C122-B018-01EN

PRIMEQUEST
510A

SYSTEM DESIGN GUIDE

FUJITSU

FOR SAFE OPERATION

This manual contains important information regarding the use and handling of this product. Read this manual thoroughly. Pay special attention to the section "NOTE ON SAFETY" Use the product according to the instructions and information available in this manual. Keep this manual handy for further reference.

Fujitsu makes every effort to prevent users and bystanders from being injured or from suffering damage to their property. Use the product according to this manual.

ABOUT THIS PRODUCT

This product is designed and manufactured for use in standard applications such as office work, personal device, household appliance, and general industrial applications. This product is not intended for use in nuclear-reactor control systems, aeronautical and space systems, air traffic control systems, mass transportation control systems, medical devices for life support, missile launch control systems or other specialized uses in which extremely high levels of reliability are required, the required levels of safety cannot be guaranteed, or a failure or operational error could be life-threatening or could cause physical injury (referred to hereafter as "high-risk" use). You shall not use this product without securing the sufficient safety required for high-risk use. If you wish to use this product for high-risk use, please consult with sales representatives in charge before such use.

RADIO FREQUENCY INTERFERENCE STATEMENT

The following notice is for EU users only.

WARNING: This is a product which meets Class A of EN55022. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.

The following notice is for USA users only.

This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference in which case the user will be required to correct the interference at his own expense.

Laser standards

This equipment includes Class 1 laser products and complies with FDA Radiation Performance Standards, 21 CFR 1040.10 and 1040.11, and the International Laser Safety Standards IEC60825-1: 2001.

TRADEMARK ACKNOWLEDGEMENTS

- Microsoft, Windows, MS, Windows NT, and Windows Server are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark or registered trademark of Linus Torvalds in the United States and other countries.
- Red Hat, RPM, and all Red-Hat-based marks and logos are trademarks or registered trademarks of Red Hat, Inc. in the United States and other countries.
- SUSE is a registered trademark of Novell Inc. in the United States and other countries.
- Intel, Xeon, and Itanium are trademarks or registered trademarks of Intel Corporation.
- Ethernet is a registered trademark of Fuji Xerox Co., Ltd. and Xerox Corporation in the United States and other countries.
- All other product names mentioned herein are the trademarks or registered trademarks of their respective owners.
- System and product names in this manual are not always noted with trademark or registered trademark symbols (™), (®).

TERMS AND CONDITIONS

The product includes software provided by third parties in addition to that provided by Fujitsu Ltd. You are granted permission to use the third parties' software subject to the terms and conditions below. If you acquire the source code of the software to which the following terms and conditions apply, refer to LICENSE1_EN.pdf and LICENSE2_EN.pdf, which are provided with the 'PRIMEQUEST Manuals' (C122-E013-C2).

THIS SOFTWARE IS PROVIDED "AS IS" AND FUJITSU LIMITED MAKES NO WARRANTIES, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER WHATSOEVER REGARDING TO THIS SOFTWARE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE.

IN NO EVENT SHALL FUJITSU LIMITED BE LIABLE FOR ANY CLAIM FROM A THIRD PARTY, OR SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE, COPYING, MODIFICATION OR DISTRIBUTION OF THIS SOFTWARE.

Fujitsu Ltd, April 2006

C122-B018-01EN

Revision History

(1/1)

Edition	Date	Revised section (Added/ Deleted/ Altered) (Note)	Details
01	2008-05-20	—	—

Note: In this table, the revised section is indicated by its section number in the current edition.
An asterisk (*) indicates a section in the previous edition.

Preface

This manual describes conditions and points for consideration for PRIMEQUEST system operation design and provides related essential notes.

The manual is intended for system administrators. Read the manual together with the reference manuals cited in it.

This section explains

Structure and Contents of This Manual

This manual is organized as described below.

CHAPTER 5 Redundant Configuration

Describes the redundant configuration of main unit components, including the redundant LANs and HDDs.

CHAPTER 6 Hot Swapping

Describes hot swapping of hardware components of the main unit.

CHAPTER 7 Clustering

Describes clustering and provides essential notes on configuring a cluster system.

CHAPTER 8 Backup and Restoration

Describes system maintenance by explaining the necessity of backups, how to back up and restore the PRIMEQUEST system, and other such topics.

CHAPTER 9 Considerations for Maintenance

Describes the system maintenance considerations that must be made during system design.

Appendix A Software Supplied with PRIMEQUEST Hardware

Describes the software supplied with the PRIMEQUEST.

Glossary

Explains the terms used in this manual.

Index

Provides keywords and corresponding reference page numbers so that the reader can easily search for items in this manual as necessary.

Other Reference Manuals

The following manuals are provided for reference:

a) PDF manuals included on the *PRIMEQUEST Manuals* CD-ROM disk (C122-E013-C2)

Title	Description	Manual code
PRIMEQUEST 510A System Design Guide	Explains requirements, considerations, and notes on the system operation design of the PRIMEQUEST 510A.	C122-B018EN
PRIMEQUEST 510A Installation Planning Manual	Explains specifications and requirements for installation sites that are applicable to the installation of the PRIMEQUEST 510A.	C122-H003EN
PRIMEQUEST 500A/500/ 400 Series Installation Manual	Explains the setup of the PRIMEQUEST, including the preparation for the installation, initial settings, and software installation.	C122-E001EN
PRIMEQUEST 510A Reference Manual: Basic Operation/GUI/ Commands	Explains operations, setup methods, and the system management method that are required for the system operation of the PRIMEQUEST. The explanation covers basic operations and functions of the MMB, PSA, and EFI.	C122-E096EN
PRIMEQUEST 500A/500/ 400 Series Reference Manual: Tools/Operation Information	Explains system maintenance, Hot Plug, REMCS, and LEDs and other information required for system operation. Also, the manual provides supplementary information such as information on the physical locations of components.	C122-E074EN
PRIMEQUEST 500A/500/ 400 Series Reference Manual: Messages/Logs	Explains measures to be taken against problems that occur during operation and describes various types of messages.	C122-E004EN
SPARC Enterprise/ PRIMEQUEST Common Installation Planning Manual	Explains basic information and policy on installation planning and facilities planning that are required for the installation of the SPARC Enterprise series and PRIMEQUEST series.	C120-H007EN

b) Printed manual

For the printed manual (sold separately), contact your certified service engineer.

- PRIMEQUEST 500A/500/400 Series Installation Manual (C122-E001EN)

Abbreviations

In this manual, the product names are abbreviated as follows:

Long title	Abbreviations
Red Hat® Enterprise Linux® AS (v.4 for Itanium) (*1)	Red Hat (*3)
Red Hat® Enterprise Linux® 5 (for Intel Itanium) (*2)	
Red Hat® Enterprise Linux® AS (v.4 for Itanium)	RHEL-AS4(IPF)
Red Hat® Enterprise Linux® 5 (for Intel Itanium)	RHEL5(IPF) (*4)
SUSE™ Linux Enterprise Server 9 for Itanium Processor Family	SUSE
SUSE™ Linux Enterprise Server 10 for Itanium Processor Family	
SUSE™ Linux Enterprise Server 9 for Itanium Processor Family	SUSE9
SUSE™ Linux Enterprise Server 10 for Itanium Processor Family	SUSE10
Microsoft® Windows Server® 2003, Enterprise Edition for Itanium-based Systems	Windows Windows Server 2003
Microsoft® Windows Server® 2003, Datacenter Edition for Itanium-based Systems	
Microsoft® Windows Server® 2008 for Itanium-Based Systems	Windows Windows Server 2008

*1: Supports Red Hat® Enterprise Linux® AS (v4.6 for Itanium) or later.

*2: Supports Red Hat® Enterprise Linux® 5.1 (for Intel Itanium) or later.

*3: Version-independent abbreviation

*4: A description in the form of "RHEL5.x (IPF)" indicates an updated version.

Text Conventions

This manual uses the following fonts and symbols to express specific types of information.

Fonts/symbols	Meaning	Example
Italic	Indicates names of manuals.	See the *PRIMEQUEST 510A System Design Guide*.
" "	Indicates names of chapters, sections, items, buttons, or menus.	See CHAPTER 5, "Redundant Configuration"
[]	Indicates window names, window button names, tab names, and dropdown menu selections.	Click the [OK] button.

Syntax of the Command Line Interface (CLI)

The command syntax is described below.

Command syntax

The command syntax is as follows:

- A variable that requires input of a value must be enclosed in < >.
- An optional element must be enclosed in [ ].
- A group of options for an optional keyword must be enclosed in [ ] and delimited by |.
- A group of options for a mandatory keyword must be enclosed in { } and delimited by |.

The command syntax is shown in a frame such as this one.

Notes Regarding Notations Used in This Manual

- Items marked with "Linux" apply to both Red Hat ® Enterprise Linux ® AS (v.4 for Itanium), SUSE™ Linux Enterprise Server 9 for Itanium Processor Family, and SUSE™ Linux Enterprise Server 10 for Itanium Processor Family (*).
  * For details, contact your Fujitsu certified service engineer.

Conventions for Alert Messages

This manual uses the following conventions to show alert messages. An alert message consists of an alert signal and alert statements.

WARNING	This indicates a hazardous situation that *could* result *in serious personal injury* if the user does not perform the procedure correctly.
CAUTION	This indicates a hazardous situation that *could* result *in minor or moderate personal injury* if the user does not perform the procedure correctly. This signal also indicates that damage to the product or other property *may* occur if the user does not perform the procedure correctly.
IMPORTANT	This indicates information that could help the user to use the product more effectively.

Alert messages in the text

In the text, alert messages are indented to distinguish then from regular text. A wider space precedes and follows the message to show where the message begins and ends.

WARNING

Certain tasks in this manual should only be performed by a certified service engineer. Users must not perform these tasks. Incorrect operation of these tasks may cause electric shock, injury, or fire.

- Installation and reinstallation of all components, and initial settings
- Removal of front or top covers
- Mounting/de-mounting of optional internal devices

Environmental Requirements for Using This Product

This product is a computer which is intended to be used in a computer room. For details on the operational environment, see the *PRIMEQUEST 510A Installation Planning Manual* (C122-H003EN).

Reader Feedback

- The screen images in this manual may be different from the actual screen images.
- If you find any errors or unclear statements in this manual, please fill in the "Reader's Comment Form" sheet at the back of this manual and forward it to the address indicated at the bottom of the sheet.
- This manual is subject to revision without prior notice.
- The PDF version of this manual is best viewed in Adobe® Reader® with a magnification of 100% and Single Page for the page layout.

NOTE ON SAFETY

Alert Labels

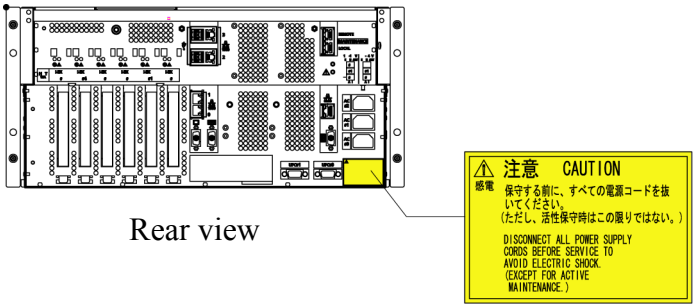
The following labels are attached to this product:

These labels provide information to the users of this product.

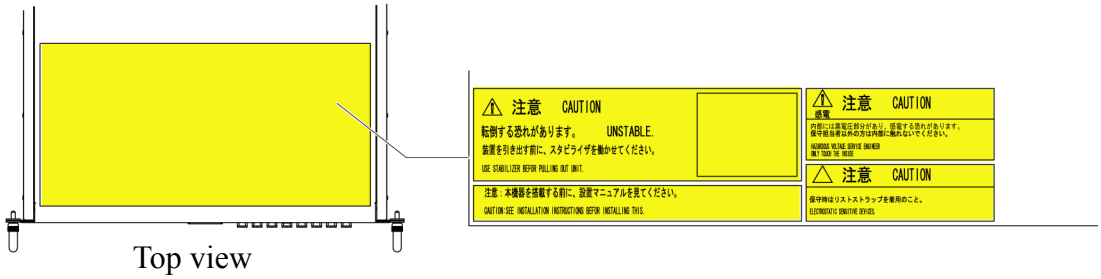
⚠CAUTION

Do not peel off the labels.

- Main unit



Rear view



Top view

Product Handling

Maintenance

⚠WARNING

Certain tasks in this manual should only be performed by a certified service engineer. Users must not perform these tasks. Incorrect operation of these tasks may cause electric shock, injury, or fire.

- Installation and reinstallation of all components, and initial settings
- Removal of front or top covers
- Mounting/de-mounting of optional internal devices
- Plugging or unplugging of external interface cards
- Maintenance and inspections (repairing, and regular diagnosis and maintenance)

⚠CAUTION

The following tasks regarding this product and the optional products provided from Fujitsu should only be performed by a certified service engineer. Users must not perform these tasks. Incorrect operation of these tasks may cause malfunction.

- Unpacking optional adapters and such packages delivered to the users

Remodeling/Rebuilding

⚠CAUTION

Do not make mechanical or electrical modifications to the equipment.
Using this product after modifying or overhauling may cause unexpected injury or damage to the property, the user, or bystanders.

Contents

Figures

Tables

CHAPTER 1 PRIMEQUEST Overview

This chapter provides an overview of PRIMEQUEST as follows:

1.1 Basics of PRIMEQUEST

PRIMEQUEST is a server platform based on the latest IA architecture.

The PRIMEQUEST series includes a number of high-reliability and high-performance technologies that Fujitsu has long cultivated over the years. It can thus serve large-scale mission-critical systems, including those supporting social infrastructures or enterprise business operations.

PRIMEQUEST is a server platform that includes, as its central processing units (CPUs), Itanium Processor Family processors (hereafter referred to as IPF processors or simply as CPUs), which are 64-bit processors manufactured by Intel Corporation.

This manual describes PRIMEQUEST 510A.

The PRIMEQUEST 510A houses up to 4 CPUs and can be equipped with up to 64 GB of memory, internal hard disks with a maximum capacity of 1176 GB, PCI Express slots that can be expanded to up to 6 slots.

This model supports Linux and Windows, operating systems (hereafter sometimes abbreviated as OSs) that are available in 64-bit processor configurations.

Figure 1.1 shows the appearance of the PRIMEQUEST 510A.

Figure 1.1 PRIMEQUEST 510A

1.1.1 PRIMEQUEST conceptual diagram

The PRIMEQUEST cabinet contains hardware elements. These hardware elements are interconnected within the cabinet and work independently of other elements and, at the same time, cooperate with each other to serve as an integral system.

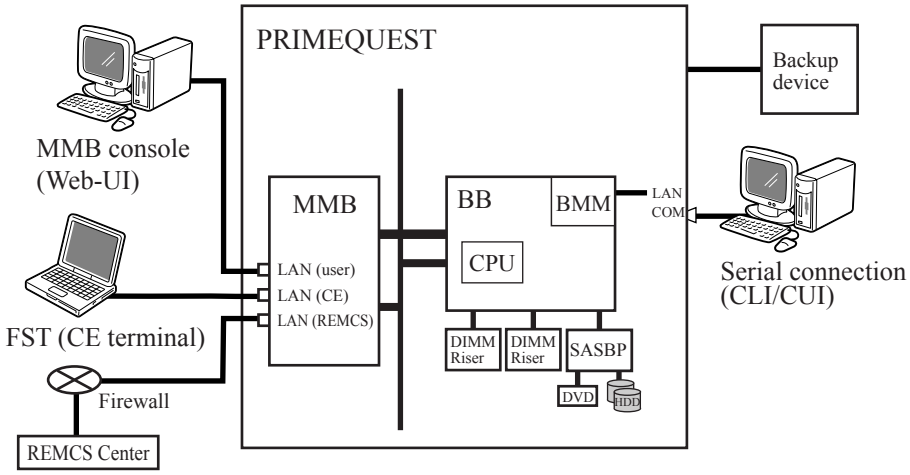


Figure 1.2 Concept of PRIMEQUEST configuration

PRIMEQUEST contains software resources, such as an operating system and application programs, as well as hardware resources, such as CPU and I/O devices. PRIMEQUEST can be connected to an external storage device via a PCI card. It can also interact with the operating system via a remotely connected or serially connected console.

The Management Board (MMB) is a feature that manages the PRIMEQUEST hardware as a whole. The MMB can be connected to the MMB console, the Field Support Tool (FST), and the Internet via a LAN. The MMB console is a console used to configure or display the hardware resource settings. The FST is a console used by maintenance personnel to maintain the PRIMEQUEST hardware. It is connected to the Remote Customer Support System (REMCS) via the Internet or a P-to-P connection to remotely monitor the unit status.

1.1.2 Product lineup

The PRIMEQUEST 510A is a server that accommodates up to 4 CPUs.

Table 1.1 Hardware specifications

Item		Specification
CPU	Type	Itanium
	Freq./L3cache	1.60 GHz/24 MB, 1.42 GHz/12 MB
	Number of CPUs	Up to 4 CPUs
Base board (SB)		1
Chipset		PRIMEQUEST-dedicated chipset
Memory (maximum capacity)		64 GB
Built-in I/O	LAN	4 × Gigabit Ethernet
	External port	Video, USB × 4, Serial × 1
	DVD-ROM drive	1
HDD		Up to 8 drives: 1176 GB (147 GB × 8 slots)
Number of PCI slots		Up to 6 PCI Express slots
Server management		MMB
Redundant configuration		PSU, HDD, PCI card, fan
Hot swapping		PSU, fan, HDD, PCI card
External dimensions (WxDxH) (mm)		482 × 740 × 219 (5V)
Weight		60 kg
Power consumption		1.17 KVA
Power supply condition		Power supply type: AC Voltage: 100 to 120 VAC ± 10%, 200 to 240 VAC ± 10% Number of phases: Single phase Frequency: 50 Hz/60 Hz, - 2 %/- 4 %

MMB: Management board
PSU: Power supply unit

1.2 PRIMEQUEST Features

PRIMEQUEST has the following three features:

- High performance
  The system provides the high-speed multiprocessor and high throughput.
- High reliability and high availability
  PRIMEQUEST-series machines detect errors and failures in their own systems, and they correct minor errors and isolate faulty components to provide continuous system operation. Faulty components can be easily located and replaced, enabling quick system recovery.
- High operability
  All server management operations can be centrally managed from the MMB Web-UI.

Figure 1.3 shows the technologies underlying the three features. The following sections describe the overview of these technologies.

