FUJITSU

Fujitsu Software
Technical Computing Suite V4.0L20

Job Operation Software Administrator's Guide for Maintenance

J2UL-2458-02ENZ0(12)
September 2024

Preface

Purpose of This Manual

This manual describes the operations during hardware and software maintenance in a system with Technical Computing Suite installed. It also describes how to change the system configuration and how to collect investigation materials when troubleshooting.

Intended Readers

This manual is intended for system administrators and cluster administrators who handle system maintenance and troubleshooting.
The manual assumes readers have the following knowledge:

- Basic Linux knowledge
- Overall understanding of the Job Operation Software, obtained from the "Job Operation Software Overview"
- Understanding of the installed system, obtained from procedures in the "Job Operation Software Setup Guide"
- Understanding of the system management function, obtained from the "Job Operation Software Administrator's Guide for System Management"
- Understanding of the job operation management function, obtained from the "Job Operation Software Administrator's Guide for Job Management"

Organization of This Manual

This manual is organized as follows.

Chapter 1 About Performing Maintenance Work

This chapter describes maintenance work performed by the administrator.

Chapter 2 Hardware Maintenance

This chapter describes types of hardware maintenance and their procedures.

Chapter 3 Software Maintenance

This chapter describes types of software maintenance and their procedures.

Chapter 4 Changing the System Configuration

This chapter describes procedures for changing the system configuration.

Chapter 5 Collecting Investigation Materials

This chapter describes how to collect investigation materials.

Chapter 6 Backing Up and Restoring a Node

This chapter describes procedures for backing up and restoring a node.

Chapter 7 Rebuilding a Node

This chapter describes cases requiring a node rebuild and their procedures.

Notation Used in This Manual

Representation of Units

The following table lists the prefixes representing units in this manual. Basically, disk size is represented as a power of 10, and memory size is represented as a power of 2. Be careful about specifying sizes when displaying or entering commands.

Prefix	Value	Prefix	Value
K (kilo)	10^3	Ki (kibi)	2^{10}
M (mega)	10^6	Mi (mebi)	2^{20}
G (giga)	10^9	Gi (gibi)	2^{30}
T (tera)	10^{12}	Ti (tebi)	2^{40}

Prefix	Value	Prefix	Value
P (peta)	10^{15}	Pi (pebi)	2^{50}

Notation of Model Names

In this manual, the computer that based on Fujitsu A64FX CPU is abbreviated as "FX server", and FUJITSU server PRIMERGY as "PRIMERGY server" (or simply "PRIMERGY").

Also, specifications of some of the functions described in the manual are different depending on the target model. In the description of such a function, the target model is represented by its abbreviation as follows:

[FX]: The description applies to FX servers.
[PG]: The description applies to PRIMERGY servers.

Administrators

The Job Operation Software has different types of administrator: system administrator, cluster administrator, and job operation administrator. Unless otherwise noted, the descriptions in this manual apply to functions for system administrators and cluster administrators. For this reason, the term "administrator" usually means a system administrator and a cluster administrator.

Prompts in Command Input Examples

means that the command must be executed with administrator privileges (superuser) for the OS.

Path Names of Commands

In the examples of operations, commands in the /bin, /usr/bin, /sbin, or /usr/sbin directory may not be represented by absolute paths.

Symbols in This Manual

This manual uses the following symbols.

Note

The Note symbol indicates an item requiring special care. Be sure to read these items.

See

The See symbol indicates the reference source of detailed information.

Information

The Information symbol indicates a reference note related to the Job Operation Software.

Trademarks

- Linux(R) is the registered trademark of Linus Torvalds in the U.S. and other countries.
- Red Hat and Red Hat Enterprise Linux are registered trademarks of Red Hat, Inc. in the U.S. and other countries.
- Other company and product names in this manual are trademarks or registered trademarks of their respective owners.

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.

Date of Publication and Version

Version	Manual Code
September 2024, Version 2.12	J2UL-2458-02ENZ0(12)
September 2023, Version 2.11	J2UL-2458-02ENZ0(11)
March 2023, Version 2.10	J2UL-2458-02ENZ0(10)
September 2022, Version 2.9	J2UL-2458-02ENZ0(09)
March 2022, Version 2.8	J2UL-2458-02ENZ0(08)
November 2021, Version 2.7	J2UL-2458-02ENZ0(07)
August 2021, Version 2.6	J2UL-2458-02ENZ0(06)
March 2021, Version 2.5	J2UL-2458-02ENZ0(05)
January 2021, Version 2.4	J2UL-2458-02ENZ0(04)
December 2020, Version 2.3	J2UL-2458-02ENZ0(03)
September 2020, Version 2.2	J2UL-2458-02ENZ0(02)
June 2020, Version 2.1	J2UL-2458-02ENZ0(01)
March 2020, Second version	J2UL-2458-02ENZ0(00)
January 2020, First version	J2UL-2458-01ENZ0(00)

Update history

Changes	Location	Version
Improved description of the arguments for the --failback option of the paclstmgr command.	2.3.5 3.5.4	2.12
Corrected errors.	-	
Improved recovery of PRIMERGY server compute nodes after dump collection.	5.2.2	2.11
Added instructions if node state is "Failed" after completing a backus or restore.	6.3.2 6.3.3	2.10
Added Infiniband and network device maintenance items in figure 2.1. Added a maintenance procedure for netwrok device.	2.1.1 2.1.5 2.5	2.9
Added a procedure to check the job memory free status (FX server only) when a node in a cluster enters software maintenance mode.	3.3.2.1	
Changed the description of initrd creation when updating packages.	3.6.2.4	
Changed the description of service stop on the target node.	3.6.5.2	
Corrected errors.	-	
Changed the procedure for OS Minor Version Update(For FX server, removed the procedure for removing and reapplying interconnect related packages).	3.6.5.2	2.8
Added a note for OS Minor Version Update.		
Improved the procedure for applying/updating/deleting/downgrading packages to FX servers.	3.6.1.3 3.6.1.5 3.6.2.3 3.6.2.5 3.6.3.1 3.6.3.3	2.7

Changes	Location	Version
	3.6.4.1 3.6.4.3 3.6.5.2	
Added instructions for maintaining both nodes of a redundant file server system at the same time.	3.3	2.6
Added instructions for applying an OS update package of RedHat to the target with an FX server.	3.6.2.3 3.6.2.5	
Added instructions to remove the FEFS setting before creating a disk image for FX server backups.	6.2.1.2 6.2.1.3	
Added instructions for updating the module definition files required to update RHEL 8 OS packages.	3.6.1.1	2.5
Added a procedure for reconfiguring job resource management functions in a cluster to the restore procedure.	6.2.1.3 6.3.3	
Added the information about using a system management node, a compute cluster management node, and a login node on one node.	3.3.2.2 3.5.3.1	2.4
The method of specifying the range when entering and releasing hardware maintenance mode has been changed.	2.2.2.1 2.2.2.3 2.2.4.1 2.2.4.3	2.3
Added caveats regarding the transition to hardware maintenance mode.	2.2.2.1 2.2.3.1 2.2.4.1	
Added items to check the power condition during maintenance.	2.2.2.1 2.2.3 2.2.4 6.2.1.2 6.2.1.3	2.2
Added a note for maintenance operations.	2.2.2.3 3.3.2	
Added a note for restarting maintaining the active node and standby node.	3.3 3.5.3.2	
Added instructions for OS Minor Version Update.	3.6.5 7.3.3.1	
Added a note for maintenance operations that result in no active global I/O nodes in the GIO group.	2.2.2.1 2.2.2.3 3.3.1.1 3.5.3.2	2.1
Changed FEFS service startup order after software maintenance.	3.5.2	
Added a note about the paclstmgr --recover option.	3.5.3.2	
Added notes for the kernel package updates, deletes and downgrades.	3.6.2.3 3.6.3.1 3.6.4.1	
Corrected errors.	-	
The communication path of a job can now be dynamically changed when a Tofu interconnect link goes down.	2.1.2.1	2
The look has changed according to product upgrades.	-	

Contents

Chapter 1 About Performing Maintenance Work

1.1 System Maintenance

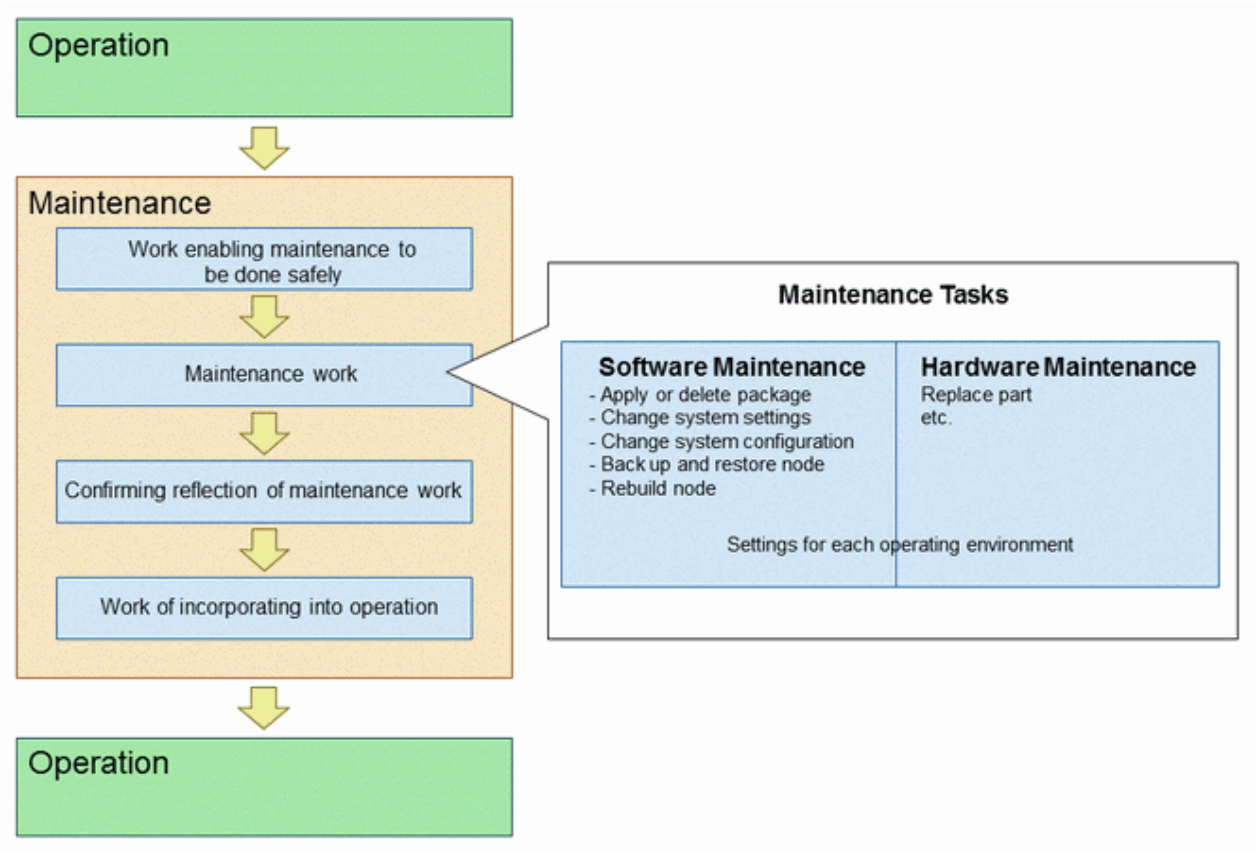
System maintenance refers to hardware maintenance, for replacing a part when a system disk failure or other hardware fault occurs, and software maintenance, such as for applying a fix package to a node.

The subsequent chapters describe procedures for safe and quick system maintenance by a system administrator.

1.2 Maintenance Workflow

Follow the overall flow of maintenance work shown below when performing maintenance on the whole system or part of the system.
If trouble occurs in system operation, see "Job Operation Software Troubleshooting" and isolate the problem. Then, perform maintenance as shown in the following flow.

Figure 1.1 System Maintenance Flow



- Work enabling maintenance to be done safely

  Isolate the maintenance target node from job operation and transition the node to maintenance mode so that it will not be allocated to any job.

  When isolating the maintenance target node from job operation, you need to also consider whether stopping the node will affect job operations on other nodes.

- Maintenance work

  Perform software or hardware maintenance.

  Software maintenance includes applying a package to a node, changing settings, and changing the system configuration. Hardware maintenance means replacing parts in nodes and BoBs.

- Work of confirming reflection of maintenance work

  After performing maintenance work, determine whether incorporating the target node into operation will lead to another problem.

  Also, confirm that the maintenance work does not have an adverse effect on other nodes in the system.

- Work of incorporating into operation

  After you have been able to confirm the maintenance work reflected in the node, incorporate the node into job operation again.

The subsequent chapters describe the work performed sequentially for the respective maintenance.

Chapter 2 Hardware Maintenance

The administrator performs maintenance work for failures that occur in system hardware. This chapter describes hardware maintenance with the Job Operation Software.

2.1 Hardware Maintenance Concept

This section describes the concept of hardware maintenance with the Job Operation Software.

2.1.1 Hardware Maintenance Categories

The following table shows the categories of hardware maintenance.
For the corresponding categories, see the sections listed under "Reference."

Table 2.1 Hardware Maintenance

Category	Description	Reference
FX server maintenance	This maintenance is performed when a part associated with the FX server fails or when the FX server firmware is updated.	2.1.2 FX server Maintenance
PRIMERGY server maintenance	This maintenance is performed when a part associated with the PRIMERGY server fails. This also applies to maintenance performed on an I/O disk unit used with the global file system.	2.1.3 PRIMERGY server Maintenance
InfiniBand maintenance	This maintenance is performed when the event of a component failure related to InfiniBand.	2.1.4 InfiniBand Maintenance
Network device maintenance	This maintenance is performed to replace the management network and the network devices (Network switches, LAN cables, etc.) that make up the control network, or to update the firmware.	2.1.5 Network Devices Maintenance

2.1.2 FX server Maintenance

With the FX server, the system monitoring function detects a hardware fault or software fault when any of the following parts fails:

- All nodes
  - CPU
  - Memory
  - ICC cable
  - AOC
- Boot I/O node
  - PCI card and cable
  - GbE cable
  - SSD (Sysvol)
- Global I/O node
  - IB-HCA card and cable
  - PCI card and cable
- Storage I/O node
  - PCI card and cable

    - SSD
- BoB unit
    - BMC (Baseboard Management Controller)
    - LAN cable (control network)
    - PSU
- Shelf
- Rack
- Other
    - Switch

Likewise, stoppages due to power exceeding the threshold and failures due to water leakage or condensation are also treated as hardware faults.

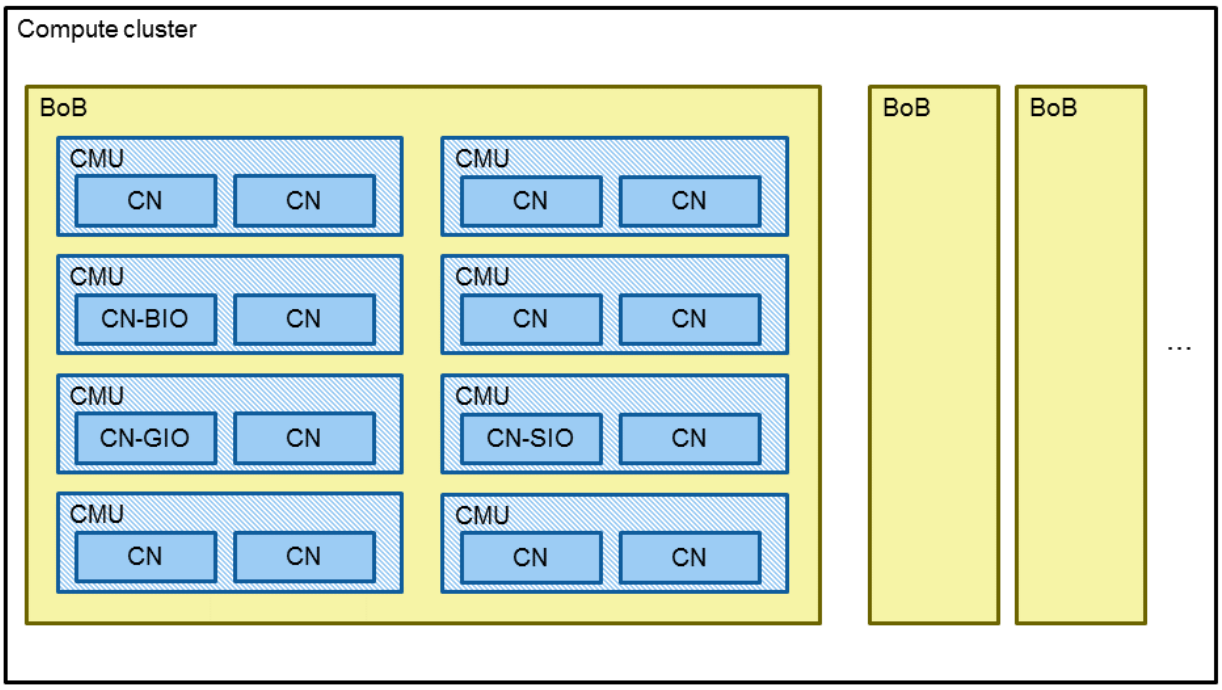
Information

If a shelf or rack fails, all nodes in that shelf and rack are treated as hardware faults.

FX server maintenance must take the hardware configuration into consideration.

The following figure shows the FX server hardware configuration.

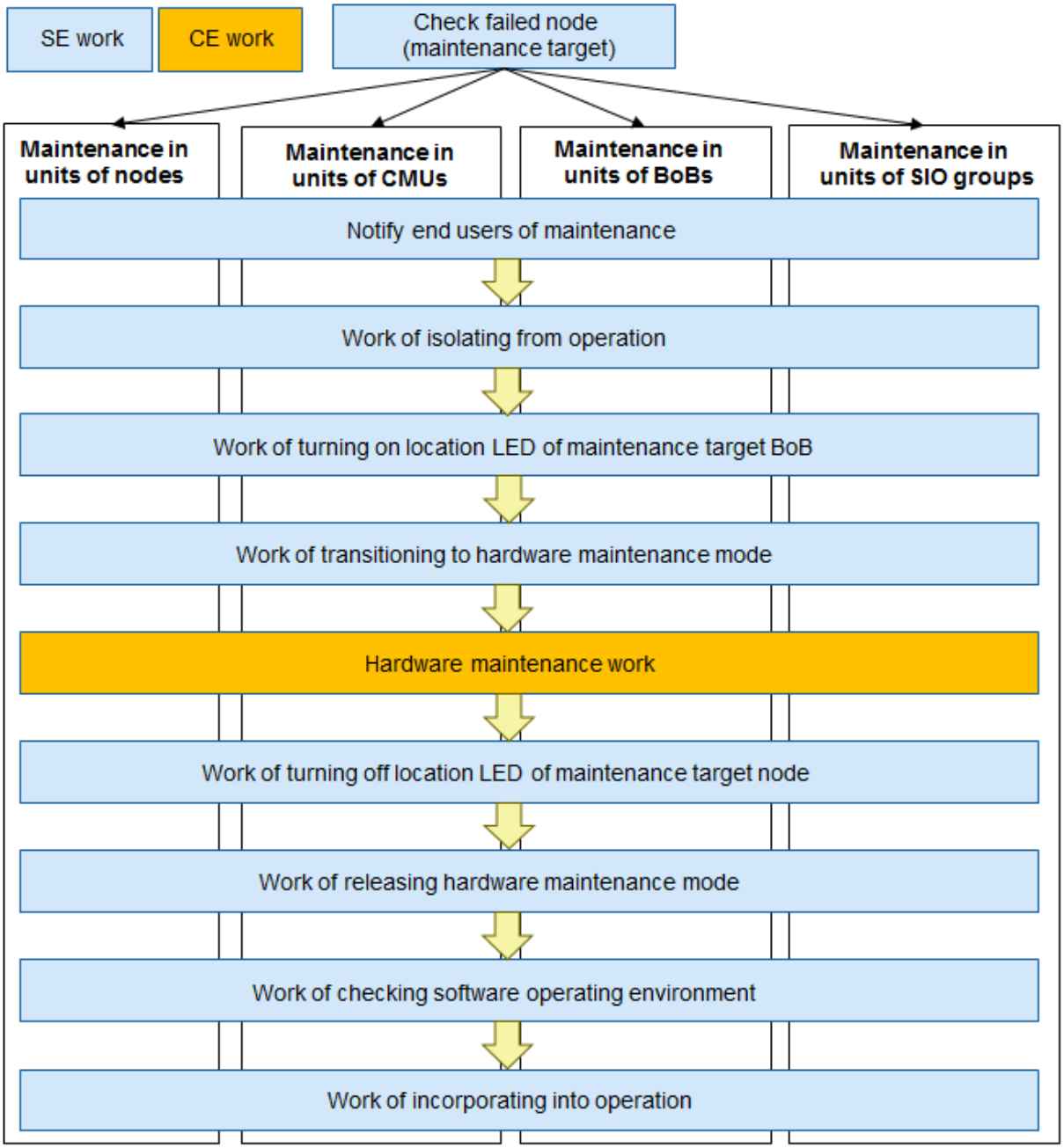
Figure 2.1 FX server Hardware Configuration



On the hardware configuration, the FX server consists of two nodes in each CMU. If a part fails, you will need to power off the failed node or the CMU before taking corrective action such as replacing hardware.

Therefore, for maintenance on a part that requires the CMU to be stopped, the minimum unit for maintenance targeting the part is two nodes. Also, depending on the node type in maintenance on a node, the maintenance target range changes to a range larger than a node or CMU, which also changes the effective range of maintenance.
The following conceptual image shows the hardware maintenance workflow and maintenance range for the FX server.

Figure 2.2 FX server Hardware Maintenance Workflow



Check a failed node as described in "2.1.2.1 Checking a Failed Node (Maintenance Target)." Perform maintenance by type of failed node and by part (in the four units shown above).

This section describes the concept of each part of the workflow.

2.1.2.1 Checking a Failed Node (Maintenance Target)

Use the pashowclst command to check the FX server status. If the REASON column shows one of the following reasons for a node, that node may possibly have failed.

Table 2.2 FX server Hardware Faults

REASON	Description
ResetRequest	The node must be restarted because memory on the node failed. The Job Operation Software waits for the end of a job running on that node and no longer assigns a new job to that node.
ReservedAlarm	A severe error occurred at the node, leading to performance degradation or a decrease in reliability. The Job Operation Software waits for the end of a job running on that node and no longer assigns a new job to that node.
RouterFatal	A fatal error occurred at the Tofu port. The Job Operation Software will kill jobs running on that node and no longer assign new jobs to that node.
PortRouterFatal	A fatal error occurred at the Tofu port. If the job running on the relevant node has an instruction to change the communication path when a Tofu interconnect link goes down, the Job Operation Software waits for the job to end. After the job ends, no new jobs are allocated to that node. If the job has an instruction not to change the communication path when a Tofu interconnect link goes down, the Job Operation Software forcibly terminates the job. After that, no new jobs are allocated to that node.
PortReservedAlarm	An error occurred at the Tofu port, leading to performance degradation or a decrease in reliability. The Job Operation Software waits for the end of a job running on that node and no longer assigns a new job to that node.
Alarm	A fatal error occurred at the node. The Job Operation Software will kill jobs running on that node and no longer assign new jobs to that node.

Also, depending on the part being checked (like the GbE cable of the boot I/O node), the pashowclst command may not display the above results in the REASON column but instead display "SoftError" in the STATUS column or "NodeDown" in the REASON column.

The minimum unit for maintenance targeting FX server nodes is one CMU (two nodes), so the following node types should be detected as maintenance targets. One is the node type of the failed node, and the other is the node type of another node in the same CMU as the failed node.

The following table shows the correspondence between node states and maintenance procedures for FX server node failures.

Check for the relevant node type and text displayed by the pashowclst command, and see the sections listed under "Maintenance to Perform" column.

Note

A hyphen (-) shown in "Map of FX server Maintenance Work" below does not mean the pashowclst command displays the hyphen. Instead, it means the text displayed there does not require your attention.

You do not need to be aware of the items represented by a hyphen, to identify the maintenance to be done.

BIO: Boot I/O node
GIO: Global I/O node
SIO: Storage I/O node
CN: Compute node

Table 2.3 Map of FX server Maintenance Work

Failure Range	Displayed by pashowclst				Maintenance to Perform
	STATUS	REASON	PWR_STATUS	ARCH_STATUS	
BIO	Disable	ResetRequest	-	ICC_Running	2.1.2.3 Maintenance in Units of BoBs
	Disable	ReservedAlarm	-	ICC_Running	
	Disable	RouterFatal	-	ICC_Disable	

Failure Range	Displayed by pashowclst				Maintenance to Perform
	STATUS	REASON	PWR_STATUS	ARCH_STATUS	
	Disable	PortRouterFatal	-	ICC_Disable	
	Disable	PortReservedAlarm	-	ICC_Disable	
	Disable	Alarm	-	ICC_Running	
	SoftError	NodeDown	-	ICC_Running	
GIO	Disable	ResetRequest	-	ICC_Running	2.1.2.2 Maintenance in Units of CMUs (*2)
	Disable	ReservedAlarm	-	ICC_Running	
	Disable	RouterFatal	-	ICC_Disable	
	Disable	PortRouterFatal	-	ICC_Disable	
	Disable	PortReservedAlarm	-	ICC_Disable	
	Disable	Alarm	-	ICC_Running	
SIO	Disable	ResetRequest	-	ICC_Running	2.1.2.4 Maintenance in Units of SIO Groups
	Disable	ReservedAlarm	-	ICC_Running	
	Disable	RouterFatal	-	ICC_Disable	
	Disable	PortRouterFatal	-	ICC_Disable	
	Disable	PortReservedAlarm	-	ICC_Disable	
	Disable	Alarm	-	ICC_Running	
CN	-	-	-	-	2.1.2.2 Maintenance in Units of CMUs
BoB (*1)	-	-	Unknown	-	2.1.2.3 Maintenance in Units of BoBs
	Disable	PortRouterFatal	-	ICC_Disable	
Firmware update	-	-	-	-	2.1.2.3 Maintenance in Units of BoBs

(*1) The table shows the states of all the nodes contained in the BoB.
(*2) If the pashowclst command shows "!" for the FEFSSR service in the SRV_STATUS column, perform the work in "2.1.4 InfiniBand Maintenance." (The InfiniBand maintenance range is "maintenance in units of nodes.")

Perform maintenance procedures according to the above maintenance work map. The next sections describe the concepts of the individual procedures.

2.1.2.2 Maintenance in Units of CMUs

Perform this maintenance work when hardware fails either in a compute node not also serving as an I/O node or in a global I/O node.

Maintenance in units of CMUs requires the two nodes in the target CMU to be isolated from operation and powered off.

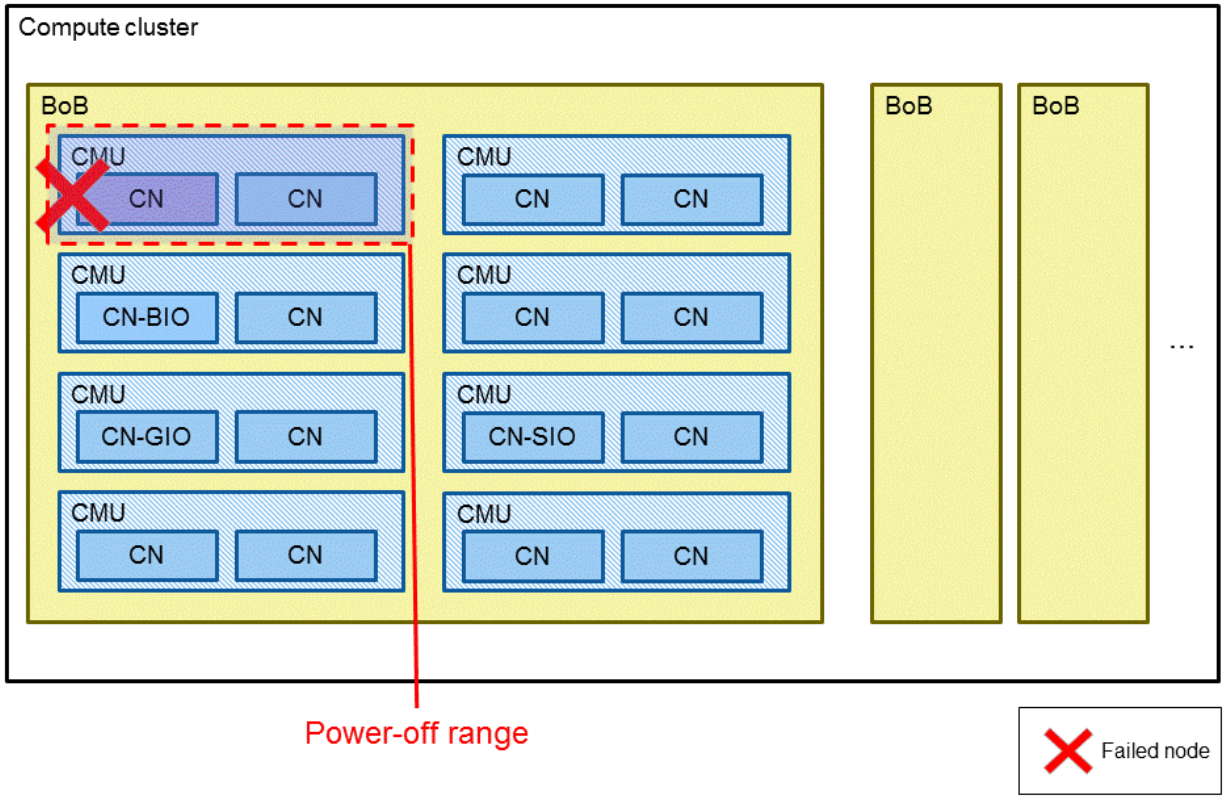
After being isolated from operation, the two target nodes are not allocated any more jobs, nor can any jobs using the interconnect for these nodes be executed. Also, depending on how jobs are allocated, the maintenance work may affect the jobs.

