "hvshut", see the pages of these commands respectively.

- Set up "remote file copy" and "remote command execution" for the RMS Wizard. See the notes on "5.1.1 Setting Up CF and CIP." If the cluster interconnect is not protected by security, cancel the "remote file copy" and "remote command execution" settings on all the cluster nodes after setting up the cluster applications.

6.7.1 Starting RMS Wizard

Execute the "hvw" command.

The following is an example of starting the RMS Wizard with the configuration file name (testconf):

```
# /opt/SMAW/SMAWRrms/bin/hvw -n testconf
```

Note

While executing the hvw command, do not exit the hvw command forcibly, for example, by exiting a terminal. Doing so may corrupt the configuration information file and require the cluster application to be created again.

6.7.2 Setting Up userApplication

This section explains how to configure a cluster application.

There are two types of cluster applications, namely, standby operation and scalable operation.

Note that the term "userApplication" has the same meaning as "cluster application."

Note

About the name of userApplication

The character string set by ApplicationName menu of the hvw command is converted to lower case, and used for the cluster application name.

ApplicationName must satisfy all the conditions below:

- Must be a combination of uppercase letters, numbers, and "_" (underscore).
- Must start with an uppercase letter.
- Up to 14 letters.
- To the identifier, do not specify the same name as the configuration file name specified to the argument of hvw command described in "6.7.1 Starting RMS Wizard."

6.7.2.1 Creating Standby Cluster Applications

This section explains how to configure a cluster application for standby operation.

Operation Procedure:

1. Select "Application-Create" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 3
```

2. Select "STANDBY" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                           7) SCALABLE
 2) QUIT                           8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 8
```

3. Next, "turnkey wizard STANDBY" will be output. Select "Machines+Basics" and then set up userApplication.

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP1
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

4. The userApplication setup page will appear. Set up the following for the userApplication:

   - Nodes that constitute the userApplication
   - Attributes of the userApplication

   Set up the nodes that constitute userApplication by selecting "Machines[number]" and then a SysNode name on the subsequent screen that is displayed.

   The procedures for setting up the nodes that constitute a userApplication and cluster application priority are explained for each topology, below.

Topology	How to set up userApplication configuration nodes and cluster application priority
1:1 standby	In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up.

Topology	How to set up userApplication configuration nodes and cluster application priority
	In "Machines[1]," specify a SysNode that is in standby status or Offline when the userApplication first starts up
Mutual standby	**(For the first userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. In "Machines[1]," specify a SysNode that is in standby status or Offline when the userApplication first starts up. **(For the second userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. For this SysNode, specify the SysNode specified for "Machines[1]" when the first userApplication was set up. In "Machines[1]," specify a SysNode that is in standby status or Offline when the userApplication first starts up. Specify the SysNode specified in "Machines[0]" when the first userApplication was set up.
N:1 standby	**(For the first userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. In "Machines[1]," specify a SysNode that is in standby status or Offline when the userApplication first starts up. **(For the second or subsequent userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. For this, specify a SysNode other than that previously specified for "Machines[0]" or "Machines[1]" when the userApplication was set up. In "Machines[1]," specify a SysNode that is in standby status or Offline when the userApplication first starts up. For this, specify the same SysNode as that previously specified in "Machines[1]" when the userApplication was set up.
Cascaded	In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. For "Machines[1]" or later, specify a SysNode that is in standby status or Offline when the userApplication first starts up. State transition occurs in ascending order of the numbers specified for "Machines[number]." Example) When there are four nodes, state transition occurs in the order shown below: "Machines[0]" -> "Machines[1]" -> "Machines[2]" -> "Machines[3]"
Priority transferring	**(For the first userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. For "Machines[1]" or later, specify a SysNode that is in standby status or Offline when the userApplication first starts up. **(For the second or subsequent userApplication)** In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. For this, specify a SysNode other than that previously specified in "Machines[0]" when the userApplication was set up.

Topology	How to set up userApplication configuration nodes and cluster application priority
	For "Machines[1]" or later, specify a SysNode that is in standby status or Offline when the userApplication first starts up. State transition occurs in ascending order of the numbers specified in "Machines[number]." Example) When there are four nodes, state transition occurs in the order shown below: "Machines[0]" -> "Machines[1]" -> "Machines[2]" -> "Machines[3]"

Set up the attributes of the userApplication as follows:

Attribute	Setup value	Contents	Remark
AutoStartUp	yes	Automatically starts up the cluster application when RMS is started.	To create a cluster application in standby operation that constitutes scalable operation, set "AutoStartUp" to "no." For information on how to create scalable cluster applications, see "6.7.2.2 Creating Scalable Cluster Applications."
AutoSwitchOver	HostFailure| ResourceFailure| Shutdown	Automatically performs failover if a node or resource fails or when the node is stopped.	Do not set a value in the single-node cluster operation.
HaltFlag	yes	The shutdown facility forcibly stops the application if another error (double errors) occurs during failover.	To ensure safe operation, always set "yes." Set [No] in the single-node cluster operation.
StandbyTransitions	ClearFaultRequest| StartUp| SwitchRequest	Monitor the states of the resources on the standby node performing standby operation.	This setting must be made when you are monitoring the states of the Gls resources on the standby node by using the redundant line control function of GLS.
LicenseToKill	yes	Sets up an exclusive relationship between cluster applications.	For information on exclusive relationships, see the following.
AutoBreak	yes	Cancels an exclusive relationship between cluster applications.	
PartialCluster	0	Allows userApplication to start even if RMS on all the nodes configuring userApplication is not started.	On cluster applications for a standby operation, "0" is set unless otherwise directed.

Information

For more information and the list of attributes settable to userApplication, refer to "12.1 Attributes available to the user" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Note

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".

To set up exclusive relationships between cluster applications, you must set up the following.
For details on exclusive relationships between applications, see "6.7.7 Exclusive Relationships Between Cluster Applications."

Create multiple cluster application groups between which an exclusive relationship can be established. Exclusive control is established between the cluster applications within a single group.

Up to 52 groups of A to Z or a to z can be specified. "20X" and "10X" are fixed values. Therefore, you must always specify either "20X" or "10X" after the group.

- Example) When the cluster application is included in group A and the job priority is high

```
A20X
```

- Example) When the cluster application is included in group A and the job priority is low

```
A10X
```

Note

Exclusive relationships between cluster applications can be established only when the operation is being performed with two or more cluster applications. When the operation is to be performed with one cluster application, do not set up any relationships between cluster applications.

- When a cluster application with a high job priority is to be used

  Select "LicenseToKill" and then enter the following into ">>" that is output after "FREECHOICE" has been selected.

```
Group 20X
```

- When a cluster application with a low job priority is to be used

  Select "LicenseToKill" and then enter the following into ">>" that is output after "FREECHOICE" has been selected.

```
Group 10X
```

- When an exclusive relationship is to be cancelled, set up the following:

```
"LicenseToKill" : "no"

"AutoBreak"     : "yes"
```

After the completion of setup, select "SAVE+EXIT."

Note

Operator intervention requests and error resource messages are displayed only when the AutoStartUp and PersistentFault attributes are set to yes(1). When the operator intervention and error resource messages are to be displayed, set yes(1) for the AutoStartUp and PersistentFault attributes. For information on the operator intervention and error resource messages, see "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

Information

The following scripts can be registered to userApplication. For more information on each script, refer to "12 Appendix-Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

- FaultScript
- PreCheckScript
- PreOnlineScript
- PostOnlineScript
- PreOfflineScript
- OfflineDoneScript

Do not use a tilde (~) for the command path or the argument set to each script.

5. Confirm if the setup information is correct, and then select "SAVE+EXIT."

```
Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=0)
26) (ScriptTimeout=)
Choose the setting to process: 3
```

6. "turnkey wizard STANDBY" is output. Specify the settings for each resource.

```
Consistency check ...
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                        9) Procedure:BasicApplication(-)
 2) -                          10) Symfoware(-)
 3) SAVE+EXIT                  11) Procedure:SystemState3(-)
 4) -                          12) Procedure:SystemState2(-)
 5) ApplicationName=APP1       13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)      14) IpAddresses(-)
 7) CommandLines(-)            15) LocalFileSystems(-)
 8) Procedure:Application(-)   16) Gds:Global-Disk-Services(-)
Choose the setting to process:
```

6.7.2.2 Creating Scalable Cluster Applications

This section explains how to register a cluster application in scalable operation.

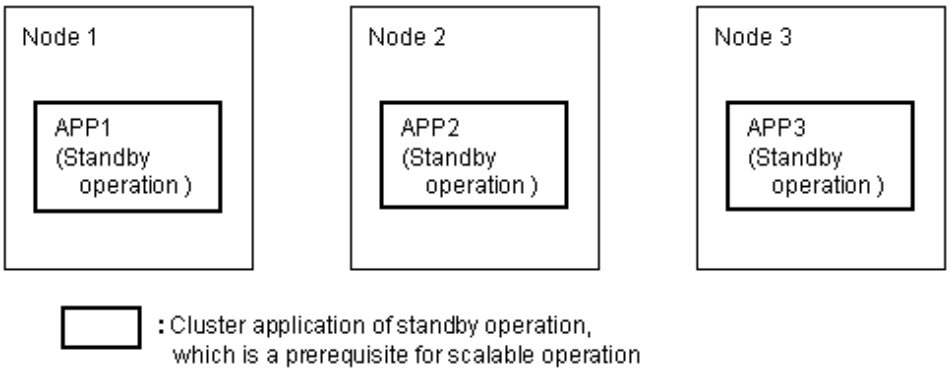
Preparing standby cluster applications

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.

Example 1) Preparing for scalable operation

When you create a cluster application in a scalable operation, you must first create a cluster application in a standby operation, which is a prerequisite for scalable operation.

If the cluster application of scalable operation is to run on three nodes, create a cluster application of standby operation on each of those nodes (the node is for operation only and has no standby).

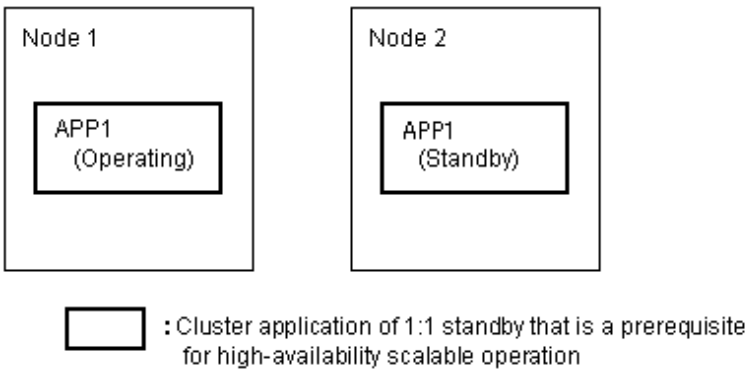


When you create a cluster application for standby operation, which is a prerequisite for scalable operation, set up only "Machines[0]."

Example 2) Preparing for high-availability scalable operation

To create a high-availability scalable cluster application, you must first create a cluster application for standby operation, which is a prerequisite for high-availability scalable operation.

If the cluster application for high-availability scalable operation is 1:1 standby, create a cluster application for 1:1 standby.



Note

To create a cluster application in standby operation that constitutes scalable operation, set "AutoStartUp" to "no." To start the cluster applications automatically when you start RMS, set the value of "AutoStartUp" to "yes" when you create a cluster application as part of scalable operation.

The procedure for setting up the node of a cluster application in a standby operation, which is a prerequisite for scalable operation, is as shown below.

Topology	How to set up userApplication configuration nodes
Scalable	In "Machines[0]," specify a SysNode that is Online when the userApplication first starts up. Since standby is not included, you do not need to set up "Machines[1]" and any subsequent items.
High-availability scalable	Note that the set-up method varies depending on the topology of the standby operation that is a constituent factor of the cluster application in scalable operation. For information on making this setting, see how to set up the topology of each standby operation.

For information on how to create standby cluster applications, see "6.7.2.1 Creating Standby Cluster Applications."

After you complete the setup of standby operation, which is a prerequisite for scalable operation, you must create the cluster application of scalable operation as explained below.

Creating scalable cluster applications

Operation Procedure:

1. Select "Application-Create" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 3
```

2. Select "SCALABLE" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 7
```

3. "turnkey wizard "SCALABLE"" is output. Select "Machines+Basics" and set up the userApplication.

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "SCALABLE" (APP2:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP2
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

4. The userApplication setup screen is output. Specify the following settings for the userApplication:

   - Nodes where the userApplication is configured
   - userApplication attributes

   Set up the nodes where the userApplication is configured as follows:

   - Specify all SysNode names where the cluster application is configured (standby operation) in "Machines[*number*]".

   Refer to the following when setting the userApplication attributes:

Attribute	Setting	Description	Remarks
AutoStartUp	yes	Automatically starts the cluster application when RMS is started.	

5. For "Settings of application type," select "SAVE+EXIT."

```
Consistency check ...

Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ShutDown)
17) (PreserveState=yes)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=0)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=1)
26) (ScriptTimeout=)
Choose the setting to process: 3
```

6. "turnkey wizard "SCALABLE"" is output. Select "Controllers."

```
Consistency check ...
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "SCALABLE" (APP2:not yet consistent)
 1) HELP               4) -                        7) Controllers(-)
 2) -                  5) ApplicationName=APP2
 3) SAVE+EXIT          6) Machines+Basics(app2)
Choose the setting to process: 7
```

7. "Settings of application type" is output. Select "AdditionalAppToControl."

```
Consistency check ...
Yet to do: assign at least one application to control
Yet to do: configure at least one controlled application without the M flag

Settings of application type "Controller" (not yet consistent)
 1) HELP              4) REMOVE+EXIT              7) (FaultScript=)
 2) -                 5) ControlPolicy=SCALABLE   8) (ApplicationSequence=)
 3) SAVE+EXIT         6) AdditionalAppToControl   9) (StateChangeScript=)
Choose the setting to process: 6
```

8. Select a cluster application (standby operation) that allows scalable operation.

Information

All of a cluster application of standby operation is displayed with lowercase characters.

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) app1
 5) app2
Choose an application to control: 4
```

9. Select "SAVE+RETURN" from "Set global flags for scalable."

```
Set *global* flags for all scalable (sub) applications: app1
Currently set: TIMEOUT (T180)
 1) HELP                              5) MONITORONLY(M)
 2) -                                 6) TIMEOUT(T)
 3) SAVE+RETURN
 4) DEFAULT
Choose one of the flags: 3
```

10. To allow scalable operation with multiple cluster applications (standby operation), repeat steps 7. to 9.

11. Set up the order in which cluster applications are started up (standby operation). When you start the cluster applications, start from the one with the smallest startup sequence number. When stopping, from the one with the largest startup sequence number. Cluster applications with the same startup sequence number must start up or stop in parallel.

Note

If you do not need to set up a startup sequence number, you do not have to perform the procedure described below.

1. Select "(ApplicationSequence=)" from "Settings of application type."

```
Settings of application type "Controller" (consistent)
 1) HELP                               7) Controllers[0]=T180:app1
 2) -                                  8) Controllers[1]=T180:app2
 3) SAVE+EXIT                          9) (FaultScript=)
 4) REMOVE+EXIT                       10) (ApplicationSequence=)
 5) ControlPolicy=SCALABLE            11) (StateChangeScript=)
 6) AdditionalAppToControl
Choose the setting to process: 10
```

2. Select "FREECHOICE."

```
 1) HELP
 2) RETURN
 3) NONE
 4) FREECHOICE
Set the application sequence: 4
```

3. Enter the startup sequence number, and then press the return key.

   - Enter the cluster application with the highest startup sequence number first.
   - If the startup sequence numbers are different, input a single colon (:) between the cluster applications.
   - If the startup priority numbers are the same, input a single space between the cluster applications.

   Note

   The cluster application for standby operation must be entered entirely in lowercase characters.

   The following is an example in which the startup sequence of app1 is the first, followed by app2 and then app3 (app2 and app3 have the same startup sequence number).

```
 1) HELP
 2) RETURN
 3) NONE
 4) FREECHOICE
Set the application sequence: 4
      >> app1:app2 app3
```

12. Select "SAVE+EXIT" from "Settings of application type."

```
Settings of application type "Controller" (consistent)
 1) HELP                              7) Controllers[0]=T180:app1
 2) -                                 8) Controllers[1]=T180:app2
 3) SAVE+EXIT                         9) (FaultScript=)
 4) REMOVE+EXIT                      10) (ApplicationSequence=app1:app2)
 5) ControlPolicy=SCALABLE           11) (StateChangeScript=)
 6) AdditionalAppToControl
Choose the setting to process: 3
```

When two or more cluster applications for scalable operation are to be created, repeat steps 1. to 12.

6.7.3 Setting Up Resources

This section explains how to register resources to the userApplication that was set up in the previous section.

You can register the following resources:

- **Cmdline resources**

  You can use Cmdline resources to set up script files or commands as resources. The Cmdline resources are required to generate the state transition of userApplication along with the stop of user applications, and conversely, to start or stop ISV applications or user applications along with the state transition of the userApplication.

- **Fsystem resources**

  Used when you mount a file system along with userApplication startup.

  Note

  To use a file system in a class created by GDS as an Fsystem resource, you must register the Gds resource to the same userApplication.

- **Gds resources**

  Used when you start and stop a disk class to be defined by GDS by linking it with the userApplication.

- **Gls resources**

  Used when you set up a takeover IP address that is to be defined in a userApplication with the redundant line control function of GLS, or when you set a takeover IP address in a userApplication with the single line control function.

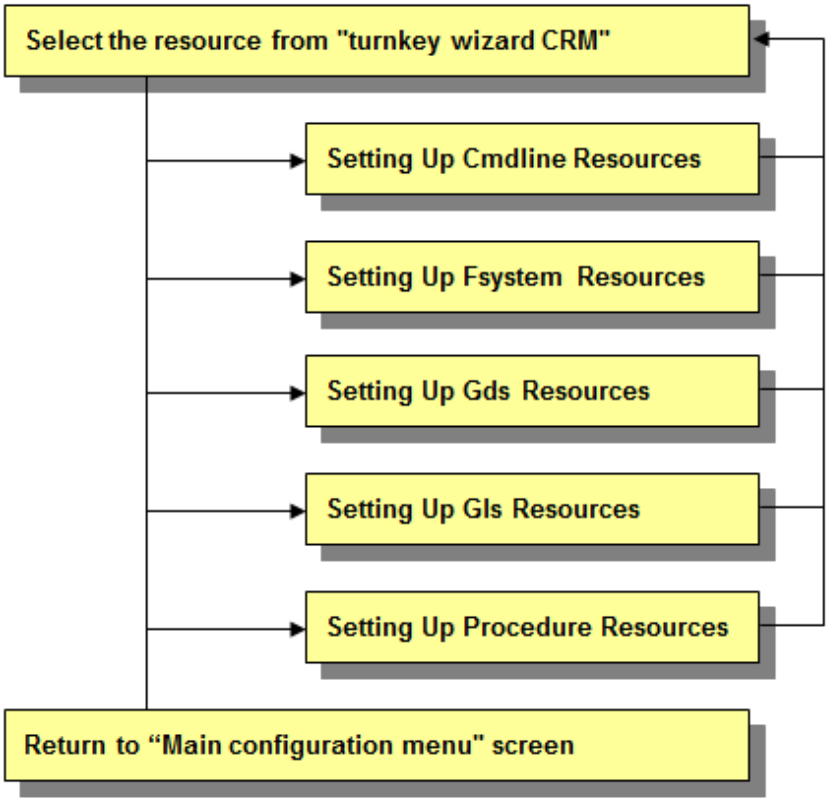
- **Takeover network resource**

  Used when you set up a takeover IP address on a single network interface without using GLS.

- **Procedure resources**

  Used when you register a state transition procedure in the userApplication.

Resource setup flow



6.7.3.1 Setting Up Cmdline Resources

This section describes the procedure for creating Cmdline resources.

For notes when setting Cmdline resources, see "6.11 Notes When Setting Cmdline Resources."

Operation Procedure:

1. Select "CommandLines" from "turnkey wizard STANDBY."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          9) Procedure:BasicApplication(-)
 2) -                            10) Symfoware(-)
 3) SAVE+EXIT                    11) Procedure:SystemState3(-)
 4) -                            12) Procedure:SystemState2(-)
 5) ApplicationName=APP1         13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)        14) IpAddresses(-)
 7) CommandLines(-)              15) LocalFileSystems(-)
 8) Procedure:Application(-)     16) Gds:Global-Disk-Services(-)
Choose the setting to process:7
```

2. "CommandLines" will appear. Select "AdditionalStartCommand."

```
CommandLines (Cmd_APP1:not yet consistent)
1) HELP                          4) REMOVE+EXIT
2) -                             5) AdditionalStartCommand
3) SAVE+EXIT                     6) (ReturnCodes)
Choose the setting to process: 5
```

3. Select "FREECHOICE" and then enter the full path of the StartCommand. If you need to specify arguments, delimit them with blanks.

   StartCommand is executed during Online processing to start user applications.

   [StartCommand exit codes]

   StartCommand has the following exit codes:

   0: Normal exit. The Online processing is successfully done.

   Other than 0: Abnormal exit. The Online processing fails. When the script exits with the cord other than 0, the resource will enter Faulted.

```
1) HELP
2) RETURN
3) FREECHOICE
Set a start command: 3
       >>
```

Note

The following characters cannot be used in the script path and the arguments that set for StartCommand, and StopCommand and CheckCommand to be described later.

```
=    \    ~    %    @    &
```

If you need to use those characters, describe them within the script that sets to Cmdline resources.

4. "CommandLines" will appear. If you need to stop the user programs, select "StopCommands."

   StopCommand is executed during Offline processing to stop user applications.

   You do not always have to set up the StopCommand.

   [StopCommand exit codes]

   StopCommand has the following exit codes:

   0: Normal exit. The Offline processing is successfully done.

Other than 0: Abnormal exit. The Offline processing fails. When the script exits with the cord other than 0, the resource will enter Faulted.

If you do not use StopCommand, start from step 6.

```
CommandLines (Cmd_APP1:consistent)
 1) HELP                              7) StopCommands[0]=none
 2) -                                 8) CheckCommands[0]=none
 3) SAVE+EXIT                         9) CheckCommandTimeouts[0]=none
 4) REMOVE+EXIT                      10) Flags[0]=DT300
 5) AdditionalStartCommand           11) (ReturnCodes)
 6) StartCommands[0]='xxxx'
Choose the setting to process:7
```

Note

If "none" is set to StopCommands, regardless of the settings of Flags, LIEOFFLINE attribute is enabled and CLUSTEREXCLUSIVE is disabled. In this status, the Cmdline resource is started and monitored.

5. Select "FREECHOICE" and then enter the full path of StopCommand. If you need to specify arguments, delimit them with blanks.

```
1) HELP
2) RETURN
3) FREECHOICE
4) NONE
Set a start command: 3
       >>
```

6. "CommandLines" will appear. Select "CheckCommands."

CheckCommand is executed periodically to notify RMS of the state of the user applications.

[CheckCommand exit codes]

CheckCommand has the following exit codes:

0: Online.

Other than 0: Offline.

If you enable the ALLEXITCODES attribute, the script that is defined in CheckCommand provides more detailed state of the user applications. Change the attribute in step 8.

```
CommandLines (Cmd_APP1:consistent)
 1) HELP                              7) StopCommands[0]='yyyy'
 2) -                                 8) CheckCommands[0]=none
 3) SAVE+EXIT                         9) CheckCommandTimeouts[0]=none
 4) REMOVE+EXIT                      10) Flags[0]=DT300
 5) AdditionalStartCommand           11) (ReturnCodes)
 6) StartCommands[0]='xxxx'
Choose the setting to process:8
```

7. Select "FREECHOICE" and then enter the full path of the CheckCommand. If you need to specify arguments, delimit them with blanks.

```
1) HELP
2) RETURN
3) FREECHOICE
4) NONE
Set a start command: 3
      >>
```

8. Change the attribute.

   Change the attribute to suit the purpose. To change the attribute, select "Flags[0]." For details on the attribute, see "Table 6.2 Attributes of the Cmdline resource" in "6.11 Notes When Setting Cmdline Resources."

   Note

   If you enable the "NULLDETECTOR" attribute, CheckCommand is not started from RMS. For hot-standby operation, enable the following two attributes;

   - STANDBYCAPABLE

     RMS executes Standby processing of the resources on all the nodes where the userApplication is Offline.

   - ALLEXITCODES

     Check script provides the detailed state of the resource with the exit code.

     For further details about the hot-standby operation settings, see "6.6 Setting Up Online/Offline Scripts."

9. Finally, select "SAVE+EXIT."

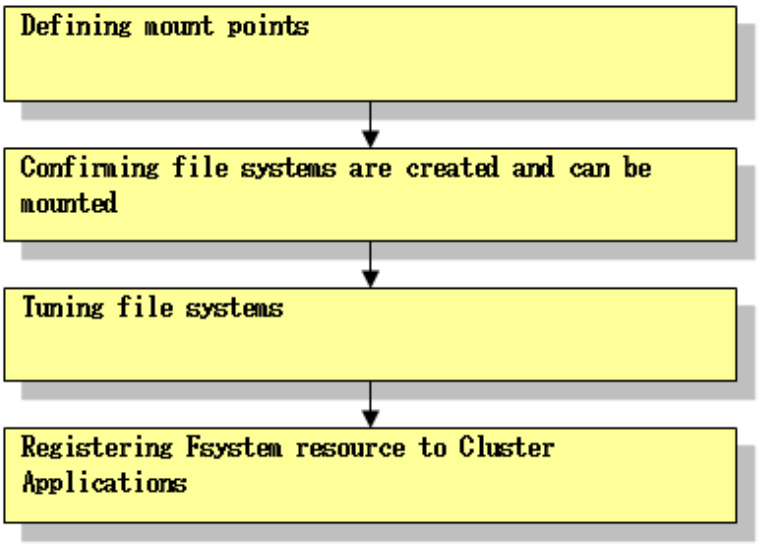
```
CommandLines (Cmd_APP1:consistent)
 1) HELP                              7) StopCommands[0]='yyyy'
 2) -                                 8) CheckCommands[0]=none
 3) SAVE+EXIT                         9) CheckCommandTimeouts[0]=none
 4) REMOVE+EXIT                      10) Flags[0]=DT300
 5) AdditionalStartCommand           11) (ReturnCodes)
 6) StartCommands[0]='xxxx'
Choose the setting to process:3
```

6.7.3.2 Setting Up Fsystem Resources

An Fsystem resource must be set up if you want to mount a file system when userApplication is started.

For notes when setting Fsystem resources, see "6.12 Notes When Setting Fsystem Resource."

You need to set up the Fsystem resource according to the flow below:



Note

The file system on the volume of LVM (Logical Volume Manager) cannot be controlled in Fsystem resource.

1. Defining mount points

   Define mount points for file systems that are used as resources in the /etc/fstab.pcl file on all the nodes where userApplication is configured. Each line must begin with "#RMS#."

   The description format of the first field to the sixth field in a line beginning with "#RMS#" is the same as that of the /etc/fstab file of RHEL. For details, refer to the official documentation of RHEL.

   Example: /etc/fstab.pcl file

```
#RMS#/dev/sdd1          /mnt/swdsk1          ext3     defaults        0 0
#RMS#/dev/sdd2          /mnt/swdsk2          ext4    nodelalloc      0 0
#RMS#/dev/sdd3          /mnt/swdsk3          xfs     defaults       0 0
```

   If you plan to use GDS volumes, you need to define the /etc/fstab.pcl file as follows.

   Example: /etc/fstab.pcl file

```
#RMS#/dev/sfdsk/class0001/dsk/volume0001 /mnt/swdsk1 ext3  defaults    0 0
#RMS#/dev/sfdsk/class0001/dsk/volume0002 /mnt/swdsk2 ext4  nodelalloc  0 0
#RMS#/dev/sfdsk/class0001/dsk/volume0003 /mnt/swdsk3 xfs   defaults    0 0
```

   ### Note

   - If you have defined the same device or mount point in the /etc/fstab file, those definitions can be removed by making them into comment lines. If those definitions are remained, userApplications may fail to be started normally.
   - Ext4 and xfs are used to make the allocation of the disk area more efficiently, and to improve the writing performance, by using their "Delayed Allocation" feature. As a result of the implementation of "Delayed Allocation", there is a possibility that a part of data is lost by OS panic or power supply interruption of servers, because the sojourn time on the memory of data that should be stored on the disk becomes longer.

     When a program has to guarantee writing immediately after writing in file system, the application which writes the file should issue the fsync() call. Refer to Storage Administration Guide of the Red Hat, Inc. for "Delayed allocation."
   - For the directory paths that are specified as the mount points, specify any paths that do not include symbolic links.

2. Starting the 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
```

3. Confirming file systems are created and can be mounted

   It is necessary to create file systems on disk partitions that are used as resources. Refer to Storage Administration Guide of the Red Hat, Inc. for "Create file systems".

   Example: Creating the ext3 file system

   Create the file system.

```
# /sbin/mkfs -t ext3 /dev/sdd1
```

   Check if the file systems can be mounted.

```
# /bin/mount -t ext3 /dev/sdd1 /mnt/swdsk1
# /bin/umount /mnt/swdsk1
```

   Example: Creating the ext4 file system

   Create the file system.

```
# /sbin/mkfs.ext4 /dev/sdd2
```

   Check if the file systems can be mounted.

```
# /bin/mount -t ext4 /dev/sdd2 /mnt/swdsk2
# /bin/umount /mnt/swdsk2
```

   Example: Creating the xfs file system

   Create the file system.

```
# /sbin/mkfs.xfs /dev/sdd3
```

   Check if the file systems can be mounted.

```
# /bin/mount -t xfs /dev/sdd3 /mnt/swdsk3
# /bin/umount /mnt/swdsk3
```

4. Tuning file systems

   Set up the cluster environment in according to their file system types.

   - Forcible file system check prevention (recommended for ext3 and ext4)

     If ext3 or ext4 is used for a file system, the file system might forcibly be checked during Online processing of a switching file system. It is part of the ext3 and ext4 specification that file systems are checked when a certain number of mounting has been executed since the last file system check, or a certain period of time has passed.

     When the file systems are forcibly checked along with node startup or failover, timeout occurs due to file system Online processing, and PRIMECLUSTER startup or failover might fail.

     It is necessary to prevent the file systems from being checked by executing the following command for all the ext3 and ext4 switching files.

     Example: Configuring and confirming the prevention of file systems from being checked

```
# /sbin/tune2fs -c0 -i0 /dev/sdd1
```

     After executing the above command, check if "Maximum mount count :-1", "Check interval:0"is displayed using the following command:

```
# /sbin/tune2fs -l /dev/sdd1 | /bin/grep "Maximum mount count"
Maximum mount count:      -1
# /sbin/tune2fs -l /dev/sdd1 | /bin/grep "Check interval"
Check interval:           0 (<none>)
```

If the forcible file system check is prevented, file systems might corrupt due to failures such as disk errors and kernel bug. These failures cannot be detected through file system logging and journaling. The file system corruption might cause data corruption.

To prevent this, execute the "fsck - f" command to enable the file system forcible check during periodic maintenance.

- Set Delayed Allocation disabled. (Only for ext4)

  For ext4 file systems, Delayed Allocation feature can be disabled by specifying nodelalloc for the mount attribute.

  You need to specify mount attribute field in the /etc/fstab.pcl file as follows.

```
#RMS#/dev/sdd2        /mnt/swdsk2          ext4    nodelalloc     0  0
```

5. Stopping the GDS volume (Only when Step 2 has already been implemented)

   Stop the GDS volume started in Step 2.

   Example: Stopping the volume volume0001 of the disk class class with a command

```
# /usr/sbin/sdxvolume -F -c class -v volume0001
```

6. Registering Fsystem resource to Cluster Applications

   1. Select "LocalFileSystems" from "turnkey wizard STANDBY."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                                 9) Procedure:BasicApplication(-)
 2) -                                   10) Symfoware(-)
 3) SAVE+EXIT                           11) Procedure:SystemState3(-)
 4) -                                   12) Procedure:SystemState2(-)
 5) ApplicationName=APP1                13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)               14) IpAddresses(-)
 7) CommandLines(-)                     15) LocalFileSystems(-)
 8) Procedure:Application(-)            16) Gds:Global-Disk-Services(-)
Choose the setting to process: 15
```

   2. Select "AdditionalMountPoint."

```
File systems (Lfs_APP1:not yet consistent)
1) HELP                  4) REMOVE+EXIT           7) (Timeout=180)
2) -                     5) AdditionalMountPoint
3) SAVE+EXIT             6) (Filter=)
Choose the setting to process: 5
```

   3. The mount points, which are defined in the above-mentioned /etc/fstab.pcl file for preliminary setup, will appear. Select the mount point to be set in userApplication.

```
1) HELP                                 6) /mnt/swdsk2
2) RETURN                               7) /mnt/swdsk3
3) FREECHOICE
4) ALL
5) /mnt/swdsk1
Choose a mount point: 5
```

   4. Select "SAVE+RETURN."

```
Set flags for mount point: /mnt/swdsk1 Currently set: LOCAL,AUTORECOVER (LA)
1) HELP                  4) DEFAULT               7) SHARE(S)
2) -                     5) SYNC(Y)               8) MONITORONLY(M)
3) SAVE+RETURN           6) NOT:AUTORECOVER(A)
Choose one of the flags: 3
```

5. If you register multiple mount points, repeat steps 2 to 4 for each mount point. After you have registered all necessary mount points, Select "SAVE+EXIT."

```
File systems (Lfs_APP1:consistent)
 1) HELP                                6) MountPoints[0]=LA:/mnt/swdsk1
 2) -                                   7) MountPoints[1]=LA:/mnt/swdsk2
 3) SAVE+EXIT                           8) MountPoints[2]=LA:/mnt/swdsk3
 4) REMOVE+EXIT                         9) (Filter=)
 5) AdditionalMountPoint               10) (Timeout=180)
Choose the setting to process:  3
```

6.7.3.3 Preliminary Setup for Gds Resources

[Prerequisites]

If you need to set up a Gds resource, you must first set up a shared volume.
Also, before you make the settings required for the Gds resources, execute the following command on either node.

This command operation is required for linking the specified Gds resources with the status of the userApplication and to start and stop the GDS volume. In addition, this command sets all the nodes sharing the volume to Offline status.

```
# /opt/SMAW/SMAWRrms/bin/hvgdsetup -a [class-name]
 ...
Do you want to continue with these processes ? [yes/no] y
```

Information

To check the setup status of a shared volume, execute the following command:

```
# /opt/SMAW/SMAWRrms/bin/hvgdsetup -l
```

Note

- If the preliminary setup is not performed, the cluster application is set to Inconsistent status. For details, see "Cluster applications become "Inconsistent"." in "Cluster System Related Error" of "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."
- This operation must not be performed when a GFS shared file system is used.

6.7.3.4 Setting Up Gds Resources

After completing the preliminary setup for the Gds resources, register the Gds resources to the userApplication. The procedure for setting Gds resources to a userApplication is described below.

Operation Procedure:

1. Select "Gds:Global-Disk-Services" from "turnkey wizard STANDBY."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) -                                10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(-)                  15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:16
```

2. Select "AdditionalDiskClass" from "Volume management."

   If you want to register multiple resources, repeat steps 2 to 3 for each resource.

   Example) When you register the class [CL] of the shared volume in userApplication:

```
Volume management (Gds_APP1:not yet consistent)
 1) HELP                 4) REMOVE+EXIT          7) (StandbySupport=no)
 2) -                    5) AdditionalDiskClass  8) (AutoRecover=no)
 3) SAVE+EXIT            6) (ClassNameFilter=)   9) (Timeout=1800)
Choose the setting to process:5
```

3. A list of the cluster resources that can be registered in userApplication appears.

   - In the case of the multiple-node cluster operation
     Select the cluster resource.

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) CL
Choose the setting to process:4
```

   - In the case of the single-node cluster operation
     Enter the created class name of GDS by selecting "FREECHOICE", since the list of the cluster resources is not displayed.

4. Select "SAVE+EXIT."

```
Volume management (Gds_APP1:consistent)
 1) HELP                 5) AdditionalDiskClass   9) (AutoRecover=no)
 2) -                    6) DiskClasses[0]=CL    10) (Timeout=1800)
 3) SAVE+EXIT            7) (ClassNameFilter=)
 4) REMOVE+EXIT          8) (StandbySupport=no)
Choose the setting to process:3
```

6.7.3.5 Setting Up Gls Resources

[Prerequisites]

Before setting up the Gls resources, you must first set up the virtual interfaces and takeover virtual interfaces. For details, see "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

Operation Procedure:

1. Select "Gls:Global-Link-Services" from "turnkey wizard STANDBY."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) -                                10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(-)                  15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:13
```

2. Select "AdditionalTakeoverIpaddress" and then set the takeover IP address.

   If you need to register multiple resources, repeat steps 2 to 4 for each resource.

```
Gls (Gls_APP1:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) -                             5) AdditionalTakeoverIpaddress
 3) SAVE+EXIT                     6) (Timeout=60)
Choose the setting to process: 5
```

3. The takeover IP addresses are displayed as options.

   Select one.

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 10.34.214.185
Choose a takeover IP address for Gls: 4
```

4. Set up the operation mode of the resource.

   The operation modes of the resources are "AUTORECOVER(A)" (selective value) and "NOT:AUTORECOVER(N)" (default value). If you select "SAVE+RETURN," the default value "NOT:AUTORECOVER(N)" will be set. Select "AUTORECOVER(A)" if an attempt to recover the resource is to be made for a given duration (default: 60 seconds) when a resource error occurs, or select "NOT:AUTORECOVER(N)" if switchover to another node is to be performed immediately.

```
Set a flag for takeover IP address: 10.34.214.185
Currently set:
 1) HELP                          5) AUTORECOVER(A)
 2) -
 3) SAVE+RETURN
 4) DEFAULT
Choose additionally one of the flags: 3
```

5. To save the Gls resource settings and then exit, select "SAVE+EXIT."

   You can change the timeout value of the Gls resource by selecting "(Timeout=60)" and setting any value (seconds).

```
Gls (Gls_APP1:consistent)
 1) HELP                              5) AdditionalTakeoverIpaddress
 2) -                                 6) TakeoverIpaddress[0]=N,10.34.214.185
 3) SAVE+EXIT                         7) (Timeout=60)
 4) REMOVE+EXIT
Choose the setting to process: 3
```

See

By setting up the value in the StandbyTransition attribute when the cluster application is created, Gls resources on the standby node can be switched to the "Standby" state and the state of the Gls resources on the standby node can be monitored. For information on how to make this setting, see "6.7.2.1 Creating Standby Cluster Applications."

6.7.3.6 Setting Up Takeover Network Resources

[Prerequisites]

You need to configure the following information before using takeover network resources.

1. Define an IP address and a host name to be taken over between nodes

   First, define a host address to be taken over between nodes. Next, allocate a host name to this IP address. The allocated host name should be defined in the /etc/hosts file for the node which uses the takeover IP address.

2. Define a target host

   If you want to monitor the reachability of your network using ICMP while a takeover IP address is active, define a host name and an IP address of the device (e.g. router or layer 3 hub) which enables ICMP communication between more than one node in the /etc/hosts file. This process can be skipped if you do not need to monitor the reachability of the network.

3. Confirming a network interface

   Confirm the network interface (e.g. eth2) on which the takeover IP address will be active is being recognized by the operating system and available.

4. Define a setting file for the cluster

   Add the takeover IP address to the following file:
   /usr/opt/reliant/etc/hvipalias
   Enter entries according to the following rules.

```
<node name> <takeover> <interface> <netmask/prefix>
```

```
<node name> : CF node name of the node which uses the takeover IP address
<takeover>  : Host name of the takeover IP address
<interface> : Network interface name on which the takeover IP address will be activated
<netmask/prefix>   : Netmask for the takeover IP address (for IPv4), or network prefix length (for
IPv6)
```

Example

When an IPv4 address for the host "takeover" (netmask 255.255.255.0) is taken over between two nodes (node0 and node1) on the network interface eth2, define as follows (specify the 8-digit netmask in hexadecimal).

```
node0 takeover eth2 0xffffff00
node1 takeover eth2 0xffffff00
```

When an IPv6 address for the host "takeover6" (network prefix length: 64) is taken over on the network interface eth3, define as follows.

```
node0 takeover6 eth3 64
node1 takeover6 eth3 64
```

Note

- An IPv6 link local address cannot be used as a takeover network resource. Moreover, it cannot be used as a communication destination of reachability monitoring.
- When defining a host name in the /etc/hosts file, do not assign the same host name to the IPv4 address and the IPv6 address.

Operation Procedure:

1. Select "IpAddresses" of turnkey wizard "STANDBY"

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
1) HELP                                 9) Procedure:BasicApplication(-)
2) -                                   10) Symfoware(-)
3) SAVE+EXIT                           11) Procedure:SystemState3(-)
4) -                                   12) Procedure:SystemState2(-)
5) ApplicationName=APP1                13) Gls:Global-Link-Services(-)
6) Machines+Basics(app1)               14) IpAddresses(-)
7) CommandLines(-)                     15) LocalFileSystems(-)
8) Procedure:Application(-)            16) Gds:Global-Disk-Services(-)
Choose the setting to process:14
```

2. When you have previously specified the target host to monitor its network reachability using ICMP, select "AdditionalPingHost" and specify that target host.

```
Ipaddresses and ipaliases (Adr_APP1:not yet consistent)
1) HELP                  4) REMOVE+EXIT            7) (NeedAll=yes)
2) -                     5) AdditionalInterface    8) (Timeout=60)
3) SAVE+EXIT 6) AdditionalPingHost 9) (InterfaceFilter=)
Choose the setting to process:6
```

   The target host name registered in the process of prerequisites will be shown as an option. Select the host name you have previously specified.

```
1) HELP
2) RETURN
3) FREECHOICE
4) router
5) l3hub
6) takeover
Choose another trusted host to ping:4
```

   When you finish specifying the target host, you will be brought back to the previous screen. Since you are required to specify more than one target host, you need to select "AdditionalPingHost" again to add another target host on the previous screen.

3. Select "AdditionalInterface" to set up the takeover IP address.

   When you have more than one IP address, you need to repeat this process for each IP address.

```
Ipaddresses and ipaliases (Adr_APP1:not yet consistent)
1) HELP                                 7) PingHostPool[0]=router
2) -                                    8) PingHostPool[1]=l3hub
3) SAVE+EXIT                            9) (NeedAll=yes)
4) REMOVE+EXIT                         10) (Timeout=60)
5) AdditionalInterface                 11) (InterfaceFilter=)
6) AdditionalPingHost
Choose the setting to process:
```

1. Takeover IP address registered in the process of prerequisites will be shown as an option.

   Select the host name for the takeover IP address you have previously specified.

```
1) HELP
2) RETURN
3) FREECHOICE
4) router
5) l3hub
6) takeover
Choose an interface name:6
```

2. Specify the operation mode for the takeover IP address.

```
Currently set: VIRTUAL,AUTORECOVER (VA)
1) HELP                  4) DEFAULT               7) MONITORONLY(M)
2) -                     5) BASE(B)               8) PING(P)
3) SAVE+RETURN           6) NOT:AUTORECOVER(A)
Choose one of the flags:
```

   - DEFAULT

     If you choose "DEFAULT", all values will revert back to their default values.

   - BASE, VIRTUAL

     This attribute is effective only when using an IPv4 address. When using an IPv6 address, do not change this attribute. The default value is "VIRTUAL".

     - BASE

     If you specify "BASE", activation/deactivation of the takeover IPv4 address and activation/deactivation of the physical interface (for example, eth2) are performed at the same time. "BASE" will be shown on "Currently set" and "5) VIRTUAL" is shown on the menu page.

     - VIRTUAL

     If you specify "VIRTUAL", activation/deactivation of the takeover IPv4 address and activation/deactivation of the logical interface (for example, eth2:1) are performed at the same time. "BASE" will be shown on the menu page. "VIRTUAL" will be shown on "Currently set" and "5) BASE" is shown on the menu page.

     You must activate the IPv4 address on the physical interface (for example, eth2) where the logical interface will be created beforehand because the takeover IPv4 address with this attribute specifies the IPv4 address for the logical interface. To activate the IPv4 address on the physical interface beforehand, make settings so that the IPv4 address is activated on the physical interface at startup of the operating system, or register the takeover IPv4 address with "BASE" attribute with the same takeover network resource.

   - AUTORECOVER, NOT:AUTORECOVER

     If you reactivate the takeover IP address, specify this attribute. The default value is "AUTORECOVER".

     - AUTORECOVER

     If you specify "AUTORECOVER" and the network interface goes down or becomes unreachable due to an error, it will try to activate the takeover IP address only once. "AUTORECOVER" will be shown on "Currently set"and "6) NOT:AUTORECOVER" is shown on the menu page. When the activation of the takeover IP address fails, it will be notified to the cluster.

     - NOT:AUTORECOVER

     If you specify "NOT:AUTORECOVER", the "AUTORECOVER" setting will be disabled. "NOT:AUTORECOVER" will be shown on "Currently set" and "AUTORECOVER" is shown on the menu page."

   - BASIC-IF

     You cannot use this attribute. Do not change.

   - MONITORONLY, NOT:MONITORONLY

     - MONITORONLY

If you specify "MONITORONLY" and the network interface goes down or becomes unreachable due to an error, the error will not be notified to the cluster. "MONITORONLY" will be shown on "Currently set" and "7) NOT:MONITORONLY" is shown on the menu page. If you specify this attribute, a switchover due to a takeover IP address failure will not occur.

- NOT:MONITORONLY

If you specify "NOT:MONITORONLY", the "MONITORONLY" setting will be disabled. "NOT:MONITORONLY" will be shown on "Currently set" and "7) MONITORONLY" is shown on the menu page.

At least one out of all the takeover IP addresses you have registered to the takeover network resources should be set to "NOT:MONITORONLY".

- PING

By setting this attribute, you can specify the previously configured target host for the takeover IP address. Select the target host name to be monitored which you have set in the process of prerequisites.

```
1) HELP
2) RETURN
3) router(000)
4) l3hub
Choese a ping host of the pool ():3
```

4. Change "Timeout" if needed.

   Use "Timeout" to specify the timeout value (in seconds) for the system to wait until all registered takeover IP addresses become active or inactive. When any takeover IP address does not become active or inactive after the timeout value elapses, the resource will notify the error to the cluster. The default value is 60 seconds. Specify the value bigger than 45 seconds.

- NeedAll, InterfaceFilter

  You cannot use these attributes. Do not change.

6.7.3.7 Setting Up Procedure Resources

The procedure resource setting is used to register the state transition procedure resource of the products supported by PRIMECLUSTER in userApplication.

[Prerequisites]

To create a procedure resource, you must first create a state transition procedure and register the procedure to the resource database.

For details, see "D.1 Registering a Procedure Resource."

Operation Procedure:

1. Select "Procedure:XXXXXXXXXX" from "turnkey wizard STANDBY."

   Example of registering cluster resources of the BasicApplication class to a userApplication:

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) -                                10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(-)                  15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:9
```

2. Select "AdditionalProcedureResource" from "Application detail Resource wizard."

   To register multiple resources, execute steps 2 to 4 for each resource.

```
BasicApplication Procedure (ProBApp_APP1:not yet consistent)
1) HELP                              3) REMOVE+EXIT
2) -                                 4) AdditionalProcedureResource
Choose the setting to process:4
```

3. The list of cluster resources will appear. Select one.

Note

If a cluster resource does not appear on this screen, it indicates that the cluster resource has not been registered in the resource database. Confirm whether the cluster resource has been registered on each node of the userApplication, which is designed with "6.7.2 Setting Up userApplication." Register cluster resources if they are not registered. For details on the "clgettree" command, see the manual pages of this command. For details on registering the cluster resource in the resource database, see "D.1 Registering a Procedure Resource."

```
1) HELP
2) RETURN
3) -
4) rscl
Choose the resource: 4
```

4. You can change the following on this screen. If necessary, select "SAVE+RETURN" from "Application detail Resource wizard" after that.

   - Timeout value of the state transition procedure

     The default value is 1,800 seconds. If you use a machine that requires more than 1800 seconds for timeout, you need to change the value by selecting "TIMEOUT."

   - Priority within the resource class

     The default value is specified by -p option when the state transition procedure resource is registered with the "claddprocrsc" command. If the -p option is not specified, 65535 is used as the default value. If you register multiple resources of the same class in the cluster application and specify the order of Online and Offline, change this value by selecting "PRIORITY" from the following screen. The resources will be Online in the ascending order, and will be Offline in the descending order.

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P1)
1) HELP
2) -
3) SAVE+RETURN
4) TIMEOUT
5) PRIORITY
Choose one of the flags:
```

6.7.4 Generate and Activate

This section explains how to execute Generate and Activate. You need to confirm first that the cluster application has been correctly created.

Operation Procedure:

1. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

2. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

Note

Do not execute "Configuration-Activate" simultaneously on multiple nodes which constitute the cluster.

6.7.5 Registering the Cluster Service of a PRIMECLUSTER-compatible product

If the resources registered to a userApplication are for a PRIMECLUSTER-compatible product, register the resources to the cluster service according to the procedure described below.

Operation Procedure

1. Register the cluster service of the PRIMECLUSTER-compatible product.
   Execute the following command in a node that is part of the cluster system.
   This step is not necessary if PRIMECLUSTER Wizard for NAS is used.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

Note

- Make sure to register the cluster service for the PRIMECLUSTER-compatible products other than PRIMECLUSTER Wizard for NAS. Otherwise, they do not operate correctly.
- To find out the PRIMECLUSTER-compatible products, see "Appendix A PRIMECLUSTER Products."

6.7.6 Attributes

See

For information on the attributes, see "12 Appendix - Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

6.7.7 Exclusive Relationships Between Cluster Applications

If exclusive relationships for the cluster applications are set up, the cluster applications with exclusive relationships can be prevented from operating simultaneously in one cluster node. 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.

Exclusive control is established between cluster applications within a single group.

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.
- For information on setting up an exclusive relationship, see "6.7.2.1 Creating Standby Cluster Applications."
- The cluster application in which the exclusive relationship is set transits to the Standby state according to the StandbyTransitions attribute.

Note

When the cluster application state is Faulted on a node, cluster applications in exclusive relationships on that node cannot be made operational by newly starting the cluster applications. Cluster applications started later will be stopped regardless of job priority.

The reason for this is that possibly not all resources under the control of the cluster application in the Faulted state could be stopped.

In such a case, clear the Faulted state of the cluster application to bring it to the Offline state, and then start the cluster applications that are in exclusive relationships.

For information on how to clear the Faulted state of 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

The example below explains the exclusive relationship between cluster applications within a single group. Even when there are multiple groups, the operation within a single group is the same.

When the job priorities of the cluster applications with an exclusive relationship are different

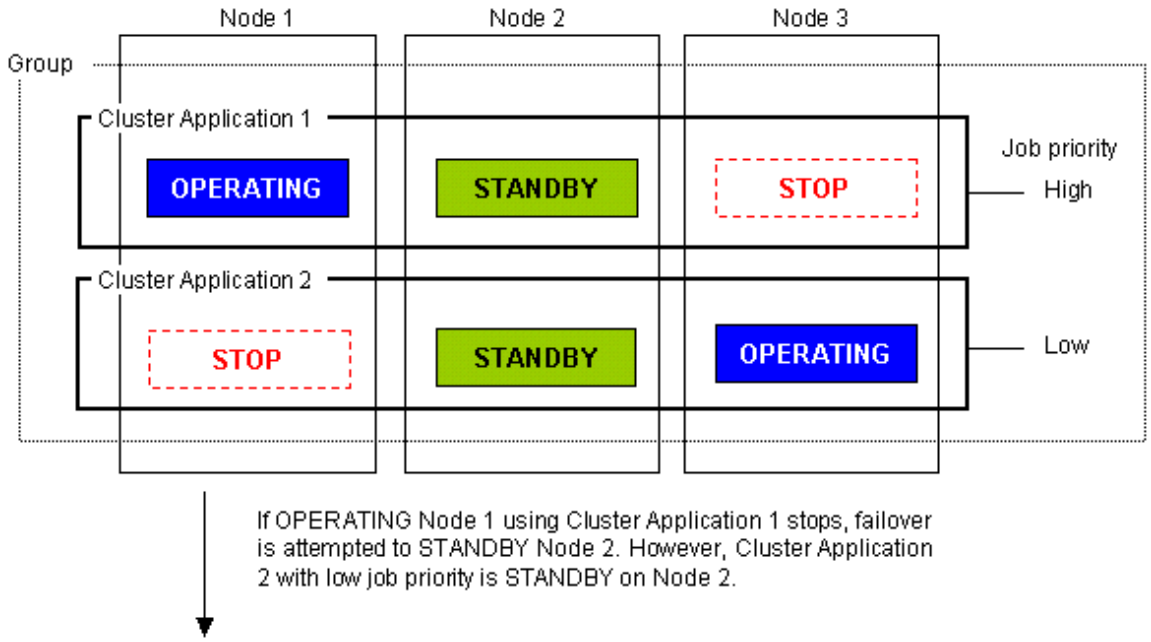
Cluster applications with the highest job priority take the top priority for startup on the nodes on which the cluster applications with high job priority are running or on the nodes to which the cluster applications with high job priority are failed over. Therefore, cluster applications running with low priorities will be forcibly exited.

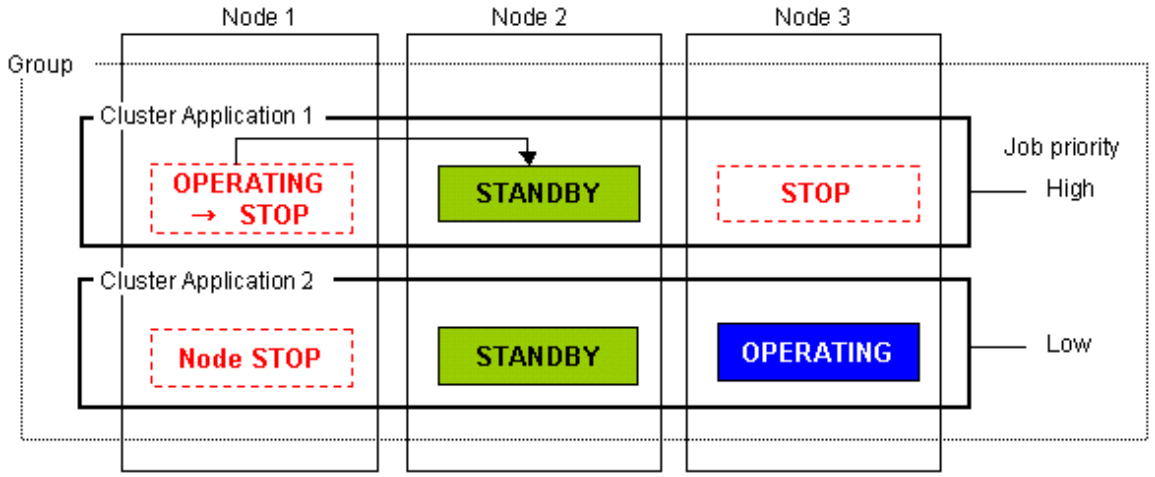
The states indicated in the following figure are as follows:

- OPERATING: The cluster application is in the Online state.
- STANDBY: The cluster application is in the Standby state.
- STOP: The cluster application is in the Offline state.

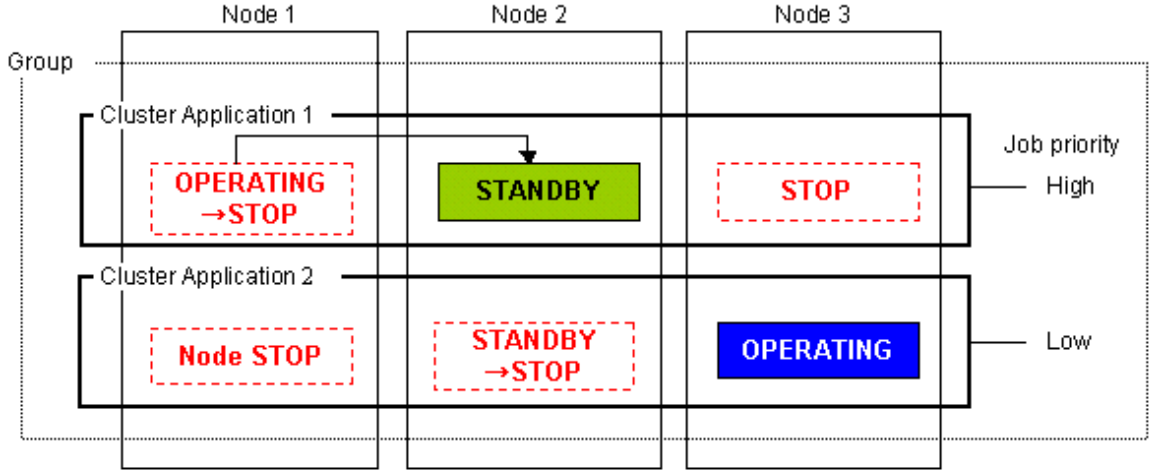
Failover of the cluster application with a high job priority

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.

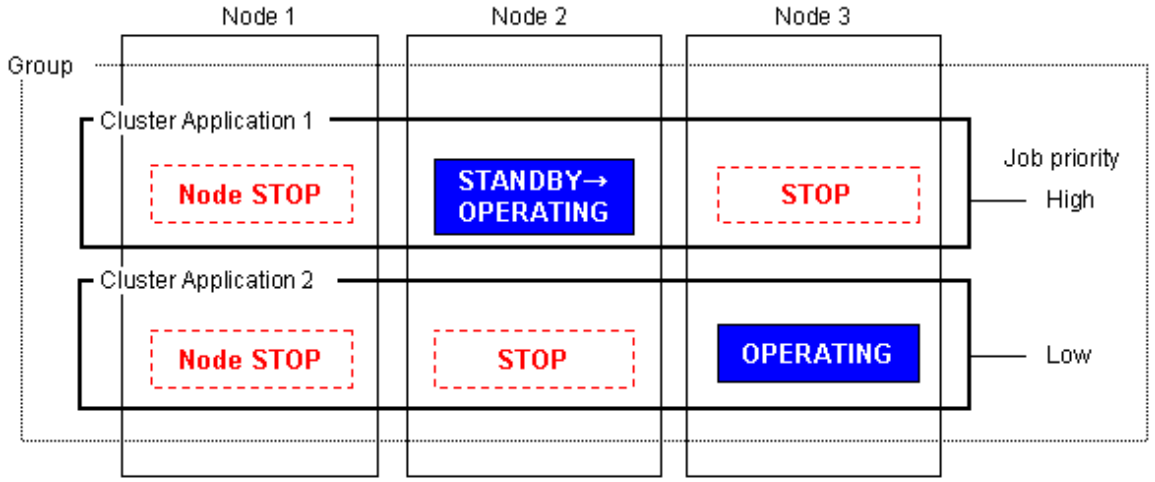




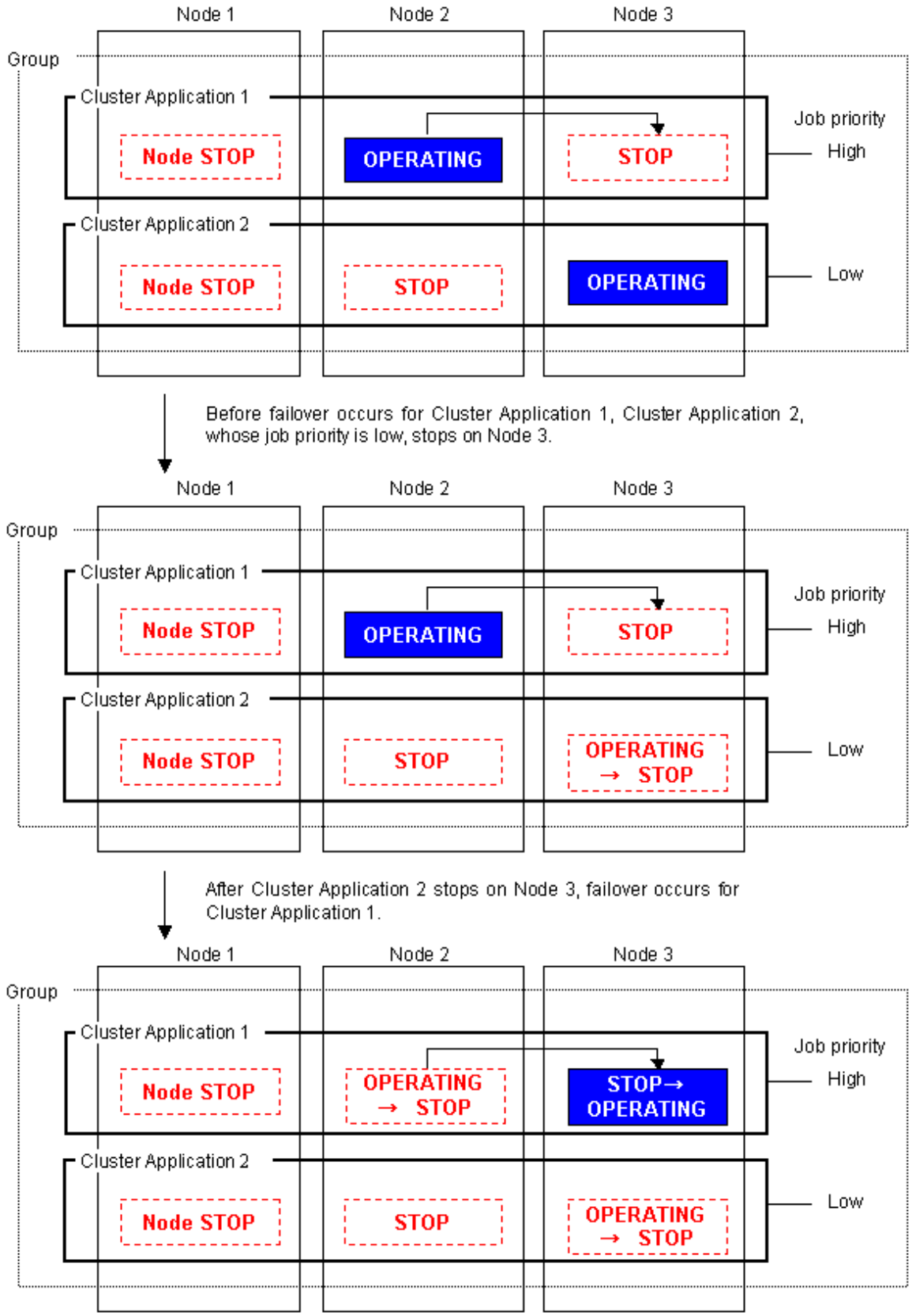
Before failover occurs for Cluster Application 1 whose job priority is high, Cluster Application 2, whose job priority is low, stops.



After Cluster Application 2 stops on Node 2, failover occurs for Cluster Application 1 whose job priority is high.

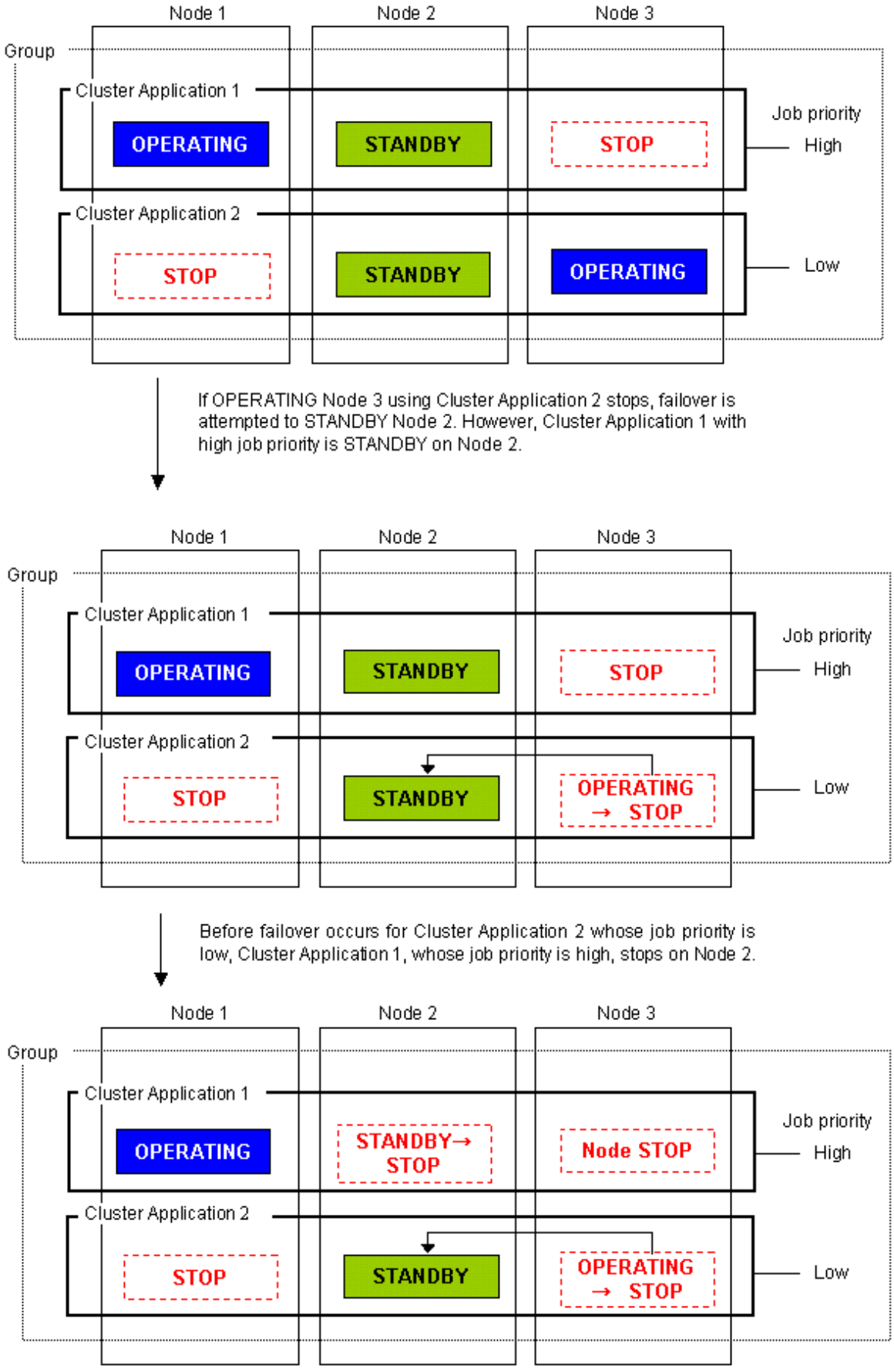


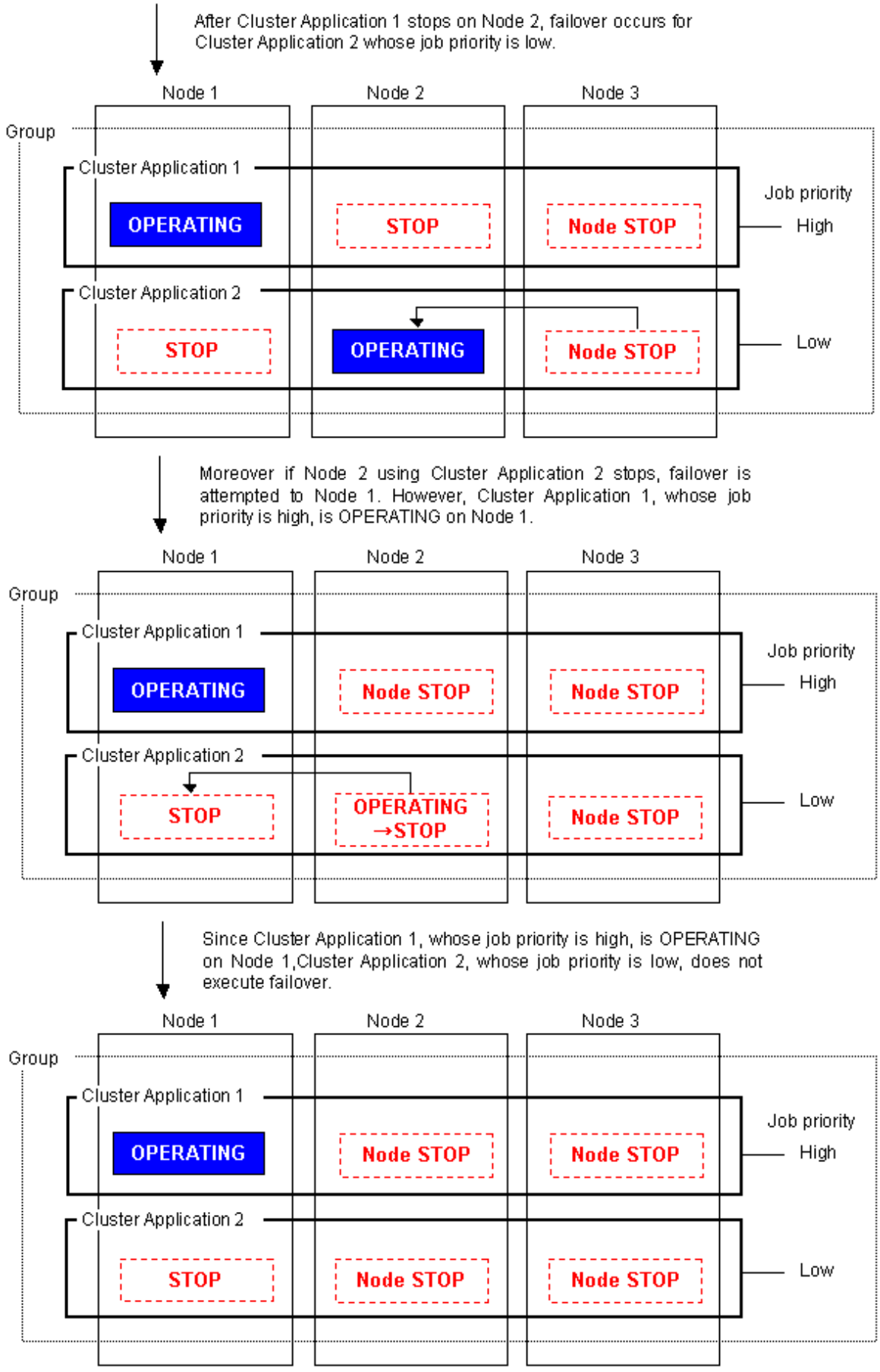
Moreover if Node 2 using Cluster Application 1 stops, failover is attempted to Node 3. However, Cluster Application 2 is OPERATING on Node 3.



Failover of the cluster application with a low job priority

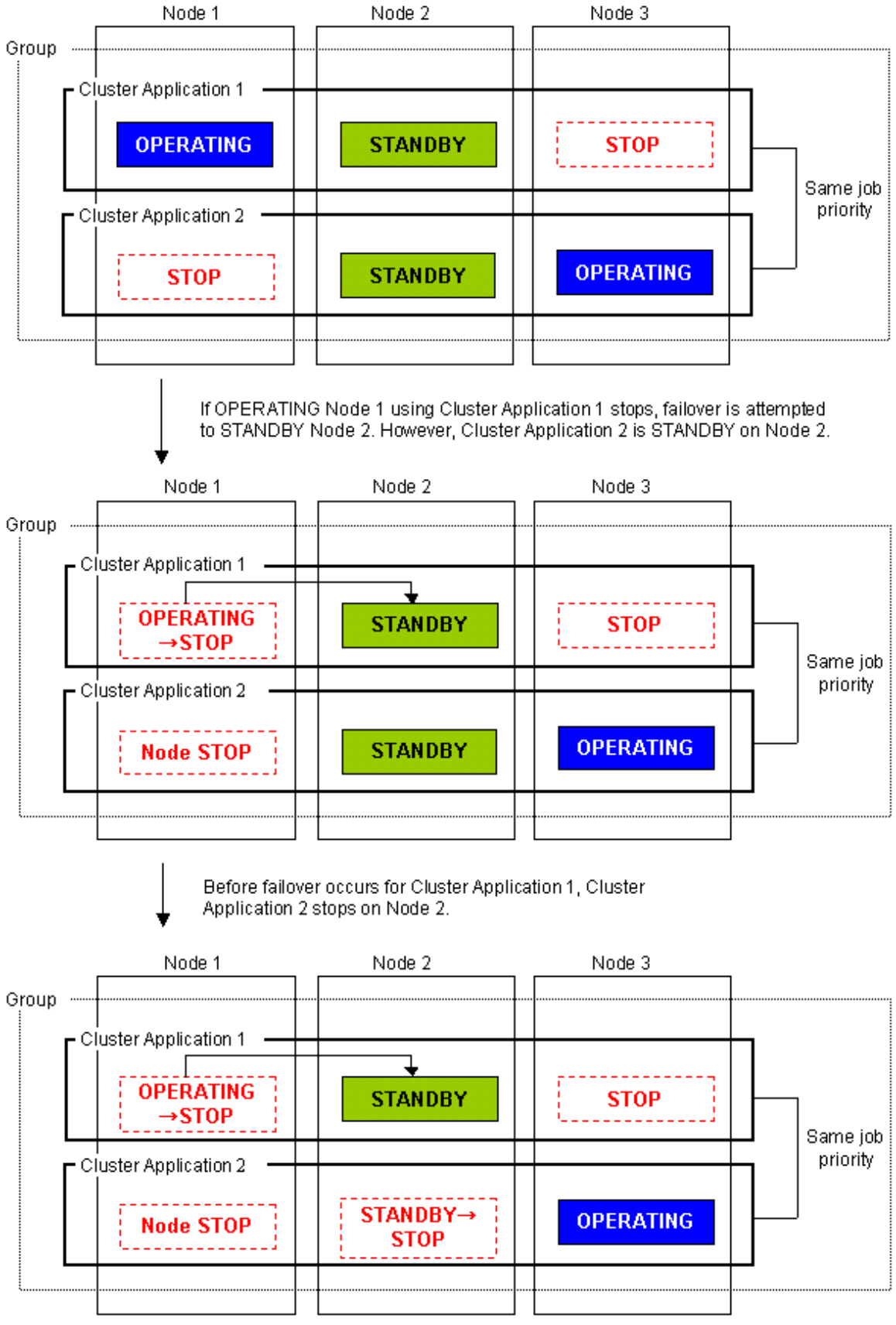
Failover occurs for a cluster application with a low job priority only when there is no cluster application with a high job priority included on the node to which the cluster application with a low job priority is to be failed over.

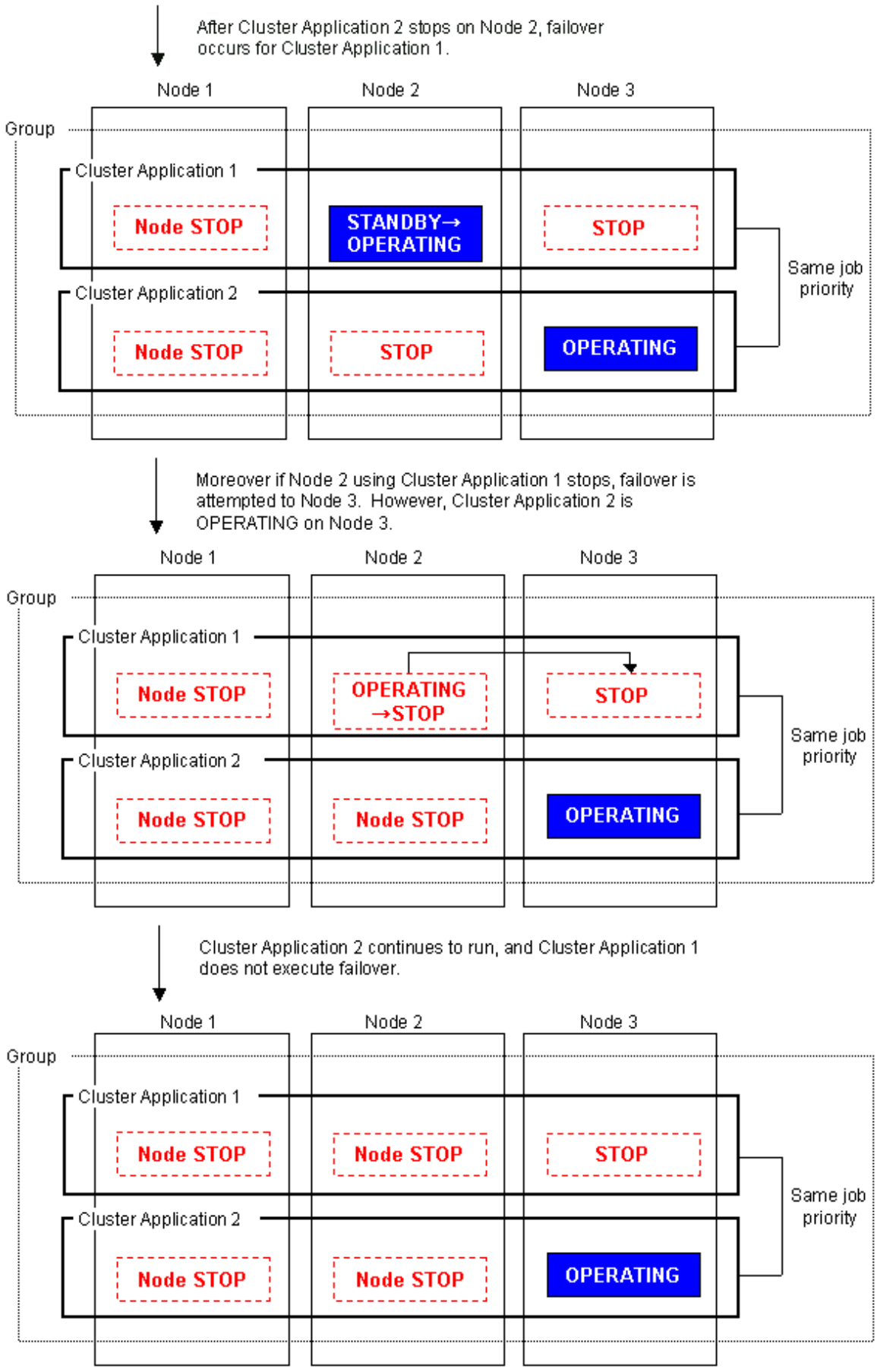

Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
STANDBY
STOP
Job priority
High
Cluster Application 2
STOP
STANDBY
OPERATING
Low
If OPERATING Node 3 using Cluster Application 2 stops, failover is attempted to STANDBY Node 2. However, Cluster Application 1 with high job priority is STANDBY on Node 2.
Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
STANDBY
STOP
Job priority
High
Cluster Application 2
STOP
STANDBY
OPERATING → STOP
Low
Before failover occurs for Cluster Application 2 whose job priority is low, Cluster Application 1, whose job priority is high, stops on Node 2.
Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
STANDBY→ STOP
Node STOP
Job priority
High
Cluster Application 2
STOP
STANDBY
OPERATING → STOP
Low


After Cluster Application 1 stops on Node 2, failover occurs for Cluster Application 2 whose job priority is low.
Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
STOP
Node STOP
Job priority
High
Cluster Application 2
STOP
OPERATING
Node STOP
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.
Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
Node STOP
Node STOP
Job priority
High
Cluster Application 2
STOP
OPERATING
→STOP
Node STOP
Low
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.
Node 1
Node 2
Node 3
Group
Cluster Application 1
OPERATING
Node STOP
Node STOP
Job priority
High
Cluster Application 2
STOP
Node STOP
Node STOP
Low

When the job priorities of cluster applications with an exclusive relationship are the same

The operation of the cluster applications that are already running will be continued. On the node on which cluster applications are already running, cluster applications that subsequently start up will be stopped.




After Cluster Application 2 stops on Node 2, failover occurs for Cluster Application 1.
Node 1
Node 2
Node 3
Group
Cluster Application 1
Node STOP
STANDBY→ OPERATING
STOP
Cluster Application 2
Node STOP
STOP
OPERATING
Same job priority
Moreover if Node 2 using Cluster Application 1 stops, failover is attempted to Node 3. However, Cluster Application 2 is OPERATING on Node 3.
Node 1
Node 2
Node 3
Group
Cluster Application 1
Node STOP
OPERATING →STOP
STOP
Cluster Application 2
Node STOP
Node STOP
OPERATING
Same job priority
Cluster Application 2 continues to run, and Cluster Application 1 does not execute failover.
Node 1
Node 2
Node 3
Group
Cluster Application 1
Node STOP
Node STOP
STOP
Cluster Application 2
Node STOP
Node STOP
OPERATING
Same job priority

6.8 Setting Up the RMS Environment

When using RMS, you need to check "Setup (initial configuration)"of PRIMECLUSTER Designsheets 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 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 readme under Tool/clchkcluster on DVD of the PRIMECLUSTER product package.

6.10 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.

6.10.1 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

1. 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.

2. 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		AutoStartUp=no	
		StandbyTransitions		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 exists 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 x ("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. Changing the setting with the hvw command

   Follow the procedure below:

   1. Start up RMS Wizard with the hvw command.
   2. Select "Application-Create" from "Main configuration menu."