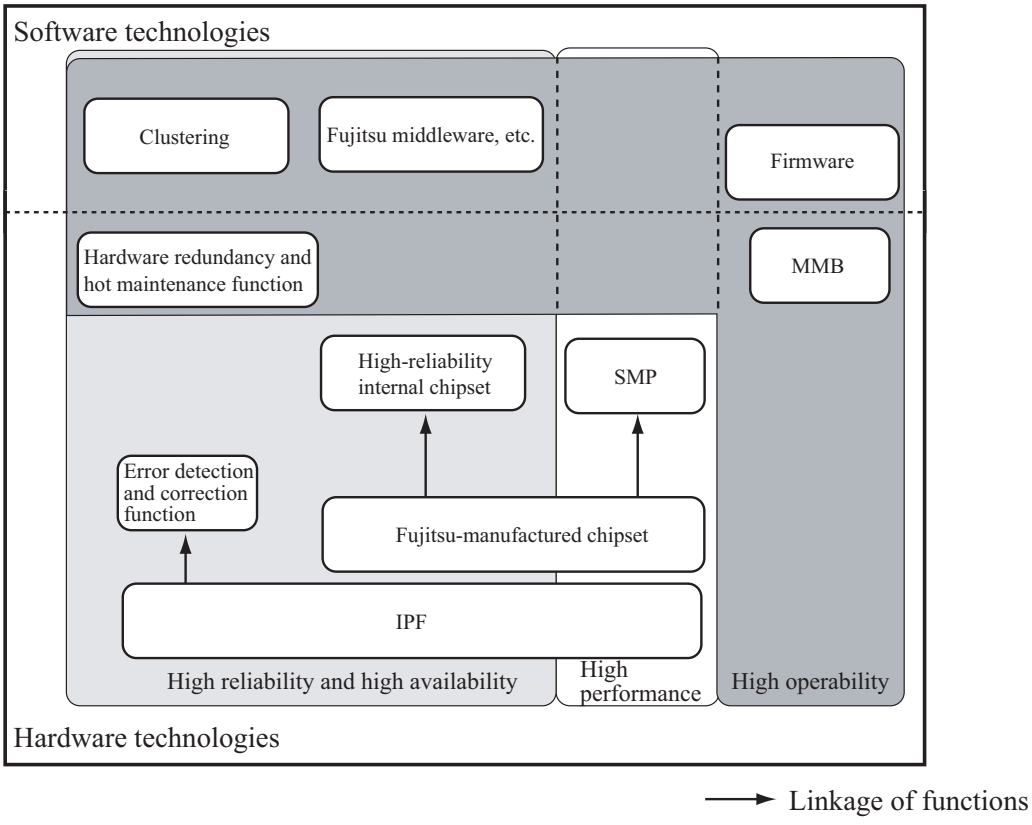


Figure 1.3 PRIMEQUEST technologies

1.2.1 Hardware technologies

This section describes the following essential hardware technologies shown in Figure 1.3.

- Technologies underlying high performance
  - PRIMEQUEST-dedicated chipset
  - IPF
  - Symmetric multi-processing (SMP) technology
- Technologies underlying high reliability and high availability
  - Hardware redundancy and hot swapping function
- Technologies underlying high operability
  - MMB (management board)

1.2.1.1 PRIMEQUEST-dedicated chipset

The chipset is an LSI device that controls data transfer among CPUs, memory, and IO Units.

The PRIMEQUEST-series machines contain this new chipset that has been developed by Fujitsu for improved performance and increased reliability.

The following three achievements are made by the PRIMEQUEST-dedicated chipset:

- SMP

  The high-speed memory access technology in the chipset suppresses deteriorations in performance in a large-scale SMP configuration, resulting in a high-performance system with high scalability.

- High reliability of the internal chipset

  By using highly reliable technologies (e.g., duplicated internal chipset) originally developed for mainframes, the chipset has greatly increased reliability.

1.2.1.2 IPF

The PRIMEQUEST-series machines contain IPF, which are Intel 64-bit processors.

The IPF features reliability and performance levels, as described below, suitable for large-scale mission-critical systems:

- High reliability

  The reliability, availability and serviceability (RAS) function is enhanced by the error detection and correction function incorporated for cache memory and other components. This improves business continuity.

- High performance

  The IPF improves performance because its 64-bit processors use an expanded memory address space and have higher computing speeds.

- Openness

  Since the IPF processors are open-standard IA processors, a wide range of IPF independent software vendor (ISV) and independent hardware vendor (IHV) products can be used, thereby expanding the selection of products for system construction.

1.2.1.3 Symmetric multi-processing (SMP) technology

Each processor is the same as any other processor in a multiprocessor system using SMP technology, so processing can be assigned to any processor. Scalability can thus be obtained with applications unchanged.

The PRIMEQUEST 510A supports an SMP system configuration that supports up to four CPUs.

1.2.1.4 Hardware redundancy and hot maintenance function

The main components in the PRIMEQUEST-series machines are made redundant to improve business continuity.

Also, PRIMEQUEST supports the hot maintenance function under certain conditions, so that maintenance components can be disconnected and installed during system operation. However, replacing components is a task handled by Fujitsu maintenance personnel.

Hardware redundant configuration

The redundant configurations of the main hardware components are described below:

- Internal chipset — The important control part of the PRIMEQUEST-dedicated internal chipset is duplicated.
- Power supply unit — Redundant power supply units are provided, so if one power supply unit fails, operation can continue with another power supply unit. A dual power feed has been prepared as an option.
- Cooling fan unit — Redundant cooling fan units are also provided, so if one fan unit becomes faulty, operation can continue with another fan unit cooling the system.
- Other main components — Different types of LANs are duplicated (standard), and if one system fails, operation can continue with the other system.

Hot maintenance function

The hot maintenance function disconnects a unit from the system for maintenance work on the unit, without stopping the system operation.

Hot maintenance is possible because the main components and important units in a PRIMEQUEST-series machine are in redundant configurations, as described below:

- As standard, the power supply units (PSUs) and fans support hot maintenance.
- HDD, PCI card, and other devices is possible under specific conditions.

1.2.1.5 Management board (MMB)

No dedicated management server need be installed because the PRIMEQUEST cabinet contains the MMB, which has the server management function.

The MMB functions include hardware status monitoring, displaying configuration and error information, network environment management, and power control.

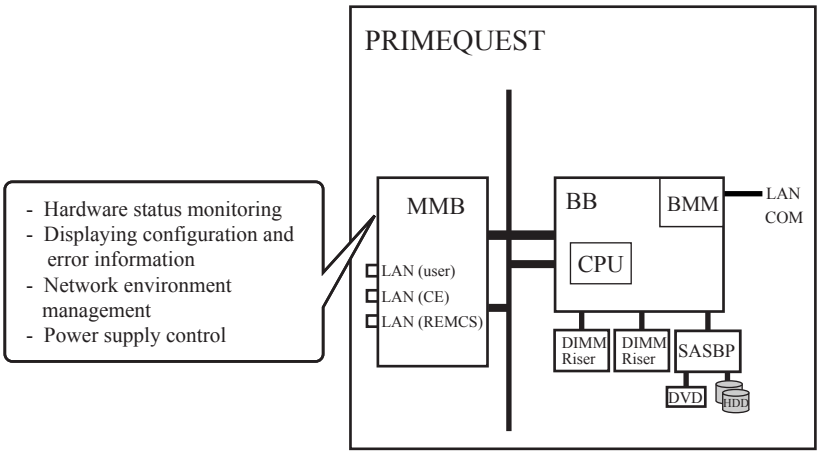


Figure 1.4 MMB management function

The built-in MMB provides the following advantages:

- Reduced management costs

  Management costs can be greatly reduced because the server management function does not require a dedicated management server or dedicated software.

- Centralized management

  The MMB can centrally monitor, control, and set up hardware resources of the server.

1.2.2 Software technologies

The following software is adopted by the PRIMEQUEST-series machines to increase reliability, availability, and operability:

- Firmware
- Linux
- Windows
- Clustering software
- Fujitsu middleware products

1.2.2.1 Firmware

The firmware listed below is the main hardware management and operation firmware installed on PRIMEQUEST-series machines.

For details, see CHAPTER 3, "General Software Configuration."

- MMB firmware
  Monitoring of all hardware and centralized operation
- PSA (PRIMEQUEST Server Agent: system management software)
  Monitoring of hardware

1.2.2.2 Linux

PRIMEQUEST supports Red Hat Enterprise Linux, which are platforms adopted as global standards. It also supports SUSE Linux Enterprise Server, mainly for overseas markets.

Table 1.2 Linux OS features

Growth potential	Linux's strong growth potential is underscored by the following characteristics: • Not dependent on the principles and business of any particular enterprise (OS vendor) • Continuous technological improvement and development of new technologies by many developers and researchers
High reliability	Objective of continuous increases in reliability, such as with the following high-reliability functions: • Quick fixes to security holes because engineers around the world tackle defects • Logging of detailed hardware error information • Enhanced maintenance functions (trace function and dump function) • High-speed and high-reliability cluster switching synchronized with hardware and cluster software
Long-time support	Support can be provided for a long time because Linux is open-source software and its source is open to the public.

1.2.2.3 Windows

PRIMEQUEST supports Windows Server 2003, which supports 64-bit processors, and will support the next-generation of Windows when it is released.

1.2.2.4 Clustering software

The Linux OS supports clustering using PRIMECLUSTER.

Windows Server 2003 supports clustering using the OS standard function (Microsoft Cluster Service).

The availability of the whole system can be increased in a cluster configuration. If a failure occurs in the active system or the cabinet, the current task can be handed to a standby system or cabinet, thereby providing high availability. This function also enables the automatic detection of failures in important system resources ranging from hardware to middleware and applications, implements autonomous control such as failover, and thereby enables operation to continue safely. Even the work done in periodic maintenance and scheduled stoppages such as a system configuration change can be performed without affecting the resources in use. As a result, the service availability time is greater.

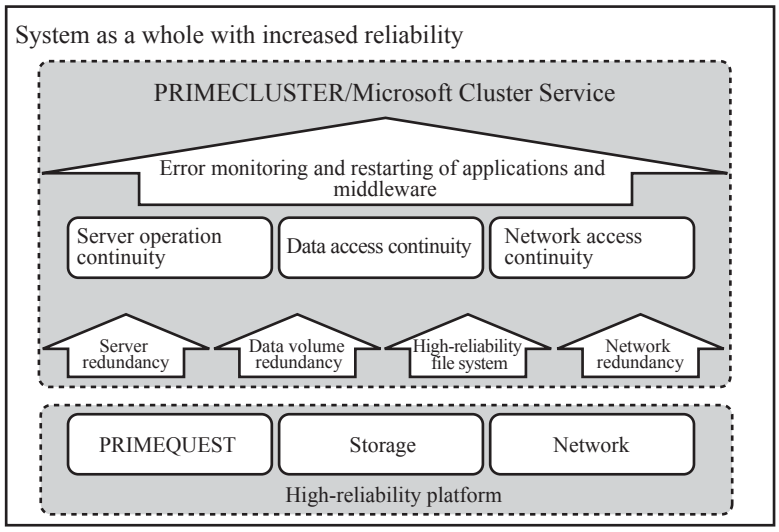


Figure 1.5 Clustering

1.2.2.5 Fujitsu middleware products

A high-reliability and high-availability system can be constructed by linking the PRIMEQUEST-series machines with the following middleware products that use Fujitsu's high-reliability and high-performance technologies and accumulated know-how:

- Interstage..... Business integration software
- Symfoware..... Database software
- Systemwalker..... Integrated operation management software

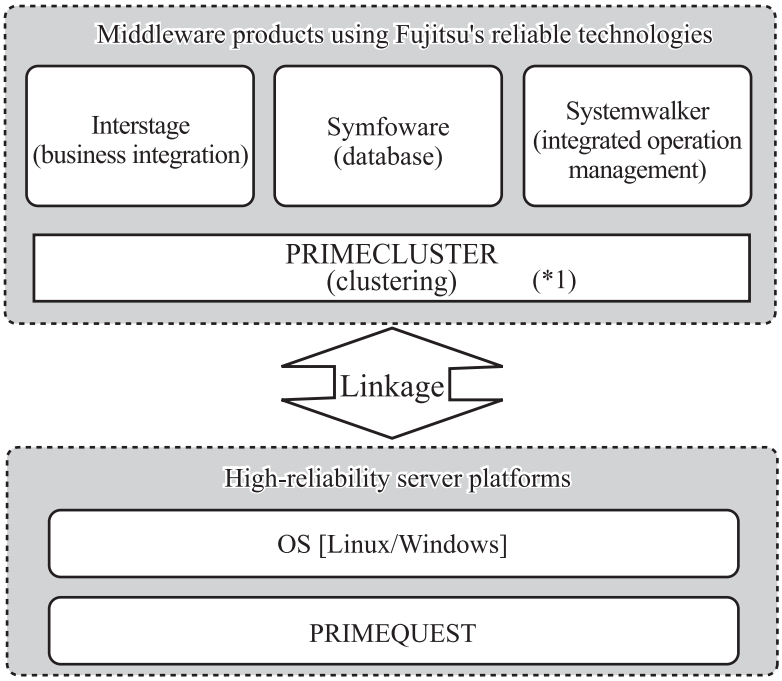


Figure 1.6 Fujitsu middleware products

*1 This is supported only for Linux.

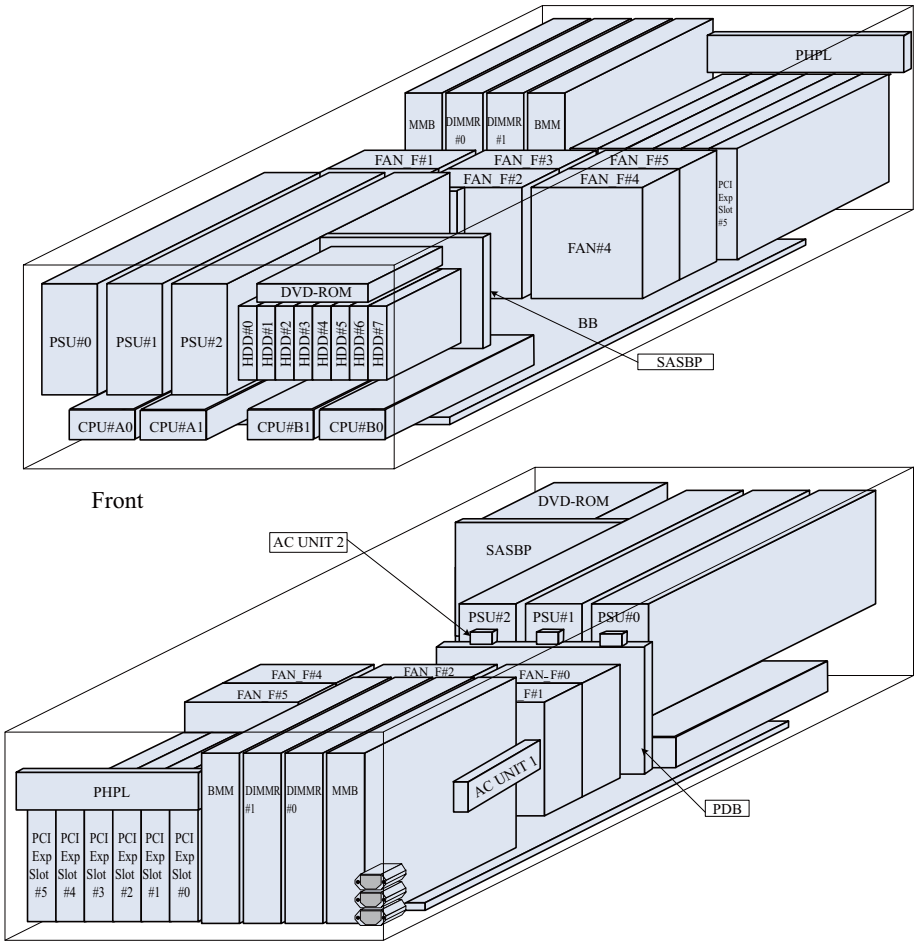
CHAPTER 2 General Hardware Configuration

This chapter describes the following hardware devices that compose a PRIMEQUEST-series machine:

- Main unit
  - Cabinet
  - Base Board (BB) (CPU, memory, BMM, PCI Express slot)
  - SASBP
  - HDD
  - DVD
  - LED
  - Management Board (MMB)
  - Power Supply Unit (PSU)

The configurations of PRIMEQUEST components are shown below.

Physical configuration



PSU : Power Supply Unit
CPU : Central Processing Unit
HDD : Hard Disk Drive
FAN : Fan
MMB : Management Board

DIMMR : Riser board for installing DIMM
BMM : BMC Module
BB : Base Board
PCI Exp Slot : PCI Express slot
PDB : Power Distribution Board

Figure 2.1 PRIMEQUEST 510A physical configuration (full package)

Table 2.1 Types and maximum numbers of boards and components

Board/component		Maximum number	Remarks
BB		1	
	CPU	4	At least one CPU is required.
	Memory (DIMM)	Maximum size: 64 GB (16 × 4 GB)	At least two memory modules are required.
	BMM	1	
	PCI Express slot	6	
HDD		8	-
MMB		1	-
DVD drive		1	-
PSU	100 VAC	3	The minimum is two.
	200 VAC	2	The minimum is two.
FAN (*1)		6	-

*1 5+1 redundancy

2.1 Cabinet

The main unit of the cabinet contains base boards (BB), and other components.

The cabinet also contains the management board (MMB) (dedicated unit for server management) for controlling and monitoring components in the cabinet. The MMB can be operated through a Web interface from a general-purpose personal computer.

Figure 2.2 shows the appearance of the cabinet with its front and rear covers removed (example of the full package).

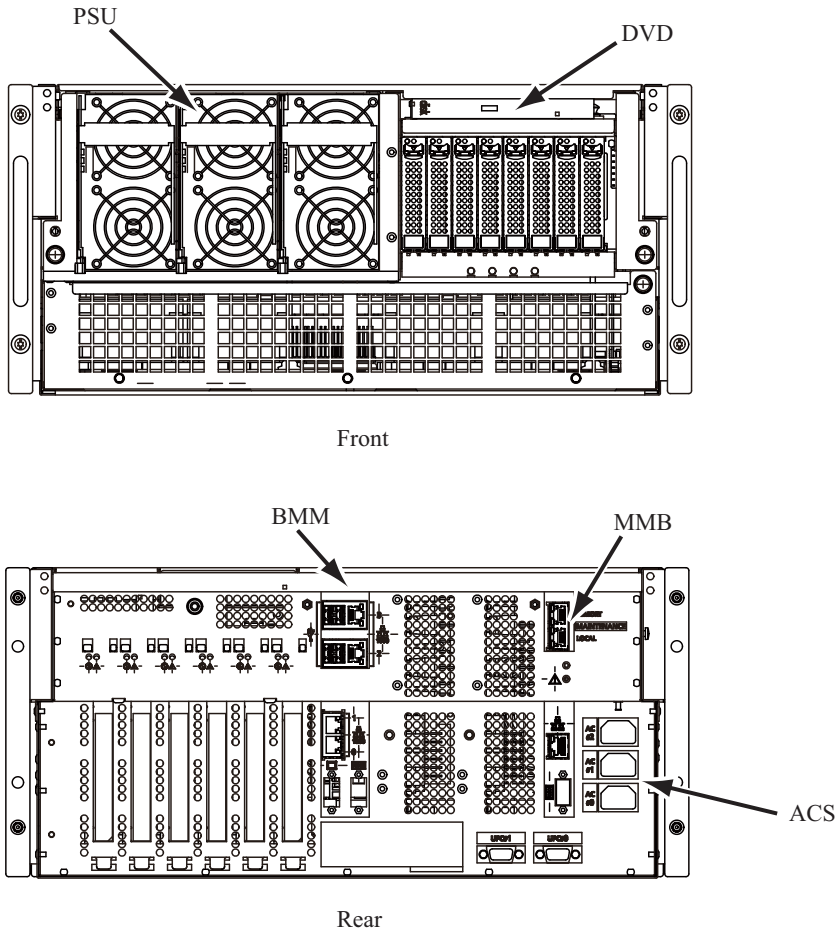


Figure 2.2 Main unit

2.2 Base Board (BB)

CPUs, memory, and various chipsets are mounted on the BB:

- The BB contains one to four CPUs and one set (two modules) to eight sets (16 modules) of memory.
- A DIMM Riser is used to mount DIMMs (memory). Up to eight DIMMs can be mounted on one DIMM Riser, and the system has two DIMM Risers. Therefore, 16 DIMMs can be mounted in total.