Note

Jobs may operate by using non-contiguous mode. Concurrent maintenance performed for multiple CMUs may cause those jobs to remain in the QUEUED state and not start execution, even if a sufficient number of nodes are available. This is because communication paths between nodes belonging to different Tofu units are blocked. For this reason, perform concurrent maintenance for only those CMUs that have the same B coordinate value, which means that communication paths between nodes belonging to different Tofu units will not be blocked.

The following is the range of power outages for maintenance on a CMU basis.

Figure 2.3 Maintenance in Units of CMUs

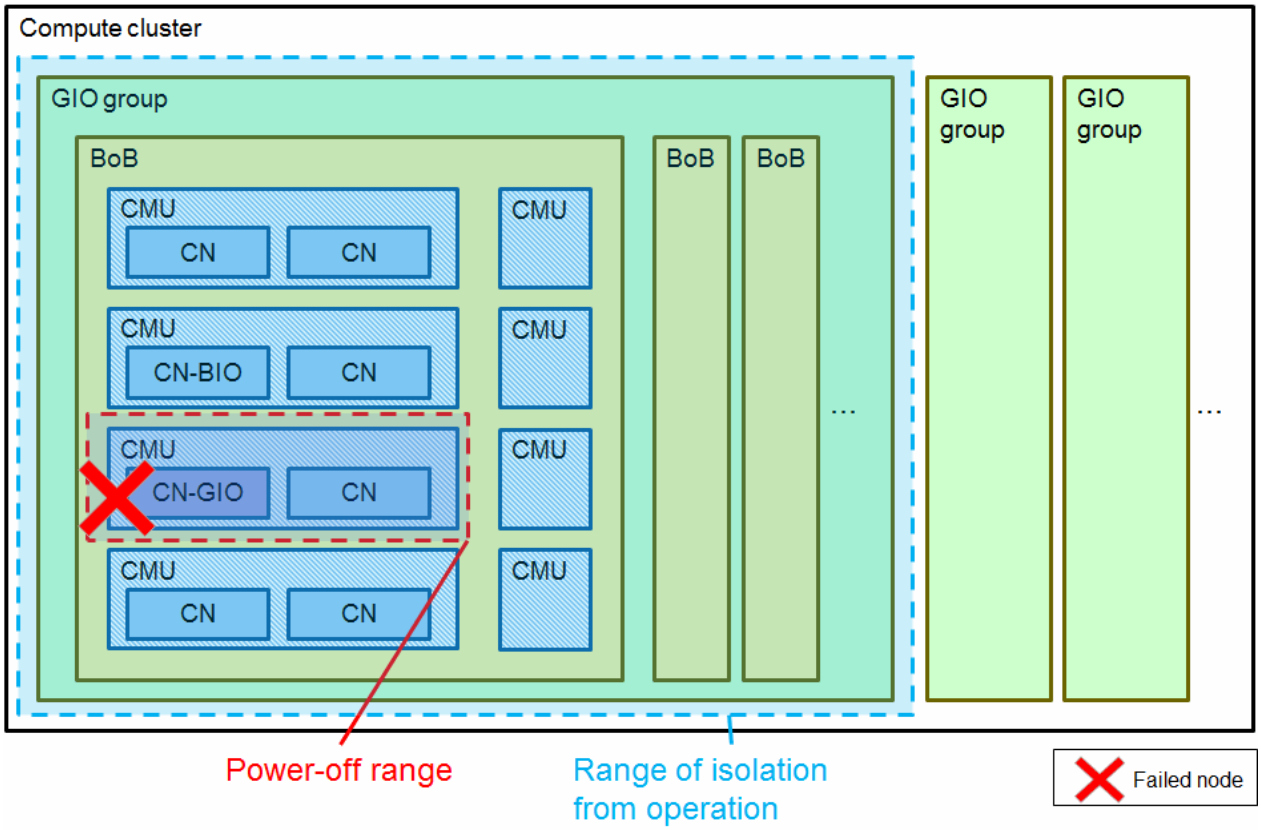


Note

- If the boot I/O node is inside the power-off range, the maintenance range is a BoB.
  For details on maintenance in units of BoBs, see "2.1.2.3 Maintenance in Units of BoBs."
- If a storage I/O node is inside the power-off range, the maintenance range is an SIO group.
  For details on maintenance in units of SIO groups, see "2.1.2.4 Maintenance in Units of SIO Groups."
- Suppose that a global I/O node is inside the power-off range and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, if the target global I/O node belongs to the GIO group, jobs can neither be allocated

to nor executed on nodes in the GIO group. For this reason, the global I/O node must be isolated from operation so that jobs are not allocated to nodes in the target GIO group.

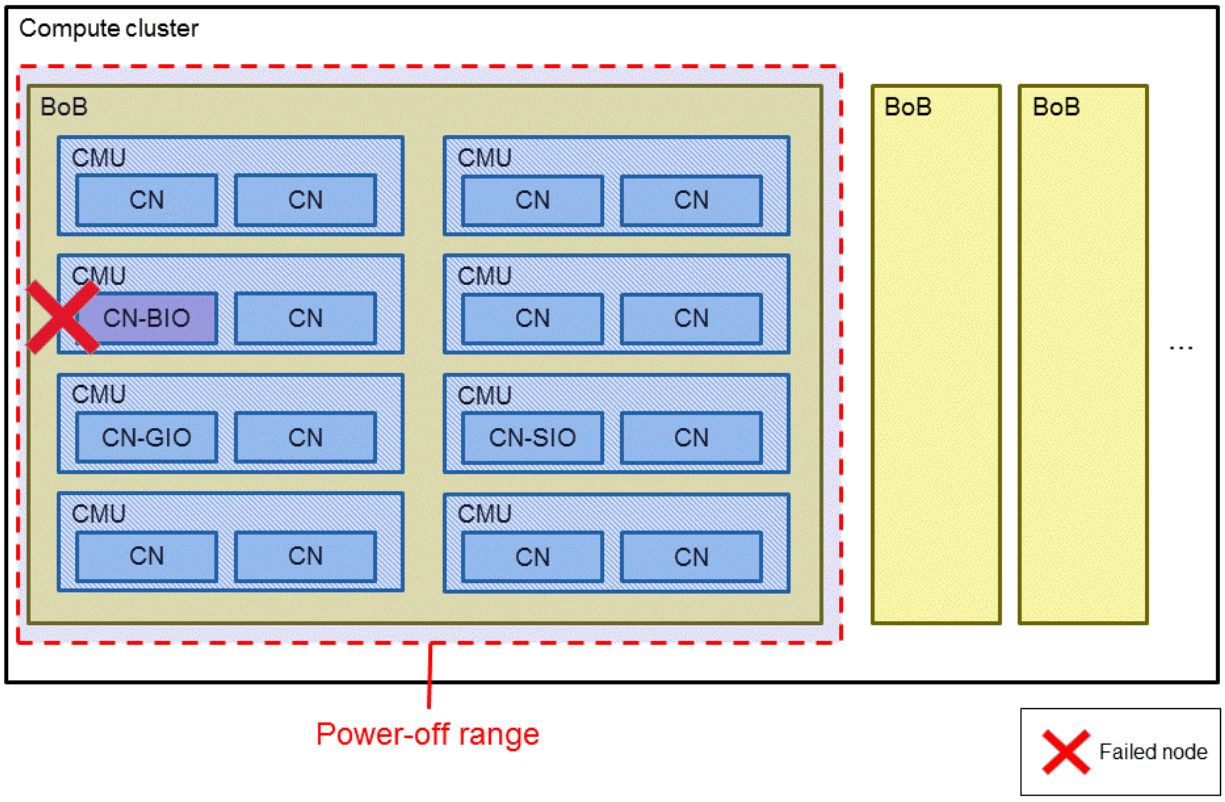
Figure 2.4 Maintenance in Units of CMUs Containing a Global I/O Node



2.1.2.3 Maintenance in Units of BoBs

Perform this maintenance work when replacing the Sysvol or when hardware fails either in the boot I/O node or in a node in the same CMU as the boot I/O node. Maintenance in units of BoBs requires the nodes belonging to the same BoB as the maintenance target node to be isolated and powered off in order to stop the boot I/O node.

Figure 2.5 Maintenance in Units of BoBs



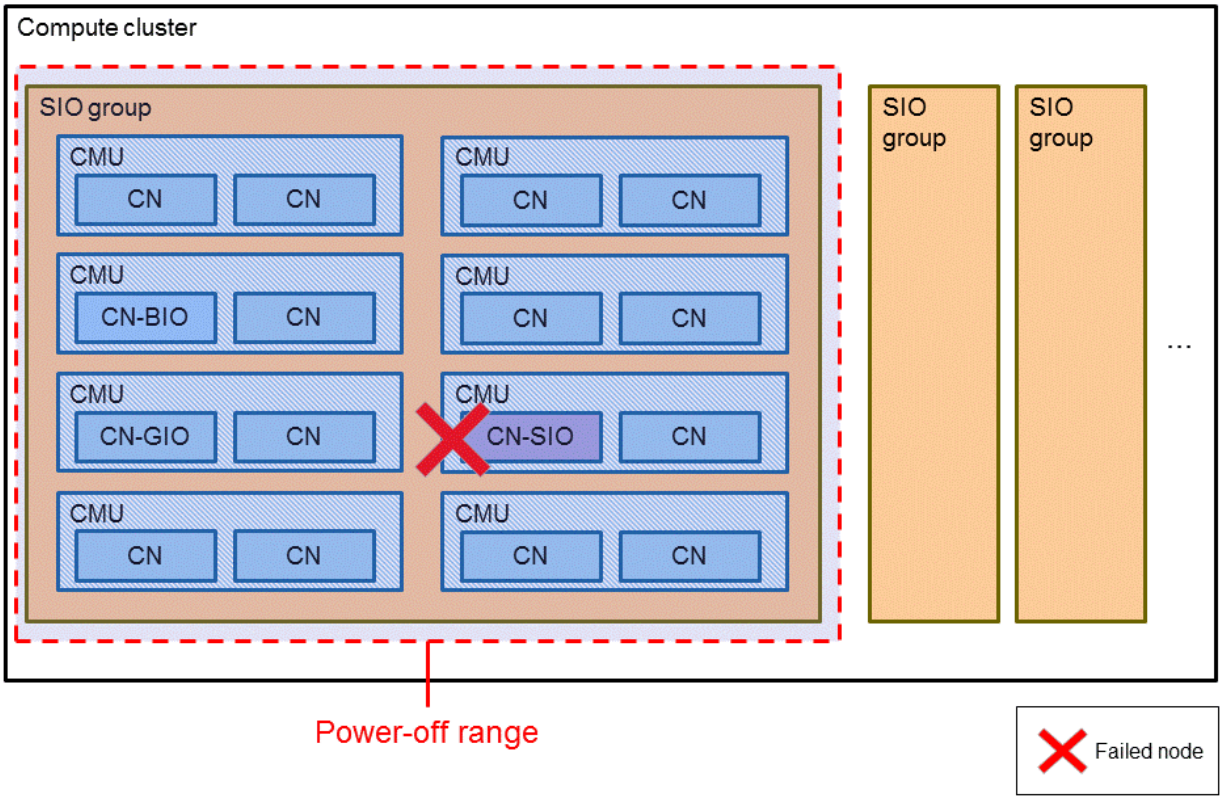
Note

- If a storage I/O node is inside the power-off range, the maintenance range is an SIO group.
  For details on maintenance in units of SIO groups, see "2.1.2.4 Maintenance in Units of SIO Groups."
- Suppose that a global I/O node is inside the power-off range and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, if the target global I/O node belongs to the GIO group, jobs can neither be allocated to nor executed on nodes in the GIO group. For this reason, the global I/O node must be isolated from operation so that jobs are not allocated to nodes in the target GIO group.

2.1.2.4 Maintenance in Units of SIO Groups

Perform this maintenance work when hardware fails either in a storage I/O node or in a node belonging to the same CMU as the storage I/O node.
Maintenance in units of SIO groups requires the nodes belonging to the same SIO group as the maintenance target node to be isolated and powered off in order to stop the storage I/O node.

Figure 2.6 Maintenance in Units of SIO Groups



Note

Suppose that a global I/O node is inside the power-off range and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, if the target global I/O node belongs to the GIO group, jobs can neither be allocated to nor executed on nodes in the GIO group. For this reason, the global I/O node must be isolated from operation so that jobs are not allocated to nodes in the target GIO group.

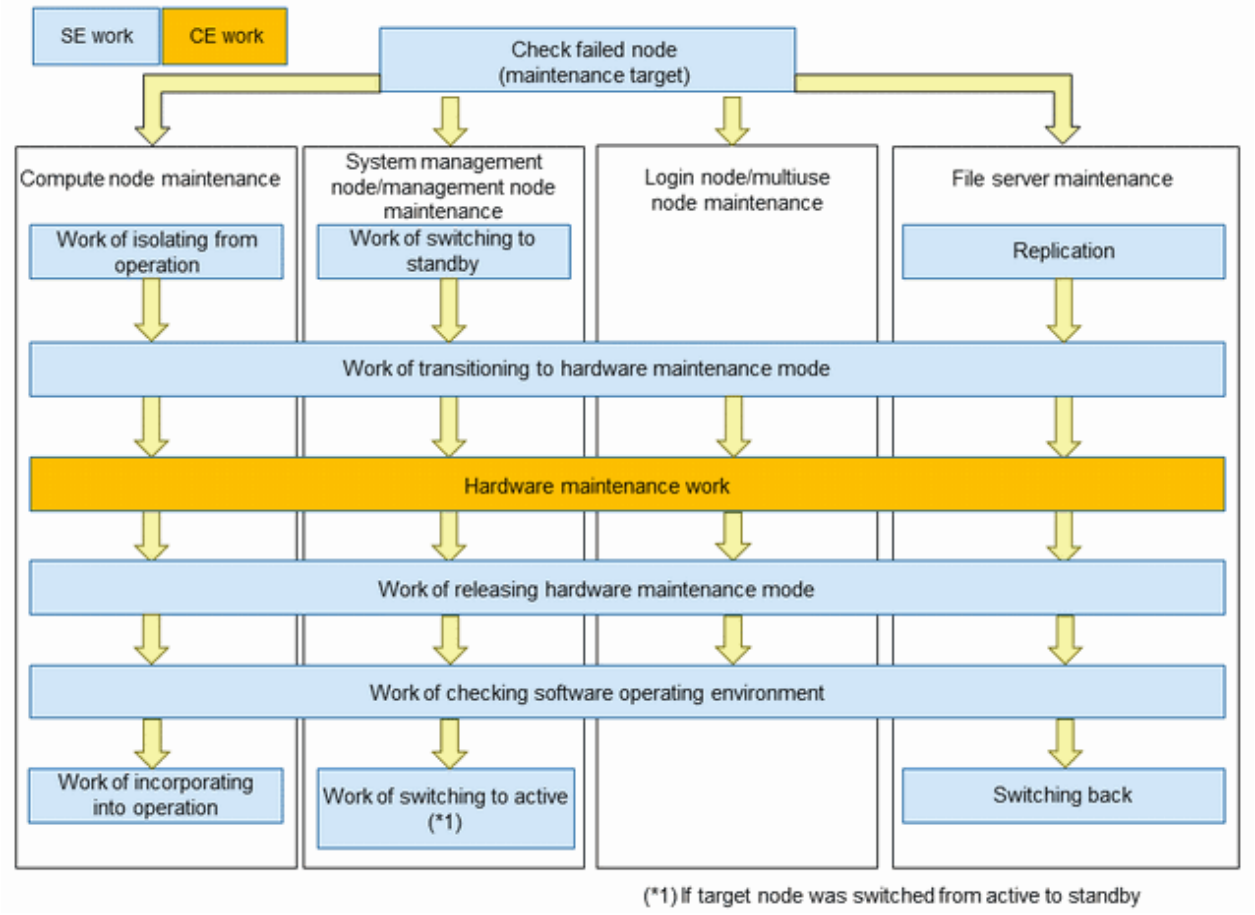
2.1.3 PRIMERGY server Maintenance

With the PRIMERGY server, the system monitoring function detects a software fault when any of the following parts fails:

- LAN
- CPU
- InfiniBand (IB)
- Storage unit (ETERNUS, etc.)
- Memory
- Disk
- Other (switch, power supply unit, etc.)

The following figure shows the PRIMERGY server hardware maintenance workflow.

Figure 2.7 PRIMERGY server Hardware Maintenance Workflow



This section describes the concept of each part of the workflow.

2.1.3.1 Checking a Failed Node (Maintenance Target)

Use the pashowclst command to check the PRIMERGY server status. If the STATUS column shows "SoftError" or REASON column shows "NodeDown" or "SrfDown" for a node, a hardware fault may possibly have occurred on that node.

The minimum unit for maintenance targeting PRIMERGY server is one node, so the node type that should be detected as the maintenance target is the node type of the failed node.

The following table shows the correspondence between node states and maintenance procedures for PRIMERGY server failures.

Table 2.4 Map of PRIMERGY server Maintenance Work

Maintenance Target Node	Maintenance Procedure to Perform
System management node (*1)	2.1.3.3 System Management Node and Management Node Maintenance
Storage cluster management node	2.1.3.3 System Management Node and Management Node Maintenance
MGS node (*2)	2.1.3.5 File Server Node Maintenance
MDS node	2.1.3.5 File Server Node Maintenance
OSS node	2.1.3.5 File Server Node Maintenance
Compute cluster management node	2.1.3.3 System Management Node and Management Node Maintenance
Compute cluster sub management node	2.1.3.3 System Management Node and Management Node Maintenance
Login node	2.1.3.4 Login Node and Multiuse Node Maintenance
Compute node	2.1.3.2 Compute Node Maintenance
Multiuse node	2.1.3.4 Login Node and Multiuse Node Maintenance

Maintenance Target Node	Maintenance Procedure to Perform
File system	2.1.3.5 File Server Node Maintenance

(*1) Includes a system management node that also serves as a storage cluster node.
(*2) Includes an MGS node that also serves as an MDS node and OSS node, and an MGS node that also serves as an MDS node.

See

For each node with either "!" or "x" shown for the IC service, perform the work in "2.1.4 InfiniBand Maintenance."

If "!" or "x" is shown for the status of the FEFS service or FEFSSR service, an I/O unit in the global file system or InfiniBand may possibly have failed. Perform the work in "2.1.3.5 File Server Node Maintenance" for I/O unit maintenance in the global file system or "2.1.4 InfiniBand Maintenance" for InfiniBand maintenance.

The next sections describe PRIMERGY server maintenance by node type.

2.1.3.2 Compute Node Maintenance

For compute node maintenance, isolate the maintenance target node from operation and power it off. After that, perform the maintenance.

See

For the maintenance procedures for PRIMERGY server compute nodes, see "2.3.1 Operations During Compute Node Maintenance."

2.1.3.3 System Management Node and Management Node Maintenance

The system management node and management nodes (compute cluster management node, compute cluster sub management node, and storage cluster management node) can be configured redundantly for high availability.
Before performing hardware maintenance on the active node in such configurations, perform failover to the standby node. This allows job operations to continue.

Note

- Due to a failure or maintenance work, you may be stopping both nodes configured redundantly as a pair. Or you may encounter a configuration where nodes do not have redundancy. In either case, when maintaining the target node, the affected node must be stopped.
  - In the case of the system management node, the effective range is a system.
  - In the case of the compute cluster management node, the effective range is a cluster.
  - In the case of the compute cluster sub management node, the effective range is a node group.
  - In the case of the storage cluster management node, the effective range is a cluster.
- When stopping operations in a storage cluster, you need to also stop operations in the compute cluster and multiuse cluster that use the storage cluster.

See

For the maintenance procedures for the system management node, see "2.3.2 Operations During System Management Node Maintenance ."
For the maintenance procedures for management nodes, see "2.3.3 Operations During Management Node Maintenance."

2.1.3.4 Login Node and Multiuse Node Maintenance

For maintenance on the login node or a multiuse node, power off the target node. After that, perform the maintenance.

If you stop the login node without another login node on hand, the resulting effect on job operations includes end users not being able to submit jobs.

For the maintenance procedures for the login node, see "2.3.4 Operations During Login Node Maintenance." For the maintenance procedures for multiuse nodes, see "2.3.6 Operations During Multiuse Node Maintenance."

2.1.3.5 File Server Node Maintenance

The services and daemons for the MGS, MDS, and OSS nodes are in a mutual standby configuration.

Perform replication to the node paired with the maintenance target node in this configuration before performing maintenance work. This allows job operations to continue.

Furthermore, to perform maintenance on any ETERNUS used in the global file system, replication must be done for the connected node. Since the entire system shares the global file system targeted for maintenance, it is necessary to stop use of the file system across the whole system.

For the maintenance procedures for file server nodes, see "2.3.5 Operations During File Server Node Maintenance."

2.1.4 InfiniBand Maintenance

Perform InfiniBand (called "IB" below) maintenance when replacing the HCA, IB cable, or IB switch.

Using commands and also the system log messages output when a hardware fault occurs, check for the cause of the hardware fault, and perform maintenance work.

For the maintenance procedures for replacing the HCA, IB cable, or IB switch, see "2.4 Operations During InfiniBand Maintenance."

2.1.5 Network Devices Maintenance

Perform network devices maintenance when replacing the network switches, routers, or LAN cables.

During network device maintenance, communication using the network to be maintained will not be possible. Perform maintenance on the affected node in advance.

For the maintenance procedures for replacing the network device, see "2.5 Operations During Network Device Maintenance."

2.2 Operations During FX server Maintenance

This section describes the operations to be performed during maintenance when the FX server fails.

2.2.1 Checking a Failed Node

This section describes procedures for detecting a failed node and determining the extent of maintenance required in the system.

Perform this work on the active system management node.

1. Checking a failed node

   Check the run status of nodes with the pashowclst command.
   The following example checks the run status of nodes whose boot group ID is 0xFFF01 in the cluster named "compute."

```
# pashowclst -c compute --bootgrp 0xFF01 -v
[ CLST: compute ]
[ BOOTGRP: 0xFF01 ]
NODE         NODETYPE  STATUS   REASON         PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Running  -              os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002  SIO,CN    Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003  GIO,CN    Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010007  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010008  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010009  CN        Disable  ReservedAlarm  os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000A  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000B  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000C  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000D  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000E  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000F  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010010  CN        Running  -              os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

   The above example shows a failed node with node ID 0xFF010009. Perform one of the following procedures according to the node type of the failed node:

   - Boot I/O node: "2.2.3 Maintenance in Units of BoBs"
   - Storage I/O node: "2.2.4 Maintenance in Units of SIO Groups"
   - Other than above: Perform step 2.

2. Checking nodes in the same CMU

   If the failed node checked in step 1 is a global I/O node or compute node, check the nodes in the same CMU.

```
# pashowclst -c compute --cmu 0xFF010009 -v
[ CLST: compute]
[ CMU: 0xFF010009]
NODE         NODETYPE  STATUS     REASON          PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Running    -               os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010009   CN        Disable    ReservedAlarm   os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

In the above example, the node with node ID 0xFF010001 is a node in the same CMU as the node detected in step 1.
Perform the following procedures according to the type of node in the same CMU as the failed node:

- Boot I/O node: "2.2.3 Maintenance in Units of BoBs"
- Storage I/O node: "2.2.4 Maintenance in Units of SIO Groups"
- Global I/O node or compute node: "2.2.2 Maintenance in Units of CMUs"

2.2.2 Maintenance in Units of CMUs

To perform maintenance for the FX server in units of CMUs, perform the procedures below.

Perform this work on the active system management node.

2.2.2.1 Isolating From Operation and Powering Off

Isolate the nodes in the targeted CMU from operation, transition them to hardware maintenance mode, and power them off.
The following three methods are ways toward that end:

- By waiting for running jobs to end
- By using another node to re-execute running jobs
- By setting a deadline schedule

Perform the steps in one of the above methods before isolating the nodes from operation and transitioning them to hardware maintenance mode.

[By waiting for running jobs to end]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

   Note

   Suppose that the maintenance target CMU contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that nodes in the GIO group should be isolated from job operation.

   Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE        NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1  GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2  GIO,CN    Stopped  -       off         ICC_Running  -
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node
   gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
   gionodeid2: Node ID of a global I/O node in the GIO group

2. Isolating the target CMU from operation

   Using the --disable option in the paclstmgr command, isolate the maintenance target CMU from operation (transition to the Disable state). At this time, also specify the --ic option. As a result, processing waits for the end of not only running jobs on the target nodes but also jobs using the interconnect built on the target nodes. If operation does not permit submission of jobs in non-contiguous mode, you can omit specifying the --ic option.

```
# paclstmgr -c clstname --disable --ic --cmu nodeid
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

Note

Suppose that the maintenance target CMU contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the nodes in the GIO group, not the CMU, must be isolated from operation.

To isolate the nodes in the GIO group from operation, execute the following on the command line.

```
# paclstmgr -c clstname --disable --cmu nodeid --giogrp --ic
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

3. Confirming isolation from operation

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "Manual" in the REASON column, and "ICC_Disable" in the ARCH_STATUS column by the nodes isolated in step 2. If you omitted the --ic option in step 2, the ARCH_STATUS column should show "ICC_Running" for the nodes.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON          PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005   CN        Disable    Manual          os-running    ICC_Disable
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000D   CN        Disable    Manual          os-running    ICC_Disable
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

4. Turning on the location LED of the target BoB

Note

Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.

To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

5. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the target CMU to hardware maintenance mode.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

Note

If the maintenance target nodes contain a global I/O node, and the maintenance operation causes a global I/O node that is running to no longer exist in the GIO group, transit the node in the GIO group to hardware maintenance mode.

Determine the boot group range for the target GIO group.
The boot group range is determined from the node ID of the boot I/O node as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype BIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: BIO ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Disable   Manual    os-running   ICC_Disable
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
...
0xFF180001  BIO,CN    Disable   Manual    os-running   ICC_Disable
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
```

clstname: Cluster name
nodeid: Node ID of the node to be maintained

When you run the paclstmgr command, you specify a range with the --bootgrp option.
In the above example, run the command line as follows.

```
# paclstmgr -c clstname --bootgrp 0xFF01-0xFF18 --hard-mainte
```

clstname: Cluster name

Note

Depending on the hardware failure, the target node may not be able to stop normally.
Use the force-off option of the papwrctl command.

6. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "off" in the REASON column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005   CN        Disable    HardMaintenance   off           ICC_Stopped    -
0xFF01000D   CN        Disable    HardMaintenance   off           ICC_Stopped    -
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

7. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid all-status
NODE       PWR_STATUS     ICC CMU BoB
```

```
0xFF010005 off            off off on
0xFF01000D off            off off on
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

[By using another node to re-execute running jobs]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

   Note

   Suppose that the maintenance target CMU contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that nodes in the GIO group should be isolated from job operation.

   Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE        NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1  GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2  GIO,CN    Stopped  -       off         ICC_Running  -
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node
   gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
   gionodeid2: Node ID of a global I/O node in the GIO group

2. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the range for maintenance work to hardware maintenance mode. If you also specify the -k option at this time, the command immediately suspends running jobs and jobs using the interconnect, and re-executes them on another node.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte -k
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

   Note

   - Suppose that the maintenance target CMU contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the nodes in the GIO group are automatically isolated from operation when the paclstmgr command with the -hard-mainte and -k options transitions the maintenance targets to hardware maintenance mode. (The nodes that are not maintenance targets can use the interconnect and are not powered off.)
   - Depending on the hardware failure, the target node may not be able to stop normally.
     Use the force-off option of the papwrctl command.

3. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "off" in the REASON column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
```

```
[ CMU: nodeid ]
NODE          NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005    CN        Disable    HardMaintenance   off           ICC_Stopped    -
0xFF01000D    CN        Disable    HardMaintenance   off           ICC_Stopped    -
```

4. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010005 off            off off on
0xFF01000D off            off off on
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

5. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

   clstname: Cluster name
   bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

[By setting a deadline schedule]

1. Notifying the end users of maintenance

   Notify the end users about setting a deadline schedule on the target nodes for maintenance.