```
sweetpea: Main configuration menu, current configuration: aaa
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: 3
```

3. Select "Controller" from "Application selection menu."

```
Edit: Application selection menu (restricted):
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) APP1
 6) Cmd_APP1
 7) Controller
 8) app1
Application Name: 7
```

4. Select "Controllers" from "Settings of application type."

```
Settings of application type "Controller" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=Controller
 5) ControlPolicy=SCALABLE
 6) AdditionalAppToControl
 7) Controllers[0]=T180:app1
 8) (FaultScript=)
 9) (ApplicationSequence=)
10) (StateChangeScript=)
Choose the setting to process: 7
```

5. Select "SELECTED."

```
 1) HELP
 2) RETURN
 3) NONE
 4) FREECHOICE
 5) SELECTED:app1
Set the application to control: 5
```

6. Select "TIMEOUT(T)" from "Set *global* flags for all scalable (sub) applications."

```
Set *global* flags for all scalable (sub) applications: app1
Currently set: TIMEOUT (T180)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) DEFAULT
 5) MONITORONLY(M)
 6) TIMEOUT(T)
Choose one of the flags: 6
```

7. Select "FREECHOICE" and enter the setting value (when entering 2340).

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 180
Set an appropriate timeout: 3
        >> 2340
```

8. Select "SAVE+RETURN" from "Set *global* flags for all scalable (sub) applications."

```
Set *global* flags for all scalable (sub) applications: app1
Currently set: TIMEOUT (T2340)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) DEFAULT
 5) MONITORONLY(M)
 6) TIMEOUT(T)
Choose one of the flags: 3
```

9. Select "SAVE+EXIT" from "Settings of application type."

```
Settings of application type "Controller" (consistent)
 1) HELP
 2) NO-SAVE+EXIT
 3) SAVE+EXIT
 4) ApplicationName=Controller
 5) ControlPolicy=SCALABLE
 6) AdditionalAppToControl
 7) Controllers[0]=T2340:app1
 8) (FaultScript=)
 9) (ApplicationSequence=)
10) (StateChangeScript=)
Choose the setting to process: 3
```

See

For detailed operation on how to change RMS Wizard and attributes, see "10.3 Changing a Cluster Application" or "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

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

Note that 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 "0." First, start the RMS Wizard.

```
node1# /opt/SMAW/SMAWRrms/bin/hvw -n testconf
```

Note

Change testconf based on your environment.

For details, see "11.1 Changing the Operation Attributes of a userApplication."

1. Select "Application-Edit" from "Main configuration menu."
2. Select the appropriate cluster application (userApplication) to change its configuration in "Application selection menu."
3. Select "Machines+Basics" in "turnkey wizard."
4. Select "ShutdownPriority."
5. Select "FREECHOICE" to enter 0.
6. Select "SAVE+EXIT" in "Machines+Basics."
7. Select "SAVE+EXIT" in "turnkey wizard."
8. Select "RETURN" on "Application selection menu."
9. Select "Configuration-Generate."
10. Select "Configuration-Activate."

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

[PRIMERGY]

Panic the operational node with the NMI switch or keyboard operation in the main device.

[PRIMEQUEST]

Collect dumps of the operational node with Web-UI to stop it.

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 use SNMP manager to monitor cluster system

If any error occurs in the resources registered in the userApplication of a cluster, SNMP Trap will be sent to the server which SNMP manager runs on, thus the cluster system will be able to be monitored.

See

For details of this function, see "14.12 SNMP Notification of Resource Failure" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Set the FaultScript attribute of userApplication to "To be specified by the hvsnmptrapsend command" as follows.

Prechecking

Check if the net-snmp-utils package provided by the OS has been installed on all the nodes of the cluster which uses this function. If it has not been installed, you need to install it.

Example

```
# rpm -q net-snmp-utils
net-snmp-utils-5.5-41.el6.i686
```

Confirm that the SNMP manager supports version 2c of SNMP in the SNMP Trap destination. Moreover, check the community names that the SNMP manager can receive beforehand.

Setup procedure

1. Start up RMS Wizard with the hvw command.
2. Select "(FaultScript=)" from the "Machines+Basics" menu of the userApplication which monitors resource errors.

```
Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji2RMS
 8) Machines[1]=fuji3RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (AutoBreakMaintMode=no)
25) (HaltFlag=yes)
26) (PartialCluster=0)
27) (ScriptTimeout=)
Choose the setting to process: 14
```

See

For information on how to set up userApplication with the RMS Wizard, see "6.7.2.1 Creating Standby Cluster Applications" and "10.3 Changing a Cluster Application."

3. Select "FREECHOICE" and execute the following command.

```
/opt/SMAW/bin/hvsnmptrapsend <community> <host>
```

```
<community> Specify the SNMP community.
<host>      Specify the destination of SNMP trap.
```

```
 1) HELP
 2) RETURN
 3) NONE
 4) FREECHOICE
```

```
Enter the command line to start upon fault processing: 4
       >> /opt/SMAW/bin/hvsnmptrapsend community snmprvhost
```

Note

When the Fault script has been registered already, create a new script for executing both the Fault script command and the hvsnmptrapsend command, and register this script in the Fault script.

4. Confirm that "FaultScript" of the "Machines+Basics" menu has been set.

```
Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji2RMS
 8) Machines[1]=fuji3RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript='/opt/SMAW/bin/hvsnmptrapsend~community~snmprvhost')
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (AutoBreakMaintMode=no)
25) (HaltFlag=yes)
26) (PartialCluster=0)
27) (ScriptTimeout=)
Choose the setting to process:
```

5. See "6.7.4 Generate and Activate" and execute the "Configuration-Generate" and "Configuration-Activate" processes.

6.10.2 Notes on Configuration

Do not use reserved words for cluster application names

If you use a reserved word for a cluster application name, RMS cannot be configured properly.

Do not use the following reserved words in addition to numbers and types of characters limited in PRIMECLUSTER Installation and Administration Guide.

<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
```

6.11 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.

The following table indicates attributes can be set to the Cmdline resources.

Table 6.2 Attributes of the Cmdline resource

Attribute	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

Attribute	Outline
	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 attribute is set to "No," the exit code of the Check script is interpreted as follows: 0: Online Other than 0: Offline If the attribute is set to "Yes," the exit code is interpreted as follows. 0: online 1: offline 2: faulted 3: unknown 4: standby 5: onlinewarning 6: offlinefaulted The default value is "No." * Do not use them as return values within the script because the values displayed in 3, 5, 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 one 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 attempt is performed just once. The default value is "Yes."
MONITORONLY	This attribute controls whether the status of userApplication becomes Faulted or not when the resource becomes Faulted. If this attribute is set to "Yes," userApplication does not become Faulted even if the resource becomes Faulted. Set "No" to at least one Cmdline resource that is registered in userApplication. The default value is "No."
STANDBYCAPABLE	If the attribute is set to "Yes," RMS sets the StandbyCapable attribute to "1" for this resource. For detailed information regarding this attribute, see "12 Appendix - Attributes" of "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide." The default value is "No."

Attribute	Outline
REALTIME	If the attribute is set to "No," the Check script is started in the TS class. If the attribute is set to "Yes," the Check script is started in the RT class. Note that the highest priority is assigned to the process started in the RT class in the operating system. Thus, the bugs of the script or commands may give a large effect on system performance. The default value is "No."
TIMEOUT	This attribute sets a timeout interval (seconds) to start and stop programs. The default value is "300."

Note

When PRIMECLUSTER products are not specified, do not change ReturnCodes of the Cmdline resource.

6.11.1 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
	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

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->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
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

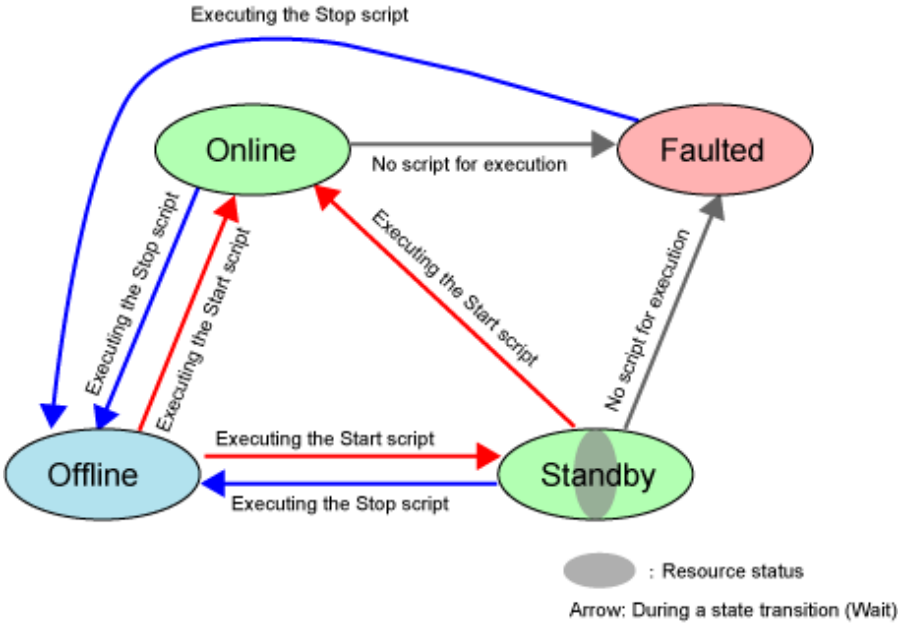
For the environment variable that can be referred to within a script, see "6.11.2.1.2 Environment Variables can be referred to within the Start and Stop Scripts."

6.11.1.1 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 "6.11.1.5 Flow of the Cmdline Resource Operation."

Figure 6.1 Scripts to be executed during a state transition

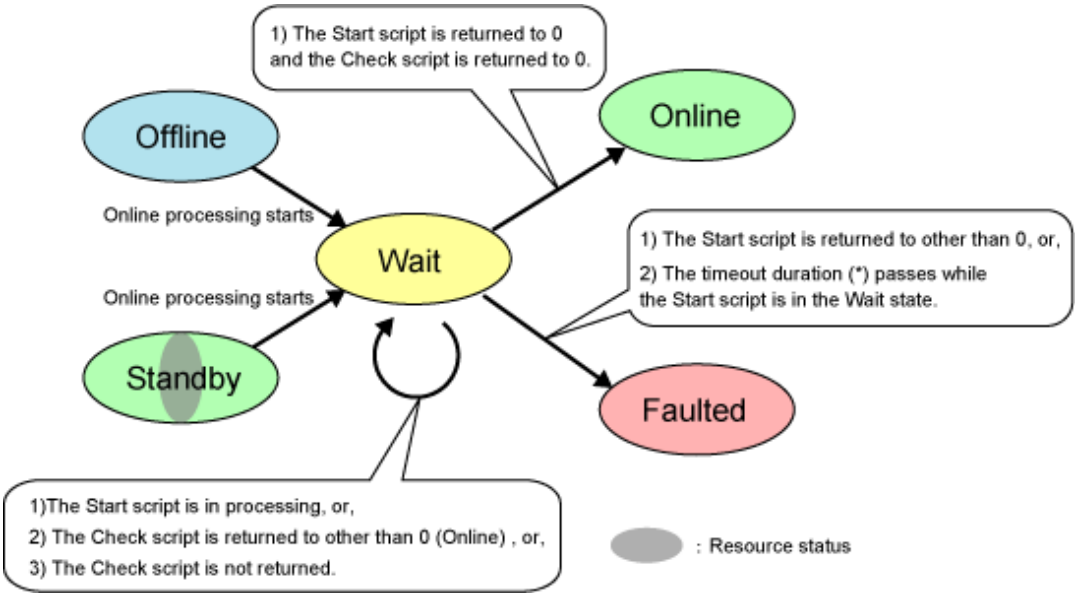


6.11.1.2 Script States When Online

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 "6.11.1.5 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 "6.11.2.2.3 Check Script Exit Code."

Figure 6.2 Scripts states when Online



(*) For a timeout, see "6.11.2.1.5 Timeout of Scripts."

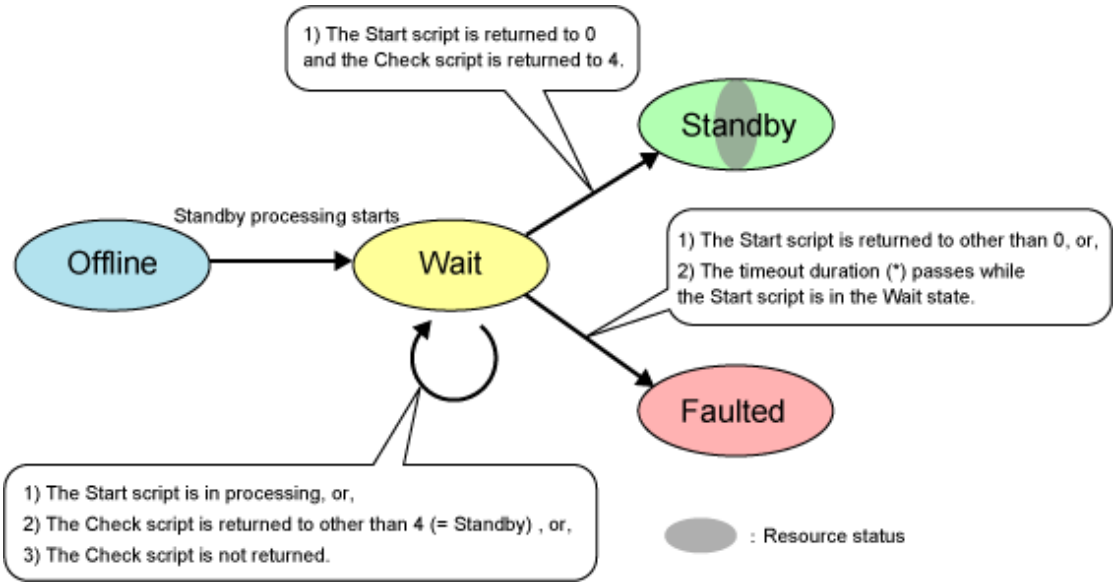
6.11.1.3 Script States When Standby

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 "6.11.1.5 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 "6.11.2.2.3 Check Script Exit Code."

Figure 6.3 Script states when Standby



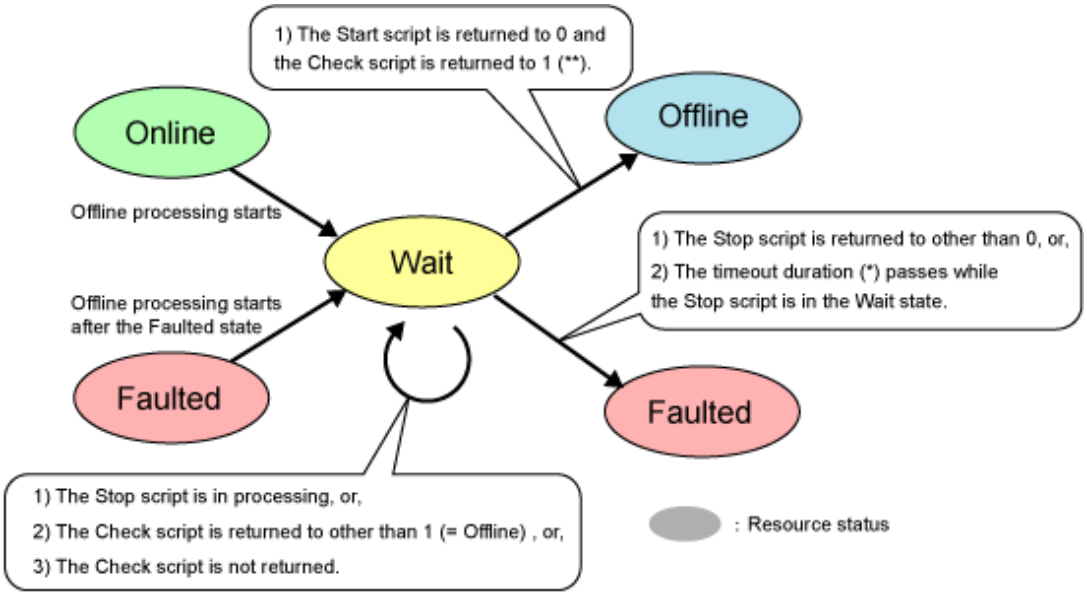
(*) For a timeout, see "6.11.2.1.5 Timeout of Scripts."

6.11.1.4 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 "6.11.1.5 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 "6.11.2.2.3 Check Script Exit Code."

Figure 6.4 Script states when Offline



(*) For a timeout, see "6.11.2.1.5 Timeout of Scripts."

(**) It is when ALLEXITCODES is set. For details, see the Outline of the ALLEXITCODES attribute in Table 3.1.

6.11.1.5 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

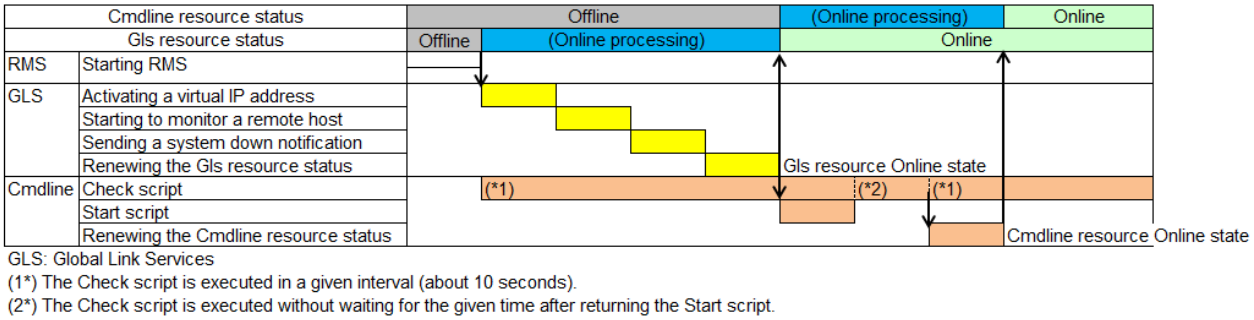
    The Check script is executed in a given interval (about 10 seconds) after starting RMS. After that, the Start script is executed. Then, the Check script is executed without waiting for the given time after returning the Start script. After the Start script is normally returned and the Check script is 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.

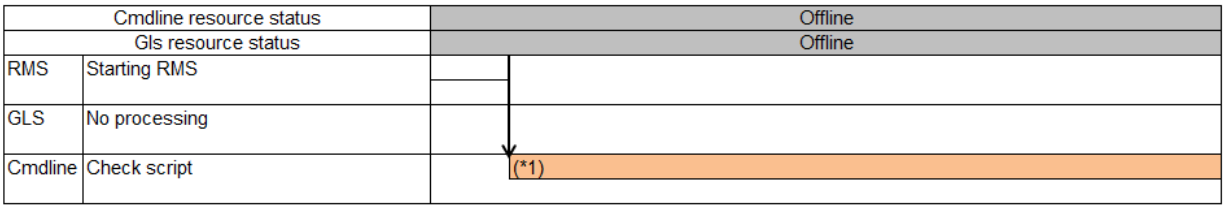
- At RMS startup Standby system (Offline->Offline)

  - 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

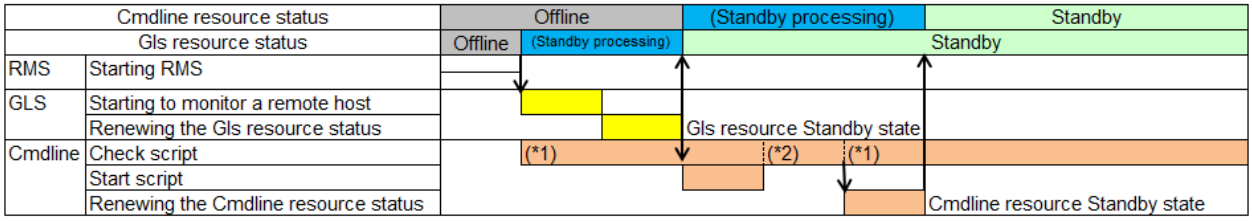
    The Check script is executed in a given interval (about 10 seconds) after starting RMS. After that, the Start script is executed. Then, the Check script is executed without waiting for the given time after returning the Start script. After the Start script is normally returned and the Check script is 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.

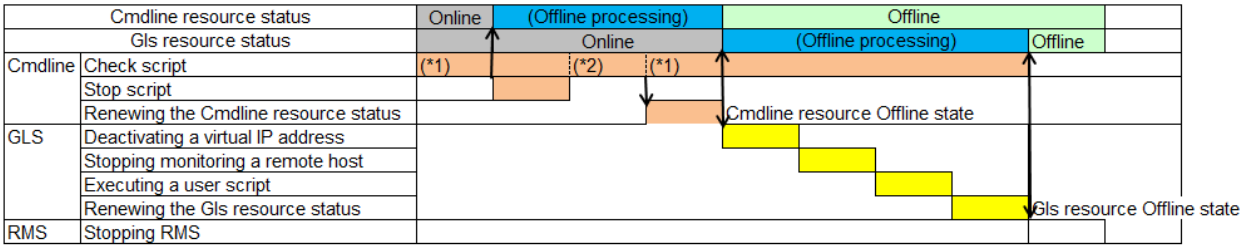
- At RMS stop Operational system (Online->Offline)

  - 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 Stop script.

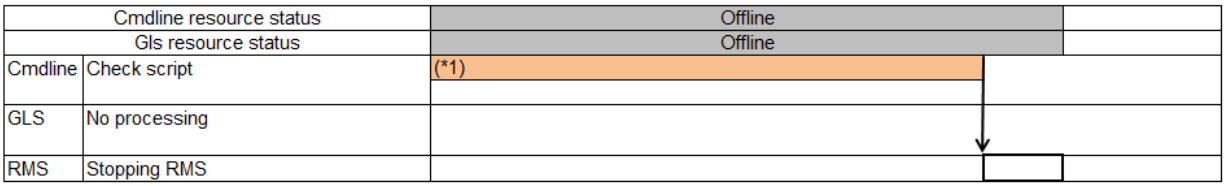
- At RMS stop Standby system (Offline->Offline)

  - 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).

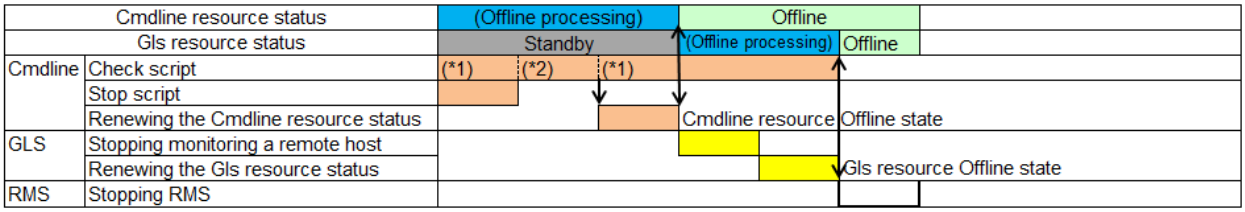
- At RMS stop Standby system (Standby->Offline)

  - 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.

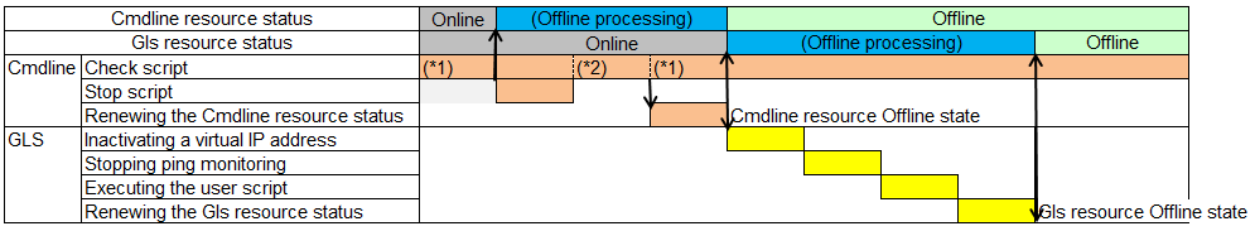
- At switchover Operational system (Online->Offline)

  - 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.

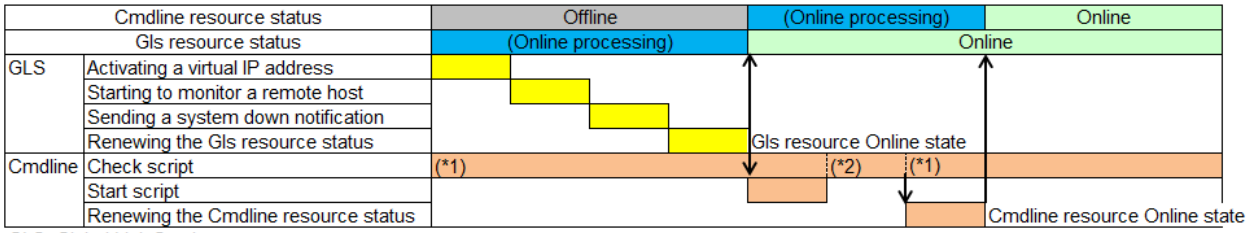
- At switchover Standby system (Offline->Online)

  - 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.

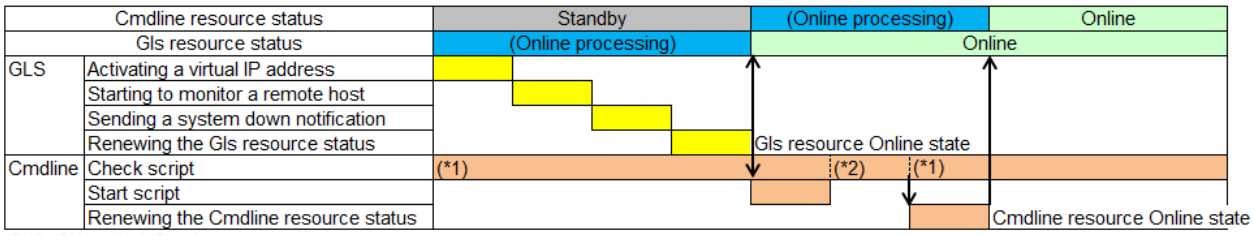
- At switchover Standby system (Standby->Online)

  - 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.

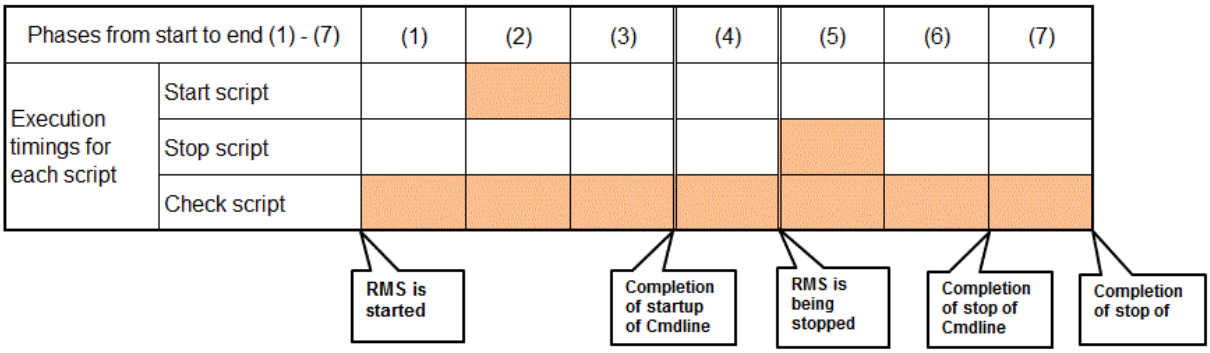
6.11.1.6 Operation for Each Exit Code of the Check Script

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

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	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

6.11.2 Notes When Creating Scripts

This section describes notes when creating scripts.

6.11.2.1 start and stop Scripts

6.11.2.1.1 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 6.5 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 6.6 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
```

6.11.2.1.2 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 "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

6.11.2.1.3 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.

6.11.2.1.4 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, the Online script of Cmdline resource where the NULLDETECTOR flag is set is executed even if the resource is started at the termination of maintenance mode.

6.11.2.1.5 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.

6.11.2.2 Check Script

6.11.2.2.1 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 6.11.2.1.1 Examples of start and stop Scripts."

Figure 6.7 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 6.11.2.1.1 Examples of start and stop Scripts."

Figure 6.8 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
```

6.11.2.2.2 Environment Variables that can be referred to within the Check Scripts

The following environment variables are set when executing the Check script. These environment variables can be referred to within the script.

- HV_APPLICATION
- HV_NODENAME

See

For outlines on these environment variables, see "Table 6.5 Environment variables that can be referred to within the Start and Stop scripts."

And, 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."

6.11.2.2.3 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.
4	Standby state. If a target to be monitored has been Standby state, use this exit code.

Note

Since the exit codes other than the above indicate the specific status, use these codes only when applicable products are specified in the environment that uses PRIMECLUSTER products.

6.11.2.2.4 Timeout of Check Script

If the cluster application needs to be switched due to a resource error when the Check script processing is not completed within the specified time, set the CheckCommandTimeouts attribute of the Cmdline resource. Specify the time to determine a resource error if the Check script processing is not completed in seconds. By default, "none" is set, which means that a resource error does not occur.

The settable value is "none" or in the range of 45 seconds to 3600 seconds.

6.11.3 Notes on Scripts

- The execute permission for each script is user: root and group: root.
- Environment variables set in each server ("/etc/profile" or "etc/bashrc", for example) 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.
- When multiple Cmdlines are registered in userApplication, it is performed in the order of registering Cmdline when starting userApplication. On the other hand, when stopping it, it is performed in the opposite order of registering Cmdline. The example is as follows.
  The resource registered first is Command[0], the resource registered next is Command[1].
  Those resources are started and stopped in the following order.

  At startup

  ```
  StartCommands[0]
  StartCommands[1]
  ```

  At stop

  ```
  StopCommands[1]
  StopCommands[0]
  ```

- 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.

6.12 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)

6.12.1 Monitoring Fsystem

Before describing notes on the Fsystem resource, this section describes the Fsystem detector.
The Fsystem detector is as follows:

- 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 is set in line with the definition of /etc/fstab.pcl.
  - I/O to the file system is performed properly while it is mounted.

When creating an Fsystem resource, the following file is temporarily created to verify access to a mount point.

<mount-point>/.TMPRMSMOUNTPOINTPROBE

This file must not be accessed as a target for backups or the like.

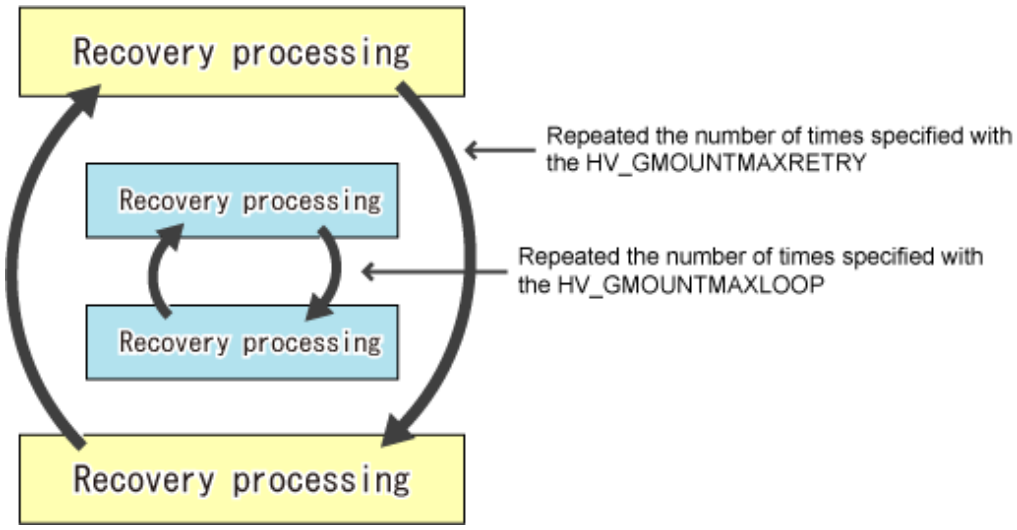
6.12.2 Fsystem Resource Attribute

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."

"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

6.12.3 File System on the Shared Disk Device

According to the type of file system on the shared disk device, perform the following settings and corrective actions.

See

The type of file system that can be used on the shared disk device varies depending on the OS. For details on the file system and notes on use, see "Linux user guide" of each OS.

6.12.3.1 Corrective Actions for the Forced File System Check

If ext3 or ext4 is used for a file system, the file system might forcibly be checked during Online processing of a switching file system. It is part of the ext3 and ext4 specification that file systems are checked when a certain number of mounting has been executed since the last file system check, or a certain period of time has passed.

When the file systems are forcibly checked along with node startup or failover, timeout occurs due to file system Online processing, and PRIMECLUSTER startup or failover might fail.

It is necessary to prevent the file systems from being checked by executing the following command for all the ext3 and ext4 switching files.

```
# tune2fs -c0 -i0 <device_name>
```

Example

```
# tune2fs -c0 -i0 /dev/sdi1
```

After executing the above command, check if "Maximum mount count :-1", "Check interval:0"is displayed using the following command:

```
# tune2fs -l /dev/sdi1
[snip]
Mount count: 10
Maximum mount count: -1
[snip]
Check interval: 0 (<none>)
[snip]
```

Note

If the forcible file system check is prevented, file systems might corrupt due to failures such as disk errors and kernel bug. These failures cannot be detected through file system logging and journaling. The file system corruption might cause data corruption. To prevent this, execute the "fsck - f" command to enable the file system forcible check during periodic maintenance.

6.12.3.2 Corrective Actions for delayed allocation

If ext4 or xfs is used for a file system, take the following corrective actions for delayed allocation.

See

Ext4 and xfs are used to make the allocation of the disk area more efficient, and to improve the writing performance, using their "Delayed Allocation" feature. As a result of the implementation of "Delayed Allocation", there is a possibility that a part of data is lost by OS panic or power supply interruption of servers, because the sojourn time on the memory of data that should be stored on the disk becomes longer.

For the details of delayed allocation, see Storage Administration Guide of the Red Hat, Inc.

- ext4

  The delayed allocation can be set disable by specifying nodelalloc for mount option in ext4. Specify the mount option of /etc/fstab.pcl file as follows.