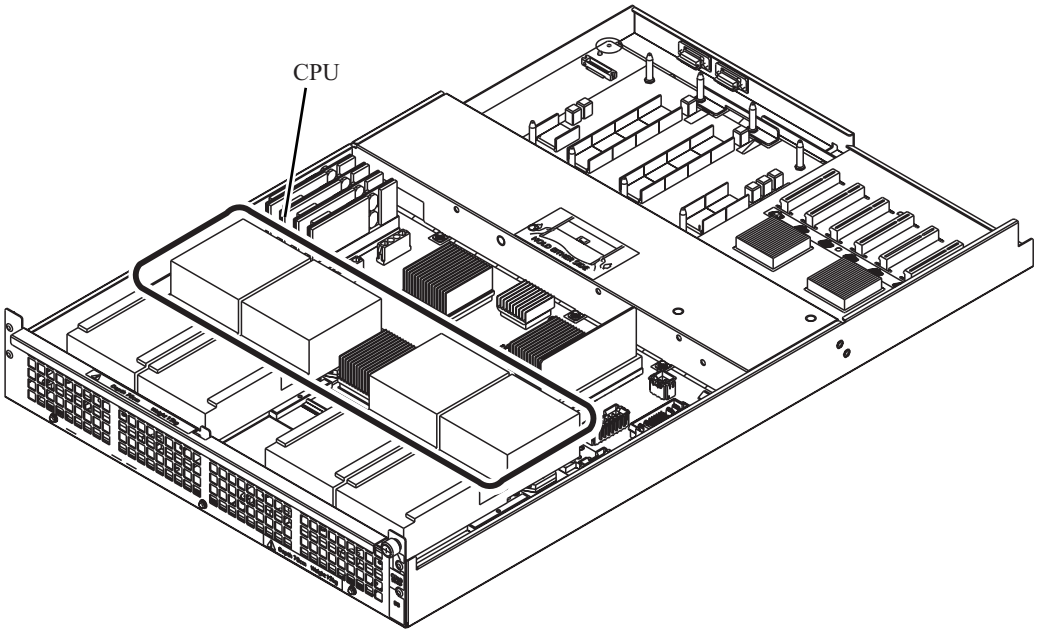


Figure 2.3 BB

2.2.1 CPU

This section describes the CPU.

- The CPU uses the Itanium. The reliability, availability, and serviceability (RAS) function is enhanced by the error detection and correction function incorporated for cache memory, etc. This improves business continuity.
- The Itanium improves performance because its 64-bit processors use an expanded memory address space and have higher computing speeds.
- One to four CPUs can be mounted.

2.2.2 Memory

This section describes the memory:

- The memory module (DIMM) uses double data rate synchronous dynamic random access memory (DDR2 SDRAM).
- The memory module can be added in sets of two modules of the same type.
- Single 4-bit block error correction - double 4-bit block error detection (S4EC-D4ED) is used as the error correction function for high system reliability. S4EC-D4ED is the DIMM error check function for 16 bytes (128 bits).
- The memory module is handled in sets of two DIMMs mounted in the same channel and the same slot on a LDX.
- One set (two modules) to eight sets (16 modules) can be mounted on the DIMM Risers mounted on one BB.

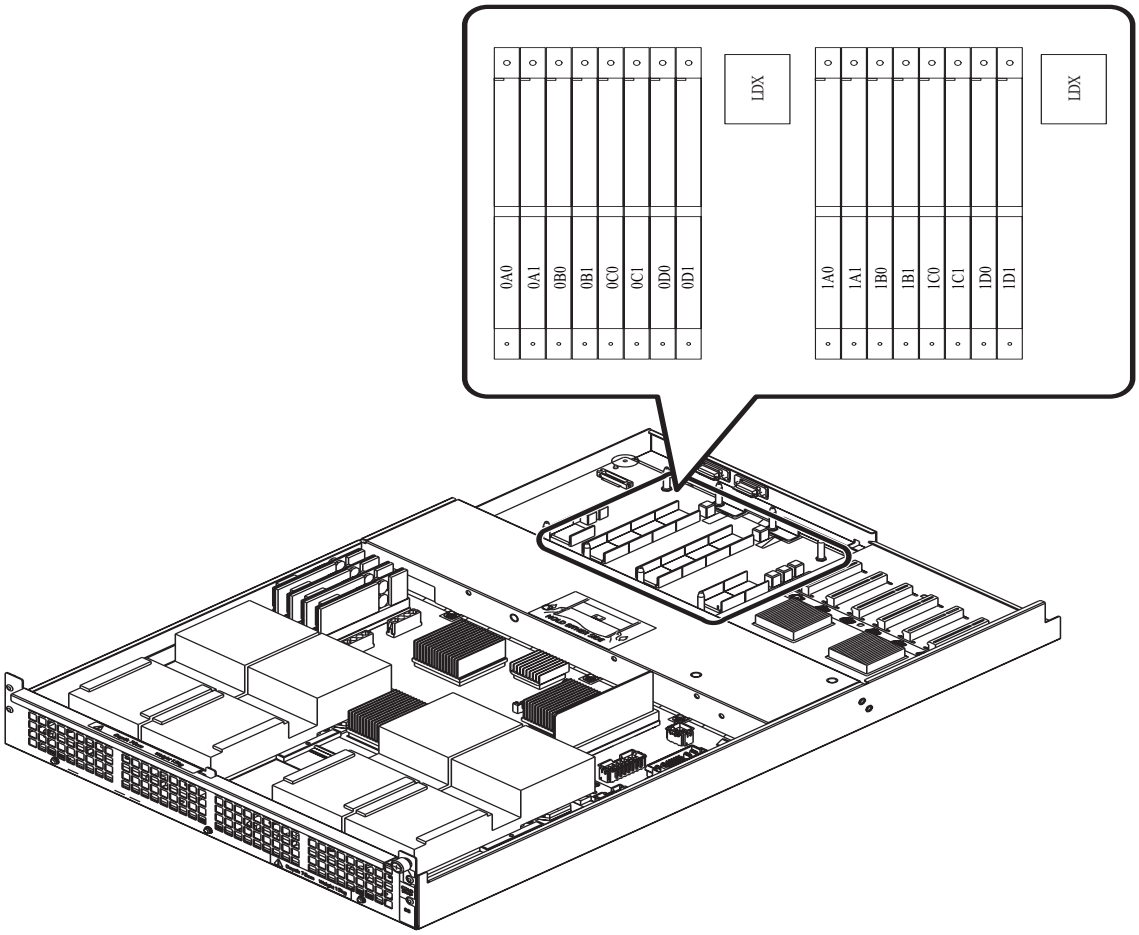


Figure 2.4 Memory (DIMM) slot numbers and locations

Notes on mounting DIMMs

Note the following points on mounting DIMMs:

- DIMMs must be added in sets of two DIMMs. One set is the minimum configuration unit.
- The two DIMMs in each set must be the same type.
- If the initial diagnosis detects a fixed one-bit error in a DIMM, isolation of the DIMM is by the unit of addition (two modules). For this reason, it is recommended to mount two sets of DIMMs or more (four modules or more).
- Different types of DIMMs can be mixed. For details, contact your Fujitsu certified service engineer.

2.2.3 BMM

The BMM is a board that manages input and output in the system. One VGA port, two USB ports, four LAN ports, and one COM port are mounted on the BMM.

Table 2.2 External interfaces for I/O resources

External interface	Quantity
USB	4
RS-232C	1
VGA	1
LAN	4

2.2.4 PCI Express slot

The PCI Express slot is a slot in which a PCI Express card is installed:

- Six PCI Express slots can be mounted in the PRIMEQUEST 510A.
- Whether a PCI Express card is hot swappable depends as follows on the OS installed on the partition:
  - Linux
    The PCI Express card inserted into a PCI card cassette is hot swappable.
  - Windows Server 2008
    See the PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide (C122-E087EN).
  - Windows Server 2003
    Hot swapping is not available.
- Each slot is configured with eight short lanes.

2.2.5 Compatible PCI Express cards

PRIMEQUEST supports the following PCI Express cards.

Remarks: For the latest information about the range of PCI Express cards supported, check with a Fujitsu certified service engineer.

Table 2.3 Compatible PCI Express cards

Type	Function
Fibre Channel card	Fibre Channel (4 Gbps × 1 port)
	Fibre Channel (4 Gbps × 2 ports)
Network-related PCI card	1000Base-T card (dual channel)
	1000Base-SX card (single channel)
SCSI card	Ultra320 SCSI card (dual channel)

2.3 SASBP

The SASBP has functions for HDD monitoring and LED lighting control. HDDs and DVD drives are mounted on the BB through the SASBP.

2.4 HDD

Each HDD contains a 2.5-inch hot-pluggable Serial Attached SCSI (SAS) drive. The cabinet can accommodate up to eight HDDs, which are controlled by the SAS controllers.

2.5 DVD

The mounted DVD drive is used for OS installation.

2.6 LED

PRIMEQUEST has a four-LED display function.
Each LED status indicates the unit status and whether a failure has occurred.

- Displaying the power status of the main unit (Power-LED)
- Indicating whether a unit is faulty (Alarm-LED)
- Identifying the unit under operation (Location-LED)
- Displaying the MMB status, for MMB status checks from outside the main unit (MMB-Ready-LED)

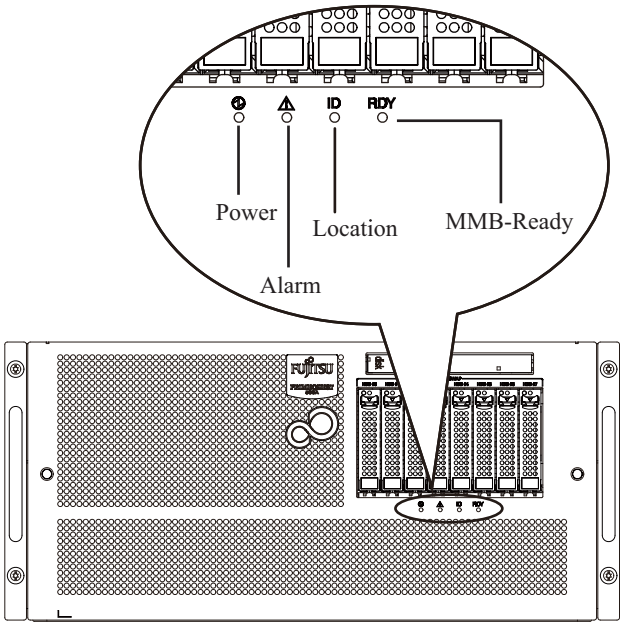


Figure 2.5 LED

Table 2.4 LED display

LED	Color	LED status	Unit status
Power-LED	Green	Off	Power to unit is off
		On	Power to unit is on
Alarm-LED	Orange	Off	Normal status
		On	Unit failure
Location-LED	Blue	Off	Unit cannot be recognized
		On/blinking	Unit is recognized
Ready-LED	Green	Off	Power to unit is off
		On	MMB is active
		Blinking	MMB is being initialized

2.7 Management Board (MMB)

The MMB is integrated in the main unit, and it manages the unit as a whole.

The MMB has a specific CPU and memory separate from the BB. MMB firmware runs on the MMB.

The MMB firmware provides a Web interface. As a result, a server can be managed and operated from a connected PC and other connected devices via a Web browser. Since the MMB centrally manages the whole system, unit management does not require any specific console or software.
The MMB has the following management functions:

- Cabinet power control
- Configuring settings, displaying the status, and various other operations from the MMB Web-UI
- Remote control of unit power and unit resets
- Monitoring sensors and other devices
- Reset processing
- Scheduling operations
- REMCS operations
- Setting the Alarm E-mail

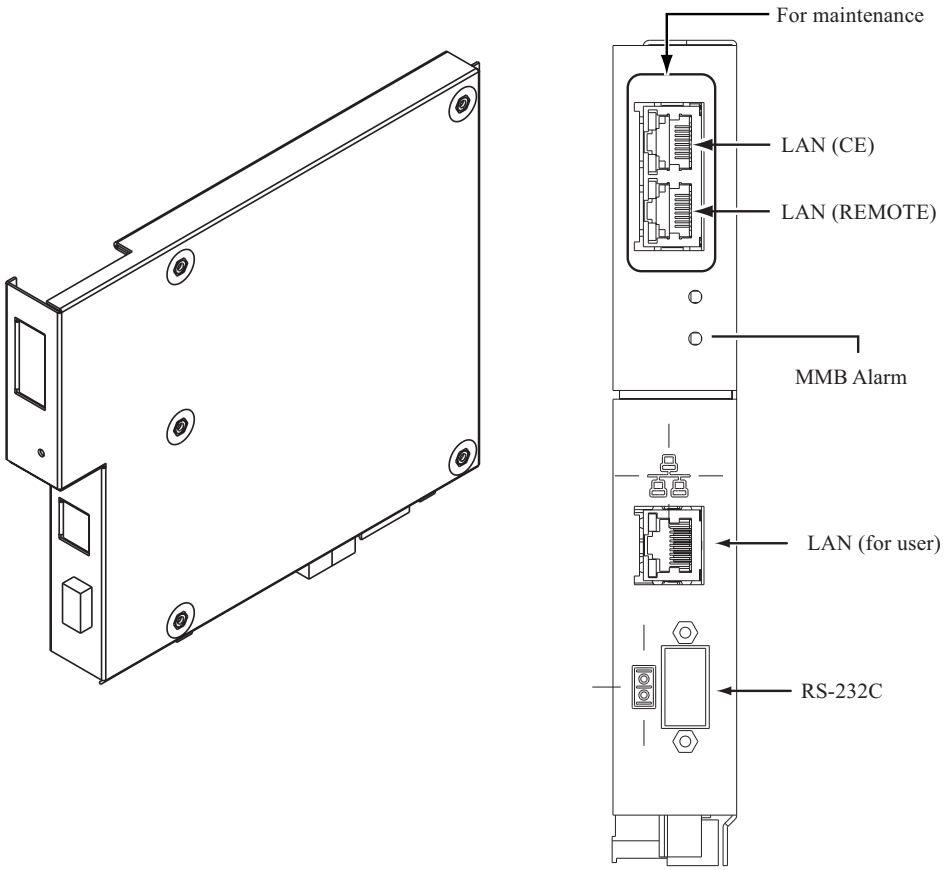


Figure 2.6 MMB

Table 2.5 MMB external interfaces

External interface		Count	Remarks
RS-232C		1	Used when the Fujitsu certified service engineer performs the setup of the equipment. It is not used during usual operation.
LAN	100Base-TX	3	• Used for console connection: user port • Used for maintenance - [CE]: port of the Fujitsu certified service engineer - [REMOTE]: REMCS port

The MMB includes a built-in hub that has the VLAN function. The VLAN function controls communication between the REMCS port and a partition, thereby achieving a high security level. For details, see Section 4.1.1.1, "Management LAN."

Remarks: When Speed/Duplex for the LAN port on the MMB is set to an item other than "AUTO," use a cross cable to connect a switching hub to the MMB LAN port.

2.8 Power Supply Unit (PSU)

Each PSU converts AC input to 48 VDC. The PRIMEQUEST 510A supports 200 V as an option in addition to 100 V.

This unit implements a single power feed of 100 V without redundancy. It supports a dual power feed option.

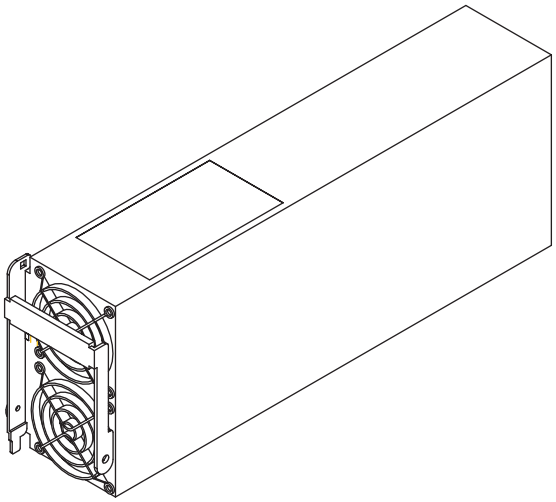


Figure 2.7 PSU exterior view

Table 2.6 PSU specifications

Item		Specification
Output	Rated voltage	48 VDC
	Rated power	1450 W
Size		69.5 mm x 123.5 mm x 344.3 mm
Mass		3.3 kg

Table 2.7 Number of SBs and IO Units mounted and PSU configuration

Input power supply	Feeding method	Number of PSUs mounted	PSU configuration
100 VAC	Single power feed	3	2+1 redundancy
200 VAC	Single power feed	3	2+1 redundancy
	Dual power feed	4	2+2 redundancy (2 per feed)

CHAPTER 3 General Software Configuration

This chapter describes the PRIMEQUEST software configuration.

As shown in Figure 3.1, the software configuration consists of MMB firmware installed on the management board (MMB) and the installed software, such as the OS (Linux or Windows), PRIMEQUEST Server Agent (PSA), and related software.

System management operations and configuration are performed from the MMB Web-UI. All system management operations, including operations related to PSA, can be performed from a Web browser running on a general-purpose PC through linkage with a Web server using the MMB firmware.

PRIMEQUEST supports large-scale enterprise environments that require high reliability, by ensuring close linkage between hardware and software.