Note

Suppose that the maintenance target CMU contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that a deadline schedule should be set for the nodes in the GIO group.

Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE         NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1   GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2   GIO,CN    Stopped  -       off         ICC_Running  -
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node
gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
gionodeid2: Node ID of a global I/O node in the GIO group

2. Setting a deadline schedule

   Set a deadline schedule on the nodes in the range for maintenance work.
   For details on how to set a deadline schedule, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

3. Checking the status of the deadline schedule

   Using the pashowclst command, confirm state transitions to "DeadlineSchedule" in the REASON column by the CMU nodes in the range for maintenance work.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005   CN        Running   DeadlineSchedule  os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000D   CN        Running   DeadlineSchedule  os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

   Subsequently, the nodes will enter the period of the set deadline schedule. Then, confirm state transitions to "Disable" in the STATUS column, "DeadlineSchedule" in the REASON column, and "ICC_Disable" in the ARCH_STATUS column by the compute nodes in the range for maintenance work.

4. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

   clstname: Cluster name
   bootgid: Boot group ID

   See

   For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

5. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the target CMU to hardware maintenance mode.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

6. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "off" in the REASON column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005   CN        Disable    HardMaintenance   off           ICC_Stopped    -
0xFF01000D   CN        Disable    HardMaintenance   off           ICC_Stopped    -
```

7. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010005 off            off off on
0xFF01000D off            off off on
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

2.2.2.2 Hardware Maintenance Work

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

Note

Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.
Perform the procedure "2.2.2.3 Incorporating Into Operation" to restart the nodes.

2.2.2.3 Incorporating Into Operation

After completing maintenance work on a CMU, incorporate the nodes in the CMU into operation.

1. Turning off the location LED of the target BoB

   If you turned on the location LED of the target BoB in "2.2.2.1 Isolating From Operation and Powering Off," turn them off with the paledctl command.

```
# paledctl -c clstname --bootgrp bootgid off
```

   clstname: Cluster name
   bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

2. Releasing hardware maintenance mode

   Release hardware maintenance mode on the nodes in the target CMU, and power them on.

   When releasing hardware maintenance mode, leave the target nodes isolated from operation (Disable state). In step 3 below, you will be checking whether the nodes in the target CMU can be incorporated into operation.

   Specify the --disable and --ic options in addition to the --recover option in the paclstmgr command to release hardware maintenance mode.

   At this time, the command also simultaneously powers on the nodes in the target CMU.

```
# paclstmgr -c clstname --cmu nodeid --recover --disable --ic
```

Note

- If you performed maintenance in units of GIO group, specify a range in the --bootgrp option.

```
# paclstmgr -c clstname --bootgrp 0xFF01-0xFF18 --recover --disable --ic
```

clstname: Cluster name

- If hardware maintenance in units of CMU, be sure to recovery in units of CMU. When you perform a recovery in units of BoB or node, the Tofu port does not link up and does not start normally.

3. Checking the status after releasing hardware maintenance mode

   Using the pashowclst command, confirm that the STATUS column shows "Disable," REASON column shows "Manual," and ARCH_STATUS column shows "ICC_Disable" for the target node.

   Services have started by this time.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON           PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010005   CN        Disable    Manual           os-running    ICC_Disable
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000D   CN        Disable    Manual           os-running    ICC_Disable
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

Information

To release hardware maintenance mode while incorporating the target nodes into operation at the same time, specify only the --recover option in the paclstmgr command. After executing the command, proceed to step 6.

4. Checking the software environment

   From the software perspective, check whether the started nodes in the maintenance target CMU are ready to be incorporated into operation.
   Execute the pachecker command to check the environment.
   For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."

5. Incorporating into operation

   Using the --enable option in the paclstmgr command, incorporate the nodes in the maintenance target CMU into operation.

```
# paclstmgr -c clstname --cmu nodeid --enable
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

Note

If you performed maintenance in units of GIO groups, execute the following command line:

```
# paclstmgr -c clstname --enable --cmu nodeid --giogrp
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

If you set a deadline schedule in "[By setting a deadline schedule]" in "Isolating From Operation and Powering Off," use the padeadline command to cancel the setting.
For details on how to cancel the setting, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

6. Checking the node status

   Using the pashowclst command, confirm state transitions to "Running" in the STATUS column by the nodes in the target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010105   CN        Running    -                 os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF01000D   CN        Running    -                 os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

When all the services on a target node are running, the STATUS column shows "Running" for that node.

2.2.3 Maintenance in Units of BoBs

Each boot I/O node is mounted to a BoB. To perform maintenance for the CMU of a boot I/O node, isolate the relevant BoB from operation.
To perform maintenance for the FX server in units of BoBs, perform the procedures below.
Perform this work on the active system management node.

2.2.3.1 Isolating From Operation and Powering Off

Isolate the maintenance target BoB from operation, transition it to hardware maintenance mode, and power it off.
The following three methods are ways toward that end:

- By waiting for running jobs to end
- By using another node to re-execute running jobs
- By setting a deadline schedule

Perform the steps in one of the above methods before isolating the BoB from operation and transitioning it to hardware maintenance mode.

[By waiting for running jobs to end]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

   Note

   Suppose that the maintenance target BoB contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that nodes in the GIO group should be isolated from job operation.

   Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE         NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1   GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2   GIO,CN    Stopped  -       off         ICC_Running  -
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node
   gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
   gionodeid2: Node ID of a global I/O node in the GIO group

2. Isolating the target BoB from operation

   Maintenance in units of BoBs requires the maintenance target BoB to be powered off. For this reason, isolate the nodes in the BoB from operation (transition to the Disable state) by using the --disable option in the paclstmgr command.

```
# paclstmgr -c clstname --disable --bootgrp bootgid
```

   clstname: Cluster name
   bootgid: Boot group ID

Note

Suppose that the maintenance target BoB contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the nodes in the GIO group, not the BoB, must be isolated from operation.

To isolate the nodes in the GIO group from operation, execute the following on the command line.

```
# paclstmgr -c clstname --disable -n nodeid --giogrp
```

clstname: Cluster name
nodeid: Node ID of the maintenance target

3. Confirming isolation from operation

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "Manual" in the REASON column by the nodes isolated in step 2.

```
# pashowclst -c clstname --bootgrp bootgid -v
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002  SIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003  GIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   bootgid: Boot group ID

4. Isolating the target CMU from operation

   Perform step 2 in "[By waiting for running jobs to end]" in "2.2.2.1 Isolating From Operation and Powering Off" in "2.2.2 Maintenance in Units of CMUs" to isolate the nodes in the maintenance target CMU from operation (transition to the Disable state).

   Using the pashowclst command, confirm state transitions to "ICC_Disable" in the ARCH_STATUS column by the nodes in the maintenance target CMU.

```
# pashowclst -c clstname --bootgrp bootgid -v
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Disable   Manual    os-running   ICC_Disable
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
```

```
0xFF010002  SIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003  GIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

clstname: Cluster name
bootgid: Boot group ID

5. Turning on the location LED of the target BoB

Note

Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.

To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

6. Transitioning to hardware maintenance mode

Using the pashowclst command, confirm state transitions by the isolated nodes to "Disable" in the STATUS column. Then, transition the CMU of the target nodes to hardware maintenance mode by using the --hard-mainte option in the paclstmgr command.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte
```

clstname: Cluster name
nodeid: Node ID of the maintenance target

Note

- This operation handles the power-off processing of the maintenance target BoB.
- Depending on the hardware failure, the target node may not be able to stop normally.
  Use the force-off option of the papwrctl command.

7. Confirming hardware maintenance mode

Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the node specified in step 6.
Also confirm that the PWR_STATUS column shows "off" for the nodes isolated in step 2.

```
# pashowclst -c clstname --bootgrp bootgid -v
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010002   SIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010003   GIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010004   CN        Disable    Manual            off           ICC_Running    -
0xFF010005   CN        Disable    Manual            off           ICC_Running    -
0xFF010006   CN        Disable    Manual            off           ICC_Running    -
...
```

clstname: Cluster name
bootgid: Boot group ID

8. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --bootgrp bootgid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            off off on
0xFF010002 off            on  on  on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

clstname: Cluster name
bootgid: Boot group ID

[By using another node to re-execute running jobs]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

Suppose that the maintenance target BoB contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that nodes in the GIO group should be isolated from job operation.
Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE         NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1   GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2   GIO,CN    Stopped  -       off         ICC_Running  -
```

nodeid: Node ID of the maintenance target node
gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
gionodeid2: Node ID of a global I/O node in the GIO group

2. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the range for maintenance work to hardware maintenance mode. If you also specify the -k option at this time, the command immediately suspends running jobs and jobs using the interconnect, and re-executes them on another node.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte -k
```

clstname: Cluster name
nodeid: Node ID

- Suppose that the maintenance target BoB contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. The nodes in the GIO group are automatically isolated from operation when the paclstmgr command with the -hard-mainte and -k options transitions the maintenance target nodes to hardware maintenance mode. (The nodes that are not maintenance targets can use the interconnect. Also, excluding the BoB of the maintenance target nodes, BoBs are not powered off.)
- Depending on the hardware failure, the target node may not be able to stop normally.
  Use the force-off option of the papwrctl command.

3. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the node specified in step 2.
   Also confirm that the PWR_STATUS column shows "off" for the nodes in the same BoB as the node specified in step 2.

   The following example assumes the specified node ID in step 2 is 0xFF010001.

```
# pashowclst -c clstname --bootgrp 0xFF01 -v
[ CLST: clstname ]
[ BOOTGRP: 0xFF01 ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010002   SIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010003   GIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010004   CN        Disable    Manual            off           ICC_Running    -
0xFF010005   CN        Disable    Manual            off           ICC_Running    -
0xFF010006   CN        Disable    Manual            off           ICC_Running    -
...
```

   clstname: Cluster name
   bootgid: Boot group ID

4. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --bootgrp bootgid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            off off on
0xFF010002 off            on  on  on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

   clstname: Cluster name
   bootgid: Boot group ID

5. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

[By setting a deadline schedule]

1. Notifying the end users of maintenance

   Notify the end users about setting a deadline schedule on the target nodes for maintenance.

Note

Suppose that the maintenance target BoB contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that a deadline schedule should be set for the nodes in the GIO group.

Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE        NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1  GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2  GIO,CN    Stopped  -       off         ICC_Running  -
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node
gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
gionodeid2: Node ID of a global I/O node in the GIO group

2. Setting a deadline schedule

   Set a deadline schedule on the nodes in the range for maintenance work.
   For details on how to set a deadline schedule, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

3. Checking the status of the deadline schedule

   Using the pashowclst command, confirm state transitions to "DeadlineSchedule" in the REASON column by the BoB nodes in the range for maintenance work.

```
# pashowclst -c clstname --bootgrp bootgid -v
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE       NODETYPE STATUS    REASON            PWR_STATUS   ARCH_STATUS  SRV_STATUS
0xFF010001 BIO,CN   Running   DeadlineSchedule  os-running   ICC_Running  (*)
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002 SIO,CN   Running   DeadlineSchedule  os-running   ICC_Running  (*)
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
...
```

clstname: Cluster name
bootgid: Boot group ID

(*) The places marked with "(*)" show line feeds inserted due to space restrictions. Actually, the separated output appears on a single line.

Subsequently, the nodes will enter the period of the set deadline schedule. Then, confirm state transitions to "Disable" in the STATUS column and "DeadlineSchedule" in the REASON column by the BoB nodes in the range for maintenance work.

Also confirm state transitions by the maintenance target nodes to "ICC_Disable" in the ARCH_STATUS column.

4. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

   clstname: Cluster name
   bootgid: Boot group ID

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

5. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the target CMU to hardware maintenance mode.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

This operation handles the power-off processing of the maintenance target BoB.

6. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "off" in the REASON column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the maintenance target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   CN        Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010009   CN        Disable    HardMaintenance   off           ICC_Stopped    -
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the node specified in step 5.
   Also confirm that the PWR_STATUS column shows "off" for the nodes in the BoB of the node specified in step 5.

   The following example assumes the specified node ID in step 5 is 0xFF010001.

```
# pashowclst -c clstname --bootgrp bootgid -v
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010002   SIO,CN    Disable    DeadlineSchedule  off           ICC_Running    -
```

```
0xFF010003   GIO,CN    Disable    DeadlineSchedule  off          ICC_Running    -
0xFF010004   CN        Disable    DeadlineSchedule  off          ICC_Running    -
0xFF010005   CN        Disable    DeadlineSchedule  off          ICC_Running    -
0xFF010006   CN        Disable    DeadlineSchedule  off          ICC_Running    -
...
```

clstname: Cluster name
bootgid: Boot group ID

7. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --bootgrp bootgid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            off off on
0xFF010002 off            on  on  on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

clstname: Cluster name
bootgid: Boot group ID

2.2.3.2 Hardware Maintenance Work

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

Note

- Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.
  Perform the procedure "2.2.3.3 Incorporating Into Operation" to restart the nodes.
- If the BMC was replaced, the replacement BMC must be configured for FX server control. After BMC replacement, perform the work as an SE in "FX server Control Setting" in the "Job Operation Software Administrator's Guide for System Management."

See

For hardware maintenance procedures related to rebuilding by restoring a disk image, see "6.2 Backing Up and Restoring the FX server."

2.2.3.3 Incorporating Into Operation

After completing maintenance work on a CMU involving isolating a BoB from operation, incorporate the nodes in the CMU and BoB into operation.

1. Turning off the location LED of the maintenance target BoB

   If you turned on the location LED of the maintenance target BoB in "2.2.3.1 Isolating From Operation and Powering Off," turn them off with the paledctl command.

```
# paledctl -c clstname --bootgrp bootgid off
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

2. Releasing hardware maintenance mode on the target CMU

   Perform step 2 in "2.2.2.3 Incorporating Into Operation" in "2.2.2 Maintenance in Units of CMUs" to release hardware maintenance mode on the nodes in the maintenance target CMU.
   At this time, the command also simultaneously starts the nodes in the target BoB. Including the nodes in the maintenance target CMU, the nodes in the target BoB are left isolated from operation.

3. Confirming that maintenance mode was released on the target CMU and the BoB nodes started

   Confirm that maintenance mode has been released on the target CMU, the BoB nodes have started, and the nodes in the BoB including the target CMU have been left isolated from operation.

   Services have started on the BoB nodes by this time.

```
# pashowclst -c clstname -v --bootgrp bootgid
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE         NODETYPE  STATUS     REASON    PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    Manual    os-running    ICC_Disable
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002   SIO,CN    Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003   GIO,CN    Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   bootgid: Boot group ID

4. Checking the software environment

   From the software perspective, check whether the started nodes in the maintenance target CMU and BoB are ready to be incorporated into operation.
   Execute the pachecker command to check the environment.
   For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."

5. Incorporating the target BoB into operation

   Using the --enable option in the paclstmgr command, incorporate the nodes in the maintenance target BoB into operation.

```
# paclstmgr -c clstname --bootgrp bootgid --enable
```

   clstname: Cluster name
   bootgid: Boot group ID

Note

- If you isolated the GIO group in "2.2.3.1 Isolating From Operation and Powering Off," incorporate the nodes in the GIO group into operation.

```
# paclstmgr -c clstname -n nodeid --giogrp --enable
```

nodeid: Node ID of the maintenance target

- If you set a deadline schedule in "[By setting a deadline schedule]" in "2.2.3.1 Isolating From Operation and Powering Off," use the padeadline command to cancel the setting.

  For details on how to cancel the setting, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

6. Checking the node status

   Using the pashowclst command, confirm state transitions to "Running" in the STATUS column and "ICC_Running" in the ARCH_STATUS column by the nodes in the target BoB. Also confirm state transitions to "os-running" in the PWR_STATUS column by the nodes in the target BoB.

```
# pashowclst -c clstname -v --bootgrp bootgid
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE         NODETYPE  STATUS     REASON    PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Running    -         os-running    ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002   SIO,CN    Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003   GIO,CN    Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   bootgid: Boot group ID

   When all the services on a target node are running, the STATUS column shows "Running" for that node.

Note

If FX server power control with the Job Operation Software is unavailable due to a BMC failure or the like, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

2.2.4 Maintenance in Units of SIO Groups

Each storage I/O node is mounted to an SIO group. To perform maintenance for the CMU of a storage I/O node, isolate the relevant SIO group from operation.
To perform maintenance for the FX server in units of SIO groups, perform the procedures below.
Perform this work on the active system management node.

2.2.4.1 Isolating From Operation and Powering Off

Isolate the maintenance target SIO group from operation, transition it to hardware maintenance mode, and power it off.

The following three methods are ways toward that end:

- By waiting for running jobs to end
- By using another node to re-execute running jobs
- By setting a deadline schedule

Perform the steps in one of the above methods before isolating the SIO group from operation and transitioning it to hardware maintenance mode.

[By waiting for running jobs to end]

1. Checking the maintenance range

   Specify the node ID of the maintenance target node in the -n option in the pashowclst command, and additionally specify the --siogrp option, to check the nodes in the same SIO group as the maintenance target.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    SoftError SrvDown   os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(a),PWRD(o)
0xFF010002  SIO,CN    Disable   Alarm     off          ICC_Running    -
0xFF010003  GIO,CN    SoftError SrvDown   os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(a),PWRD(o)
0xFF010004  CN        SoftError SrvDown   os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010005  CN        SoftError SrvDown   os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010006  CN        SoftError SrvDown   os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
...
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

   To perform maintenance in units of SIO groups, power off the listed nodes through the above operation.

2. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

Note

   Suppose that a global I/O node is inside the range checked in step 1 and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that nodes in the GIO group should be isolated from job operation.

   Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE        NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
gionodeid1  GIO,CN    Running  -       os-running  ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2  GIO,CN    Stopped  -       off         ICC_Running  -
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node
   gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
   gionodeid2: Node ID of a global I/O node in the GIO group

3. Isolating the target SIO group from operation

   Maintenance in units of SIO groups requires the target SIO group to be powered off. For this reason, isolate nodes from operation (transition to the Disable state) by using the --disable option in the paclstmgr command.

```
# paclstmgr -c clstname --disable --cmu nodeid --siogrp
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

Note

Suppose that a global I/O node is inside the range checked in step 1 and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the node in the GIO group, not the SIO group, must be isolated from operation.

To isolate the nodes in the GIO group from operation, execute the following on the command line.

```
# paclstmgr -c clstname --disable -n nodeid --giogrp
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

4. Confirming isolation from operation

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "Manual" in the REASON column by the nodes in the maintenance target SIO group.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(a),PWRD(o)
0xFF010002  SIO,CN    Disable   Alarm     off          ICC_Running    -
0xFF010003  GIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(a),PWRD(o)
0xFF010004  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010005  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010006  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
...
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

5. Isolating the target CMU from operation

   Perform step 2 in "[By waiting for running jobs to end]" in "2.2.2.1 Isolating From Operation and Powering Off" in "2.2.2 Maintenance in Units of CMUs" to isolate the nodes in the maintenance target CMU from operation (transition to the Disable state).

   Using the pashowclst command, confirm state transitions from the status confirmed in step 4 to "ICC_Disable" in the ARCH_STATUS column by the nodes in the maintenance target CMU.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(a),PWRD(o)
0xFF010002  SIO,CN    Disable   Manual    off          ICC_Disable    -
0xFF010003  GIO,CN    Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(a),PWRD(o)
0xFF010004  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010005  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
0xFF010006  CN        Disable   Manual    os-running   ICC_Running
PLE(o),NRD(o),FEFS(a),PWRD(o)
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

6. Turning on the location LED of the target BoB

Note

Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.

To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

7. Transitioning to hardware maintenance mode

Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column by the nodes in the maintenance target SIO group. Then, transition the CMU of the target nodes to hardware maintenance mode by using the --hard-mainte option in the paclstmgr command.

```
# paclstmgr -c clstname --hard-mainte --cmu nodeid
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

Note

- This operation handles the power-off processing of the maintenance target SIO group.
- If you want to move the entire SIO group to hardware maintenance mode, switch to hardware maintenance mode with the --bootgrp option.

  ```
  # paclstmgr -c clstname --bootgrp bootgid --hard-mainte
  ```

  clstname: Cluster name
  bootgid: Boot group ID

- Depending on the hardware failure, the target node may not be able to stop normally.
  Use the force-off option of the papwrctl command.

8. Confirming hardware maintenance mode

Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the maintenance target CMU.
Also confirm that the PWR_STATUS column shows "off" for the nodes you isolated in Step 3.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE         NODETYPE  STATUS     REASON           PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    Manual           off           ICC_Running    -
```

```
0xFF010002   SIO,CN    Disable    HardMaintenance   off          ICC_Stopped    -
0xFF010003   GIO,CN    Disable    Manual            off          ICC_Running    -
0xFF010004   CN        Disable    Manual            off          ICC_Running    -
0xFF010005   CN        Disable    Manual            off          ICC_Running    -
0xFF010006   CN        Disable    Manual            off          ICC_Running    -
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

9. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid --siogrp all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            on  on  on
0xFF010002 off            off off on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

[By using another node to re-execute running jobs]

1. Checking the maintenance range

   Perform step 1 in "[By waiting for running jobs to end]" to check the maintenance range.

2. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

   
   Note

   Suppose that a global I/O node is inside the range checked in step 1 and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the node in the GIO group, not the SIO group, must be isolated from operation.

   Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE         NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS   SRV_STATUS
gionodeid1   GIO,CN    Running   -         os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2   GIO,CN    Stopped   -         off          ICC_Running   -
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node
gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
gionodeid2: Node ID of a global I/O node in the GIO group

3. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the range for maintenance work to hardware maintenance mode. If you also specify the -k option at this time, the command immediately suspends running jobs and jobs using the interconnect, and re-executes them on another node.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte -k
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

- This operation handles the power-off processing of the nodes in the maintenance target SIO group.
- Suppose that the maintenance target SIO group contains a global I/O node and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, the nodes in the GIO group are automatically isolated from operation when the paclstmgr command with the -hard-mainte and -k options transitions the maintenance targets to hardware maintenance mode. (The nodes that are not maintenance targets can use the interconnect. Also, excluding the SIO group of the maintenance target nodes, SIO groups are not powered off.)
- Depending on the hardware failure, the target node may not be able to stop normally.
  Use the force-off option of the papwrctl command.

4. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the node specified in step 3.
   Also confirm that the PWR_STATUS column shows "off" for the nodes in the SIO group of the node specified in step 1.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010002   SIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010003   GIO,CN    Disable    Manual            off           ICC_Running    -
0xFF010004   CN        Disable    Manual            off           ICC_Running    -
0xFF010005   CN        Disable    Manual            off           ICC_Running    -
0xFF010006   CN        Disable    Manual            off           ICC_Running    -
...
```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

5. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid --siogrp all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            on  on  on
0xFF010002 off            off off on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

6. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

[By setting a deadline schedule]

1. Notifying the end users of maintenance

   Notify the end users about setting a deadline schedule on the target nodes for maintenance.

Note

Suppose that a global I/O node is inside the range checked in step 1 and the active global I/O node in the GIO group is powered off, so the active node does not exist in the group. In this situation, notify the end user that a deadline schedule should be set for the nodes in the GIO group.

Check the global I/O nodes in the target GIO group as follows.

```
# pashowclst -c clstname --cmu nodeid --giogrp --nodetype GIO -v
[ CLST: clstname ]
[ CMU: nodeid ]
[ NODETYPE: GIO ]
NODE        NODETYPE  STATUS    REASON    PWR_STATUS   ARCH_STATUS   SRV_STATUS
gionodeid1  GIO,CN    Running   -         os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
gionodeid2  GIO,CN    Stopped   -         off          ICC_Running   -
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node
gionodeid1: Node ID of a global I/O node in the same CMU as the maintenance target node
gionodeid2: Node ID of a global I/O node in the GIO group

2. Setting a deadline schedule

   Set a deadline schedule on the nodes in the range for maintenance work.
   For details on how to set a deadline schedule, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

3. Checking the status of the deadline schedule

   Using the pashowclst command, confirm state transitions to "DeadlineSchedule" in the REASON column by the SIO group nodes in the range for maintenance work.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Running    DeadlineSchedule  os-running    ICC_Running    -
0xFF010002   SIO,CN    Running    DeadlineSchedule  os-running    ICC_Running    -
0xFF010003   GIO,CN    Running    DeadlineSchedule  os-running    ICC_Running    -
0xFF010004   CN        Running    DeadlineSchedule  os-running    ICC_Running    -
0xFF010005   CN        Running    DeadlineSchedule  os-running    ICC_Running    -
0xFF010006   CN        Running    DeadlineSchedule  os-running    ICC_Running    -
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

Subsequently, the nodes will enter the period of the set deadline schedule. Then, confirm state transitions to "Disable" in the STATUS column and "DeadlineSchedule" in the REASON column by the SIO group nodes in the range for maintenance work.

Also confirm state transitions by the target nodes to "ICC_Disable" in the ARCH_STATUS column.

4. Turning on the location LED of the target BoB

   To locate the target BoB for maintenance work by a Fujitsu customer engineer (CE), use the paledctl command to turn on the location LED of the BoB targeted for maintenance.

```
# paledctl -c clstname --bootgrp bootgid on
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

5. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the target CMU to hardware maintenance mode.

```
# paclstmgr -c clstname --cmu nodeid --hard-mainte
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

Note

This operation handles the power-off processing of the maintenance target SIO group.

6. Confirming hardware maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "off" in the REASON column, and "ICC_Stopped" in the ARCH_STATUS column by the nodes in the maintenance target CMU.

```
# pashowclst -c clstname --cmu nodeid -v
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010002   SIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF01000A   CN        Disable    HardMaintenance   off           ICC_Stopped    -
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column, "HardMaintenance" in the REASON column, "off" in the PWR_STATUS column, and "ICC_Stopped" in the ARCH_STATUS column by the node specified in step 4.
Also confirm that the PWR_STATUS column shows "off" for the nodes in the range specified in step 3.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE         NODETYPE  STATUS     REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    DeadlineSchedule  off           ICC_Running    -
0xFF010002   SIO,CN    Disable    HardMaintenance   off           ICC_Stopped    -
0xFF010003   GIO,CN    Disable    DeadlineSchedule  off           ICC_Running    -
0xFF010004   CN        Disable    DeadlineSchedule  off           ICC_Running    -
```

```
0xFF010005   CN        Disable    DeadlineSchedule  off          ICC_Running    -
0xFF010006   CN        Disable    DeadlineSchedule  off          ICC_Running    -
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

7. Confirming power status

   Use the papwrctl command to confirm that the PWR_STATUS, ICC, and CMU columns of the nodes in the CMU being maintained are transitioned to "off".

```
# papwrctl -c clstname --cmu nodeid --siogrp all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            on  on  on
0xFF010002 off            off off on
0xFF010003 off            on  on  on
0xFF010004 off            on  on  on
0xFF010005 off            on  on  on
0xFF010006 off            on  on  on
...
```

clstname: Cluster name
nodeid: Node ID within the maintenance target CMU

2.2.4.2 Hardware Maintenance Work

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

Note

- Check the status with the pashowclst command, If the REASON column shows "ResetRequest" for the maintenance target node, this task is unnecessary because parts replacement work by a Fujitsu customer engineer (CE) is not required.
  Perform the procedure "2.2.4.3 Incorporating Into Operation" to restart the nodes.
- If the BMC was replaced, the replacement BMC must be configured for FX server control. After BMC replacement, perform the work as an SE in "Control Setting for FX server" in the "Job Operation Software Administrator's Guide for System Management."

See

To rebuild a system by restoring a disk image because of hardware maintenance, see "6.2 Backing Up and Restoring the FX server."

2.2.4.3 Incorporating Into Operation

After completing maintenance work on a CMU involving isolating an SIO group from operation, incorporate the nodes in the CMU and SIO group into operation.

1. Turning off the location LEDs of the maintenance target SIO group

   If you turned on the location LEDs of the maintenance target SIO group in "2.2.4.1 Isolating From Operation and Powering Off," turn them off with the paledctl command.

```
# paledctl -c clstname --bootgrp bootgid off
```

clstname: Cluster name
bootgid: Boot group ID

See

For details on the paledctl command, see "Controlling the Location LED" in the "Job Operation Software Administrator's Guide for System Management."

2. Releasing hardware maintenance mode on the target CMU

   Perform step 2 in "2.2.2.3 Incorporating Into Operation" in "2.2.2 Maintenance in Units of CMUs" to release hardware maintenance mode on the nodes in the maintenance target CMU.

   At this time, the command also simultaneously starts the nodes in the target SIO group. Including the nodes in the maintenance target CMU, the nodes in the target SIO group are left isolated from operation.

Note

When the entire SIO group is moved to hardware maintenance mode, release hardware maintenance mode with the --bootgrp option.