```
#RMS#/dev/sdd2        /mnt/swdsk2          ext4     nodelalloc      0 0
```

- xfs

  The delayed allocation cannot be set disable when xfs is used. Therefore, in order to prevent a part of data not be lost by OS panic or power supply interruption of servers, the application should immediately issue the fsync() call after writing to guarantee writing in the file system.

6.12.4 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.

6.12.5 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 the 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

   ```
   # /bin/df -k
   Filesystem           1K-blocks      Used Available Use% Mounted on
   /dev/sda2             20315844   7474340  11792864  39% /
   /dev/sda1               256666     25466    217948  11% /boot
   tmpfs                   971664         0    971664   0% /dev/shm
   ```

   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

    1. Restoring the file system (only if necessary)

        To restore the file system, use the fsck command. If the file system to be maintained exists in the volume controlled by GDS, execute the fsck command on the node that the GDS volume has started in Step 3.

        See

        For how to restore the file system with the fsck command or e2fsck command, see the Online manual page for Linux (man fsck or man e2fsck).

    2. Mounting the file system (only if necessary)

        Mount the file system with the mount command.

        The device name of the file system controlled by the Fsystem resource has been described in the /etc/fstab.pcl file. Refer to the /etc/fstab.pcl file to mount the file system.

        Example: Checking the contents of the /etc/fstab.pcl file with the cat command

        ```
        # /bin/cat /etc/fstab.pcl
        #RMS#/dev/sfdsk/class0001/dsk/volume0001 /mnt/swdsk1 ext3 noauto 0 0
        ```

        Example: Mounting the file system of the mountpoint /mnt/swdsk1 controlled by the Fsystem resource

        ```
        # /bin/mount -t ext3 /dev/sfdsk/class0001/dsk/volume0001 /mnt/swdsk1
        ```

    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: Unmounting the file system mounted in /mnt/swdsk1

        ```
        # /bin/umount /mnt/swdsk1
        ```

5. Stopping the GDS volume

    Stop the GDS volume started in Step 3.

    Example: 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 the nodes

    Start RMS on all cluster nodes.

    Example: Starting RMS on all the nodes configuring a cluster from any one of nodes with a command

    ```
    # /opt/SMAW/SMAWRrms/bin/hvcm -a
    ```

Part 3 Operations

Chapter 7 Operations..... 249

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
Referring 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

  The states of the resources that are managed with the cluster resource management facility are displayed.

- RMS main window

  Use this screen to monitor the state of the cluster system and to manage cluster applications and resources.

- MSG main window

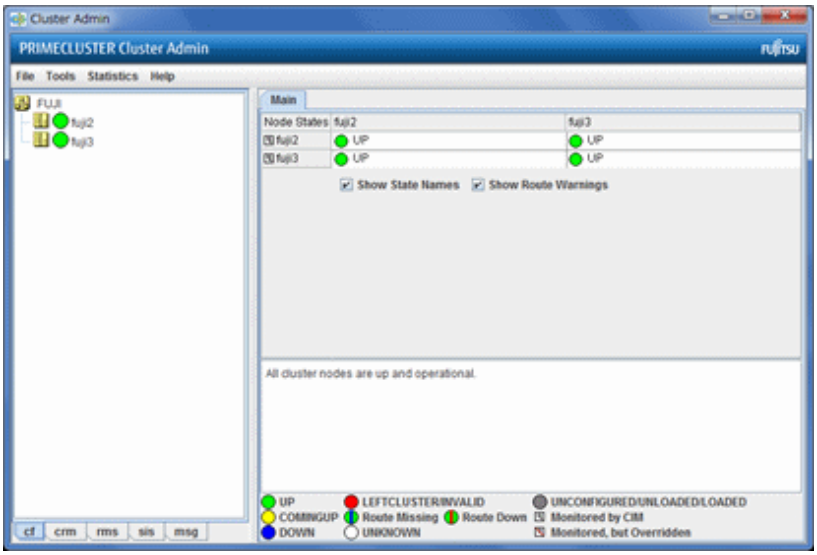
  This screen displays messages for cluster control.

See

For instructions on displaying each screen, see "4.5.3 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 "Chapter 4 GUI administration" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

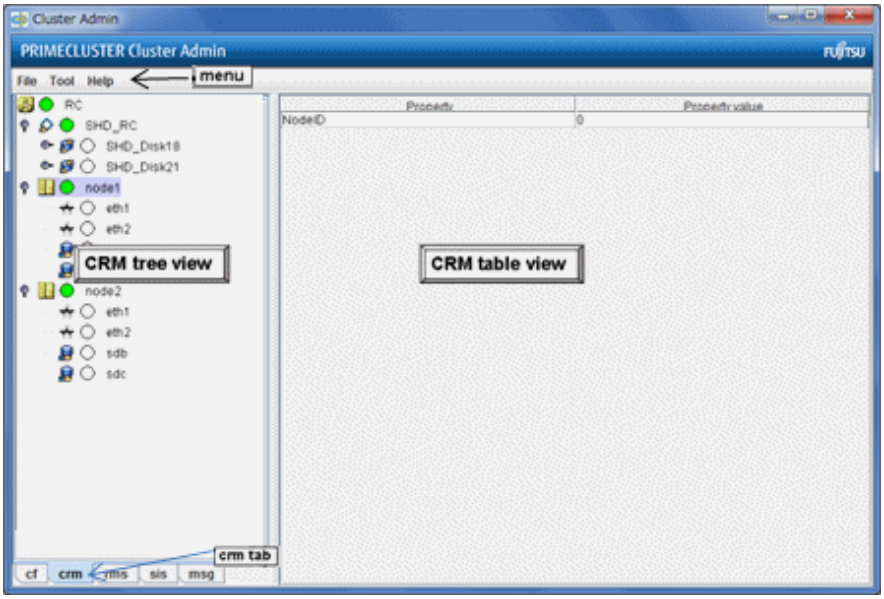
Note

The node states may be displayed as Unknown. In this case, exit the Web-Based Admin View screen and restart. If the node states in the CF main window of Cluster Admin are still displayed as Unknown, check the node states by using cftool -n.

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 "4.5.3 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
	Shared disk device
	IP address Network interface
	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 do not need to take any corrective actions.

Cluster states

The following cluster states are displayed.

Icon	Icon color	Outline	Details
	Green	ON	All the 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 node is LEFTCLUSTER in CF, it becomes ON in the CRM tree view.
- After you execute the "clinitreset" command to initialize the resource database, the resource states displayed in the CRM main window differ according to the actual resource states. Execute "clinitreset", restart the nodes, and then close the browser and redisplay the screen. Execute the "clinitreset" 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 on 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
Request Resource activation	*Tool* - *Start*	SDX_DC (*2)	wvroot clroot cladmin
Request Resource deactivation	*Tool* - *Stop*	SDX_DC (*2)	wvroot clroot cladmin
Exit Cluster Admin screen	*File* - *Exit*	All No selection	All
View Help	*Help* - *Content* (*3)	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 Only the disk resources that are registered to Global Disk Services are enabled.

*3 Help for the CRM main window is displayed with a separate browser from the browser that displays Help for CF, RMS, and SIS.

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 "4.3.1 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."

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.

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.

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).
 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.
 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

Icon/ resource class name	Attributes	Meaning/attribute value (Top: Meaning, Bottom: Attribute value)
		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.
	WebView	This item indicates the network interface to be used by Web-Based Admin View.
		If Web-Based Admin View is being used, USE is set. If not, UNUSE is set.
 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.

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

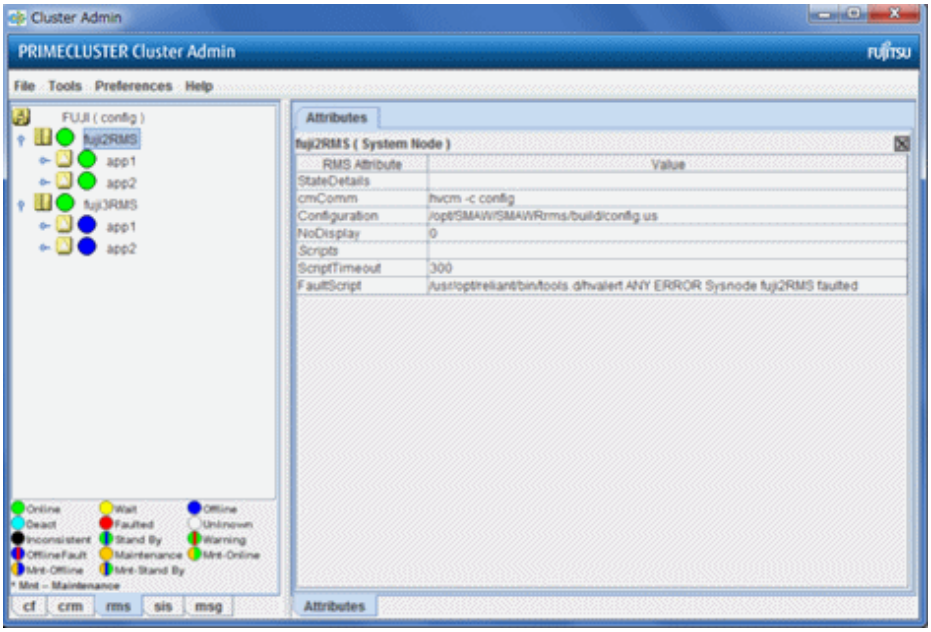
The RMS tree displays the configuration information of the cluster in a hierarchical format. The tree has the following levels:

- 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."

Node state display

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.

Icon	Icon color	Outline	Details
	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.

Note

The node states in the RMS main window of Cluster Admin may be displayed as Unknown. In this case, exit the Web-Based Admin View screen and restart. If the node states in the RMS main window of Cluster Admin are still displayed as Unknown, check the node states by using hvdisp -a.

State display of other objects

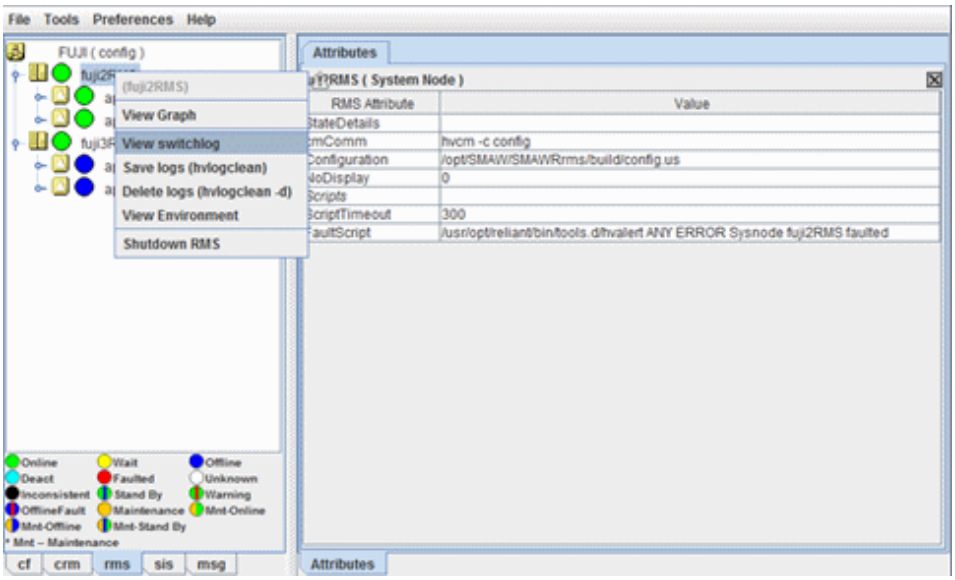
The following state icons are displayed with object icons that fall under parent objects.

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.

Icon	Icon color	Outline	Details
	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 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.

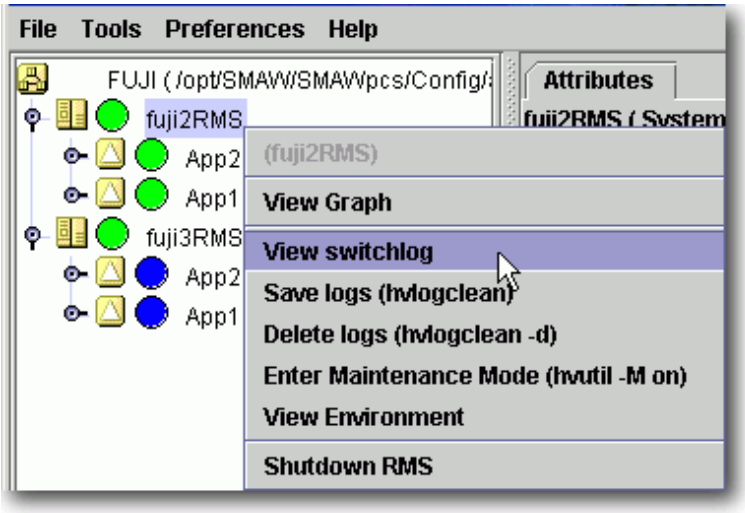
7.1.3.2 Configuration information or object attributes

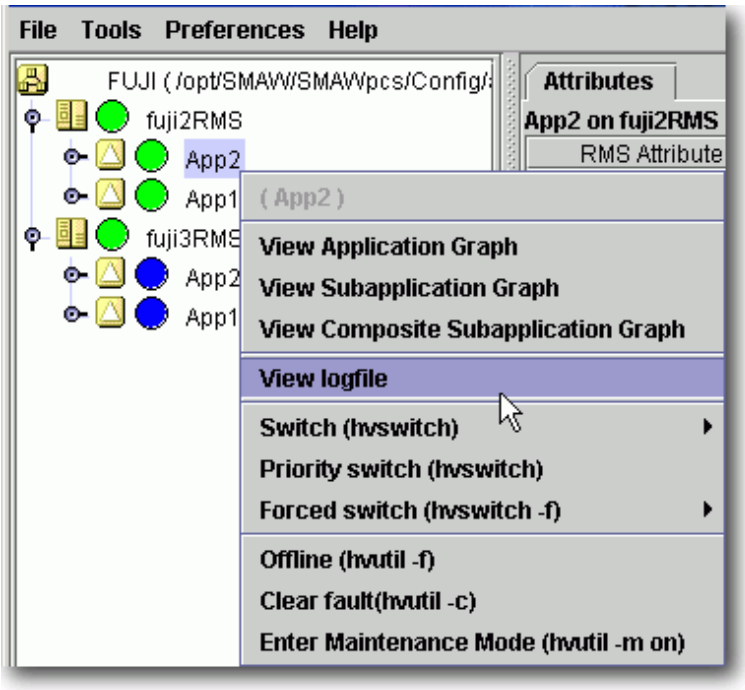
View the configuration information for the individual objects by left-clicking the object in the RMS tree. The properties are displayed in a tabular format on the right-hand side panel of the RMS main window.

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*.





Note

If you have built a cluster system in the following environments, these logs cannot be displayed from Cluster Admin.

- FUJITSU Cloud Service K5 environment
- Environment that uses Firewall

Display all RMS log files (/var/opt/SMAWRrms/log/) using a common UNIX text editor such as vi.

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 click *Start RMS -> all available nodes*. When the confirmation screen is displayed, click *OK*.

    You can also start RMS on individual nodes directly.

    1. Choose the node you want to start from the cluster tree in the RMS main window.
    2. Right-click on the node and select [Start RMS] from the pop-up menu.

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 *Tool* pull-down menu on the RMS main window or right-click the system node, and then select the shutdown mode on the screen that appears next.

    1. Choose either a specific node or all the nodes.

    2. Choose to stop all cluster applications, leave the applications running, or forcibly shutdown the node(s).

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 Resource 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 the 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 Resource 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 Resource 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 Resource Information" and "7.3.3 Concurrent Viewing of Node and Cluster Application States."

7.2.2.5 Clearing the Wait State of a Node

A node becomes the Wait state during state transition. 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 Resource 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 becomes 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 becomes 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 on all the nodes.
- 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 the 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 "6.11.2.1.4 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 Resource 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 Resource 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 "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

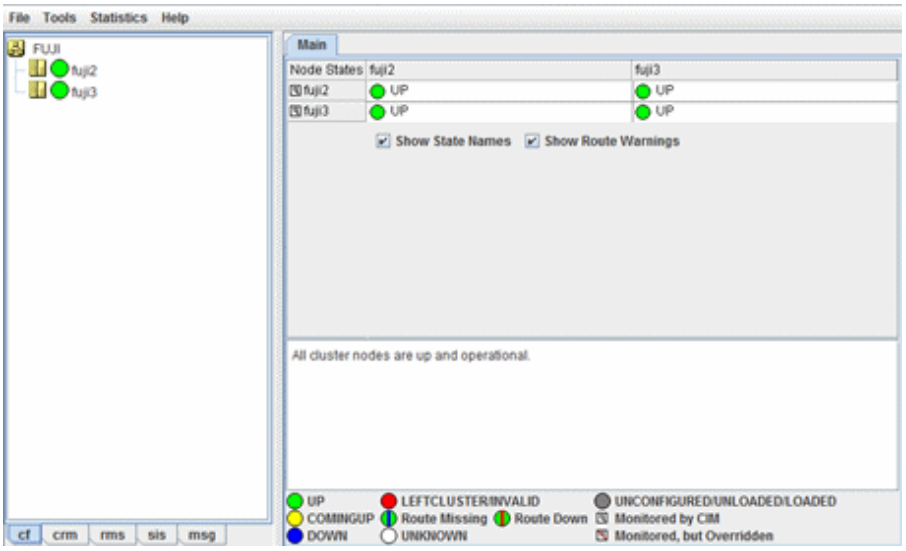
The pop-up context menu can be displayed by right-clicking the icon of the RMS graph. For details on the RMS graph, see "7.3.5 Viewing Detailed Resource Information."

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 a node on the CF tree. The node state will appear in the right panel.

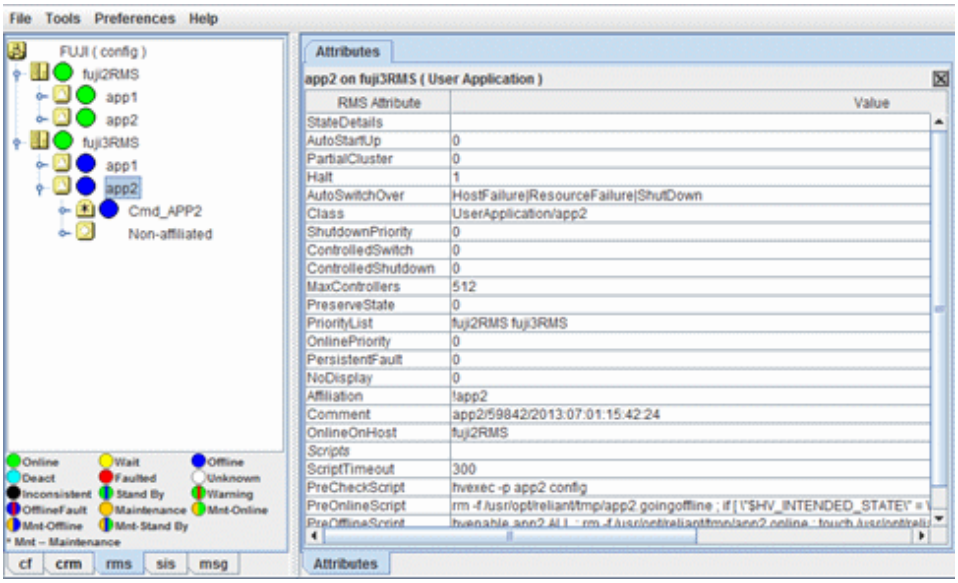


The node indicates one of the following states:

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

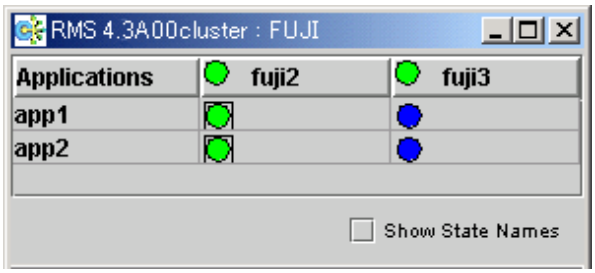
See "State display of other objects" in "7.1.3.1 RMS Tree."

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.

Note

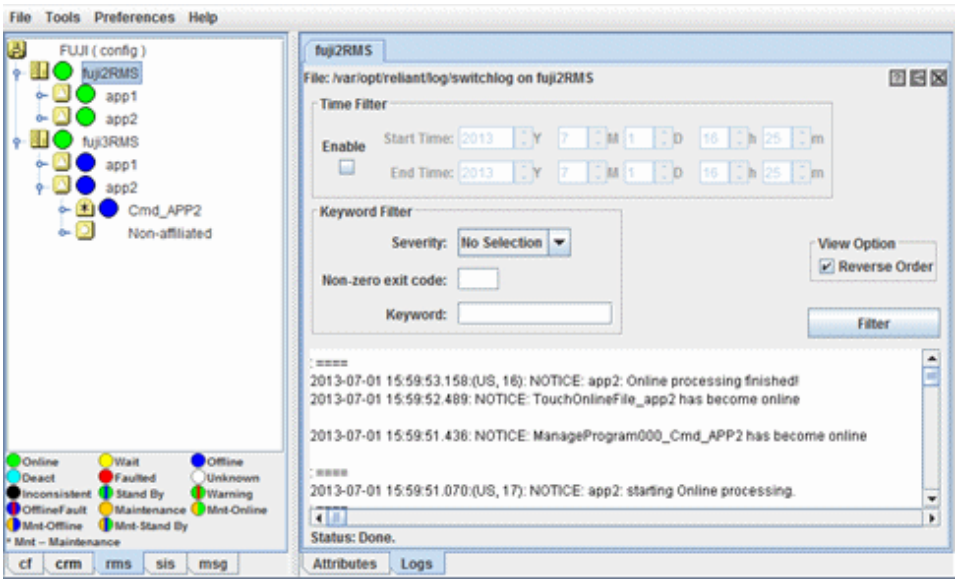
If you have built a cluster system in the following environments, these logs cannot be displayed from Cluster Admin.

- FUJITSU Cloud Service K5 environment
- Environment that uses Firewall

Display all RMS log files (/var/opt/SMAWRrms/log/) using a common UNIX text editor such as vi.

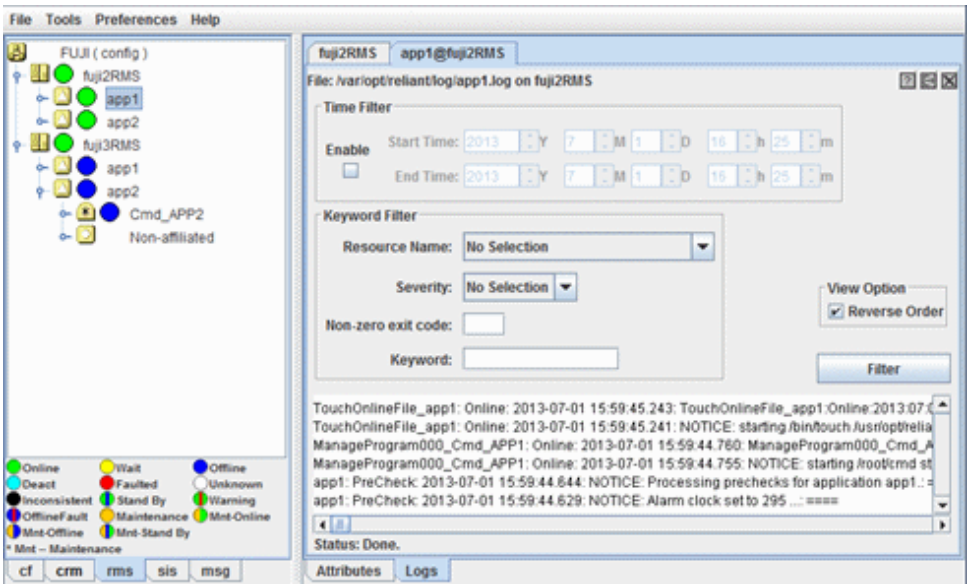
7.3.4.1 Viewing switchlogs

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 Resource Information

Use RMS graphs to display detailed resource 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.

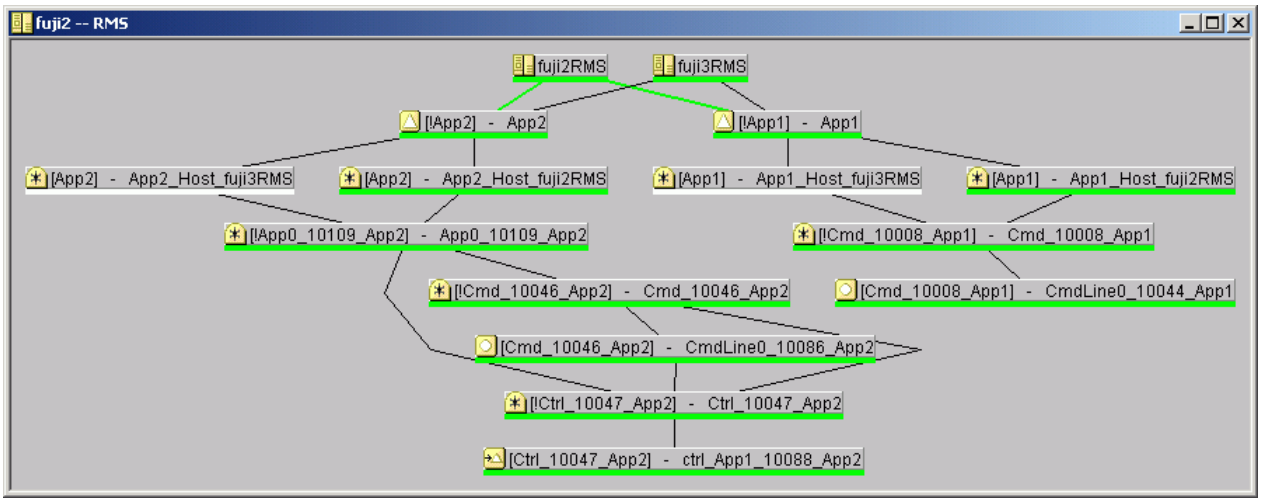
- Sub-application graph

  Lists all sub-applications used by a given application and shows the connections between the sub-applications.

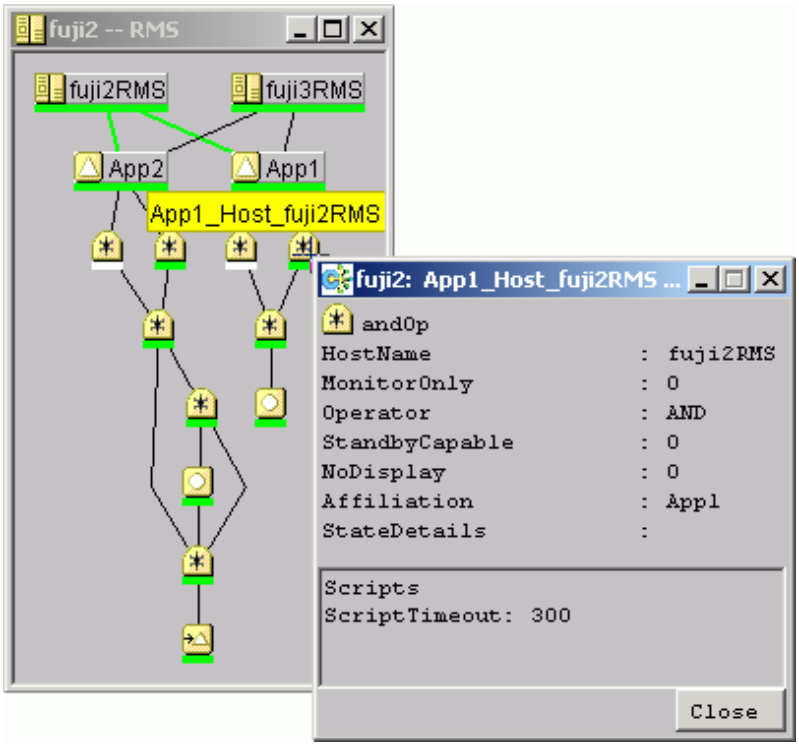
- Composite sub-applications graph

  Shows all sub-applications that the application depends on directly or indirectly.

RMS graphs



If you left-click the target object, the attributes of the object will be displayed on a pop-up screen.

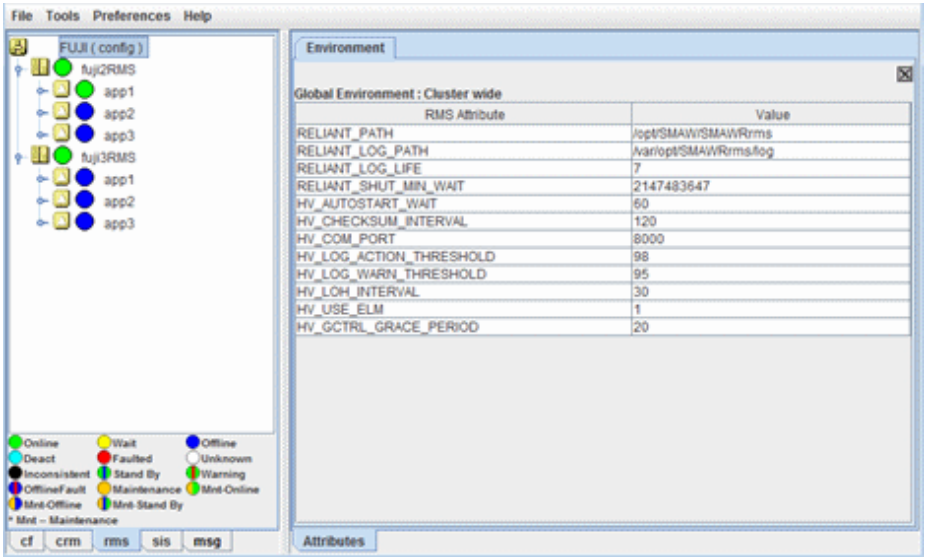


See

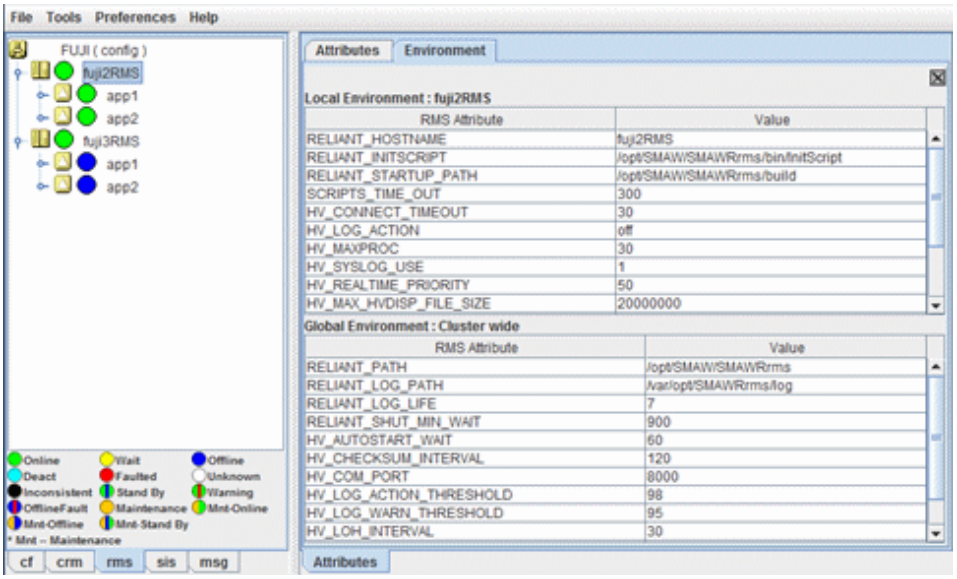
See "7.2 Using RMS graphs" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

7.3.6 Displaying environment variables

Right-click a cluster in the RMS tree window and select *View Environment*. The local and global variables are displayed.



Right-click a node in the RMS tree, and select *View Environment*. The local 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

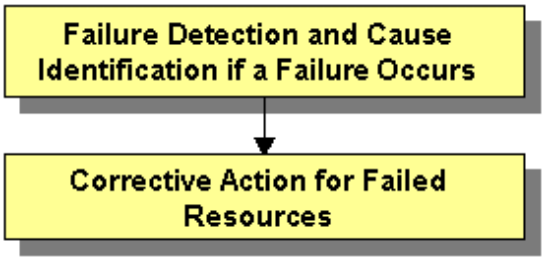
A hardware or software failure might occur while the cluster system is running. If a failure occurs in a resource, a message indicating that a failure occurred is displayed in the console screen. Based on this message, you need to identify the faulted resource using the CF and RMS main window and take 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 failed.

7.4.1 Corrective Action in the event of a resource failure

This section describes the corrective actions to take when a resource failure occurs.



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 (g) below are relevant to the "Failure confirmation features list" given below:

Failure detection

Normally, the RMS main window (a) 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:

  - The color of the icons in the RMS main window (a) changes.
  - A message is output to the msg main window (c), Syslog(f), and the console (g).

- 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:

  - A message is output to Syslog(f) and the console (g).

- 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(f) and the console (g). 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 list" to detect the failure.

Cause identification

You can also use the function that detected the failure and the features listed in "Failure confirmation features list" below to identify the faulted resource that caused the failure.

Failure confirmation features list

Failure confirmation features		Manual reference
(a)	**RMS main window** The RMS tree and the RMS cluster table can be used from this screen.	7.1.3 RMS Main Window
(b)	**CF main window** The CF tree can be used from this screen.	7.1.1 CF Main Window
(c)	**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.	-
(d)	Application log	7.3.4.2 Viewing application logs
(e)	switchlog	7.3.4.1 Viewing switchlogs
(f)	Syslog	-
(g)	**Console*** Messages that are displayed on the console or Syslog can be checked.	PRIMECLUSTER Messages
(h)	GDS GUI	PRIMECLUSTER Global Disk Services Configuration and Administration Guide

***Console**

- The operator intervention request messages (message numbers: 1421, 1423), incurred when RMS is not started on all the nodes, are displayed only when yes(1) is set for the AutoStartUp attribute of the userApplication. For information on the userApplication attribute, see "12 Appendix - Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."
- The operator intervention request messages (message numbers: 1422, 1423) and the error resource messages incurred after a resource or system error occurs are displayed only when yes(1) is set for the PersistentFault attribute of the userApplication. For information on the userApplication attribute, see "12 Appendix - Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."
- The operator intervention request and error resource messages are displayed by using the "clwatchlogd" daemon to monitor switchlog. You need to send the SIGHUP signal to clwatchlogd when you change the value of RELIANT_LOG_PATH that is defined in the "hvenv.local" file. When clwatchlogd receives this signal, clwatchlogd acquires the latest value of RELIANT_LOG_PATH. After you change RELIANT_LOG_PATH, you must start RMS.

Note

When you check the message of a resource failure, a resource with the "MONITORONLY" attribute may be in the fault state even if the cluster application is in the Offline state. Check whether there are any resources in the fault state. Especially, check that Fsystem resources are not in the fault state.

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."

   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.

(Example) Symfoware

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."

7.4.1.3 Recovery of Failed Cluster Interconnect

The following problems can cause cluster interconnect failures.

- Hardware error
  - Error on LAN card, hub, or cable
  - Connection error
- Network configuration error
  - Configuration error on IP address, netmask, or routing information, etc.

Contact your system administrator on the network configuration error. The following section describes how to fix hardware related errors.

If any heartbeat error on the cluster interconnect is detected, either of the following messages will be output to the /var/log/messages file.

```
"CF: Problem detected on cluster interconnect NIC_NAME to node NODE_NAME: missing heartbeat replies.
(CODE)"
"CF: Problem detected on cluster interconnect NIC_NAME to node NODE_NAME: ICF route marked down.
(CODE)"
```

"*NIC_NAME*" indicates the network interface card on which the error is detected.

"*NODE_NAME*" indicates the CF node name on which the error is detected.

"*CODE*" indicates the necessary information to determine the cause.

When either of the above messages is output to the file, follow the steps below.

Corrective action

1. Determining the failed node

   Confirm that each device is working properly. You can also use the ping command to determine the failed node and its location.

   Note

   When an error on the entire cluster interconnects (all interconnects for every node) occurs, the cluster system forcibly shut down all the nodes except one which has the highest survival priority.

   For details on survival priority, see "5.1.2 Setting up the Shutdown Facility."

   If an error on an active node (e.g. LAN card error of a node on which an active cluster application resides) occurs, you must stop the node before fixing it. To minimize the down time, make sure to follow the steps below before performing "Step 2. Performing maintenance tasks."

   1. Stopping a node in the "Online" state

      Before performing the maintenance task, stop the node on which "Online" cluster application resides.

2. Starting the forcefully terminated node

   Start the node which was forcefully terminated by the cluster system and make the cluster application back to the "Online" state. For details on how to start a cluster application, see "7.2.1.1 Starting RMS."

   Be sure to check that the node, which is described in Step 1. Stopping a node in the "Online" state, is completely stopped before performing this step.

2. Performing maintenance tasks

   After determining the cause of the error, perform the following maintenance task depending on the category of error.

   Note

   For a LAN card error, the failed node must be stopped to perform the maintenance task.

   For an error on cables or hubs, you can perform the maintenance task with the node being active.

   - When the error was caused by your LAN card or cable

     If the cable is unplugged, plug in properly.
     If the cable is properly plugged, your LAN card might be the cause. Contact field engineers.

   - When the error was caused by a hub

     If the power is off , push the power button.
     If the power is on, there is a possibility the hub is broken down. Contact field engineers.

3. Recovery

   To recover the partial failure of the cluster interconnect, skip to "Step 2. Cluster interconnect recovery" below.

   1. Starting all the nodes

      Start all the nodes.

   2. Cluster interconnect recovery

      Use the ping command to confirm if nodes can communicate each other through the failed cluster interconnect.

   After confirming that the cluster interconnect is recovered successfully, clear the "Faulted" state of the cluster application as necessary. For details on the operation, see "7.2.2.4 Bringing Faulted Cluster Application to available state."

7.4.2 Corrective Action in the event of the LEFTCLUSTER state when the virtual machine function is used

If the host OS becomes the panic state or hangs up when the virtual machine is used, the LETCLUSTER state may occur. This section describes the corrective actions in this case.

7.4.2.1 When the host OS becomes the panic state

1. When the host OS becomes the panic state, and the host is restored after that, go to the step 3.
2. Check the state of the host OS and restore the host OS.
3. If a failed node (guest OS) is in the LEFTCLUSTER state, perform the procedure in "5.2 Recovering from LEFTCLUSTER" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."
4. Check that the node joins the cluster.

7.4.2.2 When the host OS hangs up

1. Check that the node (guest OS) is actually stopped. If it is running, stop it.
2. Check the state of the host OS and restore the host OS.

3. If a failed node is in the LEFTCLUSTER state, perform the procedure in "5.2 Recovering from LEFTCLUSTER" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

4. Check that the node joins the cluster.

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 the nodes simultaneously

When executing the hvshut -l command on all the nodes simultaneously, RMS will not be stopped and occasionally the timeout and hvshut command times out or hangs up.
When stopping RMS on all the 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 the nodes simultaneously and the hvshut command times out, stop or restart all the nodes. In addition, if the hvshut command hangs up, stop RMS forcibly using the hvshut -f command, and then stop or restart 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 restart 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 /iSCSI 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 on any of the nodes

Set the node to panic status manually with one of the following methods.

- Press <Alt> + <SysRq> + <C> on the system console.

- Press the NMI button.

For details, see "Linux user guide."

When you start cluster applications manually or confirm the message of a resource failure, check whether a resource with the "MONITORONLY" attribute has been in the fault state.

If you start or switch over cluster applications before the failure of the resource with the "MONITORONLY" attribute is solved, cluster inconsistencies or data corruption may occur.

When you set Firewall and use the state module in Firewall, do not restart the iptables service or the ip6tables service during PRIMECLUSTER operation.

When using the state module in Firewall, restarting the iptables service or the ip6tables service triggers initializing information of the communication status, and subsequent communication may not work correctly. Neither applications nor PRIMECLUSTER can work correctly, when you change the setting of Firewall, perform one of the following operations:

- Restarting the cluster node
- Reflecting the change by iptables-restore or ip6tables-restore

The following error messages may be output to the console and syslog during system startup in RHEL6 environment

The following messages may be output to the console and syslog during system startup in RHEL6 environment. This does not disrupt ongoing operation.

```
kernel: Disabling lock debugging due to kernel taint
kernel: clonltrc: module license 'Proprietary' taints kernel.
kernel: symsrv: module license 'Proprietary' taints kernel.
kernel: symsrv: applying 16k kernel stack fix up
kernel: cf: module license 'Proprietary' taints kernel.
kernel: cf: applying 16k kernel stack fix up
kernel: sha: module license 'Proprietary' taints kernel.
```

The following error messages may be output to the console and Syslog during system startup in RHEL7 environment

The following messages may be output to the console and Syslog during system startup in RHEL7 environment. This does not disrupt ongoing operation.

```
kernel: Request for unknown module key 'FUJITSU Software: Fujitsu BIOS DB FJMW Certificate:
Hexadecimal, forty-digit' err -11
kernel: Disabling lock debugging due to kernel taint
kernel: clonltrc: module license 'Proprietary' taints kernel.
kernel: clonltrc: module verification failed: signature and/or required key missing - tainting kernel
kernel: sfdsk_lib: module verification failed: signature and/or required key missing - tainting kernel
kernel: sha: module license 'Proprietary' taints kernel.
kernel: sha: module verification failed: signature and/or required key missing - tainting kernel
kernel: symsrv: module license 'Proprietary' taints kernel.
kernel: symsrv: applying kernel_stack fix up
kernel: symsrv: module verification failed: signature and/or required key missing - tainting kernel
kernel: cf: applying kernel_stack fix up
kernel: poffinhibit_ipdv: module verification failed: signature and/or required key missing -
tainting kernel
```

7.5.1 Notes on Switching a Cluster Application Forcibly

When Forced switch request (Forced startup) of a cluster application or a resource is issued, RMS overrides all safety checks and starts the cluster application or the 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.

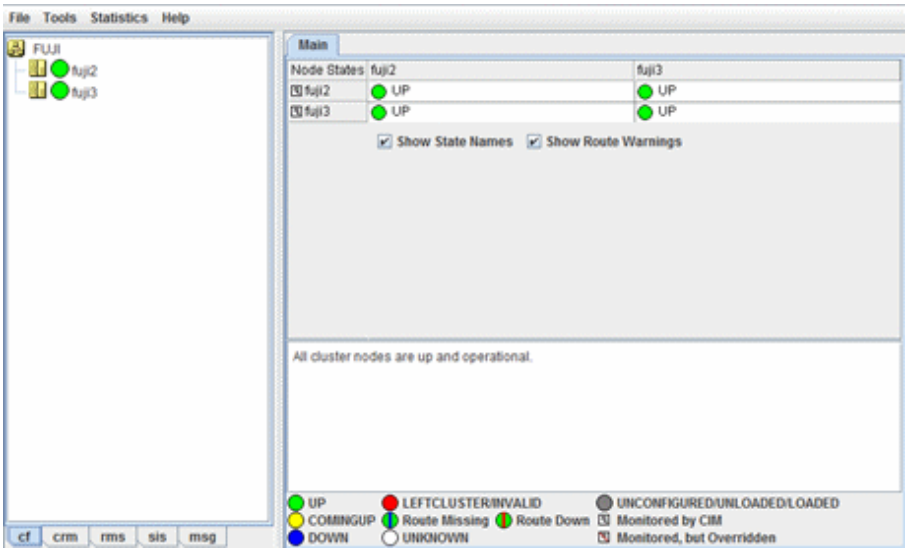
To perform forced startup of a cluster application or a resource safely, check whether RMS is running on all the nodes in the cluster before starting forced startup according to the following procedure, and if there are the nodes on which RMS is not running, then shut down the nodes.

1. Check the node state by one of the following methods:

   - Execute the cftool -n command on all the nodes.

```
fuji2# cftool -n
Node  Number State       Os      Cpu
fuji2 1      UP          Linux   EM64T
fuji3 2      UP          Linux   EM64T
```

   - Check the CF tree of the Cluster Admin.



2. 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 (CF) Configuration and Administration Guide."

   - 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 OS in single-user mode.

3. Check whether some nodes on which RMS is not running exist among the nodes on which the cluster application or the resource will be forcibly started by one of the following methods:

   - Execute the hvdisp -a command on nodes where the cluster application or the resource will be started and check that the state of objects whose Type is SysNode is Online.

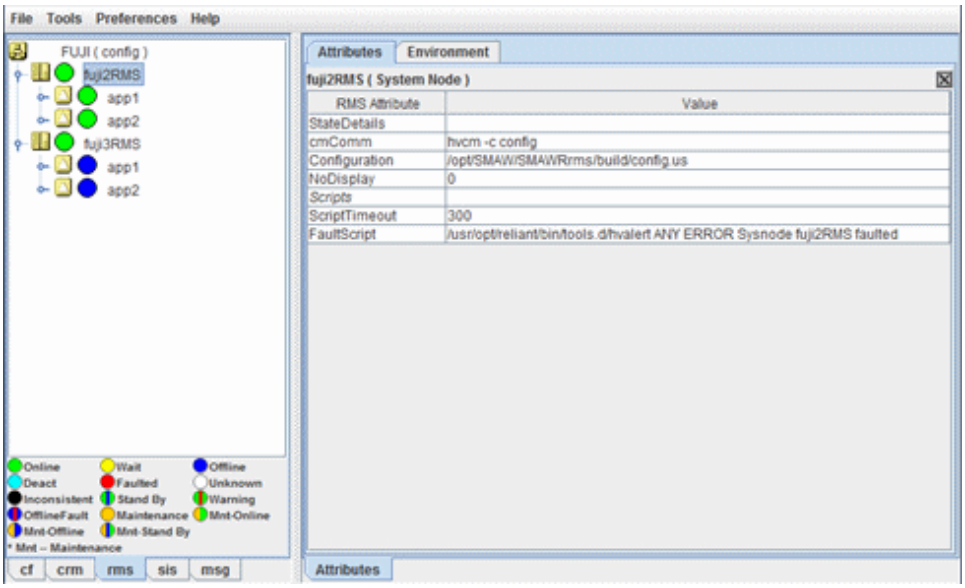
```
fuji2# hvdisp -a

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.



4. If nodes which satisfy the following conditions exist, shut down the operating system of the nodes, or restart OS in single-user mode.
   - The node state is UP, and
   - The state of SysNode is not Online.
5. Execute the Forced switch (hvswitch -f) to forcibly start the cluster application or the 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 7.2 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)	- 4.1A40 or earlier 45 seconds - 4.2A00 or later 600 seconds

(*1): When using the monitoring agent of PRIMECLUSTER, the monitoring agent detects it immediately

(*2): In the environment where the ELM heartbeat (RMS heartbeat) is available, the ELM heartbeat detects it immediately (the ELM heartbeat is available in 4.2A00 or later as default).

(*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 sub directories under RELIANT_LOG_PATH whose update time (ctime) is older than the number of days specified 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.....283
Chapter 9 Changing the Cluster System Environment.....299
Chapter 10 Configuration change of Cluster Applications.....312
Chapter 11 Changing the Operation Attributes of a Cluster System.....338

Chapter 8 Changing the Cluster System Configuration

This chapter explains some configuration nodes of PRIMECLUSTER system, and how to add, delete, and change hardware.

Before adding the cluster application or the resource, check "Design (the number of resources)" of PRIMECLUSTER Designsheets to verify that the number of resource objects and the number of detectors that can be set in the whole PRIMECLUSTER system do not exceed their maximum values.

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 "6.9 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
- System board

Note

- When you change a system board, reconfigure BMC or iRMC used by the shutdown facility.
- When you change a system board or a network interface card, do not restart the network.

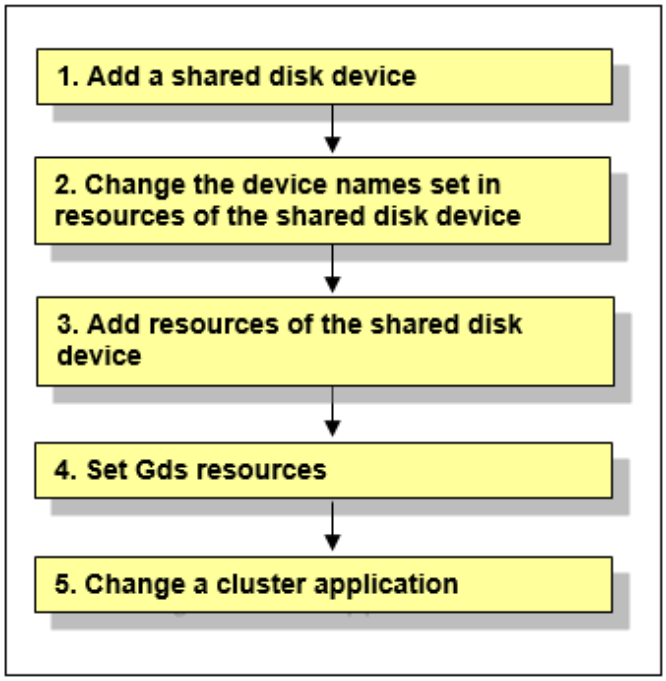
8.1.1 Adding Hardware

This section describes how to add hardware.

8.1.1.1 Adding a shared disk device

The following describes how to add a shared disk device.

Figure 8.1 Procedure to add a shared disk device



Information

You must stop RMS during performing "5. Change a cluster application."

However, you do not need to stop RMS if all the following conditions are met because performing "5. 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.

Operation Procedure:

1. Add a shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to add a shared disk device.

2. Change the device names set in resources of the shared disk device.

   Update the device names set in the resources of the existing shared disk device to the current device names.

   Execute the following command. For *filepath,* specify an empty file with absolute path.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

   ### Note

   When SDX_UDEV_USE=off is described in the GDS configuration file /etc/opt/FJSVsdx/sdx.cf, do not execute the clautoconfig command.

3. Add resources of the shared disk device.

   Register resources corresponding to the added shared disk device to the resource database.

See

To register resources, see "5.1.3.2 Registering Hardware Devices."

4. Set up Gds resources.

   To use 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 GDS Configuration Setup," "6.7.3.3 Preliminary Setup for Gds Resources," and "6.7.3.4 Setting Up Gds Resources."

5. Change a cluster application.

   Change a cluster application to add the following resources related to the added shared disk device to the cluster application.

   - Fsystem resource
   - Gds resource

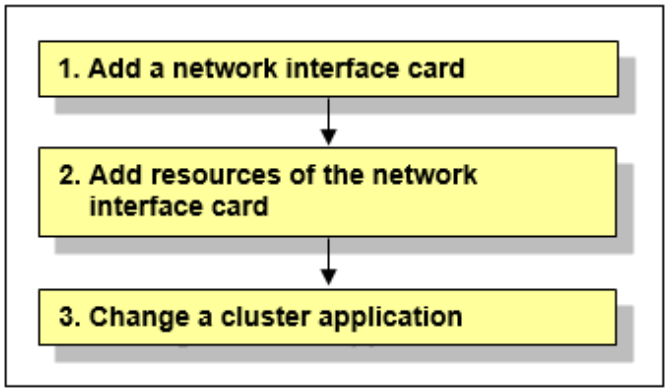
See

For information on how to change a cluster application, see "10.3 Changing a Cluster Application."

8.1.1.2 Adding a Network Interface Card Used for the Public LAN and the Administrative LAN

This section describes how to add a network interface card used for the public LAN and the Administrative LAN.

Figure 8.2 Procedure to add a network interface card



Operation Procedure:

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 to the resource database.

See

To register resources, see "5.1.3.2 Registering Hardware Devices."

3. Change a cluster application.

   Change a cluster application to add the following resources related to the added network interface card to the cluster application.

   - Takeover network resource
   - Gls resource

   See

   For information on how to change a cluster application, see "10.3 Changing a Cluster Application."

8.1.1.3 Adding Hardware by DR (Dynamic Reconfiguration)

This section explains the procedure for adding a system board by DR during PRIMECLUSTER system operation.

If a system board is added by DR, this might affect the PRIMECLUSTER monitoring facility resulting in node elimination.

If DR needs to be used, stop the cluster monitoring facility beforehand with the following procedure:

1. Execute the "hvshut" command on each node to stop PRIMECLUSTER RMS as follows. Answer "yes," then only RMS will stop. The cluster application will remain running.

```
# 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

NOTICE: User has been warned of 'hvshut -L' and has elected to proceed.
```

   Add the following line to the end of the "/opt/SMAW/SMAWRrms/bin/hvenv.local" file on each node.

```
export HV_RCSTART=0
```

   It is necessary to perform the procedure above so that RMS will not automatically start immediately after OS startup.

2. Execute the "sdtool" command on each node to stop PRIMECLUSTER SF as follows.

```
# sdtool -e
LOG3.013806902801080028   11   6    30   4.4A00      SMAWsf          : RCSD returned a
successful exit code for this command
```

3. Perform the following operation on each node to change the timeout value of PRIMECLUSTER CF:

   - Add the following line to the "/etc/default/cluster.config" file.

```
CLUSTER_TIMEOUT "600"
```

   - Execute the following command.

```
# cfset -r
```

   - Check whether or not the timeout value is valid.

```
# cfset -g CLUSTER_TIMEOUT
>From cfset configuration in CF module:
Value for key: CLUSTER_TIMEOUT --->600
#
```

4. Use DR.

See

For DR operation, refer to the related hardware manual.

5. Perform the following operation on each node to return the timeout value of PRIMECLUSTER CF to the default value:

   - Change the value of CLUSTER_TIMEOUT defined in "/etc/default/cluster.config" file earlier to 10.

     Before change

     ```
     CLUSTER_TIMEOUT "600"
     ```

     After change

     ```
     CLUSTER_TIMEOUT "10"
     ```

   - Execute a following command.

     ```
     # cfset -r
     ```

   - Check whether or not the timeout value is valid.

     ```
     # cfset -g CLUSTER_TIMEOUT
     >From cfset configuration in CF module:
     Value for key: CLUSTER_TIMEOUT --->10
     #
     ```

6. Execute the "sdtool" command on each node to start the PRIMECLUSTER SF.

   ```
   # sdtool -b
   ```

7. Check if PRIMECLUSTER SF is running. (The following indicates an output example of a two-node configuration)

   ```
   # sdtool -s
   Cluster Host    Agent               SA State      Shut State  Test State  Init State
   ------------    -----               --------      ----------  ----------  ----------
   node0           SA_mmbp.so          Idle          Unknown     TestWorked  InitWorked
   node0           SA_mmbr.so          Idle          Unknown     TestWorked  InitWorked
   node1           SA_mmbp.so          Idle          Unknown     TestWorked  InitWorked
   node1           SA_mmbr.so          Idle          Unknown     TestWorked  InitWorked
   ```

8. Execute the "hvcm" command on each node to start PRIMECLUSTER RMS.

   ```
   # hvcm
   Starting Reliant Monitor Services now
   ```

9. RMS must be running on all the nodes. Check if each icon indicating the node state is green (Online) in the RMS main window of Cluster Admin.

   Finally, remove the following line from "/opt/SMAW/SMAWRrms/bin/hvenv.local" file on each node.

   ```
   export HV_RCSTART=0
   ```

Note

- If you plan to use DR, be sure to verify a cluster system during cluster configuration using the above steps.
- If a node failure (such as a node panic or reset) or a hang-up occurs due to hardware failure and so on during step 1 through 7, you need to follow the procedure below to start the cluster application, which was running on the node where DR is used, on a standby node.

  1. If a hang-up occurs, stop the failed node forcibly, and then check that the node is stopped.

2. Mark the node DOWN by executing the "cftool" command on any of the nodes where a failure has not been occurred and specifying the node number and CF node name for failed nodes. However, if the state of the failed node is not LEFTCLUSTER, wait until the node becomes LEFTCLUSTER, and then execute the "cftool -k" command.

```
# cftool -n
Node  Number State         Os       Cpu
node0 1       UP           Linux    EM64T
node1 2       LEFTCLUSTER  Linux    EM64T
# cftool -k
This option will declare a node down. Declaring an operational
node down can result in catastrophic consequences, including
loss of data in the worst case.
If you do not wish to declare a node down, quit this program now.

Enter node number: 2
Enter name for node #2: node1
cftool(down): declaring node #2 (node1) down
cftool(down): node node1 is down
# cftool -n
Node  Number State        Os        Cpu
node0 1       UP          Linux     EM64T
node1 2       DOWN        Linux     EM64T
#
```

3. Perform Steps 5 through 9 on all the nodes where no failure occurred, and then start RMS. If the cluster application is in an active standby configuration, execute the "hvswitch -f " command to force the cluster application to go Online. For details on the "hvswitch" command, see the description of the -f option of the online manual page for the command.

```
# hvswitch -f userApplication
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]:yes
#
```

4. After restoring the failed node, perform step 5 through 9 on the appropriate node to start RMS.

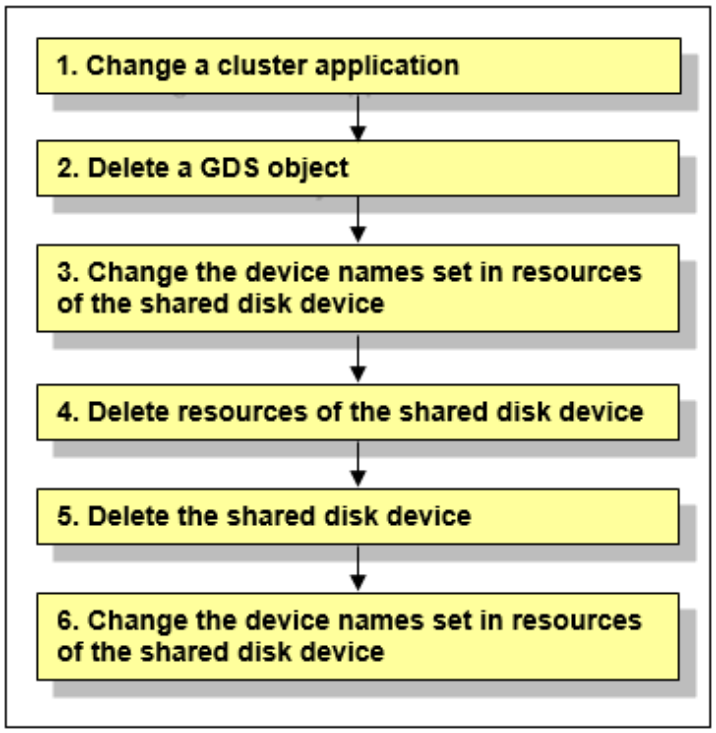
8.1.2 Deleting Hardware

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



Operation Procedure:

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 resources, see "10.3 Changing a Cluster Application" and "10.5 Deleting a Resource."

2. Delete a GDS object.

   Delete a GDS object related to the shared disk device to be deleted.

   See

   For deleting a GDS object, see "Removing Configuration" of "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

3. Change the device names set in resources of the shared disk device.

   Before deleting resources, update the device names set in the resources to the current device names. Execute the following command. For *filepath*, specify an empty file with absolute path.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

4. Delete resources of the shared disk device.

   Delete resources of the registered shared disk device by using the "cldelrsc" command.

   For details on the "cldelrsc" command, see the manual page.

After executing the "cldelrsc" command, execute the following command to inform that resources are deleted to GDS.

Specify the full path of an empty file for *filepath*.

```
# /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
```

Note

- When the shared disk device, from which resources are to be deleted, is 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 a GDS class, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."
- When SDX_UDEV_USE=off is described in the GDS configuration file /etc/opt/FJSVsdx/sdx.cf, do not execute the clautoconfig command.

5. Delete the shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to delete the shared disk device.

6. Change the device names set in resource of the shared disk device.

   By deleting the shared disk device, any device name of the shared disk device which has not been deleted may be changed. To modify the device name of the resource of the shared disk device according to the correct device name, execute the following command. Specify the full path of an empty file for *filepath*.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

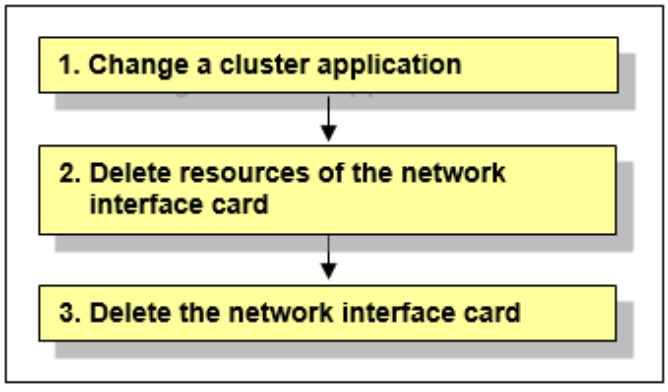
   Note

   When SDX_UDEV_USE=off is described in the GDS configuration file /etc/opt/FJSVsdx/sdx.cf, do not perform Step 6.

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



Operation Procedure:

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 resources, see "10.3 Changing a Cluster Application" and "10.5 Deleting a Resource."

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.

8.1.2.3 Removing System Board by Dynamic Reconfiguration

This section explains the procedure for removing a system board by using Dynamic Reconfiguration (DR) during PRIMECLUSTER system operation.

If a system board is hot-removed by DR, this might affect the PRIMECLUSTER monitoring facility resulting in node elimination. If DR needs to be used, stop the cluster monitoring facility beforehand with the following procedure:

Note

A system board equipped with I/O cannot be removed by DR. Before removing a system board, also make sure to estimate that the ongoing operation can be continued even after the amount of CPU and memory is decreased.

1. Execute the "hvshut" command on each node to stop PRIMECLUSTER RMS as follows. Answer "yes," then only RMS will stop. The cluster application will remain running.

```
# 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

NOTICE: User has been warned of 'hvshut -L' and has elected to proceed.
```

   Add the following line to the end of the "/opt/SMAW/SMAWRrms/bin/hvenv.local" file on each node.

```
export HV_RCSTART=0
```

   It is necessary to perform the procedure above so that RMS will not automatically start immediately after OS startup.

2. Execute the "sdtool" command on each node to stop the PRIMECLUSTER shutdown facility as follows.

```
# sdtool -e
LOG3.013806902801080028   11   6    30   4.4A00      SMAWsf          : RCSD returned a successful
exit code for this command
```

3. Perform the following operation on each node to change the timeout value of PRIMECLUSTER CF:

   - Add the following line to the "/etc/default/cluster.config" file.

```
CLUSTER_TIMEOUT "600"
```

- Execute the following command.

```
# cfset -r
```

- Check whether or not the timeout value is valid.

```
# cfset -g CLUSTER_TIMEOUT
>From cfset configuration in CF module:
Value for key: CLUSTER_TIMEOUT --->600
#
```

4. Use DR.

See

For DR operation, refer to the related hardware manual.

5. Perform the following operation on each node to return the timeout value of PRIMECLUSTER CF to the default value.

   - First, change the value of CLUSTER_TIMEOUT defined in "/etc/default/cluster.config" file earlier to 10.

     Before change:

```
CLUSTER_TIMEOUT "600"
```

     After change:

```
CLUSTER_TIMEOUT "10"
```

   - Execute the following command.

```
# cfset -r
```

   - Check whether or not the timeout value is valid.

```
# cfset -g CLUSTER_TIMEOUT
>From cfset configuration in CF module:
Value for key: CLUSTER_TIMEOUT --->10
#
```

6. Execute the "sdtool" command on each node to start the PRIMECLUSTER shutdown facility.

```
# sdtool -b
```

7. Check if the PRIMECLUSTER shutdown facility is running. (The following indicates an output example of a two-node configuration.)

```
# sdtool -s
```

```
Cluster Host    Agent               SA State      Shut State  Test State  Init State
------------    -----               --------      ----------  ----------  ----------
node0           SA_mmbp.so          Idle          Unknown     TestWorked  InitWorked
node0           SA_mmbr.so          Idle          Unknown     TestWorked  InitWorked
node1           SA_mmbp.so          Idle          Unknown     TestWorked  InitWorked
node1           SA_mmbr.so          Idle          Unknown     TestWorked  InitWorked
```

8. Execute the "hvcm" command on each node to start PRIMECLUSTER RMS.

```
# hvcm
Starting Reliant Monitor Services now
```

9. RMS must be running on all the nodes. Check if each icon indicating the node state is green (Online) in the RMS main window of Cluster Admin.

   Finally, remove the following line from "/opt/SMAW/SMAWRrms/bin/hvenv.local" file on each node.

```
export HV_RCSTART=0
```

Note

- If you plan to use DR, be sure to verify a cluster system during cluster configuration using the above steps.
- If a node failure (such as a node panic or reset) or a hang-up occurs due to hardware failure and so on during step 1 through 7, you need to follow the procedure below to start the cluster application, which was running on the node where DR is used on a standby node.

  1. If a hang-up occurs, stop the failed node forcibly, and then check that the node is stopped.
  2. Mark the node DOWN by executing the "cftool" command on any of the nodes where a failure does not occur and specifying the node number and CF node name for failed nodes. However, if the state of the failed node is not LEFTCLUSTER, wait until the node becomes LEFTCLUSTER, and then execute the "cftool -k" command.

```
# cftool -n
Node  Number State         Os       Cpu
node0 1       UP            Linux    EM64T
node1 2       LEFTCLUSTER   Linux    EM64T
# cftool -k
This option will declare a node down. Declaring an operational
node down can result in catastrophic consequences, including
loss of data in the worst case.
If you do not wish to declare a node down, quit this program now.

Enter node number: 2
Enter name for node #2: node1
cftool(down): declaring node #2 (node1) down
cftool(down): node node1 is down
# cftool -n
Node  Number State        Os        Cpu
node0 1       UP           Linux     EM64T
node1 2       DOWN         Linux     EM64T
#
```

  3. Perform Steps 5 through 9 on all the nodes where no failure occurred, and then start RMS. If the cluster application is in an active standby configuration, execute the "hvswitch -f " command to force the cluster application to go Online. For details on the "hvswitch" command, see the description of the -f option of the online manual page for the command.

```
# hvswitch -f userApplication
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]:yes
#
```

  4. After restoring the failed node, perform step 5 through 9 on the appropriate node to start RMS.

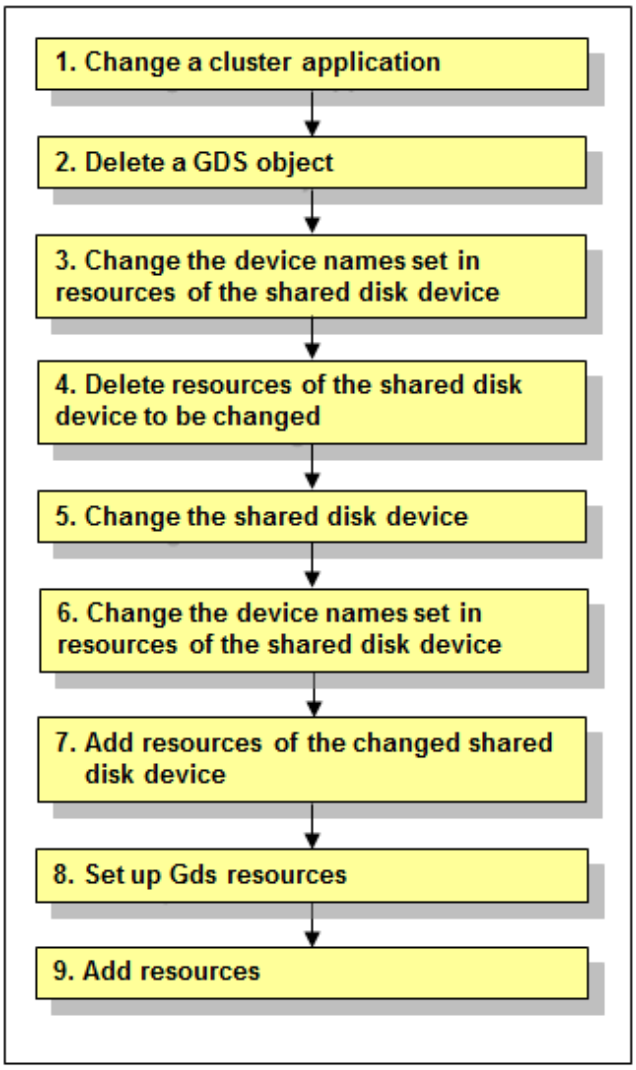
8.1.3 Changing Hardware

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 resources of the target shared disk device beforehand. After the change, you need to add the resources of the changed shared disk device.

Figure 8.5 Changing a shared disk device



GDS: Global Disk Services

Operation Procedure:

1. Change a cluster application.

   Delete the following resources, which are using the shared disk device to be changed, from the cluster application:

   - Fsystem resource
   - Gds resource

   See

   For details on how to change the cluster application configuration and delete resources, see "10.3 Changing a Cluster Application" and "10.5 Deleting a Resource."

2. Delete a GDS object.

   Delete a GDS object related to the shared disk device to be changed.

   See

   For deleting a GDS object, see "Removing Configuration" of "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

3. Change the device names set in resources of the shared disk device.

   Before deleting resources, update the device names set in the resources to the current device names. Execute the following command. For *filepath*, specify an empty file with absolute path.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

4. Delete resources of the shared disk device to be changed.

   Delete resources of the registered shared disk device by using the "cldelrsc" command.

   For details on the "cldelrsc" command, see the manual page.

   After executing the "cldelrsc" command, execute the following command to inform that resources are deleted to GDS.

   Specify the full path of an empty file for *filepath*.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

   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 a GDS class, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."
   - When SDX_UDEV_USE=off is described in the GDS configuration file /etc/opt/FJSVsdx/sdx.cf, do not execute the clautoconfig command.

5. Change the shared disk device.

   See "12.2 Maintenance Flow" and ask field engineers to change the shared disk device.

6. Change the device names set in resources of the shared disk device.

   Before adding resources to the changed shared disk device, update the device names set in the resources to the new device names. Execute the following command. For *filepath*, specify an empty file with absolute path.

   ```
   # /etc/opt/FJSVcluster/bin/clautoconfig -f filepath
   ```

   Note

   When SDX_UDEV_USE=off is described in the GDS configuration file /etc/opt/FJSVsdx/sdx.cf, do not execute the clautoconfig command.

7. Add resources of the changed shared disk device.

   Register resources corresponding to the changed shared disk device to the resource database.

   See

   For information on how to register the resource database, see "5.1.3.2 Registering Hardware Devices."

8. Set up Gds resources.

   To use GDS, set up GDS and create Gds resources.

   See

   For information on how to set up GDS and create Gds resources, see "6.3 GDS Configuration Setup" and "6.7.3.4 Setting Up Gds Resources."

9. Add resources.

   If you have deleted Fsystem resources in Step 1, add Fsystem resources.

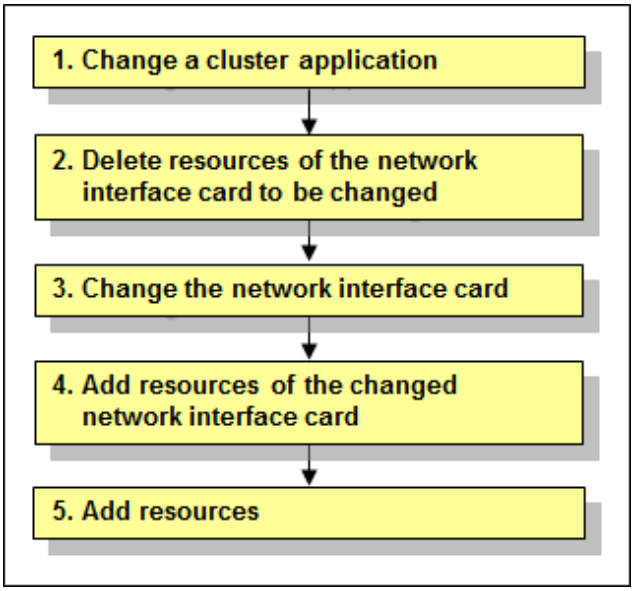
   See

   To add resources, see "6.7.3 Setting Up Resources."

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 resources of the target network interface card beforehand. After the change, you need to add resources of the network interface card.

Figure 8.6 Procedure to change a network interface card



Operation Procedure:

1. Change a cluster application.

   Delete the following resources, which are using the network interface card to be changed, from the cluster application:

   - Takeover network resource
   - Gls resource

   See

   For details on how to change the cluster application configuration and delete resources, see "10.3 Changing a Cluster Application" and "10.5 Deleting a Resource."

2. 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.

3. Change the network interface card.

   See "12.2 Maintenance Flow" and ask field engineers to change the network interface card.

4. Add resources of the changed network interface card.

   Register resources corresponding to the changed network interface card to the resource database.

   See

   For information on how to register the resource database, see "5.1.3.2 Registering Hardware Devices."

5. Add resources.

   If you have deleted takeover network resources and Gls resources in Step 1, add takeover network resources and Gls resources.

   See

   To add resources, see "6.7.3 Setting Up Resources."

8.1.3.3 Changing NIC of CIP

Below is an example when changing from eth3 to eth4.

Note

A network interface card used for cluster interconnects cannot be replaced using PCI Hot Plug. Stop the node and then replace the network interface card.

Procedure when changing from eth3 to eth4

1. Stop CF on all the nodes configuring a cluster.

   For how to stop CF, see "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

2. Check interfaces currently used by executing the following command on all the nodes.

   # **cfconfig -g**
   The own node name the cluster name eth3

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 eth4**

5. Check that the interfaces currently used has been changed by executing the following command on all the nodes.
   # **cfconfig -g**
   The own name the cluster name eth4 (Check that eth4 has been displayed).

6. In the environment where the shutdown agent SA_icmp for VMware environment is used, if the cluster interconnect is used to check whether the node is alive or not, modify /etc/opt/SMAW/SMAWsf/SA_icmp.cfg on each node.

   See

   For details, see "H.2.3.3 Setting Up the Shutdown Facility (when using I/O fencing function)."

7. Start CF on all the nodes configuring a cluster
    1. Log in to Web-Based-Admin View.
    2. Open Cluster Admin.
    3. Select the cf tab.
    4. Select the driver load.
    5. Select all check boxes (drivers) of a pop-up menu and check "OK."
    6. After starting CF of the first machine (initial connection), select "Start CF" from the CF of the second machine. Then, select the check boxes (drivers) of the pop-up and click "OK."
    7. Check that both nodes are Online on cf in Cluster Admin. In addition, check that each connector is UP.
    8. Finish Cluster Admin.
    9. Log out from Web-Based-Admin View.

Chapter 9 Changing the Cluster System Environment

This chapter describes how to change the configuration information and environmental settings of PRIMECLUSTER system.

Before adding the cluster application or the resource, check "Design (the number of resources)" of PRIMECLUSTER Designsheets to verify that the number of resource objects and the number of detectors that can be set in the whole PRIMECLUSTER system do not exceed their maximum values.

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 "6.9 Checking the Cluster Environment."

9.1 Changing the Cluster Configuration information

9.1.1 Changing a Node Name

The following explains how to change the node name after building a PRIMECLUSTER system.

Changing a node name may have a serious impact on the system. Therefore, make this change only when it is absolutely necessary.

Operation Procedure:

1. Stop the CF on the node whose node name is to be changed.

   For information on how to stop CF, see "4.6 Starting and stopping CF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

2. On the node whose node name is changed, change the old host name in the /etc/hosts file to the new host name.

   (Example)

   [Before change]

   ```
   10.20.30.40    node1
   ```

   [After change]

   ```
   10.20.30.40    nodeA
   ```

3. On the node whose node name is changed, change the old host name in the /etc/sysconfig/network file (for RHEL6) and the /etc/hostname file (for RHEL7) to the new host name.

   (Example) for RHEL6

   [Before change]

   ```
   HOSTNAME=node1
   ```

   [After change]

   ```
   HOSTNAME=nodeA
   ```

   (Example) for RHEL7

   [Before change]

   ```
   node1
   ```

   [After change]

```
nodeA
```

4. Restart OS on the node whose node name is changed.

```
# /sbin/shutdown  -r now
```

After restarting OS, execute the following procedure for the other node.

5. After r restarting the system, change the old host name in the /etc/hosts file on the other node to the new host name.

Note

If the host name is set in the shutdown facility, correct the "/etc/opt/SMAW/SMAWsf/rcsd.cfg" file on each node. For details, see "5.1.2 Setting up the Shutdown Facility."

6. Restart Web-Based Admin View.

See

For information on how to restart Web-Based Admin View, see "PRIMECLUSTER Web-Based Admin View Operation Guide."

9.1.2 Changing the SF Node Weight

The following explains how to change the SF node weight after building a PRIMECLUSTER system.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Modify /etc/opt/SMAW/SMAWsf/rcsd.cfg on each node.

Modify the following weight values.

```
CFNameX,weight=weight,admIP=myadmIP: agent=SA_xxx,timeout=timeout
```

Since the node weight affects the survival priority, see "5.1.2.1 Survival Priority" to determine the value to be set.

3. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

9.2 Changing the Network Environment

This section explains how to change the IP address setting if the IP address of a public LAN or administrative LAN changes after the PRIMECLUSTER system is installed. Note that when you change an IP address, do not change the host name.

Note

If you use the virtual machine function, this section explains the Public / administrative LAN of the guest OS.

9.2.1 Changing the IP Address of the Public LAN

The following describes how to change an IP address when the IP address of the public LAN has been changed.

Operation Procedure:

1. Execute the following command on one of the cluster nodes to stop RMS operation.

```
# hvshut -a
```

2. Execute the following commands on all the nodes to start the system in single-user mode.

   [For RHEL6]

```
# /sbin/shutdown now
```

   [For RHEL7]

```
# /usr/bin/systemctl set-default rescue.target
# /sbin/shutdown -r now
```

3. Mount the local file system on all the nodes.

   (Example)

```
# /bin/mount -a -t ext3
```

4. Edit the "/etc/hosts" file, and change the IP address on each node.

5. Change the IP address of the public LAN.

   For details on how to change the IP address, see the Linux documentation.

6. If the administrative LAN is shared with the public LAN, execute the following command on each node to save rcsd.cfg.

```
# mv /etc/opt/SMAW/SMAWsf/rcsd.cfg /etc/opt/SMAW/SMAWsf/rcsd.cfg.work
```

7. If an IP address used by Web-Based Admin View also needs to be changed as the IP address of the public LAN is changed, change it on each node.

   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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

8. If a takeover IP address must be changed (when the takeover IP address is changed after installation, or when the takeover IP address is changed due to transfer of the node), correct the IP address being used as the takeover IP address in the "/etc/hosts" file of each node.

   When you have created takeover network resources, and change the subnet mask value due to the change of the public LAN, you also need to edit the /usr/opt/reliant/etc/hvipalias file.

   See

   For information on how to edit the /usr/opt/reliant/etc/hvipalias file, see "6.7.3.6 Setting Up Takeover Network Resources."

9. If a Gls resource is created, restrict the automatic startup of RMS so that RMS does not start automatically when the node is restarted.

   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 10.

- If "1" is set, execute the following command to restrict the automatic startup of RMS.

```
# hvsetenv HV_RCSTART 0
# hvsetenv HV_RCSTART
0 <- Check "0" is output.
```

10. Restart OS on all the nodes.

    [For RHEL6]

    ```
    # /sbin/shutdown -r now
    ```

    [For RHEL7]

    ```
    # /usr/bin/systemctl set-default multi-user.target
    # /sbin/shutdown -r now
    ```

11. If the takeover IP address is defined by GLS, see "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

12. If you change the settings that restrict the automatic startup of RMS in Step 9, return the file back to its previous settings.

    ```
    # hvsetenv HV_RCSTART 1
    # hvsetenv HV_RCSTART
    1 <- Check "1" is output.
    ```

    Note

    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/hosts and the GLS environment settings, and re-register the Gls resource.

    For details, see "6.2.1 GLS Setup" and "6.7.3.5 Setting Up Gls Resources" in this manual, and "PRIMECLUSTER Global Link Services Configuration and Administration Guide: Redundant Line Control Function."

13. If rcsd.cfg is saved in Step 6, execute the following command.

    ```
    # mv /etc/opt/SMAW/SMAWsf/rcsd.cfg.work /etc/opt/SMAW/SMAWsf/rcsd.cfg
    ```

    After that, set the shutdown facility again.

    See

    For details, see "5.1.2 Setting up the Shutdown Facility."

9.2.2 Changing the IP Address of the Administrative LAN

The following describes how to change an IP address when the IP address of the administrative LAN has been changed.