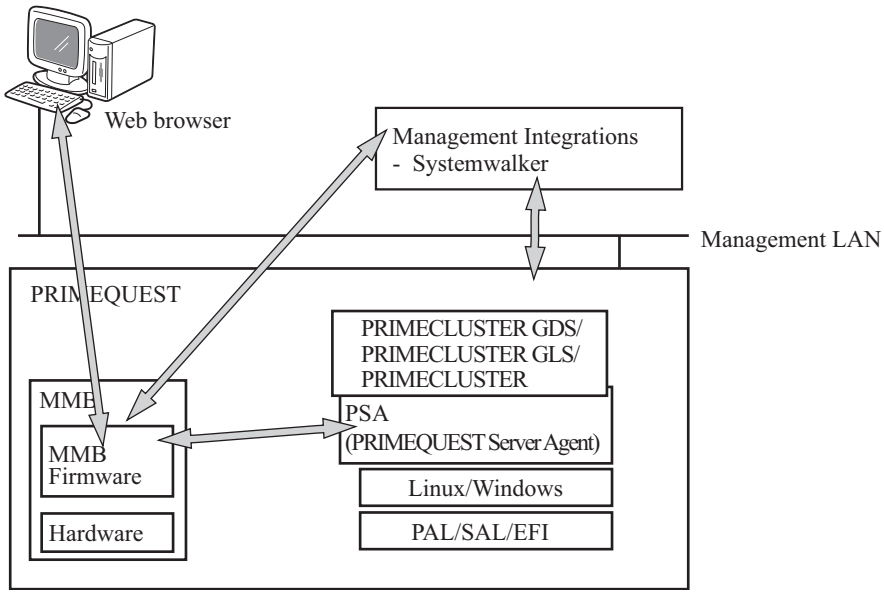


Figure 3.1 PRIMEQUEST software configuration

3.1 Firmware

This section describes the different types of PRIMEQUEST firmware.

3.1.1 Processor Abstraction Layer (PAL)

PAL is part of the Itanium architecture, and it abstracts the functions unique to IPF CPU hardware to higher layers.

PAL encapsulates processor functions that may vary from one IPF CPU to another. It gives a consistent view of the hardware to SAL and the OS, which are located on a higher layer.

The PAL functions include diagnosis, initialization, and error processing of the CPU hardware.

3.1.2 System Abstraction Layer (SAL)

SAL is firmware that abstracts the platform and provides the OS with the specified interface.

The SAL functions include diagnosis and initialization of the PRIMEQUEST platform including ASIC and memory configuration control.

3.1.3 Extensible Firmware Interface (EFI)

EFI is PRIMEQUEST boot firmware.

EFI provides a boot environment that is not dependent on any platform for an OS such as Linux and Windows.

The EFI functions include loading EFI drivers, diagnosis and initialization of the OS memory area, write/read diagnosis of I/O registers, boot device selection, and setting the order of boot devices.

3.1.4 BMC firmware

BMC firmware is incorporated in the BB. BMC firmware manages the BB configuration, and monitors and controls hardware in mounted units.

3.1.5 MMB firmware

MMB firmware is incorporated in the MMB. MMB firmware is linked with different types of firmware (e.g., PAL/SAL/EFI and BMC firmware) for operations such as managing the hardware configuration, hardware monitoring, and power control.

For details on the MMB, see CHAPTER 4, "Hardware System Management."

3.2 OS (Linux or Windows)

PRIMEQUEST supports the following OSs.

- Linux
  - Red Hat® Enterprise Linux® AS (v.4 for Itanium)
  - Red Hat® Enterprise Linux® 5 (for Intel Itanium)
  - SUSE™ Linux Enterprise Server 9 for Itanium Processor Family (*1)
  - SUSE™ Linux Enterprise Server 10 for Itanium Processor Family (*1)

  *1: Supported mainly for markets outside Japan.
- Microsoft Windows Server 2003
  - Microsoft®Windows Server® 2003, Enterprise Edition for Itanium-based Systems
  - Microsoft®Windows Server® 2003, Datacenter Edition for Itanium-based Systems
  - Microsoft® Windows Server® 2008 for Itanium-Based Systems

3.2.1 Linux

Linux supports large-scale enterprise environments that require high reliability:

(1) Support of large-scale systems

Linux supports construction of large-scale systems by increasing scalability with the following functions:

- 64-bit virtual space

  Large-scale business applications can be developed without consideration of the virtual space size.

- Support of large-scale SMP

  A high-performance system with a large-scale SMP configuration can be constructed using high-speed interconnect technology.

- Increased number of connected devices

  Large-scale storage systems can be managed.

- Support of large-capacity storage

  Systems that handle large-volume data can be managed.

(2) Higher reliability and availability

Linux works together with hardware to increase reliability and availability as follows:

- Enhancing drivers

  Error handling is enhanced, thereby increasing reliability.

- Machine Check Abort (MCA) information logging function

  Information on recoverable errors is always collected in a log and used for protection against the errors and maintenance.

- PCI hot plugging

  PCI cards that support PCI hot plugging are hot-swappable.

- Guaranteed uniqueness of device names

  Each device name is guaranteed to be unique even if a device is removed or replaced because of a disk failure.

3.2.2 Windows

PRIMEQUEST supports Microsoft Windows Server 2003 and its subsequent versions.

3.3 Supplied Software

The software programs supplied as standard with the PRIMEQUEST-series machines are described below.

For details on the supplied software, see Appendix A, "Software Supplied with PRIMEQUEST Hardware."

3.3.1 PSA (system management software)

PSA handles hardware failure monitoring, configuration management, and other tasks.

For details on the PSA, see CHAPTER 4, "Hardware System Management."

SIRMS

When REMCS is used for operation, SIRMS collects software configuration information on each partition and troubleshooting materials in case of a software failure.

3.3.2 Drivers

SCSI-related drivers, LAN-related drivers, and other types of drivers are supplied.

3.3.3 SystemcastWizard Lite

SystemcastWizard is used to install OSs, back up systems, and restore systems.

3.3.4 System data output tool (fjsnap)

fjsnap collects the system information (e.g., configuration information, operation information, definition information, and logs) required for a failure investigation in the event of a PRIMEQUEST system failure. This software can run in Linux.

3.4 Linked Software

This section describes optional software that can be used in linkage with the PRIMEQUEST-series machines.

3.4.1 PRIMECLUSTER

PRIMECLUSTER is software for cluster operations. When Linux is installed as its operating system, install these programs as necessary. Cluster operations can be performed between cabinets in the PRIMEQUEST system.

For details on cluster operations, see CHAPTER 7, "Clustering."

For details on PRIMECLUSTER, see the PRIMECLUSTER manuals.

3.4.2 PRIMECLUSTER GDS

When Linux is installed as its operating system, PRIMECLUSTER GDS is volume management software that helps increase the availability of data stored in disk drive units and facilitates operation management of the data. PRIMECLUSTER GDS protects disk data against hardware failures and operating errors by users, and it provides assistance in operation management of disk drive units.

For details on the redundant configuration, see CHAPTER 5, "Redundant Configuration."

For details on PRIMECLUSTER GDS, see the PRIMECLUSTER GDS manuals.

3.4.3 PRIMECLUSTER GLS

When Linux is installed as its operating system, PRIMECLUSTER GLS is software that uses multiple network interface cards (NICs) to ensure redundancy of the network transmission paths connected to its system in order to provide high reliability in all communication.

For details on the redundant configuration, see CHAPTER 5, "Redundant Configuration."

For details on PRIMECLUSTER GLS, see the PRIMECLUSTER GLS manuals.

3.4.4 PRIMEQUEST System Disk Mirror for Windows (PSDM)

PRIMEQUEST System Disk Mirror for Windows (referred to below as PSDM) is software that provides a redundant configuration for a PRIMEQUEST system by mirroring the system disk on which the Windows OS is installed. It thereby enhances system availability.

Mirroring is a function for data redundancy, duplicating disk data to multiple disks. This ensures continuous accessibility to a valid copy of the data even if failures occur on some of the disks.

The advantages provided by PSDM include mirroring in units of individual disks and dynamic setting of EFI boot entries.

For details on PSDM, see the *PRIMEQUEST System Disk Mirror for Windows User's Guide* (B1FN-5771).

3.4.5 Systemwalker

Systemwalker is integrated operation management software that manages systems, networks, resources, etc., to support Internet-based businesses.

For details on Systemwalker, see the Systemwalker manuals.

3.4.6 Backup and restoration software

The following backup and restoration software can be used in linkage with the PRIMEQUEST-series machines:

- PRIMECLUSTER GDS Snapshot
- ETERNUS SF AdvancedCopy Manager
- VERITAS NetBackup™
- NetVault
- NetWorker

For details on backup and restoration, see CHAPTER 8, "Backup and Restoration."

Remarks:

Access the Fujitsu Web site to confirm the latest information on linkage software.

CHAPTER 4 Hardware System Management

This chapter describes the basic mechanism of hardware system management, the functions of the management board (MMB) that is the core component that performs total hardware management, and the functions of the PRIMEQUEST Server Agent (PSA) that monitors the system in the BB.

- Basic Mechanism
- MMB Functions
- PSA Functions

4.1 Basic Mechanism

Figure 4.1 shows the basic mechanism for hardware system management in PRIMEQUEST series machines.

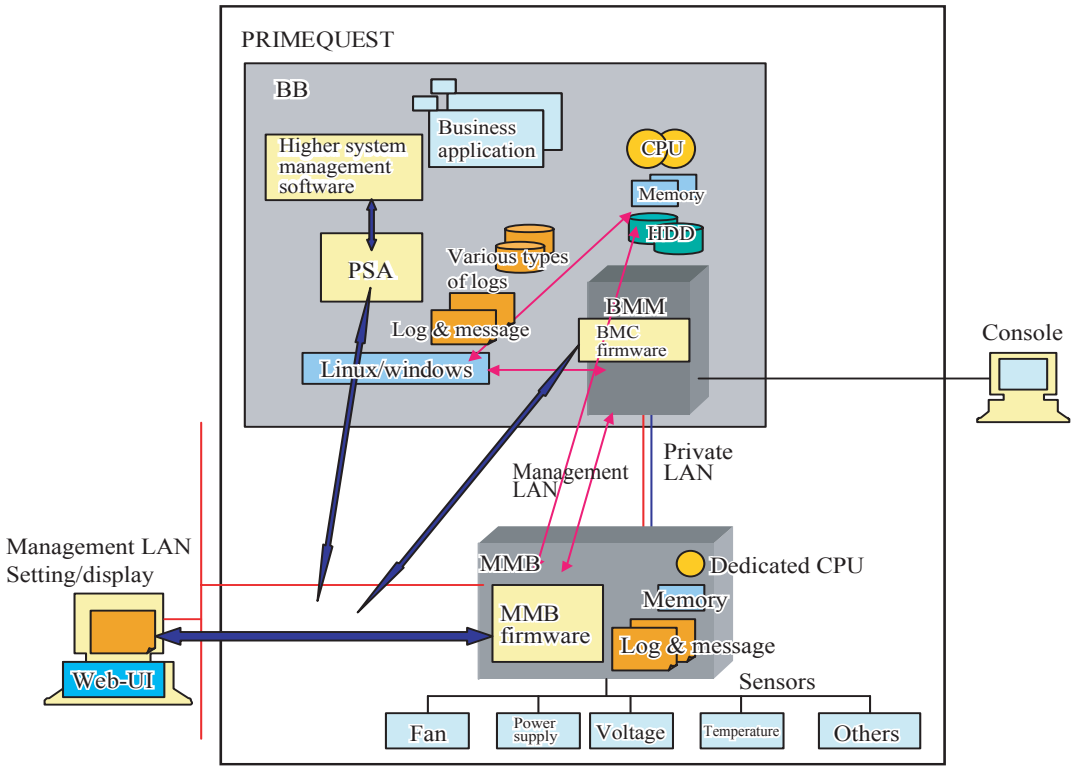


Figure 4.1 Basic mechanism

The MMB performs total management of the PRIMEQUEST hardware system.

The MMB manages the mounted components (BB, chip sets, power supply units, fans, LANs, etc.), and controls the system clock, management LAN and private LAN. The MMB sets various types of operations such as console redirection. These functions can be operated from the MMB Web-UI.

As shown in Figure 4.1, the MMB is connected to each hardware component through the LANs. The MMB is also connected with various types of sensors through the internal bus. The MMB firmware always monitors and manages the entire system by linking with the BMC firmware and the PSA, by using the networks in the cabinet.

The PSA runs on Linux or Windows. The PSA monitors the configuration and state of the connected hardware units such as the BB.

Monitoring targets include all of the hardware units incorporated such as the CPUs and memory of the BB, the HDDs and PCI slots.

The PSA links with the MMB, and is operated from the MMB Web-UI.

To monitor the hardware, information is exchanged by using an information path separated from the information path of the business system, that is, the private LAN and management LAN. Not depending on the business system, a particular information system with extensive multiple connections encompassing the hardware unit is used. More specifically, various types of firmware around the MMB firmware acquire information detected by the OS and various types of drivers and information detected by various types of sensors and prompts users to take corrective action if necessary.

Monitoring items are set and changed and details such as management values are set from the MMB Web-UI.

This console information is input from and output to the COM port of the BMM.

The PRIMEQUEST series has a function that redirects the console information to the management LAN having the external interface through the firmware and private LAN (console redirection). This function enables the console to be operated from a general-purpose PC connected to the management LAN.

For details of the console redirection function, see Section 4.2.4, "Console redirection."

Processing for the MMB and PSA can be executed from the MMB Web-UI. Rather than from a dedicated console, this can be done from a general-purpose PC connected to the management LAN.

Accordingly, functions used by the system administrator or user during operation through the MMB and MMB Web-UI are operated through the Web-UI.

The main functions of the MMB are described in Section 4.2, "MMB Functions."

The main functions of the PSA are described in Section 4.3, "PSA Functions."

4.1.1 LAN configuration (management LAN/private LAN/business LAN)

The PRIMEQUEST cabinet contains three different LANs in addition to the internal bus connection. Each of these LANs has a different use.

These three LANs are the management LAN, private LAN, and business LAN. They are separated for security and load distribution purposes.

The management LAN is provided for system management purposes and has external ports for the user (system administrator), Fujitsu certified service engineer, and remote customer support system (REMCS) on the MMB.

The private LAN is provided for internal control of the system.

PRIMEQUEST is connected to an external LAN through a LAN card. This LAN is called the business LAN.

Fujitsu recommends configuring the management LAN and business LAN on different subnets to prevent a problem on the management LAN or business LAN from stopping operations or business.

4.1.1.1 Management LAN

The configuration of the management LAN is shown in Figure 4.2.

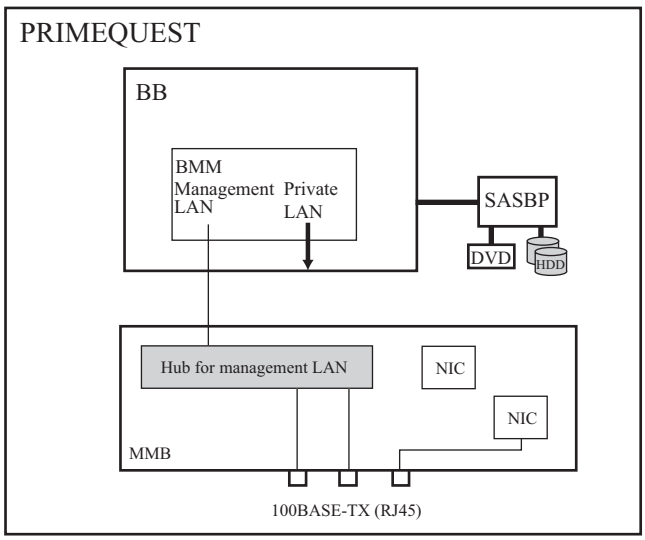


Figure 4.2 Management LAN

To manage the PRIMEQUEST system, the management LAN connects the MMB to the system in the BB, and the MMB to a LAN outside the cabinet. The MMB operates and controls the system in the BB through the management LAN. The system in the BB sends its log information in the OS and hardware configuration information to the MMB through the management LAN.

Applications of the management LAN are described below.

- Communication between the MMB and the system in the BB
  Sends and receives information about Web linkage, SNMP linkage, and REMCS linkage between the MMB and the PSA managing the system in the BB
- Operator console (Web-UI, SSH, etc.)
  Connected as the console of the MMB and the system in the BB
- Management console
  Used for linkage with the operation management system of an external server.
- Integrated monitoring software linkage
  Used for linkage with integrated monitoring software such as Systemwalker
- Firmware update
  Used for updating various types of firmware from the MMB
- Time synchronization
  Used for synchronizing the time of each OS while the MMB operates as the NTP server

- External connection

  The following three ports are provided for external connection of the management LAN.

  - User port

    For management purposes such as connecting an external management server and using the operation console
  - Fujitsu certified services engineer’s port

    For connection to a maintenance terminal
  - REMCS port

    For connection to the REMCS Center

The ports for REMCS connections can be selected based on the customer's requirements as follows:

- P-P connection

  Connect a dial-up router directly to the REMCS port.
  - Si-R30/130 dial-up router (10BASE-T4 port on the LAN side)
  - Si-R 170 broadband router (10BASE-T4 port on the LAN side)

  * For the connection to either of the above routers, the settings of the MMB port must be 10 Mbps and half-duplex fixed.
- Internet connection

  For an REMCS connection established through the network that is connected to a user port through a firewall, the REMCS port need not be used.

  For an REMCS connection established through a network that is not connected to a user port because of security requirements, use the REMCS port.

Speed/Duplex setting for the management LAN port

Set "AUTO" for Speed/Duplex of the USER port.

Remarks: If Speed/Duplex of the USER port is set to a value other than "AUTO," the Auto MDI/MDI-X function is disabled. To prevent this problem, use the following cable:
To connect a switching hub device: Crossover cable
To directly connect a PC: Straight cable

Also, note the following about the REMCS port:

- The REMCS port does not have the Auto MDI-X function.
  Use the following cable regardless of the Speed/Duplex setting:
  To connect a switching hub device: Straight cable
  To directly connect a PC: Crossover cable
  Note: A different cable is used compared with that of the USER port. Be careful.
- If Speed/Duplex of an external device connected to the REMCS port is set to a value other than "AUTO," also set "AUTO" for Speed/Duplex of the REMCS port. (Be sure to set the same value as that of the external device.)

The management LAN function

The hub for the management LAN can be found on the MMB. On this hub, cables for network of the system in the BMM, user port, Fujitsu certified service engineer's port, and REMCS port are concentrated.

The VLAN function of the PRIMEQUEST management LAN provides the following two types of mode. You can select the mode appropriate for the operation method you are using.

- Management LAN mode

(1) Communication connection mode (Port enable mode)
This mode enables communication between ports of the system in the BMM. However, communication between the REMCS/Fujitsu certified service engineer's port and the user port, and communication between the REMCS/Fujitsu certified service engineer's port and BMM ports are not supported. Y in Table 4.1 indicates that communication between the corresponding ports is available.

Table 4.1 Communication connection mode (Port enable mode)

No VLAN mode	User port	CE maintenance port	REMCS port	MMB	BMM ports
User port	Y	N	N	Y	Y
Fujitsu certified service engineer's port	N	Y	N	Y	N
REMCS port	N	N	Y	Y	N
MMB	Y	Y	Y	Y	Y
Ports of the system in the BMM	Y	N	N	Y	Y

Y: Communication possible, N: Communication impossible

(2) Communication disconnection mode (Port disable mode)
This mode is provided for enhancing security.
This mode disconnects communication of the management LAN with BMM ports. Intra-cabinet communication between an MMB and BMM port is not possible. Y in Table 4.2 indicates that communication between the corresponding ports is available.