```
# paclstmgr -c clstname --bootgrp bootgid --recover --disable --ic
```

clstname: Cluster name
bootgid: Boot group ID

3. Confirming that maintenance mode was released on the target CMU and the SIO group nodes started

   Confirm that maintenance mode has been released on the target CMU, the SIO group nodes have started, and the nodes in the SIO group including the target CMU have been left isolated from operation.

   Services have started on the SIO group nodes by this time.

```
# pashowclst -c clstname -v --cmu nodeid --siogrp
[ CLST: clstname ]
[ CMU: nodeid ]
NODE         NODETYPE  STATUS     REASON    PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Disable    Manual    os-running    ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002   SIO,CN    Disable    Manual    os-running    ICC_Disable
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003   GIO,CN    Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006   CN        Disable    Manual    os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

4. Checking the software environment

   From the software perspective, check whether the started nodes in the maintenance target CMU and SIO group are ready to be incorporated into operation.
   Execute the pachecker command to check the environment.
   For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."

5. Incorporating the target SIO group into operation

   Using the --enable option in the paclstmgr command, incorporate the nodes in the target SIO group into operation.

```
# paclstmgr -c clstname --cmu nodeid --siogrp --enable
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

Note

- If you isolated the GIO group in "2.2.4.1 Isolating From Operation and Powering Off," incorporate the nodes in the GIO group into operation.

```
# paclstmgr -c clstname -n nodeid --giogrp --enable
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

- If you set a deadline schedule in "[By setting a deadline schedule]" in "2.2.4.1 Isolating From Operation and Powering Off," use the padeadline command to cancel the setting.
  For details on how to cancel the setting, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

6. Checking the node status

Using the pashowclst command, confirm state transitions to "Running" in the STATUS column and "ICC_Running" in the ARCH_STATUS column by the nodes in the target CMU. Also confirm state transitions to "os-running" in the PWR_STATUS column by the nodes in the target SIO group.

```
# pashowclst -c clstname -n nodeid --siogrp
[ CLST: clstname ]
[ NODE: nodeid ]
NODE        NODETYPE  STATUS    REASON       PWR_STATUS   ARCH_STATUS    SRV_STATUS
0xFF010001  BIO,CN    Running   -            os-running   ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002  SIO,CN    Running   -            os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003  GIO,CN    Running   -            os-running   ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004  CN        Running   -            os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005  CN        Running   -            os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006  CN        Running   -            os-running   ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

clstname: Cluster name
nodeid: Node ID of the maintenance target node

When all the services on a target node are running, the STATUS column shows "Running" for that node.

Note

If FX server power control with the Job Operation Software is unavailable due to a BMC failure or the like, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

2.3 Operations During PRIMERGY server Maintenance

This section describes operations during hardware maintenance for PRIMERGY server.

2.3.1 Operations During Compute Node Maintenance

For maintenance on a PRIMERGY server compute node, prevent jobs from being allocated to the node. After that, perform the maintenance

Perform this work on the active system management node.

2.3.1.1 Isolating From Operation and Powering Off

Isolate the target node from operation, transition it to hardware maintenance mode, and power it off.
The following three methods are ways to complete the task:

- By waiting for running jobs to end
- By using another node to re-execute running jobs
- By setting a deadline schedule

Perform the steps in one of the above methods before isolating the node from operation and transitioning it to hardware maintenance mode.

[By waiting for running jobs to end]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target compute node from job operation for the purpose of maintenance.

2. Isolating the target node from operation

   Using the --disable option in the paclstmgr command, isolate the maintenance target compute node from operation (transition to the Disable state). At this time, processing waits for the end of running jobs on the target compute node.

   ```
   # paclstmgr -c clstname --disable -n nodeid1,nodeid2,...
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

3. Confirming isolation from operation

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "Manual" in the REASON column by the maintenance target compute node.

   ```
   # pashowclst -c clstname -n nodeid1,nodeid2,...
   [ CLST: clstname ]
   [ NODE: nodeid1 ]
   NODE        NODETYPE  STATUS   REASON         PWR_STATUS    ARCH_STATUS    SRV_STATUS
   nodeid1     CN        Disable  Manual         on            -
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   [ NODE: nodeid2 ]
   NODE        NODETYPE  STATUS   REASON         PWR_STATUS    ARCH_STATUS    SRV_STATUS
   nodeid2     CN        Disable  Manual         on            -
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   ...
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

4. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the node to hardware maintenance mode.

   ```
   # paclstmgr -c clstname --hard-mainte -n nodeid1,nodeid2,...
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

5. Confirming hardware maintenance mode

   Confirm that the numbers shown under COMPLETE and TOTAL for the hardware maintenance status by the pashowclst command are equal values.

   ```
   # pashowclst -c clstname -n nodeid1,nodeid2,... --hard-mainte
   [ CLST: clstname ]
   [ NODE: nodeid1 ]
   NODE     COMPLETE   TOTAL
   ```

```
nodeid1  1          1
[ NODE: nodeid2 ]
NODE     COMPLETE   TOTAL
nodeid2  1          1
...
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

[By using another node to re-execute running jobs]

1. Notifying the end users of maintenance

   Notify the end users about isolating the target compute nodes from job operation for the purpose of maintenance.

2. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the nodes in the range for maintenance work to hardware maintenance mode. If you also specify the -k option at this time, the command immediately suspends running jobs and jobs using the interconnect, and re-executes them on another node.

   ```
   # paclstmgr -c clstname --hard-mainte -k -n nodeid1,nodeid2,...
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

3. Confirming hardware maintenance mode

   Confirm that the numbers shown under COMPLETE and TOTAL for the hardware maintenance status by the pashowclst command are equal values.

   ```
   # pashowclst -c clstname -n nodeid1,nodeid2,... --hard-mainte
   [ CLST: clstname ]
   [ NODE: nodeid1 ]
   NODE     COMPLETE   TOTAL
   nodeid1  1          1
   [ NODE: nodeid2 ]
   NODE     COMPLETE   TOTAL
   nodeid2  1          1
   ...
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

[By setting a deadline schedule]

1. Notifying the end users of maintenance

   Notify the end users about the setting of a deadline schedule on the target compute nodes for maintenance.

2. Setting a deadline schedule

   Set a deadline schedule on the nodes in the range for maintenance work.
   For details on how to set a deadline schedule, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

3. Confirming isolation from operation

   Using the pashowclst command, confirm state transitions to "DeadlineSchedule" in the REASON column by the nodes in the range for maintenance work.

   The following example checks the status of specific nodes.

   ```
   # pashowclst -c clstname -n nodeid1,nodeid2,...
   [ CLST: clstname ]
   [ NODE: nodeid1 ]
   NODE        NODETYPE  STATUS   REASON            PWR_STATUS  ARCH_STATUS  SRV_STATUS
   nodeid1     CN        Running  DeadlineSchedule  on          -
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   ```

```
[ NODE: nodeid2 ]
NODE        NODETYPE  STATUS   REASON            PWR_STATUS  ARCH_STATUS  SRV_STATUS
nodeid2     CN        Running  DeadlineSchedule  on          -
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

Subsequently, the nodes will enter the period of the set deadline schedule. Then, confirm state transitions to "Disable" in the STATUS column and "DeadlineSchedule" in the REASON column by the compute nodes in the range for maintenance work.

4. Transitioning to hardware maintenance mode

   Using the pashowclst command, confirm state transitions by the isolated nodes to "Disable" in the STATUS column. Then, transition the target nodes to hardware maintenance mode by using the --hard-mainte option in the paclstmgr command.

```
# paclstmgr -c clstname --hard-mainte -n nodeid1,nodeid2,...
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

5. Confirming hardware maintenance mode

   Confirm that the numbers shown under COMPLETE and TOTAL for the hardware maintenance status by the pashowclst command are equal values.

```
# pashowclst -c clstname -n nodeid1,nodeid2,... --hard-mainte
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE     COMPLETE   TOTAL
nodeid1  1          1
[ NODE: nodeid2 ]
NODE     COMPLETE   TOTAL
nodeid2  1          1
...
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

2.3.1.2 Hardware Maintenance Work

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

2.3.1.3 Incorporating Into Operation

After completing maintenance work on a node, incorporate the node into operation.

1. Releasing hardware maintenance mode on the target node

   Release hardware maintenance mode on the target node, and power it on.

   When releasing hardware maintenance mode, leave the target node isolated from operation (Disable state). In step 3 below, you will be checking whether the node can be incorporated into operation.

   Specify the --disable option in addition to the --recover option in the paclstmgr command to release hardware maintenance mode.

   At this time, the command also simultaneously powers on the target nodes.

```
# paclstmgr -c clstname -n nodeid1,nodeid2, ... --recover --disable
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

Information

- If you set a deadline schedule, the node is automatically incorporated into operation upon reaching the set end date and time after hardware maintenance mode is released. However, even after hardware maintenance mode is released, the compute node may remain in the Disable state, so it will not be immediately incorporated into operation.
  To immediately incorporate the node into operation without waiting for the set end date and time, cancel the set deadline schedule.
- To incorporate the node into operation directly after hardware maintenance mode is released, specify only the --recover option in the paclstmgr command.

2. Confirming that hardware maintenance mode was released

   Confirm that hardware maintenance mode has been released and that the node has started.

```
# pashowclst -c clstname -n nodeid1,nodeid2,...
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   CN         Disable    Manual   on           -              PLE(o),NRD(o),FEFS(o),PWRD(o)
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   CN         Disable    Manual   on           -              PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

3. Checking the software environment

   From the software perspective, check whether the started node (the maintenance target) is ready to be incorporated into operation. Execute the pachecker command to check the environment.
   For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."

4. Incorporating into operation

   Using the --enable option in the paclstmgr command, incorporate the target node into operation.

```
# paclstmgr -c clstname -n nodeid1,nodeid2, ... --enable
```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

5. Checking the node status

   Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

```
# pashowclst -c clstname -n nodeid1,nodeid2,...
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   CN         Running    -        on           -              PLE(o),NRD(o),FEFS(o),PWRD(o)
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   CN         Running    -        on           -              PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

   When all the services on a target node are running, the STATUS column shows "Running" for that node.

2.3.2 Operations During System Management Node Maintenance

Hardware maintenance on the system management node in a redundant configuration is performed on the standby system management node.

Even if the system management node also serves as the storage cluster management node, use this procedure for maintenance.

Perform this work on the active system management node.

1. Preparation

   a. Checking whether the node is active or standby

      Check whether the maintenance target system management node is the active node or standby node.

      Check the display by the pashowclst command. In the STATUS column, "(A)" indicates an active node, and "(S)" indicates a standby node.

```
# pashowclst -c clstname -m --nodetype SMM
[ CLST: clstname ]
[ NODETYPE: SMM ]
NODEGRP     NODETYPE       NODE         STATUS  NODE         STATUS
SMMGRP      SMM            nodeid1      (A)     nodeid2      (S)
```

      clstname: Cluster name
      nodeid1, nodeid2: Node IDs of the system management node

   b. Switching from the active node to the standby node

      If the maintenance target node is the active node, switch it to the standby node by using the --failover option in the paclstmgr command.

```
# paclstmgr -c clstname --failover nodeid1
```

      clstname: Cluster name
      nodeid1: Node ID of the active node

   c. After the switching, confirm that the status is as shown below.

```
# pashowclst -c clstname -m --nodetype SMM
[ CLST: clstname ]
[ NODETYPE: SMM ]
NODEGRP     NODETYPE       NODE         STATUS  NODE         STATUS
SMMGRP      SMM            nodeid1      (S)     nodeid2      (A)
```

      clstname: Cluster name
      nodeid1, nodeid2: Node IDs of the system management node

   d. Transitioning to hardware maintenance mode

      After step b is done, the active node becomes the standby node. Therefore, log in the new active node to perform the following steps.
      Using the --hard-mainte option in the paclstmgr command, transition the maintenance target node (which was switched to the standby node) to hardware maintenance mode, and power it off.

```
# paclstmgr -c clstname -n nodeid1 --hard-mainte
```

      clstname: Cluster name
      nodeid1: Node ID of the maintenance target node

      Using the pashowclst command, confirm the transition to hardware maintenance mode.
      Confirm that STATUS column shows "Disable", REASON column shows "HardMaintenance", and PWR_STATUS column shows "off".

```
# pashowclst -c clstname -n nodeid1
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE        NODETYPE      STATUS        REASON              PWR_STATUS        ARCH_STATUS
```

```
SRV_STATUS
nodeid1     SMM          Disable      HardMaintenance     off              -
-
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

2. Performing hardware maintenance work

   For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

3. Incorporating into operation

   Using the --recover option in the paclstmgr command, release hardware maintenance mode on the target node, and incorporate it into operation. The command simultaneously powers on the target node.

```
# paclstmgr -c clstname -n nodeid1 --recover
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

Using the pashowclst command, confirm the incorporation into operation.
Confirm that STATUS column shows "Running".

```
# pashowclst -c clstname -n nodeid1
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE        NODETYPE      STATUS        REASON              PWR_STATUS        ARCH_STATUS
SRV_STATUS
nodeid1     SMM           Running       -                   on                -
PWRD(o)
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

Information

To return the nodes to their original active or standby states before the work began, perform the following step again: "b. Switching the active node to the standby node" in "1. Preparation."

2.3.3 Operations During Management Node Maintenance

Hardware maintenance on a management node (compute cluster management node, compute cluster management sub node, or storage cluster management node) in a redundant configuration is performed on the standby node.

See

If the storage cluster management node also serves as the system management node, perform maintenance with the procedure described in "2.3.2 Operations During System Management Node Maintenance ."

Perform this work on the active system management node.

1. Preparation

   a. Checking whether the node is active or standby

      Check whether the maintenance target node is the active node or standby node.
      Check the display by the pashowclst command. In the STATUS column, "(A)" indicates an active node, and "(S)" indicates a standby node.

```
# pashowclst -c clstname -m --nodetype nodetype
[ CLST: clstname ]
[ NODETYPE: nodetype ]
```

```
NODEGRP    NODETYPE      NODE        STATUS  NODE        STATUS
MASTERGRP  nodetype      nodeid1     (A)     nodeid2     (S)
```

clstname: Cluster name
nodetype: Target node type
nodeid1, *nodeid2*: Node IDs of the target management node

b. Switching from the active node to the standby node

If the maintenance target node is the active node, switch it to the standby node by using the --failover option in the paclstmgr command.

```
# paclstmgr -c clstname --failover nodeid1
```

clstname: Cluster name
nodeid1: Node ID of the active node

After the switching, confirm that the status is as shown below.

```
# pashowclst -c clstname -m --nodetype nodetype
[ CLST: clstname ]
[ NODETYPE: nodetype ]
NODEGRP    NODETYPE      NODE        STATUS  NODE        STATUS
MASTERGRP  nodetype      nodeid1     (S)     nodeid2     (A)
```

clstname: Cluster name
nodetype: Target node type
nodeid1, nodeid2: Node IDs of the target management node

c. Transitioning to hardware maintenance mode

Using the --hard-mainte option in the paclstmgr command, transition the maintenance target node (which was switched to the standby node) to hardware maintenance mode, and power it off.

```
# paclstmgr -c clstname -n nodeid1 --hard-mainte
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

2. Performing hardware maintenance work

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

3. Incorporating into operation

Using the --recover option in the paclstmgr command, release hardware maintenance mode on the target node, and incorporate it into operation. The command simultaneously powers on the target node.

```
# paclstmgr -c clstname -n nodeid1 --recover
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

4. Checking the node status

Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

```
# pashowclst -c clstname -n nodeid1,nodeid2
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   nodetype   Running    -        on           -              services
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   nodetype   Running    -        on           -              services
```

clstname: Cluster name
nodetype: Target node type

nodeid1, *nodeid2*: Node IDs of the maintenance target management node
services: Run status of services on a target node

When all the services on a target node are running, the STATUS column shows "Running" for that node.

Information

To return the nodes to their original active or standby states before the work began, perform the following step again: "b. Switching the active node to standby node" in "1. Preparation."

2.3.4 Operations During Login Node Maintenance

Hardware maintenance on a login node is performed after the node is powered off.

1. Preparation

   a. Notifying the end users of maintenance

      Even after a failure, the login node may continue operating by degrading the fault location. In such cases, in order to stop the login node, notify the end users so that they log out of the target login node by the scheduled stop time.

      The following example executes the wall command on the login node to notify all the users on the node.

      ```
      # wall "message"
      ```

      Message: Notification message

   b. Preventing new logins

      As the scheduled stop time approaches, make the following setting to prevent new logins by end users.

      ```
      # touch /etc/nologin
      ```

      After this operation, only the root user can log in.

   c. Transitioning to hardware maintenance mode

      Execute the paclstmgr command with the --hard-mainte option on the active system management node to transition the login node to hardware maintenance mode and power it off.

      ```
      # paclstmgr -c clstname -n nodeid --hard-mainte
      ```

      clstname: Cluster name
      nodeid: Node ID of the target node

2. Performing hardware maintenance work

   For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

3. Incorporating into operation

   Using the --recover option in the paclstmgr command on the active system management node, release hardware maintenance mode on the target node, and incorporate it into operation. The command simultaneously powers on the target node.

   ```
   # paclstmgr -c clstname -n nodeid --recover
   ```

   clstname: Cluster name
   nodeid: Node ID of the target node

4. Checking the node status

   Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

   ```
   # pashowclst -c clstname -n nodeid
   [ CLST: clstname ]
   [ NODE: nodeid  ]
   ```

```
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS   SRV_STATUS
nodeid    LN         Running    -        on           -             PLE(o),FEFS(o)
```

clstname: Cluster name
nodeid: Node ID of the target node

5. Canceling suppression of new logins

   Cancel suppression of new logins by end users.
   The following example deletes the file created in step 1 b.

```
# rm /etc/nologin
```

6. Notifying the end users of work completion

   Notify the end users that the maintenance work has ended.

2.3.5 Operations During File Server Node Maintenance

Hardware maintenance on a file server node (MGS node, MDS node, or OSS node) in a mutual standby configuration is performed after replication of the FEFSSR service to the paired node in the mutual standby configuration.

Perform this work on the active system management node.

1. Preparation

   a. Replicating the FEFSSR service

      Using the --failover option in the paclstmgr command, replicate the FEFSSR service to the paired node in the mutual standby configuration.

```
# paclstmgr -c clstname --failover nodeid1
```

      clstname: Cluster name
      nodeid1: Node ID of the maintenance target node

      (*) The node ID displayed at this time (*nodeid2*) is the destination node for the replicated service.

Note

      The MGS and MDS nodes may be configured redundantly as pairs of active and standby nodes. Replicate the FEFSSR service only if the maintenance target node is the active node.
      For the procedure for checking whether the maintenance target is the active or standby MGS/MDS node, see "Switching a Node" in the "Job Operation Software Administrator's Guide for System Management."

   b. Checking the status of the FEFSSR service

      Using the pashowclst command, confirm that replication of the FEFSSR service has completed. If the SRV_STATUS column shows "w" for the FEFSSR service on the target node, the service has been replicated on the paired node in the mutual standby configuration. In other words, the target node has switched to standby.

```
# pashowclst -c clstname -n nodeid1
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE       NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS   SRV_STATUS
nodeid1    nodetype   SoftError  SrvDown  on           -             FEFSSR(w)
```

      clstname: Cluster name
      nodeid1: Node ID of the maintenance target node
      nodetype: Target node type

Information

If a fault is found in the target node before replication of the FEFSSR service, the node is forcibly stopped. In this case, execute the pashowclst command, and confirm that the node has stopped. If the PWR_STATUS column shows "off" for the node, it has stopped.

You can also check the destination node for the replicated service to see whether or not the FEFSSR service has been replicated. If the SRV_STATUS column shows "*" for the FEFSSR service on that node, the node is operating with the replicated service.

```
# pashowclst -c clstname -n nodeid2
[ CLST: clstname ]
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   nodetype   Running    -        on           -              FEFSSR(*)
```

clstname: Cluster name
nodeid2: Node ID of the service replication destination
nodetype: Target node type

c. Transitioning to hardware maintenance mode

Using the --hard-mainte option in the paclstmgr command, transition the replication source node for the FEFSSR service to hardware maintenance mode, and power it off.

```
# paclstmgr -c clstname --hard-mainte -n nodeid1
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

2. Performing hardware maintenance

For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

3. Incorporating into operation

Using the --recover option in the paclstmgr command, release hardware maintenance mode on the target node. The command simultaneously powers on the target node.

```
# paclstmgr -c clstname -n nodeid1 --recover
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node

The target node starts in the standby state.
Check the output as follows.

```
# pashowclst -c clstname -n nodeid1,nodeid2
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   nodetype   SoftError  SrvDown  on           -              FEFSSR(w)
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   nodetype   Running    -        on           -              FEFSSR(*)
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of the target file server node
nodetype: Target node type

4. Recovering the replicated service

Assuming you replicated the service in the preparation, recover the service on the node not currently running the service, by using the --failback option in the paclstmgr command.

```
# paclstmgr -c clstname --failback nodeid1
```

clstname: Cluster name
nodeid1: Node ID of the maintenance target node (The node ID of a node that became the standby system node when the --failover option in the paclstmgr command was executed)

5. Checking the node status

   Using the pashowclst command, confirm state transitions to "Running" in the STATUS column by the target node and to "o" in the SRV_STATUS column by the FEFSSR service.

```
# pashowclst -c clstname -n nodeid1,nodeid2
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   nodetype   Running    -        on           -              FEFSSR(o)
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   nodetype   Running    -        on           -              FEFSSR(o)
```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of the target file server node
   nodetype: Target node type

   When all the services on a target node are running, the STATUS column shows "Running" for that node.

2.3.6 Operations During Multiuse Node Maintenance

Hardware maintenance on a multiuse node is performed after the node is powered off.

Perform this work on the active system management node.

1. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the target node to hardware maintenance mode, and power it off.

```
# paclstmgr -c clstname -n nodeid --hard-mainte
```

   clstname: Cluster name
   nodeid: Node ID of the target node

2. Performing hardware maintenance work

   For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

3. Incorporating into operation

   Using the --recover option in the paclstmgr command, release hardware maintenance mode on the target node, and incorporate it into operation. The command simultaneously powers on the target node.

```
# paclstmgr -c clstname -n nodeid --recover
```

   clstname: Cluster name
   nodeid: Node ID of the target node

4. Checking the node status

   Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

```
# pashowclst -c clstname -n nodeid
[ CLST: clstname ]
[ NODE: nodeid ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid    nodetype   Running    -        on           -              -
```

   clstname: Cluster name
   nodeid: Node ID of the target node
   nodetype: Node type name of the target node

2.4 Operations During InfiniBand Maintenance

InfiniBand maintenance consists of the following:

- HCA maintenance
- IB cable maintenance
- IB switch maintenance

For the above-described maintenance, power off the connected nodes. After that, perform the hardware maintenance.

Perform this work on the active system management node.

Procedures for hardware maintenance vary depending on the node type of the maintenance target.

Perform the work in "2.3.2 Operations During System Management Node Maintenance ," "2.3.3 Operations During Management Node Maintenance," and "2.3.5 Operations During File Server Node Maintenance" as the maintenance procedures for the system management node, management nodes, and file server nodes, respectively.

For nodes other than the above, perform the following procedure.

1. Isolating from operation

   If the maintenance target node is a compute node, isolate it from operation.

   ```
   # paclstmgr -c clstname -n nodeid --disable
   ```

   clstname: Cluster name
   nodeid: Node ID of the target node

2. Transitioning to hardware maintenance mode

   Using the --hard-mainte option in the paclstmgr command, transition the maintenance target node to hardware maintenance mode, and power it off.

   ```
   # paclstmgr -c clstname -n nodeid --hard-mainte
   ```

   clstname: Cluster name
   nodeid: Node ID of the target node

3. Performing hardware maintenance work

   For maintenance on hardware, contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

4. Incorporating into operation

   Using the --recover option in the paclstmgr command, release hardware maintenance mode on the target node, and incorporate it into operation. The command simultaneously powers on the target node.

   ```
   # paclstmgr -c clstname -n nodeid --recover
   ```

   clstname: Cluster name
   nodeid: Node ID of the target node

5. Checking the node status

   Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

   ```
   # pashowclst -c clstname -n nodeid
   [ CLST: clstname ]
   [ NODE: nodeid ]
   NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
   nodeid    nodetype   Running    -        on           -              services
   ```

   clstname: Cluster name
   nodetype: Node type name of the target node
   nodeid: Node ID of the target node
   services: Run status of services on the target node

2.5 Operations During Network Device Maintenance

Network maintenance consists of the following:

- Reboot of network devices
- Replace of network devices
- Firmware updates for network devices

These maintenance procedures determine the scope of the impact of network devices maintenance and address affected nodes before performing network maintenance.

2.5.1 Maintaince of the Management Network

If you perform maintenance on the management network, you will not be able to communicate using the management network. Therefore, the job operation software does not operate normally. The nodes affected by the maintenance of the management network must be addressed by the job operation software to minimize the impact.

In particular, you should use caution when performing maintenance that prevents communication between the system management node or between the compute cluster management node and each node in the cluster.

If you perform maintenance on the management network while the job operation software service is running, the system monitoring function might detect a software error.

The node to be handled differs depending on the range of influence of the maintenance of the network device. Determine the scope of impact on nodes that will be unable to communicate using the management network due to network device maintenance.

Table 2.5 Scope of impact of network device maintenance

Node that will not be able to communicate	Scope of impact of maintenance
System management node System management node also serving as storage cluster management node	This applies to all nodes of all clusters.
Storage cluster management node	This applies to all storage cluster management nodes and subordinate nodes.
Compute cluster management node	Compute cluster administration node and all nodes (compute node, login node and etc) in the same cluster.
Compute cluster sub management node	This applies to the compute cluster sub management node and subordinate compute nodes.
MGS node MGS node also serving as MDS node and OSS node MGS node also serving as MDS node MDS node OSS node	The target node is a node that cannot communicate. It also affects all nodes that use the storage cluster.
Node serving as both compute node and boot I/O node(FX server)	This applies to the node serving as both compute node and boot I/O node and the subordinate compute nodes (including nodes serving multiple roles).
Compute node (PRIMERGY server) Login node Multiuse node	The target node is a node that cannot communicate. It also affects all nodes that use the storage cluster.

After identifying the node to be maintained using the list above, refer to the procedure below for each node type to be maintained.

- For maintenance in which communication failure occurs for a short time (within 1 minute)

  It is assuming that network outages due to maintenance such as resetting network devices or replacing LAN cables will be completed in a short time.

  If communication using the management network is not possible for a short time (1 minute or less), perform maintenance on the management network without taking action on the job operation software.

Information

Even for maintenance that can cause a short (less than one minute) loss of communication, the system monitoring feature may detect a node down. If the system monitoring function detects a node down, transit the node to hardware maintenance mode and isolate the node from operation. Then incorporate the node into the operation.

Perform the work in "2.3 Operations During PRIMERGY server Maintenance" as the maintenance procedures for PRIMERGY server.

Perform the work in "2.2.3 Maintenance in Units of BoBs" as the maintenance procedures for FX server.

- For maintenance that causes a long communication loss

  It is assumed that a network outage due to maintenance such as network device replacement or firmware update continues for a long time.

  To minimize the impact of network maintenance, all affected nodes are moved to software maintenance mode in advance, and maintenance is performed after the service of job operation software is stopped.

  For the software maintenance procedures, see "Chapter 3 Software Maintenance."

Note

When moving to software maintenance mode, be sure to stop the job operation software service by specifying the option to stop the service.

2.5.2 Maintaince of the Control Network

Control network maintenance has the following effects.

- Getting the power state of a node
- Power control (node power ON/OFF)
- console connection
- Acquisition of power status
- Collecting a dump file

When the system management node cannot communicate using the control network due to maintenance of the control network, all nodes except the operation system management node are shifted to the software maintenance mode and the service of the job operation software is stopped.

If maintenance of the control network does not affect the system management nodes, move only the affected nodes to software maintenance mode and stop the job operations software service.

For the software maintenance procedures, see "Chapter 3 Software Maintenance."

Chapter 3 Software Maintenance

This chapter describes software maintenance with the Job Operation Software.

3.1 Software Maintenance Concept

This section describes the concept of software maintenance, such as applying fix packages (called "packages" below) to correct problems and add functions or updating configuration files.

3.1.1 Maintenance Work Range

The range of software maintenance varies depending on what is applied to correct a problem or add a function. For this reason, follow the provided instructions (README, etc.) for corrections in the effective range of maintenance.

The following work ranges for software maintenance conform to the effective range of corrections.