Operation Procedure:

1. Execute the following command on one of the cluster nodes to stop RMS operation.

   ```
   # hvshut -a
   ```

2. Execute the following commands on all the nodes to start the system in single-user mode.

   [For RHEL6]

   ```
   # /sbin/shutdown now
   ```

   [For RHEL7]

```
# /usr/bin/systemctl set-default rescue.target
# /sbin/shutdown -r now
```

3. Mount the local file system on all the nodes.

   (Example)

```
# /bin/mount -a -t ext3
```

4. Edit the "/etc/hosts" file, and change the IP address on each node.
5. Change the IP address of the administrative LAN.

   For details on how to change the IP address, see the Linux documentation.

6. If the administrative LAN is shared with the public LAN, execute the following command on each node to save rcsd.cfg.

```
# mv /etc/opt/SMAW/SMAWsf/rcsd.cfg /etc/opt/SMAW/SMAWsf/rcsd.cfg.work
```

7. If an IP address used by Web-Based Admin View also needs to be changed as the IP address of the administrative LAN is changed, change it on each node.

   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 "PRIMECLUSTER Web-Based Admin View Operation Guide."

8. Restart OS on all the nodes.

   [For RHEL6]

```
# /sbin/shutdown -r now
```

   [For RHEL7]

```
# /usr/bin/systemctl set-default multi-user.target
# /sbin/shutdown -r now
```

9. If rcsd.cfg is saved in Step 6, execute the following command.

```
# mv /etc/opt/SMAW/SMAWsf/rcsd.cfg.work /etc/opt/SMAW/SMAWsf/rcsd.cfg
```

   After that, set the shutdown facility again.

   See

   For details, see "5.1.2 Setting up the Shutdown Facility."

9.2.3 Changing a CIP Address

This section describes how to change the IP address if the IP address of interconnect is changed after installation of the PRIMECLUSTER system.

Operation Procedure:

1. Start all the nodes that constitute the cluster system.

   If the nodes are already operating, you do not need to restart them.

2. Stop CF on all the nodes that constitute the cluster system.

   For information on how to stop CF, see "4.6 Starting and stopping CF" in "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 "1.2 CIP configuration file" in "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(5) file.

   Perform this procedure on all the nodes constituting the cluster system.

6. In the environment where the shutdown agent SA_icmp for VMware environment is used, if the cluster interconnect is used to check whether the node is alive or not, modify /etc/opt/SMAW/SMAWsf/SA_icmp.cfg on each node.

   See

   For details, see "H.2.3.3 Setting Up the Shutdown Facility (when using I/O fencing function)."

7. Start CF on all the nodes constituting the cluster system.

   For instructions on how to start CF, see "4.6 Starting and stopping CF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

8. Use the ciptool command to confirm that the IP address of CIP was changed.

```
# /opt/SMAW/SMAWcf/bin/ciptool -a
```

See

For details on the "ciptool" command, see the manual page describing "ciptool".

9.2.4 Changing the Subnet Mask of CIP

To change a subnet mask of CIP, perform the following procedure:

1. Migrate all the 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 the 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 the nodes in multi-user mode.

9.2.5 Changing the MTU Value of a Network Interface Used for Cluster Interconnects

This section describes how to change the MTU value of a network interface used for cluster interconnects.

1. Stop CF on all the nodes that constitute the cluster.

   For information on how to stop CF, see "4.6 Starting and stopping CF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

2. Change the MTU value of a network interface used for cluster interconnects.

3. Start CF on all the nodes that constitute the cluster.

   For information on how to start CF, see "4.6 Starting and stopping CF" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

Note

The MTU value of a network interface used for cluster interconnects must be the same on all the nodes. If there is a different value on a node, the node cannot join the cluster.

9.3 Changing Option Hardware Settings

9.3.1 Changing MMB Settings

This section explains how to change MMB settings.

9.3.1.1 Changing the MMB IP Address

This section explains how to change the MMB IP address.

Note

- Change the MMB IP address on each node.
- Repeat the operation procedure described below to change more than one MMB IP address.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Execute the following command on the node in which IP address is changed to stop MMB asynchronous monitoring daemons.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl stop
```

3. Change the IP address according to the MMB procedure.
4. On the node in which the IP address was changed, execute the following command to start MMB asynchronous monitoring daemons and the shutdown facility.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl start
# sdtool -b
```

5. After the shutdown facility started in Step 4, start the shutdown facility on the remaining nodes.

```
# sdtool -b
```

6. Execute the following command on all the nodes and check that the shutdown facility operates normally.

```
# sdtool -s
```

Note

There is a possibility that the hardware settings are not correct when any of the following statuses are displayed though changing the settings of the shutdown facility is completed.

- InitFailed is displayed in Init State.

- Unknown or TestFailed is displayed in Test State.

In this case, review the hardware settings.

9.3.1.2 Changing the User Name and Password for Controlling the MMB with RMCP

This section explains how to change the user name and password for controlling the MMB with RMCP.

Operation Procedure

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. According to the procedures of MMB, change the user name and password to control MMB by RMCP. If you change the user name and password for several nodes, change them for all the nodes.
3. By executing the following command, change the user name and password of MMB information for MMB shutdown facility. If the user name and the password are to be changed on multiple nodes, change the values on all the nodes to be changed.

```
# /etc/opt/FJSVcluster/bin/clmmbsetup -m [-u user-name] [-p]
```

4. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

9.3.2 Changing IPMI Settings

This section describes changing IPMI settings.

9.3.2.1 Changing the IPMI IP Address

This section explains how to change the IPMI IP address.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Change the IP address according to the IPMI procedure.
3. Define the changed IP address in the Shutdown Agent configuration file.

For details on how to define the configuration file, see "5.1.2.3.3 Setting up IPMI Shutdown Agent."

4. Execute the following command on any node to apply changes of the configuration file.

```
# /etc/opt/FJSVcllkcd/bin/panicinfo_setup
```

After the following message is displayed, select "I."

```
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
```

5. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

6. Execute the following command on all the nodes and check that the shutdown facility operates normally.

```
# sdtool -s
```

Note

There is a possibility that the settings of the agent or hardware are not correct when any of the following statuses are displayed though changing the settings of the shutdown facility is completed.

- InitFailed is displayed in Init State.
- Unknown or TestFailed is displayed in Test State.

In this case, review the settings of the agent or hardware.

9.3.2.2 Changing the User Name and Password for iRMC

This section explains how to change the user name and password for iRMC.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Change the user name and password according to the procedure for IPMI.
3. Encrypt the password.

```
# /opt/SMAW/SMAWsf/bin/sfcipher -c
Enter Password:
Re-Enter Password:
D0860AB04E1B8FA3
```

4. Define the changed user name and the encrypted password for iRMC in the Shutdown Agent configuration file.

See

For details on how to define the configuration file, see "5.1.2.3.3 Setting up IPMI Shutdown Agent."

5. Execute the following command on any node to apply changes of the configuration file.

```
# /etc/opt/FJSVcllkcd/bin/panicinfo_setup
```

After the following message is displayed, select "I."

```
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
```

6. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

7. Execute the following command on all the nodes and check that the shutdown facility operates normally.

```
# sdtool -s
```

Note

There is a possibility that the settings of the agent or hardware are not correct when any of the following statuses are displayed though changing the settings of the shutdown facility is completed.

- InitFailed is displayed in Init State.
- Unknown or TestFailed is displayed in Test State.

In this case, review the settings of the agent or hardware.

9.3.3 Changing Blade Settings

This section describes changing Blade settings.

9.3.3.1 Changing the IP Address of the Management Blade

This section explains how to change the IP address of the management blade.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Change the IP address according to the procedure of the management blade.
3. Define the changed IP address in the Shutdown Agent configuration file.

See

For details on how to define the configuration file, see "5.1.2.3.4 Setting up Blade Shutdown Agent."

4. Execute the following command on any node to apply changes of the configuration file.

```
# /etc/opt/FJSVcllkcd/bin/panicinfo_setup
```

After the following message is displayed, select "I."

```
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
```

5. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

6. Execute the following command on all the nodes and check that the shutdown facility operates normally.

```
# sdtool -s
```

Note

There is a possibility that the settings of the agent or hardware are not correct when any of the following statuses are displayed though changing the settings of the shutdown facility is completed.

- InitFailed is displayed in Init State.
- Unknown or TestFailed is displayed in Test State.

In this case, review the settings of the agent or hardware.

9.3.3.2 Changing the Slot Number of Server Blades

This section explains how to change the slot number of server blades.

Operation Procedure:

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Change the slot position according to procedure for the server blade.

3. Define the changed slot number of the server blade in the Shutdown Agent configuration file.

See

For details on how to define the configuration file, see "5.1.2.3.4 Setting up Blade Shutdown Agent."

4. Execute the following command on any node to apply changes of the configuration file.

```
# /etc/opt/FJSVcllkcd/bin/panicinfo_setup
```

After the following message is displayed, select "I."

```
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ?
```

5. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

6. Execute the following command on all the nodes and check that the shutdown facility operates normally.

```
# sdtool -s
```

Note

There is a possibility that the settings of the agent or hardware are not correct when any of the following statuses are displayed though changing the settings of the shutdown facility is completed.

- InitFailed is displayed in Init State.
- Unknown or TestFailed is displayed in Test State.

In this case, review the settings of the agent or hardware.

9.4 Changing Virtual Machine Settings

This section describes how to change the VMGuest settings when you have changed the Host OS setting.

9.4.1 Changing Host OS Settings (KVM environment)

This section describes how to change the settings of the shutdown facility when changing the settings of the host OS in the environment where the KVM virtual machine function is used.

9.4.1.1 Changing the IP address of the Host OS

The following describes how to change the settings when you have changed the host OS IP address of the virtual machine after introducing the PRIMECLUSTER system into a KVM environment.

Operation procedure

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Define the changed IP address in the Shutdown Agent configuration file.

See

For details on how to define the configuration file, see "5.1.2.5.2 Setting up libvirt Shutdown Agent."

3. For the host OS IP addresses (ip-address) you want to change, log in as a shutdown facility user on all guest OSes (nodes) in advance, as you need to authenticate yourself (create the RSA key), which is required when using SSH for the first time.

```
# ssh -l user 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"
```

4. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

9.4.1.2 Changing the Password of the Host OS Account (Shutdown Facility)

The following describes how to change the settings when you have changed the password for the user for the shutdown facility set in the Shutdown Agent configuration file after introducing the PRIMECLUSTER system into a KVM environment.

Operation procedure

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Encrypt the password.

```
# /opt/SMAW/SMAWsf/bin/sfcipher -c
Enter Password:
Re-Enter Password:
Xh+kSlJ8nlQ=
```

3. Define the encrypted password in the Shutdown Agent configuration file.

See

For details on how to define the configuration file, see "5.1.2.5.2 Setting up libvirt Shutdown Agent."

4. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

9.4.1.3 Changing the Settings in /etc/sysconfig/libvirt-guests

This section explains the procedure for changing the settings in /etc/sysconfig/libvirt-guests after installing the PRIMUCLUSTER system in a KVM environment.

Operation procedure

1. Execute the following command on all the nodes to stop the shutdown facility.

```
# sdtool -e
```

2. Change the settings in /etc/sysconfig/libvirt-guests.

See

For details on the settings in /etc/sysconfig/libvirt-guests, see "Setting the guest OS in the host OS (in a KVM environment)" for each virtual environment shown below:

- When building a cluster system between guest OSes on one host OS, see "3.2.1.2 Host OS setup (after installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes without using Host OS failover function, see "3.2.2.2 Host OS setup (after installing the operating system on guest OS)."
- When building a cluster system between guest OSes on multiple host OSes using Host OS failover function, see "3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)."

3. Execute the following command on all the nodes to start the shutdown facility.

```
# sdtool -b
```

Chapter 10 Configuration change of Cluster Applications

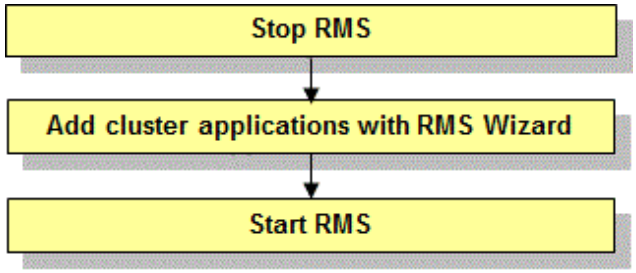
This chapter describes how to change the configuration of cluster applications.

Before adding the cluster application or the resource, check "Design (the number of resources)" of PRIMECLUSTER Designsheets to verify that the number of resource objects and the number of detectors that can be set in the whole PRIMECLUSTER system do not exceed their maximum values.

10.1 Adding cluster applications

This chapter explains how to add cluster applications.

Operation flow



Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Add cluster applications with the RMS Wizard.

   1. Log in to any one of the cluster nodes using system administrator access privileges.

   2. Start up the RMS Wizard.

      Execute the "hvw -n *configuration* file" command. Specify the name of the configuration file in which the configuration is defined.

      The following example shows how to start up RMS Wizard with the configuration file name "testconf."

      ```
      # /opt/SMAW/SMAWRrms/bin/hvw -n testconf
      ```

   3. Set up the userApplication that you want to add.

      Set up the userApplication that you want to add, and register the resources as described in "6.7.2 Setting Up userApplication" and "6.7.3 Setting Up Resources."

3. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

4. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

5. Select "QUIT" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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
```

6. Check the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

7. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

8. Start RMS.

   Start RMS as described in "7.2.1.1 Starting RMS."

10.2 Deleting a Cluster Application

This section explains how to delete a cluster application and its resources.

Note

Be sure to stop RMS of all the nodes before deleting a cluster application and its resources. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

10.2.1 Deleting the Hardware Resource

This section explains how to delete the resources of the following hardware.

- Shared disk device
- Network interface card

Procedure

1. Stop RMS of all the nodes.

   If RMS is activated, stop RMS as explained in "7.2.1.2 Stopping RMS."

2. Delete the hardware resource.

   Use "cldelrsc" command to delete the hardware resource that was registered.

   See the relevant manual pages for details on this command.

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" command.
- If the shared disk for which resources are to be deleted is registered to a GDS class, first delete the shared disk from the GDS class, and then delete the resources of the shared disk. For instructions on how to delete a shared disk from a GDS class, refer to "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

10.2.2 Deleting a userApplication

This section explains how to delete the userApplication.

Note

- If you delete a userApplication, all the resources registered to the userApplication will also be deleted.
- If Gds resources are registered to the userApplication to be deleted, bring the Gds volume online. See "10.5.1 Settings made when deleting a Gds resource."

Operation Procedure:

1. Log in to any one of the cluster nodes using system administrator access privileges.
2. Start the RMS Wizard.

   Execute the "hvw -n *configuration file*" command. Specify a name of the configuration file in which the userApplication is defined.

The following example shows how to start RMS Wizard with the configuration file name "testconf."

```
# /opt/SMAW/SMAWRrms/bin/hvw -n testconf
```

3. Select "Application-Remove" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 5
```

4. Select the userApplication that you want to delete from the "Application selection menu."

   The following example shows how to select APP2.

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
6) APP2
Application Name: 6
```

   Enter "yes" in response to the following message.

```
  About to remove all data of APP2,
      Please confirm this by typing yes: yes
```

Note

When deleting a cluster application that is performing standby operation as a component of the cluster application in scalable operation, change the resources of the Controller after deleting the cluster application that is performing standby operation. For details on how to change the resource of the Controller, see "10.3 Changing a Cluster Application."

5. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

6. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

7. Select "QUIT" from the "Main configuration menu" to exit from the RMS Wizard.

```
node1: Main configuration menu, current configuration: testconf
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
```

If all userApplications are deleted, you do not have to take the remaining steps.

8. Check the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

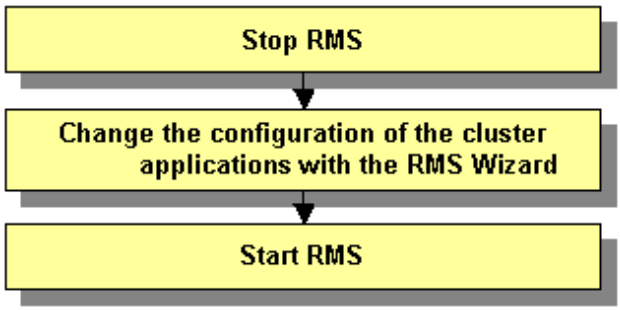
9. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

10.3 Changing a Cluster Application

This section explains how to modify a cluster application by following operations:

Operation flow



Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Change the configuration of the cluster applications with the RMS Wizard.

   1. Log in to any one of the cluster nodes using system administrator access privileges.
   2. Start up the RMS Wizard.

      Execute the "hvw -n *configuration* file" command. Specify the name of the configuration file in which the configuration is defined.

      The following example shows how to start up RMS Wizard with the configuration file name "testconf."

      ```
      # /opt/SMAW/SMAWRrms/bin/hvw -n testconf
      ```

   3. Select "Application-Edit" from the "Main configuration menu."

      ```
      node1: Main configuration menu, current configuration: testconf
      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
      ```

      When a cluster application that is performing standby operation is to be changed

      1. Select the userApplication that needs modification of configuration from "Application selection menu." If more than one selection item is displayed, select userApplication written in capital letters. The following example shows how to select "APP1."

         ```
         Edit: Application selection menu (restricted):
         1) HELP
         2) QUIT
         3) RETURN
         4) OPTIONS
         5) APP1
         Application Name: 5
         ```

2. When "turnkey wizard" appears, select what you want to change from the following table.

Contents to be changed	Selection item
Identifier used as userApplication name and resource name (The cluster application name and the cluster resource name are specified based on ApplicationName.)	"ApplicationName"
Attribute of the userApplication	"Machines+Basics"
Cmdline resource configuration	"CommandLines"
Fsystem resource configuration	"LocalFileSystems"
Gds resource configuration	"Gds:Global-Disk-Services"
Gls resource configuration	"Gls:Global-Link-Services"
Procedure resource configuration	"Procedure:XXXXXX"

For details on the operation when you select above items, see "6.7 Setting Up Cluster Applications." After you change the configuration, select "SAVE+EXIT" to return to the "Application selection menu." After that, select "RETURN" to return to the "Main configuration menu."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) READONLY                         10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(Cmd_APP1)           15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:
```

The following example shows how to change the attribute of "AutoStartUp" of the userApplication setting from "no" to "yes":

1. Select "Machines+Basics" from "turnkey wizard."

2. Select "AutoStartUp."

```
Machines+Basics (appl:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=0)
26) (ScriptTimeout=)
Choose the setting to process:
```

3. Select "yes."

```
1) HELP
2) RETURN
3) yes
4) no
Set the AutoStartUp mode: 3
```

4. Confirm that "AutoStartUp" is changed to "yes," and then select "SAVE+EXIT."

```
Machines+Basics (appl:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=yes)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=0)
26) (ScriptTimeout=)
Choose the setting to process:3
```

5. Select "SAVE+EXIT" from "turnkey wizard."

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) -                                10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(Cmd_APP1)           15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:3
```

6. Select "RETURN" from the "Application selection menu."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 3
```

When a cluster application in a scalable operation is to be changed

Note

For information on how to change a cluster application performing standby operation and which forms part of a cluster application in a scalable operation, see "When a cluster application that is performing standby operation is to be changed."

1. Select the userApplication to be reconfigured from "Application selection menu." If more than one selection item is displayed, select userApplication written in capital letters. The following example shows how to select "APP3."

```
Edit: Application selection menu (restricted):
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) APP1
 6) APP2
 7) APP3
Application Name: 7
```

2. Select "Controllers" from the "turnkey wizard SCALABLE" menu.

```
Settings of turnkey wizard "SCALABLE" (APP3:consistent)
 1) HELP            4) -                      7) Controllers(Ctl_APP3)
 2) -               5) ApplicationName=APP3
 3) SAVE+EXIT       6) Machines+Basics(app3)
Choose the setting to process:7
```

3. "Settings of application type "Controller"" is displayed. Select one of the following according to the contents to be changed:

Contents to be changed	Resource to be selected
Start sequence of the cluster application (standby operation)	"ApplicationSequence"
Deletion of the application (standby operation)	"Controllers[*]"

[Supplement]

A number is specified in the "*" mark included in "Controllers[*]". Select the cluster application in a standby operation that you want to delete. You can delete a cluster application in a standby operation by specifying "NONE" on the screen after the selection.
For details on the operation to be performed after making the above selection, see "6.7 Setting Up Cluster Applications." After you change the configuration, select "SAVE+EXIT" to return to the "Application selection menu." After that, select "RETURN" to return to the "Main configuration menu."

The following is an example in which the "AutoStartUp" attribute of the userApplication is changed to "yes" from "no."

1. Select "Machines+Basics" from the "turnkey wizard "SCALABLE"" menu.

```
Settings of turnkey wizard "SCALABLE" (APP3:consistent)
 1) HELP            4) -                      7) Controllers(Ctl_APP3)
 2) -               5) ApplicationName=APP3
 3) SAVE+EXIT       6) Machines+Basics(app3)
Choose the setting to process:6
```

2. Select "(AutoStartUp=no)" from the "Machines+Basics" menu.

```
Machines+Basics (app3:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ShutDown)
17) (PreserveState=yes)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=0)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=1)
26) (ScriptTimeout=)
Choose the setting to process:15
```

3. Select "yes."

```
 1) HELP
 2) RETURN
 3) yes
 4) no
Set the AutoStartUp mode:3
```

4. Check that "AutoStartUp" has been changed to "yes," and then select "SAVE+EXIT."

```
Machines+Basics (app3:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=yes)
16) (AutoSwitchOver=HostFailure|ShutDown)
17) (PreserveState=yes)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=0)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=1)
26) (ScriptTimeout=)
Choose the setting to process: 3
```

5. Select "SAVE+EXIT" from the "turnkey wizard "SCALABLE"" menu.

```
Settings of turnkey wizard "SCALABLE" (APP3:consistent)
 1) HELP           4) -                    7) Controllers(Ctl_APP3)
 2) -              5) ApplicationName=APP3
 3) SAVE+EXIT      6) Machines+Basics(app3)
Choose the setting to process:3
```

6. Select "RETURN" from the "Application selection menu."

```
Edit: Application selection menu (restricted):
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) APP1
 6) APP2
 7) APP3
Application Name:3
```

3. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

4. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

5. Select "QUIT" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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
```

6. Check the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

7. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.
   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

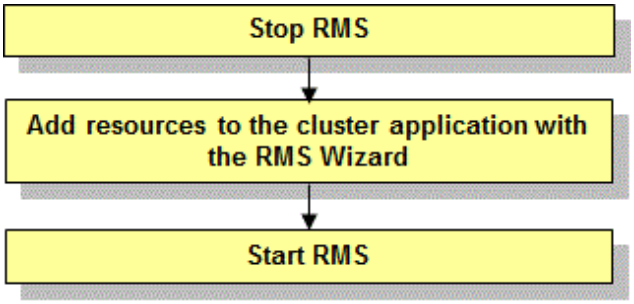
8. Start RMS.

   Start RMS as described in "7.2.1.1 Starting RMS."

10.4 Add Resources

This chapter describes the method for adding resources to the cluster applications.

Operation flow



Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Register the new resources to the cluster application with the RMS Wizard.

   1. Log in to any one of the cluster nodes using system administrator access privileges.
   2. Start up the RMS Wizard.

      Execute the "hvw -n configuration file" command. Specify the name of the configuration file in which the configuration is defined.

      The following example shows how to start up RMS Wizard with the configuration file name "testconf."

      ```
      # /opt/SMAW/SMAWRrms/bin/hvw -n testconf
      ```

   3. Select "Application-Edit" from the "Main configuration menu."

      ```
      node1: Main configuration menu, current configuration: testconf
      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
      ```

   4. Select a registered userApplication for adding resources from the "Application selection menu."

      The following example shows how to select "APP1."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 5
```

5. Register the added resources.

   See "6.7.3 Setting Up Resources" and register the added resources.

3. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

4. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

5. Select "QUIT" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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
```

6. Check the cluster service for the PRIMECLUSTER-compatible product.

   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

7. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.

   Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

8. Start RMS.

   Start RMS as described in "7.2.1.1 Starting RMS."

10.5 Deleting a Resource

This section explains how to delete a resource.

Note

- When deleting a Gds resource, you need to bring the Gds volume Online.
  See "10.5.1 Settings made when deleting a Gds resource."
- When deleting a procedure resource, first delete the procedure resource from the cluster resource management facility after deleting the procedure resource from the cluster application. For details on how to delete a procedure resource from the cluster resource management facility, see "D.3 Deleting a Procedure Resource."
- When deleting an Fsystem resource, delete the mount point that was being used as the resource (mount point of the line beginning with "#RMS#") from /etc/fstab.pcl on all the nodes.
- When deleting takeover network resource, delete entries added at the time of setting up takeover network resource from the following environment files:
  - /usr/opt/reliant/etc/hvipalias
  - /etc/hosts
- To delete the resource (Gds resource or Fsystem resource) that controls the shared disk in the VMware environment where the I/O fencing function is used, make sure that userApplication is Offline on all the nodes before stopping RMS.
  If an error such as a resource failure or an OS panic has occurred right before stopping RMS, take the following steps first and then delete the resource:

  1. Remove the cause of a fault or an error.

2. Start userApplication once and then stop it.
3. Make sure that userApplication stopped in step 2 becomes Offline successfully.

Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Log in to any one of the cluster nodes using system administrator access privileges.
3. Start the RMS Wizard.

   Execute the "hvw -n *configuration file*" command. Specify the name of the configuration file in which the resource is defined.

   The following example shows how to start RMS Wizard with the configuration file name "testconf."

```
# /opt/SMAW/SMAWRrms/bin/hvw -n testconf
```

4. Select "Application-Edit" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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
```

5. Select the userApplication in which the resource is registered from the "Application selection menu." The following example shows how to select "APP1."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 5
```

6. Use the "turnkey wizard" to select the resource.

   To delete all the same type of resources, select "REMOVE+EXIT" from the screen displayed after selecting the resource.
   To delete only some of the same type resources, select and delete the target resources one by one. After that, select "SAVE+EXIT". After that, select "SAVE+EXIT" and go back to "Application selection menu". Then, select "Return" to go back to "Main configuration menu".

```
Settings of turnkey wizard "STANDBY" (APP1:consistent)
 1) HELP                          9) Procedure:BasicApplication(-)
 2) -                            10) Symfoware(-)
 3) SAVE+EXIT                    11) Procedure:SystemState3(-)
 4) -                            12) Procedure:SystemState2(-)
 5) ApplicationName=APP1         13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)        14) IpAddresses(-)
 7) CommandLines(Cmd APP1)       15) LocalFileSystems(-)
 8) Procedure:Application(-)     16) Gds:Global-Disk-Services(-)
Choose the setting to process:
```

7. Select "Configuration-Generate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 7
```

8. Select "Configuration-Activate" from the "Main configuration menu."

```
node1: Main configuration menu, current configuration: testconf
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: 8
```

9. Select "QUIT" from the "Main configuration menu" to exit from the RMS Wizard.

```
node1: Main configuration menu, current configuration: testconf
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
```

10. Check the cluster service for the PRIMECLUSTER-compatible product.
    Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

11. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.
    Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

10.5.1 Settings made when deleting a Gds resource

When the Gds resource is to be deleted, you must also notify the cluster resource management facility (CRM) that the Gds resources can no longer be used for the cluster application.

Operation Procedure:

Execute the following command on the node on which you want to delete resources.

```
# /opt/SMAW/SMAWRrms/bin/hvgdsetup -d [class-name]
```

10.6 Changing Resources

This section explains how to change the resources used by the cluster application.

There are the following situations for changing resources.

- Changing host name and IP address of takeover network resource
- Changing devices of file systems controlled by the Fsystem resource

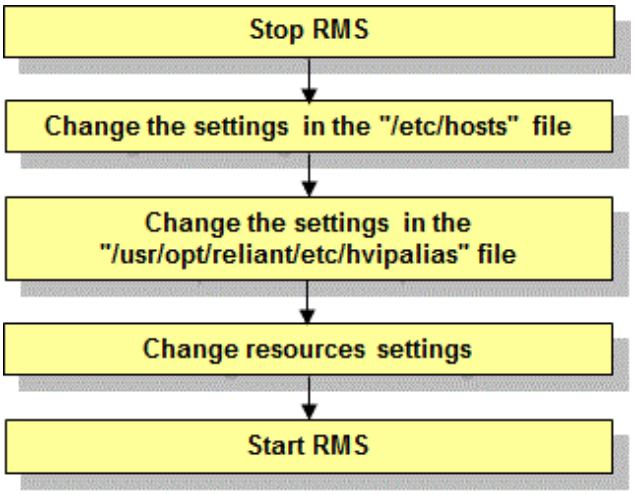
Point

It is possible to change the resources only when RMS is stopped.

10.6.1 Changing Host Names and IP Addresses of Takeover Network Resource

This section explains the procedure for changing the host name and the IP address of the takeover network resource.

Operation flow



Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Change the settings in the "/etc/hosts" file.

   Change the host name and IP address in the "/etc/hosts" file on all the nodes that use takeover network resource.

3. Change the settings in the "/usr/opt/reliant/etc/hvipalias" file.

   See "6.7.3.6 Setting Up Takeover Network Resources" and change the settings in the "hvipalias" file on all the nodes that use takeover network resource.

   There are two items that may need to be changed as follows.

   ```
   <node name> : When the host name is changed, the CF node name also needs to be changed.
                 Change the value of this field to the modified CF node name.
   <takeover>  : Change this host name when the host name associated with the takeover IP address
                 was changed.
   ```

4. Change the setting with the RMS Wizard.

   Execute the "hvw" command in any node that is part of the cluster system.

   Change the host name settings for the takeover network resource from the GUI menu displayed in the RMS Wizard.

   Note

   For changing only the IP addresses of takeover network resource but not the host names, it is not necessary to use the RMS Wizard.

   See

   For details on changing settings with the RMS Wizard, see "8.5 Changing the Operation Attributes of a userApplication."

   1. Select "Application-Edit" from the "Main configuration menu."

      ```
      node1: 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
```

2. Select the userApplication that needs modification of the configuration from the "Application selection menu."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 5
```

3. Select "IpAddresses" in "turnkey wizard."

```
Consistency check ...

Settings of turnkey wizard "STANDBY" (APP1:consistent)
 1) HELP                                 9) Procedure:BasicApplication(-)
 2) READONLY                            10) Symfoware(-)
 3) SAVE+EXIT                           11) Procedure:SystemState3(-)
 4) -                                   12) Procedure:SystemState2(-)
 5) ApplicationName=APP1                13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)               14) IpAddresses(Adr_APP1)
 7) CommandLines(-)                     15) LocalFileSystems(-)
 8) Procedure:Application(-)            16) Gds:Global-Disk-Services(-)
Choose the setting to process: 14
```

4. Select Interfaces[X] to set the host name to be changed from the "Ipaddresses and ipaliases menu."

```
Consistency check ...

Ipaddresses and ipaliases (Adr_APP1:consistent)
 1) HELP                                 7) Interfaces[0]=VAProuter,l3hub:takeover1
 2) NO-SAVE+EXIT                         8) PingHostPool[0]=router
 3) SAVE+EXIT                            9) PingHostPool[1]=l3hub
 4) REMOVE+EXIT                         10) (NeedAll=yes)
 5) AdditionalInterface                 11) (Timeout=60)
 6) AdditionalPingHost                  12) (InterfaceFilter=)
Choose the setting to process: 7
```

5. Select the changed host name associated with the takeover IP address.

```
 1) HELP                  6) node2RMS
 2) RETURN                7) takeover2
 3) NONE
 4) FREECHOICE
 5) node1RMS
Choose an interface name: 7
```

6. Select "SAVE+RETURN."

```
Set flags for interface: takeover2
Currently set: VIRTUAL,AUTORECOVER,PING (VAProuter,l3hub)
1) HELP                  4) DEFAULT               7) MONITORONLY(M)
2) -                     5) BASE(B)               8) NOT:PING(P)
3) SAVE+RETURN           6) NOT:AUTORECOVER(A)
Choose one of the flags: 3
```

7. Make sure that the changed host name is displayed in Interfaces[X] in the "Ipaddresses and ipaliases menu."

```
Ipaddresses and ipaliases (Adr_APP1:consistent)
 1) HELP                                7) Interfaces[0]=VAProuter,l3hub:takeover2
 2) NO-SAVE+EXIT                        8) PingHostPool[0]=router
 3) SAVE+EXIT                           9) PingHostPool[1]=l3hub
 4) REMOVE+EXIT                        10) (NeedAll=yes)
 5) AdditionalInterface                11) (Timeout=60)
 6) AdditionalPingHost                 12) (InterfaceFilter=)
Choose the setting to process:
```

8. If you have to change multiple objects, repeat Steps 4. to 7. for each object. After completing all changes, select "SAVE +EXIT."

9. Select "SAVE+EXIT" in "turnkey wizard."

```
Settings of turnkey wizard "STANDBY" (APP1:consistent)
 1) HELP                                9) Procedure:BasicApplication(-)
 2) -                                  10) Symfoware(-)
 3) SAVE+EXIT                          11) Procedure:SystemState3(-)
 4) -                                  12) Procedure:SystemState2(-)
 5) ApplicationName=APP1               13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)              14) IpAddresses(Adr_APP1)
 7) CommandLines(-)                    15) LocalFileSystems(-)
 8) Procedure:Application(-)           16) Gds:Global-Disk-Services(-)
Choose the setting to process: 3
```

10. Select "RETURN" on "Application selection menu."

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 3
```

11. Select "Configuration-Generate" and "Configuration-Activate" on "Main configuration menu."

    See "6.7.4 Generate and Activate" for information on Configuration-Generate and Configuration-Activate.

12. Select "QUIT" to exit from the RMS Wizard.

5. Start RMS.

   Start RMS as described in "7.2.1.1 Starting RMS."

10.6.2 Changing the Devices of File systems Controlled by the Fsystem Resource

This section explains how to change devices of file systems controlled by the Fsystem resource.

Note

In the VMware environment where the I/O fencing function is used, make sure that userApplication is Offline on all the nodes before stopping RMS.

If an error such as a resource failure or an OS panic has occurred right before stopping RMS, take the following steps first and then change the device:

1. Remove the cause of a fault or an error.
2. Start userApplication once and then stop it.

3. Make sure that userApplication stopped in step 2 becomes Offline successfully.

Operation Procedure:

1. Stop RMS of all the nodes.

   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.

2. Change the settings in the "/etc/fstab.pcl" file.

   See "6.7.3.2 Setting Up Fsystem Resources" and change the device that is described in the first field of the "/etc/fstab.pcl" file.

   When you do this, do not change "#RMS#" at the beginning.

3. Execute the following procedures described in "6.7.3.2 Setting Up Fsystem Resources."

   - Confirm that the file system can be mounted.
   - Carry out tuning of the file system.

4. Start RMS.

   Start RMS as described in "7.2.1.1 Starting RMS."

10.7 Adding file system to the shared disk by Dynamic Changing Configuration

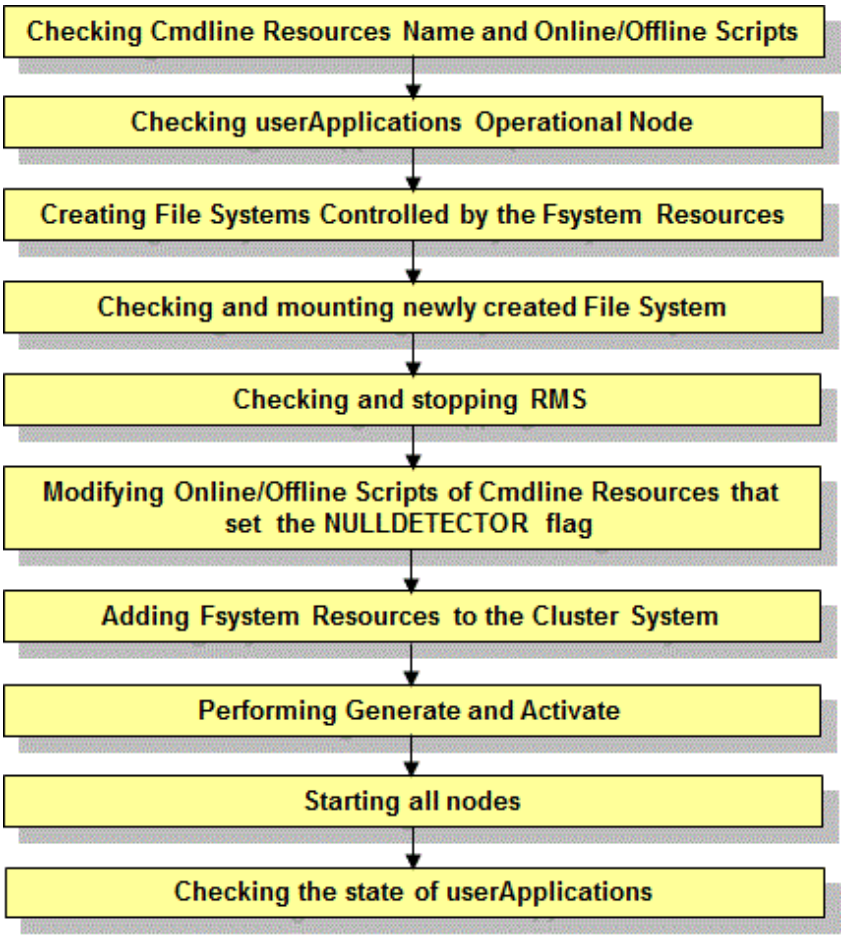
This section describes add Fsystem resources without stopping jobs.

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



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 "6.11.2.1.4 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 starting the volume of GDS and creating file system, see "6.7.3.2 Setting Up Fsystem Resources."

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

According to the following Step 8, specify an example to add the following line to the /etc/fstab.pcl file.

```
#RMS#/dev/sfdsk/class0001/dsk/volume0004 /mnt/swdsk4 ext3 noauto 0 0
```

Execute the command below in the operational node to mount file system.

```
# /sbin/mount -t ext3 /dev/sfdsk/class0001/dsk/volume0004 /mnt/swdsk4
```

After mounting, execute the command below to check that if the mount point is displayed (if the file system is mounted).

```
# /sbin/mount | /bin/grep "/mnt/swdsk4 "
/dev/sfdsk/class0001/dsk/volume0004 on /mnt/swdsk4 type ext3 (rw)
```

Additionally, check that the file system is not mounted on the standby node.

5. Stop RMS.

   Execute the hvshut -L command on all the 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
```

6. Check the stop of RMS.

   Execute the hvdisp -a command on all the 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 "6.11.2.1.4 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 "6.7.3.2 Setting Up Fsystem Resources."

   1. Defining mount point

   4. Tuning of file system

   6. Registering cluster application of Fsystem resources

   When the mount point controlled by the Fsystem resource is created on the new class of GDS, execute it based on the procedures described in "6.7.3.3 Preliminary Setup for Gds Resources" and "6.7.3.4 Setting Up Gds Resources."

9. Perform Generate and Activate.

   For details of performing Generate and Activate, See the procedure of "6.7.4 Generate and Activate."

10. Start RMS on all the nodes.

    Execute the hvcm - a command on any one node to start RMS on all the nodes.

```
# hvcm -a
```

11. Check the state of userApplications.

    Execute the hvdisp -a command on all the 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 10 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.

Chapter 11 Changing the Operation Attributes of a Cluster System

11.1 Changing the Operation Attributes of a userApplication

PRIMECLUSTER allows operation attributes to be set by CUI operation, according to the needs of the user. Change the operation attributes to match the desired operation.

- Operation attribute types

  For further details about the operation attribute of the userApplication, see "12 Appendix - Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

- Change methods

  - "11.1.1 Changing the Operation Attributes (CUI)"

    Explains how to change the operation attributes of the userApplication.

  - "11.2 Changing the RMS Environment Variables "

    Explains how to change the RMS environment variables.

Note

Be sure to stop RMS before you change the operation attributes of userApplication. For instructions on stopping RMS, see "7.2.1.2 Stopping RMS."

11.1.1 Changing the Operation Attributes (CUI)

This section explains how to change the userApplication attributes with CUI.

For further details about the operation attribute specified in step 8, see "12 Appendix - Attributes" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

Operation Procedure:

Note

"Application" on the CUI screen indicates a cluster application.

1. Log in to any one of the cluster nodes using system administrator access privileges.
2. Stop RMS.
   If RMS is running, see "7.2.1.2 Stopping RMS" and stop RMS.
3. Execute the "hvw" command.

   ```
   # /opt/SMAW/SMAWRrms/bin/hvw -n <configuration file>
   ```

   Specify an optional Configuration file name in <*configuration file*>.

4. Select "Application-Edit" from the main menu of CUI. Enter a number and then press the *Enter* key.

Figure 11.1 Main menu

```
apple: 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:
```

5. Select the userApplication for which you want to change the operation attributes from the "Application selection menu."

   The following example shows how to select "APP1."

Figure 11.2 Cluster application selection menu

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 5
```

6. When turnkey wizard STANDBY appears, select "Machines+Basics" and then change the operation attributes of the userApplication.

Figure 11.3 turnkey wizard "STANDBY"

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                         9) Procedure:BasicApplication(-)
 2) -                           10) Symfoware(-)
 3) SAVE+EXIT                   11) Procedure:SystemState3(-)
 4) -                           12) Procedure:SystemState2(-)
 5) ApplicationName=APP1        13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)       14) IpAddresses(-)
 7) CommandLines(Cmd APP1)      15) LocalFileSystems(-)
 8) Procedure:Application(-)    16) Gds:Global-Disk-Services(-)
Choose the setting to process:6
```

7. Select the operation attribute that you want to change from "Machines+Basics."

Figure 11.4 Machines+Basics

```
Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=fuji01RMS
 8) Machines[1]=fuji02RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=no)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (HaltFlag=no)
25) (PartialCluster=0)
26) (ScriptTimeout=)
Choose the setting to process:
```

To change the value:
Select the item that you want to change. Enter a number and then press the *Enter* key.
(Example)
Choose the setting to process: **20** <RETURN>
The "Value setup menu," as shown in step 8, will be displayed.

Select "SAVE+EXIT." Enter a number and then press the *Enter* key.
Choose the setting to process: **3** <RETURN>
"turnkey wizard "STANDBY"" of step 9 will be displayed.

8. Set up the value from the value setup menu.

Figure 11.5 Value setup menu (Example of OnlinePriority)

```
1) HELP
2) RETURN
3) 0
4) 1
Enable Online Priority (Active Standby) feature for this application:
```

Select a setup value. Enter a number and then press the *Enter* key.

Point

Select "RETURN" to return to the previous menu.

If there are multiple attributes to be changed, repeat steps 7 and 8 for each attribute.

If the attribute is other than "OnlinePriority," the menu number in step 8 will be different from that in this example.

9. Select "SAVE+EXIT" from the "Machines+Basics" screen to return to the "turnkey wizard "STANDBY."

Figure 11.6 turnkey wizard "STANDBY"

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          9) Procedure:BasicApplication(-)
 2) -                            10) Symfoware(-)
 3) SAVE+EXIT                    11) Procedure:SystemState3(-)
 4) -                            12) Procedure:SystemState2(-)
 5) ApplicationName=APP1         13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)        14) IpAddresses(-)
 7) CommandLines(Cmd APP1)       15) LocalFileSystems(-)
 8) Procedure:Application(-)     16) Gds:Global-Disk-Services(-)
Choose the setting to process:
```

Select "SAVE+EXIT" to return to the cluster application selection menu.

10. Select "RETURN" to return to the main menu.

Figure 11.7 Cluster application selection menu

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 3
```

11. Select "Configuration-Generate" and then "Configuration-Activate" from the main menu.
Content changes will be enabled on all the cluster nodes.

Figure 11.8 Main menu

```
apple: 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:
```

```
apple: 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:
```

Figure 11.9 Configuration distribution (Example of executing Configuration-Activate)

```
About to activate the configuration config ...
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..................................... done
hvbuild using /usr/opt/reliant/build/wizard.d/config/config.us
About to distribute the new configuration data to hosts:
test node1RMS,test node2RMS
The new configuration was distributed successfully.
About to put the new configuration in effect ... done.
The activation has finished successfully.
Hit CR to continue
```

When the processing is successfully done, the message "The activation has finished successfully" appears. If this message is not displayed, the modified information contains incorrect settings. Check and correct the settings.

12. Press the *Enter* key to return to the main menu.

Figure 11.10 Main menu

```
apple: 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:
```

13. Select "QUIT" to terminate the processing.

14. Check the cluster service for the PRIMECLUSTER-compatible product.
    Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig -c
```

15. If the results of the cluster service check for the PRIMECLUSTER-compatible product shows that the "clrwzconfig" command output message 8050, re-register the cluster service for the PRIMECLUSTER-compatible product.
    Execute the following command in any node that is part of the cluster system.

```
# /etc/opt/FJSVcluster/bin/clrwzconfig
```

16. Start up RMS and userApplication from Cluster Admin.

Information

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."

11.2 Changing the RMS Environment Variables

The environment variables are configured in the following RMS environment files.

You can change the environment variables by editing /opt/SMAW/SMAWRrms/bin/hvenv.local of the RMS environment files and set or change the values.

See

- For details 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 Time 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 "PRIMECLUSTER 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 "PRIMECLUSTER 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 the 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 the 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 both nodes.
   # **hvshut -a**

2. Change /usr/opt/reliant/etc/CONFIG.rms on both 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 both nodes.
   # **hvcm -a**

4. Check if RMS has started with the option specified in Step 2.
   # **hvdisp -h**
   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.

Chapter 12 Maintenance of the PRIMECLUSTER System.....347

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.

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:

When stop maintenance is to be performed

1. All the nodes of the running PRIMECLUSTER system shall be stopped by the 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 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" 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" 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(8)" 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 start up the cluster application, see "7.2.2.1 Starting a Cluster Application." For details on how to execute failover/failback, see "7.2.2.3 Switching a Cluster Application."

12.3 Software Maintenance

This section provides notes on intensive correction and formal repair, as well as the procedure for applying them to the PRIMECLUSTER system.

12.3.1 Notes on Applying Corrections to the PRIMECLUSTER System

Note the following when you apply intensive correction 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 the 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 an intensive 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.

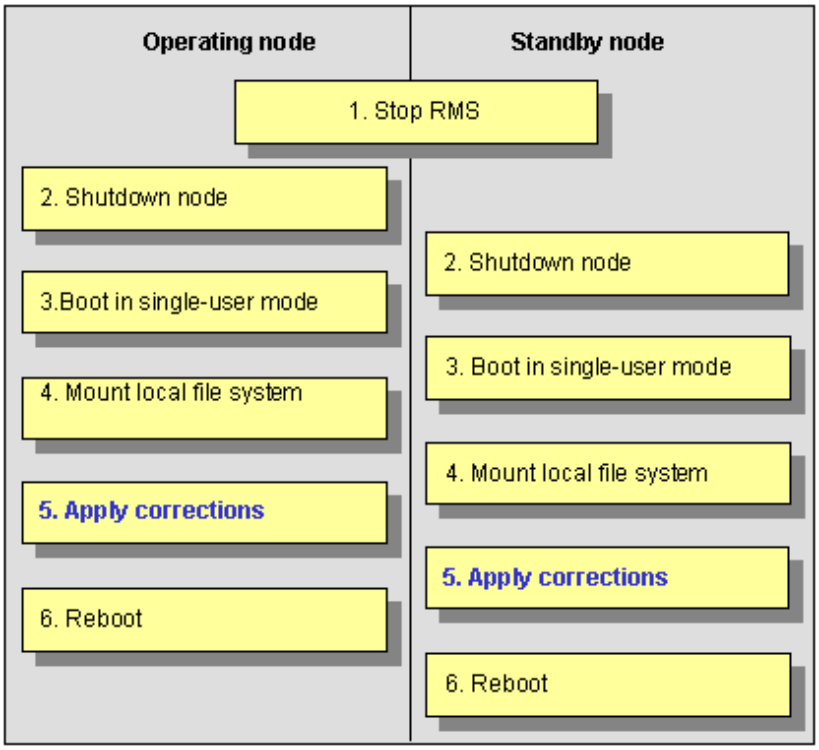
12.3.2 Overview of the Correction Application Procedure

This section provides an overview of the procedure for applying corrections such as an intensive 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 the 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.

   Mount the required local file system on all the nodes.

5. Apply corrections.

   Apply the corrections that were copied to the local file system in advance.

6. Restart.

   After applying the corrections, boot the nodes by using **shutdown -r**.

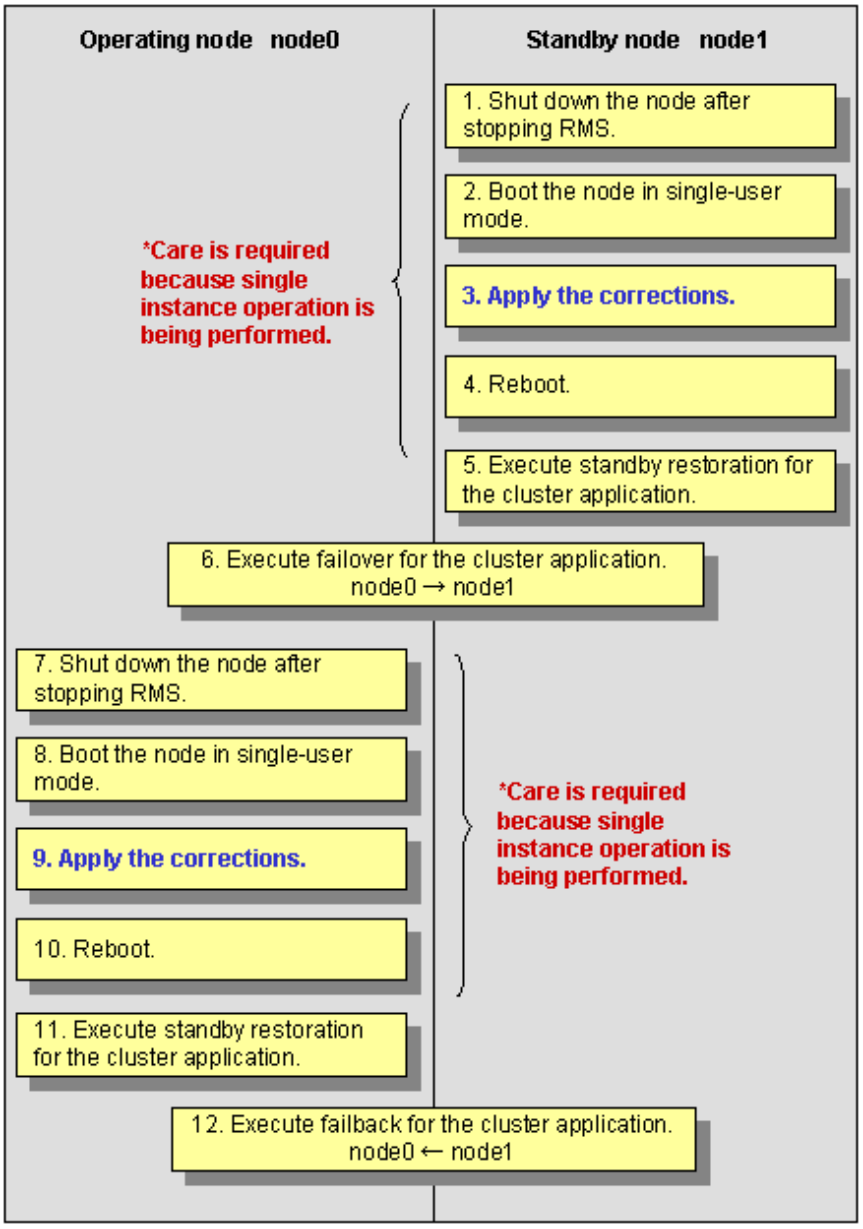
Note

- 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. Restart the standby node (node1).

   Restart the node.

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), switch all cluster applications to the standby node (node1) with the "**hvswitch**" command.

   If there is a scalable cluster application, switch all standby cluster applications under the scalable cluster application first, then switch the scalable cluster application.

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. Restart the operating node (node0).

    Restart the node to apply the corrections to the system.

11. Execute standby restoration for the operating node.

    Since the shutdown 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.

- 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."

Appendix A PRIMECLUSTER Products

PRIMECLUSTER products are as follows:

- PRIMECLUSTER Wizard for Oracle *1
- PRIMECLUSTER Wizard for NAS
- PRIMECLUSTER Wizard for NetVault
- PRIMECLUSTER Wizard for NetWorker
- PRIMECLUSTER Wizard for SAP HANA
- Interstage Application Server Enterprise Edition
- Symfoware Server
- Symfoware Server Hot Standby Option
- Systemwalker Centric Manager Enterprise Edition Server License (for manager)
- Systemwalker Centric Manager Enterprise Edition Server License (for agent)
- Systemwalker Operation Manager Enterprise Edition Server License
- Symfoware Server Enterprise Extended Edition
- Symfoware Server Parallel Cluster Option
- Systemwalker Service Quality Coordinator Enterprise Edition
- Systemwalker Centric Manager Enterprise Edition
- Systemwalker Operation Manager Enterprise Edition
- Systemwalker Software Delivery Enterprise Edition

*1 For correspondence of Oracle Database, refer to the manual for PRIMECLUSTER Wizard for Oracle.

See

For details on the version levels of PRIMECLUSTER products and the range of support, see the manual of each product.

Appendix B Manual Pages

This appendix provides online manual page lists for CF, CIP, operator intervention, PAS, cluster resource management facility, RMS, shutdown facility (SF), tracing failed resource, SIS, Web-Based Admin View, procedure resource, 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**

Note

In some cases, "(1M)" may be output as the section number of the manual page that is displayed with the man command.
Should this occur, assume the section number to be "(8)."

B.1 CF

System administrator

Command	Function
cfconfig	Configures or unconfigures a node for a PRIMECLUSTER cluster.
cfregd	CF registry synchronization daemon
cfset	Applies or modifies /etc/default/cluster.config entries into the CF module.
cftool	Prints the node communications state of a node or the cluster.
changeng	Replaces a node group definition.
deleteng	Deletes a node group.
descng	Replaces a node group explanation.
detailng	Displays the dynamic expansion of a node group.
newng	Creates a new node group.
rcqconfig	Configures or starts the quorum operation of a cluster system.
rcqquery	Acquires the state of consistency (quorum) of the cluster.
showng	Displays the name and definition of the node group.

B.2 CIP

System administrator

Command	Function
cipconfig	Starts or stops CIP 2.0.
ciptool	Retrieves CIP information about local and remote nodes in the cluster.

File format

File	Format
cip.cf	CIP configuration file format

B.3 Operator Intervention

System administrator

Command	Function
clreply	Responds to an operator intervention request message.

B.4 PAS

System administrator

Command	Function
mipcstat	MIPC statistics

B.5 Cluster Resource Management Facility

System administrator

Command	Function
clautoconfig	Executes automatic resource registration.
clbackuprdb	Saves the resource database.
clinitreset	Resets the resource database.
clrestorerdb	Restores the resource database.
clsetparam	Checks the connections of shared disk units and sets up the operation for automatic resource registration.
clsetup	Sets up the resource database.
clstartrsc	Activates a resource (GDS only).
clstoprsc	Deactivates a resource (GDS only).
clsyncfile	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
clgettree	Outputs tree information for the resource database.

B.6 RMS

System administrator

Command	Function
hvassert	Asserts (tests for) an RMS resource state.
hvcm	Starts the RMS configuration monitor.

Command	Function
hvconfig	Displays or saves the RMS configuration file.
hvdisp	Displays RMS resource information.
hvdispall	Displays RMS resource information on all the nodes.
hvdump	Collects debugging information about RMS.
hvlogclean	Cleans the RMS log files.
hvshut	Shuts down RMS.
hvswitch	Switches control of an RMS user application or resource to another host.
hvutil	Manipulates the availability of an RMS resource.

File format

File	Format
hvenv.local	RMS local environment valuables file

B.7 Shutdown Facility (SF)

System administrator

Command	Function
cldevparam	Changes and displays the tunable operation environment for asynchronous monitoring.
clmmbmonctl	Starts, stops, restarts, and displays the operating system of the MMB asynchronous monitoring daemon.
clmmbsetup	Registers, changes, deletes, and displays MMB information.
clvmgsetup	Registers, changes, deletes, and displays host OS information.
sdtool	Interface tool for shutdown daemon
rcsd	Shutdown daemon for shutdown manager

File format

File	Format
rcsd.cfg	Configuration file for shutdown daemon
SA_ipmi.cfg	Configuration file for IPMI Shutdown Agent
SA_blade.cfg	Configuration file for blade Shutdown Agent

B.8 Tracing Failed Resource

System administrator

Command	Function
cldispfaultrsc	Outputs a list of the current failed resources

B.9 SIS

System administrator

Command	Function
dtcpadmin	Starts the SIS administration utility.
dtcpd	Starts the SIS daemon for configuring VIPs.
dtcpdbg	Displays SIS debugging information.
dtcpstat	Displays state information on SIS.

B.10 Web-Based Admin View

System administrator

Command	Function
fjsvwvbs	Starts or stops Web-Based Admin View.
wvCntl	Starts, stops, or gets debugging information for Web-Based Admin View.
wvGetparam	Displays the Web-Based Admin View environment variables.
wvSetparam	Sets the Web-Based Admin View environment variables.
wvstat	Displays the operating state of Web-Based Admin View.

B.11 Procedure Resource

System administrator

Command	Function
claddprocrsc	Registers an application resource that uses a state transition procedure.
cldelproc	Deletes a state transition procedure.
cldelprocrsc	Deletes an application resource that uses state transition procedure.
clgetproc	Gets a state transition procedure.
clsetproc	Registers a state transition procedure.
clsetprocrsc	Changes the registered information of an application resource that uses a state transition procedure.

User command

Command	Function
cldspproc	Outputs information on the resource that uses the state transition procedure.

B.12 RMS Wizards

System administrator

Command	Function
clrwzconfig	Sets up the linking function between the PRIMECLUSTER resource manager and the middleware products after the RMS configuration definitions are activated.

RMS Wizards and RMS Application Wizard

The RMS Wizard manual will be saved in the following directory when the SMAWRhvdo package is installed.

```
/usr/doc/packages/SMAWRhv-do/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 the nodes that construct the cluster and the cluster management servers. Then, contact your customer support representative.

1. PRIMECLUSTER investigation information

   - Use fjsnap, FJQSS or pclsnap to collect information.
   - Retrieve the system dump.
   - Collect the Java Console on the clients.

     See "B.2.2 Java console" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

   - Collect a hard copy of the client screens.

     See "B.2.3 Screen hard copy" in "PRIMECLUSTER Web-Based Admin View Operation Guide."

2. Investigation information for the failed application
3. Crash Dump

   If the failed node is restartable, manually collect a crash dump before restarting it. Crash dump will be useful for troubleshooting if the failure is OS related.

   Example) If the failover occurred due to an unexpected resource failure

   After the failover of the cluster application is complete, collect a crash dump on the node where the resource failure occurred.

   For details on the crash dump, see "C.1.3 Crash Dump."

4. 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 the 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, FJQSS or pclsnap.

C.1.1 Executing the fjsnap or pclsnap Command

The fjsnap or pclsnap command is a tool which gathers system information necessary for analyzing the trouble at a time. When the trouble occurs by the PRIMECLUSTER system, the cause can be pursued by collecting necessary error information by the fjsnap or pclsnap command.

You can execute this command as follows:

1. Log in with system administrator authority.
2. Execute the "fjsnap" or "pclsnap" command.

   - For fjsnap

     /usr/sbin/fjsnap -a *output*

   - For pclsnap

     /opt/FJSVpclsnap/bin/pclsnap -a *output*

- The file name which becomes an output destination of system information collected by using the fjsnap or pclsnap command for output is specified.
- The following messages may be output to a switchlog and /var/log/messages when the fjsnap or pclsnap command is executed while one or more cluster nodes are stopped. However, no action is required for these messages.

  ```
  (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 included in the "FJSVsnap" package.

For details on the "pclsnap" command, see the "README" file included in the "FJSVpclsnap" package.

Information

Execution timings for the fjsnap or pclsnap command

- For problems that occur during operation, for example, if an error message is output, execute the "fjsnap" or "pclsnap" command immediately after the problem occurs.
- If the "fjsnap" or "pclsnap" command cannot be executed because the system hangs, collect a crash dump. Then start the system in single user mode, and execute the "fjsnap" or "pclsnap" command.
  For information on how to collect a crash dump, see "C.1.3 Crash Dump."
- 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" or "pclsnap" command.
- If investigation information cannot be collected because the "fjsnap" or "pclsnap" command results in an error, or the "fjsnap" or "pclsnap" command does not return, then collect a crash dump.

C.1.2 FJQSS (Information Collection Tool)

Collecting Information by FJQSS (Information Collection Tool)

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 input "[Enter]".

   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.
documents/fjqss-manual_sollnx/index_en.html

C.1.3 Crash Dump

Check the crash dump directory for a crash dump created after the switchover had occurred. The time the dump was collected can be found by referring to the time stamp using, for example, the "ls(1)" command.

- If a crash dump after the switchover is found

  Save the crash dump.

- If a crash dump after the switchover is not found

  If the failed node is restartable, manually collect a crash dump before restarting it.

Information

Crash dump directory

A crash dump is stored as a file on the node in which the error occurred.

If your guest OS has been forcefully stopped by the shutdown facility or the guest OS has been panicked in the environment where the KVM virtual machine function is used, the crash dump will be stored in the following directory for the host OS.

```
/var/crash/<shutdown time of the guest OS (YYYYMMDDHHMMSS)>.<Domain name for the guest OS>.core
```

Example: node1 was forcefully stopped at 12:34:56 on 20th April, 2011

/var/crash/20110420123456.node1.core

C.1.4 PSA Information

If a problem occurs in PRIMEQUEST 1000 Series shown below, collect the PSA information in addition to the PRIMECLUSTER failure investigation information.

Server model

PRIMEQUEST 1400S/1400E/1400L/1800E/1800L

Execute "getopsa" command.

```
/opt/FJSVpsa/sh/getopsa <filename>
```

Example:

```
/opt/FJSVpsa/sh/getopsa /tmp/node1_getopsa
```

See

For details on the "getopsa" command, see "PRIMEQUEST 1000 Series Tool Reference."

C.1.5 SVmco Information

If a problem occurs in PRIMEQUEST shown below, collect the SVmco information in addition to the PRIMECLUSTER failure investigation information.

Server model

- PRIMEQUEST 1400S2/1400E2/1400L2/1800E2/1800L2
- PRIMEQUEST 2000 Series

Execute "getosvmco" command.