Table 4.2 Communication disconnection mode
(Port disable mode)

Port disable mode	User port	Fujitsu certified service engineer's port	REMCS port	MMB	BMM ports
User port	Y	N	N	Y	N
Fujitsu certified service engineer's port	N	Y	N	Y	N
REMCS port	N	N	Y	Y	N
MMB	Y	Y	Y	Y	N
Ports of the system in the BMM	N	N	N	N	Y

Y: Communication possible, N: Communication impossible

4.1.1.2 Private LAN

The configuration of the private LAN is shown in Figure 4.3.

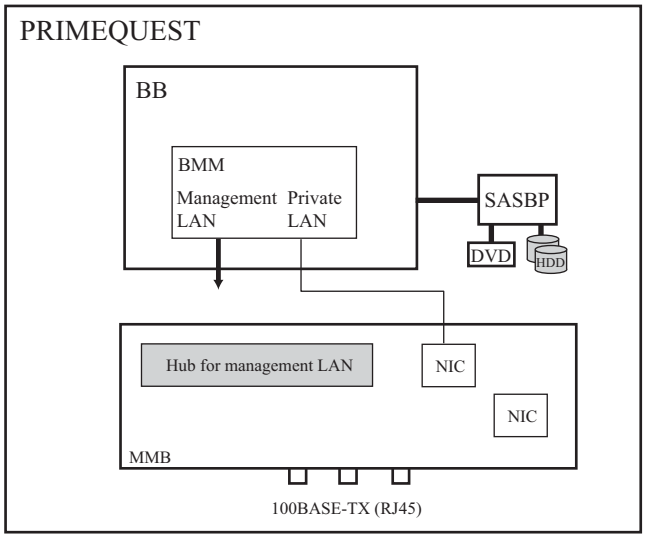


Figure 4.3 Private LAN

The private LAN is an internal control LAN for communication among firmware items installed in hardware components. The private LAN enables communication among the BMC firmware mounted on a BMM and the MMB firmware. The private LAN cannot be used from the OS or an applications.

4.1.1.3 Business LAN

The business LAN is used for configuring the user's business system.

- The business LAN is built in an external device connected to a 1000BASE-T card mounted in a BMM.
- The 1000BASE-SX card and other cards are mounted in PCI slots in the BMM, and the LAN is independently managed and operated.

4.1.2 IP addresses of PRIMEQUEST management LAN

The PRIMEQUEST management LAN uses the IP addresses shown in Figure 4.4.

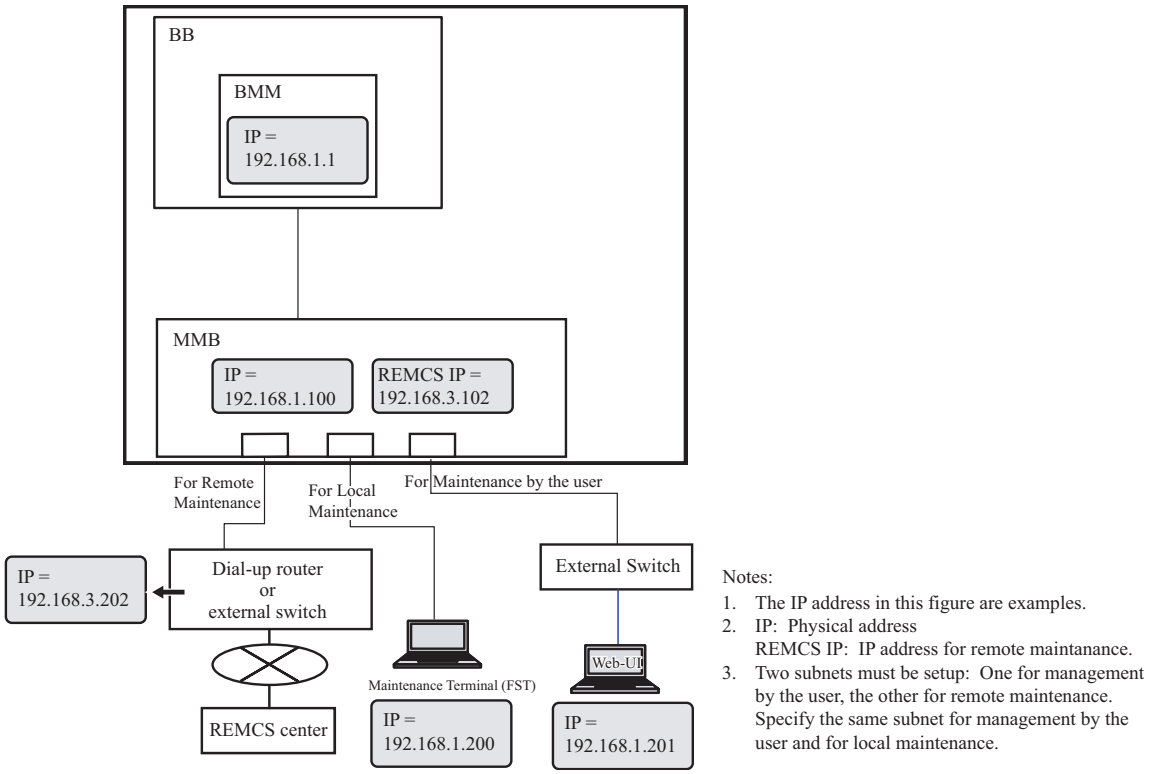


Figure 4.4 Management LAN network configuration and IP addresses

The main applications of IP addresses are described below.

- MMB IP addresses

  Each MMB has an IP address for management and a virtual IP address for a REMOTE MAINTENANCE connection.

  The IP address for management is used for MMB Web-UI operations or as the SNMP access destination from the management server.

  Also, PSA Web-UI operations are performed via the MMB Web-UI.

- IP address set by the OS

  An IP address must be specified for the operating system. In physical terms, the IP address corresponds to a NIC in a BMM.

  This IP address is used to control linkages, such as the Web-UI linkage between PSA and the MMB, SNMP linkage, and REMCS linkage. It is also used to link PRIMECLUSTER to the MMB.

  Note: If the switching hub that supports a loop prevention function, such as the spanning tree protocol or domain separation, is used for an external switch, suspend the loop prevention function by setting disable for the spanning tree protocol of the connection port between the switching hub and the main unit or by setting on to the domain separation.

  - The MMB transmits packets in communication of the following types: NTP, alarm e-mail, REMCS, and SNMP trap.
    Under the following conditions, firewall, mail, and other servers must be configured such that the packets with the IP addresses of MMB can pass through:
    - The destination server is an external server outside the firewall.
    - IP addresses are restricted by the mail server used.

    Also, if the REMOTE MAINTENANCE port is used under the above conditions, the servers must be configured such that the packets with the IP addresses for REMOTE MAINTENANCE can pass through.

4.2 MMB Functions

The MMB functions are indicated by ◆ MMB in Figure 4.5 in which all of the PRIMEQUEST functions are shown.

This section describes the MMB functions.

The MMB functions include the switching function for using the PSA functions.

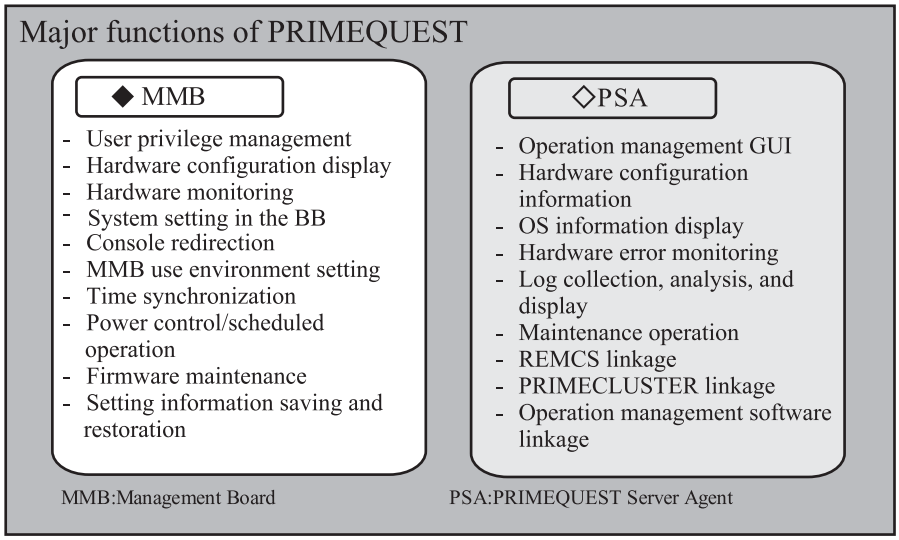


Figure 4.5 Main MMB functions

4.2.1 User privilege management

The user privileges necessary for accessing the MMB are classified as follows. System account elements such as a user name and a user privilege need to be registered. Users can use functions corresponding to their access privileges.

- User .....A user having this privilege is allowed only to reference the system state. The user cannot set system configuration information and cannot turn on and off power supplies of the system.
- CE.....A user having this privilege is allowed to reference the system state and to perform maintenance. The user cannot manage users and change network settings.
- Operator .....A user having this privilege is allowed to reference and set a system state. The user cannot manage users and change the configuration of a management LAN.
- Administrator .....A user having this privilege is allowed to execute all operations.

4.2.2 Hardware configuration display

This function displays the configuration and state of hardware.

This function displays the configuration of all components including fans and power supply units mounted in the PRIMEQUEST system and units in the components such as CPUs, memory, and HDDs.

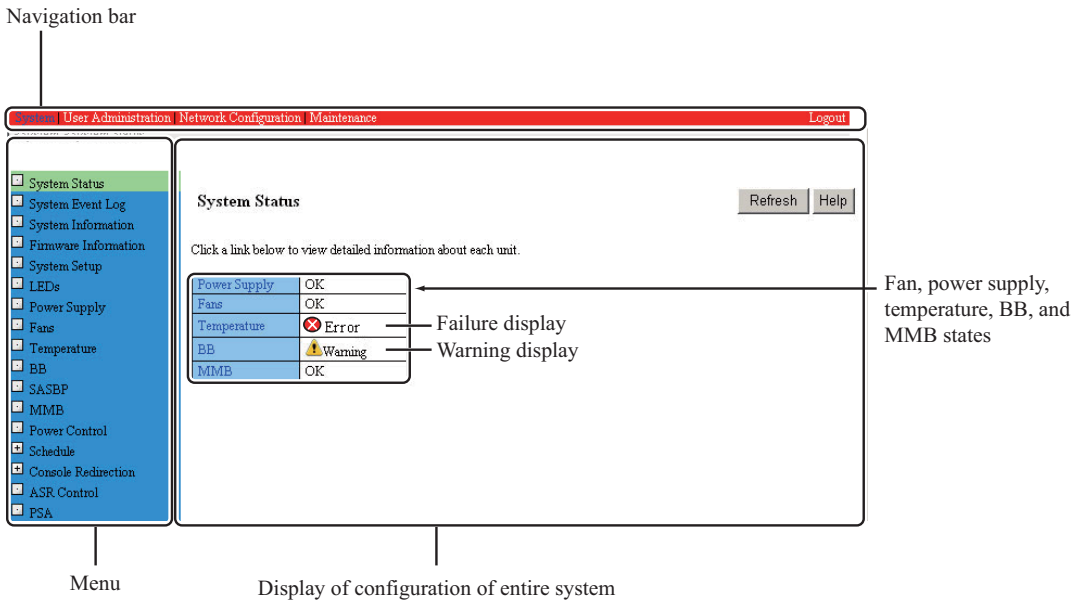


Figure 4.6 Example of the configuration display window

Figure 4.6 shows an example of displaying the configuration of the entire system. The states in which the configuration elements of the system are placed are also displayed in the configuration information. For example, whether the configuration elements are in the normal or abnormal state is displayed. In some cases where these elements are in the abnormal state, the system administrator or person in charge of maintenance needs take corrective action. Detailed information can be displayed by clicking the link in the display item field.

4.2.3 Hardware monitor

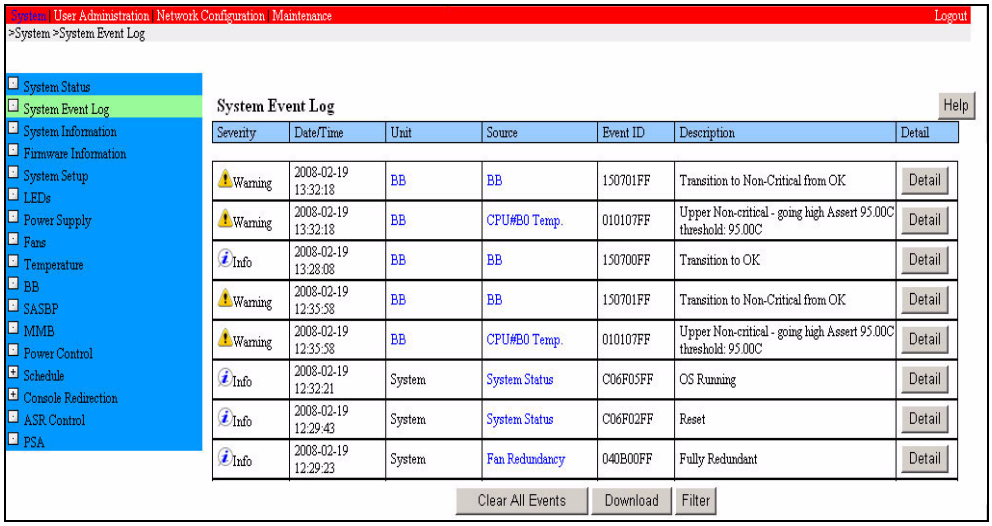
This function monitors for hardware failures and errors.

Hardware failures and errors are detected by various types of check mechanisms and sensors. For each detected failure or error, corrective action must be taken according to the settings of the corresponding components. All logs including logs about processing that was able to proceed following a failure or error occurrence (such as parts replacement and parts disconnection) are collected and error messages are issued.

Failures and errors are displayed by LEDs according to their severity. In addition, different types of reports are supported as follows: reports to applications for operation management following SNMP trap configuration, mail reports to the system administrator following alarm e-mail configuration, and reports to the REMCS Center following REMCS() registration.

To reduce the amount of displayed information, all logs and messages can be filtered prior to their display. The conditions for notifying information on abnormal states that have been detected can also be set.

An example of the display of system event logs is shown in Figure 4.7, and an example of the filter selection window is shown in Figure 4.8.



Severity	Date/Time	Unit	Source	Event ID	Description	Detail
Warning	2008-02-19 13:32:18	BB	BB	150701FF	Transition to Non-Critical from OK	Detail
Warning	2008-02-19 13:32:18	BB	CPU#B0 Temp.	010107FF	Upper Non-critical - going high Assert 95.00C threshold: 95.00C	Detail
Info	2008-02-19 13:28:08	BB	BB	150700FF	Transition to OK	Detail
Warning	2008-02-19 12:35:58	BB	BB	150701FF	Transition to Non-Critical from OK	Detail
Warning	2008-02-19 12:35:58	BB	CPU#B0 Temp.	010107FF	Upper Non-critical - going high Assert 95.00C threshold: 95.00C	Detail
Info	2008-02-19 12:32:21	System	System Status	C06F05FF	OS Running	Detail
Info	2008-02-19 12:29:43	System	System Status	C06F02FF	Reset	Detail
Info	2008-02-19 12:29:23	System	Fan Redundancy	040B00FF	Fully Redundant	Detail

Figure 4.7 Example of the system event log display window

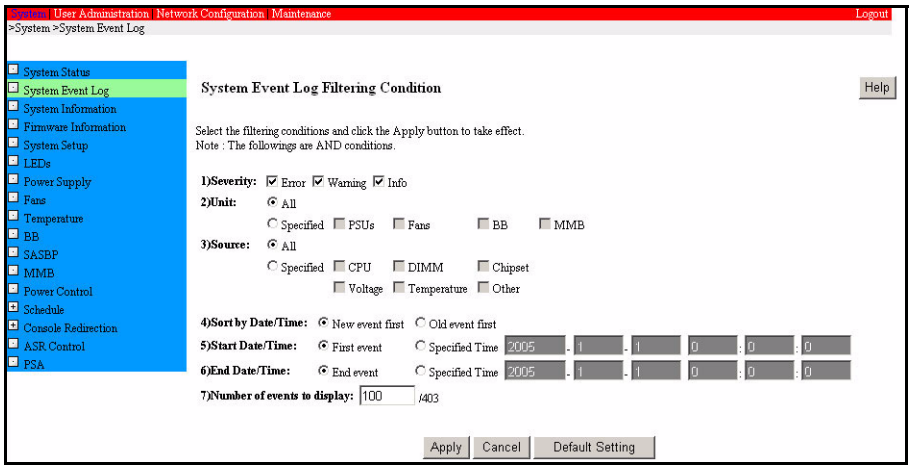


Figure 4.8 Example of the filter condition setting window

4.2.4 Console redirection

This function sets I/O destinations of console information through a management LAN.

The two methods below are available for OS input and output:

- Using the COM port of the BMM
- Using the management LAN through the MMB

Figure 4.9 shows the flow of console information sent when the information is redirected to the management LAN.

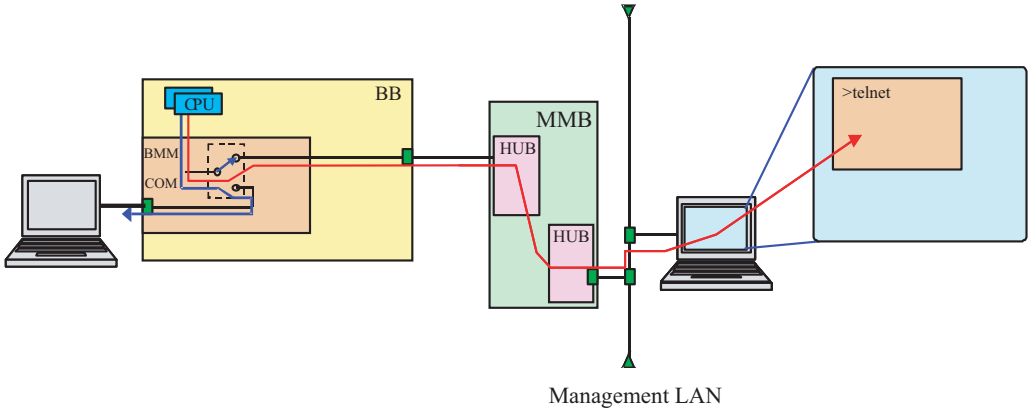


Figure 4.9 Flow of redirected information

When the information is redirected to the management LAN, a terminal connected to the management LAN can be used as a console.