Table 3.1 Software Maintenance Work Units

Range	Description
Entire system	The range of maintenance in this case covers an entire system. Examples include system reinstallation. Job operations must be stopped.
Cluster	The range of maintenance in this case covers an entire cluster. Examples include maintenance on both the active and standby compute cluster management nodes or storage cluster management nodes. This also applies if the above-described nodes are not in redundant configuration. Job operations must be stopped in the target cluster. To perform maintenance on an entire storage cluster, job operations must be stopped in the compute cluster used. This is because the compute cluster is also affected.
Node group	The range of maintenance in this case is confined to a node group. Examples include maintenance on both the active and standby compute cluster sub management nodes. With the target node group isolated from operation, you can perform partial maintenance while job operations continue.
SIO group	The range of maintenance in this case is confined to the nodes in an SIO group. Examples include maintenance for a storage I/O node. With the target SIO group isolated from operation, you can perform partial maintenance while job operations continue.
GIO group	The range of maintenance in this case is confined to the nodes in a GIO group. Examples include maintenance for a global I/O node. With the target GIO group isolated from operation, you can perform partial maintenance while job operations continue.
Boot group	The range of maintenance in this case is confined to the nodes in a boot group. Examples include maintenance for a boot I/O node. With the target boot group isolated from operation, you can perform partial maintenance while job operations continue.
Resource group	Maintenance in this case is in units of resource groups, considering the performance effect on jobs. With the nodes belonging to the maintenance target resource group isolated from operation, you can perform partial maintenance while job operations continue.
File system	Maintenance in this case is done on MGS, MDS, and OSS nodes. The maintenance target file system is stopped and unmounted. You can perform partial maintenance while job operations continue by using a file system other than the maintenance target.
Node (redundant configuration)	Maintenance in this case is done on nodes (system management node, compute cluster management node, compute cluster sub management node, MGS node, MDS node, OSS node) in a redundant configuration. By performing maintenance on the standby node, you allow job operations to continue during the

Range	Description
	maintenance. If the target node is not in a redundant configuration, job operations must be stopped in the unit of a cluster or file system.
Node (non-redundant configuration)	Maintenance in this case is done on the login node, a compute node, or a multiuse node. With the maintenance target node isolated from operation, you can perform maintenance while job operations continue by using a node other than the maintenance target.

Information

"Rolling Update" refers to performing partial maintenance on some compute nodes while continuing job operation within the cluster without stop the system or the entire cluster.
For the availability of rolling updates, see "3.1.2 Applying Packages Through Rolling Updates."

3.1.2 Applying Packages Through Rolling Updates

The following prerequisites apply to rolling updates of fix packages for the Job Operation Software.

- Rolling updates are not permitted for upgrades to a higher version or level.
- Rolling updates are only permitted to use packages for the previous version.

For software maintenance that cannot be done with a rolling update, the software maintenance work unit is a cluster as shown in "3.1.1 Maintenance Work Range."

The conditions for using a rolling update are written under Description in the packages provided for the Job Operation Software.
Use the -qpi option in the rpm command to see the Description entry in a package.

```
# rpm -qpi FJSVpxpsm.xxxx.rpm

...(*)

Description :
The Technical Computing Suite System Management Foundation is a program for monitoring and control the
job scheduling system.
This package contains the monitoring daemon (pxmonitor) and some control commands.
Version coexistence: possible
  Previous version: 3.0.0 and over
  Same version range: node(redundant)
After updating: service restart
Relation package: none
```

xxxx: Character string representing the version number and architecture

(*) The displayed information includes the package name and version.

Version coexistence

Whether a mixture of versions in the same environment is permitted (possible: yes, impossible: no).
If this item shows "possible," the package can be applied through a rolling update.

Previous version

Previous version number.
The package can be applied through a rolling update only when updating from the version number written here.
If the [Version coexistence] item shows "impossible," this item is not displayed.

Use the -qi option in the rpm command to check the version number of the currently applied package.

Same version range

Permitted range for a mixture of versions.
The package is applied through a rolling update according to the unit written here.
If the [Version coexistence] item shows "impossible," this item is not displayed.

Table 3.2 Possible Ranges for a Mixture of Versions

[Same version range] Value	Meaning
node group	Node group
sio group	SIO group
gio group	GIO group
boot group	Boot group
resource group	Resource group
file system	File system
node(redundant)	Node (redundant configuration)
node(not redundant)	Node (non-redundant configuration)

After updating

Necessity of a restart after package application.
After this package is applied, one of the following operations is required.

Table 3.3 Operations Required After Package Application

[After updating] Value	Restart Method
unnecessary to restart	Restart not required
service restart	Service restart
reboot	Node restart (Cold reboot)
cold reboot	

Note

To apply multiple packages that have different values shown at [After updating], implement either of the restart methods closest to the bottom of the above table. For example, if [After updating] shows "service restart" and "cold reboot" for two packages, the "cold reboot" package has priority over the "service restart" package when the packages are applied. For this reason, restart the node. When you restart the node, a service restart is unnecessary.

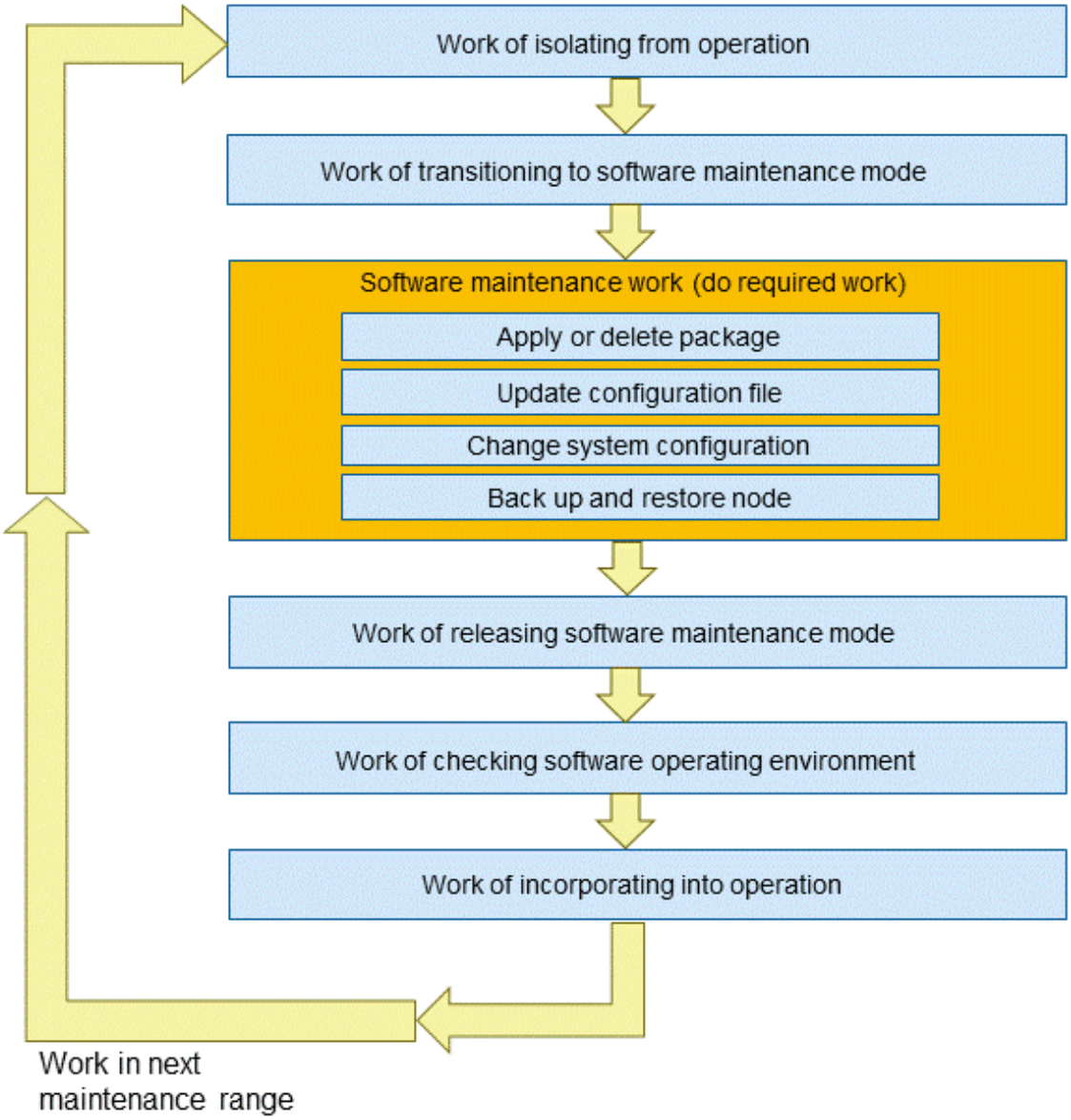
Relation package

Written name of the package that must be applied at the same time.
If this item shows "none," no package must be applied at the same time.

3.2 Software Maintenance Workflow

Perform software maintenance in the following procedure.

Figure 3.1 Software Maintenance Workflow



The next sections respectively describe the preparation for software maintenance, maintenance work, and post-maintenance work.

3.3 Preparation for Software Maintenance

Before software maintenance, such as applying or deleting a package or updating a configuration file, the following work must be done so that maintenance can be done safely on the nodes in the range for maintenance work.

- Isolating from operation

  Isolate the maintenance target nodes from operation by node type.

  - Maintenance on a compute node

    If the maintenance target is a compute node, the nodes in the range for maintenance work have to be in a state where they are not executing jobs. Perform the work in "3.3.1.1 Isolating a Compute Node From Operation ."

  - Maintenance on a management node

    If the maintenance target is a redundantly configured management node (compute cluster management node, compute cluster management sub node, or storage cluster management node), set the node to standby. Perform the work in "3.3.1.2 Isolating Redundantly Configured Nodes From Operation."

When maintaining the active node and standby node at the same time, disable the failover function. If the failover function is not suppressed, an error may be detected when recovering from the software maintenance mode.

For information about disabling the automatic failover, see "Setting Automatic Failover" in the "Job Operation Software Administrator's Guide for System Management".

- Maintenance on a file server node or a file system

  If the maintenance target is a redundantly configured file server node or file system, replicate the node. Perform the work in "3.3.1.3 Isolating a File Server Node From Operation."

  When maintaining the active node and standby node at the same time, disable the failover function. If the failover function is not suppressed, an error may be detected when recovering from the software maintenance mode.

  For information about disabling the automatic failover, see "Setting Automatic Failover" in the "Job Operation Software Administrator's Guide for System Management."

- Maintenance on the login node or a multiuse node

  The work of isolation from operation is unnecessary. Perform the work in "3.3.2 Transitioning to Software Maintenance Mode" directly.

- Transitioning to software maintenance mode

  Transition the maintenance target node to software maintenance mode and stop the Job Operation Software service on the node so that it is ready for maintenance work. Perform the work in "3.3.2 Transitioning to Software Maintenance Mode."

- Stopping the FEFS service

  To perform FEFS maintenance in the range for maintenance work or to restart a node that uses FEFS, stop the FEFS service so that neither FEFS maintenance nor the file system is affected. Perform the work in "3.3.3 Stopping the FEFS Service ."

- Starting TFTP and HTTP services

  If you perform maintenance work (Apply or remove packages, back up or restore nodes, rebuild nodes) using the installation function, perform "3.3.4 Starting TFTP and HTTP Services."

Perform the work in "Isolating from operation," "Transitioning to software maintenance mode," and "Stopping the FEFS service" in this order.

If the maintenance range includes the system management node, isolate the nodes in the cluster from operation, and transition them to software maintenance mode. Then, transition the system management node to software maintenance mode.

If you perform the work in a different order than described above, the system may possibly stop operating. Perform the work in the above procedures in the order described.

This section describes procedures for preparing for software maintenance.

Perform the work on the active system management node, unless otherwise noted.

3.3.1 Isolating From Operation

To perform software maintenance, isolate the maintenance target node from operation to prepare for a transition to software maintenance mode. For the range of the maintenance work, see "3.1.1 Maintenance Work Range."

Perform the procedures described below for the respective ranges of maintenance to transition to software maintenance mode.

3.3.1.1 Isolating a Compute Node From Operation

If the maintenance range includes a compute node, the nodes in the range for maintenance work have to be in a state where they are not executing jobs.
The following three methods are ways toward that end:

- By waiting for running jobs to end

- By using another node to re-execute running jobs
- By setting a deadline schedule

Perform the steps in one of the above methods before isolating the maintenance target node from operation.

[By waiting for running jobs to end]

Wait for the running job to end.
Perform this work on the active system management node.

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

   Note

   If the maintenance target nodes contain a global I/O node, and the maintenance operation causes a global I/O node that is running to no longer exist in the GIO group, notify the end user that the node in the GIO group should be isolated from job operation. Check the global I/O nodes in the target GIO group as follows.

   ```
   # pashowclst -c clstname -n nodeid --giogrp --nodetype GIO -v
   [ CLST: clstname ]
   [ CMU: nodeid ]
   [ NODETYPE: GIO ]
   NODE         NODETYPE  STATUS   REASON  PWR_STATUS  ARCH_STATUS  SRV_STATUS
   gionodeid1   GIO,CN    Running  -       os-running  ICC_Running
   PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
   gionodeid2   GIO,CN    Stopped  -       off         ICC_Running  -
   ```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node
   gionodeid1, *gionodeid2*: Node ID of a global I/O node within the same GIO group of the maintenance target node

2. Isolating the nodes in the range for maintenance work from operation

   Using the --disable option in the paclstmgr command, isolate the nodes in the range for maintenance work from operation. Wait until the end of the jobs running at this time.

   The following example checks the status of specific nodes.

   ```
   # paclstmgr -c clstname -n nodeid1,nodeid2,... --disable
   ```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

   For details on the nodes in the range for maintenance work, see "3.1.1 Maintenance Work Range."

   Note

   If the maintenance target nodes contain a global I/O node, and the maintenance operation causes a global I/O node that is running to no longer exist in the GIO group, isolate the nodes in the GIO group of the maintenance target node from job operation.
   To isolate the nodes in the GIO group from job operation, execute the following on the command line.

   ```
   # paclstmgr -c clstname --disable -n nodeid --giogrp
   ```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

3. Confirming the isolation from operation

   Using the -v option in the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "Manual" in the REASON column by the nodes in the range for maintenance work.

```
# pashowclst -c clstname -v
```

clstname: Cluster name

[By using another node to re-execute running jobs]

Suspend the running job, and re-execute it on another node.

1. Notifying the end users of maintenance

   Notify the end users about isolating the target nodes from job operation for the purpose of maintenance.

Since this procedure both isolates a node from operation and transitions it to software maintenance mode, see "3.3.2 Transitioning to Software Maintenance Mode."

[By setting a job deadline schedule]

1. Notifying the end users of maintenance

   Notify the end users about setting a deadline schedule on the target nodes for the purpose of maintenance.

2. Setting a deadline schedule

   Set a deadline schedule on the nodes in the range for maintenance work.
   For details on how to set a deadline schedule, see "Cluster deadline scheduling management" in the "Job Operation Software Administrator's Guide for Job Management."

3. Checking the status of the deadline schedule

   Using the pashowclst command on the system management node, confirm state transitions to "Running" in the STATUS column and "DeadlineSchedule" in the REASON column by the nodes in the range for maintenance work.

   The following example checks the status of specific nodes.

```
# pashowclst -c clstname -n nodeid1,nodeid2,...
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE        NODETYPE  STATUS   REASON            PWR_STATUS  ARCH_STATUS  SRV_STATUS
nodeid1     CN        Running  DeadlineSchedule  on          -
PLE(o),NRD(o),FEFS(o),PWRD(o)
[ NODE: nodeid2 ]
NODE        NODETYPE  STATUS   REASON            PWR_STATUS  ARCH_STATUS  SRV_STATUS
nodeid2     CN        Running  DeadlineSchedule  on          -
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   nodeid1, nodeid2: Node IDs of maintenance target nodes

   Subsequently, the nodes will enter the period of the set deadline schedule. Then, confirm state transitions to "Disable" in the STATUS column and "DeadlineSchedule" in the REASON column by the compute nodes in the range for maintenance work.

3.3.1.2 Isolating Redundantly Configured Nodes From Operation

Management nodes (compute cluster management node, compute cluster management sub node, and storage cluster management node) may be in a redundant configuration. If the maintenance target is the active management node in that configuration, you need to switch it with the standby node so that maintenance is performed on the standby node.

If the target is a storage cluster management node that also serves as a system management node, see "3.3.2.2 Transitioning the System Management Node to Software Maintenance Mode."

Perform this work on the active system management node.

1. Checking whether the node is active or standby

   Check whether the maintenance target node is the active node or standby node.

   Check the display by the pashowclst command. In the STATUS column, "(A)" indicates an active node, and "(S)" indicates a standby node.

```
# pashowclst -c clstname -m --nodetype nodetype
[ CLST: clstname ]
[ NODETYPE: nodetype ]
NODEGRP    NODETYPE      NODE        STATUS  NODE        STATUS
MASTERGRP  nodetype      nodeid1     (A)     nodeid2     (S)
```

clstname: Cluster name
nodetype: Node type of the maintenance target node
nodeid1, nodeid2: Node IDs of redundantly configured nodes

2. Switching from the active node to the standby node

   If the maintenance target node is the active node, switch it to the standby node by using the --failover option in the paclstmgr command.

```
# paclstmgr -c clstname --failover nodeid1
```

clstname: Cluster name
nodeid1: Node ID of a redundantly configured node

After the switching, confirm that the status is as shown below.

```
# pashowclst -c clstname -m --nodetype nodetype
[ CLST: clstname ]
[ NODETYPE: nodetype ]
NODEGRP    NODETYPE      NODE        STATUS  NODE        STATUS
MASTERGRP  nodetype      nodeid1     (S)     nodeid2     (A)
```

clstname: Cluster name
nodetype: Node type of the maintenance target node
nodeid1, nodeid2: Node IDs of redundantly configured nodes

3.3.1.3 Isolating a File Server Node From Operation

Replicate the FEFSSR service on a file server node (MGS node, MDS node, OSS node) in a mutual standby configuration to the paired node in the configuration.

Perform this work on the active system management node.

1. Replicating the FEFSSR service

   Using the --failover option in the paclstmgr command, replicate the FEFSSR service on the maintenance target node to the paired node in the mutual standby configuration.

```
# paclstmgr -c clstname --failover nodeid1
```

clstname: Cluster name
nodeid1: Node ID of a redundantly configured node

(*) The node ID displayed at this time (*nodeid2*) is the destination node for the replicated service.

Confirm that the end code of the paclstmgr command is 0.

2. Checking the status of the FEFSSR service (in a multi-rack configuration)

   Using the -v option in the pashowclst command, confirm that replication of the FEFSSR service has completed. If the SRV_STATUS column shows "w" for the FEFSSR service on the target node, the service has been replicated on the paired node in the mutual standby configuration. In other words, the target node has switched to standby.

```
# pashowclst -c clstname -n nodeid1
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid1   MGS        SoftError  SrvDown  on           -              FEFSSR(w)
```

clstname: Cluster name
nodeid1: Node ID of a redundantly configured node

Information

If a fault is found in the target node before replication of the FEFSSR service, the node is forcibly stopped. In this case, execute the pashowclst command with the -v option, and confirm that the node has stopped. If the PWR_STATUS column shows "off" for the node, it has stopped.

Note

The MGS and MDS nodes may be configured redundantly as pairs of active and standby nodes. Replicate the FEFSSR service only if the maintenance target node is the active node.
For the procedure for checking whether the maintenance target is the active or standby MGS/MDS node, see "Switching a Node" in the "Job Operation Software Administrator's Guide for System Management."

You can also check the destination node for the replicated service to see whether or not the FEFSSR service has been replicated. If the SRV_STATUS column shows "*" for the FEFSSR service on that node, the node is operating with the replicated service.

```
# pashowclst -c clstname -n nodeid2
[ CLST: clstname ]
[ NODE: nodeid2 ]
NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
nodeid2   MGS        Running    -        on           -              FEFSSR(*)
```

clstname: Cluster name
nodeid2: Node ID of a redundantly configured node

3.3.2 Transitioning to Software Maintenance Mode

Transition the maintenance target node to software maintenance mode, and stop the Job Operation Software service on the node to enable maintenance operations there.

If the maintenance range includes the system management node, first transition the nodes in the cluster to software maintenance mode, and then transition the system management node to software maintenance mode.

Note

The FX server internally frees the target node's job memory during the transition to software maintenance mode. At the time of a software maintenance, be sure to specify the --soft-mainte option for paclstmgr and confirm that you are in software maintenance mode.

If you restart with the papwrctl command or the like in the software maintenance mode state, the service starts, and the job memory is acquired. If you do not release the software maintenance mode state and want to apply a patch to the kernel and then reboot, and then continue another patch application, execute the paclstmgr command again with the --soft-mainte option after the reboot to free up job memory.

3.3.2.1 Transitioning the Nodes in a Cluster to Software Maintenance Mode

This section shows the procedure for transitioning the nodes in a cluster to software maintenance mode.
Perform this work on the active system management node.

1. Transitioning to software maintenance mode

   Using the --soft-mainte option in the paclstmgr command, transition the target nodes to software maintenance mode.

```
# paclstmgr -c clstname -n nodeid1,nodeid2,... --soft-mainte --service-stop
```

clstname: Cluster name
nodeid1, nodeid2: Node IDs of maintenance target nodes

Information

- Depending on the maintenance tasks, such as applying some packages, you may be able to perform maintenance without stopping the Job Operation Software service. If so, specify the --no-service-stop option in the --soft-mainte option in the paclstmgr command so that you can perform maintenance without stopping the Job Operation Software service.

  To determine the necessity of restarting the service when applying a Job Operation Software package, see "3.1.2 Applying Packages Through Rolling Updates."

- To perform the work in [By using another node to re-execute running jobs] in "3.3.1.1 Isolating a Compute Node From Operation ," specify the -k option in addition to those in the command line shown above.

2. Confirming software maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "SoftMaintenance" in the REASON column by the maintenance target nodes.

```
# pashowclst -c clstname -v
[ CLST: clstname ]
NODE        NODETYPE    STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
...
0xFFFF0010  CN          Disable   SoftMaintenance   on            -              -
0xFFFF0011  CN          Disable   SoftMaintenance   on            -              -
...
```

   clstname: Cluster name

Information

If you specified the --no-service-stop option in step 1, the Job Operation Software service does not stop, so the SRV_STATUS column shows that the service is running.

```
# pashowclst -c clstname -v
[ CLST: clstname ]
NODE        NODETYPE    STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
...
0xFFFF0010  CN          Disable   SoftMaintenance   on            -
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFFFF0011  CN          Disable   SoftMaintenance   on            -
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name

3. Checking the job memory free status (FX server only)

   Check that the job memory is freed, and that the memory limit is unlimited.

   This procedure is not required for PRIMERGY servers.

   The following is how to check.

   - Checking the job memory status

     Check that the job memory is released at the compute node.

```
# pmexe -c clstname scope --nodetype CN --stdout \
--compare "cat /proc/self/status | grep Mems_allowed_list"
```

     clstname: Cluster name
     scope: Range options for limiting the range

Check the job memory free status according to the displayed information. In this case, the following information is displayed:

- `Mems_allowed_list: 4-7`
  Job memory is released correctly
- `Mems_allowed_list: 0-3`
  Job memory is not released correctly

- Checking memory limits

  Check that the compute node has unlimited memory.

```
# pmexe -c clstname scope --nodetype CN --stdout \
--compare "cat /sys/fs/cgroup/memory/user.slice/memory.limit_in_bytes"
```

  clstname: Cluster name
  scope: Range options for limiting the range

  Check the memory limit status according to the displayed information. In this case, the following information is displayed:

  - `9223372036854710272`
    Memory limit is unlimited
  - Other than `9223372036854710272`
    Memory limit is in effect

  If the job memory is not released correctly or the memory limit is not unlimited, restart the OS and confirm that the Job Operation Software service has started, and transit software maintenance mode (Use the paclstmgr command with the --soft-mainte option).

Note

If a node in the range for maintenance work has a user-defined service, transition the node to software maintenance mode, and then manually stop the user-defined service. For details on user-defined services, see "Settings of the User-Defined Service" in the "Job Operation Software Administrator's Guide for System Management."

3.3.2.2 Transitioning the System Management Node to Software Maintenance Mode

When you perform maintenance on the system management node, the maintenance range described in "3.1.1 Maintenance Work Range" may be the following:

- System
- Cluster
- Node (redundant configuration)

Perform this work on the active system management node.

[If the system management node is redundant]

System management node maintenance is performed on the standby node.

1. Transitioning to software maintenance mode

   Using the --soft-mainte option in the paclstmgr command, transition the target node to software maintenance mode.

```
# paclstmgr -c clstname -n nodeid --soft-mainte --service-stop
```

   clstname: Cluster name
   nodeid: Node ID of the target node

Information

Depending on the maintenance tasks, including applying some packages, you may be able to perform maintenance without stopping the Job Operation Software service. If so, specify the --no-service-stop option in the --soft-mainte option in the paclstmgr command so that you can perform maintenance without stopping the Job Operation Software service.

To determine the necessity of restarting the service when applying a Job Operation Software package, see "3.1.2 Applying Packages Through Rolling Updates."

2. Confirming software maintenance mode

   Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "SoftMaintenance" in the REASON column by the standby system management node.

```
# pashowclst -c clstname -n 0x00000002
[ CLST: clstname ]
[ NODE: 0x00000002 ]
NODE        NODETYPE    STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0x00000002  SMM         Disable   SoftMaintenance   on            -              -
```

   clstname: Cluster name

Information

If you specified the --no-service-stop option in step 1, the Job Operation Software service does not stop, so the SRV_STATUS column shows that the service is running.

```
# pashowclst -c clstname -n 0x00000002
[ CLST: clstname ]
[ NODE: 0x00000002 ]
NODE        NODETYPE    STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
0x00000002  SMM         Disable   SoftMaintenance   on            -              PWRD(o)
```

   clstname: Cluster name

[If the system management node is not redundant]

Stop job operations in the entire system, and then perform the following procedure.

1. Stopping the Job Operation Software service of the system management node

   System managed node cannot be transitioning to software maintenance mode.
   Stop the Job Operation Software services on the system management node with the paclstmgr command --service option.

```
# paclstmgr -c clstname -n nodeid --service stop
```

   clstname: Cluster name
   nodeid: Node ID of the target node

3.3.3 Stopping the FEFS Service

Perform the steps shown below when applying the following packages:

- FEFS client package
- FEFS server package

If you are not applying any of the above packages, these steps are unnecessary.

Execute the fefs_sync command to stop the FEFS service on the nodes in the range for maintenance work.
Perform the following work as required. The steps performed will vary depending on the applied package.

See

For details on the fefs_sync command, see the *FEFS User's Guide*.

Perform this work on the active system management node.

- When applying the FEFS client package

  Stop the FEFS service on the maintenance target nodes in the compute cluster or multiuse cluster in order to apply the FEFS client package. If you are not applying the FEFS client package, this step is unnecessary.
  Using the --stop option in the fefs_sync command, stop the service on the nodes in the range for maintenance work.

```
# fefs_sync --stop --compute=clstname --nodelist=filename
```

  clstname: Compute or multiuse cluster name
  filename: File listing node IDs

- When applying the FEFS server package

  To perform maintenance in the unit of a file system as described in "3.1.1 Maintenance Work Range," perform these steps because the file system must be stopped for maintenance. If you are not applying the FEFS server package, these steps are unnecessary.

  Execute the fefs_sync command to unmount the maintenance target file system.

  Note

  When stopping the file system, notify users so that they do not execute jobs using this file system.

  Using the --umount option in the fefs_sync command, unmount the maintenance target file system.

```
# fefs_sync --umount --compute=clstname --fsname=fsname1
# fefs_sync --umount --compute=clstname --fsname=fsname2
...
```

  clstname: Compute or multiuse cluster name
  fsname1, *fsname2*: Names of the maintenance target file systems

  Stop the FEFS service and FEFSSR service on the maintenance target nodes.
  Using the --stop option in the fefs_sync command, stop the services on the nodes in the range for maintenance work.

```
# fefs_sync --stop --storage=clstname --nodelist=filename
```

  clstname: Storage cluster name
  filename: File listing node IDs

3.3.4 Starting TFTP and HTTP Services

Performe this procedure to apply or remove packages, back up or restore a node, or rebuild a node.
If not, this procedure is not required.

```
# pmexe -c clstname --nodetype SMM,CCS "systemctl start httpd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl start xinetd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl enable httpd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl enable xinetd"
```

clstname: Compute cluster name

3.4 Software Maintenance Work

Perform software maintenance work.

To perform the work listed below, see the corresponding sections.

Table 3.4 Software Maintenance

Maintenance Work	Corresponding Section
Apply or delete package	3.6 Applying or Deleting a Package
Update configuration file	3.7 Changing System Settings
Change system configuration	Chapter 4 Changing the System Configuration
Back up and restore node	Chapter 6 Backing Up and Restoring a Node

3.5 Incorporating Into Operation After Software Maintenance

The maintenance target has been isolated from operation in the preparation for software maintenance. After software maintenance, incorporate it into operation.
This section describes procedures for incorporating the maintenance target into operation.
Perform the work on the active system management node, unless otherwise noted.