```
/opt/fujitsu/SVmco/sh/getosvmco <filename>
```

Example:

```
/opt/fujitsu/SVmco/sh/getosvmco /tmp/node1_getosvmco
```

See

For details on the "getosvmco" command, see the following manuals:

- PRIMEQUEST 1000 Series

  "PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual"

- PRIMEQUEST 2000 Series

  "PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual"

C.2 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 "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 "5.2 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.

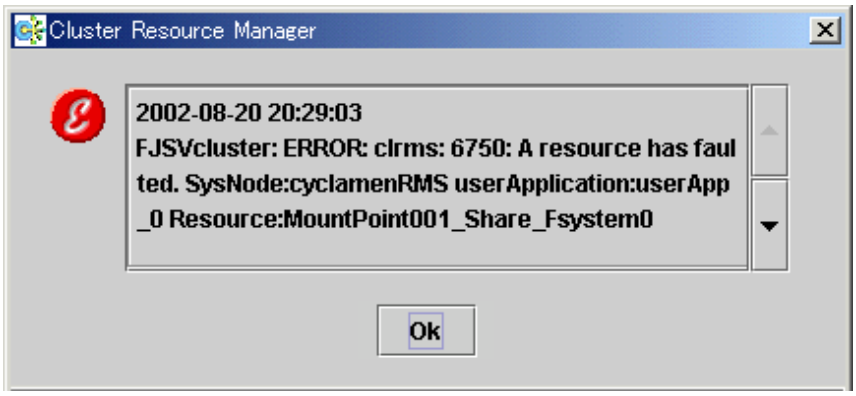
When this value is changed, you need to send the "SIGHUP" signal to clwatchlogd. When clwatchlogd receives this signal, clwatchlogd acquires the latest value of RELIANT_LOG_PATH. After completing the above processing, start RMS.

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.2.1 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 "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 "3.2 CRM View Messages" in "PRIMECLUSTER Messages" and for information on how to display previous messages, refer to "C.2.2 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 "4.3.3.3 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.

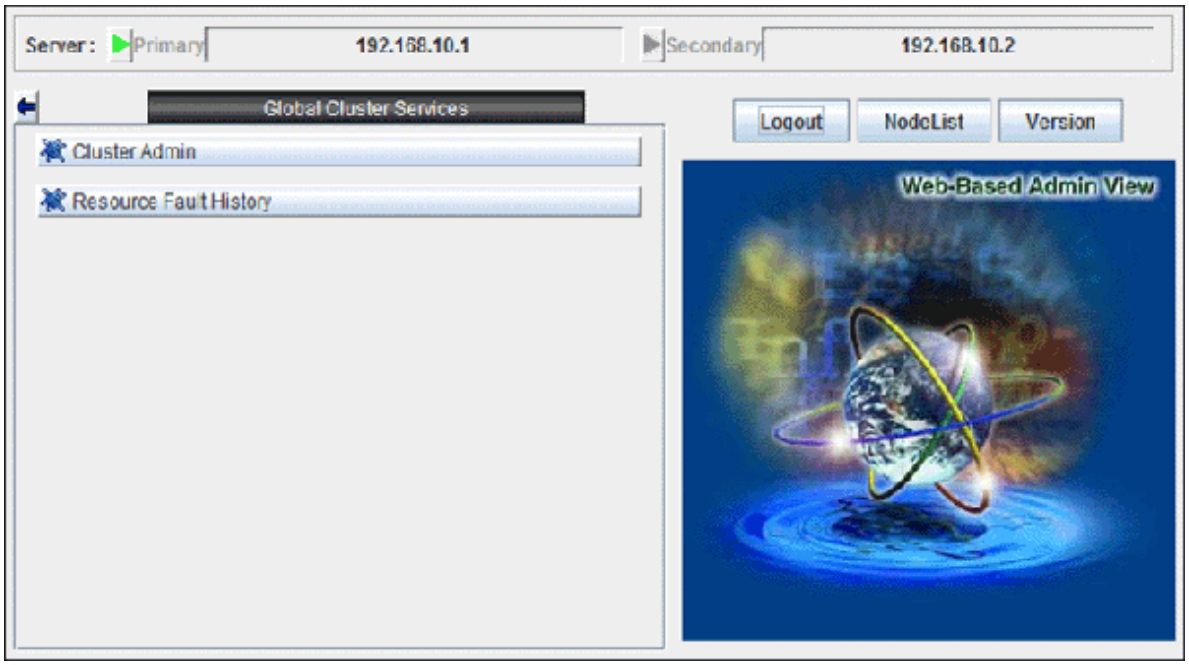
Each management server administers its fault resource messages. If you change the management server after confirming the message, the same message will be displayed again. To delete these messages, select *Cluster Admin* by using the GUI of *Web-Based Admin View* after closing *Cluster Admin*, and then open *Cluster Admin* again.

C.2.2 Resource Fault History

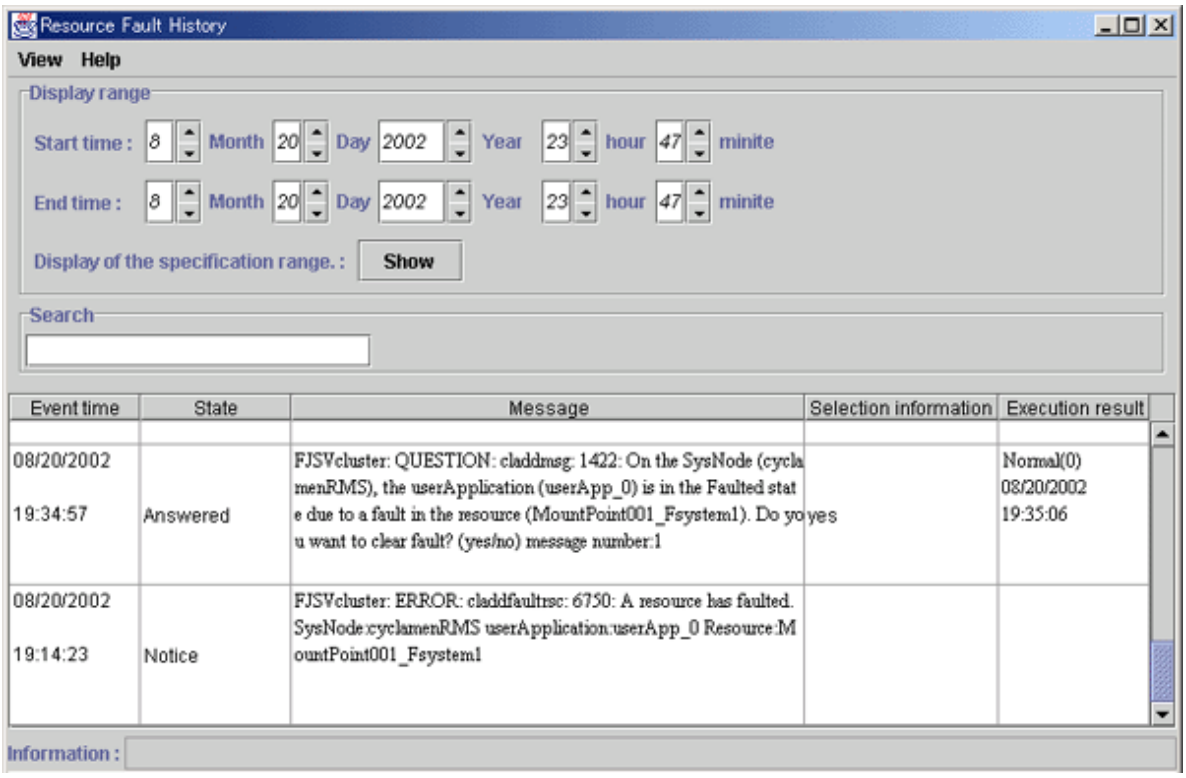
Display the "Resource Fault History" screen, in which the resource fault history is displayed, in the following procedure.

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.2.3 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" command, nothing will be displayed.
- Execution result - The result and time of the response processing are displayed.

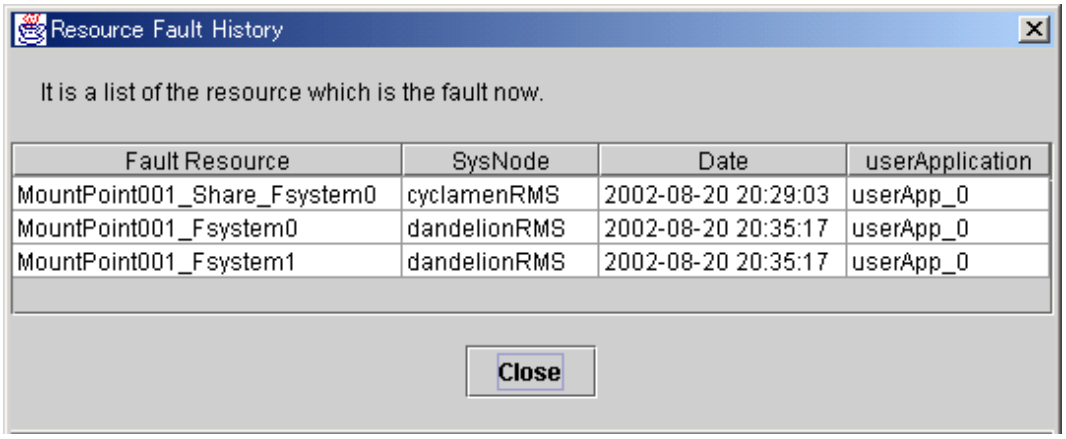
Information field

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.2.3 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" command.

C.3 PRIMECLUSTER Log Files

This appendix describes PRIMECLUSTER log files.

C.3.1 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/devmmbd

/var/opt/FJSVcluster/cores/FJSVcldev/devmmbmond

/var/opt/FJSVcluster/cores/FJSVcldev/devmmbmonitord

/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/devmmbd

devmmbd

Under /var/opt/FJSVcluster/cores/FJSVcldev/devmmbmond

devmmbmond

Under /var/opt/FJSVcluster/cores/FJSVcldev/devmmbmonitord

devmmbmonitor

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_nfs
hvdet_prmd
hvdet_execproc

C.3.2 core File Configuration

C.3.2.1 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.3.2.2 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.3.3 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.

- If RMS is run for one or more days with changing log level, configure the cron job settings to execute the hvlogclean command in order to avoid shortage of disk space caused by RMS log files. For details, see "C.3.4 Rotation and Deletion of RMS Log Files."

C.3.4 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 (RMS) 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 Registering, Changing, and Deleting State Transition Procedure Resources for PRIMECLUSTER Compatibility

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.

D.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 the 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" command to register the state transition procedure.

   See

   For details on the "clsetproc" command, see the manual page.

   Example

   To register the "/tmp/program" state transition procedure as program (file name) to the BasicApplication class

   ```
   # /etc/opt/FJSVcluster/bin/clsetproc -c BasicApplication -m program /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" command to register the procedure resource.

   See

   For details on the "claddprocrsc" command, see the manual page for claddprocrsc .

   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
```

D.2 Changing a Procedure Resource

This section explains how to change a procedure resource.

D.2.1 Changing a state transition procedure

Take the following steps to change the state transition procedure on all the 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" command to retrieve the state transition procedure.

   See

   For details on the "clgetproc" 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 "/tmp" directory.
   - The file name registered in the BasicApplication class is program.

   ```
   # /etc/opt/FJSVcluster/bin/clgetproc -c BasicApplication -f /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

   Register the state transition procedure by using the "clsetproc" command.

   For registration, specify the "-o" option to overwrite the state transition procedure.

   Example

   To register the "/tmp/program" state transition procedure as program (file name) to the BasicApplication class

   ```
   # /etc/opt/FJSVcluster/bin/clsetproc -c BasicApplication -m program -o /tmp/program
   ```

D.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 the nodes where the state transition procedure is registered.

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 a procedure resource again.

For more details, see "Chapter 10 Configuration change of Cluster Applications."

Operation Procedure:

1. Log in with the system administrator authority to the node in which the startup priority of state transition procedure is to be changed.
2. Delete the procedure resource of the cluster application.

   For deleting the procedure resource of the cluster application, refer to "10.5 Deleting a Resource."
3. Execute the "clsetprocrsc(1M)" command to change the startup priority of the state transition procedure used by the procedure resource.

   After performing this step on all the nodes where the procedure resource is registered, go to the next step.

   See

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

   Example

   When changing the startup priority of the state transition procedure 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
   ```
4. Register the procedure resource to the cluster application.

   For registering the procedure resource to the cluster application, refer to "6.7.3.7 Setting Up Procedure Resources."

D.2.3 Changing registration information of a procedure resource

Take the following steps to change the registration information of the procedure resource on all the nodes where the procedure resource to be changed is registered.

Note

To change the registration information of the procedure resource, you need to delete the procedure resource with the procedure for changing the cluster application configuration and create the procedure resource again.

For more details, see "Chapter 10 Configuration change of Cluster Applications."

Operation Procedure:

1. Log in with the system administrator authority to the node in which the registration information of procedure resource is to be changed.
2. Delete the procedure resource of the cluster application.

   For deleting the procedure resource of the cluster application, refer to "10.5 Deleting a Resource."
3. Execute the "clsetprocrsc(1M)" command to change the registration information of the procedure resource.

   After performing this step on all the nodes where the procedure resource is registered, go to the next step.

See

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

Example

When the procedure resource with the following conditions receives a state transition request of START RUN BEFORE in addition to START RUN AFTER and STOP RUN BEFORE;

- The resource key of the procedure resource is SDISK,
- This procedure resource uses the state transition procedure registered in the BasicApplication class, and
- The procedure resource is registered in the node (NODE1)

```
# /etc/opt/FJSVcluster/bin/clsetprocrsc -n SDISK -c BasicApplication -s NODE1 -K BEFORE,AFTER
-S BEFORE
```

4. Register the procedure resource to the cluster application.

   For registering the procedure resource to the cluster application, refer to "6.7.3.7 Setting Up Procedure Resources."

D.3 Deleting a Procedure Resource

Take the following steps to delete a procedure resource on all the 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" command to delete the procedure resource.

See

For details on the "cldelprocrsc" 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" command to delete the state transition procedure.

See

For details on the "cldelproc" 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 E Configuration Update Service for SA

This appendix explains Configuration Update Service for SA.

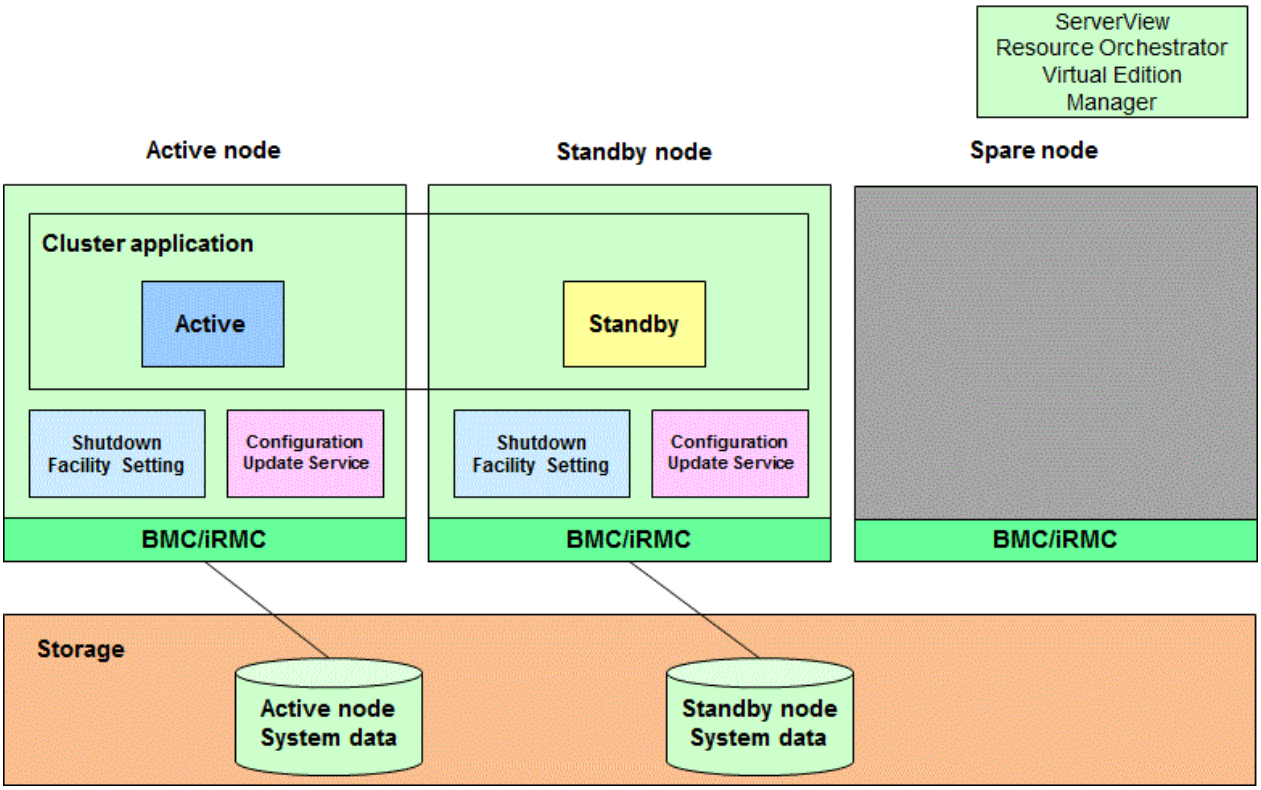
E.1 Feature Description

This function is enabled when building a cluster system in combination with ServerView Resource Orchestrator Virtual Edition.

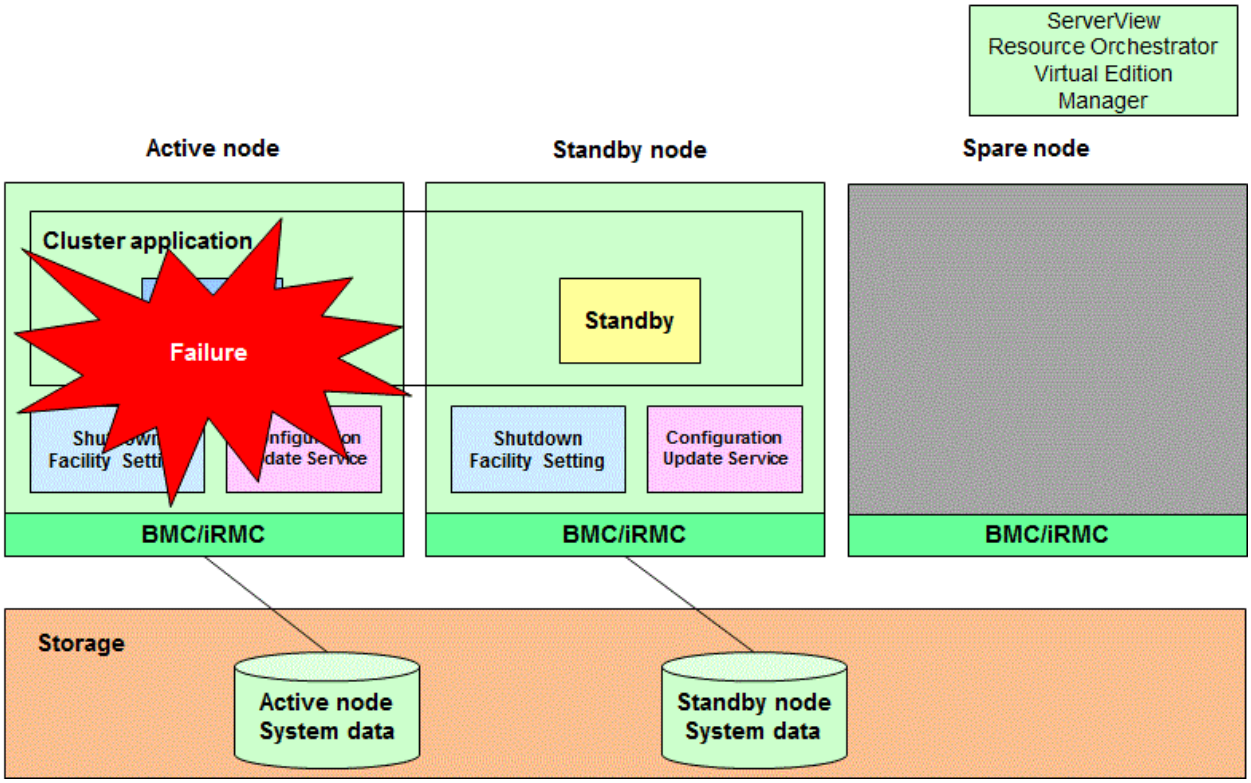
This configuration provides active, standby and spare nodes. Normally, cluster applications are started on the active node. In the event of a failure on the active node, the cluster applications fail over to the standby node.

After this, the spare node is started up using data from the storage, thus keeping the cluster configuration alive and ensuring even higher availability.

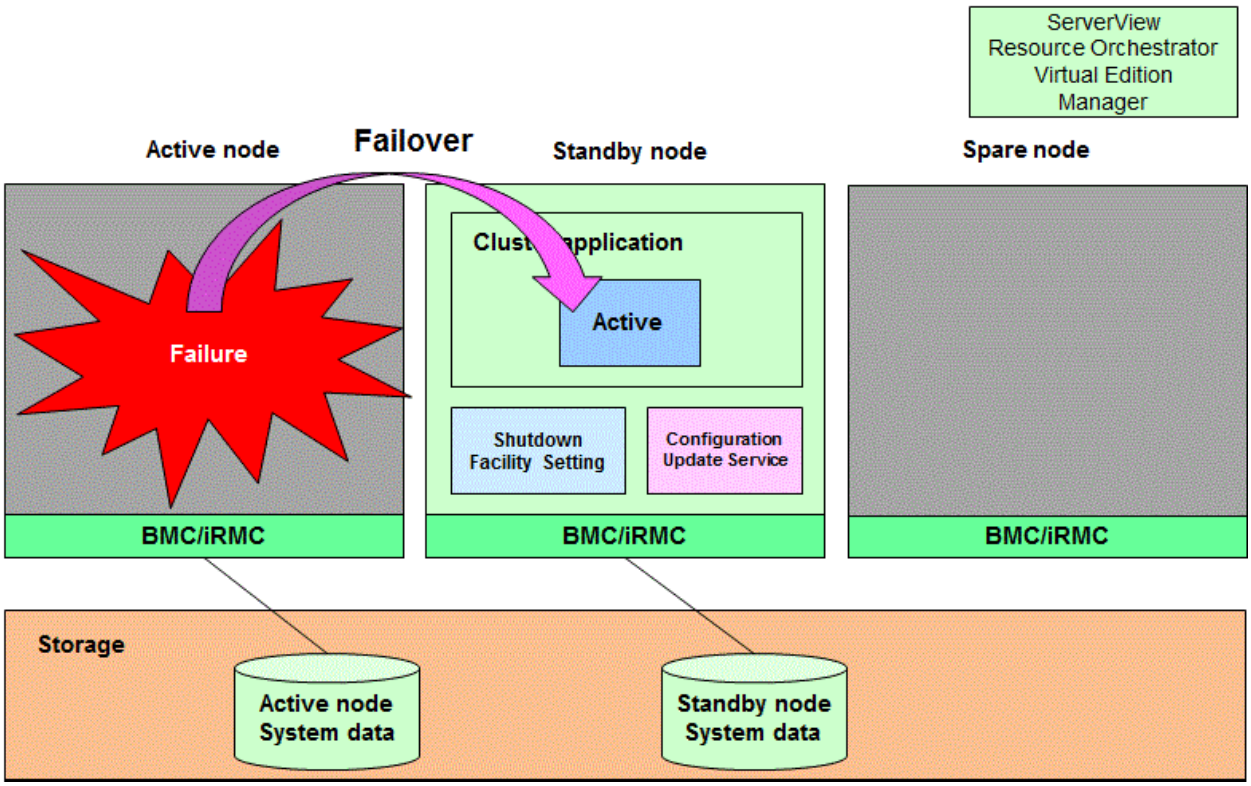
Configuration Update Service for SA is a function that automatically retrieves the BMC or iRMC IP address of the spare node and then updates the configurations of shutdown agents on the local node and other cluster nodes when starting the operating system.



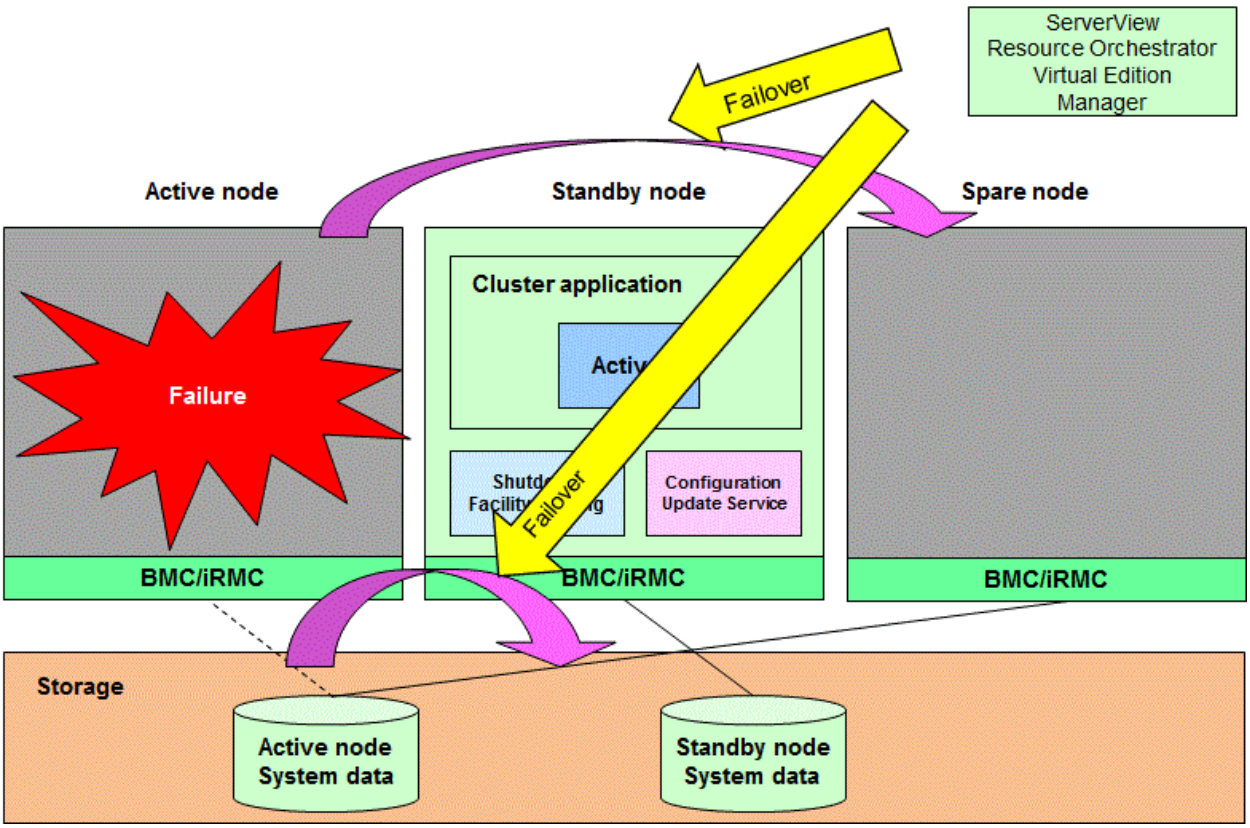
(1) A failure occurs on the active node, going to stop.



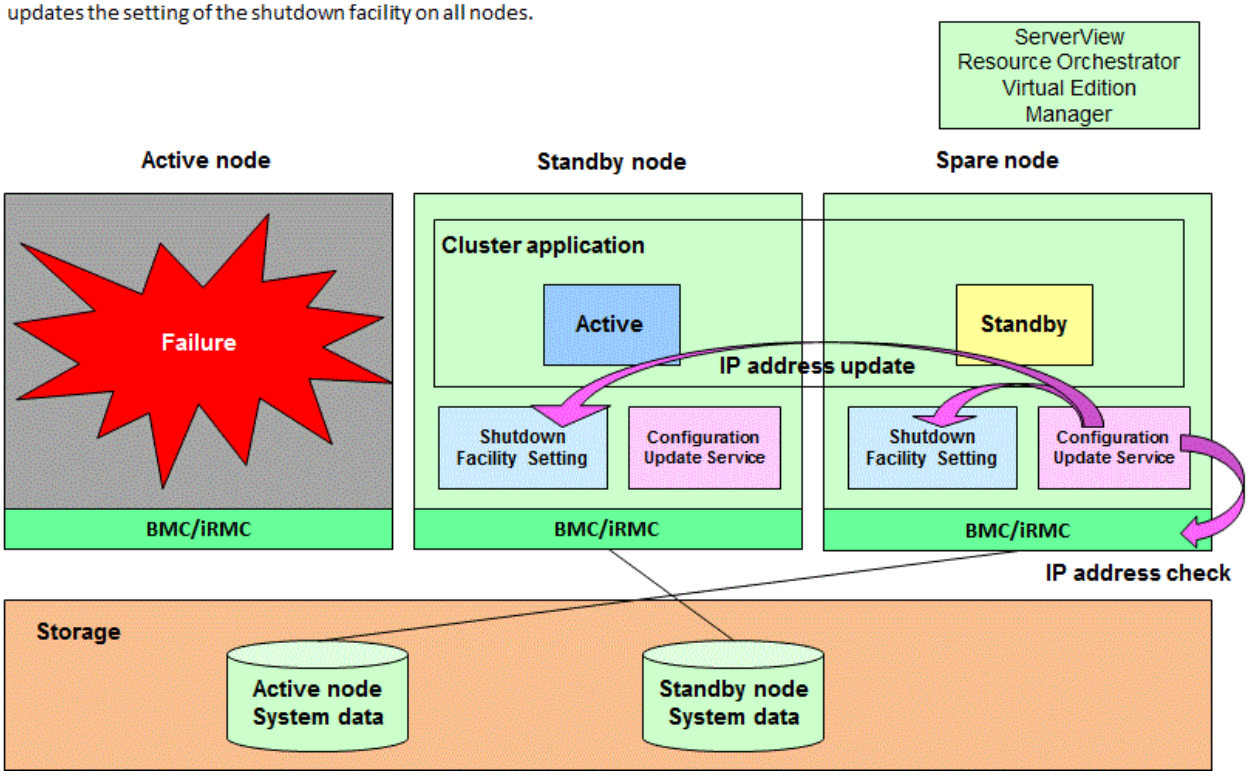
(2) Detects the stop of active node, the cluster applications fail over to the standby node.



(3) The manager of ServerView Resource Orchestrator Virtual Edition, connects system data of the active node to the spare node, and then starts the spare node.



(4) When starting the spare node, Configuration Update Service for SA checks the BMC or iRMC IP address, and then updates the setting of the shutdown facility on all nodes.



Note

- Use the same user name and password for BMC or iRMC on every node.
- If the PersistentFault attribute of RMS is set to "1," the Fault information is kept even if RMS is started on a normal spare node. (The default value of the PersistentFault attribute is "0.")
- When you update the configuration file for the shutdown agent, the updated configuration file is distributed to nodes in which the communication is available. The file is not distributed to nodes in which operation is stopped or the network communication is not available.

  In addition, when you start multiple nodes simultaneously, the configuration file for the shutdown agent is updated and distributed on multiple nodes at the same time. In this case, inconsistencies may occur in the information of the configuration file for the shutdown agent stored in each node.

  To check that correct information is distributed to all the nodes, execute the following command on any node when all the nodes are activated.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -s
```

  When the information that is output by the command is different between nodes, restore the service according to the procedure in "E.6 Restoration."

E.2 Operation Environment

You need the following environment to use Configuration Update Service for SA:

- Server model

  Models using the IPMI shutdown agent (SA_ipmi)

  ### See

  For details on models using the IPMI shutdown agent, see "5.1.2 Setting up the Shutdown Facility."

  ### Note

  When using Configuration Update Service for SA, available IP address for BMC or iRMC is only IPv4 address.

- Operating system

  The following operating systems are supported:

  - Red Hat Enterprise Linux 6 (for x86)
  - Red Hat Enterprise Linux 6 (for Intel64)
  - Red Hat Enterprise Linux 7 (for Intel64)

  ### Note

  This service is not available in a virtual machine environment.

- Required package

  - Red Hat Enterprise Linux 6, Red Hat Enterprise Linux 7
    - OpenIPMI
    - ipmitool

Check that the packages described above are installed by executing the rpm command. Install packages if they are not installed.

Packages are included in the installation media for the operating system.

E.3 Configuration

This section describes how to set up this service.

E.3.1 Startup Configuration for the IPMI Service

Perform the startup configuration for the IPMI service so that Configuration Update Service for SA can use the IPMI service when obtaining the BMC or iRMC IP address of the server.

In the RHEL6 environment

1. Starting the IPMI service

   Execute the following command on all the nodes to check the startup status of the IPMI service.

```
# /sbin/service ipmi status
```

   Execute the following command on all the nodes in which the IPMI service is not activated to start the IPMI service.

```
# /sbin/service ipmi start
Starting ipmi drivers:                                     [  OK  ]
```

2. Setting the run level of the IPMI service

   Check the current run level of the IPMI service on all the nodes.

```
# /sbin/chkconfig --list ipmi
ipmi            0:off   1:off   2:off   3:on    4:off   5:off   6:off
```

   Note

   Copy the value when the run level is set to "on." The above example shows that the run level is 3. This value is required for canceling this service and restoring the environment.

   Execute the following command on all the nodes to read the IPMI service on startup.

```
# /sbin/chkconfig ipmi on
```

   Information

   You can set "on" to run levels only that you want to activate this service. In this case, specify run levels in the range from 2 to 5.

In the RHEL7 environment

1. Starting the IPMI service

   Execute the following command on all the nodes to check the startup status of the IPMI service.

   When the content of "Active:" item is "inactive", the IPMI service has not been started, go to Step 2.

   When the content of "Active:" item is "active", the IPMI service has been started, go to Step 3.

```
# /usr/bin/systemctl status ipmi.service
ipmi.service - IPMI Driver
Loaded: loaded (/usr/lib/systemd/system/ipmi.service; disabled)
Active: inactive (dead)
```

2. Execute the following command to start the IPMI service on all the nodes where the IPMI service has not been started.

```
# /usr/bin/systemctl start ipmi.service
```

3. Enable the IPMI service.

   Confirm that the IPMI service is "enabled" on all the nodes.

```
# /usr/bin/systemctl list-unit-files --type=service | grep ipmi.service
ipmi.service                                disabled
```

   As "disabled" is displayed in the above situation, you need to execute the following command on all the nodes to read the IPMI service on startup.
   If "enabled" is displayed, it is not necessary to execute the following command.

```
# /usr/bin/systemctl enable ipmi.service
```

E.3.2 Activating Configuration Update Service for SA

Execute the sfsacfgupdate command to activate this service.

See

For details on the sfsacfgupdate command, see "E.7 sfsacfgupdate."

E.3.2.1 Startup Configuration for Update Service for SA

Execute the following command on all the nodes to activate this service.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -e
```

Information

In the RHEL6 environment

If you set "on" to run levels only that you want to activate this service in Step 2 in "E.3.1 Startup Configuration for the IPMI Service," specify the values of the run levels.

- Example: When activating this service in run levels 3 and 5

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -e 35
```

When run levels are omitted, this service is activated in run levels 2 to 5.

E.3.2.2 Checking the Configuration

Check the setup status on all the nodes.

[In the RHEL6 environment]

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -c
Configuration file exists.              [ OK ]
ipmitool command exists.                [ OK ]
ipmi service has been started.          [ OK ]
ipmi service's run level :
0:off   1:off   2:on    3:on    4:on    5:on    6:off
Configuration Update Service's run level :
0:off   1:off   2:on    3:on    4:on    5:on    6:off
```

[In the RHEL7 environment]

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -c
Configuration file exists.              [ OK ]
ipmitool command exists.                [ OK ]
ipmi service has been started.          [ OK ]
ipmi service state.                     [ enabled ]
Configuration Update Service state.     [ enabled ]
```

The following describes items to be displayed and contents to be checked.

Configuration file exists.

The existence of the configuration file for the shutdown agent is displayed.

Check that the status is "OK."

If the status is "NG," the configuration file for the shutdown agent does not exist. Review the configuration of the shutdown agent.

ipmitool command exists.

The existence of the ipmitool command is displayed.

Check that the status is "OK."

If the status is "NG," the ipmitool command does not exist. Check that the ipmitool command is correctly installed.

ipmi service has been started.

The startup status of the IPMI service is displayed.

Check that the status is "OK."

If the status is "NG," the IPMI service is not activated. Perform Step 1 in "In the RHEL6 environment" or "In the RHEL7 environment" of "E.3.1 Startup Configuration for the IPMI Service" again.

ipmi service's run level :

The statuses of run levels for the IPMI service are displayed.

Check that the status of each run level you set in Step 2 in "In the RHEL6 environment" of "E.3.1 Startup Configuration for the IPMI Service" is "on."

If the status of each run level is different, perform Step 2 in "In the RHEL6 environment" of "E.3.1 Startup Configuration for the IPMI Service" again.

Configuration Update Service's run level :

The status of the run level for Configuration Update Service for SA is displayed.

Check that the status is "on" in all or any levels from 2 to 5.

When the following error message is output, perform "E.3.2.1 Startup Configuration for Update Service for SA" again.

```
Configuration Update Service's run level :
sfsacfgupdate: ERROR: "sfsacfgupdate -e [<levels>]" is not executed.
```

ipmi service state.

To show the state of the IPMI service is enabled or disabled.

Confirm the state of the IPMI service is "enabled".

If the state of the IPMI service is "disabled", perform Step 3 in "In the RHEL7 environment" of "E.3.1 Startup Configuration for the IPMI Service" again.

Configuration Update Service state.

To show the configuration update service state of the shutdown agent is enabled or disabled.

Confirm the configuration update service state of shutdown agent is "enabled".

If the configuration update service state of shutdown agent is "disabled", or the following error message is output, perform "E.3.2.1 Startup Configuration for Update Service for SA" again.

```
sfsacfgupdate: ERROR: "sfsacfgupdate -e " is not executed.
```

Note

In the RHEL6 environment

Check that run levels, which are "on" in "Configuration Update Service's run level" are also "on" in "ipmi service's run level."

If the status of each run level is not identical, any setting may be incorrect. Review the processes of Step 2 in "In the RHEL6 environment" of "E.3.1 Startup Configuration for the IPMI Service" and "E.3.2.1 Startup Configuration for Update Service for SA."

E.3.2.3 Checking the BMC or iRMC IP Address and the Configuration Information of the Shutdown Agent

To check the BMC or iRMC IP address and the configuration information of the shutdown agent, execute the sfsacfgupdate command on any node.

Check that the following information is consistent with the displayed contents.

- BMC or iRMC IP address of each node ("BMC IP Address :")
- BMC or iRMC IP address of each node stored in the configuration file for the shutdown agent in each node. ("Configuration file :")

Example) Three-node cluster with nodeA, nodeB, and nodeC

The BMC IP address of each node is as follows:

```
nodeA: 10.20.30.41
nodeB: 10.20.30.42
nodeC: 10.20.30.43
```

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -s
Node : nodeA
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.41
 Configuration file :
  nodeA 10.20.30.41
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeB
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.42
 Configuration file :
  nodeA 10.20.30.41
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeC
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.43
 Configuration file :
  nodeA 10.20.30.41
  nodeB 10.20.30.42
  nodeC 10.20.30.43
```

The following describes displayed items.

Node :

The node name is displayed.

Node status :

The startup status of the node is displayed.

When the node is running, the status is "UP." For other than "UP," the subsequent information is not displayed.

Configuration Update Service status :

The setup status of Configuration Update Service for SA is displayed.

If no problem is found in "E.3.2.2 Checking the Configuration," the status is "ENABLE." For other than "ENABLE," the subsequent information is not displayed.

BMC IP Address :

The current BMC or iRMC IP address is displayed.

Configuration file :

The BMC or iRMC IP address of each node stored in the current configuration file for the shutdown agent is displayed.

E.4 Operation Check

This section describes how to check the operation of this service.

E.4.1 Operation Check by Restarting the System

This service operates on operating system startup. Therefore, you need to restart OS to check the operation.

The following describes how to check the operation by temporarily and manually updating the configuration file for the shutdown agent and restarting the system.

1. Backing up the configuration file for the shutdown agent

   Back up the SA_ipmi.cfg file on a node, hereafter referred to as nodeA, for which you check the operation.

```
# cp -p /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg.bk
```

2. Updating the configuration file for the shutdown agent

   Change the BMC or iRMC IP address of nodeA in the configuration file for the shutdown agent in nodeA to an unused IP address.

```
# vi /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg
~~~
nodeA 10.20.30.41:user:pass cycle
   The new address is as follows:
nodeA 255.255.255.255:user:pass cycle <- Change to an unused IP address
```

   Note

   When you change the IP address, the following message may be displayed on Syslog. As a result of execution of sdtool -s, the state of SA_ipmi may be "TestFailed," however, there is no problem.

```
SMAWsf : SA SA_ipmi to test host <node> failed
```

3. Restarting the system

   Restart nodeA.

```
# shutdown -r now
```

4. Checking the configuration file for the shutdown agent

   Check that the BMC or iRMC IP address of nodeA is updated in the configuration file for the shutdown agent in nodeA.

```
# vi /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg
~~~
nodeA 10.20.30.41:user:pass cycle
```

5. Deleting the backup file

   Delete the backed up configuration file for the shutdown agent on nodeA.

```
# rm -f  /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg.bk
```

E.5 Cancellation

The following describes how to cancel this service.

E.5.1 Deactivating Configuration Update Service for SA

Execute the following command on all the nodes to deactivate the configuration of this service.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -d
```

E.5.2 Restoring the Startup Configuration of the IPMI Service

When you do not need the IPMI service, return the run level of the IPMI service to the status before Step 2 in "E.3.1 Startup Configuration for the IPMI Service" is performed.

E.6 Restoration

This section describes restoration methods if correct information is not distributed to all the nodes when this service operates.

E.6.1 Restoration Method When Correct Information is not Distributed to All the Nodes

If communication is not available in other nodes when Configuration Update Service for SA operates, the configuration file for the shutdown agent is not distributed to other nodes.

In this case, consistency of the information stored in each node is not ensured and the shutdown agent cannot operate normally.

The following example shows the restoration methods in such a case.

If distribution of the configuration file for the shutdown agent to nodeB fails because the BMC IP address of nodeA is changed and nodeB is stopped:

nodeA: 10.20.30.41 -> Changed to 10.20.30.51
nodeB: 10.20.30.42 (Stopped)
nodeC: 10.20.30.43

1. Checking the message

   On nodeA, for which the BMC IP address is changed, check that any of the following messages are output on Syslog.

```
sfsacfgupdate: ERROR: Failed to copy the backup of <file> on node <node>.
```

```
sfsacfgupdate: ERROR: Failed to distribute <file> to node <node>.
```

```
sfsacfgupdate: ERROR: Failed to change the access permission of <file> on node <node>.
```

```
sfsacfgupdate: ERROR: Failed to change the group of <file> on node <node>.
```

```
sfsacfgupdate: ERROR: Failed to change the owner of <file> on node <node>.
```

If any of the above messages are output, the process for *<node>* has failed.

2. Checking the stopped node

   If all the other nodes stop while Configuration Update Service for SA is operating, messages in Step 1 are not output. Check if any nodes stop.

3. Restoring the stopped node

   Check that the status of the stopped node and restore it.

4. Checking the current status

   Execute the following command on any node to check the current status.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -s
Node : nodeA
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.51 <- Changed from 10.20.30.41
 Configuration file :
  nodeA 10.20.30.51  <- Updated with the changed information on nodeA
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeB
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.42
 Configuration file :
  nodeA 10.20.30.41 <- Not updated with the changed information on nodeB
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeC
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  11.22.33.46
 Configuration file :
  nodeA 10.20.30.51 <- Updated with the changed information on nodeC
  nodeB 10.20.30.42
  nodeC 10.20.30.43
```

   In the above example, you can see the BMC IP address of nodeA is not updated with the changed information in the configuration file for the shutdown agent stored in nodeB.

5. Restoring the information

   Execute the following command on a node storing the correct information to restore the status.

   In this case, execute the command on nodeA or nodeC because the information stored in nodeA and nodeC is correct.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -r
```

6. Checking the status after restoration

   Execute the following command on any node to check the current status.

   In the following example, you can see the BMC IP address of nodeA is updated with the changed information in the configuration file for the shutdown agent stored in nodeB.

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -s
Node : nodeA
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.51
 Configuration file :
  nodeA 10.20.30.51
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeB
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  10.20.30.42
 Configuration file :
  nodeA 10.20.30.51 <- Updated with the changed information on nodeB
  nodeB 10.20.30.42
  nodeC 10.20.30.43

Node : nodeC
 Node status : UP
 Configuration Update Service status : ENABLE
 BMC IP Address :
  11.22.33.46
 Configuration file :
  nodeA 10.20.30.51
  nodeB 10.20.30.42
  nodeC 10.20.30.43
```

E.7 sfsacfgupdate

The following describes how to use the sfsacfgupdate command.

Name

sfsacfgupdate(8) - Management of Configuration Update Service for SA

Synopsis

/opt/SMAW/SMAWsf/bin/sfsacfgupdate {-e [*<levels>*]|-d|-c|-s|-r}

Feature description

This command manages Configuration Update Service for SA.

When Configuration Update Service for SA is activated, the configuration information of the shutdown agent is automatically updated on operating system startup. Execute this command with the system administrator authority.

Options

-e

Activates Configuration Update Service for SA.

Specify the value of run levels 2 to 5 which you want to activate for *<levels>*. You can specify several run levels.

For example, when you specify "-e 35," run levels 3 and 5 will be activated.

When you omit the value, all run levels from 2 to 5 will be activated.

-d

Deactivates Configuration Update Service for SA.

-c

Checks the setup status of Configuration Update Service for SA.

-s

Displays the configuration information of the shutdown agent stored in all the nodes.

-r

Restores the configuration information of the shutdown agent.

Example

```
# /opt/SMAW/SMAWsf/bin/sfsacfgupdate -c [Return]
Configuration file exists.              [ OK ]
ipmitool command exists.                [ OK ]
ipmi service has been started.          [ OK ]
ipmi service's run level :
0:off   1:off   2:on    3:on    4:on    5:on    6:off
Configuration Update Service's run level :
0:off   1:off   2:on    3:on    4:on    5:on    6:off
#
```

Exit status

0 : Normal exit

Other than 0 : Abnormal exit

E.8 Output Message (Syslog)

The following shows the messages output to Syslog.

---

sfsacfgupdate: ERROR: <*command*> command failed. return_value=<*value*>.

Content:

<command> abnormally ended with the return value *<value>.*

Corrective action:

Copy this message, and then contact field engineers.

---

sfsacfgupdate: ERROR: Could not find <*file*>.

Content:

<file> does not exist.

Corrective action:

Create *<file>.*

---

sfsacfgupdate: ERROR: Could not find ipmitool command.

Content:

The ipmitool command does not exit.

Corrective action:

Install the ipmitool command.

---

sfsacfgupdate: ERROR: ipmi service doesn't start.

Content:

The ipmi service does not start.

Corrective action:

Start the ipmi service.

sfsacfgupdate: ERROR: <*file*> is invalid.

Content:

Contents described in <*file*> include incorrect information.

Corrective action:

Check the contents in <*file*> and enter the correct information.

sfsacfgupdate: ERROR: Reading the Shutdown Agent configuration failed.

Content:

Reading the configuration file for the shutdown agent failed.

Corrective action:

Review the contents of the configuration file for the shutdown agent, and check if the correct information is entered.

sfsacfgupdate: ERROR: Failed to copy the backup of <*file*> on node <*node*>.

Content:

Copying the backup of <*file*> failed on <*node*>.

Corrective action:

Check that the communication with <*node*> is available. After restoring the state of <*node*>, execute this command with the -r option and restore the configuration information of the shutdown agent.

sfsacfgupdate: ERROR: <*file*> generation failed.

Content:

Creating a file failed.

Corrective action:

Copy this message, and then contact field engineers.

sfsacfgupdate: ERROR: Failed to distribute <*file*> to node <*node*>.

Content:

Distributing <*file*> to <*node*> failed.

Corrective action:

Check that the communication with <*node*> is available. After restoring the state of <*node*>, execute this command with the -r option and restore the configuration information of the shutdown agent.

sfsacfgupdate: ERROR: Failed to change the access permission of <*file*> on node <*node*>.

Content:

Changing the mode of <*file*> failed on <*node*>.

Corrective action:

Check that the communication with <*node*> is available. After restoring the state of <*node*>, execute this command with the -r option and restore the configuration information of the shutdown agent.

sfsacfgupdate: ERROR: Failed to change the group of <*file*> on node <*node*>.

Content:

Changing the group of <*file*> failed on <*node*>.

Corrective action:

Check that the communication with <*node*> is available. After restoring the state of <*node*>, execute this command with the -r option and restore the configuration information of the shutdown agent.

sfsacfgupdate: ERROR: Failed to change the owner of <*file*> on node <*node*>.

Content:

Changing the owner of <*file*> failed on <*node*>.

Corrective action:

Check that the communication with <*node*> is available. After restoring the state of <*node*>, execute this command with the -r option and restore the configuration information of the shutdown agent.

Appendix F Using PRIMECLUSTER in a KVM Environment

This appendix explains PRIMECLUSTER used in KVM environment.

F.1 Controlling and monitoring a guest OS by a cluster application on a host OS.

Adding the hvlibvirt script to the Cmdline resource of a cluster application on a host OS enables a guest OS to be started and shut down in response to start and shut down of the cluster application. The cluster application can also monitor the guest OS status. By the virsh command, the hvlibvirt script executes following procedures that are set as arguments: starting and shutting down the guest OS, and monitoring the guest OS status.

Specify each script Start, Stop, and Check to configure the Cmdline resource that uses the hvlibvirt script.

<Start script>

```
/opt/SMAW/bin/hvlibvirt -c -z <dom_name> -t <timeout>
```

<Stop script>

```
/opt/SMAW/bin/hvlibvirt -u -z <dom_name> -t <timeout>
```

<Check script>

```
/opt/SMAW/bin/hvlibvirt -m -z <dom_name> -t <timeout>
```

Specify the domain name of the target guest OS for <*dom_name*>.

Specify the timeout value to shut down the guest OS by seconds for <*timeout*>. The script shuts down the guest OS in offline processing. When the shutdown process is not completed beyond the shutdown time specified by <*timeout*>, use the destroy command of virsh (virsh destroy <*dom_name*>) to shut down the guest OS.

Set flags for the Cmdline resource as follows. See "6.11 Notes When Setting Cmdline Resources" for more information.

- NULLDETECTOR

  Disabled (to enable Check script)

- STANDBYCAPABLE

  Disabled (Standby is disabled)

- ALLEXITCODES

  Disabled (Standby is disabled)

- TIMEOUT

  The default value is 300 seconds. Set the timeout duration to be longer than the time until the boot/shutdown sequence of the guest OS completes.

See "6.7.3.1 Setting Up Cmdline Resources" to set the Cmdline resource.

Information

Execute virsh command as below to check the domain name of the guest OS.

(Example) The domain name of the guest OS is domain 1

```
# virsh list --all
 Id   Name                Status
----------------------------------
  0 Domain-0             Active
  - domain1              Shutoff
```

Appendix G Using the Migration Function in a KVM Environment

This appendix describes design, prerequisites and operations when using the Migration function in a KVM environment.

G.1 Design

Following three types of the Migration function can be used for a cluster system in a KVM environment:

- Live Migration

  Transferring an active guest OS.

- Offline Migration

  Transferring a suspended guest OS.

- Migration by Export/Import

  Exporting/Importing the XML setup files of stopped guest OSes.

For the cluster configurations which are available for the KVM migration function, see "2.2.1 Virtual Machine Function."

Note

In the migrated guest OS, virtio block storages are added under the device name "vdpcl". Note the following points to add virtio block storages for migration.

- Keep the number of virtio block storages in guest OSes within 27 devices except the device (vdpcl) to be added for migration.
- Do not use "vdpcl" for the device name of virtio block storages in guest OSes.

G.2 Prerequisites

This section describes the prerequisites for the migration function in a KVM environment.

G.2.1 Without using the Host OS failover function

Perform the following procedure on guest OSes in which the Migration is performed and all host OSes.

You need to perform this procedure only once and not for each Migration.

1. Creating host OS information files (guest OS)

   Execute the following command under any directory on one of the cluster nodes of the guest OS to create an information file of the host OS. After executing this command, a file named "sfkvmmigrate.img.*hostname*" will be created in the current directory.

   Execute the command several times to create information file of all host OSes.

   If you have already performed this procedure, you do not have to perform Step 1 through 3.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigratesetup -c -i hostip -g hostname
   ```

   hostip

   Specify the IP address of the host OS.

   Available IP address formats are IPv4 and IPv6.

   IPv6 link local addresses are not available.

   hostname

   Specify the host name of the host OS.

2. Distributing the host OS information file (host OS)

   Forward the host OS information file created in Step 1 to each host OS and change the file name to "sfkvmmigrate.img." Then, place it to "/var/opt/SMAWsf".

```
# mkdir -p /var/opt/SMAWsf
# cp sfkvmmigrate.img.hostname /var/opt/SMAWsf/sfkvmmigrate.img
```

3. Changing privileges (host OS)

   Execute the following commands to change the privileges of the host OS information file distributed in Step 2.

```
# chown root:root /var/opt/SMAWsf/sfkvmmigrate.img
# chmod 644 /var/opt/SMAWsf/sfkvmmigrate.img
```

4. Setting up the guest OSes (host OS/guest OS)

   Perform the following procedure on all guest OSes.
   It is alternative to perform following procedure on all guest OSes at a time or one by one.

   1. Stopping of guest OS

      Execute the following command on the guest OS to stop the guest OS.

```
# /sbin/shutdown -P now
```

   2. Settings to look up host OS information

      On the host OS where the guest OS is stopped, execute the following command to enable the guest OS to look up the host OS information file.

      [RHEL6]

```
# virsh attach-disk domain /var/opt/SMAWsf/sfkvmmigrate.img vdpcl --cache none --mode
readonly --shareable --persistent
```

      [RHEL7]

```
# virsh attach-disk domain /var/opt/SMAWsf/sfkvmmigrate.img vdpcl --cache none --mode
readonly --persistent
```

      domain

      Specify the domain name of the guest OS.

   3. Startup of guest OS

      Start the guest OS.

5. Creating the user ID in the destination host OS (host OS)

   Create the user ID in the destination host OS.

   For the detailed procedure, see "3.2.2.2 Host OS setup (after installing the operating system on guest OS)."

6. Login to the destination host OS (guest OS)

   From all guest OSes, authenticate yourself (create the RSA key) on the destination host OS in advance, which is required when using SSH for the first time.

   Log in to the destination host OS from all guest OSes with the host OS account specified in libvirt shutdown agent.

```
# ssh -l user 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
```

G.2.2 Using the Host OS failover function

Perform the following procedure on guest OSes in which the Migration is performed and all host OSes.

You need to perform this procedure only once and not for each Migration.

1. Creation of kvmguests.conf file (host OS)

   Perform the following procedure on all host OSes to create the kvmguest.conf file. The file named kvmguests.conf must be the same on all host OSes. For PRIMEQUEST, if the file is already created when the host OS failover function was set, it is not required to perform the procedure again.

   1. Check the setting information.

      When performing migration, log in to the guest OS (cluster node) via SSH to change the settings of the shutdown facility.

      Prior to the settings, confirm the following information that are required for the settings.

      - IP address of the guest OS
      - Domain name of the guest OS
      - Cluster name of the guest OS
      - CF node name of the guest OS

   2. Create the user (when logging in to the guest OS not as a root user).

      Create the user (when logging in to the guest OS not as a root user).

      Take the following steps on the guest OS to be migrated.

      1. Create the login user.

         Set the user password with seven-bit ASCII characters except the following characters.

         ```
         >  <  "  /  \  =  !  ?  ;  ,  &
         ```

      2. Set the sudo command so that the created user can execute the command as a root user.

         Execute the visudo command by using the sudo command. Describe the following setting in the displayed setting file.

         ```
         <User created in step (1)>    ALL=(root) NOPASSWD: ALL
         ```

   3. Encrypt the password.

      Execute the sfcipher command to encrypt the user password (for the user created as a root user or the user created in step 2) to log in to the guest OS via SSH.

      For information on how to use the sfcipher command, see the "sfcipher" manual page.

      ```
      # sfcipher -c
      Enter User's Password:
      Re-enter User's Password:
      D0860AB04E1B8FA3
      ```

   4. Create /etc/opt/FJSVcluster/etc/kvmguests.conf.

      Create /etc/opt/FJSVcluster/etc/kvmguests.conf with the following contents.

      Create the kvmguests.conf file as a root user. Set the permission as 600.

      ```
      guest-name host-cfname guest-clustername guest-cfname guest_IP guest_user guest_passwd
        :
      ```

      - Enter the information of one node in one line.
      - Delimit each item with a single space.

- The kvmguests.conf file must be the same on all cluster nodes.

```
guest-name          :Specify the domain name of the guest OS to be migrated.
host-cfname         :Specify the CF node name of the host OS in which "guest-name"
                     is running.
                     If you execute "cftool -l" on the host OS in which "guest-name"
                     is running, you can confirm the CF node name of the node.
guest-clustername   :Specify the cluster name of the guest OS.
                     If you execute "cftool -c" on the guest OS, you can confirm
                     the cluster name of the node.
guest-cfname        :Specify the CF node name of the guest OS.
                     If you execute "cftool -l" on the guest OS, you can confirm
                     the CF node name of the node.
guest_IP            :Specify the IP address of the guest OS.
                     Available IP address formats are IPv4 and IPv6 addresses.
                     IPv6 link local addresses are not available.
guest_user          :Specify the user name for logging in to the guest OS.
                     Specify the user created as a root user or the created in step 2.
guest_passwd        :Specify the user password for logging in to the guest OS.
                     Specify the password encrypted in step 3.
```

Example: In a two-node configuration between guest OSes, two cluster systems are configured

```
guest11 cfhost1 cluster1 cfguest11  10.20.30.50 user1 D0860AB04E1B8FA3
guest12 cfhost2 cluster1 cfguest12  10.20.30.51 user2 D0860AB04E1B8FA3
guest21 cfhost1 cluster2 cfguest21  10.20.30.60 user3 D0860AB04E1B8FA3
guest22 cfhost2 cluster2 cfguest12  10.20.30.61 user4 D0860AB04E1B8FA3
```

5. Confirm the log in to the guest OS.

   The shutdown facility accesses the target node with SSH during migration. Therefore, you need to authenticate yourself (create the RSA key) in advance, which is required when using SSH for the first time.

   Check that you can connect to all the guest OSes (nodes) which are defined to /etc/opt/FJSVcluster/etc/kvmguests.conf via SSH as a root user. Execute the command as a root user.

```
# ssh -l user 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."
```

6. Check the setting in /etc/opt/FJSVcluster/etc/kvmguests.conf.

   Execute the sfkvmtool command on all the host OSes to make sure that the settings in /etc/opt/FJSVcluster/etc/kvmguests.conf are correct. If the settings are correct, the following message is output.

```
# /opt/SMAW/SMAWsf/bin/sfkvmtool -c
NOTICE: The check of configuration file succeeded.
```

   If a message other than above is output, review the setting in /etc/opt/FJSVcluster/etc/kvmguests.conf.

7. Start the shutdown facility.

   Check that the shutdown facility has already been started on all the nodes.

```
# sdtool -s
```

   If the shutdown facility has already been started, execute the following on all the nodes to restart it.

```
# sdtool -e
# sdtool -b
```

   If the shutdown facility has not been started, execute the following on all the nodes to start it.

```
# sdtool -b
```

2. Registration of host OS information (host OS)

   Execute the following command on the all host OSes to register the host OS information.

```
# /opt/SMAW/SMAWsf/bin/sfkvmmigratesetup -c -i hostip [-w off]
```

   hostip

   Specify the IP address of the host OS on which this command was executed.

   Available IP address formats are IPv4 and IPv6.

   IPv6 link local addresses are not available.

   -w off

   Specify this option if the weights of the guest OS shutdown facility and that of the host OS shutdown facility should not be linked when migrating the guest OS.
   Without this option, linkage of the weights of the guest OS shutdown facility and the host OS shutdown facility is enabled when migrating the guest OS.
   This option must be the same on all host OSes.

3. Setting up guest OSes (host OS/guest OS)

   Perform following procedure on all guest OSes.
   It is alternative to perform following procedure on all guest OSes at a time or one by one.

   1. Stopping of guest OS

      Execute the following command on the guest OS to stop the guest OS.

```
# /sbin/shutdown -P now
```

   2. Settings to look up host OS information

      On the host OS where the guest OS is stopped, execute the following command to enable the guest OS to look up the host OS information file.

```
# /opt/SMAW/SMAWsf/bin/sfkvmmigratesetup -s domain
```

      domain

      Specify the domain name of the guest OS.

   3. Startup of guest OS

      Start the guest OS.

4. Creating the user ID in the destination host OS (host OS)

   Create the user ID in the destination host OS.

   For the detailed procedure, see "3.2.3.1.4 Host OS setup (after installing the operating system on guest OS)."

5. Login to the destination host OS (guest OS)

   Log in to the destination host OS from all guest OSes and authenticate yourself (create the RSA key) in advance, which is required when using SSH for the first time.

   Log in to the destination host from all guest OSes with the host OS account specified in libvirt shutdown agent.

```
# ssh -l user 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
```

G.3 Operation

This appendix describes operation with the migration function in a KVM environment.

G.3.1 When performing Live Migration

G.3.1.1 When not using the Host OS failover function

G.3.1.1.1 Operations before Live Migration

This section describes operations before Live Migration in a KVM environment.

1. Changing cluster settings (guest OS)

   Change the cluster settings before Live Migration.

   Execute the following command on the guest OS that is to be migrated.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -p source-domain -g
   ```

   source-domain

   Domain name of guest OS to be migrated

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (10 seconds to 600 seconds)
   - Stop of the shutdown facility

G.3.1.1.2 Operations after Live Migration

This section describes operations after Live Migration in a KVM environment.

1. Changing cluster settings (guest OS)

   Change the cluster settings after Live Migration.

   Execute the following command on the migrated guest OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -u source-domain -g
   ```

   source-domain

   Domain name of migrated guest OS

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (600 seconds to 10 seconds)
   - Settings of the shutdown facility (IP address of host OS, CF node name of host OS, weight of SF)
   - Settings of the Host OS failover function (CF node name of host OS)
   - Startup of the shutdown facility
   - Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes of the guest OS to check that the cluster settings are correct after Live Migration.

   ```
   # sdtool -s
   ```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

   Perform the procedure from Step 1 again.

G.3.1.2 When using the Host OS failover function

G.3.1.2.1 Operations before Live Migration

This section describes operations before Live Migration in a KVM environment.

1. Changing cluster settings (host OS)

   Change the cluster settings before Live Migration.

   Execute the following command on the cluster node of the original host OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -p source-domain
   ```

   source-domain

   Domain name of guest OS that is to be migrated

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (10 seconds to 600 seconds)
   - Stop of the shutdown facility

G.3.1.2.2 Operations after Live Migration

This section describes operations after Live Migration in a KVM environment.

1. Changing cluster settings (host OS)

   Change the cluster settings after Live Migration.

   Execute the following command on the original host OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -u source-domain
   ```

   source-domain

   Domain name of migrated guest OS

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (600 seconds to 10 seconds)
   - Settings of the shutdown facility (IP address of host OS, CF node name of host OS, weight of SF)
   - Settings of the Host OS failover function (CF node name of host OS)
   - Startup of the shutdown facility

2. Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes to check that the cluster settings are correct after Live Migration.

   ```
   # sdtool -s
   ```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

   Perform the procedure from Step 1 again.

G.3.2 When performing Offline Migration

G.3.2.1 When not using the Host OS failover function

G.3.2.1.1 Operations before Offline Migration

This section describes operations before Offline Migration in a KVM environment.

1. Changing cluster settings (guest OS)

   Change the cluster settings before Offline Migration.

   Execute the following command on the guest OS that is to be migrated.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -p source-domain -t CFtimeout -g
   ```

   source-domain

   Domain name of guest OS to be migrated

   CFtimeout

   Timeout of CF cluster interconnect (seconds)

   For the value of *CFtimeout*, specify (real time of Offline Migration + 300 seconds of tolerance time for processing delay).

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (10 seconds to *CFtimeout* seconds)
   - Stop of the shutdown facility

G.3.2.1.2 Operations after Offline Migration

This section describes operations after Offline Migration in a KVM environment.

1. Changing cluster settings (guest OS)

   Change cluster settings after Offline Migration.

   Execute the following command on the migrated guest OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -u source-domain -g
   ```

   source-domain

   Domain name of migrated guest OS

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (value specified before Offline Migration [seconds] to 10 seconds)
   - Settings of the shutdown facility (IP address of host OS, CF node name of host OS, weight of SF)
   - Settings of the Host OS failover function (CF node name of host OS)
   - Startup of the shutdown facility

2. Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes to check that the cluster settings are correct after Offline Migration.

   ```
   # sdtool -s
   ```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

Perform the procedure from Step 1 again.

G.3.2.2 When using the Host OS failover function

G.3.2.2.1 Operations before Offline Migration

This section describes operations before Offline Migration in a KVM environment.

1. Changing cluster settings (host OS)

   Change the cluster settings before Offline Migration.

   Execute the following command on the original host OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -p source-domain -t CFtimeout
   ```

   source-domain

   Domain name of guest OS to be migrated

   CFtimeout

   Timeout of CF cluster interconnect (seconds)

   For the value of *CFtimeout*, specify (real time of Offline Migration + 300 seconds of tolerance time for processing delay).

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (10 seconds to *CFtimeout* seconds)
   - Stop of the shutdown facility

G.3.2.2.2 Operations after Offline Migration

This section describes operations after Offline Migration in a KVM environment.

1. Changing cluster settings (host OS)

   Change the cluster settings after Offline Migration.

   Execute the following command on the original host OS.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigrate -u source-domain
   ```

   source-domain

   Domain name of migrated guest OS

   After executing this command, the cluster settings on all the nodes of clusters among the guests specified by *source-domain* will be changed as follows:

   - Timeout of CF cluster interconnect (value specified before Offline Migration [seconds] to 10 seconds)
   - Settings of the shutdown facility (IP address of host OS, CF node name of host OS, weight of SF)
   - Settings of the Host OS failover function (CF node name of host OS)
   - Startup of the shutdown facility

2. Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes to check that the cluster settings are correct after Offline Migration.

   ```
   # sdtool -s
   ```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

Perform the procedure from Step 1 again.

G.3.3 When performing Migration by Export/Import

G.3.3.1 When not using the Host OS failover function

G.3.3.1.1 Operations before Migration by Export/Import

Operations before Migration by Export/Import in a KVM environment are not required.

G.3.3.1.2 Operations after Migration by Export/Import

This section describes operations after Migration by Export/Import in a KVM environment.

1. Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes to check that the cluster settings are correct after Migration by Export/Import.

```
# sdtool -s
```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

   Perform the procedure in "G.3.1.2.2 Operations after Live Migration."

G.3.3.2 When using the Host OS failover function

G.3.3.2.1 Operation before Migration by Export/Import

Operations before Migration by Export/Import in a KVM environment are not required.

G.3.3.2.2 Operation after Migration by Export/Import

This section describes operations after Migration by Export/Import in a KVM environment.

1. Checking the status of the shutdown facility (guest OS)

   Execute the following command on all the nodes to check that the cluster settings are correct after Migration by Export/Import.

```
# sdtool -s
```

   Note

   If TestFailed or InitFailed is displayed, there is a possibility that the settings of the shutdown facility were not changed.

   Perform the procedure in "G.3.1.2.2 Operations after Live Migration."

G.4 Changing Settings

This section describes the procedures to change the settings when using the migration function in KVM environment.

G.4.1 Canceling Prerequisites

When the migration of the guest OS is no longer necessary or before uninstalling PRIMECLUSTER from the host OS, take the following steps to cancel the prerequisites for using the migration function.