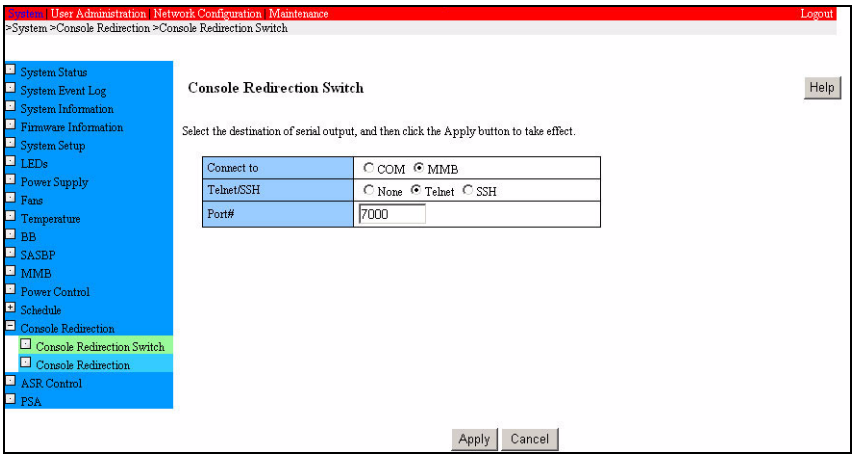


Figure 4.10 [Console Redirection Switch] window

4.2.5 MMB use environment setting

This function sets and changes the MMB use environments.

- User management
  For information on the user privilege for accessing the MMB, see Section 4.2.1, "User privilege management."
- Making and changing the network environment settings
  In this kind of operation, the MMB usage environments for the HTTP and Telnet is set.

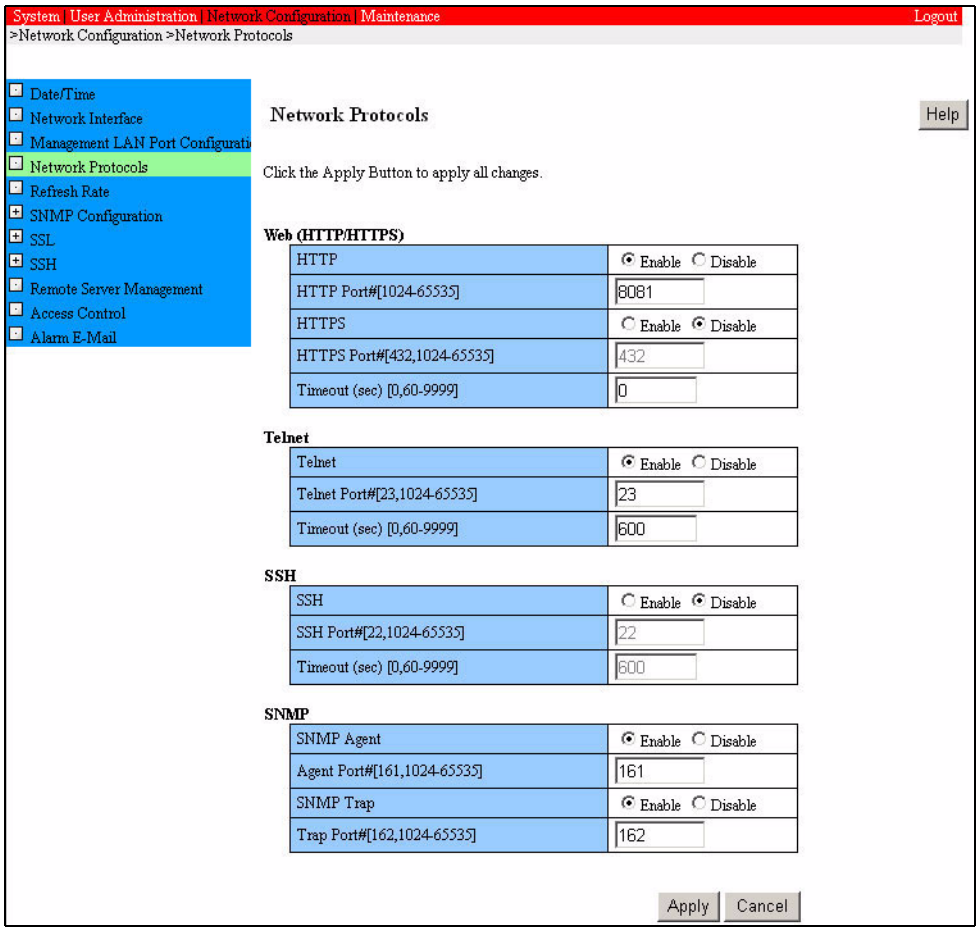


Figure 4.11 Example of setting the usage environment

- Access control

  To ensure security, an IP filter for permitting access to the MMB is set up. (Usable IP addresses are specified for each protocol.) Access to the MMB is allowed only from the set IP addresses.

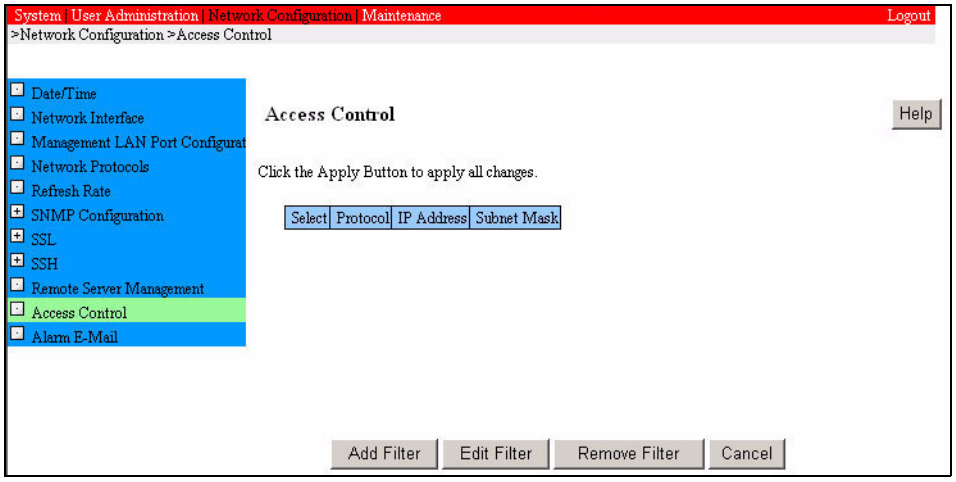


Figure 4.12 Example of the access control settings window

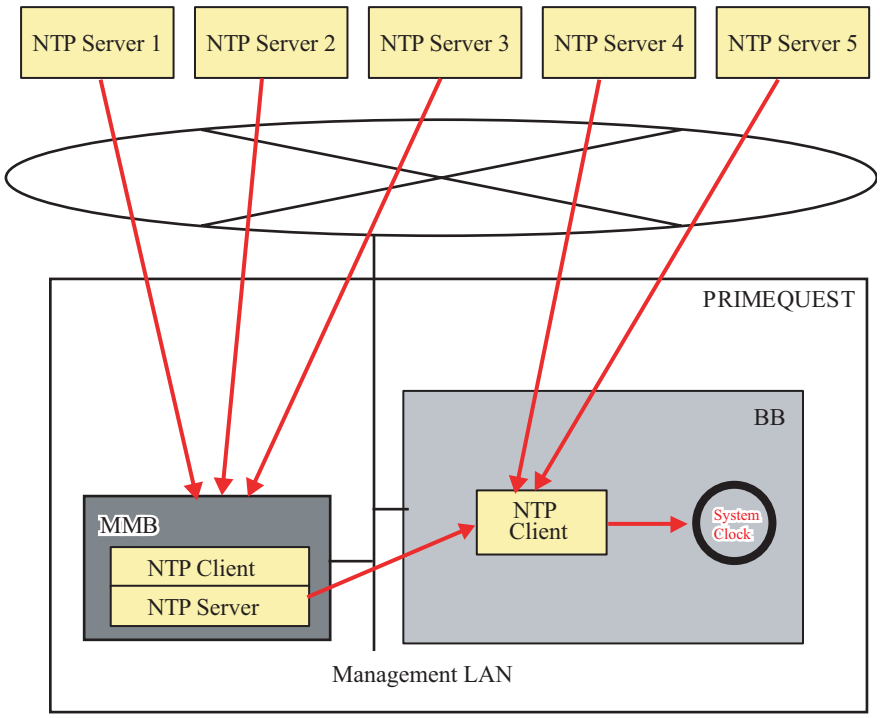
- Secure Sockets Layer (SSL) support

  Web and Telnet access is ciphered in the SSL. Secret keys and electronic certificates are generated in this layer.

4.2.6 Time synchronization

As shown in Figure 4.13, synchronization with the NTP server time is possible by accessing the external NTP server.

However, the MMB need not be specified for the NTP servers. For stable NTP operation, specify multiple NTP servers from the NTP client (Fujitsu recommends at least three).



Note: NTP server 1 to NTP server 5 are NTP servers using very high-precision clocks for connections to the Internet or intranets.

Figure 4.13 Time synchronization image

4.2.7 Power control/scheduled operation

This function sets a schedule for turning on and off the power supply and automatically controls the operation in accordance with the set schedule. Figure 4.14 shows an example of the display for scheduled operation.

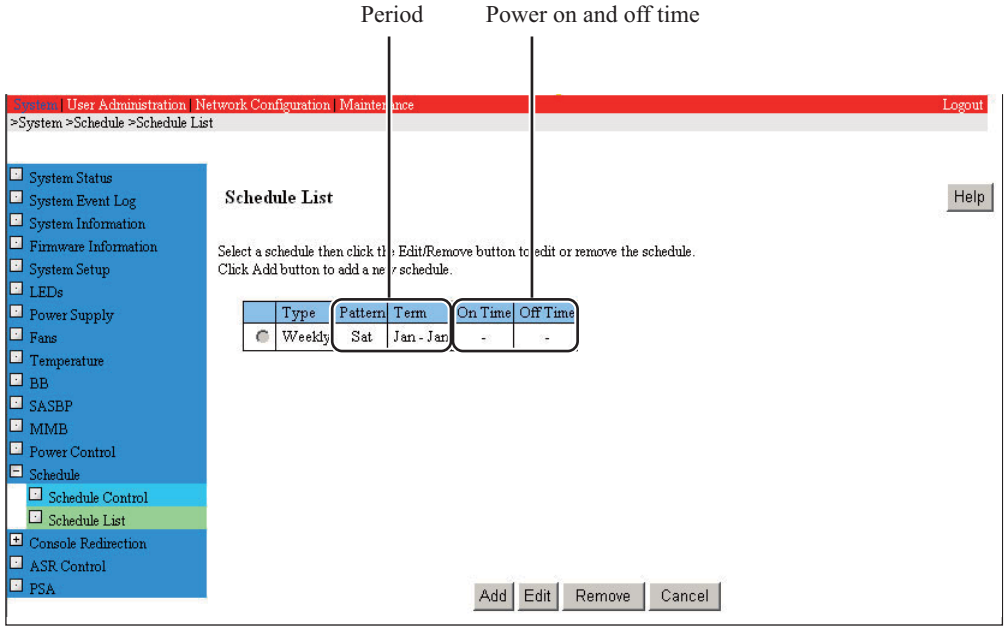


Figure 4.14 Example of the scheduled operation display window

4.2.8 Firmware maintenance

This function updates various types of firmware. The update is performed by Fujitsu certified service engineer.

The following firmware components are subject to update:

- MMB firmware (installed in the MMB)
- BMC firmware
- PAL/SAL/EFI firmware

4.2.9 Setting information save and restore

This function saves and restores EFI setting information and MMB setting information.

The MMB provides a save and restore function for EFI setting information as shown in Figure 4.15.

This function enables the following operations:

- After EFI settings are made from the EFI SETUP screen, the EFI setting information can be saved.
- If a failure occurs in a BMM and the BMM is swapped, the saved EFI information enables restoration of the original state.

Saved information can be stored on a remote terminal. Data saved on a remote terminal can be restored.

In addition to the function for the EFI configuration information, a save and restore function for saving and restoring MMB configuration information is provided as well.

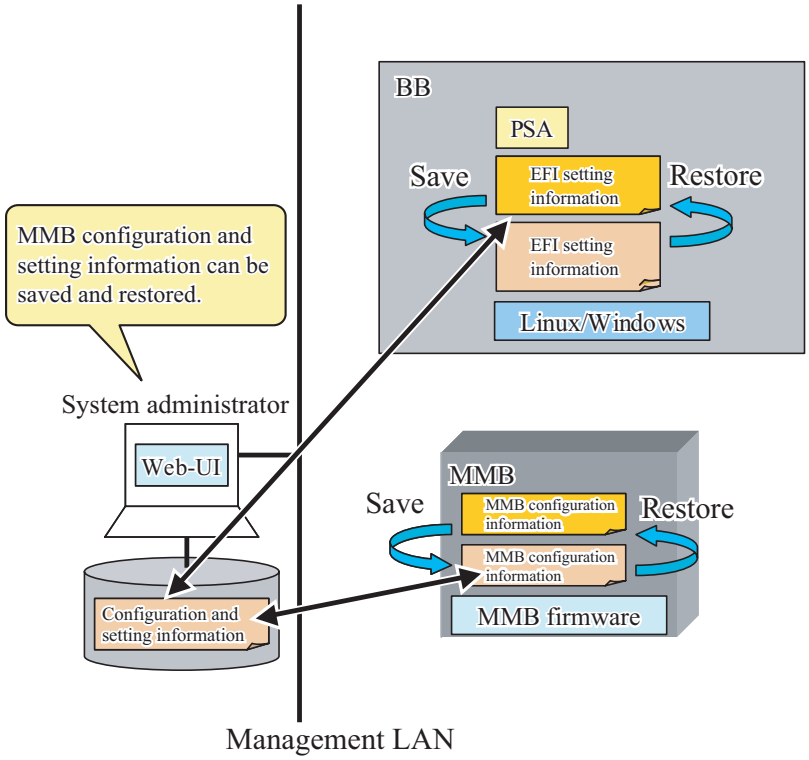


Figure 4.15 Schematic of configuration information save and restore operation

4.3 PSA Functions

The PSA functions are indicated by ◆ PSA in Figure 4.16 which shows all of the PRIMEQUEST functions.

This section describes each PSA function.

Remarks: If a PRIMEQUEST series machine is operated, be sure to install the PSA. If the PSA is not installed, the following restrictions apply:

- Notification of errors of I/O devices (PCI cards, HDDs, etc.) and trap notification to the administrator cannot be performed.
- Monitoring by using the watchdog function cannot be performed.
- The following error notification through failure prediction and trap notification to the administrator cannot be performed:
  - Correctable error count of a CPU, DIMM, or chip set, that exceeds the threshold
  - Exceeded threshold in the S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology system) monitoring of an HDD
- The operation management software cannot collect information in the OS.
- When a REMCS contract is entered, software errors are not reported.
- HDD cannot be hot-maintained. The system in the BB must be stopped during maintenance.
- PRIMECLUSTER linkage cannot be used.

Note: When using this function under Windows, do not stop the Windows print spooler service.

For the management of the hardware configuration, Windows Management Instrumentation (WMI) is used in the processing for collecting information from the operating system. However, when the print spooler service is stopped, an error is reported to WMI and the information is not collected correctly.

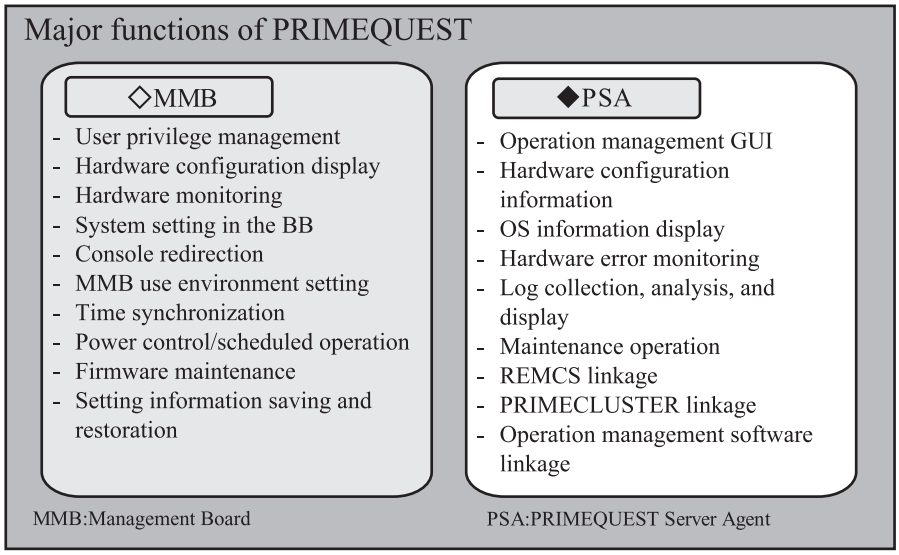


Figure 4.16 PSA functions

4.3.1 Operation management GUI

This function is a Web-UI function that can be used to manage the OS operation.

By linking the PSA installed in the OS to MMB firmware, the system in the BB can be displayed and operated via the MMB Web-UI without a Web server function at the BMM system end.

As shown in Figure 4.17, the MMB firmware has the Common Gateway Interface (CGI-WebGateCGI) which is used for communication with the Web server function. The PSA in the OS has the operation management GUI functions comprising the WebGate and HTML template group.

When receiving a request from a user, the WebGateCGI communicates with the WebGate by TCP/IP and delivers a corresponding HTML. The WebGate acquires information from the data source (configuration information, etc.) in accordance with the request and embeds the information in an HTML template.

In this manner, the Web-UI function for managing the OS information operation is provided.

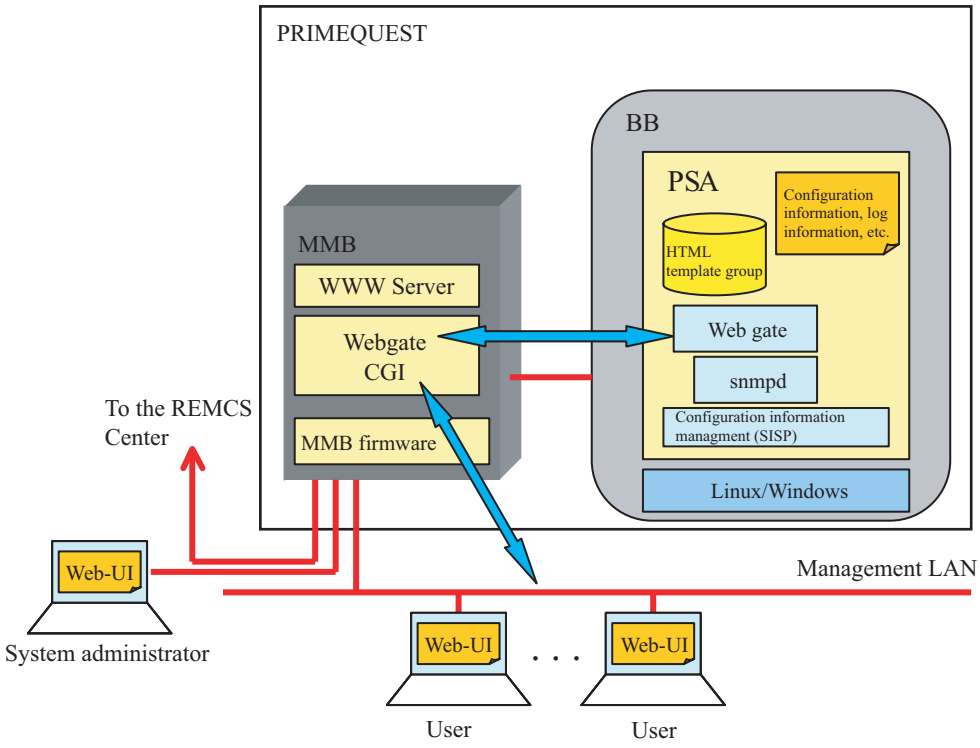


Figure 4.17 Web-UI function

The following functions, which enable script operation and operation in general by using the command line from an OS console, are provided by the Command Line Interface (CLI). For details, see the *PRIMEQUEST 510A Reference Manual: Basic Operation/GUI/Commands* (C122-E096EN).