3.5.1 Stopping TFTP and HTTP Services

If you performed the procedure in "3.3.4 Starting TFTP and HTTP Services" before performing any maintenance, stop the TFTP and HTTP services by performing the following work.

```
# pmexe -c clstname --nodetype SMM,CCS "systemctl stop httpd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl stop xinetd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl disable httpd"
# pmexe -c clstname --nodetype SMM,CCS "systemctl disable xinetd"
```

clstname: Compute cluster name

3.5.2 Starting the FEFS Service

Start the FEFS service on nodes in the range of the maintenance done.
If you performed the procedure in "3.3.3 Stopping the FEFS Service " before maintenance work, recover FEFS by performing the following work.

Note

- If the [After updating] item of the applied package shows "reboot" or "cold reboot," this step is unnecessary.
  Proceed to "3.5.3 Returning From Software Maintenance Mode" to recover the maintenance target nodes.
- If you applied the FEFS server package and the FEFS client package at the same time, perform [When the FEFS server package has been applied], then [When the FEFS client package has been applied] to start the FEFS service.

Perform this work on the active system management node.

- When the FEFS server package has been applied

  If you performed maintenance on a file server node, perform these steps to recover the node.

  Start the FEFSSR service on the maintenance target nodes.
  Using the --start option in the fefs_sync command, start the service on the nodes in the range for maintenance work.
  The following example executes the fefs_sync command on specific nodes.

  ```
  # fefs_sync --start --storage=clstname --nodeid=nodeid1,nodeid2,...
  ```

  clstname: Cluster name
  nodeid1, nodeid2: Node IDs
  Execute the following commands to mount the maintenance target file systems.

  Using the --mount option in the fefs_sync command, mount the maintenance target file systems.
  The following example executes the fefs_sync command on a specific boot group.

```
# fefs_sync --mount --compute=clstname --fsname=fsname1
# fefs_sync --mount --compute=clstname --fsname=fsname2
...
```

clstname: Cluster name
fsname1, *fsname2*: Names of the maintenance target file systems

- When the FEFS client package has been applied

  Start the FEFS service and FEFSSR service on the maintenance target nodes.
  Using the --start option in the fefs_sync command, start the services on the nodes in the range for maintenance work.

  The following example executes the fefs_sync command on a specific boot group.

```
# fefs_sync --start --compute=clstname --bootgrp bootgid
```

clstname: Cluster name
bootgid: Boot group ID

3.5.3 Returning From Software Maintenance Mode

3.5.3.1 Recovering the System Management Node

[If the system management node is redundant]

If the maintenance targets include the system management node, recover the standby system management node from software maintenance mode.
Using the --recover option in the paclstmgr command, incorporate the maintenance targets into operation.
At this time, select a node recovery method by specifying an additional option according to the necessity of restarting the maintenance targets.

Table 3.5 Sub Options of the --recover Option

Sub Option	Recovery Method
--no-restart	Incorporates the nodes without a restart. If the Job Operation Software service has been stopped, the service does not start.
--service-restart	Restarts the service and incorporates the nodes into operation.
--cold-reboot Or sub option omitted	Performs a cold reboot and incorporates the nodes into operation.

If the Job Operation Software service was stopped during a transition to software maintenance mode, do not specify the --no-restart option. Otherwise, the maintenance mode is released while the Job Operation Software service stays stopped.

Perform this work on the active system management node.

The following example incorporates the standby system management node *nodeid* into operation through a cold reboot.

```
# paclstmgr -c clstname -n nodeid --recover --cold-reboot
```

clstname: Cluster name
nodeid: Node ID

Using the pashowclst command, confirm a state transition by the target node to "Running" in the STATUS column.

```
# pashowclst -c clstname -v --nodetype SMM
[ CLST: clstname ]
[ NODETYPE: SMM ]
NODE         NODETYPE  STATUS     REASON       PWR_STATUS    ARCH_STATUS    SRV_STATUS
```

```
0x00000001   SMM       Running    -           on          -           PWRD(o)
0x00000002   SMM       Running    -           on          -           PWRD(o)
```

clstname: Cluster name

When all the services on a target node are running, the STATUS column shows "Running" for the node.

[If the system management node is not redundant]

Since the system management function command cannot be used when the Job Operation Software service is stopped, restart the OS with the shutdown command.

```
# shutdown -r now
```

Note

if the system uses a single node serving as all of the system management node, compute cluster management node, and login node, the Job Operation Software service on all compute nodes in the cluster must be restarted for recovery.

See "How to Restart the Node Serving as All of System Management Node, Compute Cluster Management Node, and Login Node" in "Chapter 3 Maintenance Work Problems" of the manual "Job Operation Software Troubleshooting" to restart the Job Operation Software service on all compute nodes in the cluster and incorporate compute nodes into operation.

3.5.3.2 Recovering the Nodes in a Cluster

Return the nodes in the cluster back from software maintenance mode. If the maintenance targets include a storage cluster, perform work in the following order:

1. Storage cluster
2. Compute cluster and multiuse cluster

Note

Using the pashowclst command, confirm that the STATUS column shows "Running" for all the nodes in the storage cluster, before recovering the compute cluster.

Using the --recover option in the paclstmgr command, incorporate the maintenance targets into operation.

At this time, select a node recovery method by specifying an additional option according to the necessity of restarting the maintenance targets.

Table 3.6 Sub Options of the --recover Option

Sub Option	Recovery Method
--no-restart	Incorporates the nodes without a restart. If the Job Operation Software service has been stopped, the service does not start.
--service-restart	Restarts the service and incorporates the nodes into operation.
--cold-reboot Or sub option omitted	Performs a cold reboot and incorporates the nodes into operation.

Note

- If the Job Operation Software service was stopped during a transition to software maintenance mode, do not specify the --no-restart option. Otherwise, the maintenance mode is released while the Job Operation Software service stays stopped.

- Each sub-option of the --recover option behaves differently if the specified range of nodes includes nodes that are not in software maintenance mode.
  See "Recovering a Node" in the "Job Operation Software Administrator's Guide for System Management."

Before incorporating the target nodes into operation, use the software environment check function to confirm that the nodes are ready for operation.

The following example shows steps up to the step of incorporating all the nodes in a cluster into operation.

Perform this work on the active system management node.

1. Releasing software maintenance mode

   Release software maintenance mode on the maintenance target nodes, and power on the nodes or restart services.

   When releasing software maintenance mode, keep compute nodes isolated from operation (Disable state). In step 3 below, you will be checking whether the nodes in the target CMU can be incorporated into operation.

   Specify the --disable option in addition to the --recover option in the paclstmgr command to release software maintenance mode.

   The following example starts the entire cluster through a cold reboot.

   The cluster starts at this time with compute nodes remaining in the Disable state.

   ```
   # paclstmgr -c clstname -a --recover --cold-reboot --disable
   ```

   clstname: Cluster name

   If you performed maintenance in units of GIO group, execute the following command line.

   ```
   # paclstmgr -c clstname -n nodeid --giogrp --recover --cold-reboot --disable
   ```

   clstname: Cluster name
   nodeid: Node ID of the maintenance target node

2. Confirming recovery from software maintenance mode

   Confirm that software maintenance mode has been released and that the nodes are powered on or services have restarted. For the nodes, confirm that the pashowclst command shows "Disable" in the STATUS column and "Manual" in the REASON column.

   ```
   # pashowclst -c clstname -v --bootgrp bootgid
   [ CLST: clstname ]
   [ BOOTGRP: bootgid ]
   NODE         NODETYPE  STATUS   REASON      PWR_STATUS    ARCH_STATUS    SRV_STATUS
   0xFF010001   BIO,CN    Disable  Manual      os-running    ICC_Running
   PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
   0xFF010002   SIO,CN    Disable  Manual      os-running    ICC_Running
   PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
   0xFF010003   GIO,CN    Disable  Manual      os-running    ICC_Running
   PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
   0xFF010004   CN        Disable  Manual      os-running    ICC_Running
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   0xFF010005   CN        Disable  Manual      os-running    ICC_Running
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   0xFF010006   CN        Disable  Manual      os-running    ICC_Running
   PLE(o),NRD(o),FEFS(o),PWRD(o)
   ...
   ```

   clstname: Cluster name
   bootgid: Boot group ID

3. Checking the software environment

   From the software perspective, check whether the started nodes (the maintenance targets) are ready to be incorporated into operation. Execute the pachecker command to check the environment.
   For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."

4. Incorporating into operation

   Using the --enable option in the paclstmgr command, incorporate the targets into operation. The targets are the compute nodes in the Disable state at this point in time.

   The following example incorporates a specific boot group into operation.

```
# paclstmgr -c clstname --bootgrp bootgid --enable
```

   clstname: Cluster name
   bootgid: Boot group ID

   Note

   If you performed maintenance in units of GIO group, execute the following command line.

```
# paclstmgr -c clstname --enable -n nodeid --giogrp
```

   clstname: Cluster name
   nodeid: Node ID within the maintenance target CMU

5. Checking the status

   Using the pashowclst command, confirm state transitions by the target nodes to "Running" in the STATUS column.

   The following example checks the status of a specific boot group.

```
# pashowclst -c clstname -v --bootgrp bootgid
[ CLST: clstname ]
[ BOOTGRP: bootgid ]
NODE         NODETYPE  STATUS     REASON    PWR_STATUS    ARCH_STATUS    SRV_STATUS
0xFF010001   BIO,CN    Running    -         os-running    ICC_Running
PLE(o),SRD(o),NRD(o),FEFS(o),PWRD(o)
0xFF010002   SIO,CN    Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010003   GIO,CN    Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFSSR(o),FEFS(o),PWRD(o)
0xFF010004   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010005   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
0xFF010006   CN        Running    -         os-running    ICC_Running
PLE(o),NRD(o),FEFS(o),PWRD(o)
...
```

   clstname: Cluster name
   bootgid: Boot group ID

   When all the services on a target node are running, the STATUS column shows "Running" for the node.

To perform a rolling update on the nodes in a redundant configuration, recover the node where maintenance was done so that it is the active node again, and then perform maintenance on the node in the next maintenance range.
Perform work from "3.3 Preparation for Software Maintenance" on the node in the next maintenance range.

Note

If the failover function is suppressed, enable the failover function.

For information about enabling failover, see "Setting Automatic Failover" in the "Job Operation Software Administrator's Guide for System Management."

3.5.4 Failback the System

If you replicated the FEFSSR service in "3.3.1.3 Isolating a File Server Node From Operation", perform the following steps to failback the system.

1. Recovering the replicated service

   Recover the service on the node not currently running the service, by using the --failback option in the paclstmgr command.

   ```
   # paclstmgr -c clstname --failback nodeid1
   ```

   clstname: Cluster name
   nodeid1: Node ID of the maintenance target node (The node ID of a node that became the standby system node when the --failover option in the paclstmgr command was executed)

2. Checking the node status

   Using the pashowclst command, confirm state transitions to "Running" in the STATUS column by the target node and to "o" in the SRV_STATUS column by the FEFSSR service.

   ```
   # pashowclst -c clstname -n nodeid1,nodeid2
   [ CLST: clstname ]
   [ NODE: nodeid1 ]
   NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
   nodeid1   nodetype   Running    -        on           -              FEFSSR(o)
   [ NODE: nodeid2 ]
   NODE      NODETYPE   STATUS     REASON   PWR_STATUS   ARCH_STATUS    SRV_STATUS
   nodeid2   nodetype   Running    -        on           -              FEFSSR(o)
   ```

   clstname: Cluster name
   nodeid1,*nodeid2*: Node IDs of the target file server node
   nodetype: Target node type

   When all the services on a target node are running, the STATUS column shows "Running" for that node.

3.6 Applying or Deleting a Package

This section describes procedures for applying and deleting a package.

Information

This procedure uses the yum command as an example. If you want to use another package manager, follow the instructions for using that package manager.

3.6.1 Applying a Package

Use a repository created by the installation function to apply a package.
This section describes procedures for registering a package with a repository and applying the package.

3.6.1.1 Registering a Package

Register a package with a repository created by the installation function.
Perform this work on the active system management node.

1. Confirming the path to the package storage directory

   Using the repo subcommand in the pxinst command, confirm the repository for registering packages.

```
# pxinst repo list
REPONAME      PATH                              PKGDIR     OS      VERSION    ARCH
--------      ----                              ------     --      -------    ----
reponame      repopath                          dirname    os      version    arch
...
```

reponame: Repository name
repopath: Path to the repository
dirname: Directory where packages are located in the repository
os: OS type
version: OS version
arch: Architecture

Check the line showing the repository used by nodes to apply packages.
"*repopath*/*dirname*" is the storage path of the packages.

Information

Depending on the OS version, PKGDIR displays multiple directories separated by commas.

```
# pxinst repo list
REPONAME      PATH                    PKGDIR                              OS      VERSION  ARCH
--------      ----                    ------                              --      -------  ----
rhel80        /repos/rhel80            AppStream/Packages,BaseOS/Packages      rhel    8.0
aarch64
```

In the sample output above, either "/ repos/rsel 80/AppStream/Packages" or "/ repos/rsel 80/BaseOS/Packages" is the path where the package is stored.

When a job operation software package or a user adds a package, store it in the directory shown at the top "/ repos/rsel 80/AppStream/Packages".

If you are adding an OS update package, store it in the distributor's public repository.

2. Adding a package

   Place the package in the directory confirmed in step 1.

```
# cp pkgname repopath/dirname
```

pkgname: Package name
repopath: Path of the repository
dirname: Directory where stores packages in a repository

[Procedure for Update the module definition file]

If you add an OS update package of RedHat to the AppStream/Packages directory of the RHEL 8 repository, you also need to update the module definition file as well as add the package.

Before updating the module definition file, create a local mirror of the Red Hat repository on the subscripted server.

Extract the module definition file from the file on the local mirror of the Red Hat repository and place it in the target repository on the system management node.

The following example expands the module definition file on the subscripted server and places it in the target repository on the system management node.

```
# ssh server_ip zcat mirror_path/repodata/*-modules.yaml.gz > repopath/pxinst/
modules_AppStream.yaml
```

server_ip: IP address of the subscripted server
mirror_path: Local mirror path of the AppStream repository
repopath: Repository path of the system management node

See

For subscription servers, see "Appendix C Preparing Subscripted Servers" in "Job Operation Software Setup Guide."

3. Updating the repository

   Use the repo subcommand in the pxinst command to update the repository.
   This operation updates the metadata of the repository.

```
# pxinst repo update -R reponame
[INFO] INST 0800 pxinst repo completed successfully.(update repository)
```

   reponame: Repository name

4. Synchronizing the repository

   Reflect the changes in the repository to the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync repo
...
[INFO] INST 1805 pxinst File replication completed.
```

3.6.1.2 Deleting Cached Information

Before applying a package to the target node, use the pmexe command to delete cached information that is stored by package manager on the node.
Perform this work on the system management node.
If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.

The following is an example of deleting cache information.

```
# pmexe --chroot -c clstname scope "yum clean all"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

clstname: Cluster name
scope: Range options for limiting the range
resultfile: Execution result file
Check the execution result file to verify that no errors are output in the command results for each node.

Note

Because the diskless node of the FX server NFS-mounts the space on the boot I/O node as rootfs, rpmdb can become corrupted when you run commands that manipulate rpmdb, such as the rpm or yum commands.

To reduce the possibility of rpmdb corruption, if you want to run commands that manipulate rpmdb, such as the rpm or yum commands, use the pmexe command with the --chroot option.

Also, because the pachecker command references rpmdb internally, do not run the pachecker command while running commands that manipulate rpmdb, such as the rpm or yum commands.

3.6.1.3 Stopping the Compute Node

If the target includes an FX server, use the papwrctl command to power off the compute node.
After stopping the destination node, specify the -v option of the pashowclst command to confirm that the PWR_STATUS column for the destination node is set to "off".

```
# papwrctl -c clstname scope --model FT --nodetype CN --excludetype BIO off
```

clstname: Cluster name
scope: Range options for limiting the range

3.6.1.4 Applying a Package

Use the pmexe command to apply a package on the target node.
Perform this work on the system management node.
If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.
The following example applies the FJSVpxinst package.

```
# pmexe --chroot -c clstname scope "yum install FJSVpxinst -y"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

clstname: Cluster name
scope: Range options for limiting the range
resultfile: Execution result file
Check the execution result file to verify that no errors are output in the command results for each node.

Note

Because the diskless node of the FX server NFS-mounts the space on the boot I/O node as rootfs, rpmdb can become corrupted when you run commands that manipulate rpmdb, such as the rpm or yum commands.

To reduce the possibility of rpmdb corruption, if you want to run commands that manipulate rpmdb, such as the rpm or yum commands, use the pmexe command with the --chroot option.

Also, because the pachecker command references rpmdb internally, do not run the pachecker command while running commands that manipulate rpmdb, such as the rpm or yum commands.

3.6.1.5 Starting the Compute Node

If the target includes an FX server, use the papwrctl command to power on the compute node.
After starting the destination node, specify the -v option of the pashowclst command to confirm that the PWR_STATUS column for the destination node is set to "os-running".

```
# papwrctl -c clstname scope --model FT --nodetype CN --excludetype BIO on
```

clstname: Cluster name
scope: Range options for limiting the range

Note

If you want to continue software maintenance, use the paclstmgr command to instruct the compute node to go back to the software maintenance state and stop the service.

This is to stop again services and job memory settings that were restarted by node startup.

```
# paclstmgr -c clstname scope --model FT --nodetype CN --excludetype BIO --soft-mainte --service-stop
-k
```

clstname: Cluster name
scope: Range options for limiting the range

3.6.2 Updating a Package

Use a repository created by the installation function to update a package.
This section describes procedures for registering a package with a repository and updating the package.

3.6.2.1 Registering a Package

Perform the procedures in "3.6.1.1 Registering a Package".

3.6.2.2 Deleting Cached Information

Perform the procedures in "3.6.1.2 Deleting Cached Information".

3.6.2.3 Stopping the Compute Node

Perform the procedures in "3.6.1.3 Stopping the Compute Node".

3.6.2.4 Updating a Package

Use the pmexe command to update a package on the target node.
Perform this work on the system management node.
If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.
The following example updates the FJSVpxinst package.

```
# pmexe --chroot -c clstname scope "yum update FJSVpxinst -y"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

clstname: Cluster name
scope: Range options for limiting the range
resultfile: Execution result file
Check the execution result file to verify that no errors are output in the command results for each node.

Note

Because the diskless node of the FX server NFS-mounts the space on the boot I/O node as rootfs, rpmdb can become corrupted when you run commands that manipulate rpmdb, such as the rpm or yum commands.

To reduce the possibility of rpmdb corruption, if you want to run commands that manipulate rpmdb, such as the rpm or yum commands, use the pmexe command with the --chroot option.

Also, because the pachecker command references rpmdb internally, do not run the pachecker command while running commands that manipulate rpmdb, such as the rpm or yum commands.

Note

If you update the FJSVpxinst package or update the kernel-related OS update package from Red Hat, you must re-create the initrd (initial RAM disk for OS installation) that is required to install the OS.
Perform the following procedures.

1. Creating an initrd
   Re-create the initrd for all repositories that have updated the package.
   For the procedure, see "Creating the initrd [SMM]" in "Job Operation Software Setup Guide."

2. Configuring ServerView Suite
   Re-import the kernel modules and supplied software provided by the ServerView Suite required by the PRIMERGY server.
   Perform this procedure only if you configured ServerView Suite for the PRIMERGY server repository during deployment.
   For the procedure, see "Configuring ServerView Suite [SMM]" in "Job Operation Software Setup Guide."

Note

If you update the kernel package on FX server, you must update the link to the kernel used by the diskless node with the link to the new kernel before rebooting the system.

Perform the following steps:

```
# pmexe -c clstname --nodetype BIO scope "/opt/FJSVxos/setting/link_for_nio.sh"
```

clstname: Cluster name
scope: Range options for limiting the range

3.6.2.5 Starting the Compute Node

Perform the procedures in "3.6.1.5 Starting the Compute Node".

3.6.3 Deleting a Package

This section describes procedures for deleting a package from the target node and deleting a package located in a repository.

3.6.3.1 Stopping the Compute Node

Perform the procedures in "3.6.1.3 Stopping the Compute Node".

3.6.3.2 Deleting a package

Use the pmexe command to delete a package applied to the target node.
If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.
The following example deletes the FJSVpxinst package.

```
# pmexe --chroot -c clstname scope "yum remove FJSVpxinst -y"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

clstname: Cluster name
scope: Range options for limiting the range
resultfile: Execution result file
Check the execution result file to verify that no errors are output in the command results for each node.

Note

Because the diskless node of the FX server NFS-mounts the space on the boot I/O node as rootfs, rpmdb can become corrupted when you run commands that manipulate rpmdb, such as the rpm or yum commands.

To reduce the possibility of rpmdb corruption, if you want to run commands that manipulate rpmdb, such as the rpm or yum commands, use the pmexe command with the --chroot option.

Also, because the pachecker command references rpmdb internally, do not run the pachecker command while running commands that manipulate rpmdb, such as the rpm or yum commands.

Note

If you delete the kernel package on FX server, you must update the link to the kernel used by the diskless node with the link to the current kernel before rebooting the system.

Perform the following steps:

```
# pmexe -c clstname --nodetype BIO scope "/opt/FJSVxos/setting/link_for_nio.sh"
```

clstname: Cluster name
scope: Range options for limiting the range

3.6.3.3 Starting the Compute Node

Perform the procedures in "3.6.1.5 Starting the Compute Node".

3.6.3.4 Deleting the package from a repository

As required, delete a stored package in the repository.

1. Confirming the directory where the package is located
   Using the repo subcommand in the pxinst command, confirm the repository for registering packages.

```
# pxinst repo list
REPONAME      PATH                          PKGDIR     OS      VERSION    ARCH
--------      ----                          ------     --      -------    ----
reponame      repopath                      dirname    os      version    arch
...
```

   reponame: Repository name
   repopath: Path to the repository
   dirname: Directory where the package is located in the repository
   os: OS type
   version: OS version
   arch: Architecture

   You will be deleting the package from the node. Check the line showing the repository used by the node.
   "*repopath*/*dirname*" is the directory where the package is located.

2. Deleting the package
   Delete the package from the directory located in step 1.

```
# rm repopath/dirname/pkgname
```

   repopath: Path to the repository
   dirname: Directory where the package is located in the repository
   pkgname: Package name

3. Updating the repository
   Use the repo subcommand in the pxinst command to update the repository.
   This operation updates the metadata of the repository.

```
# pxinst repo update -R reponame
[INFO] INST 0800 pxinst repo completed successfully.(update repository)
```

   reponame: Repository name

4. Synchronizing the repository
   Reflect the changes in the repository to the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync repo
[INFO] INST 1805 pxinst File replication completed.
```

3.6.4 Downgrading a Package

This section describes procedures for downgrading a package on the target node and deleting the package from the repository.

3.6.4.1 Stopping the Compute Node

Perform the procedures in "3.6.1.3 Stopping the Compute Node".

3.6.4.2 Downgrading a Package

Use the pmexe command to downgrade a package applied to the target node.
Perform this work on the system management node.
If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.
The following example downgrades the FJSVpxinst package.

```
# pmexe --chroot -c clstname scope "yum downgrade FJSVpxinst -y"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

clstname: Cluster name
scope: Range options for limiting the range
resultfile: Execution result file
Check the execution result file to verify that no errors are output in the command results for each node.

Note

Because the diskless node of the FX server NFS-mounts the space on the boot I/O node as rootfs, rpmdb can become corrupted when you run commands that manipulate rpmdb, such as the rpm or yum commands.

To reduce the possibility of rpmdb corruption, if you want to run commands that manipulate rpmdb, such as the rpm or yum commands, use the pmexe command with the --chroot option.

Also, because the pachecker command references rpmdb internally, do not run the pachecker command while running commands that manipulate rpmdb, such as the rpm or yum commands.

Note

If you downgrade the kernel package on FX server, you must update the link to the kernel used by the diskless node with the link to the new kernel before rebooting the system.

Perform the following steps:

```
# pmexe -c clstname --nodetype BIO scope "/opt/FJSVxos/setting/link_for_nio.sh"
```

clstname: Cluster name
scope: Range options for limiting the range

3.6.4.3 Starting the Compute Node

Perform the procedures in "3.6.1.5 Starting the Compute Node".

3.6.4.4 Deleting the package from a repository

Delete the corresponding version of the package stored in the repository. If you do not delete the package from the repository, the downgraded package will be updated the next time you update the package.

Perform the procedure in "3.6.3.4 Deleting the package from a repository".

3.6.5 OS Minor Version Update

If you are updating a minor OS version on your system, delete an existing repository with a previous minor version and create a new repository and replace the one you want to use.

3.6.5.1 Create Repository

Replace the old minor version repository with the new minor version repository.

1. Confirm Existing Repository

   Using the repo subcommand in the pxinst command, confirm the repository for the OS for which you want to update minor versions.

```
# pxinst repo list
REPONAME PATH     PKGDIR  OS VERSION ARCH
-------- ----     ------  -- ------- ----
reponame repopath dirname os version arch
...
```

   reponame: Repository name

   repopath: Path to the repository

   dirname: Directory where packages are located in the repository

   os: OS type

   version: OS version

   arch: Architecture

   Confirm the "reponame", "repopath", "dirname", and "arch" for the target repository.

2. Backing Up an Existing Repository

   Back up the repository path repopath confirmed in step 1.

   The backup destination bk_path can be any directory.

Note

   If the backup destination is a production system managed node, pay attention to disk usage.

```
# mkdir -p bk_path/
# cp -pr repopath bk_path/
```

   repopath: Path to the repository

   bk_path: Path to backup to

3. Deleting an Existing Repository

   Using the repo subcommand in the pxinst command, delete the repository confirmed in step 1.

```
# pxinst repo delete -R reponame
```

   reponame: Repository name confirmed in step 1

4. Creating a New Repository

   Create a new OS minor version repository.

   The new repository name specifies the repository name reponame from step 1.

   [When the creation repository is used on a System management node]

```
# pxinst repo create -R reponame -S filepath --enable
```

   reponame: Repository name confirmed in step 1

   filepath: Path to the ISO image file for the new minor version of the OS

   [If you do not want to use the creation repository on a System management node]

```
# pxinst repo create -R reponame -S filepath
```

   reponame: Repository name confirmed in step 1

   filepath: Path to the ISO image file for the new minor version of the OS

5. Packaging Registration for the Technical Computing Suite Package Kit

   For the new repository created in step 4, extract the Technical Computing Suite package kit provided on the Technical Computing Suite DVD and copy the packages to the new repository.

```
# mkdir work/
# cd work/
# tar zxf packagekit_path
# find packagekit/ -name "*arch*.rpm" | grep "COMMON" | xargs -i cp {} repopath/dirname
# find packagekit/ -name "*arch*.rpm" | grep "kernel" | xargs -i cp {} repopath/dirname
# find packagekit/ -name "*noarch*.rpm" | grep "arch" | grep "COMMON" | xargs -i cp {} repopath/
dirname
# find packagekit/ -name "*noarch*.rpm" | grep "arch" | grep "kernel" | xargs -i cp {} repopath/
dirname
```

   packagekit_path: Path to the package kit

   repopath: Path to the repository confirmed in step 1

   dirname: Directory where packages are located in the repository confirmed in step 1

   arch: Architecture confirmed in step 1

   kernel: Number of kernel versions for GA version of old repository

Information

Depending on the OS version, the dirname confirmed in step 1 may be more than one directory, for example, "AppStream/Packages, BaseOS/Packages".

In the above case, set "AppStream/Packages" to dirname.

6. Register the RPM package file for the Technical Computing Suite

   Register RPM packages of cumulative patches for the Technical Computing Suite and any additional patches required to update OS minor versions in the new repository.

```
# cp pkgfile repopath/dirname
```

   pkgfile: Technical Computing Suite RPM files

   repopath: Path to the repository confirmed in step 1

   dirname: Directory where packages are located in the repository confirmed in step 1

Information

Depending on the OS version, the dirname confirmed in step 1 may be more than one directory, for example, "AppStream/Packages, BaseOS/Packages".

In the above case, set "AppStream/Packages" to dirname.

7. Updating the repository

   Use the repo subcommand in the pxinst command to update the repository.

   This operation updates the metadata of the repository.

```
# pxinst repo update -R reponame
[INFO] INST 0800 pxinst repo completed successfully.(update repository)
```

   reponame: Repository name confirmed in step 1

8. Synchronizing the repository

   Reflect the changes in the repository to the standby system management node and compute cluster sub management node.

If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync repo
...
[INFO] INST 1805 pxinst File replication completed.
```

3.6.5.2 Updating a Package

Update the package on the target node.