Without using the Host OS failover function

1. Setting up the guest OS (host OS/guest OS)

   Take the following steps on the guest OS when the migration for this OS is no longer necessary.

   You can perform this procedure on multiple guest OSes at the same time, or on each guest OS one after another.

   1. Stopping of guest OS

      Execute the following command on the guest OS to stop the guest OS.

      ```
      # /sbin/shutdown -P now
      ```

   2. Canceling the settings to refer to the host OS information

      On the host OS where the guest OS is stopped, execute the following command to cancel the settings to refer to the host OS information file from the guest OS.

      ```
      # virsh detach-disk domain vdpcl --persistent
      ```

      domain

      Specify the domain name of the guest OS.

   3. Startup of guest OS

      Start the guest OS.

2. Deleting the host OS information file (host OS)

   If the migration is no longer necessary for all the guest OSes, execute the following command on each host OS to delete the host OS information file.

   ```
   # rm /var/opt/SMAWsf/sfkvmmigrate.img
   ```

Using the Host OS failover function

1. Setting up the guest OS (host OS/guest OS)

   Take the following steps on the guest OS when the migration for this OS is no longer necessary, or on all the guest OSes on the host OS when PRIMECLUSTER is uninstalled from this host OS.

   You can perform this procedure on multiple guest OSes at the same time, or on each guest OS one after another.

   1. Stopping of guest OS

      Execute the following command on the guest OS to stop the guest OS.

      ```
      # /sbin/shutdown -P now
      ```

   2. Canceling the settings to refer to the host OS information

      On the host OS where the guest OS is stopped, execute the following command to cancel the settings to refer to the host OS information file from the guest OS.

      ```
      # /opt/SMAW/SMAWsf/bin/sfkvmmigratesetup -r domain
      ```

      domain

      Specify the domain name of the guest OS.

   3. Startup of guest OS

      Start the guest OS.

2. Deleting the host OS information file (host OS)

   If the migration is no longer necessary for all the guest OSes, execute the following command on each host OS to delete the host OS information file.

   ```
   # /opt/SMAW/SMAWsf/bin/sfkvmmigratesetup -d
   ```

Appendix H Using PRIMECLUSTER in a VMware Environment

This appendix explains how to use PRIMECLUSTER in a VMware environment.

See

For details on VMware, see the documentation for VMware.

Note

Supported configuration

- The following environments and functions are not supported:
  - Cluster configuration between ESXi hosts with different versions
  - N-Port ID Virtualization (NPIV)
- Support for multipath software from third parties, contact field engineers.

Using VMware functions

- Remove a virtual machine that installs PRIMECLUSTER from targets for cluster functions (such as VMware HA, VMware FT, VMware DRS, and VMware DPM) of VMware.
- The following functions are not available in a virtual machine in which PRIMECLUSTER is to be installed.
  - Migration with VMware vCenter Converter
  - Snapshot of VMware
  - Backup by Data Protection
- Following hot swap operations cannot be performed for the virtual machine hardware.
  - Increasing disk size
  - Increasing memory
  - Increasing CPU
  - Using snapshot
  - Over committing of memory that causes virtual swap or memory ballooning

H.1 Cluster Systems in a VMware Environment

When using PRIMECLUSTER in a VMware environment, clustering (virtual machine function) between guest OSes on multiple ESXi hosts are available.

When an error occurs on a guest OS within a VMware environment, applications on that guest OS will no longer work. With PRIMECLUSTER applied to guest OSes, applications will failover from the active guest OS to a standby guest OS in the event of a failure, which creates a highly reliable guest OS environment.

Stopping virtual machine forcibly

For the cluster system in VMware environment, choose either of two functions available to stop the virtual machine forcibly which are "VMware vCenter Server functional cooperation" and "I/O fencing function."

To stop the operation node certainly and then fail over the operation when an error occurs in the guest OS or in the virtual machine, it is generally recommended to set up the forcible stop with the VMware vCenter Server functional cooperation.

However, set up the forcible stop with the I/O fencing function in the following cases:

- VMware vCenter Server is disabled, or the guest OS cannot communicate with VMware vCenter Server or cannot operate VMware vCenter Server.
- Upgrading from the VMware environment of PRIMECLUSTER 4.3A40 or earlier in which the I/O fencing function is used.
- Failing over the operation or cutting off the standby node when an error occurs in ESXi or in the server.

Note

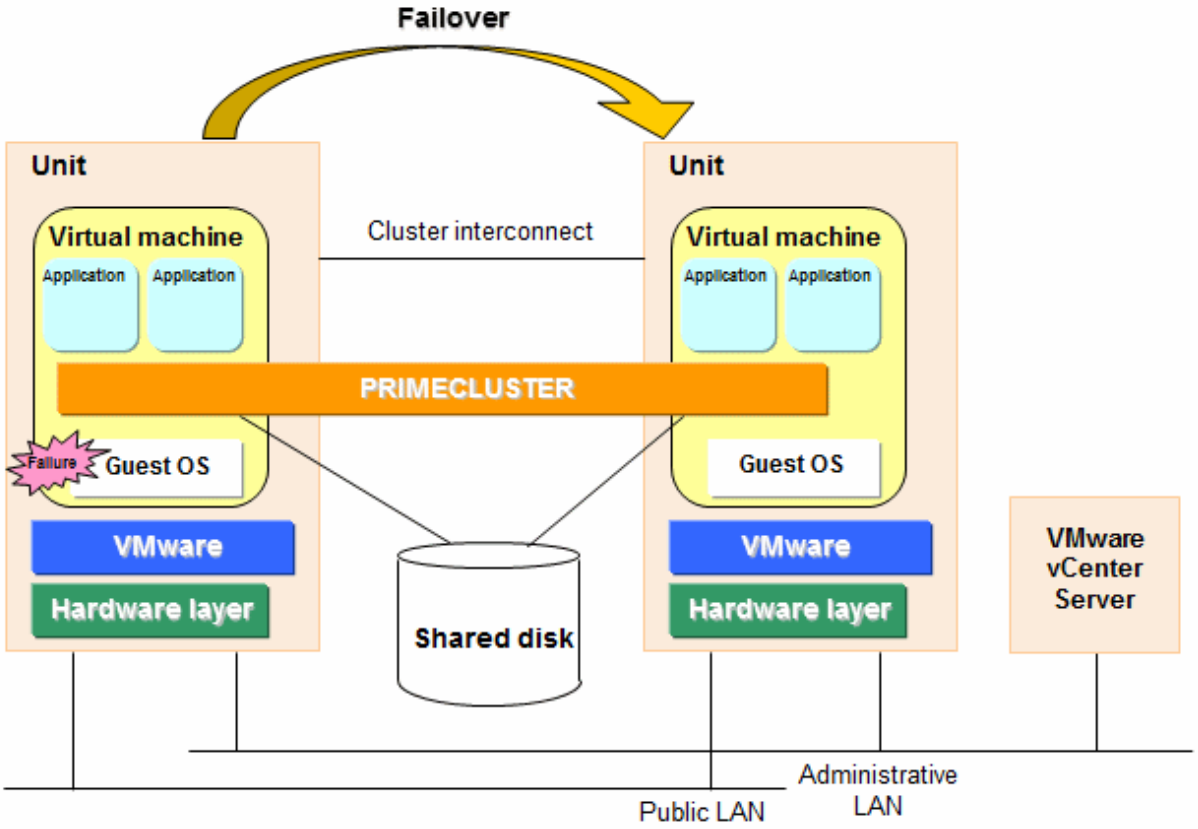
- Note the following points when using the forcible stop with the I/O fencing function:
  - The guest OS on which the cluster application is started panics regardless the survival priority if the cluster partition occurs due to failure of the cluster interconnect.
  - If the operation node panics when the operation is failed over, the status of cluster application may become Online temporarily on both operation and standby guest OSes. However, as access to the shared disk from both guest OSes at the same time is prevented, there is no impact on the operation.
- The operation is not failed over by the forcible stop with the VMware vCenter Server functional cooperation when an error occurs in ESXi or in the server. The operation becomes LEFTCLUSTER at this time.

Forcible stop with VMware vCenter Server functional cooperation (recommended)

When a failure occurs in a guest OS, the virtual machine of the error OS is powered off forcibly by linking with VMware vCenter Server. By this process, an operation can be failed over.

This process is enabled to stop a virtual machine in the cluster environment without a shared disk, or in the cluster environment between guest OSes on a single ESXi host.

Figure H.1 Cluster Systems in a VMware Environment (VMware vCenter Server functional cooperation)



Forcible stop with I/O fencing function

Use SCSI-3 Persistent Reservation as the exclusive control function to panic and stop the failed guest OS. By this operation, the operation can be switched. This process does not require VMware vCenter Server. It means that a guest OS can be panicked without any other servers besides the virtual machines that configure the cluster. However, a shared disk connected via RDM (Raw Device Mapping) and available with SCSI-3 Persistent Reservation is required.

Note

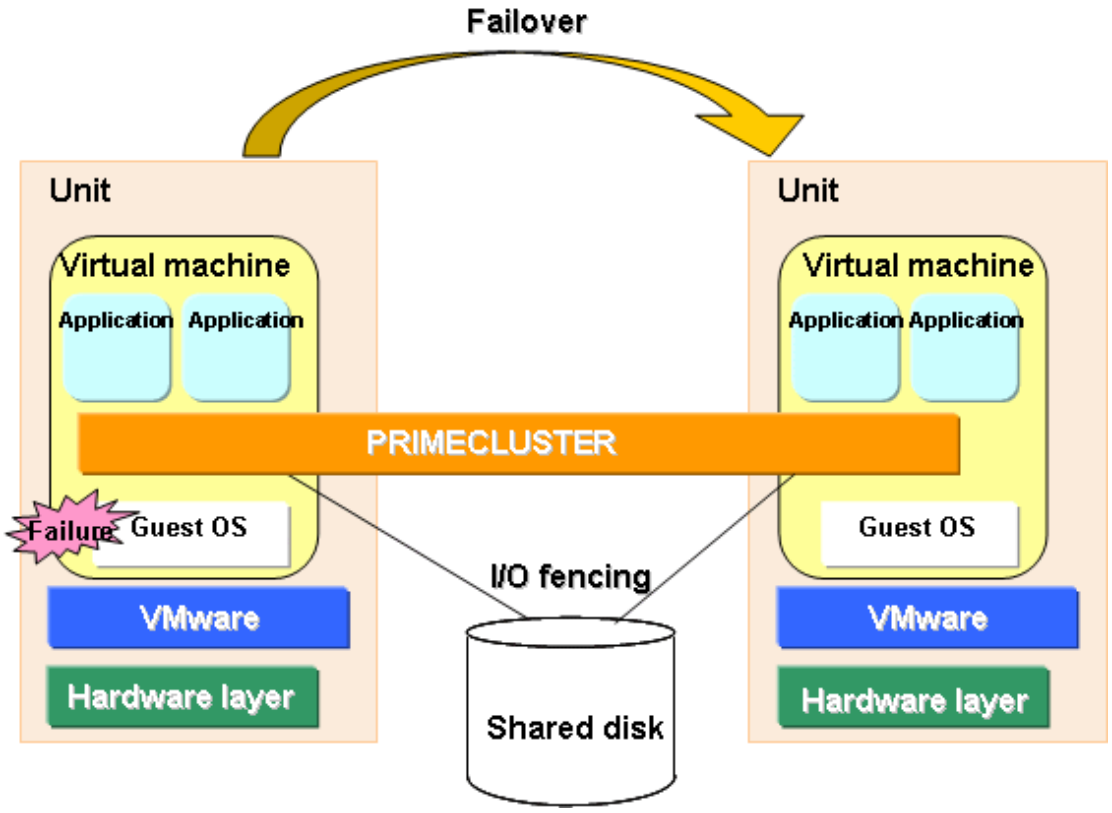
A forcible stop with the I/O fencing function is disabled in the following environments.

- Environment between guest OSes on a single ESXi host
- Environment in which the cluster application is configured with 3 or more nodes
- Environment in which multiple cluster applications that use a shared disk exist
- When using the disk configured with GDS mirroring among servers
- When using PRIMECLUSTER Wizard for SAP HANA

Information

In the cluster configuration where the I/O fencing function is used, by setting the SA_icmp shutdown agent, response from the guest OSes is checked on the network paths (administrative LAN/interconnect). The cluster application will be switched when no response is confirmed from the guest OSes. In this case, if the failed guest OS does not stop completely (when the OS is hanging, for example), both guest OSes may access the shared disk at the same time. By using SCSI-3 Persistent Reservation, the I/O fencing function prevents both guest OSes from accessing the shared disk at the same time. (To prevent the access from both guest OSes in the configuration where the VMware vCenter Server function is used, stop the failed guest OS completely before switching the guest OS.)

Figure H.2 Cluster Systems in a VMware Environment (I/O fencing function)



The comparison table below shows the forcible stop with VMware vCenter Server functional cooperation and the forcible stop with the I/O fencing function.

Item		Function to stop a virtual machine forcibly	
		VMware vCenter Server functional cooperation (recommended)	I/O fencing function
Configuration	VMware vCenter Server	Required (The guest OSes can communicate with VMware vCenter Server or operate VMware vCenter Server. Also in VMware vCenter Server, the user who is authorized to stop an operating virtual machine in the cluster must be created)	Optional
	Cluster configuration between guest OSes on a single ESXi host	Allowed	Not allowed
	Number of nodes that configure the cluster application	2 to 16 nodes	2 nodes
	Cluster application configuration	Unlimited	Allowed only one of the following configurations: - Only one cluster application - Among multiple cluster applications, only one of them contains a shared disk.
	Settings of survival priority	Allowed	Not allowed (regardless of the survival priority, a guest OS on which cluster applications are started panics)
	Shared disk	Optional Following disks are available: - Virtual disk created on the data store that can be accessed from each ESXi host - RDM (Raw Device Mapping) disk) Note: When the disk is shared between the cluster nodes, for both the virtual disk and the RDM disk, the number of shared ESXi hosts must be within 8. If the number of shared ESXi hosts is within 8, up to 16 cluster nodes can share the disk.	Required (Shared RDM (Raw Device Mapping) disk supporting SCSI-3 Persistent Reservation)
	Path policy for the Native Multipathing (NMP)	All supported	Only either of "Most Recently Used" or "Round Robin" is supported.
	PRIMECLUSTER Wizard for SAP HANA	Allowed	Not allowed

Item		Function to stop a virtual machine forcibly	
		VMware vCenter Server functional cooperation (recommended)	I/O fencing function
	Other unsupported configurations and functions	- VMware vSphere HA - VMware vSphere FT - VMware vSphere DRS - VMware vSphere DPM - Snapshot function - Backup by Data Protection - Suspending the virtual machine	- VMware vSphere HA - VMware vSphere FT - VMware vSphere DRS - VMware vSphere DPM - Snapshot function - Backup by Data Protection - Suspending the virtual machine - FCoE connection for storages - VMware vSphere vMotion - VMware vSphere Storage vMotion
Operation when an error occurs	Error in cluster interconnect	An operating node or a standby node is forcibly stopped, and an operation is failed over or the standby node is cut off.	- Only the cluster interconnect is specified for SA_icmp: An old operating node may panic due to the I/O fencing function even when an operation is failed over. - The cluster interconnect and any other networks are specified for SA_icmp: An operating node becomes LEFTCLUSTER state, and an operation is not failed over.
	Error in operating guest OS or in virtual machine	An operating node is forcibly stopped, and an operation is failed over.	An operating node panics, and an operation is failed over.
	Error in standby guest OS or in virtual machine	A standby node is forcibly stopped and then cut off.	A standby node is cut off (the standby node does not panic).
	Failure in ESXi or in server	An operation is not failed over. A node on the error ESXi becomes LEFTCLUSTER.	An operation is failed over or a standby node is cut off (an operating node panics while the standby node does not panic).
	Failure in VMware vCenter Server	A virtual machine cannot be forcibly stopped	-
	Failure in network between a virtual machine and VMware vCenter Server	A virtual machine cannot be forcibly stopped	-
	Dump collection when an error occurs	Not allowed (Forcible stop by power-off is only allowed. In this case, a cause of the error of the cluster node may not be determined.)	Allowed
Restrictions in maintenance	When using Cold Migration	None	If the migration is performed to operate two nodes that configure the cluster on a single ESXi host, an operation cannot be failed over when an error occurs either in a guest OS, a virtual machine, and the cluster interconnect.

Note

- Make sure to set either one of VMware vCenter Server functional cooperation or the I/O fencing function. A configuration with both functions or a configuration with neither of them is not allowed.

H.2 Installation

This section describes procedures for installing PRIMECLUSTER between guest OSes on multiple ESXi hosts in a VMware environment.

Note

I/O fencing function

- The I/O fencing function must be set up at the earlier stage of configuring the cluster application.
- The I/O fencing function uses the LUN on the shared disk unit registered to GDS disk class, or uses the LUN which contains the file system managed by the Fsystem resource. When using the I/O fencing function, register the Gds resource of the disk class containing the LUN or the disk, or register the Fsystem resource to the cluster application.
- The I/O fencing function cannot be used in the environment where the Gds resources and Fsystem resources are respectively registered in the multiple cluster applications.
- In the cluster application where a disk is not managed by the Fsystem resource or GDS, do not set the I/O fencing function.
- Set the path policy for the Native Multipathing (NMP) as "Most Recently Used" or "Round Robin".

  No other settings are supported.

Fsystem resource

- When using the file system that is created on the shared disk as Fsystem resources, you need to register all the file systems that are created on the same disk (LUN) or on the same disk class to the same userApplication. Due to the restriction of the I/O fencing function, you cannot create multiple file systems on one disk (LUN) or on one disk class and register each file system to the different userApplications to monitor and control them.
- In /etc/fstab.pcl file, add either of the following description formats to specify the devices of the file systems controlled by Fsystem resources.
  - When using GDS
    - /dev/sfdsk/<*disk_class_name*>/dsk/<*volume_name*>
  - Without using GDS
    - /dev/disk/by-id/ *name*
    - /dev/disk/by-path/ *name*
    - LABEL=<*file_system_label_name*>
    - UUID=<*file_system_UUID*>
    - /dev/*sd name*

H.2.1 Software Installation

Install the software required for PRIMECLUSTER on each node.

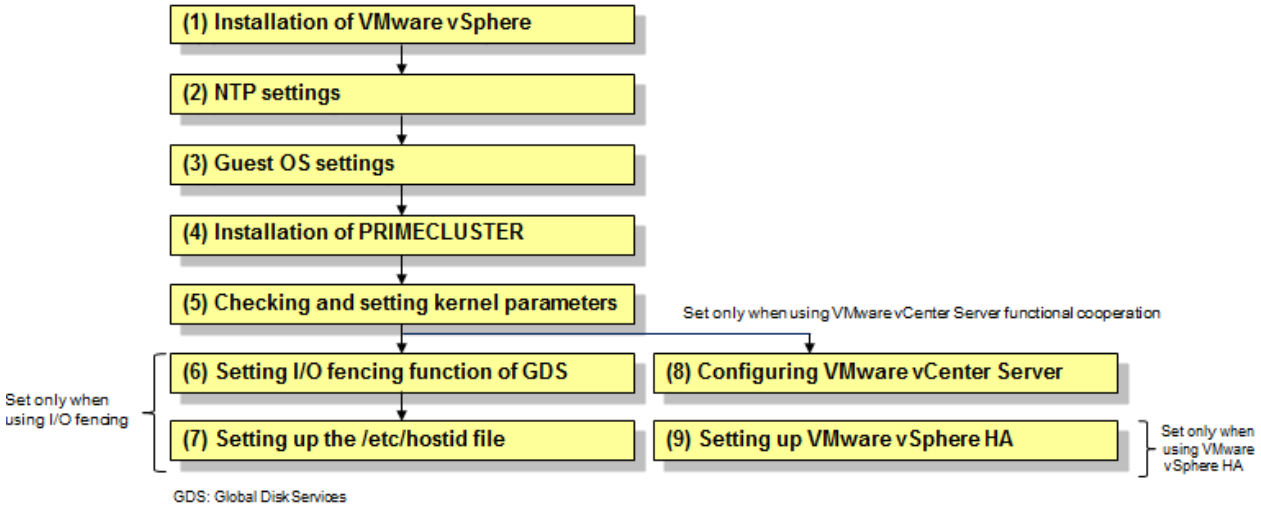
The explanation is divided into the following topics:

- Installation and configuration of related software
- Installation and environment configuration of applications

H.2.1.1 Installation and Configuration of Related Software

After installing the software related to PRIMECLUSTER, you need to take it into operation and make various settings for the OS and the hardware.

Perform the following steps as necessary.



1. Installation of VMware vSphere

   Take the following steps to set system disks and related devices, shared disks and related devices, and the virtual network.

   - Setting up system disks and related devices
     - When you create a new virtual machine by using vSphere Client or vSphere Web Client, select [Eager Zeroed] to set provisions of the system disk.
     - For types of SCSI controllers, set to "LSI Logic Parallel" or "VMware Paravirtual".
     - Set to "None" for sharing of the SCSI bus.
   - Setting up shared disks (when using the I/O fencing function)
     - When creating the first virtual machine, create shared disks taken over in the cluster system with Raw DeviceMapping (RDM). For the second virtual machine, select "Use an existing virtual disk" and specify the first RDM disk you created.
     - The Raw Device Mapping (RDM) file (.vmdk) of the shared disk taken over in the cluster system should be created in another data store besides the shared disk, and then share that data store between two ESXi hosts.
     - Set the compatibility mode if shared disk to "Physical."
     - For virtual device nodes, use a new SCSI controller which is different from the system disk.

       (Example: For the SCSI disk [SCSI(X:Y)], X indicates the controller number, and Y indicates the disk number. When the virtual device node of system disk is [SCSI(0:0)], do not use the virtual device node with the controller number 0 [SCSI(0:Y)]. Use [SCSI(1:0)] etc.)
     - Set the controller number and the disk number of virtual device nodes to be consistent among all the nodes that configure the cluster system.
     - For types of SCSI controllers, set the same type as the system disk on a guest OS.
     - For sharing SCSI buses, set to "Physical."
     - For ESXi hosts, it is necessary to mark as "Permanent Reservation" with the disk device used for shared disk of PRIMECLUSTER.

       Use the following esxcli command to mark the device as permanent reservation.

       ```
       esxcli storage core device setconfig -d <naa.id> --perennially-reserved=true
       ```

See KB1016106 in the Knowledge Base site of VMware Inc. for configuration instructions.
http://kb.vmware.com/kb/1016106

- Setting up shared disks (when using the function to stop the link with VMware vCenter Server)
  - To use the virtual disk as the shared disk, create the data store shared with each ESXi host. Create the virtual disk in this data store.
  - For virtual device nodes, use a new SCSI controller which is different from the system disk.

    (Example: For the SCSI disk [SCSI(X:Y)], X indicates the controller number, and Y indicates the disk number. When the virtual device node of system disk is [SCSI(0:0)], do not use the virtual device node with the controller number 0 [SCSI(0:Y)]. Use [SCSI(1:0)] etc.)
  - Set the controller number and the disk number of virtual device nodes to be consistent among all the nodes that configure the cluster system.
  - For types of SCSI controllers, set the same type as the system disk on a guest OS.
  - For sharing SCSI buses, set as follows:
    - In the cluster environment between guest OSes on a single ESXi host

      ```
      [Virtual]
      ```

    - In the cluster environment between guest OSes on multiple ESXi hosts

      ```
      [Physical]
      ```

- Setting up the virtual network
  - When creating the virtual machine, create at least two network systems for the cluster interconnect and connect them to different physical adapters.
  - For sharing the physical network adapter that is used as the cluster interconnect with multiple clusters, allocate a different port group to each cluster system for a vSwitch. In this case, set different VLAN ID to each port group.

When bundling the network that is specified to the interconnect by using NIC teaming of VMware, make sure to use any one of the following configurations to set the load balancing option (active-active configuration) to NIC teaming.

1. Route based on source port ID
2. Route based on source MAC hash
3. Use explicit failover order

Redundant configuration (active-standby) is enabled in any configurations other than the above configurations 1 to 3.

2. NTP settings (Guest OS)

   These settings serve to synchronize the time of each node in the cluster system configuration. Be sure to make these settings when you configure the cluster.

   Make these settings on the guest OS before you install PRIMECLUSTER.

3. Guest OS settings (Guest OS)

   Take the following steps to set the guest OS.

   - File system settings for system volume

     If an I/O device where the system volume is placed fails, a cluster failover does not occur and the system operation may continue based on the data stored on the memory.

     If you want PRIMECLUSTER to trigger a cluster failover by panicking a node in the event that an I/O device where the system volume is placed fails, set the ext3 or the ext4 file system to the system volume and perform the following setting.

Setting

Specify "errors=panic" to the mount option of each partition (the ext3 or the ext4 file system) included in the system volume.

Example: To set it in /etc/fstab (when /, /var, and /home exist in one system volume)

```
LABEL=/     /     ext3 errors=panic 1 1
LABEL=/boot /boot ext3 errors=panic 1 2
LABEL=/var  /var  ext3 errors=panic 1 3
LABEL=/home /home ext3 errors=panic 1 4
```

However, an immediate cluster failover may not become available due to taking time for an I/O error to reach the file system. The regularly writing to the system volume enhances the detection frequency of I/O error.

- Network settings

  In the guest OS in the cluster system, it is necessary to make network settings such as IP addresses for the public LAN and the administrative LAN.

  Implement these settings on the guest OS that you are going to run as a cluster.

4. Installation of PRIMECLUSTER (Guest OS)

   For installing PRIMECLUSTER, an installation script (CLI Installer) is available.

   This script method installs PRIMECLUSTER node by node on systems that already have Linux(R) and related software installed. It is also utilized for installation on cluster management servers.

   See

   For details on the installation procedure, see the Installation Guide for PRIMECLUSTER.

5. Checking and setting the kernel parameters

   Depending on the environment, the kernel parameters must be modified.

   Applicable nodes:

   All the nodes on which PRIMECLUSTER is to be installed

   Depending on the utilized products and components, different kernel parameters are required.

   Check PRIMECLUSTER Designsheets and modify the settings as necessary.

   See

   For details on the kernel parameters, see "3.1.7 Checking and Setting the Kernel Parameters."

6. Setting the I/O fencing function of GDS

   When using the I/O fencing function, set up the I/O fencing function of GDS.

   Add the following line into the /etc/opt/FJSVsdx/sdx.cf file:

   ```
   SDX_VM_IO_FENCE=on
   ```

   Applicable nodes:

   All the nodes on which PRIMECLUSTER is to be installed.

7. Setting up the /etc/hostid file

   Set hostid that is used with the I/O fencing function.

   According to the following steps, check whether setting up the /etc/hostid file is required, and then, set it up if needed.

   How to check
   Execute the hostid command and check the output.

When the output is other than "00000000," setting up the /etc/hostid file is not necessary.

```
# hostid
a8c00101
```

When the output is "00000000," follow the setting procedure below to set the host identifier (output of hostid) on all the nodes that configure the cluster. For the host identifier, specify the value unique to each node. Do not set 00000000 for the value.

Setting procedure

1. Create the /etc/hostid file.

   ```
   # touch /etc/hostid
   ```

2. Create the following python script file.
   [Contents of the file to be created]

   ```
   #!/usr/bin/python
   from struct import pack
   filename = "/etc/hostid"
   hostid = pack("I",int("0x<hhhhhhhh>",16))
   open(filename, "wb").write(hostid)
   ```

   (<*hhhhhhhh*>: Describe the intended host identifier in base 16, 8 digit numbers.)

3. Set the execute permissions to the created script file and then, execute it.

   ```
   # chmod +x <created script file name>
   # ./<created script file name>
   ```

4. Execute the hostid command to check if the specified host identifier is obtained.

   ```
   # hostid
   hhhhhhhh
   ```

   (hhhhhhhh: host identifier that is specified in the script file)

8. Configuring VMware vCenter Server

   When using VMware vCenter Server functional cooperation, configure VMware vCenter Server.

   For how to configure VMware vCenter Server, see the documentation published by VMware.

   Also take the following steps after configuring VMware vCenter Server.

   1. For VMware vCenter Server functional cooperation, add the roles to which the following authorities are applied to VMware vCenter Server:

      - Virtual machine-Interaction-Power-off
      - Virtual machine-Interaction-Power-on

      If the roles cannot be added, check the registered roles that have the above authorities.

   2. For VMware vCenter Server functional cooperation, create the user in VMware vCenter Server.
   3. Add the user created in step 2 to the authority of the virtual machine that is used as the cluster. Apply the roles that are added or checked in step 1 to this user.

Note

- If the route from the virtual machine to VMware vCenter Server is interrupted, the virtual machine cannot be forcibly stopped. In this case, configuring the route to VMware vCenter Server to be redundant is recommended.
- Do not include "\" in the virtual machine name. If it is included, the virtual machine cannot be forcibly stopped normally.

Note

- To activate the modified kernel parameters and the I/O fencing function of GDS, restart the guest OS after installation settings for related software is complete.
- When using the VMware vCenter Server functional cooperation, do not include "\" in the virtual machine name. If it is included, the virtual machine cannot be forcibly stopped normally.

H.2.1.2 Installation and Environment Configuration of Applications

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

See

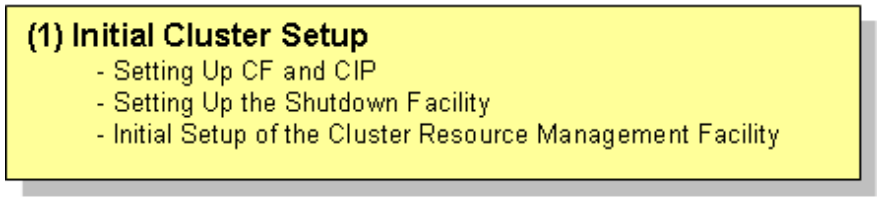
- For details on environment setup, see manuals for each application.
- For information on PRIMECLUSTER-related products supporting VMware, see the documentation for each product.

H.2.2 Preparation Prior to Building a Cluster

Refer to "Chapter 4 Preparation Prior to Building a Cluster" to make the initial cluster setup on the guest OS.

H.2.3 Building a Cluster

This section describes procedures for setting up a cluster with PRIMECLUSTER in a VMware environment.



H.2.3.1 Initial Setup of CF and CIP

Refer to "5.1.1 Setting Up CF and CIP" to set up CF and CIP on the guest OS.

H.2.3.2 Setting Up the Shutdown Facility (when using VMware vCenter Server Functional Cooperation)

For details on survival priority, see "5.1.2.1 Survival Priority."

In VMware environments, when an error occurs in a guest OS, the virtual machine of the error OS is powered off forcibly by cooperating with VMware vCenter Server. By this process, an operation can be failed over.

This section explains the method for setting up the SA_vwvmr shutdown agent as the shutdown facility.

Note

Be sure to perform the following operations on all guest OSes (nodes).

1. Encrypting the password

   Execute the sfcipher command to encrypt passwords for accessing VMware vCenter Server.

   For details on how to use the sfcipher command, see the manual page of "sfcipher."

```
# sfcipher -c
Enter User's Password:
Re-enter User's Password:
D0860AB04E1B8FA3
```

2. Setting up the shutdown agent

   Specify the shutdown agent.

   Create /etc/opt/SMAW/SMAWsf/SA_vwvmr.cfg with the following contents on all guest OSes (nodes) of the cluster:

```
# comment line
CFName: cfname1
VMName: vmname1
vCenter_IP: ipaddress1
vCenter_Port: port
user: user
passwd: passwd
# comment line
CFName: cfname2
VMName: vmname2
vCenter_IP: ipaddress2
vCenter_Port: port2
user: user
passwd: passwd
```

```
cfnameX           : Specify the CF node name.
vmnameX           : Specify the virtual machine name that controls the guest OS described
                    in CFName.
ipaddressX        : Specify the IP address of VMware vCenter Server that manages the virtual
                    machine.
                    Available IP addresses are IPv4 and IPv6 addresses.
                    IPv6 link local addresses are not available.
                    When specifying the IPv6 address, enclose it in brackets "[ ]".
                    (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
portX             : Specify the port number of VMware vCenter Server.
                    When using the default value (443), describe "vCenter_Port:". Do not specify
                    the parameter.
user              : Specify the user of VMware vCenter Server created in
                    "H.2.1.1 Installation and Configuration of Related Software."
                    When logging in with single sign-on (SSO), specify user@SSO_domain_name.
passwd            : A login password of the account specified by "user".
                    Specify the encrypted password encrypted in 1.
```

Note

- Do not change the order of each item.
- If the virtual machine name (VMName:) includes a Japanese character, use the character code UTF-8 to describe the machine name.
- One-byte space and a double-byte space is used as a different character. Use one-byte space when inserting a space in the file.
- Only the line start with "#" is treated as a comment. When "#" is in the middle of a line, this "#" is treated as a part of the setting value.

  In the following example, "vm1 # node1's virtual machine." is used as the virtual machine name.

```
...
VMName: vm1 # node1's virtual machine.
...
```

- The contents of SA_vwvmr.cfg must be the same on all the guest OSes. If not, the shutdown facility may not work correctly.

Example

- Log in with single sign-on

  When the IP address of VMware vCenter Server that manages all the virtual machines is 10.20.30.40, the port numbers are the default value, the user who connects to VMware vCenter Server is Administrator, SSO domain name is vsphere.local, and the password encrypted in step "1. Encrypting the password" is D0860AB04E1B8FA3:

```
##
## node1's information.
##
CFName: node1
VMName: vm1
vCenter_IP: 10.20.30.40
vCenter_Port:
user: Administrator@vsphere.local
passwd: D0860AB04E1B8FA3
##
## node2's information.
##
CFName: node2
VMName: vm2
vCenter_IP: 10.20.30.40
vCenter_Port:
user: Administrator@vsphere.local
passwd: D0860AB04E1B8FA3
```

- Log in without single sign-on.

  When the IP address of VMware vCenter Server that manages all the virtual machines is 10.20.30.40, the port numbers are the default value, the user who connects to VMware vCenter Server is root, and the password encrypted in step "1. Encrypting the password" is D0860AB04E1B8FA3:

```
##
## node1's information.
##
CFName: node1
VMName: vm1
vCenter_IP: 10.20.30.40
vCenter_Port:
user: root
passwd: D0860AB04E1B8FA3
##
## node2's information.
##
CFName: node2
VMName: vm2
vCenter_IP: 10.20.30.40
vCenter_Port:
user: root
passwd: D0860AB04E1B8FA3
```

3. Setting up the shutdown daemon

   Create /etc/opt/SMAW/SMAWsf/rcsd.cfg with the following contents on all guest OSes (nodes) of the cluster:

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_vwvmr,timeout=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_vwvmr,timeout=timeout
```

```
CFNameX        : CF node name of the cluster host.
weight         : Weight of the SF node.
myadmIP        : Specify the IP address of the administrative LAN for CFNameX.
                 Available IP addresses are IPv4 and IPv6 addresses.
                 IPv6 link local addresses are not available.
                 When specifying the IPv6 address, enclose it in brackets "[ ]".
                 (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
                 If you specify a host name, please make sure it is listed in /etc/hosts.
timeout        : Specify the timeout duration (seconds) of the Shutdown Agent.
                 Specify 45 for the value.
```

Note

The rcsd.cfg file must be the same on all guest OSes (nodes). Otherwise, operation errors might occur.

Example

Below is the setting examples:

```
node1,weight=1,admIP=10.0.0.1:agent=SA_vwvmr,timeout=45
node2,weight=1,admIP=10.0.0.2:agent=SA_vwvmr,timeout=45
```

4. Starting the shutdown facility

   Check that the shutdown facility has started.

   ```
   # sdtool -s
   ```

   If the shutdown facility has already started, execute the following command to restart the shutdown facility.

   ```
   # sdtool -r
   ```

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

   ```
   # sdtool -b
   ```

5. Checking the status of the shutdown facility

   Check that the status of the shutdown facility is either "InitWorked" or "TestWorked." If the displayed status is "TestFailed" or "InitFailed," check the shutdown daemon settings for any mistakes.

   ```
   # sdtool -s
   ```

H.2.3.3 Setting Up the Shutdown Facility (when using I/O fencing function)

This section explains the method for setting up the SA_icmp shutdown agent as the shutdown facility.

Note

Be sure to perform the following operations on all guest OSes (nodes).

1. Setting up the shutdown facility

   Specify the shutdown agent.

   Create /etc/opt/SMAW/SMAWsf/SA_icmp.cfg with the following contents on all guest OSes (nodes) of the cluster:

```
TIME_OUT=value
cfname:ip-address-of-node:NIC-name1,NIC-name2
```

value : Specify the interval (in seconds) for checking whether the node is alive. The recommended value is "5" (s).
cfname : Specify the name of the CF node.
ip-address-of-node : Specify the IP addresses of any one of the following networks utilized for checking whether the cfname node is alive.
Checking via multiple networks is also available.
In this case, add a line for each utilized network.
To check LAN paths, we recommend that you use multiple ones to surely determine an error.
However, if you prioritize to switch over automatically to surely determine an error, set only cluster interconnects to the LAN paths.
If only cluster interconnects are set to the LAN paths, the automatic switchover is available even though communication is disabled between cluster interconnects but available via other LAN (when you determined that the node in the communication destination is alive).
- Cluster interconnect (IP address of CIP)
- Administrative LAN
- Public LAN

Available IP addresses are IPv4 and IPv6 addresses.
IPv6 link local addresses are not available.
When specifying the IPv6 address, enclose it in brackets "[ ]".
(Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
Enter the IP address for all guest OSes (nodes) that configure the cluster system.
NIC-nameX : Specify the network interface of the local guest OS (node) utilized for checking whether the node defined by ip-address-of-node is alive.
If there is more than one, delimit them with commas (",").

Note

Registering network interfaces

- For duplicating by GLS, define all redundant network interfaces. (Example: eth0,eth1)
- If you are bonding NICs, define the bonding device behind the IP address. (Example: bond0)
- For registering the cluster interconnect, define all network interfaces that are used on all paths of the cluster interconnect. (Example: eth2,eth3)
- Do not use the takeover IP address (takeover virtual Interface).

Example

Below indicates the setting example of clusters (consisted by 2 nodes) between guest OSes on multiple ESXi hosts.

- When cluster interconnects (eth2,eth3) are set

```
TIME_OUT=5
node1:192.168.1.1:eth2,eth3
node2:192.168.1.2:eth2,eth3
```

- When the public LAN (duplicated (eth0,eth1) by GLS) and the administrative LAN (eth4) are set

```
TIME_OUT=5
node1:10.20.30.100:eth0,eth1
node1:10.20.40.200:eth4
```

```
node2:10.20.30.101:eth0,eth1
node2:10.20.40.201:eth4
```

2. Setting up the shutdown daemon

   Create /etc/opt/SMAW/SMAWsf/rcsd.cfg with the following contents on all guest OSes (nodes) of the cluster:

```
CFNameX,weight=weight,admIP=myadmIP:agent=SA_icmp,timeout=timeout
CFNameX,weight=weight,admIP=myadmIP:agent=SA_icmp,timeout=timeout
```

   CFNameX : CF node name of the cluster host.
   weight : Weight of the SF node.
   Set 1 because this value is not effective with the I/O fencing function.
   myadmIP : Specify the IP address of the administrative LAN for *CFNameX*.
   Available IP addresses are IPv4 and IPv6 addresses.
   IPv6 link local addresses are not available.
   When specifying the IPv6 address, enclose it in brackets "[ ]".
   (Example: [1080:2090:30a0:40b0:50c0:60d0:70e0:80f0])
   If you specify a host name, please make sure it is listed in /etc/hosts.
   timeout : Specify the timeout duration (seconds) of the Shutdown Agent.
   Specify the following values.
   (*TIME_OUT* + 2) X number of paths to be used for checking the survival of a node, or 20 (specify the larger value)
   TIME_OUT is the TIME_OUT value that is described in the SA_icmp.cfg.

   - When checking the survival of a node on the 1 path
     (either one of administrative LAN, public LAN, or cluster interconnects)
     (1) TIME_OUT is 18 or larger
         TIME_OUT + 2
     (2) TIME_OUT is less than 18
         20

   - When checking the survival of a node on the 2 paths
     (either two of administrative LAN, public LAN, or cluster interconnects)
     (1) TIME_OUT is 8 or larger
         (TIME_OUT + 2)X 2
     (2) TIME_OUT is less than 8
         20

   - When checking the survival of a node on the 3 paths
     (three of administrative LAN, multiple public LANs, or public LAN, or cluster interconnects)
     (1) TIME_OUT is 5 or larger
         (TIME_OUT + 2)X 3
     (2) TIME_OUT is less than 5
         20

Note

The rcsd.cfg file must be the same on all guest OSes (nodes). Otherwise, operation errors might occur.

Example

Below indicates the setting example to check survival of a node by using administrative LAN and public LAN when TIME_OUT value described in the SA_icmp.cfg is 10, in a two-node configuration.

node1,weight=1,admIP=192.168.100.1:agent=SA_icmp,timeout=24 (*)

node2,weight=1,admIP=192.168.100.2:agent=SA_icmp,timeout=24 (*)
timeout = (10 (TIMEOUT value) + 2) X 2(administrative LAN, public LAN) = 24

3. Starting the shutdown facility

   Check that the shutdown facility has started.

```
# sdtool -s
```

   If the shutdown facility has already started, execute the following command to restart the shutdown facility.

```
# sdtool -r
```

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

```
# sdtool -b
```

4. Checking the status of the shutdown facility

   Check that the status of the shutdown facility is either "InitWorked" or "TestWorked." If the displayed status is "TestFailed" or "InitFailed," check the shutdown daemon settings for any mistakes.

```
# sdtool -s
```

H.2.3.4 Initial Setup of the Cluster Resource Management Facility

Refer to "5.1.3 Initial Setup of the Cluster Resource Management Facility" to set up the resource database managed by the cluster resource management facility (hereafter referred to as "CRM") on the guest OS.

H.2.3.5 Setting Up Fault Resource Identification and Operator Intervention Request

Refer to "5.2 Setting Up Fault Resource Identification and Operator Intervention Request" to make the settings for identifying fault resources and for requesting operator intervention.

H.2.4 Building Cluster Applications

For details on how to configure cluster applications on the guest OS, see "Chapter 6 Building Cluster Applications."

If using the I/O fencing function, when configuring cluster applications, you must set up the I/O fencing function for the cluster application where the Gds resources or the Fsystem resources are registered.

H.2.4.1 Setting Up I/O Fencing Function

This section explains how to set up the I/O fencing function for all cluster applications to which Gds resources or Fsystem resources are registered.

Perform the following to set up the I/O fencing function:

- Creating Cmdline resources and setting up Online and Offline scripts
- Setting up userApplication scripts (PreOnline, OfflineDone, and Fault scripts)

1. Creating Cmdline resources and setting up Online and Offline scripts

   1. In the Cmdline resource settings, add the Start script, the Stop script, and the Check script in the following format:

      <Start script>

      /opt/SMAW/bin/hvsgpr -c

      <Stop script>

      /opt/SMAW/bin/hvsgpr -u

      <Check script>

      /opt/SMAW/bin/hvsgpr -m

To create Cmdline resources, see, "6.7.3.1 Setting Up Cmdline Resources."

2. In the attribute settings of the Cmdline resources, set the AutoRecover attribute to disabled ("0"). Do not change the default settings for other attributes.

2. Setting up userApplication scripts

1. In the cluster application settings, add the PreOnline and OfflineDone scripts in the following format.

   <PreOnline script>

   /opt/SMAW/bin/hvsgpr -r

   <OfflineDone script>

   /opt/SMAW/bin/hvsgpr -o

   1. Select "(PreOnlineScript=)" of "Machines+Basics."

```
Consistency check ...

Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=vm21RMS
 8) Machines[1]=vm22RMS
 9) (PreCheckScript=)
10) (PreOnlineScript=)
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=yes)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (AutoBreakMaintMode=no)
25) (HaltFlag=yes)
26) (PartialCluster=0)
27) (ScriptTimeout=)
Choose the setting to process:10
```

   2. Select "FREECHOICE" and enter the full path of the PreOnline script.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming ONLINE:4
        >> /opt/SMAW/bin/hvsgpr -r
```

   3. Select "(OfflineDoneScript=)" of "Machines+Basics."

```
Consistency check ...

Machines+Basics (app1:consistent)
 1) HELP
```

```
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=vm21RMS
 8) Machines[1]=vm22RMS
 9) (PreCheckScript=)
10) (PreOnlineScript='/opt/SMAW/bin/hvsgpr~-r')
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript=)
14) (FaultScript=)
15) (AutoStartUp=yes)
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (AutoBreakMaintMode=no)
25) (HaltFlag=yes)
26) (PartialCluster=0)
27) (ScriptTimeout=)
Choose the setting to process:13
```

4. Select "FREECHOICE" and enter the full path of the OfflineDone script.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming ONLINE:4
        >> /opt/SMAW/bin/hvsgpr -o
```

2. In the attribute settings of the cluster application, if the HaltFlag attribute is set to enabled ("1"), add the Fault script in the following format.

   <Fault script>

   /opt/SMAW/bin/hvsgpr -f

   1. Select "(FaultScript=)" of "Machines+Basics."

```
Consistency check ...

Machines+Basics (app1:consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) AdditionalMachine
 6) AdditionalConsole
 7) Machines[0]=vm21RMS
 8) Machines[1]=vm22RMS
 9) (PreCheckScript=)
10) (PreOnlineScript='/opt/SMAW/bin/hvsgpr~-r')
11) (PostOnlineScript=)
12) (PreOfflineScript=)
13) (OfflineDoneScript='/opt/SMAW/bin/hvsgpr~-o')
14) (FaultScript=)
15) (AutoStartUp=yes)
```

```
16) (AutoSwitchOver=HostFailure|ResourceFailure|ShutDown)
17) (PreserveState=no)
18) (PersistentFault=0)
19) (ShutdownPriority=)
20) (OnlinePriority=)
21) (StandbyTransitions=ClearFaultRequest|StartUp|SwitchRequest)
22) (LicenseToKill=no)
23) (AutoBreak=yes)
24) (AutoBreakMaintMode=no)
25) (HaltFlag=yes)
26) (PartialCluster=0)
27) (ScriptTimeout=)
Choose the setting to process:14
```

2. Select "FREECHOICE" and enter the full path of the Fault script.

```
1) HELP
2) RETURN
3) NONE
4) FREECHOICE
Enter the command line to start prior to the application becoming ONLINE:4
        >> /opt/SMAW/bin/hvsgpr -f
```

When specifying any command other than hvsgpr command in PreOnline script, OfflineDone script, and Fault script at the same time, specify the command as any one of the following examples shows:

- Separate the command by double-ampersand (&&).

Example

```
/opt/SMAW/bin/hvsgpr -r && /var/tmp/command
```

- Separate the command by semicolon (;).

Example

```
/opt/SMAW/bin/hvsgpr -o ; /var/tmp/command
```

- Create the script that runs more than one commands, and then specify that command.

Example

```
#!/bin/sh

/opt/SMAW/bin/hvsgpr -r
ret1=$?
/var/tmp/command
ret2=$?
if [ $ret1 = 0 ]; then
    exit $ret2
fi
exit $ret1
```

The table below shows how the command can be specified in each script and the notes on specifying the command.

	Separate the command by double-ampersand (&&).	Separate the command by semicolon (;).	Create the script that runs more than one commands, and then specify that command.
PreOnline script	Y (*1)	-	Y (*2)
OfflineDone script	-	Y	Y
Fault script	-	Y	Y

Y: Allowed -: Not allowed

(*1) Specify hvsgpr command as the first executed command.
The second or later command is not executed if hvsgpr command has ended abnormally.
(*2) When hvsgpr command has ended abnormally, the script's exit code must be the same with hvsgpr's exit code.

H.3 Operations

For details on functions for managing PRIMECLUSTER system operations, see "Chapter 7 Operations."

Note

- When the hvswitch -f command is executed to start or switch the cluster application, the following message is output and starting or switching of the cluster application may fail.

  ```
  ERROR: Forcibly switch request denied, unable to kill node <SysNode name>
  ```

  This message is output when the node displayed as <*SysNode name*> is in the LEFTCLUSTER state. Perform the procedure in "5.2 Recovering from LEFTCLUSTER" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide." After that, start or switch the cluster application.

- Do not perform "Suspend operation" for the virtual machine on which the cluster is running. If "Suspend" is performed by mistake, an operation may not switch automatically. In this case, power off the virtual machine on which "Suspend" is performed, and then switch the operation manually.

H.4 Changing the Configuration

For details on additions or changes of configuration nodes in the PRIMECLUSTER system and cluster applications, see "Chapter 8 Changing the Cluster System Configuration."

H.5 Maintenance

For details on items and procedures required for maintenance of the PRIMECLUSTER system, see "Chapter 12 Maintenance of the PRIMECLUSTER System."

Appendix I Startup Scripts and Startup Daemons, and Port Numbers in PRIMECLUSTER

This appendix provides explanations on scripts and daemons that are started by PRIMECLUSTER, and the port numbers being used in RHEL6.

I.1 Explanation Formats

Scripts and daemons, and port numbers 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.

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

I.2 Startup Script Lists

init

init (inittab)

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Startup daemon

/usr/sbin/sdxmond

Utilized port

None.

Remarks

None.

init (inittab)

Function

Monitoring of shutdown facility.

Effect if stopped

If shutdown facility terminates abnormally, it will not be restarted.

Startup daemon

/opt/SMAW/SMAWsf/bin/rcsd_monitor

Utilized port

None.

Remarks

None.

/etc/rc3.d

S05poffinhibit

Function

Initializing kdump shutdown agent.

Effect if stopped

Forcible stop by kdump shutdown agent is disabled.

Startup daemon

None.

Utilized port

None.

Remarks

Enabled only in physical environment.

S06clonltrc

Function

Loading the driver of the online trace.

Effect if stopped

The information for investigation of the cluster resource management facility cannot be collected.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S07clapi

Function

Beginning of online trace of the Cluster Resource Management facility (1).

Effect if stopped

The cluster cannot be started.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S07cllkcd

Function

Initializing kdump shutdown agent.

Effect if stopped

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S08clrms

Function

Beginning of online trace of the Cluster Resource Management facility (2).

Effect if stopped

The cluster cannot be started.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S12cf

Function

Loading of CF and CIP drivers.

Effect if stopped

The cluster cannot be started.

Startup daemon

/opt/SMAW/SMAWcf/bin/cfregd

Utilized port

None.

Remarks

None.

S12zcldevmon

Function

Startup of MMB asynchronous monitoring.

Effect if stopped

MMB asynchronous monitoring cannot be used.

Startup daemon

/etc/opt/FJSVcluster/sys/devmmbd
/etc/opt/FJSVcluster/sys/devmmbmond
/etc/opt/FJSVcluster/sys/devmmbmonitord
/etc/opt/FJSVcluster/sys/devmalogd

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
7000-7100 (*1)	UDP	s, r	Administrative LAN	Cluster node	7000-7100	Remote cluster node
162 (*1)	UDP	r	Administrative LAN	Local cluster node	ANY	Remote cluster node

Remarks

(*1) These ports are used when SA_mmbp and SA_mmbr are set in the Shutdown Facility on PRIMEQUEST.

S13SMAWsf

Function

Startup of Shutdown Facility.

Effect if stopped

Shutdown Facility cannot be used.

Startup daemon

/opt/SMAW/SMAWsf/bin/rcsd

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9382 (*1)	UDP	s, r	Administrative LAN	Cluster node	ANY	Remote cluster node
ANY	UDP	s, r	Administrative LAN	Cluster node	623 (*2)	BMC/iRMC
ANY	UDP	s,r	Administrative LAN	Cluster node	161 (*3)	Management blade

Remarks

These ports are used to prevent split brain.

(*1) No. 9382 is set to support the service name "sfadv."
(*2) This port is used when SA_ipmi is set in the Shutdown Facility on PRIMERGY.
(*3) This port is used when SA_blade is set in the Shutdown Facility on the Blade server.

S11hanet

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.

Startup daemon

/opt/FJSVhanet/etc/sbin/hanetctld
/opt/FJSVhanet/etc/sbin/hanetselect (*1)
/opt/FJSVhanet/etc/sbin/hanetpathmd (*2)

Utilized port

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) This daemon is started only when GS linkage mode is used.
(*2) Availability of startup and the number of processes rely on the configuration. Also, this may be suspended according to the monitoring status.

S24hanet2

Function

Startup of monitoring daemon and self check daemon.

Effect if stopped

The line monitoring function and the self-checking function cannot work.

Startup daemon

/opt/FJSVhanet/etc/sbin/hanetmond (*1)
/opt/FJSVhanet/etc/sbin/hanetselect (*2)

Utilized port

None.

Remarks

(*1) This daemon is started only when the self-checking function is used.
(*2) Availability of startup and the number of processes rely on the configuration. Also, this may be suspended according to the monitoring status.

S27SMAWsfex

Function

Starting Configuration Update Service for shutdown agent.

Effect if stopped

Configuration Update Service for shutdown agent does not work.

Startup daemon

None.

Utilized port

None.

Remarks

Only when Starting Configuration Update Service for shutdown agent is enabled.

S51cldbm

Function

Startup of cluster configuration management facility.

Effect if stopped

The cluster cannot be started.

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 cluster node
9379 (*2)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9378 (*3)	TCP	s, r	Interconnect	Cluster node	ANY	Local cluster node
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

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9383 (*7)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node
9384 (*8)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node

Remarks

(*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."
(*7) No. 9383 is set to support the service name "dcmcom2."
(*8) No. 9384 is set to support the service name "dcmlck2."

S51clrmgr

Function

Startup of Cluster Resource Management facility.

Effect if stopped

The cluster cannot be started.

Startup daemon

/etc/opt/FJSVcluster/FJSVcldbm/daemons/clrmd

Utilized port

None

Remarks

None.

S51clrwz

Function

Setting of cluster applications.

Effect if stopped

Cluster applications cannot be configured correctly, or will not work correctly.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S52sfdsk

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Startup daemon

/usr/sbin/sdxlogd
/usr/sbin/sdxservd
/usr/sbin/sdxexd

Utilized port

None.

Remarks

None.

S53clctrl

Function

Waiting for completion of startup of Cluster Resource Management facility.

Effect if stopped

The cluster cannot be started.

Startup daemon

/usr/sbin/sdxclc
/usr/sbin/sdxcle
/usr/sbin/sdxcld

Utilized port

None.

Remarks

None.

S53sfdsk2

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Startup daemon

None.

Utilized port

None.

Remarks

None.

S57sfcfsrm

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.

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
9300 (*1)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node
9200-9263 (*2)	TCP	s, r	Interconnect Administrative LAN	Cluster node	ANY	Local and remote cluster nodes

Remarks

(*1) No. 9300 is set to support the service name "sfcfsrm."
(*2) From No. 9200 to No. 9263 are set to support the service names from sfcfs-1 to sfcfs-64.

S76clprmd

Function

Startup of process monitoring facility.

Effect if stopped

Applications using the process monitoring functions will not work.

Startup daemon

/etc/opt/FJSVcluster/FJSVclapm/daemons/prmd

Utilized port

None.

Remarks

Exclusive for PRIMECLUSTER products.

S99SMAWRrms

Function

Startup of RMS.

Effect if stopped

Even if HV_RCSTART=1 is set, RMS will not start automatically at node startup.

Startup daemon

/opt/SMAW/SMAWRrms/bin/bm
/opt/SMAW/SMAWRrms/bin/hvdet_xxxx
(Detectors and applications used in cluster applications will start.)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9786 (*1)	TCP	s, r	Interconnect	Cluster node	9786	Remote cluster node
8000	UDP	s, r	Interconnect	Cluster node	8000	Remote cluster node

Remarks

(*1) No. 9786 is set to support the service name "rmshb."
If the port number overlaps with another application, change the number used in the application to resolve the conflict.

S99fjsvwvbs

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.

Startup daemon

[For nodes working as primary or secondary management servers]

/opt/SMAW/SMAWcj2re/jre/bin/java
wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0 to 2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0 to 2 processes)

[For nodes other than those described above]

wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0 to 2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0 to 2 processes)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9799 (*1)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9798 (*2)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9797 (*3)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote nodes
9796 (*4)	UDP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote nodes

Remarks

(*1) No. 9799 is set to support the service name "fjwv_c."
(*2) No. 9798 is set to support the service name "fjwv_s."
(*3) No. 9797 is set to support the service name "fjwv_n."
(*4) No. 9796 is set to support the service name "fjwv_g."
(*5) Including concurrent use with cluster nodes.
(*6) PC

S99fjsvwvcnf

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.

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

For wvcnfd of the Web-Based Admin View process, 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.

I.3 Necessary Daemons for PRIMECLUSTER to Operate

Necessary daemons other than PRIMECLUSTER for PRIMECLUSTER to operate are as follows:

- crond
- iscsid (*1)
- libvirtd (*2)
- ntpd
- radvd (*3)
- rsyslog (rsyslogd)
- snmptrapd (*4)
- tgtd (*1)

(*1) The iscsid daemon and the tgtd daemon are necessary when using the mirroring among servers.

(*2) The libvirtd daemon is necessary for the KVM environment.

(*3) The radvd daemon is necessary only when both the mode used for duplexing mode GLS is Fast switching mode and Pv6 communication is used.
(*4) The snmptrapd daemon is necessary only when MMB asynchronous monitoring is used.

Appendix J Systemd Services and Startup Daemons, and Port Numbers in PRIMECLUSTER

This appendix provides explanations on systemd services and daemons that are started by PRIMECLUSTER, and the port numbers being used in RHEL7.

J.1 Explanation Formats

Systemd services, startup daemon, and port numbers are explained with the following formats:

Name of Unit

Name of Unit.

Function

Function of Unit.

Effect if stopped

Effect if unit is stopped.

Dependence with other Units

Requires

Prerequisite Units needed by this Unit. If the Units listed here fail to start, this Unit will not be started.

Wants

Prerequisite Units needed by this Unit. If the Units listed here fail to start, this Unit will be started.

Before

Other Units started after this Unit.

After

Other Units started before this Unit.

Startup daemon

Daemon started by Unit.

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

J.2 systemd Service Lists

WantedBy: multi-user.target

fjsvclapi.service

Function

Beginning of online trace of the Cluster Resource Management facility (2).

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvclrmgr.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvclctrl.service

Function

Waiting for completion of startup of Cluster Resource Management facility.

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvclrmgr.service

Startup daemon

/usr/sbin/sdxclc
/usr/sbin/sdxcle
/usr/sbin/sdxcld

Utilized port

None.

Remarks

None.

fjsvcldbm.service

Function

Startup of Cluster Resource Management facility (1).

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvclapi.service
fjsvclrms.service
smawcf.service

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 ports

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9331 (*1)	TCP	s, r	Interconnect	Cluster node	ANY	Local cluster node

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9379 (*2)	TCP	s, r	Interconnect	Cluster node	ANY	Local and remote cluster nodes
9378 (*3)	TCP	s, r	Interconnect	Cluster node	ANY	Local cluster node
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
9383 (*7)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node
9384 (*8)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node

Remarks

(*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."
(*7) No. 9383 is set to support the service name "dcmcom2."
(*8) No. 9384 is set to support the service name "dcmlck2."

fjsvcldev.service

Function

Startup of MMB asynchronous monitoring.

Effect if stopped

MMB asynchronous monitoring cannot be used.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

poffinhibit.service
y30SVmco.service
FJSVfefpcl.service
smawcf.service
fjsvclonltrc.service

Startup daemon

/etc/opt/FJSVcluster/sys/devmmbd
/etc/opt/FJSVcluster/sys/devmmbmond

/etc/opt/FJSVcluster/sys/devmmbmonitord
/etc/opt/FJSVcluster/sys/devmalogd

Utilized ports

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
7000-7100 (*1)	UDP	s, r	Administrative LAN	Cluster node	7000-7100	Remote cluster node
162 (*1)	UDP	r	Administrative LAN	Local cluster node	ANY	Remote cluster node

Remarks

(*1) These ports are used when SA_mmbp and SA_mmbr are used in the shutdown facility on PRIMEQUEST.

fjsvcldev-clmmbmonctl.service

Function

Operation of MMB asynchronous monitoring.

Effect if stopped

None.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized ports

None.

Remarks

This service operates only when the clmmbmonctl command is started and is always in the "inactive (dead)" state.

fjsvcllkcd.service

Function

Checking the definition file for kdump.

Effect if stopped

None.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

There is no effect if it is stopped because this service operates only at the startup and the daemon does not reside.

fjsvclonltrc.service

Function

Beginning of online trace of the Cluster Resource Management facility (1).

Effect if stopped

Information necessary for the trouble investigation cannot be collected.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvclprmd.service

Function

Startup of process monitoring facility.

Effect if stopped

Applications using the process monitoring functions will not work.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

fjsvclctrl.service

Startup daemon

/etc/opt/FJSVcluster/FJSVclapm/daemons/prmd

Utilized port

None.

Remarks

Exclusive for PRIMECLUSTER products.

fjsvclrmgr.service

Function

Startup of Cluster Resource Management facility (2).

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvcldbm.service

Startup daemon

/etc/opt/FJSVcluster/FJSVcldbm/daemons/clrmd

Utilized port

None.

Remarks

None.

fjsvclrmgr2.service

Function

Startup of Cluster Resource Management facility (3).

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

fjsvclctrl.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvclrms.service

Function

Beginning of online trace of the Cluster Resource Management facility (3).

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvclonltrc.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvclrwz.service

Function

Setting of cluster applications.

Effect if stopped

Cluster applications cannot be configured correctly, or will not work correctly.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvclctrl.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvwvbs.service

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.

Dependence with other Units

Requires

None.

Wants

None.

Before

fjsvwvcnf.service

After

network.target

Startup daemon

[For nodes working as primary or secondary management servers]

/opt/SMAW/SMAWcj2re/jre/bin/java
wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0 to 2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0 to 2 processes)

[For nodes other than those described above]

wvAgent /opt/FJSVwvbs (2 processes)
/etc/opt/FJSVwvfrm/sbin/wvClEventd (0 to 2 processes)
/etc/opt/FJSVwvfrm/sbin/wvFaultEventd (0 to 2 processes)

Utilized port

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9799 (*1)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9798 (*2)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	WebView client (*6)
9797 (*3)	TCP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote nodes
9796 (*4)	UDP	s, r	Administrative LAN	Administrative server (*5)	ANY	Local and remote nodes

Remarks

(*1) No. 9799 is set to support the service name "fjwv_c."
(*2) No. 9798 is set to support the service name "fjwv_s."
(*3) No. 9797 is set to support the service name "fjwv_n."
(*4) No. 9796 is set to support the service name "fjwv_g."
(*5) Including concurrent use with cluster nodes.
(*6) PC

fjsvwvcnf.service

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.

Dependence with other Units

Requires

fjsvwvbs.service

Wants

None.

Before

None.

After

fjsvwvbs.service

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

For wvcnfd of the Web-Based Admin View process, 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.

fjsvgfsfsrm.service

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.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

fjsvclctrl.service
fjsvclrmgr2.service

WantedBy

multi-user.target

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 ports

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9300 (*1)	TCP	s, r	Interconnect	Cluster node	ANY	Remote cluster node
9200-9263 (*2)	TCP	s, r	Interconnect Administrative LAN	Cluster node	ANY	Local and remote cluster nodes

Remarks

(*1) No. 9300 is set to support the service name "sfcfsrm."
(*2) From No. 9200 to No. 9263 are set to support the service names from sfcfs-1 to sfcfs-64.

fjsvgfsfsrm2.service

Function

Stop control for monitoring facility of the GFS shared file system, unmount control for GFS shared file system.

Effect if stopped

The GFS shared file system cannot be stopped normally when the system is stopped.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

fjsvclctrl.service
fjsvclrmgr2.service
fjsvgfsfsrm.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvhanet.service

Function

Starting the daemon, activating the virtual interface, and starting the line monitoring function.

Effect if stopped

LAN cannot be duplicated by using the Redundant Line Control function.

Dependence with other Units

Requires

None.

Wants

None.

Before

network.target

After

network.service

Startup daemon

/opt/FJSVhanet/etc/sbin/hanetctld
/opt/FJSVhanet/etc/sbin/hanetselect (*1)
/opt/FJSVhanet/etc/sbin/hanetpathmd (*1)
/opt/FJSVhanet/etc/sbin/hanetmond (*2)

Utilized port (*3)

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
1807	UDP	s, r	Public LAN	Cluster node	1807	Remote node (GS)

Remarks

(*1) Availability of startup and the number of processes rely on the configuration. Also, this may be suspended according to the monitoring status.

(*2) This daemon is started only when the self-checking function is used.

(*3) The port is used only for the GS linkage mode.

fjsvsdx.service

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Dependence with other Units

Requires

None.

Wants

None.

Before

fjsvclctrl.service
fjsvsdx2.service

After

iscsi.service
iscsi-shutdown.service
target.service (*1)

Startup daemon

/usr/sbin/sdxlogd
/usr/sbin/sdxexd
/usr/sbin/sdxservd

Utilized port

None.

Remarks

(*1) The target.service has a dependency with other units only when the mirroring among servers is used.

fjsvsdx2.service

Function

Basic part of GDS.

Effect if stopped

GDS functions cannot be used.

Dependence with other Units

Requires

None.

Wants

None.

Before

fjsvsdxmon.service

After

fjsvsdx.service
fjsvclctrl.service

Startup daemon

None.

Utilized port

None.

Remarks

None.

fjsvsdxmon.service

Function

Monitoring GDS.

Effect if stopped

GDS cannot be restarted when it ends abnormally.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

fjsvsdx2.service

Startup daemon

/usr/sbin/sdxmond

Utilized port

None.

Remarks

None.

poffinhibit.service

Function

Initializing kdump shutdown agent.

Effect if stopped

Forcible stop by kdump shutdown agent is disabled.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

Enabled only in physical environment.

---

smawcf.service

Function

Loading the CF driver and the CIP driver.

Effect if stopped

The cluster cannot be started.

Dependence with other Units

Requires

None.

Wants

fjsvcldev.service

Before

smawrrms.service

After

network.target

Startup daemon

/opt/SMAW/SMAWcf/bin/cfregd

Utilized port

None.

Remarks

None.

smawrhv-to.service

Function

Initializing RMS.

Effect if stopped

The RMS function cannot be used.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawrrms.service

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

None.

smawrrms.service

Function

Startup of RMS.

Effect if stopped

Operation cannot be monitored or controlled by the cluster. The operation will be stopped if this Unit is stopped during the operation.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

network.target

Startup daemon

/opt/SMAW/SMAWRrms/bin/bm

/opt/SMAW/SMAWRrms/bin/hvdet_xxxx

(Detectors and applications used in cluster applications will start.)

Utilized ports

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9786 (*1)	TCP	s, r	Interconnect	Cluster node	9786	Remote cluster node
8000	UDP	s, r	Interconnect	Cluster node	8000	Remote cluster node

Remarks

(*1) No. 9786 is set to support the service name "rmshb."
If the port number overlaps with another application, change the number used in the application to resolve the conflict.

smawsf.service

Function

Startup of Shutdown Facility.

Effect if stopped

Shutdown Facility cannot be used.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

smawcf.service
fjsvcldev.service

Startup daemon

/opt/SMAW/SMAWsf/bin/rcsd

Utilized ports

Port	Protocol	Send/ Receive	Network	Target	Communication target	
					Port	Target
9382 (*1)	UDP	s, r	Administrative LAN	Cluster node	ANY	Remote cluster node
ANY	UDP	s, r	Administrative LAN	Cluster node	623 (*2)	BMC/iRMC
ANY	UDP	s, r	Administrative LAN	Cluster node	161 (*3)	Management blade

Remarks

These ports are used to prevent split brain.
(*1) No. 9382 is set to support the service name "sfadv."
(*2) This port is used when SA_ipmi is set in the shutdown facility on PRIMERGY.
(*3) This port is used when SA_blade is set in the shutdown facility on a blade server.

smawsf-sdtool-debugoff.service

Function

Operation of the shutdown facility.

Effect if stopped

None.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

This service operates only when the sdtool command is started and is always in the "inactive (dead)"state.

smawsf-sdtool-debugon.service

Function

Operation of the shutdown facility.

Effect if stopped

None.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

None.

Startup daemon

None.

Utilized port

None.

Remarks

This service operates only when the sdtool command is started and is always in the "inactive (dead)"state.

smawsfex.service

Function

Starting the configuration update service for SA.

Effect if stopped

None.

Dependence with other Units

Requires

None.

Wants

None.

Before

smawsf.service

After

smawcf.service

Startup daemon

None.

Utilized port

None.

Remarks

The configuration update service for SA works when the node is started only if it is activated by the sfsacfgupdate command.

smawsfmon.service

Function

Monitoring of shutdown facility.

Effect if stopped

If shutdown facility terminates abnormally, it will not be restarted.

Dependence with other Units

Requires

None.

Wants

None.

Before

None.

After

smawcf.service
smawsf.service

Startup daemon

/opt/SMAW/SMAWsf/bin/rcsd_monitor

Utilized port

None.

Remarks

None.

J.3 Necessary Services for PRIMECLUSTER to Operate

Necessary services other than PRIMECLUSTER for PRIMECLUSTER to operate are as follows:

- crond.service
- ipmi.service (*1)
- iscsi.service (*2)
- libvirtd.service (*3)
- ntpd.service, or chronyd.service
- radvd.service (*4)
- rsyslog.service
- target.service (*2)

(*1) The ipmi.service is necessary when SA_ipmi is set in the shutdown facility on PRIMERGY.

(*2) The iscsi.service and the target.service are necessary when using the mirroring among servers.

(*3) The libvirtd.service is necessary for the KVM environment.

(*4) The radvd.service is necessary only when both the mode used for duplexing GLS is Fast switching mode and IPv6 communication is used.

Appendix K Using Firewall

When using Firewall, perform either of the following procedures because the cluster may not operate normally.

- Allow the communication interface used by PRIMECLUSTER.
- Allow the port number used by PRIMECLUSTER.

This chapter provides examples for configuring Firewall by using firewalld and iptables or ip6tables.

Configure the settings in accordance with the security policy.

See

- For details on firewalld, see the man manual or other related documentation for the firewalld(1) or firewall-cmd(1) command.
- For details on iptables, see the man manual or other related documentation for the iptables(8) command.
- For details on ip6tables, see the man manual or other related documentation for the ip6tables(8) command.

When allowing the communication interface used by PRIMECLUSTER

With PRIMECLUSTER, communication interfaces are used on the administrative LAN and the cluster interconnects. Configure the settings to allow both communication interfaces.

The following is an example to allow sending and receiving the communication interface "cip0" of the cluster interconnect.

- firewalld

  The option of the firewall-cmd command which changes the settings of firewalld differs in the following two situations. One is for when an interface which is not registered in the zone is added to "zone=trusted". The other is for when an interface which is registered in another zone is changed to "zone=trusted".

  Add interface cip0 which is not originally registered in the zone to zone=trusted