- SAF_TE operation command (used by a Fujitsu SE when hot swapping disks)
- PSA start/stop command
- PSA survey data collection command
- Filter definition update command
- Local partition number acquisition command
- Serial number acquisition command
- SNMP security setting command
- Firmware information acquisition command

4.3.2 Hardware configuration management

This function displays the hardware resources.

This function displays the following configuration information:

- BB configuration
- CPU configuration (maximum number of CPUs which can be mounted, CPU mounting locations, identification information such as the CPU type)
- Memory configuration (mounting locations, detailed information such as the memory type)
- PCI configuration (PCI card mounting, PCI device mounting, detailed information such as the PCI device type, error status information)
- SCSI/FC connection unit configuration (HDD, tape drive, etc.)
- Network configuration (network interfaces, error status information)

4.3.3 OS information display

This function displays information on the OS installed.

This function displays the following information:

- OS information (OS type, OS version, and package installation information)
- Storage configuration information (device and capacity information)
- Network configuration information (interface, connection state, speed, routing information)
- OS status (operating hours and login count)
- Process list

4.3.4 Hardware error monitoring

The PSA monitors for errors reported by PCI cards and the drivers of SAS devices and other devices, the power supply of expansion file units, and FAN errors. The PSA also periodically monitors for predictive signs of failure detected by the S.M.A.R.T. (Self-Monitoring Analysis and Reporting Technology) function of hard disks. When the PSA detects an error, it performs error analysis to identify the faulty unit, records the results of the analysis as logging information, and reports the error to the MMB and the upper-level management software.

The monitoring for errors of the expansion file unit is performed in five-minute intervals.

PSA records the counts of various errors detected in the CPU, memory, chip sets, and interchip buses by the OS and firmware.

4.3.5 Log collection, analysis, and display

This function collects, analyzes, and displays logs related to the hardware error monitoring function.

When various types of events and messages are posted from firmware, various types of drivers, and the OS, this function records them in log files and takes a predefined action (notification by e-mail, notification to the REMCS, and log output). The recorded logs can be filtered based on conditions such as the time period messages are posted and the target message type, so as to display an appropriate number of log records on the screen.

The following table lists log files and information the PSA collects in the files.

Table 4.3 Log file information

Log file type	Description
Agent message log	Information related to the events (excluding event IDs 00000 to 09999 detected in the PSA) for which the PSA took action (recording to an OS log, SNMP trap, etc.) is stored in the agent log. This log can be displayed in a format supported by the GUI and can be downloaded in the form of a CSV-format file.
Error record log	Information on firmware-detected errors is recorded by the OS machine check handler. The PSA monitors the error information recorded by the OS machine check handler and stores it in an OS log file.
System event log	System event logs are logged by the MMB. The PSA periodically performs polling of the system event logs and stores them in OS log files. The stored SEL information can be displayed in a format supported by the GUI and can be downloaded in the form of a binary file.

4.3.6 Maintenance operation

This function supports hot swapping of hard disk.

The SAS controller of the hard disk used in the PRIMEQUEST series enables setting the disk power to on or off and the disk to be inserted or extracted with the SCSI Accessed Fault-Tolerant Enclosures (SAF-TE) function. The PSA provides a function by an SAF-TE operator command for safe maintenance performed by the SAF-TE function when the disk is swapped at detection of a hardware failure or at disk expansion.

Remarks: The SAF-TE operation command is supported only in Linux.

4.3.7 REMCS linkage

This function works with the MMB to report resource information and errors to the REMCS Center (remote maintenance center).

REMCS Agent reports PRIMEQUEST system errors and log information to the REMCS Center via the Internet or a P-to-P connection.

PRIMEQUEST REMCS Agent consists of the MMB firmware and the PSA and SIRMS installed in the OS. As shown in Figure 4.18, the MMB firmware monitors for errors in the entire system and, when it detects an error, reports the error to the REMCS Center. The PSA reports hardware error information and hardware configuration information detected by the OS to the REMCS Center via the MMB firmware. The PSA also reports software configuration information and software error information detected by the SIRMS to the REMCS Center via the MMB firmware.

For information on REMCS, see Chapter 7, "REMCS" in the *PRIMEQUEST 500A/ 500/400 Series Reference Manual: Tools/Operation Information* (C122-E074EN).

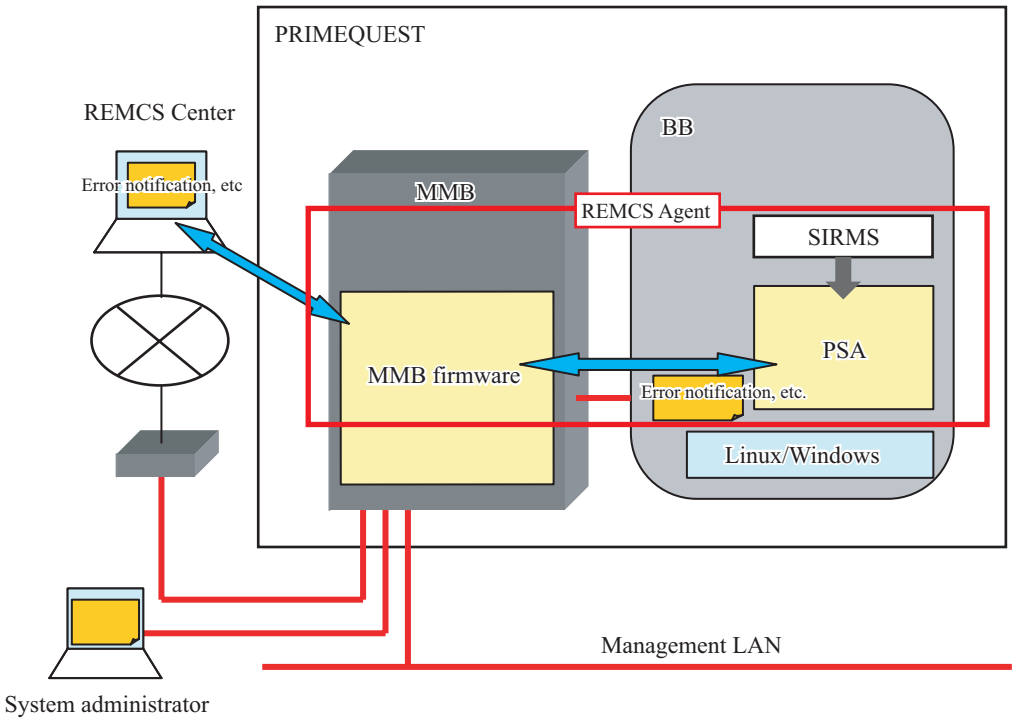


Figure 4.18 REMCS linkage

4.3.8 PRIMECLUSTER linkage

The function allows for a clustering configuration that works with PRIMECLUSTER machines. A cluster consisting of multiple nodes (PRIMEQUEST) can compose a redundant configuration that has active and standby systems.

If the active system becomes unusable such as because of a fault, the operation on the active system can be taken over by the standby system.

The PRIMECLUSTER linkage function is installed in MMB firmware and PSA to support the following functions:

- System status monitor/display .....Monitors and displays the status of the specified node.
- System status notification .....Notifies the associated nodes, which together with the local node form the cluster, of a status change of the local node.
- Event reception from another system.....Receives a status change of a remote node.

- Instruction to another system..... Issues an instruction to the specified remote node.

As shown in Figure 4.19, control information is taken over from the active system to the standby system to continue operation.

Remarks: PRIMECLUSTER linkage is supported only in Linux.

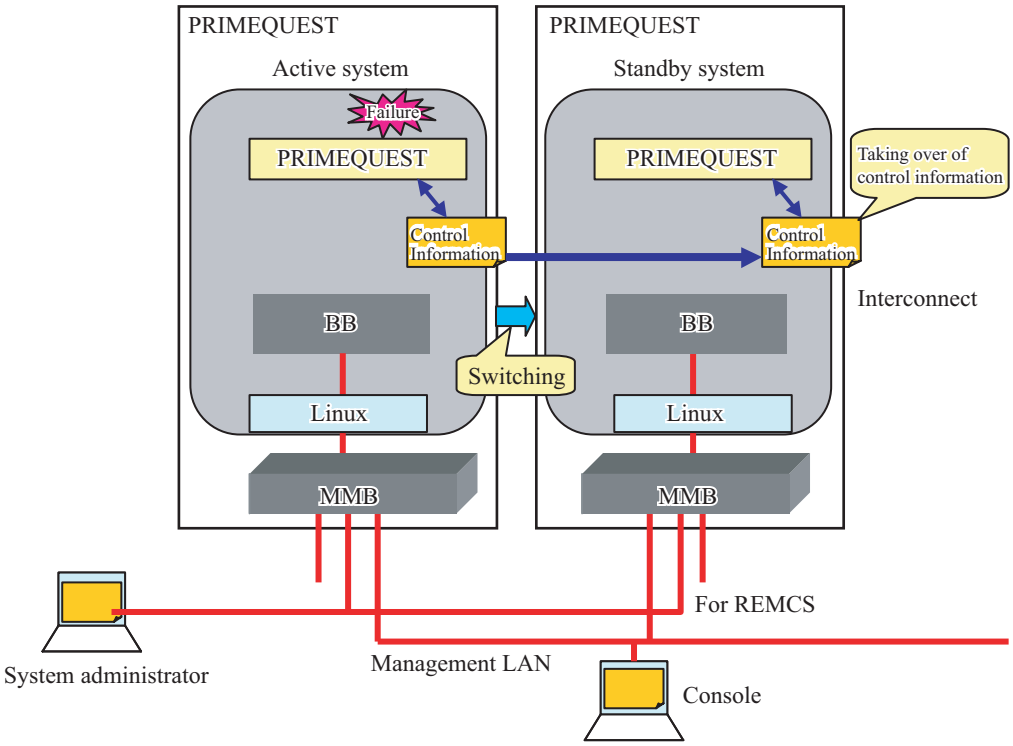


Figure 4.19 PRIMECLUSTER linkage

4.3.9 Operation management software linkage

This function enables a linkage with operation management software.

Figure 4.20 shows linkage with operation management software.

The function enables the PSA to work with operation management software such as Systemwalker. It uses SNMP (Simple Network Management Protocol) for the linkage.

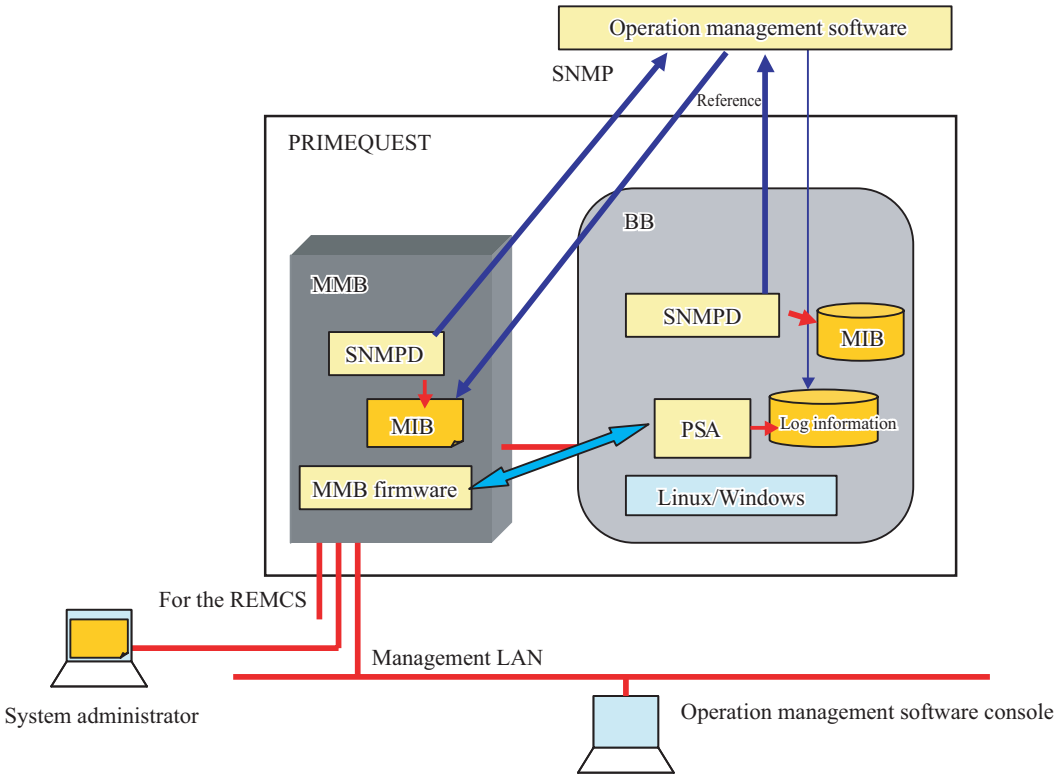


Figure 4.20 Operation management software linkage

4.3.10 Control targets of MMB user interface

The following is an overview of the MMB management and PSA management targets.

MMB management targets

- Hardware mounting information and status in cabinet
  (BB unit, PSU, MMB, etc.)
- System information display and settings (including cabinet setting, MMB, etc.)
- BB system configuration management and settings
- Maintenance operation (hot swapping of PCI cards, collection and display of logs)

PSA management targets

- Information management and operation of the system in the BB
  (PCI card, connection I/O, OS information display, and OS resource performance)
- Maintenance operation (hot swapping of disks, collection and display of logs)

Available functions

The following operations are available from GUI of PSA.

Table 4.4 Operations available from GUI of PSA

Available operation	Overview
Displaying configuration information	BB unit configuration
	CPU configuration (maximum number of CPUs which can be mounted, CPU mounting location, identification information such as the CPU type)
	Memory configuration (mounting location, detailed information such as the memory type)
	PCI configuration (PCI card mounting, PCI device mounting, detailed information such as the PCI device type, error status information)
	SCSI/FC connection unit configuration (HDD, tape drive, etc)
	Network configuration (network interfaces and error status information)
Displaying and manipulating OS information	OS information (OS type, OS version, and package installation information)
	Storage configuration information (device and capacity)
	Network configuration information (interface, connection state, speed, routine information)
	OS status (operating hours and login count)
	Process list
Maintenance operation	Displaying and saving log information (agent log)
Exporting	Exporting the current configuration and status

4.3.11 Information managed by PSA

Table 4.5 Information managed by PSA

Information type	Description	
Hardware information	Model information	• CPU information: mounting information, status, type, version, and frequency
		• Memory information: mounting information, status and type (size)
		• BB unit information: mounting information
		• PCI card information: mounting information, adaptor name, and detailed information
		• Connection I/O information: mounting information, type, and detailed information
System information	• Operating system: OS type and version number (revision number)	
	• Disk-related information: file system configuration	
	• Network-related information: Interface Name, Network Type, MAC Address, Interface Speed Current Status (up/down), Link Status (up/down), and PacketSize	
	• Other I/O information	

Note: MMB Web-UI supports the browsers listed below. If another browser is used, the Web-UI window may not be displayed correctly.

- Microsoft® IE (Internet Explorer) v5.5 (SP2) or later
- Netscape v7.02 or later

CHAPTER 5 Redundant Configuration

Redundant configuration refers to a configuration where even if some components making up the hardware configuration fail, the remaining components maintain adequate throughput.

PRIMEQUEST provides redundant configuration for the hardware configuration, such as I/O devices, transmission lines, and power units, to achieve high reliability and high availability.

This chapter describes the following types of redundant configuration adopted by PRIMEQUEST to improve reliability and availability.

For details on the PRIMEQUEST System Disk Mirror for Windows (PSDM), see the *PRIMEQUEST System Disk Mirror for Windows User's Guide* (B1FN-5771).

5.1 Redundancy of Components

Figure 5.1 shows components for which redundant configuration can be adopted.

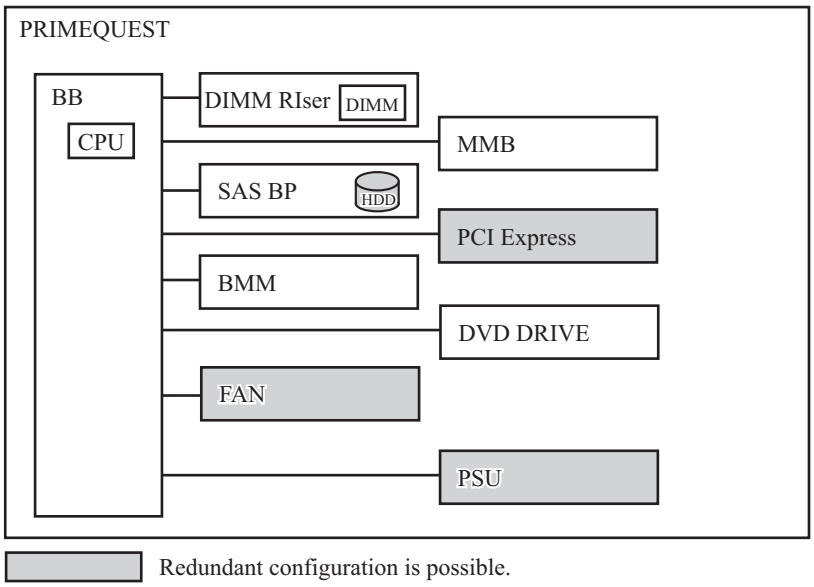


Figure 5.1 General redundant configuration

This part describes redundant configuration of the following components.

- Power Supply Unit (PSU)

  The PSU converts AC input to 48 VDC. This unit supports redundant configurations.

- FAN Unit (FAN)

  A redundant configuration can be adopted for individual fans as a basic configuration.

5.2 Redundancy of an HDD (Use of PRIMECLUSTER GDS)

5.2.1 Redundancy provided to system disk drives

PRIMEQUEST supports software mirroring of internal HDDs and system disks as an option. To support this function, optional software is required. Linux supports mirroring using PRIMECLUSTER Global Disk Services (GDS), which is provided as an optional software product. For details, see the PRIMECLUSTER GDS manual. Windows supports mirroring using the PRIMEQUEST System Disk Mirror for Windows (PSDM).

For details on the PRIMEQUEST System Disk Mirror for Windows (PSDM), see the *PRIMEQUEST System Disk Mirror for Windows User's Guide* (B1FN-5771).

Remarks: Mirroring by PRIMECLUSTER GDS does not cover the dump areas (diskdump, kdump).

HDD redundancy configuration

Figure 5.2 shows the minimum configuration for HDD redundancy. In this configuration, a redundant configuration of HDDs is implemented by connecting eight HDDs to one SAS. Redundancy, however, is not applied to SAS controllers, and other I/O devices other than HDDs.