The procedure in this section reboots the target node multiple times for the package update.

1. Deleting Cached Information

   Perform the procedures in "3.6.1.2 Deleting Cached Information".

2. Stopping the Compute Node.

   Note

   This step is necessary if the target includes an FX server.

   Perform the procedures in "3.6.1.3 Stopping the Compute Node".

3. Removing File System Related Packages

   For Kernel updates, use the pmexe command to remove file related packages.

   If the target includes an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.

   ```
   # pmexe --chroot -c clstname scope "yum -q remove -y FJSVfefs*"
   [INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
   ```

   clstname: Cluster name
   scope: Range options for limiting the range
   resultfile: Execution result file

   Check the execution result file to verify that no errors are output in the command results for each node.

4. Removing Interconnect Related Packages

   Remove interconnect related packages.

   [If the target node is an FX server]

   If the target node is an FX server, these steps are unnecessary.

   [If the target node is an PRIMERGY server]

   If the target node is a PRIMERGY server, remove the OFED package.

   Please refer to the OFED manual to remove the OFED package.

5. Stop the target node.

   Stop the target node.

   After stopping the destination node, specify the -v option of the pashowclst command to confirm that

   the PWR_STATUS column for the destination node is set to "off".

   ```
   # papwrctl -c clstname scope off
   [INFO] PSM 2081 papwrctl All power control was executed.
   ```

   clstname: Cluster name
   scope: Range options for limiting the range

6. Boot the target node.

   Boot the target node.

   [If the target node is an FX server]

   If the target node is an FX server, start only the boot I/O node.

```
# papwrctl -c clstname scope --nodetype BIO on
[INFO] PSM 2081 papwrctl All power control was executed.
```

   clstname: Cluster name
   scope: Range options for limiting the range

   [If the target node is not an FX server]

   If the target node is not an FX server, start the target node.

```
# papwrctl -c clstname scope on
[INFO] PSM 2081 papwrctl All power control was executed.
```

   clstname: Cluster name
   scope: Range options for limiting the range

7. Stopping service on the target node

   a. Checking before stopping the service on the target node

      Before stopping the service on the target node, check that the Job Operation Software service has started on the target node.

      Perform the following procedure.

      - When a superior node (CCM/CCS) is in the Running state:

        Using the -v option in the pashowclst command, confirm state transitions to "NRD(o)" in the SRV_STATUS column by the maintenance target nodes.

```
# pashowclst -v
```

      - When a superior node (CCM/CCS) is not in the Running state:

        Using the pmexe command with the systemctl command, confirm the pxnrd service has a startup status of "running".

```
# pmexe -c clstname scope --nodetype CN --stdout \
--compare "systemctl list-units --type=service | grep pxnrd"
```

        clstname: Cluster name
        scope: Range options for limiting the range

   b. Stopping service on the target node

      Transit the target node whose minor version is to be updated again to the software maintenance state.

      This is to stop again services and job memory settings that were restarted by node startup.

      Use the pasclstmagr command with --soft-mainte option for stopping the service.

      [If the target node is an FX server]

      If the target node is an FX server, only the boot I/O node is targeted

```
# paclstmgr -c clstname scope --nodetype BIO --soft-mainte --service-stop -k
```

      clstname: Cluster name
      scope: Range options for limiting the range

      [If the target node is not an FX server]

      If the target node is not an FX server, specify the target node.

```
# paclstmgr -c clstname scope --soft-mainte --service-stop -k
```

clstname: Cluster name
scope: Range options for limiting the range

c. Confirming software maintenance mode

Using the pashowclst command, confirm state transitions to "Disable" in the STATUS column and "SoftMaintenance" in the REASON column by the maintenance target nodes.

```
# pashowclst -c clstname -v
[ CLST: clstname ]
NODE        NODETYPE    STATUS    REASON            PWR_STATUS    ARCH_STATUS    SRV_STATUS
...
0xFFFF0010  CN          Disable   SoftMaintenance   on            -              -
0xFFFF0011  CN          Disable   SoftMaintenance   on            -              -
...
```

clstname: Cluster name

d. Checking after stopping the service on the target node

Check that the job memory is freed, and that the memory limit is unlimited.

Perform the following procedure.

1. Checking the job memory status

Check that the job memory is released at the compute node.

```
# pmexe -c clstname scope --nodetype CN --stdout \
--compare "cat /proc/self/status | grep Mems_allowed_list"
```

clstname: Cluster name
scope: Range options for limiting the range

Check the job memory free status according to the displayed information. In this case, the following information is displayed:

- `Mems_allowed_list: 4-7`
  Job memory is released correctly
- `Mems_allowed_list: 0-3`
  Job memory is not released correctly

2. Checking memory limits

Check that the compute node has unlimited memory.

```
# pmexe -c clstname scope --nodetype CN --stdout \
--compare "cat /sys/fs/cgroup/memory/user.slice/memory.limit_in_bytes"
```

clstname: Cluster name
scope: Range options for limiting the range

Check the memory limit status according to the displayed information. In this case, the following information is displayed:

- `9223372036854710272`
  Memory limit is unlimited
- Other than `9223372036854710272`
  Memory limit is in effect

If the job memory is not released correctly or the memory limit is not unlimited, restart the OS and confirm that the Job Operation Software service has started, and transit software maintenance mode (Use the paclstmgr command with the --soft-mainte option).

8. Deleting Cached Information

Perform the procedures in "3.6.1.2 Deleting Cached Information".

If the target is an FX server, the diskless node is down, but the task is to specify the --chroot option in the pmexe command, so the specified command runs on the boot I/O node.

9. Minor version updates

   Use the pmexe command to update the packages on the target node.

   If the target node is an FX server, update the boot I/O node and the diskless node separately to

   reduce the load on the boot I/O node.

   [If the target node is an FX server]

   If the target node is an FX server, use the --chroot option and run the command on the boot I/O node, chroot into the NFS area that the compute node mounts.

```
# pmexe --chroot -c clstname scope --nodetype BIO "yum update -y -q"
# pmexe --chroot -c clstname scope --nodetype CN --excludetype BIO "yum update -y -q"
```

   clstname: Cluster name
   scope: Range options for limiting the range

   [If the target node is not an FX server]

```
# pmexe -c clstname scope "yum update -y -q"
```

   clstname: Cluster name
   scope: Range options for limiting the range

   Note

   If many messages are output per node for commands executed by the pmexe command, the pmexe command execution does not run in parallel and takes a long time.
   Because updating the OS minor version causes the yum command to output too many messages, specify the -q option to the yum command or redirect stdout and stderr to a file on each node to suppress messages output. If the yum command does not suppress messages output, it might take 10 hours or more for the pmexe command to return.
   Do not specify the -d or -v option to the yum command.

10. Configuring Diskless Nodes

    Perform this procedure only if your FX server has been updated to a new OS minor version.

    The link to the kernel used by the diskless node must be updated with the link to the new kernel before the system reboots.

    Do the following.

```
# pmexe -c clstname --nodetype BIO scope "/opt/FJSVxos/setting/link_for_nio.sh"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

    clstname: Cluster name
    scope: Range options for limiting the range
    resultfile: Execution result file

    Check the execution result file to verify that no errors are output in the command results for each node.

11. Stop the target node.

    Stop the target node as in step 5.

12. Boot the target node.

    Boot the target node as in step 6.

13. Stopping service on the target node

    Stop service on the target node as in step 7.

14. Deleting Cached Information

    Perform the procedures in "3.6.1.2 Deleting Cached Information".

15. Reapplying Interconnect Related Packages

    Reapply the interconnect related package.

    [If the target node is an FX server]

    If the target node is an FX server, these steps are unnecessary.

    [If the target node is an PRIMERGY server]

    Reapply the OFED package if the target node is a PRIMERGY server.

    Refer to the OFED manual for applying the OFED package.

16. Applying File System Related Packages

    Applies file system related packages.

    [If the target node is a storage cluster]

    ```
    # pmexe -c clstname scope --nodetype MGS,MDS,OSS "yum install -y FJSVfefs{-2,-debuginfo,-
    modules,-osd,progs}*"
    [INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
    ```

    clstname: Cluster name
    scope: Range options for limiting the range
    resultfile: Execution result file

    Check the execution result file to verify that no errors are output in the command results for each node.

    [If the target node is a compute cluster or a System management node]

    ```
    # pmexe --chroot -c clstname scope "yum -q install -y FJSVfefs-client*"
    [INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
    ```

    clstname: Cluster name
    scope: Range options for limiting the range
    resultfile: Execution result file

    Check the execution result file to verify that no errors are output in the command results for each node.

17. Starting the Compute Node.

    Note

    This step is necessary if the target includes an FX server.

    Perform the procedures in "3.6.1.5 Starting the Compute Node".

3.6.5.3 Creating initrd

Create an initrd for the new repository.

```
# pxinst initrd create -R reponame
[INFO] INST 0800 pxinst initrd completed successfully.(create initrd)
```

reponame: Name of the repository in which to create the initrd

3.6.5.4 Configuring ServerView Suite [SMM]

This procedure should only be performed if an OS minor version update has been made to the PRIMERGY server.

Populate the new repository with kernel modules and attachments provided by ServerView Suite.

For more information, see the Deployment Guide "Configuring ServerView Suite [SMM]".

Note

Specify the ServerView Suite that corresponds to the new OS minor version.

3.6.5.5 Recreating the rootfs Image on the PRIMERGY Server

This procedure is performed when the target node of the new repository is a PRIMERGY server.

```
# pxinst rootfs create -R reponame
```

reponame: New Repository Name

3.6.5.6 Recreating the rootfs Image on the FX Server

You create rootfs for the FX server by performing a network installation on the boot I/O node.

Therefore, create the rootfs image of the FX server when you rebuild the FX server.

3.7 Changing System Settings

When changing a system configuration, such as to add a user account, you can use common settings within the system by using the batch operation command.

By using an example of adding a user, this section describes how the system reflects setting changes.

3.7.1 Adding a User

To add a user account, you must create a user account with a same user ID and group ID on a range of target nodes in the system.

Perform this work on the system management node.

Information

In order for the added user to run the job, the user must be added to the following scope.

Table 3.7 User added scope

Node Type	Need To Create Users
Login node	Yes
System management node	Yes
Compute cluster management node	Yes
Compute cluster sub management node	
Boot I/O node	Yes
Global I/O node	Yes
Storage I/O node	Yes
Compute node	Yes
Storage cluster management node	
MGS node	Yes
MDS node	Yes
OSS node	Yes
Multiuse node	Yes

1. Creating a user account on nodes in the target range

   Use the pmexe command to create a user account on nodes in the same range as for the added user.
   Same user must have the same user ID and the group ID, so use the -g and -u options to make user with the same user ID and the same group ID on the nodes in the scope.

```
# pmexe -c clstname_comp --nodetype SMM,LN,CCM,BIO,GIO,SIO,CN "useradd -g gid -u uid username"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
# pmexe -c clstname_str --nodetype MGS,MDS,OSS "useradd -g gid -u uid username"
[INFO] PSM 5082 pmexe The execution result file was output.(resultfile)
```

   clstname_comp: Compute cluster name
   clstname_str: Storage cluster name
   gid: Group ID of the added user
   uid: User ID of the added user
   username: Name of the added user

2. Creating a password for the added user

   Create a password for the added user on the login node.

```
# passwd username
```

   username: Name of the added user

Chapter 4 Changing the System Configuration

See the "Job Operation Software Setup Guide" about adding and deleting clusters and adding and deleting nodes in a cluster.

Chapter 5 Collecting Investigation Materials

If trouble occurs while the Job Operation Software is operating, the administrator needs to collect information for an investigation. This chapter describes how to collect information for an investigation.

The two types of investigation materials that must be collected are system information and dump files.

System information

The information collected by the administrator includes hardware and software configuration information, setting information, and log files.

Dump file

If a node goes down or the system hangs, the administrator collects a dump file to investigate the OS status.

The administrator is requested to collect information as described below and contact a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.

See

In the collecting materials, there are the method of collecting only the investigation materials of the node and component related to the trouble event and the method of collecting the investigation materials only of time zone to execute the job.

For these methods, see "Collection of Investigation Data by Specifying a Trouble Event" and "Collection of Investigation Data by Specifying a Job ID" in "Job Operation Software Administrator's Guide for System Management."

5.1 Nodes in Investigation Material Collection

This section describes nodes from which investigation materials are collected.

The administrator is requested to collect investigation materials from the nodes listed below. However, a dump file normally is collected only from a node that went down or experienced a system hang.

- Nodes from which investigation materials must be collected
  - System management node
  - Compute cluster management node
- Nodes from which data collection depends on where trouble occurred (when the location can be identified)

  In addition to the information from the above nodes, also collect information from the nodes where trouble occurred (when they can be identified) and the related nodes.

Table 5.1 Nodes Where Trouble Occurs (When They Can be Identified) and Related Nodes

Trouble Location		Nodes From Which Investigation Materials Must be Collected
Within compute cluster	Compute node	Compute node Compute cluster sub management node Higher-level boot I/O node [FX server]
	Boot I/O node [FX server]	Boot I/O node [FX server] Compute cluster sub management node
	Storage I/O node [FX server]	Higher-level boot I/O node [FX server] Storage I/O node [FX server] Compute cluster sub management node
	Global I/O node [FX server]	Higher-level boot I/O node [FX server] Global I/O node [FX server] Storage cluster management node MDS node

Trouble Location	Nodes From Which Investigation Materials Must be Collected
Within storage cluster	Storage cluster management node MGS node MDS node OSS node

Note

- If you are not able to identify the node where the trouble occurred, collect investigation materials from the system management node, compute cluster management node, and compute cluster sub management node.
- Information possibly may be lost as time passes. Collect materials as soon as possible after trouble occurs.
- Depending on the results of an investigation with the collected information, further investigation materials may be required. In such cases, collect investigation materials as instructed by a Fujitsu customer engineer (CE) or the Fujitsu Support Desk.
- When trouble occurs on a node with FEFS installed, see the *FEFS User's Guide* and collect the required information.

5.2 How to Collect Investigation Materials

This section describes how to collect investigation materials.

5.2.1 System Information

The system information (command execution results and files) required for an investigation can be collected in batch as investigation materials by the pasnap command. Execute the command on the system management node to batch collect investigation materials.

Note

For details on how to collect investigation materials by using the pasnap command, see "Investigation Data Collection" in the "Job Operation Software Administrator's Guide for System Management."

5.2.2 Dump File

Dump files are collected automatically when a kernel panic occurs on a node or when system monitoring detects an OS failure. Also, to manually collect a dump file, use the papwrctl command from the active system management node to issue an instruction (init) to collect a dump from the target node.

The following example issues an instruction to collect a dump from the node whose node ID is 0xFFFF0005 in the cluster cluster1.

```
# papwrctl -c cluster1 -n 0xFFFF0005 init
```

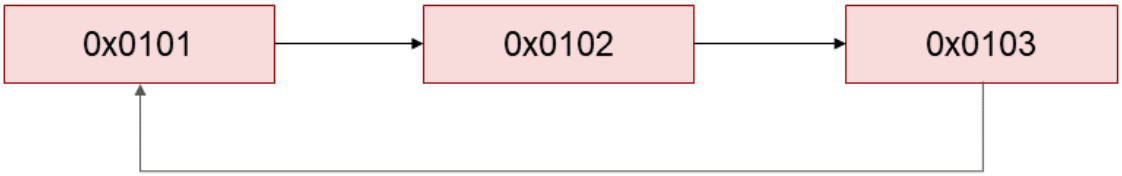
The end code of the papwrctl command is 0 when the instruction to collect a dump is successful.

Note

- To collect a dump file from the FX server boot I/O node, stop all the other nodes (nodes except this target node) in the BoB in advance. The dump file cannot be collected while the boot I/O node in a different, contiguous BoB relative to the target boot I/O node is stopped. This is because the FX server stores the dump file of the target boot I/O node on the boot I/O node in the different, contiguous BoB.

Here, the term "different, contiguous BoB" means the BoB with the adjacent boot group ID. The following example shows three boot group IDs with a cyclic relationship.

Figure 5.1 Example of the Relationship of Contiguous BoBs



The following example shows contiguous boot I/O nodes.

Table 5.2 Example of Contiguous Boot I/O Nodes

Target Boot I/O Node	Boot I/O Node in Contiguous BoB
0x01010001	0x01020001
0x01020001	0x01030001
0x01030001	0x01010001
0x01040001	0x01050001
0x01050001	0x01060001
0x01060001	0x01040001

- When restarted after dump collection, the node enters the Running state (displayed in the STATUS column by the pashowclst command) and is incorporated into operation. However, before the restart a compute node that is a PRIMERGY sever, execute the paclstmgr --disable -n <*nodeid*> from the system management node. After the node restarts, execute the paclstmgr --enable command from the system management node to incorporate it into production. For details on how to check the node status, see "Displaying Operation Status of the System" in the "Job Operation Software Administrator's Guide for System Management." For details on how to recover the node, see "Recovering a Node" in the "Job Operation Software Administrator's Guide for System Management."
- If the power-controlled nodes include nodes in a redundant configuration, disable the failover function as required. For details on disabling the failover function, see "Setting Automatic Failover" in the "Job Operation Software Administrator's Guide for System Management."

You can manage collected dump files in batch by using the padumpmgr command on the active system management node.

Examples of use are shown below.

1. Showing a list of dump files

   Using the --list option, the command can display the number of dump files and a list of the files. They are displayed by the command shown in the example below.
   You can display details for each node by specifying the -v option.
   The following example lists the number of dump files for each node in the cluster cluster1.

```
# padumpmgr --list -c cluster1 -v
```

2. Transferring dump files

   Use the --get option to transfer dump files to the system management node.

   When specifying the --get option, you need to also specify the transfer destination directory (in the -d option) for the dump files.
   The following example transfers the dump files for the boot group 0x0103 in the cluster cluster1.

```
# padumpmgr --get -c cluster1 -d /work/dump --bootgrp 0x0103
```

   In the --dumpid option, you can specify the dump files that you want to transfer.
   The following example transfers dump files (ID: 2, 3) from the node 0x01030004 in the cluster cluster1.

```
# padumpmgr --get -c cluster1 -d /work/dump/CN -n 0x01030004 --dumpid 2,3
```

Information

The dump file and dump ID on the target node can be viewed by using the padumpmgr command specifyed the -v option to the --list option, or specifyed the -n or -f option to the --list option.

The following example runs the padumpmgr command with the boot group ID 0x0101.

```
# padumpmgr --list -c cluster1 --bootgrp 0x0101 -v
[ CLST: cluster1 ]
[ BOOTGRP: 0x0101 ]
NODE       DUMPID  DATE              SIZE(M)
0x01010003 1       2018-08-25-13:43  -
0x01010004 2       2018-08-25-16:08  -
0x01010004 3       2018-08-27-10:14  -
...
```

Note

- The --get option only copies dump files to the transfer destination and does not delete them from the source nodes.
- The dump files cannot be transferred while the boot I/O node in a different, contiguous BoB relative to the target boot I/O node is stopped. In such cases, the following message is output at the dump file transfer time (when the padumpmgr command with the --get option is executed).

  ```
  [WARN] PSM 5116 padumpmgr It failed in forwarding the dump (Details)
  ```

  If this message is output, start the boot I/O node matching the above-described conditions, and then transfer the dump files again.

3. Deleting dump files

   Use the --del option to delete dump files on each node.
   The following example deletes dump files from the node whose node ID in 0xFFFF0010 in the cluster cluster1.

   ```
   # padumpmgr --del -c cluster1 -n 0xFFFF0010
   ```

4. Updating a list of dump files

   Normally, generating a dump file also automatically updates the list of dump files. However, if a fault occurs in the network between the system management node and other nodes, the list possibly may not be updated. At such times, use the --refresh option to update the list to the latest information.

   The minimum unit for the information update is a cluster. The list cannot be updated in a range smaller than a cluster, such as the unit of a node or node group.

   The following example executes the command for the cluster cluster1.

   ```
   # padumpmgr --refresh -c cluster1
   ```

Note

For FX server nodes, the --refresh option in the padumpmgr command cannot update the list of dump files. The system monitoring function automatically updates the list of dump files for FX server nodes.

Chapter 6 Backing Up and Restoring a Node

The Job Operation Software is equipped with the backup and restore functions in preparation for events such as a hard disk failure of a node. The functions back up the contents of a disk unit as a disk image and restores the backed-up disk image to the original node or another node.

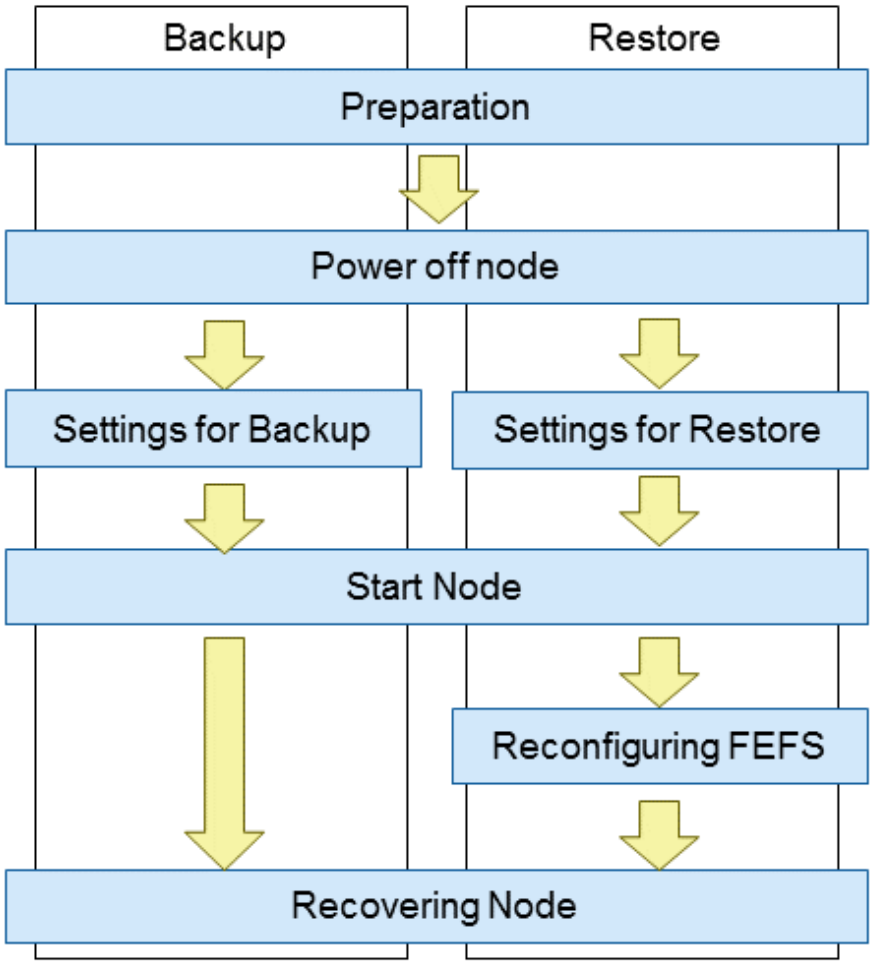
For details on the backup and restore functions, see "Details of the Backup/Restore Function" in the "Job Operation Software Administrator's Guide for System Management."

This chapter describes procedures in a maintenance workflow for backing up and restoring a node.

6.1 Workflow

The following figure shows the workflow for backing up and restoring a node.

Figure 6.1 Workflow for Backing Up and Restoring a Node



6.2 Backing Up and Restoring the FX server

This section describes procedures for backing up and restoring the FX server.

Note

- Before starting work to back up and restore a BoB belonging to a node group, confirm that the active compute cluster sub management node is running normally. The procedures use the TFTP server of the compute cluster sub management node.

- The FX server's /etc/fstab is auto-configured on restore, so changes are not retained. If you have modified /etc/fstab, reconfigure it after the restore.

6.2.1 Backing Up and Restoring a BoB

This section describes procedures for backing up and restoring an FX server BoB.

6.2.1.1 Preparation for Backup and Restore

1. Check required service status

   Use the check subcommand of the pxinst command to verify that the required services are running. Verify that "Installer node [s]" and "Relay node [s] "is displayed as "OK."

   The following is an example of a display when the service state of all nodes is normal.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]          1/1         [OK]
Relay node[s]              1/1         [OK]
[INFO] INST 1806 pxinst Service state check is complete.
```

   ipaddr1: IP address of the active system management node (management network)
   ipaddr2: IP address of the active compute cluster sub management node (management network)

   The following is an example of what you might see if you have a node with a bad service.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]           1/1          [OK]
Relay node[s]               0/1          [FAILED]
  Failed node[s]
    HOSTNAME           SERVICE
    ipaddr2            tftp, http
[WARN] INST 1501 pxinst Abnormal node exist.
```

   ipaddr1: IP address of the active system management node (management network)
   ipaddr2: IP address of the active compute cluster sub management node (management network)

   If the service is stopped or there are nodes that cannot communicate, a warning message is displayed. In addition, information of the node with error is displayed as "Failed node [s]". In this case, the system administrator must start the service on the target node. If communication with the target node is not possible, an error message is displayed in the "SERVICE" field.

2. Transitioning to maintenance mode

   To back up and restore the target nodes, transition the nodes to maintenance mode.
   For the procedure for transitioning to maintenance mode, see "3.3 Preparation for Software Maintenance."

   If the target nodes are already in maintenance mode, perform the next procedure because the preparation is complete.

Note

To perform the backup and restore procedures, the active system management node, the node for a saved disk image, and the nodes to be backed up and restored must satisfy the prerequisites of the backup and restore functions.
For the prerequisites of the backup and restore functions, see "Preconditions to Use the Backup/Restore Function" in the "Job Operation Software Administrator's Guide for System Management."

6.2.1.2 Backup Procedure

To back up the target BoB, execute the backup subcommand in the paclone command, and then restart the BoB. Perform this work on the active system management node.

1. Removing the FEFS setting for the boot group to be backed up

   Use pmexe to remove the FEFS settings for the destination node.

```
# pmexe -c clstname --bootgrp bootgid --stdout fefsconfig --cleanup
```

clstname: Cluster name
bootgid: Boot group ID

2. Powering off the target boot group

   Use the papwrctl command to power off the target BoB.

```
# papwrctl -c clstname --bootgrp bootgid off
```

clstname: Cluster name
bootgid: Boot group ID

3. Confirming that the power is off

   Use the papwrctl command to confirm that all columns on the target node have transitioned to "off".

```
# papwrctl -c clstname --bootgrp bootgid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            off off off
0xFF010002 off            off off off
0xFF010003 off            off off off
0xFF010004 off            off off off
0xFF010005 off            off off off
0xFF010006 off            off off off
...
```

clstname: Cluster name
bootgid: Boot group ID

4. Configuring settings for backup

   Specify the boot group ID of the target BoB, and execute the paclone command.

```
# paclone backup -c clstname --bootgrp bootgid -d imagedir
Do you really want to continue (y/n)? y           <- (*)
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
[INFO] INST 3201 paclone Command succeeded.(backup)
```

clstname: Cluster name
bootgid: Boot group ID
imagedir: Disk image storage directory
filepath: Path where the backup of the PXE configuration file is stored
(*) This prompts for confirmation before execution.

- The directory specified in the -d option must be set as an NFS public directory in advance. Also, read/write privileges must be granted to the root user.
- To specify a host name, the installation function must have completed registration of the node information.
- The PXE configuration file on the active system management node is overwritten when the paclone command is executed. When you rewrite the file, the unmodified PXE configuration file is stored in *filepath* shown in the following message.

  ```
  [INFO] INST 0801 paclone Created a configuration backup.(filepath)
  ```

  If you suspend the backup, place the above file stored at *filepath* in the path of the original PXE configuration file before performing step 6.

5. Synchronizing settings

   Synchronize the PXE settings, which were changed by the paclone command, with the standby system management node and compute cluster sub management node.

If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync all
[INFO] INST 1805 pxinst File replication is complete.
```

Note

If you want to suspend the backup after the settings are synchronized, place the pre-change configuration files stored in paclone in the path to the original PXE configuration files before performing step 6.
After placing, synchronize settings again.

6. Checking the status

   Use the status subcommand in the paclone command to check the backup status.
   Confirm that the OPERATION column shows "Backup," STATUS column shows "NotStarted," and STATE column shows "Waiting" for the boot I/O node in the target BoB.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
----------  ----------   --------       ---------      ------        -----        ------
clstname    0xFF010001   bio            Backup         NotStarted    Waiting      -
```

   clstname: Cluster name

7. Starting the boot I/O node

   Use the papwrctl command to start the boot I/O node.
   Start it with a PXE boot so that the backup operation begins. Specify "pxe" in the -boottype option.

```
# papwrctl -c clstname -n nodeid --boottype pxe on
```

   clstname: Cluster name
   nodeid: Node ID of the boot I/O node

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target node. If so, specify the --force-all option in the papwrctl command to start the node.