  ```
  Format:   firewall-cmd --permanent --zone=trusted --add-interface=<interface>

  Example:  firewall-cmd --permanent --zone=trusted --add-interface=cip0
  ```

  Change zone of interface cip0 which is originally registered in another zone to trusted

  ```
  Format:   firewall-cmd --permanent --zone=trusted --change-interface=<interface>

  Example:  firewall-cmd --permanent --zone=trusted --change-interface=cip0
  ```

- iptables or ip6tables

  ```
  Format:  -A INPUT -i <input-interface> -j ACCEPT
           -A OUTPUT -o <output-interface> -j ACCEPT

  Example: -A INPUT -i cip0 -j ACCEPT
           -A OUTPUT -o cip0 -j ACCEPT
  ```

When allowing the port number used by PRIMECLUSTER

See both "Appendix I Startup Scripts and Startup Daemons, and Port Numbers in PRIMECLUSTER" and "Appendix J Systemd Services and Startup Daemons, and Port Numbers in PRIMECLUSTER," and allow communication of all port numbers used by PRIMECLUSTER.

The following is an example to allow communications of some port numbers used by the cluster resource management facility between the local node and other nodes.

- firewalld

  Allow communication to specific port number

  ```
  Format:    firewall-cmd --permanent --zone=<zone> --add-port=<destination-port-number>/<tcp/udp>
  ```

```
Example:    firewall-cmd --permanent --zone=public --add-port=9383/tcp
```

Allow communication from specific port number

Command option of IPv4 and IPv6 differ from each other.

IPv4

```
Format:     firewall-cmd --permanent --direct --add-rule ipv4 filter INPUT 0 -p <tcp/udp> --sport
<source-port-number> -j ACCEPT

Example:    firewall-cmd --permanent --direct --add-rule ipv4 filter INPUT 0 -p tcp --sport 9383 -
j ACCEPT
```

IPv6

```
Format:     firewall-cmd --permanent --direct --add-rule ipv6 filter INPUT 0 -p <tcp/udp> --sport
<source-port-number> -j ACCEPT

Example:    firewall-cmd --permanent --direct --add-rule ipv6 filter INPUT 0 -p tcp --sport 9383 -
j ACCEPT
```

- iptables or ip6tables

```
Format:  -A <INPUT/OUTPUT> -p <tcp/udp> -m <tcp/udp> --dport <destination-port-number> -j ACCEPT
         -A <INPUT/OUTPUT> -p <tcp/udp> -m <tcp/udp> --sport <source-port-number> -j ACCEPT

Example: -A INPUT -p tcp -m tcp --dport 9383 -j ACCEPT
         -A OUTPUT -p tcp -m tcp --sport 9383 -j ACCEPT
         -A INPUT -p tcp -m tcp --sport 9383 -j ACCEPT
         -A OUTPUT -p tcp -m tcp --dport 9383 -j ACCEPT
```

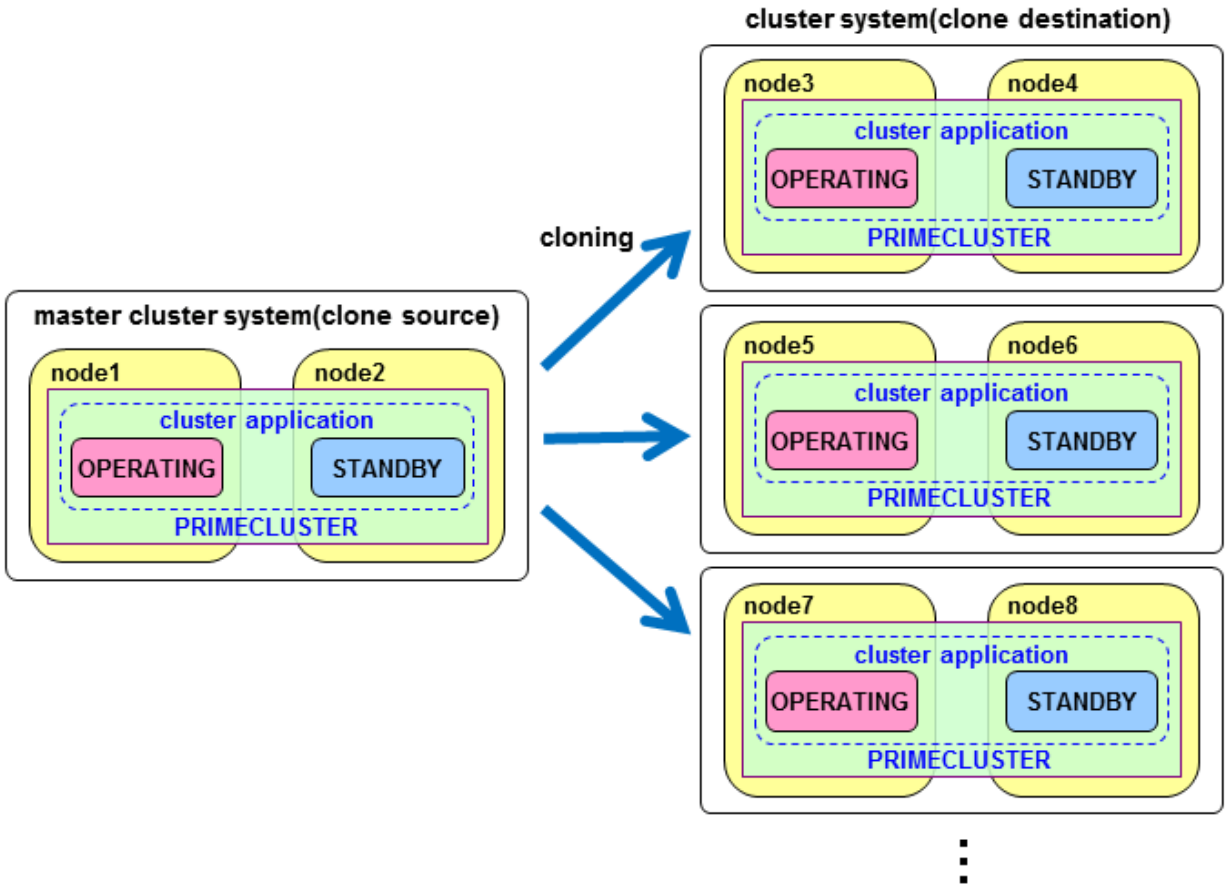
Note

- If you changed the configuration of firewalld by the '--permanent' option of firewall-cmd, restart the firewalld service.
- If you changed the configuration of iptables, perform one of the following operations instead of restarting the iptables service.
  - Restarting the cluster node
  - Reflecting the change by iptables-restore
- If you changed the configuration of ip6tables, perform one of the following operations instead of restarting the ip6tables service.
  - Restarting the cluster node
  - Reflecting the change by ip6tables-restore
- When using the state module in iptables or ip6tables, configure settings to allow communications of PRIMECLUSTER before the state module settings.
  In the following example, communications of cluster interconnects are allowed before the state module settings.

```
Example: -A INPUT -i cip0 -j ACCEPT
         -A OUTPUT -o cip0 -j ACCEPT
         -A INPUT -m state --state RELATED,ESTABLISHED -j ACCEPT
         -A INPUT -p tcp -m tcp ! --tcp-flags SYN,RST,ACK SYN -m state --state NEW -j DROP
```

Appendix L Cloning the Cluster System Environment

PRIMECLUSTER allows you to configure a new cluster system by cloning an already configured cluster system.



Note

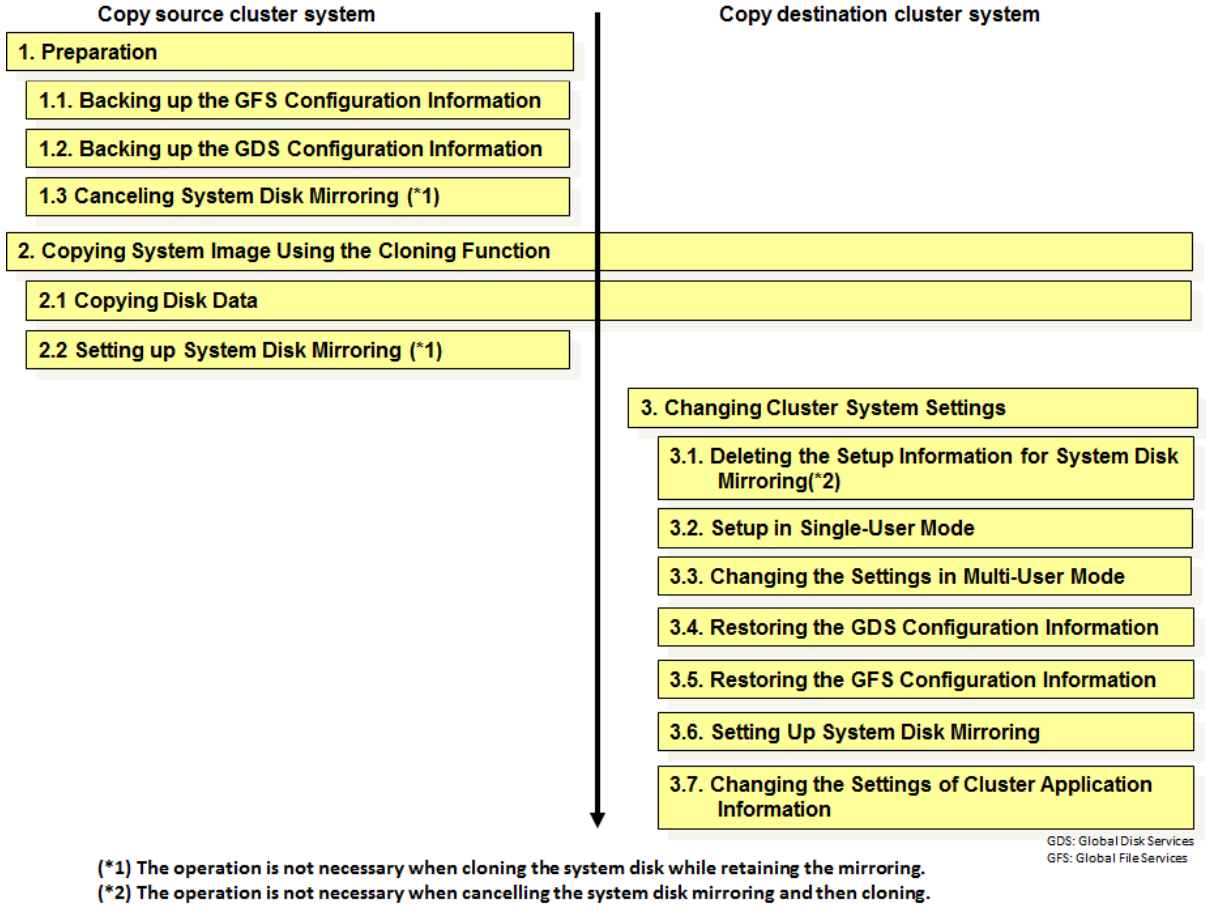
- The following items are not included in the supported range:
  - The cluster system configured in the virtual environment
  - Building a single-node cluster from multiple clusters (cases of which can be seen with Disaster Recovery and so on)
  - After building a single-node cluster, copying it to multiple 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 hardware configurations (server models and disk mounting positions) of the copy source and the copy destination are the same.
- 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 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 L.1 Procedure for Configuration by Cloning



Note

If mirroring of the system disk using GDS is set in the cluster system of the copy source, system disk mirroring must be canceled temporarily either in the source or in the destination system of copying.

This cloning method is particularly recommendable when there are multiple copy destination systems.

- When canceling the system disk mirroring on the copy source
  The procedure for canceling the system disk mirroring temporarily on the copy source and then cloning it is as follows:

  1. As described in "L.1 Preparation," cancel a system disk mirroring on the copy source.
  2. After the procedure described in "L.2 Copying System Image Using the Cloning Function," mirror the system disk again on the copy source system.
  3. As described in "L.3 Changing Cluster System Settings," make the settings for the system disk mirroring on the copy destination system.

- When canceling a system disk mirroring on the copy destination
  The cloning procedure for canceling the system disk mirroring on the copy destination is as follows:

  1. After the procedure described in "L.2 Copying System Image Using the Cloning Function," restart OS using the installation CD of the OS on the copy destination system in "L.3 Changing Cluster System Settings."
  2. Delete the configuration information for system disk mirroring.

3. After booting from the system disk, make the settings for the system disk mirroring.

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 IPMI (BMC or iRMC) or IP address of the management blade	10.20.30.200	10.20.30.201	10.20.30.202	10.20.30.203
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

L.1 Preparation

This part describes the preliminary operation executed before cloning is applied.

L.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(8) 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
...
```

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.

L.1.2 Backing up the GDS Configuration Information

This procedure is unnecessary when mirroring among servers is used.

1. 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
```

2. 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
```

L.1.3 Canceling System Disk Mirroring

For cloning after temporarily canceling system disk mirroring, unmirror the relevant system disks.

See

For procedure for canceling mirroring of system disks, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Note

This procedure is unnecessary if you carry out cloning while system disk mirroring is active.

L.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.

L.2.1 Copying Disk Data

1. Copy the system Disk

   Copy the system disk image to the destination system.

   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.

Note

When using the mirroring among servers, copy the local disk data used by the mirroring among servers in the range of a. or b.

L.2.2 Setting up System Disk Mirroring

If you canceled mirroring of the system disks as described in "L.1.3 Canceling System Disk Mirroring," the source system needs to be reverted to mirroring of the system disks afterwards.

See

For the setting up procedure, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Note

This procedure is unnecessary if you carry out cloning while system disk mirroring is active.

L.3 Changing Cluster System Settings

This section explains the procedure for changing cluster system settings required on the system image copy destination.

L.3.1 Deleting the Setup Information for System Disk Mirroring

If you copied the system disk data while mirroring was active, start up the system in rescue mode from the installation CD of the OS and delete the setup information for system disk mirroring.

See

For the method of deleting it, see "Resolution" of "System cannot be booted. (Failure of all boot disks)" in "System Disk Abnormality [EFI]" of "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

Note

- After the end of the rescue mode in this procedure, when you boot the system, start it up in single-user mode.
- This procedure is unnecessary if you carried out cloning with temporarily canceled system disk mirroring.

L.3.2 Setup in Single-User Mode

Execute the following procedure on all the nodes of the copy destination.

1. Start the system in single-user mode.
2. Change the host name.

   Change the host name in "/etc/hosts", "/etc/sysconfig/network (for RHEL6)", and "/etc/hostname (for RHEL7)".
3. Change the primary management sever, secondary management server, httpip, and mip in the Web-Based Admin View.
   1. Set the IP addresses of the primary management server and the secondary management server.

      ```
      # /etc/opt/FJSVwvbs/etc/bin/wvSetparam primary-server <IP-address-of-primary-management-server>
      # /etc/opt/FJSVwvbs/etc/bin/wvSetparam secondary-server <IP-address-of-secondary-management-server>
      ```
   2. Set httpip.

      ```
      # /etc/opt/FJSVwvbs/etc/bin/wvSetparam httpip <IP-address-used-for-a-client>
      ```
   3. Set mip.

      ```
      # /etc/opt/FJSVwvbs/etc/bin/wvSetparam mip <IP-address-for-identifying-own-host>
      ```
4. Change the CF node name, CIP/SysNode name, and the cluster name.

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. Change the string of the CF node name within the CIP/SysNode name that are described in /etc/hosts.

   [Before change]

   ```
   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 and cluster name described in /etc/default/cluster.

   [Before change]

   ```
   nodename fuji2
   clustername PRIMECLUSTER1
   device eth2
   device eth3
   ```

   [After change]

   ```
   nodename fuji4
   clustername PRIMECLUSTER2
   device eth2
   device eth3
   ```

5. Cancel the SF settings.

   Save "/etc/opt/SMAW/SMAWsf/rcsd.cfg" to the rcsd.org file.

   ```
   # mv /etc/opt/SMAW/SMAWsf/rcsd.cfg /etc/opt/SMAW/SMAWsf/rcsd.org
   ```

6. Change the node name of 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
   ```

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 the nodes.

```
# rm /var/opt/FJSVsfcfs/sfcfsrm.conf
```

8. Change the IP Address of GLS.

   If you are using GLS, perform the following settings. Change the settings on all the 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.

1. Correct the "/etc/hosts" file.

   [Before change]

```
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. Delete all settings for the takeover virtual Interface.

```
# /opt/FJSVhanet/usr/sbin/hanethvrsc delete -n all
```

3. Modify the ifcfg-eth0 file to change the IP address of the primary physical interface.

   For [primecl03]

   [Before change]

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
IPADDR=10.34.214.181
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

   [After change]

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
```

```
IPADDR=10.34.214.191
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

For [primecl04]

[Before change]

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
IPADDR=10.34.214.182
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

[After change]

```
DEVICE=eth0
BOOTPROTO=static
HWADDR=XX:XX:XX:XX:XX:XX
HOTPLUG=no
BROADCAST=10.34.214.255
IPADDR=10.34.214.192
NETMASK=255.255.255.0
NETWORK=10.34.214.0
ONBOOT=yes
TYPE=Ethernet
```

4. 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
   ```

   For [primecl04]

   ```
   # /opt/FJSVhanet/usr/sbin/hanetconfig modify -n sha0 -i 10.34.214.195 -e 10.34.214.192
   ```

5. Reregister the settings of the takeover virtual Interface.

   ```
   # /opt/FJSVhanet/usr/sbin/hanethvrsc create -n sha0
   ```

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."

9. Change the GDS settings.

   - When the mirroring among servers is not used

     Delete the local class and shared class definitions of GDS. Perform the following procedure on all the nodes.

1. Delete the settings of class name in the "/etc/opt/FJSVsdx/sysdb.d/class.db" file.

```
...
Class1          <- Delete all the lines of class 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. Delete all configuration information in the "/etc/sysconfig/devlabel" file.

```
...
/etc/opt/FJSVsdx/.devlabel/Class1/sdx_dev...    <- Delete this line
...
```

4. Delete all the directories named with the class name in "/etc/opt/FJSVsdx/.devlabel" directory.

```
# cd /etc/opt/FJSVsdx/.devlabel
# rm -rf Class1
```

- When the mirroring among servers is used

  Change the settings for iSCSI targets used in GDS. Perform the following procedure on all the nodes.

  [RHEL6]

  1. Delete the registered target portal.

     Example: The IP address used for the mirroring among servers of the copy source node is "192.168.56.20".

```
# iscsiadm -m discovery --op delete --portal 192.168.56.20
```

  2. Start the network service.

```
# service network start
```

  3. Change the IP address entries in /etc/tgt/targets.conf and the path of the by-id link.

     Example: Change the IP address to "192.168.56.21" and "192.168.56.11", and the path of the by-id link to "/dev/disk/by-id/scsi-3500000e111c56610".

     [Before change]

```
# GDS netmirror definition start
<target target-0101>
initiator-address 192.168.56.20
initiator-address 192.168.56.10
write-cache off
 <backing-store /dev/disk/by-id/scsi-3500000e111e68e00>
  scsi_id SDX-01
 </backing-store>
</target>
# GDS netmirror definition end
```

     [After change]

```
# GDS netmirror definition start
<target target-0101>
initiator-address 192.168.56.21
initiator-address 192.168.56.11
write-cache off
 <backing-store /dev/disk/by-id/scsi-3500000e111c56610>
  scsi_id SDX-01
 </backing-store>
```

```
</target>
# GDS netmirror definition end
```

4. Start the tgtd service.

```
# service tgtd start
```

5. On both nodes, change the IP addresses that are described in /etc/opt/FJSVsdx/.sdxnetmirror_ipaddr, which is the configuration file of the mirroring among servers.

   [Before change]

```
192.168.56.10
192.168.56.20
```

   [After change]

```
192.168.56.11
192.168.56.21
```

6. Establish the iSCSI session.

   For the procedure, see "Establishing iSCSI Session" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

[RHEL7]

1. Delete the registered target portal.

   Example: The IP address used for the mirroring among servers of the copy source node is "192.168.56.20".

```
# iscsiadm -m discovery --op delete --portal 192.168.56.20
```

2. Start the network service.

```
# systemctl start network.service
```

   The following message may be output, however, it does not affect the operation of GDS. No corrective action is required.

```
Error getting authority: Error initializing authority: Could not connect:
No such file or directory (g-io-error-quark, 1)
```

3. Copy the configuration information file of the iSCSI target.

```
# cp /etc/target/saveconfig.json copy_destination_file_name
```

4. Change the IP address and the path of the by-id link described in the copy destination file explained in step 3 above.

   Example: Change the IP address to "192.168.56.11", and the path of the by-id link to "/dev/disk/by-id/scsi-3500000e111c56610".

   [Before change]

```
{
  "fabric_modules": [],
  "storage_objects": [
    {
...
      "dev": "/dev/disk/by-id/scsi-3500000e111e68e00",
      "name": "store1",
      "plugin": " block ",
      "readonly": false,
      "write_back": false,
      "wwn": "4a98bfb0-7d7e-4bc8-962c-0b3cf192b214"
    }
...
          "portals": [
```

```
            {
              "ip_address": "192.168.56.20",
              "iser": false,
              "port": 3260
            }
          ],
...
```

[After change]

```
{
  "fabric_modules": [],
  "storage_objects": [
    {
...
      "dev": "/dev/disk/by-id/scsi-3500000e111c56610",
      "name": "store1",
      "plugin": "block",
      "readonly": false,
      "write_back": false,
      "wwn": "4a98bfb0-7d7e-4bc8-962c-0b3cf192b214"
    }
...
          "portals": [
            {
              "ip_address": "192.168.56.21",
              "iser": false,
              "port": 3260
            }
          ],
...
```

5. Apply the changes in the configuration information file of the iSCSI target modified in step 4 above to the target.

```
# targetctl restore file_name
```

The following message may be output, however, it does not affect the operation of GDS. No corrective action is required.

```
Unable to load target_core_user
```

6. Make sure that the iSCSI target is set correctly.

```
# targetcli ls
```

[Output example]

```
o- / ..................................................................... [...]
  o- backstores .......................................................... [...]
  | o- block .............................................. [Storage Objects: 1]
  | | o- store1  [/dev/disk/by-id/scsi-3500000e111c56610 (16.0GiB) write-thru activated]
                 (1)
  | o- fileio ............................................. [Storage Objects: 0]

  | o- pscsi .............................................. [Storage Objects: 0]
  | o- ramdisk ............................................ [Storage Objects: 0]
  o- iscsi ........................................................ [Targets: 1]
  | o- iqn.2003-01.org.linux-iscsi.node1.x8664:sn.6e665c7c1be0 ............... [TPGs: 1]
  |   o- tpg1 ................................................ [no-gen-acls, no-auth]
  |     o- acls ...................................................... [ACLs: 2]
  |     | o- iqn.1994-05.com.redhat:dee92ff9979d ..................... [Mapped LUNs: 1]
  |     | | o- mapped_lun0 ................................... [lun0 block/store1 (rw)]
  |     | o- iqn.1994-05.com.redhat:fa7eb9cf483c ..................... [Mapped LUNs: 1]
  |     |   o- mapped_lun0 ................................. [lun0 block/store1 (rw)]
  |     o- luns ...................................................... [LUNs: 1]
```

```
|     | o- lun0 ............. [block/store1 (/dev/disk/by-id/scsi-3500000e111c56610)]
                                                (2)
|     o- portals ..................................................... [Portals: 1]
|       o- 192.168.56.21:3260.............................................. [OK]
            (3)
o- loopback ..................................................... [Targets: 0]
```

Point

Make sure to confirm the command output about the following item.

- Applying the changed path (Example of output(1),(2))

- Applying the changed IP address (Example of output(3))

7. Save the target information restored in step 5.

```
# targetctl save
```

8. On both nodes, change the IP addresses that are described in /etc/opt/FJSVsdx/.sdxnetmirror_ipaddr, which is the configuration file of the mirroring among servers.

   [Before change]

```
192.168.56.10
192.168.56.20
```

   [After change]

```
192.168.56.11
192.168.56.21
```

9. Establish the iSCSI session.

   For the procedure, see "Establishing iSCSI Session" in "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

10. 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 11.

    - 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
```

11. After completing above procedure on all the nodes of the copy destination, start up all the nodes in multi-user mode.

L.3.3 Changing the Settings in Multi-User Mode

This work is to be performed after completing the settings on all the nodes in the cluster system of the copy destination in single-user mode.

1. Start all the nodes in multi-user mode.

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. Checking the CF setting item

   Check if the changed CF node name, CIP/SysNode name, and cluster name are correct.

   a. Checking the CF node name and cluster name

      Execute the cfconfig -g command on each node to check if the set CF node name and cluster name are correct.

      Example: When the CF node name used in the copy destination is fuji4, and the cluster name used in the copy destination is PRIMECLUSTER2

      ```
      # cfconfig -g
      fuji4 PRIMECLUSTER2 eth1 eth2
      ```

   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 the 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, and cluster name that are set in /etc/cip.cf, /etc/default/cluster or /etc/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, and the cluster name." of "L.3.2 Setup in Single-User Mode" again, and then restart the node.
   3. Perform "L.3.3 Changing the Settings in Multi-User Mode " again.

4. Changing 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
   ```

5. Changing the SF settings

   1. For the Blade server, change the CF node name, slot number of the server blade, the SNMP community name, and the IP address of the management blade in the /etc/opt/SMAW/SMAWsf/SA_blade.cfg file.

      Example: When changing the values as follows.

      The SNMP community name

      public -> private

      ```
      CF node name      slot number   IP address of management blade
      fuji2 -> fuji4    1 -> 3        10.20.30.200 -> 10.20.30.202
      fuji3 -> fuji5    1 -> 3        10.20.30.201 -> 10.20.30.203
      ```

      [Before change]

      ```
      community-string public
      management-blade-ip 10.20.30.200
      fuji2 1 cycle
      ```

```
management-blade-ip 10.20.30.201
fuji3 1 cycle
```

[After Change]

```
community-string private
management-blade-ip 10.20.30.202
fuji4 3 cycle
management-blade-ip 10.20.30.203
fuji5 3 cycle
```

2. For PRIMERGY, except for the Blade server, change the entries for the CF node names and the IP address for IPMI (BMC or iRMC) in "/etc/opt/SMAW/SMAWsf/SA_ipmi.cfg".

   Example: When changing the values as follows.

```
CF node name      IP address for IPMI (BMC or iRMC)
fuji2 -> fuji4    10.20.30.200 -> 10.20.30.202
fuji3 -> fuji5    10.20.30.201 -> 10.20.30.203
```

   [Before change]

```
fuji2 10.20.30.200:root:D0860AB04E1B8FA3 cycle
fuji3 10.20.30.201:root:D0860AB04E1B8FA3 cycle
```

   [After Change]

```
fuji4 10.20.30.202:root:D0860AB04E1B8FA3 cycle
fuji5 10.20.30.203:root:D0860AB04E1B8FA3 cycle
```

3. For PRIMEQUEST, execute the following procedure:

   a. Change the setting of PSA/Svmco and MMB. For details on the setting methods, see the following manuals:

      PRIMEQUEST 1000 Series

      - "PRIMEQUEST 1000 Series Installation Manual"

      - "PRIMEQUEST 1000 Series ServerView Mission Critical Option User Manual"

      PRIMEQUEST 2000 Series

      - "PRIMEQUEST 2000 Series Installation Manual"

      - "PRIMEQUEST 2000 Series ServerView Mission Critical Option User Manual"

      You need to create an RMCP user so that PRIMECLUSTER can link with the MMB units. In all PRIMEQUEST instances that make up the PRIMECLUSTER system, be sure to create a user who uses RMCP to control the MMB. To create a user who uses RMCP to control the MMB, log in to the MMB Web-UI and create the user from the "Remote Server Management" window of the "Network Configuration" menu. Create the user as shown below.

      - Set [Privilege] to "Admin".

      - Set [Status] to "Enabled".

      For details about creating a user who uses RMCP to control the MMB, see the following manuals:

      PRIMEQUEST 1000 Series

      - "PRIMEQUEST 1000 Series Tool Reference"

      PRIMEQUEST 2000 Series

      - "PRIMEQUEST 2000 Series Tool Reference"

   b. Delete the MMB information of the copy source CF node.

      Example: Delete the MMB information of fuji2, fuji3 on the copy source.

```
# /etc/opt/FJSVcluster/bin/clmmbsetup -d fuji2
# /etc/opt/FJSVcluster/bin/clmmbsetup -d fuji3
```

c. Execute the "clmmbsetup -a" command and register the MMB information of the copy destination nodes.
For information on how to use the "clmmbsetup" command, see the "clmmbsetup" manual page.

```
# /etc/opt/FJSVcluster/bin/clmmbsetup -a mmb-user
Enter User's Password:
Re-enter User's Password:
```

For *mmb-user* and *User's Password*, enter the user and password created in Step a.

d. Check that the MMB asynchronous monitoring daemon has started on all the nodes.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl
```

If "The devmmbd daemon exists." is displayed, the MMB asynchronous monitoring daemon has started.

If "The devmmbd daemon does not exist." is displayed, the MMB asynchronous monitoring daemon has not started. Execute the following command to start the MMB asynchronous monitoring daemon.

```
# /etc/opt/FJSVcluster/bin/clmmbmonctl start
```

4. Restore the saved rcsd.org file to the rcsd.cfg file.

```
# mv /etc/opt/SMAW/SMAWsf/rcsd.org /etc/opt/SMAW/SMAWsf/rcsd.cfg
```

5. Change the CF node names and the IP address of the administrative LAN (admIP) described in /etc/opt/SMAW/SMAWsf/rcsd.cfg.

Example: When changing the values as follows

```
CF node name    IP address of administrative LAN
fuji2 -> fuji4    10.20.30.100 -> 10.20.30.02
fuji3 -> fuji5    10.20.30.101 -> 10.20.30.103
```

[Before change]

```
fuji2,weight=1,admIP=10.20.30.100:agent=SA_lkcd,timeout=25:SA_ipmi,timeout=25
fuji3,weight=1,admIP=10.20.30.101:agent=SA_lkcd,timeout=25:SA_ipmi,timeout=25
```

[After change]

```
fuji4,weight=1,admIP=10.20.30.102:agent=SA_lkcd,timeout=25:SA_ipmi,timeout=25
fuji5,weight=1,admIP=10.20.30.103:agent=SA_lkcd,timeout=25:SA_ipmi,timeout=25
```

6. When kdump is used to collect the crash dump in the PRIMERGY including the Blade server, set up the kdump shutdown agent. Execute the following command on any one of the nodes.

```
# /etc/opt/FJSVcllkcd/bin/panicinfo_setup
panicinfo_setup: WARNING: /etc/panicinfo.conf file already exists.
(I)nitialize, (C)opy or (Q)uit (I/C/Q) ? <- Input I
```

7. Start up the Shutdown Facility.

```
# sdtool -b
```

Use sdtool -s to confirm whether the shutdown daemon (rcsd) is active.

```
# sdtool -s
```

By executing sdtool -s on all the nodes, the composition of the shutdown facility can be confirmed.

Confirm the shutdown facility operates normally by the display result of the sdtool -s command.

There is a possibility that the configuration settings of the agent or hardware are not correct when any of the following statuses are displayed though setting the shutdown facility is completed.

- InitFailed is displayed in Init State.
- Unknown or TestFailed is displayed in Test State.

Confirm whether the error message is output to the /var/log/messages file. Then, take corrective actions according to the content of the output message.

L.3.4 Restoring the GDS Configuration Information

Restore the GDS configuration information to the copy destination cluster system.

Note

When using the mirroring among servers, this procedure is unnecessary.

1. Deletion of shared disk resources

   If shared classes are used in the copy source, delete the class and disk resources.

   Perform this setting on any node configuring a cluster system.

   1. Delete all 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
      ...
           SHD_DISK 35 SHD_Disk35 UNKNOWN
             DISK 37 sdag UNKNOWN fuji4
             DISK 153 sdw UNKNOWN fuji5
      ...
      ```

   3. Delete all the disk resources that were checked in Step 2.

      Example: The resource IDs of the registered disk resources are "35", "37", and "153".

      ```
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 35
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 37
      # /etc/opt/FJSVcluster/bin/cldelrsc -r 153
      ```

2. Deletion of the GDS management information

   On all the nodes configuring a cluster, delete all lines that are described in the /etc/opt/FJSVsdx/sdx.udev file.

   ```
   # cat /dev/null > /etc/opt/FJSVsdx/sdx.udev
   ```

3. Restart OS on all the nodes.

4. Re-registration of shared disk resources

   If shared classes are used in the copy source, re-register the shared disks in the resource database.

   For details, see "2. Registering a shared disk" in the "5.1.3.2 Registering Hardware Devices."

5. 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.

   Example: Changing the physical disk described in the "/var/tmp/Class1.conf" configuration file from "sdb" to "sdc"

```
# sdxconfig Convert -e replace -c Class1 -p sdb=sdc -i /var/tmp/Class1.conf -o /var/tmp/
Class1.conf -e update
```

6. 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 the nodes.

   For details on the Excluded List, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

7. Restoring the object configuration information of local and shared classes

   - 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 "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

   ERROR: *device*: disk label is not matched with class *class*

8. Restart OS on all the nodes.

9. 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 7.

   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
```

10. Preliminary setup for Gds resources used in RMS

    Perform the following operation on any node.

```
# /opt/SMAW/SMAWRrms/bin/hvgdsetup -a Class1
```

11. Start the GDS volume.

    If the GDS volume stopped in Step 9. includes any GDS shared class volumes which are not registered to RMS (e.g. classes used by GFS), restart the volume manually, since it will not start automatically at the time of starting the RMS.

    Example: Starting classes used by GFS (gsf and gfs01)

```
# sdxvolume -N -c gfs
# sdxvolume -N -c gfs01
```

L.3.5 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/dsk/control file as the management partition.

   ```
   # sfcsetup -cf /dev/sfdsk/gfs/dsk/control
   ```

2. Reregister the information of the configuration node on each node.

   ```
   # sfcsetup -a /dev/sfdsk/gfs/dsk/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 "L.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(8)" command with the -p option.

   ```
   # sfcsetup -p
   /dev/sfdsk/gfs/dsk/control
   ```

   The registered node information can be confirmed by executing the "sfcsetup(8)" command without any option.

   ```
   # sfcsetup
   HOSTID          CIPNAME         MP_PATH
   80000000        fuji4RMS        yes
   80000001        fuji5RMS        yes
   ```

   The startup method of the sfcfrmd daemon can be confirmed by executing the"sfcsetup(8)" command with the -m option.

   ```
   # sfcsetup -m
   wait_bg
   ```

5. Start the sfcfrmd daemon by executing the following command on all the nodes.

   ```
   # sfcfrmstart
   ```

6. If you are not going to copy the data on the shared disk, create a GFS Shared File System.

See

For details on how to create a GFS Shared File System, see "Creating a file system" or both "Creating a file system" and "Selecting a communication protocol" in "PRIMECLUSTER Global File Services Configuration and Administration Guide."

7. If you are going to copy the data on the shared disk, restore the information of the management partition.

   Execute the shell script you edited in "L.1.1 Backing up the GFS Configuration Information" of the nodes on the copy destination servers.

```
# sh _backup_file_
get other node information start ... end
```

   Confirm that restoration of the management partition of GFS was successful by running the "sfcinfo(8)" command and the "sfcrscinfo(8)" command.

```
# sfcinfo -a
/dev/sfdsk/gfs01/dsk/volume01:
FSID special                                      size  Type  mount
   1 /dev/sfdsk/gfs01/dsk/volume01(1e721)        14422  META  -----
   1 /dev/sfdsk/gfs01/dsk/volume01(1e721)         5116  LOG   -----
   1 /dev/sfdsk/gfs01/dsk/volume01(1e721)        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
```

8. Mount the GFS Shared File System on all the nodes.

```
# sfcmntgl <mount point>
```

L.3.6 Setting Up System Disk Mirroring

To enable system disk mirroring, it is required to set up system disk mirroring on all the target nodes.

See

For details on the setting procedure, see "PRIMECLUSTER Global Disk Services Configuration and Administration Guide."

L.3.7 Changing the Settings of Cluster Application Information

Change the setting of the cluster application information. The procedures vary depending on whether GLS is used, the takeover network is used, or neither of them is used.

L.3.7.1 When Using GLS

1. Changing the cluster application information

   1. In order to change these settings with the RMS Wizard, execute the "hvw" command on any node.

```
# hvw -n config
```

   2. Select "Application-Create" from "Main configuration menu".

```
fuji4: 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
```

3. Select "APP1" from "Application selection menu".

```
Edit: Application selection menu (restricted):
1) HELP
2) QUIT
3) RETURN
4) OPTIONS
5) APP1
Application Name: 5
```

4. If you changed any IP addresses for GLS according to step 8 of "L.3.2 Setup in Single-User Mode," change the settings for the takeover IP address for Gls resources.

    1. Select "Gls:Global-Link-Services".

    ```
    Settings of turnkey wizard "STANDBY" (APP1:consistent)
     1) HELP                                9) Procedure:BasicApplication(-)
     2) READONLY                           10) Symfoware(-)
     3) SAVE+EXIT                          11) Procedure:SystemState3(-)
     4) -                                  12) Procedure:SystemState2(-)
     5) ApplicationName=APP1               13) Gls:Global-Link-Services(Gls_APP1)
     6) Machines+Basics(app1)              14) IpAddresses(-)
     7) CommandLines(-)                    15) LocalFileSystems(-)
     8) Procedure:Application(-)           16) Gds:Global-Disk-Services(-)
    Choose the setting to process: 13
    ```

    2. Select "TakeoverIpaddress" to change the takeover IP address.

        At the right side of the menu, the original IP address is displayed.

    ```
    Gls (Gls_APP1:consistent)
    1) HELP                                5) AdditionalTakeoverIpaddress
    2) NO-SAVE+EXIT                        6) TakeoverIpaddress[0]=N,10.34.214.185
    3) SAVE+EXIT                           7) (Timeout=60)
    4) REMOVE+EXIT
    Choose the setting to process: 6
    ```

    3. The IP address used after modification is displayed in the menu.

        Select the modified takeover IP address.

    ```
    1) HELP                                4) FREECHOICE
    2) RETURN                              5) SELECTED(10.34.214.185)
    3) NONE                                6) 10.34.214.195
    Choose a takeover IP address for Gls: 6
    ```

    4. Confirm that the selected IP address has been set and then select "SAVE+RETURN".

    ```
    Set a flag for takeover IP address: 10.34.214.195
    Currently set:
    1) HELP                                5) AUTORECOVER(A)
    2) -
    3) SAVE+RETURN
    ```

```
4) DEFAULT
Choose additonally one of the flags: 3
```

5. Select "SAVE+EXIT" to save the settings of Gls resources and exit the menu.

```
Gls (Gls_APP1:consistent)
1) HELP                                 5) AdditionalTakeoverIpaddress
2) NO-SAVE+EXIT                         6) TakeoverIpaddress[0]=N,10.34.214.195
3) SAVE+EXIT                            7) (Timeout=60)
4) REMOVE+EXIT
Choose the setting to process: 3
```

5. Select "SAVE+EXIT" to return to the "Application selection menu." After that, select "RETURN" to return to the "Main configuration menu."
6. Change the SysNode that configures a cluster application.
    1. Select "RMS-CreateMachine".
    2. After selecting "ALL-CF-HOSTS", check that all the SysNode names, created based on the changed CF node name "Current set", are displayed.

       At this point, the SysNode names, created based on the original CF node name, are also displayed simultaneously; however, unnecessary SysNode names are deleted in Step 9.
    3. Select "RETURN".
    4. Select "Application-Edit".
    5. Select "APP1".
    6. Select "Machines+Basics(app1)".
    7. Select "Machines[0]" and set the SysNode names based on the changed CF node name. After that, also select "Machines[1]" simultaneously
    8. Select "SAVE+EXIT" > "SAVE+EXIT" > "RETURN" to return to the menu immediately after the "hvw" command was started.
    9. After selecting "RMS-RemoveMachine", select unnecessary SysNode names in sequence to let the SysNode names, created based on the changed CF node name "Current set", only be displayed, and then select "RETURN".
7. Execute the "Configuration-Generate" and "Configuration-Activate" in sequence to check that each operation ended properly.
8. Select "QUIT" to exit the "hvw" command

2. Execute the following commands on all the nodes as required to set the automatic startup of RMS.

```
# hvsetenv HV_RCSTART 1
# hvsetenv HV_RCSTART
1    <- Check "1" is output.
```

3. Start RMS. Execute the following command on any one node.

```
# hvcm -a
```

L.3.7.2 When Using the Takeover Network

1. Changing the IP addresses and host names of public LANs with IP address takeover and node name takeover.

   Change the /usr/opt/reliant/etc/hvipalias file on all the nodes.

For details on the setting contents in the "hvipalias" file, see "6.7.3.6 Setting Up Takeover Network Resources."

Change the entries below as required.

```
<node name> : Change the value in this filed to the changed CF node name.

<takeover> :  If you changed any host names associated with takeover IP addresses, change this
host name.
```

2. Changing the cluster application information

    1. In order to change these settings with the RMS Wizard, execute the "hvw" command on any node.

    ```
    # hvw -n config
    ```

    2. Select "Application-Create" from "Main configuration menu".

    ```
    fuji4: 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
    ```

    3. Select "APP1" from "Application selection menu".

    ```
    Edit: Application selection menu (restricted):
    1) HELP
    2) QUIT
    3) RETURN
    4) OPTIONS
    5) APP1
    Application Name: 5
    ```

    4. Change the settings for the host names in the takeover network resources.

        1. Select "IpAddresses" from "turnkey wizard".

        ```
        Consistency check ...
        Settings of turnkey wizard "STANDBY" (APP1:consistent)
         1) HELP                               9) Procedure:BasicApplication(-)
         2) READONLY                          10) Symfoware(-)
         3) SAVE+EXIT                         11) Procedure:SystemState3(-)
         4) -                                 12) Procedure:SystemState2(-)
         5) ApplicationName=APP1              13) Gls:Global-Link-Services(-)
         6) Machines+Basics(app1)             14) IpAddresses(Adr_APP1)
         7) CommandLines(-)                   15) LocalFileSystems(-)
         8) Procedure:Application(-)          16) Gds:Global-Disk-Services(-)
        Choose the setting to process: 14
        ```

        2. When the "Ipaddresses and ipaliase" menu is displayed, select the "Interfaces[X]" in which the host name to be changed is set.

        ```
        Consistency check ...
        Yet to do: specify ping hosts of an interface using its P flag

        Settings of application type "Ipaddress" (consistent)
         1) HELP                               9) Interfaces[1]=V:takeover
         2) NO-SAVE+EXIT                      10) PingHostPool[0]=pinghost1
         3) SAVE+EXIT                         11) PingHostPool[1]=pinghost2
         4) ApplicationName=Adr_APP1          12) SubApplications[0]=Lfs_APP1
        ```

```
 5) AdditionalInterface                13) (NeedAll=yes)
 6) AdditionalPingHost                 14) (Timeout=60)
 7) AdditionalSubApplication           15) (InterfaceFilter=)
 8) Interfaces[0]=V:tussd2af
Choose the setting to process:
```

3. From the displayed menu, select the changed name of the host associated with the takeover IP address..

   (All host names in the "/etc/hosts" file are displayed in the menu.)

4. Select "SAVE + RETURN".

5. Check the setting value of "Interfaces [X]" to make sure that the host name at the modified position is correct.

   If there are multiple objects to be changed, repeat Steps 2 to 4 for each object.

   When all changes have been completed, select "SAVE + RETURN".

6. Select "SAVE + EXIT".

7. Select "RETURN".

5. Change the SysNode that configures a cluster application.

   1. Select "RMS-CreateMachine".

   2. After selecting "ALL-CF-HOSTS", check that all the SysNode names, created based on the changed CF node name "Current set", are displayed.

      At this point, the SysNode names, created based on the original CF node name, are also displayed simultaneously; however, unnecessary SysNode names are deleted in Step 9.

   3. Select "RETURN".

   4. Select "Application-Edit".

   5. Select "APP1".

   6. Select "Machines+Basics(app1)".

   7. Select "Machines[0]" and set the SysNode names based on the changed CF node name. After that, also select "Machines[1]" simultaneously.

   8. Select "SAVE+EXIT" > "SAVE+EXIT" > "RETURN" to return to the menu immediately after the "hvw" command was started.

   9. After selecting "RMS-RemoveMachine", select unnecessary SysNode names in sequence to let the SysNode names, created based on the changed CF node name "Current set", only be displayed, and then select "RETURN".

6. Execute the "Configuration-Generate" and "Configuration-Activate" in sequence to check that each operation ended properly.

7. Select "QUIT" to exit the "hvw" command.

3. Execute the following commands on all the nodes as required to set the automatic startup of RMS.

```
# hvsetenv HV_RCSTART 1
# hvsetenv HV_RCSTART
1    <- Check "1" is output.
```

4. Start RMS. Execute the following command on any one node.

```
# hvcm -a
```

L.3.7.3 When Using neither GLS nor the Takeover Network

1. Change the cluster application information.

   1. In order to change these settings with the RMS Wizard, execute hvw command on any one node.

```
# hvw -n config
```

2. Change the SysNode that configures a cluster application.

    1. Select "RMS-CreateMachine".
    2. After selecting "ALL-CF-HOSTS", make sure that all the SysNode names, created based on the changed CF node name "Current set", are displayed.

       At this point, the SysNode names, created based on the original CF node name, are also displayed simultaneously; however, unnecessary SysNode names are deleted in Step 9.
    3. Select "RETURN".
    4. Select "Application-Edit".
    5. Select "APP1".
    6. Select "Machines+Basics(app1)".
    7. Select "Machines[0]" and set the SysNode names based on the changed CF node name. After that, also select "Machines[1]" simultaneously.
    8. Select "SAVE+EXIT" > "SAVE+EXIT" > "RETURN" to return to the menu immediately after hvw command was started.
    9. After selecting "RMS-RemoveMachine", select unnecessary SysNode names in sequence to let the SysNode names, created based on the changed CF node name "Current set", only be displayed, and then select "RETURN".

3. Execute "Configuration-Generate" and "Configuration-Activate" in sequence and make sure that each operation ended properly.
4. Select "QUIT" to exit hvw command.

2. Execute the following commands on all the nodes as required to set the automatic startup of RMS.

```
# hvsetenv HV_RCSTART 1
# hvsetenv HV_RCSTART
1    <- Make sure that "1" is output.
```

3. Start RMS. Execute the following command on any one node.

```
# hvcm -a
```

Appendix M Changes in Each Version

This chapter explains the changes made to the specifications of PRIMECLUSTER 4.4A00.

The changes are listed in the following table.

Table M.1 List of changes

Category	Item	Version
Incompatible commands	clgettree(1) command	(Before change) PRIMECLUSTER 4.1A20 or earlier (After change) PRIMECLUSTER 4.4A00
	hvdisp command	(Before change) PRIMECLUSTER 4.1A20 or earlier (After change) PRIMECLUSTER 4.4A00
	ciptool command	(Before change) PRIMECLUSTER 4.1A30 or earlier (After change) PRIMECLUSTER 4.4A00
	sdtool command	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.4A00
	hvshut command	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	hvswitch command	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	hvdump command	(Before change) PRIMECLUSTER 4.3A30 or earlier (After change) PRIMECLUSTER 4.4A00
Incompatible functions	Posting Notification of a Resource Failure or Recovery	(Before change) PRIMECLUSTER 4.0A20 or, PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Operator Intervention Request	(Before change) PRIMECLUSTER 4.3A30 or earlier (After change) PRIMECLUSTER 4.4A00
	Shutdown Facility	(Before change) PRIMECLUSTER 4.0A20 (After change) PRIMECLUSTER 4.4A00
	Setting Up Cluster Applications	(Before change) PRIMECLUSTER 4.0A20 (After change) PRIMECLUSTER 4.4A00
	Changing the Cluster Application	(Before change) PRIMECLUSTER 4.0A20 (After change) PRIMECLUSTER 4.4A00
	State of the Main CF Table	(Before change) PRIMECLUSTER 4.0A20 (After change) PRIMECLUSTER 4.4A00
	Node state	(Before change) PRIMECLUSTER 4.1A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Operation Procedures and Displayed Items for Cluster Application Setup and Modification	(Before change) PRIMECLUSTER 4.1A30 or earlier (After change) PRIMECLUSTER 4.4A00
	Setting Up Fsystem Resources	(Before change) PRIMECLUSTER 4.3A30 or earlier (After change) PRIMECLUSTER 4.4A00

Category	Item	Version
	Client Environment for Web-Based Admin View	(Before change) PRIMECLUSTER 4.3A00 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of the Behavior of CF Startup	(Before change) PRIMECLUSTER 4.3A00 or earlier (After change) PRIMECLUSTER 4.4A00
	HV_CONNECT_TIMEOUT	(Before change) PRIMECLUSTER 4.3A00 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of the ports used by RMS	(Before change) PRIMECLUSTER 4.3A10 or earlier (After change) PRIMECLUSTER 4.4A00
	Configuring the IPMI shutdown agent	(Before change) PRIMECLUSTER 4.2A00 or later - 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Changing the port number used by the shutdown facility	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Setting up the Host OS failover function when using it in the PRIMEQUEST KVM environment	(Before change) PRIMECLUSTER 4.3A10 or later - 4.3A40 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of the target node to forcibly shut down when a heartbeat failure occurs	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Displaying Fault Traces of Resources	(Before change) PRIMECLUSTER 4.3A30 or earlier (After change) PRIMECLUSTER 4.4A00
	Change of /etc/cip.cf file	(Before change) PRIMECLUSTER 4.3A30 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes in CF over IP setting window of CF Wizard	(Before change) PRIMECLUSTER 4.3A40 or earlier (After change) PRIMECLUSTER 4.4A00
	Setting up the migration function when using it in KVM environment	(Before change) PRIMECLUSTER 4.3A40 (After change) PRIMECLUSTER 4.4A00
Incompatible message	Changes of the RMS message	(Before change) PRIMECLUSTER 4.3A00 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of the importance of the message in the RMS wizard	(Before change) PRIMECLUSTER 4.3A00 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of RMS console message	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00
	Changes of the response message for the operator intervention request	(Before change) PRIMECLUSTER 4.3A20 or earlier (After change) PRIMECLUSTER 4.4A00

M.1 Changes in PRIMECLUSTER 4.4A00 from 4.0A20

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.0A20.

- M.1.1 clgettree(1) command
- M.1.2 hvdisp command
- M.1.3 ciptool command
- M.1.4 sdtool command
- M.1.5 hvshut command
- M.1.6 hvswitch command
- M.1.7 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.0A20.

- M.1.8 Operator Intervention Request
- M.1.9 Shutdown Facility
- M.1.10 Setting Up Cluster Applications
- M.1.11 Changing the Cluster Application
- M.1.12 State of the Main CF Table
- M.1.13 Node state
- M.1.14 Operation Procedures and Displayed Items for Cluster Application Setup and Modification
- M.1.15 Setting Up Fsystem Resources
- M.1.16 Client Environment for Web-Based Admin View
- M.1.17 Changes of the Behavior of CF Startup
- M.1.18 HV_CONNECT_TIMEOUT
- M.1.19 Changes of the ports used by RMS
- M.1.20 Changes of the port number used by the shutdown facility
- M.1.21 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.1.22 Display of the resource fault trace
- M.1.23 Change of /etc/cip.cf file
- M.1.24 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.0A20.

- M.1.25 Changes of the RMS message
- M.1.26 Changes of the importance of the message in the RMS wizard
- M.1.27 Changes of RMS console message
- M.1.28 Changes of the response message for the operator intervention request

M.1.1 clgettree(1) command

Details on incompatibilities

Cluster class resource names, which are output with the "clgettree(1)" command, are changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

A fixed string "cluster" is displayed when the resource management facility is configured.

After upgrading [PRIMECLUSTER 4.4A00]

The cluster class uses the same name as the CF cluster when the resource management facility is configured.

Note

None.

M.1.2 hvdisp command

Details on incompatibilities

StateDetails information is displayed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

StateDetails information is not displayed.

After upgrading [PRIMECLUSTER 4.4A00]

StateDetails information is displayed.

Note

None.

M.1.3 ciptool command

Details on incompatibilities

Symbolic links of the ciptool commands were not made for /usr/bin.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

Symbolic links of the ciptool commands are made for /usr/bin.

After upgrading [PRIMECLUSTER 4.4A00]

Symbolic links of the ciptool commands are not made for /usr/bin.

Note

Specify /opt/SMAW/SMAWcf/bin/ciptool by full passing when you use the ciptool command.

M.1.4 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

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 4.4A00]

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.

M.1.5 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 4.0A20]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.1.6 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 4.0A20]

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 4.4A00]

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.

M.1.7 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.1.8 Operator Intervention Request

Details on incompatibilities

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 4.0A20]

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 4.4A00]

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 "4.2 Operator Intervention Messages" in "PRIMECLUSTER Messages."

M.1.9 Shutdown Facility

Details on incompatibilities

"SF node weight" is added as a new setting item for the shutdown facility of PRIMECLUSTER 4.4A00. For the same survival priority as PRIMECLUSTER 4.0A20, specify 1 in "SF node weight" for all the nodes. For details, see "5.1.2 Setting up the Shutdown Facility."

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

Survival priority is determined by "ShutdownPriority of userApplication."

After upgrading [PRIMECLUSTER 4.4A00]

Survival priority is determined by "ShutdownPriority of userApplication" and "SF node weight."

Note

None.

M.1.10 Setting Up Cluster Applications

Details on incompatibilities

If a cluster application for a standby operation is used with PRIMECLUSTER 4.4A00, set up "ClearFaultRequest|StartUp|SwitchRequest" to the StandbyTransition attribute. With this setting, a standby state transition can be performed in the same way as PRIMECLUSTER 4.0A20 for the following cases:

- When a cluster application in the Faulted state is set to operating state
- When a cluster application is started up
- When a cluster application is switched

For details on the settings of the cluster application, see "6.7 Setting Up Cluster Applications."

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

If a resource for support hot-standby operation is available, a standby state transition will be made.

After upgrading [PRIMECLUSTER 4.4A00]

A standby state transition will be performed only when a resource to support hot-standby operation is available and "ClearFaultRequest|StartUp|SwitchRequest" is set to the StandbyTransition attribute.

Note

None.

M.1.11 Changing the Cluster Application

Details on incompatibilities

If the "hvw" command is executed while RMS is running, the following menus, that could be displayed with PRIMECLUSTER 4.0A20, cannot be displayed.

- Application-Create
- Application-Edit
- Application-Remove
- Configuration-Activate

To change the cluster application, you have to stop RMS before you execute the "hvw" command. This is the same with PRIMECLUSTER 4.0A20.

If you execute the "hvw" command after stopping RMS, the same menus as PRIMECLUSTER 4.0A20 will be displayed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

If you execute the "hvw" command while RMS is running, the following menus will be displayed.

- Application-Create
- Application-Edit
- Application-Remove
- Configuration-Activate

After upgrading [PRIMECLUSTER 4.4A00]

If you execute the "hvw" command while RMS is running, the following menus will not be displayed.

- Application-Create
- Application-Edit
- Application-Remove
- Configuration-Activate

Note

To change the cluster application, you need to stop RMS before you execute the "hvw" command. This is the same with PRIMECLUSTER 4.0A20.

For details on changing the cluster application, see "10.3 Changing a Cluster Application."

M.1.12 State of the Main CF Table

Details on incompatibilities

"INVALID" is added to PRIMECLUSTER 4.4A00 as a new state of the local node in the main CF table.

For details on the main CF table, see "Chapter 4 GUI administration" in "PRIMECLUSTER Cluster Foundation (CF) Configuration and Administration Guide."

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

If the configuration of the node is invalid, "UNCONFIGURED" or "UNKNOWN" will be displayed as the state of the local node on the main CF table.

After upgrading [PRIMECLUSTER 4.4A00]

If the node configuration is invalid, "INVALID" will be displayed as the state of the local node on the main CF table.

Note

None.

M.1.13 Node state

Details on incompatibilities

If all cluster nodes are stopped, and then a cluster node is started, the state of SysNode not joining the cluster is changed in hvdisp(1M) and RMS main window.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

The state of SysNode not joining the cluster is shown as Offline.

After upgrading [PRIMECLUSTER 4.4A00]

The state of SysNode not joining the cluster is shown as Faulted.
See "StateDetails" information displayed by hvdisp (1M) command or RMS main window to identify whether Faulted state is due to not joining the cluster or abnormal shutdown of the node by a panic, for example.

	State	StateDetails
When a node is not joining the cluster	Faulted	Shutdown
Abnormal shutdown by panic. etc	Faulted	Killed

Note

None.

M.1.14 Operation Procedures and Displayed Items for Cluster Application Setup and Modification

Details on incompatibilities

The procedures and displayed items for creating userApplications and for creating and changing procedure resources have been changed.

Changes

1. Creating a userApplication for standby operation

Before upgrading [PRIMECLUSTER 4.0A20]

To create a userApplication for standby operation, select "CRM" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) CRM
 6) GENERIC
 7) SYMFOWARE
Application Type: 5
```

```
Settings of turnkey wizard "CRM"
 1) HELP               4) REMOVE+EXIT              7) Machines+Basics(-)
 2) NO-SAVE+EXIT       5) ApplicationName=APP1
 3) SAVE+EXIT          6) BeingControlled=no
Choose the setting to process: 7
```

After upgrading [PRIMECLUSTER 4.4A00]

"CRM" is not displayed in the "Application type selection menu."

To create a userApplication for standby operation, select "STANDBY" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 8
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP1
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

2. Creating a userApplication for scalable operation

Before upgrading [PRIMECLUSTER 4.0A20]

To create a userApplication for scalable operation, select "Controller" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                   9) Controller           17) Rawdisk
 2) QUIT                  10) Crm                  18) SYMFOWARE
 3) RETURN                11) Fsystem              19) Symfoware
 4) OPTIONS               12) GENERIC              20) SystemState2
 5) Application           13) Gds                  21) SystemState3
 6) BasicApplication      14) Gls                  22) foreign-code
 7) CRM                   15) Ipaddress            23) generic
 8) Cmdline               16) Lvm
Application Type: 9
```

```
Settings of application type "Controller" (not yet consistent)
 1) HELP                             5) AdditionalAppToControl
 2) NO-SAVE+EXIT                     6) (InParallel=)
 3) ApplicationName=Controller       7) (FaultScript=)
 4) ControlPolicy=FOLLOW
Choose the setting to process: 4
```

After upgrading [PRIMECLUSTER 4.4A00]

"Controller" is not displayed in the "Application type selection menu."

To create a userApplication for scalable operation, select "SCALABLE" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 7
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "SCALABLE" (APP2:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP2
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

3. Creating a procedure resource

Before upgrading [PRIMECLUSTER 4.0A20]

To create or change a procedure resource, select "CRM" from the "turnkey wizard "CRM"" menu, and then select the resource class name.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "CRM"
 1) HELP                          10) LocalFileSystems(-)
 2) -                             11) RemoteFileSystems(-)
 3) SAVE+EXIT                     12) IpAddresses(-)
 4) -                             13) Gds:Global-Disk-Services(-)
 5) ApplicationName=APP1          14) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)
 7) CommandLines(-)
 8) Controllers(-)
 9) CRM(-)
Choose the setting to process: 9
```

```
Crm (CRM_APP1:not yet consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) SystemState2(-)
 6) SystemState3(-)
 7) BasicApplication(-)
 8) Application(-)
Choose the setting to process: 7
```

After upgrading [PRIMECLUSTER 4.4A00]

The "turnkey wizard "CRM"" menu is not displayed.

To create or change a procedure resource, select "Procedure:resource-class-name" from the "turnkey wizard STANDBY" menu.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          9) Procedure:BasicApplication(-)
 2) -                            10) Symfoware(-)
 3) SAVE+EXIT                    11) Procedure:SystemState3(-)
 4) -                            12) Procedure:SystemState2(-)
 5) ApplicationName=APP1         13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)        14) IpAddresses(-)
 7) CommandLines(-)              15) LocalFileSystems(-)
 8) Procedure:Application(-)     16) Gds:Global-Disk-Services(-)
Choose the setting to process:9
```

4. Changing the timeout value of a procedure resource

Before upgrading [PRIMECLUSTER 4.0A20]

To change the timeout value, select "Flags[0]=OTtimeout-time" and then select "SCRIPTTIMEOUT."

```
Settings of "Application"
 1) HELP               4) REMOVE+RETURN          7) Flags[0]=OT1800
 2) NO-SAVE+RETURN     5) AdditionalResource     8) Priority[0]=65535
 3) SAVE+RETURN        6) ResourceName[0]=rsc1
Choose the setting to process: 7
```

```
Set Flags[0]:Currently set: STANDBYCAPABLE,SCRIPTTIMEOUT (OT1800)
 1) HELP              5) NULLDETECTOR(D)        9) MONITORONLY(M)
 2) -                 6) LIEOFFLINE(L)         10) NOT:STANDBYCAPABLE(O)
 3) SAVE+RETURN       7) CLUSTEREXCLUSIVE(X)   11) SCRIPTTIMEOUT(T)
 4) DEFAULT           8) AUTORECOVER(A)
Choose one of the flags: 11
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate script timeout:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Flags[0]=OTtimeout-time" is not displayed.

To change the timeout value, select "TIMEOUT."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 4
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate timeout:
```

5. Changing the priority in a resource class of a procedure resource.

Before upgrading [PRIMECLUSTER 4.0A20]

To change the priority in a resource class, select "Priority[0]=*priority*."

```
Settings of "Application"
 1) HELP              4) REMOVE+RETURN         7) Flags[0]=OT1800
 2) NO-SAVE+RETURN    5) AdditionalResource    8) Priority[0]=65535
 3) SAVE+RETURN       6) ResourceName[0]=rsc1
Choose the setting to process: 8
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
Set a priority:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Priority[0]=*priority*" is not displayed.

To change the priority in a resource class, select "PRIORITY."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 5
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 65535
Set an appropriate priority:
```

Note

None.

M.1.15 Setting Up Fsystem Resources

Details on incompatibilities

The file which defines the mount point of the file system used as Fsystem resource has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

The mount point was defined in /etc/fstab.

After upgrading [PRIMECLUSTER 4.4A00]

It is necessary to define the mount point in /etc/fstab.pcl.

For details, see "6.7.3.2 Setting Up Fsystem Resources."

Note

None.

M.1.16 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.1.17 Changes of the 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 4.0A20]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.0A20.

M.1.18 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds).

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.1.19 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.1.20 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.1.21 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 4.0A20]

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 4.4A00]

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.

M.1.22 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 4.0A20]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Fault Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.1.23 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.1.24 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.1.25 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

(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 4.4A00]

(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.

M.1.26 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 4.0A20]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.1.27 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

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 4.4A00]

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.

M.1.28 Changes of the response message for the operator intervention request

M.1.28.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.1.28.2 Message: 1423

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.0A20]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.2 Changes in PRIMECLUSTER 4.4A00 from 4.1A20

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A20.

- M.2.1 clgettree(1) command
- M.2.2 hvdisp command
- M.2.3 ciptool command
- M.2.4 sdtool command
- M.2.5 hvshut command
- M.2.6 hvswitch command
- M.2.7 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A20.

- M.2.8 Posting Notification of a Resource Failure or Recovery
- M.2.9 Operator Intervention Request
- M.2.10 Node state
- M.2.11 Operation Procedures and Displayed Items for Cluster Application Setup and Modification
- M.2.12 Setting Up Fsystem Resources
- M.2.13 Client Environment for Web-Based Admin View
- M.2.14 Changes of the Behavior of CF Startup
- M.2.15 HV_CONNECT_TIMEOUT
- M.2.16 Changes of the ports used by RMS
- M.2.17 Changes of the port number used by the shutdown facility
- M.2.18 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.2.19 Display of the resource fault trace
- M.2.20 Change of /etc/cip.cf file
- M.2.21 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A20.

- M.2.22 Changes of the RMS message
- M.2.23 Changes of the importance of the message in the RMS wizard
- M.2.24 Changes of RMS console message
- M.2.25 Changes of the response message for the operator intervention request

M.2.1 clgettree(1) command

Details on incompatibilities

Cluster class resource names, which are output with the "clgettree(1)" command, are changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

A fixed string "cluster" is displayed when the resource management facility is configured.

After upgrading [PRIMECLUSTER 4.4A00]

The cluster class uses the same name as the CF cluster when the resource management facility is configured.

Note

None.

M.2.2 hvdisp command

Details on incompatibilities

StateDetails information is displayed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

StateDetails information is not displayed.

After upgrading [PRIMECLUSTER 4.4A00]

StateDetails information is displayed.

Note

None.

M.2.3 ciptool command

Details on incompatibilities

Symbolic links of the ciptool commands were not made for /usr/bin.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

Symbolic links of the ciptool commands are made for /usr/bin.

After upgrading [PRIMECLUSTER 4.4A00]

Symbolic links of the ciptool commands are not made for /usr/bin.

Note

Specify /opt/SMAW/SMAWcf/bin/ciptool by full passing when you use the ciptool command.

M.2.4 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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 4.4A00]

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.

M.2.5 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 4.1A20]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.2.6 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 4.1A20]

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 4.4A00]

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.

M.2.7 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.2.8 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 4.4A00. For details, see "5.2 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 4.1A20]

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 4.4A00]

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.

M.2.9 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 4.1A20]

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 4.4A00]

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

For details, see "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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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 4.4A00]

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.

M.2.10 Node state

Details on incompatibilities

When any one of all the stopped nodes is started, the state of nodes that are separated from the cluster is changed when it is displayed by hvdisp (1M) command or output on the RMS main window.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

The nodes that are separated from the cluster show Offline state.

After upgrading [PRIMECLUSTER 4.4A00]

The nodes that are separated from the cluster show Faulted state.
See details on the state of the nodes (the value of the StateDetails attribute) that are displayed by hvdisp (1M) command or output on the RMS main window to identify if Faulted state is due to the nodes being separated from the cluster or due to an abnormal shutdown of the nodes by a panic or other errors.

	State	StateDetails
When a node is not joining the cluster	Faulted	Shutdown
Abnormal shutdown by panic. etc	Faulted	Killed

Note

None.

M.2.11 Operation Procedures and Displayed Items for Cluster Application Setup and Modification

Details on incompatibilities

The procedures and displayed items for creating userApplications and for creating and changing procedure resources have been changed.

Changes

1. Creating a userApplication for standby operation

   Before upgrading [PRIMECLUSTER 4.1A20]

   To create a userApplication for standby operation, select "CRM" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) CRM
 6) GENERIC
 7) SYMFOWARE
Application Type: 5
```

```
Settings of turnkey wizard "CRM"
 1) HELP                4) REMOVE+EXIT              7) Machines+Basics(-)
 2) NO-SAVE+EXIT        5) ApplicationName=APP1
 3) SAVE+EXIT           6) BeingControlled=no
Choose the setting to process: 7
```

   After upgrading [PRIMECLUSTER 4.4A00]

   "CRM" is not displayed in the "Application type selection menu."

   To create a userApplication for standby operation, select "STANDBY" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 8
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP1
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

2. Creating a userApplication for scalable operation

   Before upgrading [PRIMECLUSTER 4.1A20]

   To create a userApplication for scalable operation, select "Controller" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                  9) Controller          17) Rawdisk
 2) QUIT                 10) Crm                 18) SYMFOWARE
 3) RETURN               11) Fsystem             19) Symfoware
 4) OPTIONS              12) GENERIC             20) SystemState2
 5) Application          13) Gds                 21) SystemState3
 6) BasicApplication     14) Gls                 22) foreign-code
 7) CRM                  15) Ipaddress           23) generic
 8) Cmdline              16) Lvm
Application Type: 9
```

```
Settings of application type "Controller" (not yet consistent)
 1) HELP                              5) AdditionalAppToControl
 2) NO-SAVE+EXIT                      6) (InParallel=)
 3) ApplicationName=Controller        7) (FaultScript=)
 4) ControlPolicy=FOLLOW
Choose the setting to process: 4
```

After upgrading [PRIMECLUSTER 4.4A00]

"Controller" is not displayed in the "Application type selection menu."

To create a userApplication for scalable operation, select "SCALABLE" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                              7) SCALABLE
 2) QUIT                              8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 7
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "SCALABLE" (APP2:not yet consistent)
 1) HELP                              4) REMOVE+EXIT
 2) NO-SAVE+EXIT                      5) ApplicationName=APP2
 3) SAVE+EXIT                         6) Machines+Basics(-)
Choose the setting to process: 6
```

3. Creating a procedure resource

Before upgrading [PRIMECLUSTER 4.1A20]

To create or change a procedure resource, select "CRM" from the "turnkey wizard "CRM"" menu, and then select the resource class name.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "CRM"
 1) HELP                              10) LocalFileSystems(-)
 2) -                                 11) RemoteFileSystems(-)
 3) SAVE+EXIT                         12) IpAddresses(-)
 4) -                                 13) Gds:Global-Disk-Services(-)
 5) ApplicationName=APP1              14) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)
 7) CommandLines(-)
 8) Controllers(-)
 9) CRM(-)
Choose the setting to process: 9
```

```
Crm (CRM_APP1:not yet consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) SystemState2(-)
 6) SystemState3(-)
 7) BasicApplication(-)
 8) Application(-)
Choose the setting to process: 7
```

After upgrading [PRIMECLUSTER 4.4A00]

The "turnkey wizard "CRM"" menu is not displayed.

To create or change a procedure resource, select "Procedure:resource-class-name" from the "turnkey wizard STANDBY" menu.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                               9) Procedure:BasicApplication(-)
 2) -                                 10) Symfoware(-)
 3) SAVE+EXIT                         11) Procedure:SystemState3(-)
 4) -                                 12) Procedure:SystemState2(-)
 5) ApplicationName=APP1              13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)             14) IpAddresses(-)
 7) CommandLines(-)                   15) LocalFileSystems(-)
 8) Procedure:Application(-)          16) Gds:Global-Disk-Services(-)
Choose the setting to process:9
```

4. Changing the timeout value of a procedure resource

Before upgrading [PRIMECLUSTER 4.1A20]

To change the timeout value, select "Flags[0]=OTtimeout-time" and then select "SCRIPTTIMEOUT."

```
Settings of "Application"
 1) HELP               4) REMOVE+RETURN           7) Flags[0]=OT1800
 2) NO-SAVE+RETURN     5) AdditionalResource      8) Priority[0]=65535
 3) SAVE+RETURN        6) ResourceName[0]=rsc1
Choose the setting to process: 7
```

```
Set Flags[0]:Currently set: STANDBYCAPABLE,SCRIPTTIMEOUT (OT1800)
 1) HELP              5) NULLDETECTOR(D)        9) MONITORONLY(M)
 2) -                 6) LIEOFFLINE(L)         10) NOT:STANDBYCAPABLE(O)
 3) SAVE+RETURN       7) CLUSTEREXCLUSIVE(X)   11) SCRIPTTIMEOUT(T)
 4) DEFAULT           8) AUTORECOVER(A)
Choose one of the flags: 11
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate script timeout:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Flags[0]=OTtimeout-time" is not displayed.

To change the timeout value, select "TIMEOUT."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 4
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate timeout:
```

5. Changing the priority in a resource class of a procedure resource.

Before upgrading [PRIMECLUSTER 4.1A20]

To change the priority in a resource class, select "Priority[0]=*priority*."

```
Settings of "Application"
 1) HELP               4) REMOVE+RETURN         7) Flags[0]=OT1800
 2) NO-SAVE+RETURN     5) AdditionalResource    8) Priority[0]=65535
 3) SAVE+RETURN        6) ResourceName[0]=rsc1
Choose the setting to process: 8
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
Set a priority:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Priority[0]=*priority*" is not displayed.

To change the priority in a resource class, select "PRIORITY."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 5
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 65535
Set an appropriate priority:
```

Note

None.

M.2.12 Setting Up Fsystem Resources

Details on incompatibilities

The file which defines the mount point of the file system used as Fsystem resource has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

The mount point was defined in /etc/fstab.

After upgrading [PRIMECLUSTER 4.4A00]

It is necessary to define the mount point in /etc/fstab.pcl.

For details, see "6.7.3.2 Setting Up Fsystem Resources."

Note

None.

M.2.13 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.2.14 Changes of the 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 4.1A20]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.1A20.

M.2.15 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds).

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.2.16 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.2.17 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.2.18 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 4.1A20]

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 4.4A00]

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.

M.2.19 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 4.1A20]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Fault Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.2.20 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.2.21 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.2.22 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

(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 4.4A00]

(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.

M.2.23 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 4.1A20]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.2.24 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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 4.4A00]

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.

M.2.25 Changes of the response message for the operator intervention request

M. Message: 1421
2.25.1

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M. Message: 1423 2.25.2

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.3 Changes in PRIMECLUSTER 4.4A00 from 4.1A30

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A30.

- M.3.1 ciptool command
- M.3.2 sdtool command
- M.3.3 hvshut command
- M.3.4 hvswitch command
- M.3.5 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A30.

- M.3.6 Posting Notification of a Resource Failure or Recovery
- M.3.7 Operator Intervention Request
- M.3.8 Operation Procedures and Displayed Items for Cluster Application Setup and Modification
- M.3.9 Setting Up Fsystem Resources
- M.3.10 Client Environment for Web-Based Admin View
- M.3.11 Changes of the Behavior of CF Startup
- M.3.12 HV_CONNECT_TIMEOUT
- M.3.13 Changes of the ports used by RMS
- M.3.14 Changes of the port number used by the shutdown facility
- M.3.15 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.3.16 Display of the resource fault trace
- M.3.17 Change of /etc/cip.cf file
- M.3.18 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A30.

- M.3.19 Changes of the RMS message
- M.3.20 Changes of the importance of the message in the RMS wizard
- M.3.21 Changes of RMS console message
- M.3.22 Changes of the response message for the operator intervention request

M.3.1 ciptool command

Details on incompatibilities

Symbolic links of the ciptool commands were not made for /usr/bin.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

Symbolic links of the ciptool commands are made for /usr/bin.

After upgrading [PRIMECLUSTER 4.4A00]

Symbolic links of the ciptool commands are not made for /usr/bin.

Note

Specify /opt/SMAW/SMAWcf/bin/ciptool by full passing when you use the ciptool command.

M.3.2 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

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 4.4A00]

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.

M.3.3 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 4.1A30]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.3.4 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 4.1A30]

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 4.4A00]

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.

M.3.5 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.3.6 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 4.4A00. For details, see "5.2 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 4.1A30]

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 4.4A00]

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.

M.3.7 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 4.1A30]

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 4.4A00]

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

For details, see "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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

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 4.4A00]

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.

M.3.8 Operation Procedures and Displayed Items for Cluster Application Setup and Modification

Details on incompatibilities

The procedures and displayed items for creating userApplications and for creating and changing procedure resources have been changed.

Changes

1. Creating a userApplication for standby operation

Before upgrading [PRIMECLUSTER 4.1A30]

To create a userApplication for standby operation, select "CRM" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP
 2) QUIT
 3) RETURN
 4) OPTIONS
 5) CRM
 6) GENERIC
 7) SYMFOWARE
Application Type: 5
```

```
Settings of turnkey wizard "CRM"
 1) HELP               4) REMOVE+EXIT             7) Machines+Basics(-)
 2) NO-SAVE+EXIT       5) ApplicationName=APP1
 3) SAVE+EXIT          6) BeingControlled=no
Choose the setting to process: 7
```

After upgrading [PRIMECLUSTER 4.4A00]

"CRM" is not displayed in the "Application type selection menu."

To create a userApplication for standby operation, select "STANDBY" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 8
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP1
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

2. Creating a userApplication for scalable operation

Before upgrading [PRIMECLUSTER 4.1A30]

To create a userApplication for scalable operation, select "Controller" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                  9) Controller          17) Rawdisk
 2) QUIT                 10) Crm                 18) SYMFOWARE
 3) RETURN               11) Fsystem             19) Symfoware
 4) OPTIONS              12) GENERIC             20) SystemState2
 5) Application          13) Gds                 21) SystemState3
 6) BasicApplication     14) Gls                 22) foreign-code
 7) CRM                  15) Ipaddress           23) generic
 8) Cmdline              16) Lvm
Application Type: 9
```

```
Settings of application type "Controller" (not yet consistent)
 1) HELP                            5) AdditionalAppToControl
 2) NO-SAVE+EXIT                    6) (InParallel=)
 3) ApplicationName=Controller      7) (FaultScript=)
 4) ControlPolicy=FOLLOW
Choose the setting to process: 4
```

After upgrading [PRIMECLUSTER 4.4A00]

"Controller" is not displayed in the "Application type selection menu."

To create a userApplication for scalable operation, select "SCALABLE" from the "Application type selection menu."

```
Creation: Application type selection menu:
 1) HELP                          7) SCALABLE
 2) QUIT                          8) STANDBY
 3) RETURN
 4) OPTIONS
 5) DEMO
 6) GENERIC
Application Type: 7
```

```
Consistency check ...
Yet to do: process the basic settings using Machines+Basics
Yet to do: process at least one of the non-basic settings

Settings of turnkey wizard "SCALABLE" (APP2:not yet consistent)
 1) HELP                          4) REMOVE+EXIT
 2) NO-SAVE+EXIT                  5) ApplicationName=APP2
 3) SAVE+EXIT                     6) Machines+Basics(-)
Choose the setting to process: 6
```

3. Creating a procedure resource

Before upgrading [PRIMECLUSTER 4.1A30]

To create or change a procedure resource, select "CRM" from the "turnkey wizard "CRM"" menu, and then select the resource class name.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "CRM"
 1) HELP                          10) LocalFileSystems(-)
 2) -                             11) RemoteFileSystems(-)
 3) SAVE+EXIT                     12) IpAddresses(-)
 4) -                             13) Gds:Global-Disk-Services(-)
 5) ApplicationName=APP1          14) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)
 7) CommandLines(-)
 8) Controllers(-)
 9) CRM(-)
Choose the setting to process: 9
```

```
Crm (CRM_APP1:not yet consistent)
 1) HELP
 2) -
 3) SAVE+EXIT
 4) REMOVE+EXIT
 5) SystemState2(-)
 6) SystemState3(-)
 7) BasicApplication(-)
 8) Application(-)
Choose the setting to process: 7
```

After upgrading [PRIMECLUSTER 4.4A00]

The "turnkey wizard "CRM"" menu is not displayed.

To create or change a procedure resource, select "Procedure:resource-class-name" from the "turnkey wizard STANDBY" menu.

Note: Shown below is an example of registering a procedure resource of the BasicApplication class to a userApplication.

```
Settings of turnkey wizard "STANDBY" (APP1:not yet consistent)
 1) HELP                              9) Procedure:BasicApplication(-)
 2) -                                10) Symfoware(-)
 3) SAVE+EXIT                        11) Procedure:SystemState3(-)
 4) -                                12) Procedure:SystemState2(-)
 5) ApplicationName=APP1             13) Gls:Global-Link-Services(-)
 6) Machines+Basics(app1)            14) IpAddresses(-)
 7) CommandLines(-)                  15) LocalFileSystems(-)
 8) Procedure:Application(-)         16) Gds:Global-Disk-Services(-)
Choose the setting to process:9
```

4. Changing the timeout value of a procedure resource

Before upgrading [PRIMECLUSTER 4.1A30]

To change the timeout value, select "Flags[0]=OTtimeout-time" and then select "SCRIPTTIMEOUT."

```
Settings of "Application"
 1) HELP                4) REMOVE+RETURN           7) Flags[0]=OT1800
 2) NO-SAVE+RETURN      5) AdditionalResource      8) Priority[0]=65535
 3) SAVE+RETURN         6) ResourceName[0]=rsc1
Choose the setting to process: 7
```

```
Set Flags[0]:Currently set: STANDBYCAPABLE,SCRIPTTIMEOUT (OT1800)
 1) HELP               5) NULLDETECTOR(D)         9) MONITORONLY(M)
 2) -                  6) LIEOFFLINE(L)          10) NOT:STANDBYCAPABLE(O)
 3) SAVE+RETURN        7) CLUSTEREXCLUSIVE(X)    11) SCRIPTTIMEOUT(T)
 4) DEFAULT            8) AUTORECOVER(A)
Choose one of the flags: 11
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate script timeout:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Flags[0]=OTtimeout-time" is not displayed.

To change the timeout value, select "TIMEOUT."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 4
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 1800
Set an appropriate timeout:
```

5. Changing the priority in a resource class of a procedure resource.

Before upgrading [PRIMECLUSTER 4.1A30]

To change the priority in a resource class, select "Priority[0]=*priority*."

```
Settings of "Application"
 1) HELP              4) REMOVE+RETURN         7) Flags[0]=OT1800
 2) NO-SAVE+RETURN    5) AdditionalResource    8) Priority[0]=65535
 3) SAVE+RETURN       6) ResourceName[0]=rsc1
Choose the setting to process: 8
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
Set a priority:
```

After upgrading [PRIMECLUSTER 4.4A00]

"Priority[0]=*priority*" is not displayed.

To change the priority in a resource class, select "PRIORITY."

```
Set flags for Procedure resource :
Currently set: TIMEOUT (T1800), PRIORITY (P65535)
 1) HELP
 2) -
 3) SAVE+RETURN
 4) TIMEOUT
 5) PRIORITY
Choose one of the flags: 5
```

```
 1) HELP
 2) RETURN
 3) FREECHOICE
 4) 65535
Set an appropriate priority:
```

Note

None.

M.3.9 Setting Up Fsystem Resources

Details on incompatibilities 1

The file which defines the mount point of the file system used as Fsystem resource has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

The mount point was defined in /etc/fstab.

After upgrading [PRIMECLUSTER 4.4A00]

It is necessary to define the mount point in /etc/fstab.pcl.

For details, see "6.7.3.2 Setting Up Fsystem Resources."

Note

None.

Details on incompatibilities 2

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.3.10 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.3.11 Changes of the 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 4.1A30]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.1A30.

M.3.12 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds).

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.3.13 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.3.14 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.3.15 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 4.1A30]

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 4.4A00]

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.

M.3.16 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 4.1A30]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.3.17 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.3.18 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.3.19 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

(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 4.4A00]

(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.

M.3.20 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 4.1A30]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.3.21 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

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 4.4A00]

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.

M.3.22 Changes of the response message for the operator intervention request

M. 3.22.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A30]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access

by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly. Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M. Message: 1423 3.22.2

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A20]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly. Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.4 Changes in PRIMECLUSTER 4.4A00 from 4.1A40

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A40.

- M.4.1 sdtool command
- M.4.2 hvshut command
- M.4.3 hvswitch command
- M.4.4 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A40.

- M.4.5 Posting Notification of a Resource Failure or Recovery
- M.4.6 Operator Intervention Request
- M.4.7 Setting Up Fsystem Resources
- M.4.8 Client Environment for Web-Based Admin View
- M.4.9 Changes of the Behavior of CF Startup
- M.4.10 HV_CONNECT_TIMEOUT
- M.4.11 Changes of the ports used by RMS
- M.4.12 Changes of the port number used by the shutdown facility
- M.4.13 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.4.14 Display of the resource fault trace
- M.4.15 Change of /etc/cip.cf file
- M.4.16 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.1A40.

- M.4.17 Changes of the RMS message
- M.4.18 Changes of the importance of the message in the RMS wizard
- M.4.19 Changes of RMS console message
- M.4.20 Changes of the response message for the operator intervention request

M.4.1 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

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 4.4A00]

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.

M.4.2 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.

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 4.1A40]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.4.3 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 4.1A40]

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 4.4A00]

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.

M.4.4 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.4.5 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 4.4A00. For details, see "5.2 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 4.1A40]

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 4.4A00]

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.

M.4.6 Operator Intervention Request

Details on incompatibilities 1

To enable the operator intervention request, you have to set the AppWatch parameter to ON with clsetparam. With the default settings made when the cluster was installed, the operator intervention request, is disabled. For details, see "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

The operator intervention request, is always enabled.

After upgrading [PRIMECLUSTER 4.4A00]

The operator intervention request, is enabled only when the AppWatch parameter is set to ON with clsetparam. 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.

Note

For details, see "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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

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 4.4A00]

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 restartt all the nodes to validate the setting.

M.4.7 Setting Up Fsystem Resources

Details on incompatibilities 1

The file which defines the mount point of the file system used as Fsystem resource has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

The mount point was defined in /etc/fstab.

After upgrading [PRIMECLUSTER 4.4A00]

It is necessary to define the mount point in /etc/fstab.pcl.

For details, see "6.7.3.2 Setting Up Fsystem Resources."

Note

None.

Details on incompatibilities 2

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.4.8 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.4.9 Changes of the 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 4.1A40]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.1A40.

M.4.10 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds).

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.4.11 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.4.12 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.4.13 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 4.1A40]

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 4.4A00]

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.

M.4.14 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 4.1A40]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.4.15 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.4.16 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.4.17 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

(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 4.4A00]

(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.

M.4.18 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 4.1A40]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.4.19 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

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 4.4A00]

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.

M.4.20 Changes of the response message for the operator intervention request

M.4.20.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M. Message: 1423
4.20.2

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.1A40]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".

Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly. Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.5 Changes in PRIMECLUSTER 4.4A00 from 4.2A00

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A00.

- M.5.1 sdtool command
- M.5.2 hvshut command
- M.5.3 hvswitch command
- M.5.4 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A00.

- M.5.5 Posting Notification of a Resource Failure or Recovery
- M.5.6 Operator Intervention Request
- M.5.7 Setting Up Fsystem Resources
- M.5.8 Client Environment for Web-Based Admin View
- M.5.9 Changes of the Behavior of CF Startup
- M.5.10 HV_CONNECT_TIMEOUT
- M.5.11 Changes of the ports used by RMS
- M.5.12 Configuring the IPMI Shutdown Agent
- M.5.13 Changes of the port number used by the shutdown facility
- M.5.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.5.15 Display of the resource fault trace
- M.5.16 Change of /etc/cip.cf file
- M.5.17 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A00.

- M.5.18 Changes of the RMS message
- M.5.19 Changes of the importance of the message in the RMS wizard
- M.5.20 Changes of RMS console message
- M.5.21 Changes of the response message for the operator intervention request

M.5.1 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

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 4.4A00]

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.

M.5.2 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]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.5.3 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 4.4A00]

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.

M.5.4 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 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.5.5 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 4.4A00. For details, see "5.2 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 4.4A00]

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.

M.5.6 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 4.4A00]

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

For details, see "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 "5.2 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 4.4A00]

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.

M.5.7 Setting Up Fsystem Resources

Details on incompatibilities 1

The file which defines the mount point of the file system used as Fsystem resource has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The mount point was defined in /etc/fstab.

After upgrading [PRIMECLUSTER 4.4A00]

It is necessary to define the mount point in /etc/fstab.pcl.

For details, see "6.7.3.2 Setting Up Fsystem Resources."

Note

None.

Details on incompatibilities 2

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.5.8 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.5.9 Changes of the 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 4.4A00]

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. eth0).

This message is also available in 4.2A00.

M.5.10 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 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.5.11 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.5.12 Configuring the IPMI Shutdown Agent

Details on incompatibilities

The setting procedure to use the IPMI shutdown agent is added.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

The following settings were unnecessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

After upgrading [PRIMECLUSTER 4.4A00]

The following settings are necessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

Note

None.

M.5.13 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.5.14 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 4.4A00]

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.

M.5.15 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]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.5.16 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.5.17 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.2A00]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.5.18 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 4.4A00]

(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.

M.5.19 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 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.5.20 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

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 4.4A00]

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.

M.5.21 Changes of the response message for the operator intervention request

M.5.21.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has 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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.5.21.2 Message: 1423

Details on incompatibilities

Message No.1423 of the operator intervention request has 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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.6 Changes in PRIMECLUSTER 4.4A00 from 4.2A30

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A30.

- M.6.1 sdtool command
- M.6.2 hvshut command
- M.6.3 hvswitch command
- M.6.4 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A30.

- M.6.5 Posting Notification of a Resource Failure or Recovery
- M.6.6 Operator Intervention Request
- M.6.7 Setting Up Fsystem Resources
- M.6.8 Client Environment for Web-Based Admin View
- M.6.9 Changes of the Behavior of CF Startup
- M.6.10 HV_CONNECT_TIMEOUT
- M.6.11 Changes of the ports used by RMS
- M.6.12 Configuring the IPMI Shutdown Agent
- M.6.13 Changes of the port number used by the shutdown facility
- M.6.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.6.15 Display of the resource fault trace
- M.6.16 Change of /etc/cip.cf file
- M.6.17 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.2A30.

- M.6.18 Changes of the RMS message
- M.6.19 Changes of the importance of the message in the RMS wizard
- M.6.20 Changes of RMS console message
- M.6.21 Changes of the response message for the operator intervention request

M.6.1 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

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 4.4A00]

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.

M.6.2 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 4.2A30]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.6.3 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 4.2A30]

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 4.4A00]

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.

M.6.4 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.6.5 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 4.4A00. For details, see "5.2 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 4.2A30]

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 4.4A00]

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.

M.6.6 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 4.2A30]

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 4.4A00]

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

For details, see "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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

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 4.4A00]

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.

M.6.7 Setting Up Fsystem Resources

Details on incompatibilities

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.6.8 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.6.9 Changes of the 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 4.2A30]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.2A30.

M.6.10 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds) in the RHEL-AS environment, and 30 (seconds) in the RHEL5 environment.

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

There are no incompatibilities when upgrading PRIMECLUSTER from 4.2A30 for RHEL5 to 4.4A00.

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.6.11 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.6.12 Configuring the IPMI Shutdown Agent

Details on incompatibilities

The setting procedure to use the IPMI shutdown agent is added.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

The following settings were unnecessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

After upgrading [PRIMECLUSTER 4.4A00]

The following settings are necessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

Note

None.

M.6.13 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.6.14 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 4.2A30]

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 4.4A00]

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.

M.6.15 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 4.2A30]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.6.16 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.6.17 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.6.18 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

(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 4.4A00]

(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.

M.6.19 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 4.2A30]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.6.20 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

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 4.4A00]

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.

M.6.21 Changes of the response message for the operator intervention request

M.6.21.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.2A30]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may

be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M. Message: 1423 6.21.2

Details on incompatibilities

Message No.1423 of the operator intervention request has 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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.7 Changes in PRIMECLUSTER 4.4A00 from 4.3A00

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A00.

- M.7.1 sdtool command
- M.7.2 hvshut command
- M.7.3 hvswitch command
- M.7.4 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A00.

- M.7.5 Posting Notification of a Resource Failure or Recovery
- M.7.6 Operator Intervention Request
- M.7.7 Setting Up Fsystem Resources
- M.7.8 Client Environment for Web-Based Admin View
- M.7.9 Changes of the Behavior of CF Startup
- M.7.10 HV_CONNECT_TIMEOUT
- M.7.11 Changes of the ports used by RMS
- M.7.12 Configuring the IPMI Shutdown Agent
- M.7.13 Changes of the port number used by the shutdown facility
- M.7.14 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.7.15 Display of the resource fault trace
- M.7.16 Change of /etc/cip.cf file
- M.7.17 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A00.

- M.7.18 Changes of the RMS message
- M.7.19 Changes of the importance of the message in the RMS wizard
- M.7.20 Changes of RMS console message
- M.7.21 Changes of the response message for the operator intervention request

M.7.1 sdtool command

Details on incompatibilities

The number of characters displayed by "sdtool -s" or "sdtool -C" has been changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

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 4.4A00]

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.

M.7.2 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 4.3A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.7.3 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 4.3A00]

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 4.4A00]

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.

M.7.4 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.7.5 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 4.4A00. For details, see "5.2 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 4.3A00]

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 4.4A00]

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.

M.7.6 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 4.3A00]

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 4.4A00]

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

For details, see "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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

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 4.4A00]

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.

M.7.7 Setting Up Fsystem Resources

Details on incompatibilities

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.7.8 Client Environment for Web-Based Admin View

Details on incompatibilities

Linux(R) is not supported as a client environment for Web-Based Admin View by PRIMECLUSTER 4.4A00.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

Linux(R) is supported as a client environment for Web-Based Admin View.

After upgrading [PRIMECLUSTER 4.4A00]

Linux(R) is not supported as a client environment for Web-Based Admin View.

Note

None.

M.7.9 Changes of the 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 4.3A00]

CF does not start unless all of the network interfaces for the cluster interconnects are recognized.

After upgrading [PRIMECLUSTER 4.4A00]

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. eth0).

This message is also available in 4.3A00.

M.7.10 HV_CONNECT_TIMEOUT

Details on incompatibilities

The default value of the RMS local environment variables HV_CONNECT_TIMEOUT is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

The default value of HV_CONNECT_TIMEOUT is 5 (seconds) in the RHEL-AS environment, and 30 (seconds) in the RHEL5 environment.

After upgrading [PRIMECLUSTER 4.4A00]

The default value of HV_CONNECT_TIMEOUT is 30 (seconds).

Note

There are no incompatibilities when upgrading PRIMECLUSTER from 4.3A00 for RHEL5 to 4.4A00.

For details on HV_CONNECT_TIMEOUT, see "13.3 Local environment variables" in "PRIMECLUSTER Reliant Monitor Services (RMS) with Wizard Tools Configuration and Administration Guide."

M.7.11 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.7.12 Configuring the IPMI Shutdown Agent

Details on incompatibilities

The setting procedure to use the IPMI shutdown agent is added.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

The following settings were unnecessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

After upgrading [PRIMECLUSTER 4.4A00]

The following settings are necessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

Note

None.

M.7.13 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.7.14 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 4.3A00]

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 4.4A00]

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.

M.7.15 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 4.3A00]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.7.16 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.7.17 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.7.18 Changes of the RMS message

Details on incompatibilities

The RMS message (SYS, 8) logged in the syslog have been changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

(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 4.4A00]

(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.

M.7.19 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 4.3A00]

WARNING: cannot grab mount lock for dostat() check_getbdev(), returning previous state

After upgrading [PRIMECLUSTER 4.4A00]

NOTICE: cannot grab mount lock for dostat() check_getbdev(), returning previous state

Note

None.

M.7.20 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

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 4.4A00]

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.

M.7.21 Changes of the response message for the operator intervention request

M.7.21.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A00]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.7.21.2 Message: 1423

Details on incompatibilities

Message No.1423 of the operator intervention request has 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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.8 Changes in PRIMECLUSTER 4.4A00 from 4.3A10

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A10.

- M.8.1 sdtool command
- M.8.2 hvshut command
- M.8.3 hvswitch command
- M.8.4 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A10.

- M.8.5 Posting Notification of a Resource Failure or Recovery
- M.8.6 Operator Intervention Request

- M.8.7 Setting Up Fsystem Resources
- M.8.8 Changes of the ports used by RMS
- M.8.9 Configuring the IPMI Shutdown Agent
- M.8.10 Changes of the port number used by the shutdown facility
- M.8.11 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment
- M.8.12 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.8.13 Display of the resource fault trace
- M.8.14 Change of /etc/cip.cf file
- M.8.15 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A10.

- M.8.16 Changes of RMS console message
- M.8.17 Changes of the response message for the operator intervention request

M.8.1 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]

A fixed string "cluster" is displayed when the resource management facility is configured.

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 4.4A00]

The cluster class uses the same name as the CF cluster when the resource management facility is configured.

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.

M.8.2 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 4.3A10]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.8.3 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 4.3A10]

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 4.4A00]

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.

M.8.4 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 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.8.5 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 4.4A00. For details, see "5.2 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 4.4A00]

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.

M.8.6 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 4.3A10]

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 4.4A00]

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

For details, see "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 "5.2 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 4.4A00]

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.

M.8.7 Setting Up Fsystem Resources

Details on incompatibilities

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.8.8 Changes of the ports used by RMS

Details on incompatibilities

The port used by RMS is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The port number "11111" is used.

After upgrading [PRIMECLUSTER 4.4A00]

The port number "11111" is not used.

Note

None.

M.8.9 Configuring the IPMI Shutdown Agent

Details on incompatibilities

The setting procedure to use the IPMI shutdown agent is added.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

The following settings were unnecessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

After upgrading [PRIMECLUSTER 4.4A00]

The following settings are necessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

Note

None.

M.8.10 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.8.11 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment

Details on incompatibilities

When using the Host OS failover function in the PRIMEQUEST KVM environment in PRIMECLUSTER 4.4A00, it is required to set the shutdown facility on the Host OS (node).

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

When using the Host OS failover function in the PRIMEQUEST KVM environment, the shutdown facility was set on a guest OS (node).

After upgrading [PRIMECLUSTER 4.4A00]

When using the Host OS failover function in the PRIMEQUEST KVM environment, the setting of the shutdown facility is required not only on the guest OS (node) but also on the Host OS (node). This will enable you to reduce the cluster failover time between guest OSes if a failure occurs on the Host OS.

For details on the setting, see "5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)."

Note

None.

M.8.12 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 4.4A00]

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.

M.8.13 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]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.8.14 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.8.15 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.8.16 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

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 4.4A00]

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.

M.8.17 Changes of the response message for the operator intervention request

M.8.17.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.8.17.2 Message: 1423

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A10]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly.
Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.9 Changes in PRIMECLUSTER 4.4A00 from 4.3A20

Incompatible commands

The following commands of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A20.

- M.9.1 hvshut command
- M.9.2 hvswitch command
- M.9.3 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A20.

- M.9.4 Posting Notification of a Resource Failure or Recovery
- M.9.5 Operator intervention request
- M.9.6 Setting Up Fsystem Resources

- M.9.7 Configuring the IPMI Shutdown Agent
- M.9.8 Changes of the port number used by the shutdown facility
- M.9.9 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment
- M.9.10 Changes of the target node to forcibly shut down when a heartbeat failure occurs
- M.9.11 Display of the resource fault trace
- M.9.12 Change of /etc/cip.cf file
- M.9.13 Changes in CF over IP setting window of CF Wizard

Incompatible messages

The following messages of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A20.

- M.9.14 Changes of RMS console message
- M.9.15 Changes of the response message for the operator intervention request

M.9.1 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.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default and the shutdown processing of a resource by the hvshut command has not been completed in 900 (seconds), the command times out and then RMS ends abnormally. The resource does not stop and remains running at this time.

After upgrading [PRIMECLUSTER 4.4A00]

In the environment where the environment variable RELIANT_SHUT_MIN_WAIT remains in default, the hvshut command does not time out even when the shutdown processing of a resource by the command has not been completed.

Note

When using RMS, make sure to change this environment variable to suite the configuration setting.

M.9.2 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 4.3A20]

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 4.4A00]

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.

M.9.3 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 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.9.4 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 4.4A00. For details, see "5.2 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 4.4A00]

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.

M.9.5 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 4.3A20]

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 4.4A00]

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

For details, see "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 "5.2 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 4.4A00]

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.

M.9.6 Setting Up Fsystem Resources

Details on incompatibilities

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.9.7 Configuring the IPMI Shutdown Agent

Details on incompatibilities

The setting procedure to use the IPMI shutdown agent is added.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

The following settings were unnecessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

After upgrading [PRIMECLUSTER 4.4A00]

The following settings are necessary to use the IPMI shutdown agent.

- Setting the IPMI service
- Encrypting the IPMI(BMC, iRMC) password which is set in /etc/opt/SMAW/SMAWsf/SA_ipmi.cfg

Note

None.

M.9.8 Changes of the port number used by the shutdown facility

Details on incompatibilities

The port number used by the shutdown facility is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

Port number: The port number "2316" is used.

```
sfadv           2316/udp                        # SMAWsf package
```

After upgrading [PRIMECLUSTER 4.4A00]

Port number: The port number "9382" is used.

```
sfadv           9382/udp                        # SMAWsf package
```

Note

None.

M.9.9 Setting up the Host OS failover function used in the PRIMEQUEST KVM environment

Details on incompatibilities

When using the Host OS failover function in the PRIMEQUEST KVM environment in PRIMECLUSTER 4.4A00, it is required to set the shutdown facility on the Host OS (node).

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

When using the Host OS failover function in the PRIMEQUEST KVM environment, the shutdown facility was set on a guest OS (node).

After upgrading [PRIMECLUSTER 4.4A00]

When using the Host OS failover function in the PRIMEQUEST KVM environment, the setting of the shutdown facility is required not only on the guest OS (node) but also on the Host OS (node). This will enable you to reduce the cluster failover time between guest OSes if a failure occurs on the Host OS.

For details on the setting, see "5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)."

Note

None.

M.9.10 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 4.3A20]

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 4.4A00]

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.

M.9.11 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]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.9.12 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.9.13 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.9.14 Changes of RMS console message

Details on incompatibilities

Due to the additional function "M.9.2 hvswitch command," RMS console messages that are displayed when the hvswitch -f command is executed are changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

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 4.4A00]

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.

M.9.15 Changes of the response message for the operator intervention request

M. 9.15.1 Message: 1421

Details on incompatibilities

Message No.1421 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

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? (no/yes) 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.

The cluster application "userApplication" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
Forcibly start the cluster application in SysNode "SysNode"? (no/yes) Message number: *number*
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" did not start automatically. This is because all SysNodes that configure the cluster application do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access

by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly. Is it OK to forcibly start the cluster application? (no/yes) Message No: number

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M. Message: 1423 9.15.2

Details on incompatibilities

Message No.1423 of the operator intervention request has changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A20]

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? (no/yes) 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.

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
Forcibly start the cluster application in SysNode "*SysNode*" ? (no/yes) Message No: number
Warning: When cluster application is forcibly started, the safety check becomes disabled. If the operation is used incorrectly, data may be damaged and the consistency may be lost. Check that the cluster application to be forcibly started is not online in the cluster before executing the forced startup.

After upgrading [PRIMECLUSTER 4.4A00]

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 the 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*

The cluster application "*userApplication*" of SysNode "*SysNode*" has a resource "*resource*" problem. The cluster application "*userApplication*" did not start automatically because all SysNodes do not start within the given time.
You can forcibly start the cluster application in SysNode "*SysNode*".
Warning: When you forcibly start the cluster application, make sure that RMS is running on all the nodes. If there is a node without running RMS, shut down it manually before forcibly start the cluster application. Data corruption may occur due to the concurrent access by multiple nodes when performing the forced startup. To reduce the risk, the nodes without running RMS may be shut down forcibly. Is it OK to forcibly start the cluster application? (no/yes) Message No: *number*

Note

For details, see the relevant message in "PRIMECLUSTER Messages."

M.10 Changes in PRIMECLUSTER 4.4A00 from 4.3A30

Incompatible command

The following command of PRIMECLUSTER 4.4A00 is incompatible with PRIMECLUSTER 4.3A30.

- M.10.1 hvdump command

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A30.

- M.10.2 Posting Notification of a Resource Failure or Recovery
- M.10.3 Operator intervention request
- M.10.4 Setting Up Fsystem Resources
- M.10.6 Display of the resource fault trace
- M.10.7 Change of /etc/cip.cf file
- M.10.8 Changes in CF over IP setting window of CF Wizard

M.10.1 hvdump command

Details on incompatibilities

The default work directory used by the hvdump(1M) command execution is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

The default work directory is /tmp.

After upgrading [PRIMECLUSTER 4.4A00]

The default work directory is /var/tmp.

Note

None.

M.10.2 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 4.4A00. For details, see "5.2 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 4.3A30]

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 4.4A00]

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.

M.10.3 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 "5.2 Setting Up Fault Resource Identification and Operator Intervention Request."

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

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 4.4A00]

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.

M.10.4 Setting Up Fsystem Resources

Details on incompatibilities

The dedicated monitoring disk area is not required when using a shared disk device.

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

To secure the dedicated monitoring disk area was required.

After upgrading [PRIMECLUSTER 4.4A00]

The dedicated monitoring disk area is not required to be registered to the userApplication as an Fsystem resource. But when migration from earlier version, it is available to register the dedicated monitoring disk area as an Fsystem resource.

Note

None.

M.10.5 Setting up the Host OS failover function when using it in KVM environment

Details on incompatibilities

From PRIMECLUSTER 4.4A00, the user to log in to the guest OS via SSH when using the host OS failover function is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

Only the root user can log in to the guest OS via SSH.

After upgrading [PRIMECLUSTER 4.4A00]

The root user or any specified user can log in to the guest OS via SSH.

For details, see "5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)."

Note

None.

M.10.6 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 4.3A30]

After the Offline processing of the failed resource is completed, nothing is displayed in StateDetails of the failed resource object.

After upgrading [PRIMECLUSTER 4.4A00]

After the Offline processing of the failed resource is completed, "Faulted Occurred" is displayed in StateDetails of the failed resource object.

Note

None.

M.10.7 Change of /etc/cip.cf file

Details on incompatibilities

There is a change on the item that can be set in /etc/cip.cf.

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

When setting IPv4 address, option specified for the setting command such as ifconfig can be specified for CIP interface.

After upgrading [PRIMECLUSTER 4.4A00]

When setting IPv4 address, only IP address and netmask value can be specified for CIP interface.

Note

None.

M.10.8 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.3A30]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.11 Changes in PRIMECLUSTER 4.4A00 from 4.3A40

Incompatible functions

The following functions of PRIMECLUSTER 4.4A00 are incompatible with PRIMECLUSTER 4.3A40.

- M.11.1 Setting up the Host OS failover function when using it in KVM environment
- M.11.2 Changes in CF over IP setting window of CF Wizard
- M.11.3 Setting up the migration function when using it in KVM environment

M.11.1 Setting up the Host OS failover function when using it in KVM environment

Details on incompatibilities

From PRIMECLUSTER 4.4A00, the user to log in to the guest OS via SSH when using the host OS failover function is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A40]

Only the root user can log in to the guest OS via SSH.

After upgrading [PRIMECLUSTER 4.4A00]

The root user or any specified user can log in to the guest OS via SSH.

For details, see "5.1.2.5.6 Setting up the Host OS Failover Function to the Host OS (PRIMEQUEST only)."

Note

None.

M.11.2 Changes in CF over IP setting window of CF Wizard

Details on incompatibilities

From PRIMECLUSTER 4.4A00, "Auto Subnet Grouping" checkbox is deleted from CF over IP setting window. Instead, "Use Network Broadcast" checkbox is newly added.

Changes

Before upgrading [PRIMECLUSTER 4.3A40]

You can select to use or not to use the auto subnet grouping by checking/unchecking "Auto Subnet Grouping" checkbox on CF over IP setting window of CF Wizard.

After upgrading [PRIMECLUSTER 4.4A00]

You can select to use or not to use the network broadcast on CF over IP by checking/unchecking "Use Network Broadcast" checkbox on CF over IP setting window of CF Wizard.

Note

None.

M.11.3 Setting up the migration function when using it in KVM environment

Details on incompatibilities

From PRIMECLUSTER 4.4A00, the user to log in to the guest OS via SSH when using the migration function is changed.

Changes

Before upgrading [PRIMECLUSTER 4.3A40]

Only the root user can log in to the guest OS via SSH.

After upgrading [PRIMECLUSTER 4.4A00]

The root user or any specified user can log in to the guest OS via SSH.

For details, see "G.2.2 Using the Host OS failover function."

Note

None.

Appendix N Release Information

This appendix explains the outline, details, and the part changed with PRIMECLUSTER 4.4.

No	Version / Level	Change Location	Details
1	4.4 Second edition	3.3 PRIMECLUSTER Installation	Added the GDS modules to the settings of blacklist.
2	4.4 Second edition	H.1 Cluster Systems in a VMware Environment H.2 Installation	Changed the type of available shared disk with I/O fencing function.
3	4.4 Second edition	I.3 Necessary Daemons for PRIMECLUSTER to Operate J.2 systemd Service Lists J.3 Necessary Services for PRIMECLUSTER to Operate	Added the remarks when using the mirroring among servers.
4	4.4 Second edition	L.3.2 Setup in Single-User Mode	Modified the procedure for changing the GDS settings.

GDS: Global Disk Services

Glossary

AC (Access Client)

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*.

API (application program interface)

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 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.

BMC (Baseboard Management Controller)

A dedicated processor for monitoring and diagnosis of environmental factors (e.g. temperature, voltage) and parts and units.

CB

Clustering Base

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 Monito

CIP

Cluster Interconnect Protocol

class (GDS)

See *disk class*.

CLI

command-line interface

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 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 partition

The state in which communication with some of the nodes that constitute the cluster has been stopped.

Cluster Resource Management facility

Facility that manages hardware units that are shared among multiple nodes.

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 only.)

See also *simple virtual disk, striped 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.

Crash dump collection facility

The facility that collects crash dumps if an OS error (panic, etc.) occurs. The crash dump collection facility includes kdump.

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.

database node (SIS)

Nodes that maintain the configuration, dynamic data, and statistics in a SIS configuration.

See also *gateway node*, *service node*, *Scalable Internet Services*.

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, mirrored among servers, striped, or concatenated according to the type attribute (mirror, netmirror, stripe, or concatenation) of that disk group.

A disk group is sometimes simply called a "group."

DLPI

Data Link Provider Interface

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 (Event Notification Services)(CF)

See *Event Notification Services*.

environment variable (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 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.

failover (RMS, SIS)

The process by which a user application automatically transfers processes and data inherited from an operating system to a standby system because some failure has occurred.
With SIS, this process switches a failed node to a backup node. With RMS, this process is known as switchover.

See also *automatic switchover*, *directed switchover*, *switchover*, *symmetrical switchover*.

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 Linux(R) 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.

gateway node (SIS)

Gateway nodes have an external network interface. All incoming packets are received by this node and forwarded to the selected service node, depending on the scheduling algorithm for the Service.

See also *service node*, *database node*, *Scalable Internet Services*.

See *disk group*.

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 shared file system

A shared file system that allows simultaneous access from multiple Linux(R) 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.

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*.

GUI (graphical user interfac)

See *graphical user interface*.

HA (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.

hub

Star-type wiring device used for LAN or fibre channels.

ICF

Internode Communication Facility

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.

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*.

I/F

Interface

I/O

input/output

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)

See *public LAN*.

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*.

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*.

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 (Meta Data Server)

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

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.

mirroring among servers (GDS)

To mirror the local disks (such as an internal disk) of more than one server via the network.

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.

native operating system

The part of an operating system that is always active and translates system calls into activities.

netmirror group (GDS)

A group whose type is "netmirror." It is the set of disks, which will be mirrored among servers.

netmirror volume (GDS)

A volume that is created within a netmirror group. Its data will be made redundant by mirroring among servers.

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

LAN duplexed mode that is provided by GLS. The duplexed NIC is used exclusively, and LAN monitoring between the Linux(R) 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 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 (Oracle Parallel Server)

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.

OSD (operating system dependent) (CF)

See *operating system dependent*.

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 directry to the interface (for example, hme0) of a network interface card.

physical machine

A server configured with actual hardware. This is used in contrast with a virtual machine, and is also referred to as a physical server.

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 one 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 the nodes that configure the cluster system is either UP or DOWN (there is no LEFCLUSTER node).

RAO

RMS-Add on

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 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 as displayed in a system graph.

resource state (RMS)

Current state of a resource.

RMS (Reliant Monitor Services)

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*.

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.

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.

SA

Shutdown Agent. SA forcibly stops the target node by receiving instructions from the Shutdown Facility.

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.

Scalable Internet Services (SIS)

Scalable Internet Services is a TCP connection load balancer, and dynamically balances network access loads across cluster nodes while maintaining normal client/server sessions for each connection.

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, a SDX disk may be called a single disk, a keep disk, a spare disk, or an undefined disk. An SDS 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.

service node (SIS)

Service nodes provide one or more TCP services (such as FTP, Telnet, and HTTP) and receive client requests forwarded by the gateway nodes.

See also *database node*, *gateway node*, and *Scalable Internet Services*.

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 reach 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*.

single disk (GDS)

SDX disk that does not belong to a group and can be used to create a single volume.

single volume (GDS)

A volume that is created in a single disk that not belong to a group. There is no data redundancy.

SIS (Scalable Internet Services)

See *Scalable Internet Services*.

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*, *simple 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.

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 configured with PRIMEPOWER, this function turns on all other powered-off nodes and disk array unit that are connected to nodes through RCI cables.

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

A disk accessible from a virtual machine.

virtual interface (VIP)

Virtual IP address or host name that SIS uses to make multiple nodes of the cluster appear as a single-system image.

volume (GDS)

See *logical volume (GDS)*.

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

Index

[Numbers]

11 standby....30
2-tier model....40
3-tier model....41

[A]

AC....597
Access Client....597
Activating Configuration Update Service for SA....379
Activating the Cluster Interconnect....85
Adding, Deleting, and Changing Hardware....283
Adding Hardware....283
API....597
application (RMS)....597
Application building procedure and manual reference locations....138
Application Program Interface....597
application template (RMS)....597
Assigning Users to Manage the Cluster....86
attribute (RMS)....597
Attributes....203
automatic switchover (RMS)....597
AutoRecover....243
AutoStartUp....212
AutoSwitchOver....211
availability....597

[B]

base cluster foundation (CF)....597
BLADE shutdown agent....12
BM(base monitor) (RMS)....597
BMC (Baseboard Management Controller)....597
Bringing Faulted Cluster Application to available state....262
Build Flow....2
Building a cluster....98
Building Cluster Applications....138

[C]

Cancellation of Configuration Update Service for SA....383
Cascade (using one cluster application)....33
CF....96,598
CF Main Window....249
Changing a CIP Address....303
Changing a Node Name....299
Changing a Procedure Resource....370
Changing Blade Settings....308
Changing Hardware....294
Changing IPMI Settings....306
Changing the cluster system configuration....283
Changing the MMB IP Address....305
Changing the Network Environment....300
Changing the operation attributes of a userApplication....338
Changing the RMS environment variables....343
Changing the User Name and Password for Controlling the MMB with RMCP....306
Checking PRIMECLUSTER designsheets....85
Checking the BMC or iRMC IP Address and the Configuration Information of the Shutdown Agent....381
Checking the Cluster Environment....211
Checking the Configuration....379
child (RMS)....598
Clash dump....10
class (GDS)....598
Clear fault....7
Clearing the Wait State of a Node....262
Client....40
cluster....598
Cluster Admin....95
Cluster Admin functions....95
Cluster Application Operations....261
Cluster application setup....211
Cluster Foundation....598
Cluster interconnect....10
cluster interconnect (CF)....598
Cluster nodes....40
Cluster partition....598
Cluster Resource Management facility....598
Cluster states....251
Cluster Systems in a VMware Environment....401
Cmdline....221
Common....94
concatenated virtual disk....598
concatenation (GDS)....598
Concurrent Viewing of Node and Cluster Application States....267
Configuration change of Cluster Applications....312
configuration file (RMS)....599
Configuration information or object attributes....259
Configuration of Configuration Update Service for SA....378
Configuration Update Service for SA....374
Confirming Web-Based Admin View Startup....88
Corrective Action for Failed Resources....273
Corrective Action in the event of a resource failure....271
Crash Dump....360
Crash dump collection facility....599
Creating Scalable Cluster Applications....181
Creating Standby Cluster Applications....175
CRM....96
CRM Main Window....250
custom detector (RMS)....599
custom type (RMS)....599

[D]

daemon....599
database node (SIS)....599
Deactivating Configuration Update Service for SA....383
Deleting a cluster application....314
Deleting a Procedure Resource....372
Deleting a resource....327
Deleting a userApplication....314
Deleting Hardware....288
Deleting the Hardware Resource....314

Detaching Resources from Operation....347
Detailed resource information....254
Detecting a Failed Resource....361
Detector....243
detector (RMS)....599
Determining the Cluster System Operation Mode....29
Determining the failover timing of a cluster application....43
Determining the Web-Based Admin View Operation Mode....40
Development....5
DHCP....11
directed switchover (RMS)....599
disk class (GDS)....599
disk group (GDS)....599
Displayed resource types....251
Displaying environment variables....270
Double fault....211
DOWN (CF)....599
Dynamic Reconfiguration....286,291

[E]

ENS (CF)....600
Entering maintenance mode for Cluster Application....263
environment variable (RMS)....600
Environment variables....169
error detection (RMS)....600
Ethernet....600
Event Notification Services (CF)....600
exclusive relationships between cluster applications....203
Executing Standby Restoration for the Operating Job 348
Executing the fjsnap or pclsnap Command....358
Exiting the Web-Based Admin View Screen....96

[F]

Failed Resource Message....362
failover....7
Failover....211
failover (RMS, SIS)....600
Failure Detection and Cause Identification if a Failure Occurs272
Fast switching mode....600
Fault Resource List....365
fault tolerant network....600
Feature Description of Configuration Update Service for SA374
File System Creation....161
File system setup....159
Flow of Maintenance 347
fsck....245
Fsystem....243
Function Selection....15

[G]

gateway node (SIS)....600
GDS Configuration Setup....145
Generate and Activate....202
generic type (RMS)....600
GFS Shared File System....160
GFS shared file system....601
Global Cluster Services menu functions....95
Global Disk Services....601
Global File Services....601
Global Link Services....601
GLS setup....140
graph (RMS)....601
graphical user interface....601
group (GDS)....601
Guest OS setup....62,69,78
GUI....601

[H]

HaltFlag....211
Heartbeat error....215
high availability....601
highest-order group (GDS)....601
hub....601
HV_APPLICATION....237
HV_AUTORECOVER....237
HV_FORCED_REQUEST....237
HV_INTENDED_STATE....237
HV_LAST_DET_REPORT....237
HV_NODENAME....237
HV_SCRIPT_TYPE....237

[I]

Initial Cluster Setup....98
Initial Configuration Setup....129
Initial GFS Setup....159
Initial GLS setup....140
Initial RMS Setup....140
Initial setup of the cluster resource management facility....128
Initial setup of the operation management server....87
Initial setup of Web-Based Admin View....87
Installation....3,406
Installation and environment setup of applications....83
Installation procedure and manual reference sections....5
interconnect (CF)....601
Internet Protocol address....601
internode communication facility....602
IP address....602
IP aliasing....602
IPMI shutdown agent....12

[K]

kdump shutdown agent....12
Kernel parameter....10
keyword (reserved words)....602

[L]

LAN....602
latency (RMS)....602
leaf object (RMS)....602
LEFTCLUSTER (CF)....602
link (RMS)....602
local area network....602
local host....602
log file....603
logical volume (GDS)....603
low-order group (GDS)....603

[M]
MAC address....603
Maintenance Types....347
Management server....40
Manual....93
Manual Pages....353
MDS....603
message....603
message queue....603
Meta Data Server(GFS)....603
mirrored volume (GDS)....603
mirror group (GDS)....603
mirroring (GDS)....603
monitoring agent....603
Monitoring Cluster Control Messages....271
Monitoring the PRIMECLUSTER System....265
Monitoring the State of a Cluster Application....266
Monitoring the State of a Node....265
Mountpoint....245
mount point....603
Mutual standby....31

[N]
N 1 standby....32
native operating system....604
network adapter....604
network interface card....604
network partition (CF)....604
Network segment....12
NIC switching mode....604
node....604
Node failure....211
node state (CF)....604
Node states....252
NODE_SCRIPTS_TIME_OUT....237
Notes on script creation....168
NTP server....10

[O]
object (RMS)....604
object definition (RMS)....604
object type (RMS)....604
online maintenance....605
operating system dependent (CF)....605
Operating the PRIMECLUSTER System....260
Operation and Maintenance....9
Operation Check by Restarting the System....382
Operation Check for Configuration Update Service for SA....382
Operation Environment of Configuration Update Service for SA....377
Operation menu functions....93
Operation Mode Change....9
Operations....249,252
OPS....605
Oracle Parallel Server....605
OSD (CF)....605
Other resource states....252
Output Message (Syslog)....386

[P]
parent (RMS)....605
patrol diagnosis....605
physical IP address....605
physical machine....605
Planning....2
Preparation Prior to Building a Cluster....84
Preparations for starting the Web-Based Admin View screen....85
Preparing the client environment....87
primary host (RMS)....605
PRIMECLUSTER....286
PRIMECLUSTER Clustering Base....14
PRIMECLUSTER Enterprise Edition....14
PRIMECLUSTER HA Server....14
PRIMECLUSTER Installation....80
PRIMECLUSTER Lite Pack....14
PRIMECLUSTER Products....352
PRIMECLUSTER Product Selection....14
PRIMECLUSTER services (CF)....605
Priority transferring (application of N 1 standby)....34
private network address....605
private resource (RMS)....605
Product Selection....14
PSA (PRIMEQUEST Server Agent)....12
public LAN....606

[Q]
queue....606
quorum....606

[R]
redundancy....606
Registering, Changing, and Deleting State Transition Procedure Resources for PRIMECLUSTER Compatibility....369
Registering a Procedure Resource....369
Registering Hardware Devices....131
Reliant Monitor Services (RMS)....606
remote host....606
remote node....606
Replacement test....8
reporting message (RMS)....606
Reserved word....220
resource (RMS)....606
resource database (CF)....606
resource definition (RMS)....606
Resource failure....211
Resource Fault History....95,363
Resource icons....251
resource label (RMS)....606
resource state (RMS)....607
Resource states....251
Restoration Method When Correct Information is not Distributed to All Nodes....383
Restoration of Configuration Update Service for SA....383
Restoring the Startup Configuration of the IPMI Service....383
RMS....96,607
RMS command....607
RMS configuration....607

RMS graphs....270
RMS Main Window....255
RMS Operation....260
RMS Tree....255
RMS Wizard kit....607
RMS Wizard Tools....607
Rolling update....607
route....607

[S]

Sample scripts....167
SAN....607
scalability....607
Scalable Internet Services (SIS)....607
Scalable Operation....36
scope (GDS)....607
script (RMS)....608
SDX disk (GDS)....608
SDX object (GDS)....608
service node (SIS)....608
Setting the Web-Based Admin View Language....89
Setting up CF and CIP....99
Setting Up Cluster Applications....171
Setting Up Cmdline Resources....187
Setting Up Disk Units....46
Setting up fault resource identification and operator intervention request....135
Setting Up Fsystem Resources....190
Setting Up Gds Resources....194
Setting Up Gls Resources....195
Setting Up Hardware Monitoring with ServerView....47
Setting Up NTP....46
Setting Up Online/Offline Scripts....166
Setting Up Procedure Resources....200
Setting Up Resources....186
Setting Up Shared Disks....151
Setting Up System Disk Mirroring....145
Setting Up Takeover Network Resources....197
Setting Up the Application Environment....166
Setting up the browser....90
Setting Up the Cluster High-Speed Failover Function....47
Setting up the Java Plug-in....90
Setting Up the Network....46
Setting Up the RMS Environment....211
Setting up the shutdown facility....101
Setting Up userApplication....175
Setting up Web-Based Admin View when GLS is used....144
sfsacfgupdate....385
shared disk connection confirmation....608
shared resource....608
Shared resource states....252
Shutdown Facility....608
shutdown request....608
simple virtual disk....608
Single-Node Cluster Operation....37
single disk (GDS)....608
single volume (GDS)....608
SIS....96,608
Site Preparation....14
Software Installation and Setup....45
Software Maintenance....348
Spanning Tree Protocol....10
spare disk (GDS)....609
Standby Operation....30
StandbyTransitions....212
Starting a Cluster Application....261
Starting RMS....260
Starting RMS Wizard....175
Starting the Web-Based Admin View screen....91
Startup Configuration for the IPMI Service....378
Startup Configuration for Update Service for SA....379
Startup test....7
state....609
state transition procedure....609
Stop....8
Stopping a Cluster Application....261
Stopping RMS....260
Storage Area Network....609
striped group (GDS)....609
striped virtual disk....609
striped volume (GDS)....609
stripe width (GDS)....609
striping (GDS)....609
Subsystem hang....218
SVmco (ServerView Mission Critical Option)....12
Switching a Cluster Application....262
switching mode....609
Switchlogs and application logs....259
switchover....7,609
switchover (RMS)....609
symmetrical switchover (RMS)....609
synchronized power control....610
System configuration modification....282
System Design....15
System dump....10
system graph (RMS)....610

[T]

template....610
Test....6
Test for forced shutdown of cluster nodes....9
Time synchronization....10
Troubleshooting....358
type....610

[U]

UP (CF)....610
user group....610
User groups....86

[V]

Viewing application logs....268
Viewing Detailed Resource Information....269
Viewing Logs Created by the PRIMECLUSTER System....268
Viewing switchlogs....268

Viewing the PRIMECLUSTER system operation management screens..... 249
virtual interface (VIP).....610
Virtual Machine Function.....16
volume (GDS).....610
Volume setup.....152

[W]

watchdog timer monitoring..... 610
Web-Based Admin View.....610
Web-Based Admin View screen.....92
When not Using the Virtual Machine Function.....45
When Using the Virtual Machine Function.....56
Wizard (RMS)..... 610
Work process continuity.....8