```
# papwrctl -c clstname -n nodeid --force-all --boottype pxe on
```

clstname: Cluster name
nodeid: Node ID of the boot I/O node

8. Confirming backup completion

   Use the status subcommand in the paclone command to check the backup status.
   Confirm that the STATUS column shows "Completed" and STATE column shows "Succeeded" for the boot I/O node in the target BoB.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
----------  ----------   --------       ---------      ------        -----        ------
clstname    0xFF010001   bio            Backup         Completed     Succeeded    -
```

   clstname: Cluster name

9. Starting the nodes in the boot group

   Use the papwrctl command to start the nodes in the boot group.

```
# papwrctl -c clstname --bootgrp bootgid --excludetype BIO on
```

clstname: Cluster name
bootgid: Boot group ID

10. Reconfiguring FEFS

    This procedure requires that the nodes in BoB have been booted. Using the pashowclst command with -v option, confirm the PWR_STATUS column of node in the target BoB is "os-running".
    Execute the fefsconfig command on the active system management node to configure FEFS on the restored nodes.

```
# pmexe -c clstname --bootgrp bootgid --stdout --compare "/usr/sbin/fefsconfig --setup"
```

    clstname: Cluster name
    bootgid: Boot group ID

11. Starting the FEFS Service

    Start the FEFS service on the node being backed up by running the fefs_sync command on the production system managed node.
    Execute the fefs_sync command on the active system management node and then start the FEFS service on the node to be backed up.

```
# fefs_sync --start --compute=clstname --bootgrp bootgid
```

    clstname: Cluster name
    bootgid: Boot group ID

6.2.1.3 Restore Procedure

To restore the target BoB, execute the restore subcommand in the paclone command, and then restart the BoB. Perform this work on the active system management node.

1. Powering off the target nodes

   Use the papwrctl command to power off the target BoB (boot group).

```
# papwrctl -c clstname --bootgrp bootgid off
```

   clstname: Cluster name
   bootgid: Boot group ID

2. Confirming that the power is off

   Use the papwrctl command to confirm that all columns on the target node have transitioned to "off".

```
# papwrctl -c clstname --bootgrp bootgid all-status
NODE       PWR_STATUS     ICC CMU BoB
0xFF010001 off            off off off
0xFF010002 off            off off off
0xFF010003 off            off off off
0xFF010004 off            off off off
0xFF010005 off            off off off
0xFF010006 off            off off off
...
```

   clstname: Cluster name
   bootgid: Boot group ID

3. setting for restore

   Execute the restore subcommand in the paclone command.

```
# paclone restore -c clstname --bootgrp bootgid -d imagedir
Do you really want to continue (y/n)? y            <- (*)
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
[INFO] INST 3201 paclone Command succeeded.(restore)
```

   clstname: Cluster name
   bootgid: Boot group ID
   imagedir: Save directory for the disk image

filepath: Path where the backup of the PXE configuration file is stored
(*) This prompts for confirmation before execution.

Note

The PXE configuration file on the active system management node is overwritten when the paclone command is executed.
When you rewrite the file, the unmodified PXE configuration file is stored in *filepath* shown in the following message.

```
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
```

If you suspend the restore, place the above file stored at *filepath* in the path of the original PXE configuration file before performing step 6.

4. Synchronizing settings

   Synchronize the PXE settings, which were changed by the paclone command, with the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync all
[INFO] INST 1805 pxinst File replication is complete.
```

Note

If you want to suspend the restore after the settings are synchronized, place the pre-change configuration files stored in paclone in the path to the original PXE configuration files before performing step 6.
After placing, synchronize settings again.

5. Checking the status

   Use the status subcommand in the paclone command to check the restore status.
   Confirm that the OPERATION column shows "Restore," STATUS column shows "NotStarted," and STATE column shows "Waiting" for the boot I/O node in the target BoB.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
----------  ----------   --------       ---------      ------        -----        ------
clstname    0xFF010001   bio            Restore        NotStarted    Waiting      -
```

clstname: Cluster name

6. Starting the node

   Use the papwrctl command to start the boot I/O node.
   Start it with a PXE boot so that the restore operation begins. Specify "pxe" in the -boottype option.

```
# papwrctl -c clstname -n nodeid --boottype pxe on
```

clstname: Cluster name
nodeid: Node ID

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target node. If so, specify the --force-all option in the papwrctl command to start the node.

```
# papwrctl -c clstname -n nodeid --force-all --boottype pxe on
```

clstname: Cluster name
nodeid: Node ID

7. Confirming restore completion

   Use the status subcommand in the paclone command to check the restore status.
   Confirm that the STATUS column shows "Completed" and STATE column shows "Succeeded" for the boot I/O node in the target BoB.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE         DETAIL
----------  ----------   --------       ---------      ------        -----         ------
clstname    0xFF010001   bio            Restore        Completed     Succeeded     -
```

   clstname: Cluster name

8. Starting the nodes in the BoB

   Use the papwrctl command to start the nodes in the BoB.

```
# papwrctl -c clstname --bootgrp bootgid --excludetype BIO on
```

   clstname: Cluster name
   bootgid: Boot group ID

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target nodes. If so, specify the --force-all option in the papwrctl command to start the nodes.

```
# papwrctl -c clstname --bootgrp bootgid --excludetype BIO --force-all on
```

clstname: Cluster name
bootgid: Boot group ID

9. Reconfiguring for job resource management function in a cluster

   This procedure requires that the nodes in BoB have been booted.
   Using the pashowclst command with -v option, confirm the PWR_STATUS column of node in the target BoB is "os-running".

   For the procedure for setting for job resource management function in a cluster, see "Settings for job resource management function in a cluster (parsc.conf file)" in "Chapter 3 Job Operation Management Function Settings" in "Job Operation Software Administrator's Guide for Job Management."

10. Reconfiguring FEFS

    This procedure requires that the nodes in BoB have been booted.
    Using the pashowclst command with -v option, confirm the PWR_STATUS column of node in the target BoB is "os-running".

    Execute the fefsconfig command on the active system management node to configure FEFS on the restored nodes.

```
# pmexe -c clstname --bootgrp bootgid --stdout --compare "/usr/sbin/fefsconfig --setup"
```

    clstname: Cluster name
    bootgid: Boot group ID

11. Starting the FEFS Service

    Start the FEFS service on the node being backed up by running the fefs_sync command on the production system managed node.
    Execute the fefs_sync command on the active system management node and then start the FEFS service on the node to be backed up.

```
# fefs_sync --start --compute=clstname --bootgrp bootgid
```

    clstname: Cluster name
    bootgid: Boot group ID

6.2.1.4 Recovering a Node

After backing up and restoring the target nodes, recover the nodes and incorporate them into operation.
See "3.5.3 Returning From Software Maintenance Mode," and recover the nodes.
The --recover suboption of the paclstmgr command must use --no-restart.

6.3 Backing Up and Restoring a PRIMERGY server

This section describes procedures for backing up and restoring a PRIMERGY server.

Note

- If the system management node is not in a redundant configuration, the DHCP server functions are unavailable when the node stops. This means you will not be able to obtain a backup of the system management node. To back up the system management node in a non-redundant configuration, use the dump command provided by the OS.
- To restore the system management node in a non-redundant configuration, stop job operations and then perform the work.

6.3.1 Preparation for Backup and Restore

1. Check required service status

   Use the check subcommand of the pxinst command to verify that the required services are running. Verify that "Installer node [s]" and "Relay node [s] "is displayed as "OK."

   The following is an example of a display when the service state of all nodes is normal.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]          1/1         [OK]
Relay node[s]              1/1         [OK]
[INFO] INST 1806 pxinst Service state check is complete.
```

   ipaddr1: IP address of the active system management node (management network)
   ipaddr2: IP address of the active compute cluster sub management node (management network)

   The following is an example of what you might see if you have a node with a bad service.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]           1/1          [OK]
Relay node[s]               0/1          [FAILED]
  Failed node[s]
    HOSTNAME           SERVICE
    ipaddr2            tftp, http
[WARN] INST 1501 pxinst Abnormal node exist.
```

   ipaddr1: IP address of the active system management node (management network)
   ipaddr2: IP address of the active compute cluster sub management node (management network)

   If the service is stopped or there are nodes that cannot communicate, a warning message is displayed. In addition, information of the node with error is displayed as "Failed node [s]". In this case, the system administrator must start the service on the target node. If communication with the target node is not possible, an error message is displayed in the "SERVICE" field.

2. Transitioning to maintenance mode

   To back up and restore the target node, transition the node to maintenance mode.
   For the procedure for transitioning to maintenance mode, see "3.3 Preparation for Software Maintenance."

   If the target node is already in maintenance mode, perform the next procedure because the preparation is complete.

Note

To perform the backup and restore procedures, the active system management node, the node for a saved disk image, and the node to be backed up and restored must satisfy the prerequisites of the backup and restore functions.
For the prerequisites of backup and restore functions, see "Preconditions to Use the Backup/Restore Function" in the "Job Operation Software Administrator's Guide for System Management."

6.3.2 Backup Procedure

To back up the target node, execute the backup subcommand in the paclone command, and then restart the node. Perform this work on the active system management node.

1. Powering off the target node

   Use the papwrctl command to power off the target node.

```
# papwrctl -c clstname -n nodeid off
```

   clstname: Cluster name
   nodeid: Node ID

2. Confirming that the power is off

   Using the pashowclst command, confirm a state transition for the target node to "off" in the PWR_STATUS column.

```
# pashowclst -c clstname -n nodeid
[ CLST: clstname ]
[ NODE: nodeid ]
NODE         NODETYPE  STATUS    REASON            PWR_STATUS  ARCH_STATUS   SRV_STATUS
0x00000002   SMM       Disable   SoftMaintenance   off         -             -
```

   clstname: Cluster name
   nodeid: Node ID

3. Configuring settings for backup

   Execute the backup subcommand in the paclone command.

```
# paclone backup -c clstname -n nodeid -d dirname -k diskname
Do you really want to continue (y/n)? y           <- (*)
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
[INFO] INST 3201 paclone Command succeeded.(backup)
```

   clstname: Cluster name
   nodeid: Node ID

   dirname: Save directory for the disk image
   diskname: Device name of the target disk
   filepath: Path to the save directory for the backup of the PXE configuration file
   (*) This prompts for confirmation before execution.

Note

The PXE configuration file on the active system management node is overwritten when the paclone command is executed. When you rewrite the file, the unmodified PXE configuration file is stored in *filepath* shown in the following message.

```
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
```

If you suspend the backup, place the above file stored at *filepath* in the path of the original PXE configuration file before performing step 6.

Information

- The directory specified in the -d option must be set as an NFS public directory in advance. Also, read/write privileges must be granted to the root user.
- To specify a host name, the installation function must have completed registration of the node information.

4. Synchronizing settings

   Synchronize the PXE settings, which were changed by the paclone command, with the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync all
[INFO] INST 1805 pxinst File replication is complete.
```

Note

If you want to suspend the backup after the settings are synchronized, place the pre-change configuration files stored in paclone in the path to the original PXE configuration files before performing step 6.
After placing, synchronize settings again.

5. Checking the status

   Use the status subcommand in the paclone command to check the backup status.
   Confirm that the OPERATION column shows "Backup," STATUS column shows "NotStarted," and STATE column shows "Waiting" for the target node.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
----------  ----------   --------       ---------      ------        -----        ------
clstname    0x00000002   smm2           Backup         NotStarted    Waiting      -
```

   clstname: Cluster name

6. Starting a backup

   Use the papwrctl command to start the node.
   Start it with a PXE boot so that the backup operation begins. Specify "pxe" in the -boottype option.

```
# papwrctl -c clstname -n nodeid --boottype pxe on
```

   clstname: Cluster name
   nodeid: Node ID

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target node. If so, specify the --force-all option in the papwrctl command to start the node.

```
# papwrctl -c clstname -n nodeid --force-all --boottype pxe on
```

7. Confirming backup completion

   Use the status subcommand in the paclone command to check the backup status.
   Confirm that the STATUS column shows "Completed" and STATE column shows "Succeeded" for the target node.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
```

```
----------  ----------  --------  ---------  ------  ---------  ------
clstname    0x00000002  smm2      Backup     Completed  Succeeded  -
```

clstname: Cluster name

Note

If you have the standby system management node or the standby compute cluster sub management node, note the following.

The standby system management node or the standby compute cluster sub management node is unable to communicate during node startup after completing a backup. As a result, the synchronization of the configuration files from the system management node may fail, and the following message may be output to the state (DETAIL) of the backed up node.

```
# paclone status show
CLSTNAME    NODEID      HOSTNAME      OPERATION     STATUS      STATE       DETAIL
----------  ----------  --------  ---------  ------  ---------  ------
clstname    nodeid      hostname      Backup        Completed   Failed      Update PXE (*)
configuration file failed. (detail) Please change default label to "localdisk" manually.
```

detail: Detailed error message
A new line starts just after the (*) part because of the space limitation of this document, but only one line is actually displayed.

If STATUS is Completed and STATE is Failed, the configuration files failed to synchronize after the backup operation completes successfully. After confirming that the target node is started, execute "pxinst sync all" to synchronize the configuration file. Note that even if this command is successful, the STATE Failed display does not disappear, but recovery is complete. To clear the Failed display above, execute "pxinst status clear" with a hostname to clear the backup completion status display.

6.3.3 Restore Procedure

To restore the target node, execute the restore subcommand in the paclone command, and then restart the node. Perform this work on the active system management node.

1. Powering off the target node

   Use the papwrctl command to power off the target node.

   ```
   # papwrctl -c clstname -n nodeid off
   ```

   clstname: Cluster name
   nodeid: Node ID

2. Confirming that the power is off

   Using the pashowclst command, confirm a state transition for the target node to "off" in the PWR_STATUS column.

   ```
   # pashowclst -c clstname -n nodeid
   [ CLST: clstname ]
   [ NODE: nodeid ]
   NODE         NODETYPE  STATUS    REASON           PWR_STATUS  ARCH_STATUS   SRV_STATUS
   0x00000002   SMM       Disable   SoftMaintenance  off         -             -
   ```

   clstname: Cluster name
   nodeid: Node ID

3. Executing the paclone command

   Execute the restore subcommand in the paclone command.

   ```
   # paclone restore -c clstname -n nodeid -d imagedir
   Do you really want to continue (y/n)? y            <- (*)
   [INFO] INST 0801 paclone Created a configuration backup.(filepath)
   [INFO] INST 3201 paclone Command succeeded.(restore)
   ```

clstname: Cluster name
nodeid: Node ID
imagedir: Save directory for the disk image
filepath: Path where the backup of the PXE configuration file is stored
(*) This prompts for confirmation before execution.

Note

The PXE configuration file on the active system management node is overwritten when the paclone command is executed.
When you rewrite the file, the unmodified PXE configuration file is stored in *filepath* shown in the following message.

```
[INFO] INST 0801 paclone Created a configuration backup.(filepath)
```

If you suspend the restore, place the above file stored at *filepath* in the path of the original PXE configuration file before performing step 6.

4. Synchronizing settings

   Synchronize the PXE settings, which were changed by the paclone command, with the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync all
[INFO] INST 1805 pxinst File replication is complete.
```

Note

If you want to suspend the restore after the settings are synchronized, place the pre-change configuration files stored in paclone in the path to the original PXE configuration files before performing step 6.
After placing, synchronize settings again.

5. Checking the status

   Use the status subcommand in the paclone command to check the restore status.
   Confirm that the OPERATION column shows "Restore," STATUS column shows "NotStarted," and STATE column shows "Waiting" for the target node.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE        DETAIL
----------  ----------   --------       ---------      ------        -----        ------
clstname    0x00000002   smm2           Restore        NotStarted    Waiting      -
```

   clstname: Cluster name

6. Starting the node

   Use the papwrctl command to start the node.
   Start it with a PXE boot so that the restore operation begins. Specify "pxe" in the -boottype option.

```
# papwrctl -c clstname -n nodeid --boottype pxe on
```

clstname: Cluster name
nodeid: Node ID

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target node. If so, specify the --force-all option in the papwrctl command to start the node.

```
# papwrctl -c clstname -n nodeid --force-all --boottype pxe on
```

clstname: Cluster name
nodeid: Node ID

7. Confirming restore completion

   Use the status subcommand in the paclone command to check the restore status.
   Confirm that the STATUS column shows "Completed" and STATE column shows "Succeeded" for the target node.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE         DETAIL
----------  ----------   --------       ---------      ------        -----         ------
clstname    0x00000002   smm2           Restore        Completed     Succeeded     -
```

   clstname: Cluster name

If you have the standby system management node or the standby compute cluster sub management node, note the following.

The standby system management node or the standby compute cluster sub management node is unable to communicate during node startup after completing a restore. As a result, the synchronization of the configuration files from the system management node may fail, and the following message may be output to the state (DETAIL) of the restored node.

```
# paclone status show
CLSTNAME    NODEID       HOSTNAME       OPERATION      STATUS        STATE         DETAIL
----------  ----------   --------       ---------      ------        -----         ------
clstname    nodeid       hostname       Restore        Completed     Failed        Update PXE (*)
configuration file failed. (detail) Please change default label to "localdisk" manually.
```

detail: Detailed error message
A new line starts just after the (*) part because of the space limitation of this document, but only one line is actually displayed.

If STATUS is Completed and STATE is Failed, the configuration files failed to synchronize after the restore operation completes successfully. After confirming that the target node is started, execute "pxinst sync all" to synchronize the configuration file. Note that even if this command is successful, the STATE Failed display does not disappear, but recovery is complete. To clear the Failed display above, execute "pxinst status clear" with a hostname to clear the restore completion status display.

8. Reconfiguring kdump

   Reconfigure the kdump on the node being restored.
   If you have a support desk contract, you can get a dump assist tool (FJSVdumptools) and configure kdump.
   If a Support Desk is not available or if you are not using the above tools, configure the settings according to your OS documentation.

   Set the system operation to halt after collecting memory dump.

9. Reconfiguring for job resource management function in a cluster

   Reconfigure the job resource management function in a cluster for the restored node.
   For the procedure for setting for job resource management function in a cluster, see "Settings for job resource management function in a cluster (parsc.conf file)" in "Chapter 3 Job Operation Management Function Settings" in "Job Operation Software Administrator's Guide for Job Management."

10. Reconfiguring FEFS

    Configure FEFS on the restored node.
    Perform this step only when restoring following types of nodes:

    - Compute cluster management node
    - Login node
    - Compute node
    - MGS node
    - MDS node

- OSS node
- Multiuse node (only when using FEFS)

Execute the fefsconfig command on the active system management node to configure FEFS.

```
# pmexe -c clstname -n nodeid --stdout --compare "/usr/sbin/fefsconfig --setup"
```

clstname: Cluster name
nodeid: Node ID

6.3.4 Recovering a Node

After backing up and restoring the target node, recover the node and incorporate it into operation.
See "3.5.3 Returning From Software Maintenance Mode," and recover the node.

Chapter 7 Rebuilding a Node

In cases like a node failure, for example, you may need to rebuild a node.

- Hard disk replacement

  When the data on the hard disk of a node is unavailable because the disk failed or for another reason

- Software not operable

  When the Job Operation Software cannot run normally because, for example, the data in the node was accidentally deleted

This chapter describes procedures for rebuilding a node with the Job Operation Software.

7.1 How to Rebuild a Node

There are two methods for rebuilding a node: method using the backup and restore functions, and method with initial installation.

Information

If you are rebuilding a node and want to change any information that you entered in the installation design sheet, such as MAC addresses or disk partition information, delete the node, and then rebuild it by adding the node, see "Job Operation Software Installation Guide."

For details on installation design sheets, see "Creating Installation Target Node Settings" in "Overview of the Installation Function" in the "Job Operation Software Administrator's Guide for System Management."

For the procedures for deleting and adding nodes, see the "Job Operation Software Setup Guide."

7.2 Rebuilding a Node Using the Backup and Restore Functions

To rebuild a node using the backup and restore functions, see the procedures in "Chapter 6 Backing Up and Restoring a Node."

7.3 Rebuilding a Node From Initial Installation

This section describes procedures for rebuilding a node from initial installation.

7.3.1 Notes on Rebuilding a Node From Initial Installation

Note the following points on rebuilding a node from initial installation.

- Perform work as a user who has root privileges.
- Before starting work to rebuild a node in a redundant configuration, be sure to confirm that the target node has been switched from active to standby. Even for the rebuild of a node that is a standby node, make sure to check its operating status before starting work to rebuild it.
- Before starting the rebuild of a node connected to an external disk unit, disconnect the node. After completing the node rebuild, connect the devices again.

The next sections describe procedures for rebuilding a node from initial installation.

7.3.2 Preparation for Node Rebuild

This section describes the preparation for node rebuild.
The node being rebuilt must be in a state that does not affect job operations.

7.3.2.1 Check required service status

Use the check subcommand of the pxinst command to verify that the required services are running. Verify that "Installer node [s]" and "Relay node [s] "is displayed as "OK."

The following is an example of a display when the service state of all nodes is normal.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]          1/1         [OK]
Relay node[s]              1/1         [OK]
[INFO] INST 1806 pxinst Service state check is complete.
```

ipaddr1: IP address of the active system management node (management network)
ipaddr2: IP address of the active compute cluster sub management node (management network)

The following is an example of what you might see if you have a node with a bad service.

```
# pxinst check -m ipaddr1,ipaddr2
Installer node[s]           1/1          [OK]
Relay node[s]               0/1          [FAILED]
  Failed node[s]
    HOSTNAME           SERVICE
    ipaddr2            tftp, http
[WARN] INST 1501 pxinst Abnormal node exist.
```

ipaddr1: IP address of the active system management node (management network)
ipaddr2: IP address of the active compute cluster sub management node (management network)

If the service is stopped or there are nodes that cannot communicate, a warning message is displayed. In addition, information of the node with error is displayed as "Failed node [s]". In this case, the system administrator must start the service on the target node. If communication with the target node is not possible, an error message is displayed in the "SERVICE" field.

7.3.2.2 Checking the Operating Status (Active/Standby) of a Redundantly Configured Node

For the rebuild of a redundantly configured node, always be careful about switching the target node from active to standby.
Before starting the rebuild, confirm that the node has been switched from active to standby.

Perform this work on the active system management node.

Using the pashowclst command, confirm that the operating status of the target node *nodeid1* is standby (the STATUS column shows "(S)").

```
# pashowclst -m -c clstname --nodetype nodetype
[ CLST: clstname ]
[ NODETYPE: nodetype ]
NODEGRP    NODETYPE    NODE        STATUS    NODE        STATUS
SMMGRP     nodetype    nodeid2     (A)       nodeid1     (S)
```

nodeid1: Target node
nodeid2: Node paired with the target node *nodeid1*

Also, if the rebuild target is a file server node (MGS node, MDS node, or OSS node), confirm that replication of the FEFS service has completed.
The following step shows how to confirm the completion of replication of the FEFS service.

Using the -v option in the pashowclst command, confirm that the paired node has replicated the FEFS service of the target node (the SRV_STATUS column shows "w" for the FEFS service of the target node or "*" for the FEFS service of the node paired with the target node).

```
# pashowclst -c clstname --nodetype MDS -v
[ CLST: clstname ]
[ NODETYPE: MDS ]
NODE        NODETYPE      STATUS        REASON             PWR_STATUS        ARCH_STATUS       SRV_STATUS
nodeid1      MDS           Running       -                   on                -                 FEFSSR(w)
nodeid2      MDS           Running       -                   on                -                 FEFSSR(*)
```

See

If the target node has not been switched from active to standby, or if the node paired with the file server node has not replicated the FEFS service, see "Switching a Node" in the "Job Operation Software Administrator's Guide for System Management." Switch the node from the active state, or replicate the FEFS service.

7.3.2.3 Isolating the Target Node From Operation

When rebuilding a node, transition the target node to maintenance mode.
For the procedure for transitioning the node to maintenance mode, see "3.3 Preparation for Software Maintenance."

If the target node is already in maintenance mode, perform the next procedure because the preparation is complete.

7.3.3 Rebuilding a Node

This section describes the procedure for rebuilding a node.

7.3.3.1 Rebuild Procedure

Perform this work on the active system management node.

1. Powering off the node

   If the target node is running, power off the node with the papwrctl command.

   ```
   # papwrctl -c clstname -n nodeid off
   ```

   clstname: Cluster name
   nodeid: Node ID

2. Confirming that the power is off

   Using the pashowclst command, confirm a state transition for the target node to "off" in the PWR_STATUS column.

   ```
   # pashowclst -c clstname -n nodeid
   [ CLST: clstname ]
   [ NODE: nodeid ]
   NODE         NODETYPE  STATUS    REASON           PWR_STATUS  ARCH_STATUS   SRV_STATUS
   0x00000002   SMM       Disable   SoftMaintenance  off         -             -
   ```

3. Preparing a rootfs Image

   If the target node is an FX server, create a rootfs image for the FX server as described in "Preparing a rootfs Image [SMM]" in "Setup Guide".

   Then, see "Editing a rootfs Image [SMM]" in "Setup Guide" to register the rootfs image in the repository.

4. Updating node information

   To rebuild the node, update the node information held by the installation function.
   The following example updates the information, enabling image-based installation at the next start time.

   ```
   # pxinst node update --type img --hostname hostname,...
   [INFO] INST 0801 pxinst Created a configuration backup.(/tmp/pxinst/node_date)  (*)
   [INFO] INST 0800 pxinst node completed successfully.(update)
   ```

   hostname: Host name of the target node

   (*) Information before execution of the pxinst command is saved in the /tmp/pxinst/node_*date* directory.

Note

The PXE configuration file on the active system management node is overwritten when the pxinst command is executed.
When overwriting the file, the pxinst command outputs the following message and collects a backup of the PXE configuration file.
The /tmp/pxinst/node_*date* directory stores the collected backup.

```
[INFO] INST 0801 pxinst Created a configuration backup.(/tmp/pxinst/node_date)
```

If you suspend the node rebuild, place the above file stored at /tmp/pxinst/node_*date* in the path of the original PXE configuration file before performing step 6.

See

Three methods for OS installation on a node are image-based OS installation, package-based OS installation, and restoring a backup image. The system administrator can select the provisioning method. For details, see "Overview of the Installation Function" in the "Job Operation Software Administrator's Guide for System Management."

5. Synchronizing settings

   Synchronize the PXE settings, which were changed by the pxinst command, with the standby system management node and compute cluster sub management node.
   If the system does not have the standby system management node and compute cluster sub management node, this operation is unnecessary.

```
# pxinst sync all
[INFO] INST 1805 pxinst File replication is complete.
```

6. Checking the status

   Use the status subcommand in the pxinst command to check the status of the updated node.
   Confirm that the STATUS column shows "NotStarted" and STATE column shows "Waiting" for the target node.

```
# pxinst status show
HOSTNAME       OPERATION      STATUS       STATE        DETAIL
--------       ---------      ------       -----        ------
hostname       ImgProvision   NotStarted   Waiting      -
```

   hostname: Host name of the target node

7. Starting the rebuild

   Use the papwrctl command to start the node.
   Start it with a PXE boot so that the rebuild begins. Specify "pxe" in the -boottype option.

```
# papwrctl -c clstname -n nodeid --boottype pxe on
```

   clstname: Cluster name
   nodeid: Node ID

Note

When you check the run status of nodes with the pashowclst command, the REASON column may show "HardMaintenance" for the target node. If so, specify the --force-all option in the papwrctl command to start the node.

```
# papwrctl -c clstname -n nodeid --force-all --boottype pxe on
```

8. Confirming rebuild completion

   Use the status subcommand in the pxinst command to check the rebuild status.
   Confirm that the STATUS column shows "Completed" and STATE column shows "Succeeded" for the target node.

For the FX server, HOSTNAME displays the host name of the compute node that also serves as a boot I/O node.

```
# pxinst status show
HOSTNAME       OPERATION      STATUS        STATE        DETAIL
--------       ---------      ------        -----        ------
hostname       ImgProvision   Completed     Succeded     -
```

hostname: Host name of the target node

9. Reflecting cluster configuration definitions

   Execute the pxsystemadm command with the --set option to configure the Job Operation Software on the node that has the reinstalled OS.

```
# pxsystemadm --set -c clstname -n nodeid
Do you really want to continue (y/n)? y
[INFO] PSM 1082 pxsystemadm The setting was completed.
```

   clstname: Cluster name
   nodeid: Node ID

10. Configuring settings for job operation

    Configure settings for job operation on the rebuilt node.

    See "Configuring Settings for System Operation" in the "Job Operation Software Setup Guide," and configure the various settings required for job operation.

11. Checking the software environment

    From the software perspective, check whether the installed nodes are ready to be incorporated into operation.

    Execute the pachecker command, and confirm that the settings are as expected.

    For details on the pachecker command, see "Checking the Software Environment" in the "Job Operation Software Administrator's Guide for System Management."