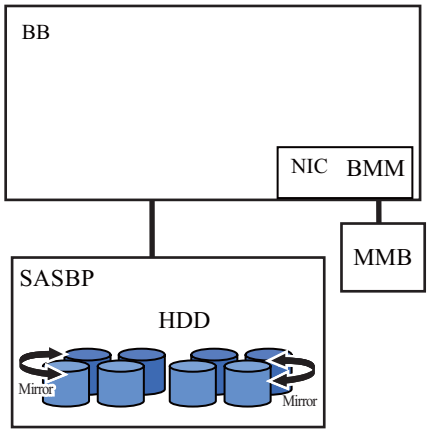


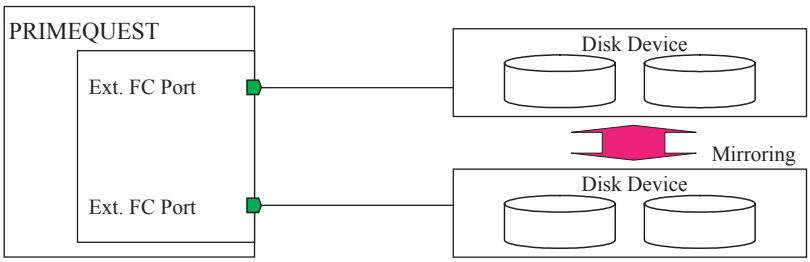
Figure 5.2 Redundancy configuration

5.2.2 Redundancy of external storage

Mirroring between cabinets

PRIMECLUSTER GDS supports mirroring. This, however, does not include the case where a disk is used as a system disk.

Figure 5.3 shows an example where mirroring between cabinets is enabled using external ports in a BB. In this example, external ports of the same BB are used. To improve system reliability, use external ports of different BBs.



Note: Use together with the multipath connection (grmpd).

Figure 5.3 Disk device mirroring

Path redundancy

To connect PRIMEQUEST series machines to ETERNUS3000, ETERNUS6000, or GR series machines over multiple paths (multipaths), use the ETERNUS multipath driver. If one of the paths in use becomes unavailable, operation is switched to another path working normally to continue operation. When this happens, an application can continue operation, since access is not interrupted. To improve access performance, use all usable paths to control the load balance (distribution of load). For details, see the manual for the ETERNUS multipath driver.

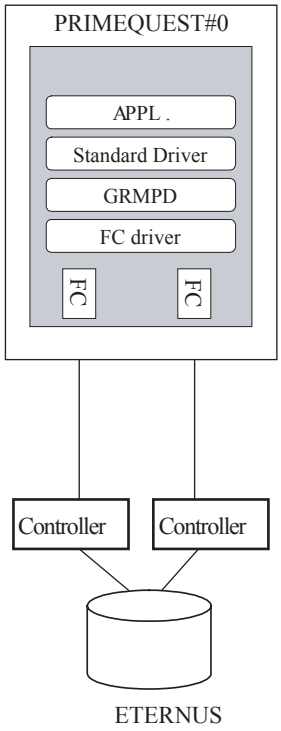


Figure 5.4 ETERNUS multipath connection

5.3 Redundancy of the Business LAN (Use of PRIMECLUSTER GLS)

This section describes the implementation of redundancy for a business LAN as related to the construction of a business system.

To implement redundancy for of a network and the network interface, at least two network interfaces and optional software to switch the interfaces are needed.

Linux supports PRIMECLUSTER Global Link Services (GLS) as optional software. Windows supports duplication using a teaming driver. This section shows an example of network redundancy that can be implemented by combining PRIMEQUEST and PRIMECLUSTER GLS. For details, see the manual for PRIMECLUSTER GLS.

5.3.1 Duplication of a transmission line between servers (high-speed switch method)

The high-speed switch method controls transmission lines by using the PRIMECLUSTER GLS-specific method. The method uses multiplexed transmission lines at the same time. When a failure occurs, a failed line is isolated and the system switches to degraded operation. Since PRIMECLUSTER GLS itself controls multiplexed transmission lines, it is possible to detect a failure at an early stage, but destination units that can be communicated with are restricted to PRIMEQUEST, PRIMERGY, PRIMEPOWER, GP7000F, FUJITSU S series, and GP-S units. When this method is used, the unit cannot communicate with a host on a different network across a router.

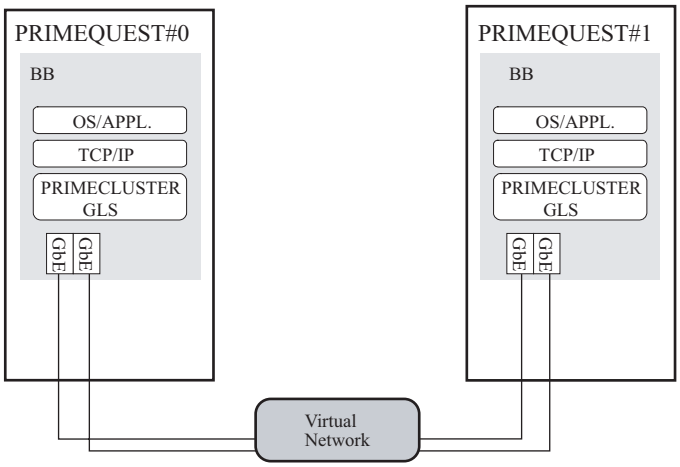


Figure 5.5 Example of implementing redundancy using the high-speed switching method

With this method, as shown in Figure 5.6, each of Network Interface Cards (NICs) is connected to a different network, and these NICs are activated and used at the same time. Transmission packets are sent to appropriate paths according to the operational status (whether operation is normal) of transmission paths.

A virtual interface is generated to logically handle multiple NICs as a single NIC. A TCP/IP application can communicate with a destination system without being aware that a network adopts redundant configuration physically by using the IP address (virtual IP address) set in the virtual interface as the own system IP address.

As the connection pattern, destination systems should be connected over the same network. It is impossible to communicate with a system on another network.

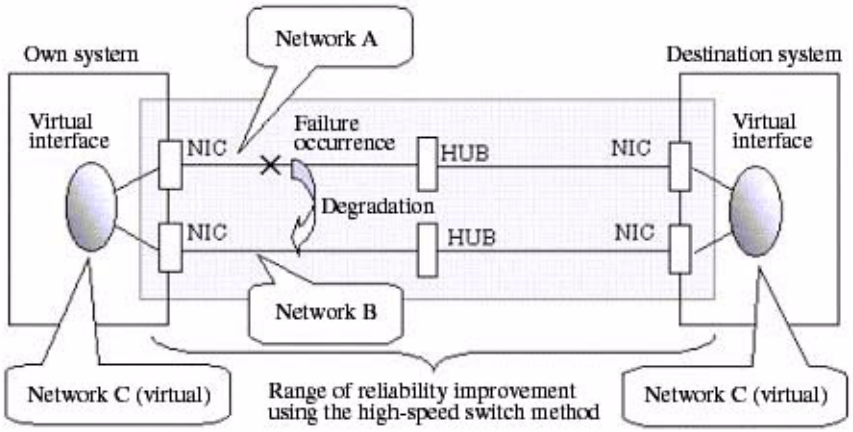


Figure 5.6 High-speed switching method

5.3.2 Duplication between a server and a hub and between switches on the same network (NIC switch method)

In the NIC switch method, as shown in Figure 5.7, duplicated Network Interface Card (NICs) are connected on the same network, and switch of transmission lines is controlled by exclusive use. With this method, communication destinations are not restricted. It is possible to communicate with hosts on the same network and on different networks. Since the range of duplication is restricted to the directly connected switches and hubs, to duplicate the entire communication path, it is necessary to duplicate routers and other network devices located halfway as well as the transmission line of the communication destination.

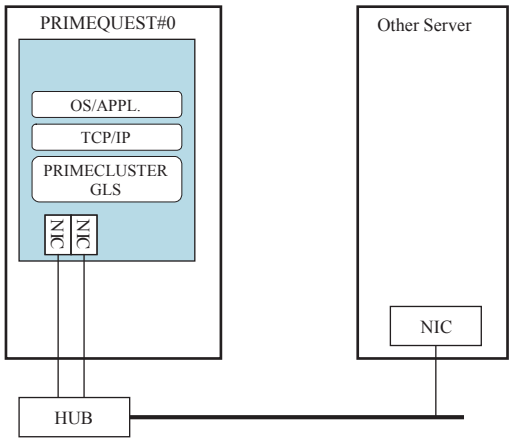


Figure 5.7 Example of implementing redundancy using the NIC switch method

With this method, as shown in Figure 5.8, duplicated NICs are connected on the same network, and switching of transmission lines is controlled by exclusive use (during normal operation, put one NIC in the "up" status for communication). A TCP/IP application can communicate with a destination system without having to take into consideration switching of NICs, by using the IP address set in the physical interface in the "up" status as the own system IP address.

As the connection pattern, duplicated NICs should be connected on the same network. It is possible to connect to a communication destination system either on the same network or on a different network via a router.

If each of the network devices in a multi-vendor environment (hub, router, and other devices) has a duplication function, the reliability of the whole system can be improved by combining devices. In this case, the range of duplication is defined separately for each vendor.

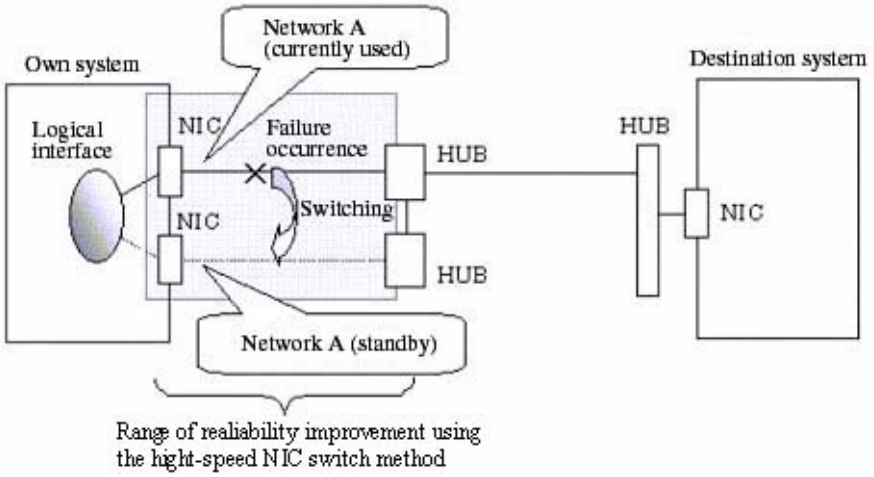


Figure 5.8 NIC switching method

5.4 Degradation Function

If a hardware abnormality is detected in the hardware power-on diagnosis or during system operation, the degradation function stops the faulty hardware and has the system inherit the operation with the remaining hardware.

Degradation reservation is the act of notifying a MMB about hardware that caused more correctable errors than the threshold in a given period, and that is likely to fail. The MMB automatically isolates such hardware at the next reboot.

Whether to use degradation varies depending on the location where the failure occurred, the nature of the failure, and the redundant configuration of a component where a failure has occurred. The following discussion includes a case where the location of a failure can be identified and a case where it cannot be identified.

Case where the location of a failure can be identified

If the location of a failure can be identified, degradation is performed as appropriate for the failed location.

Case where the location of a failure cannot be identified

If the location of a failure cannot be identified, the system is stopped.

If, after the system is stopped, the user specifies reboot, reboot or the retry of reboot is performed. If the system cannot be recovered after the specified times of retries has been reached, the system is stopped.

Remarks: It is very rare that the location of a failure cannot be identified. In most cases, the location of a failure is identified and degradation is performed.

CHAPTER 6 Hot Swapping

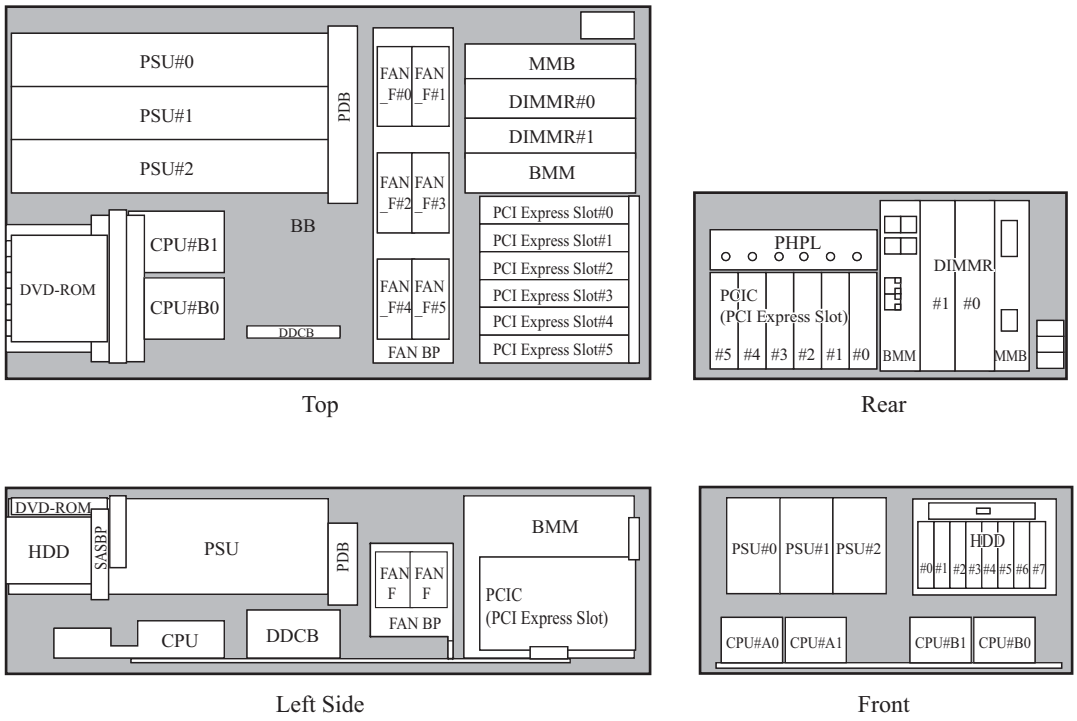
This chapter describes hot swapping of hardware components.

Hot swapping is implemented for the purpose of replacing a component that was operating but stopped due to a failure during system operation, with a normal component, without stopping the system operation. The component that had been swapped can be incorporated into the running system again. In this chapter, hot swapping also refers to deleting or adding a component without stopping the system, such as deliberately deleting a component from system operation, adding a component for expansion purposes, as well as replacing a component that was operating but stopped due to a failure.

In principle, hot swapping should be implemented by a Fujitsu certified service engineer.

6.1 Hot Swappable Components

Figure 6.1 shows the layout of components included in PRIMEQUEST series machines. Table 6.1 lists whether hot swapping and/or redundant configuration is available for components.



PSU : Power Supply Unit
CPU : Central Processing Unit
HDD : Hard Disk Drive
FAN : Fan
MMB : Management Board
DIMMR : Riser board for installing DIMM
BMM : BMC Module
BB : Base Board
PCI Express Slot : PCI Express slot
PDB : Power Distribution Board

Figure 6.1 Layout of included components

Table 6.1 Component hot swapping and/or redundant configuration availability

Item	Component	Hot swapping	Redundant configuration	Notes on redundant configuration
1	BB	N	N	
2	CPU	N	N	
3	Memory (DIMM)	N	N	
4	BMM	N	N	
5	HDD	Y (*1)	Y	Linux: Redundancy is ensured by PRIMECLUSTER GDS (software mirror) See Section 5.2, "Redundancy of an HDD (Use of PRIMECLUSTER GDS)." Windows: System disks only. Redundancy is ensured by a PSDN.
6	PCI Express Card	Y	Y	Windows Server 2003: Hot swapping of PCI cards is not supported. Windows Server 2008: See the PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide (C122-E087EN).
7	MMB	N	N	
8	DVD drive	N	N	
9	PSU	Y	Y	Hot swapping is available only if redundant configuration is adopted.
10	FAN Unit	Y	Y	

Y = Available, N = Unavailable

*1 This is supported only for Linux.

(1) BB

Hot swapping is not available because redundant configuration is not supported.

(2) CPUs

Hot swapping is not available.

If a CPU fails, the system isolates the failed CPU and restarts automatically.

Isolating the failed CPU may result in the unavailability of the normal CPUs on the BB. In this case, the entire system except the MMB stops even if it has a normal DIMM.

(3) Memory (DIMMs)

Hot swapping is not available.

If a DIMM fails, the failed DIMM is automatically isolated when the system restarts. The unit of isolation is two DIMMs including the failed DIMM (this is the same as the unit of expansion). The system automatically restarts after the DIMM failure.

Isolating the failed DIMM may result in the unavailability of the normal DIMMs on the BB. In this case, the entire system except the MMB stops even if it has a normal CPU.

(4) BMMs

Hot swapping is not available because redundant configuration is not supported.

(5) HDDs

- In a partition with Linux installed:
  A redundant configuration can be created with software for mirroring (such as PRIMECLUSTER GDS).
  If an HDD adopts the mirror configuration using mirroring software (such as PRIMECLUSTER GDS) or if deleting an HDD does not affect system operation, hot swapping is available. PSA can be used for hot swapping.
- In a partition with Windows installed:
  Redundancy is ensured for system disks only by a PSDN. Hot swapping is not supported.

(6) PCI Express cards

A redundant configuration using mirroring software such as a multipath driver can be used.

Also, in a partition with Linux installed as its OS, a PCI Express card can be hot-swapped with the PCI card inserted in a card cassette.

For a partition with Windows Server 2008 installed as its OS, see the *PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide* (C122-E087EN).

In a partition with Windows Server 2003 installed as its OS, hot swapping of PCI cards is not supported.

Remarks: For PCI Express card replacement using PCI Hot Plug, consult with a Fujitsu certified service engineer.

(7) MMBs

Redundant configuration is not adopted, and hot swapping is not available.

(8) DVD drive

Redundant configuration is not adopted, and hot swapping is not available.

(9) PSU

Redundant configuration is adopted, and hot swapping is available.

(10) FAN Unit

Redundant configuration is adopted, and hot swapping is available.

CHAPTER 7 Clustering

The chapter describes the overview of clustring.

To perform the cluster operation of the PRIMEQUEST, the clustering software PRIMECLUSTER is required. For details of PRIMECLUSTER, see the relevant PRIMECLUSTER manuals.

In Windows Server 2003, clustering is supported by Microsoft Cluster Service (MSCS), which is a standard function of the operating system. For details on Microsoft Cluster Service (MSCS) provided in Windows Server 2003, see the MSCS manuals.

7.1 The basics of clustering

Clustering is the grouping of multiple systems, each operating independently, to have them function as one system. Each of the systems forming a cluster is called a node. Even if one node fails, another node takes over the processing, thereby enabling job processing to continue.

Clustering can increase the availability of the entire system.

Availability is an index for the reliability of a system. A system in which few failures occur, which continues operation even if a partial error occurs, and which can be restored to its original state quickly following the occurrence of a system failure is called a system having high availability.

By implementing auto detection of hardware failures and errors in middleware and applications, and autonomous control such as taking over of job processing, job processing can be continued securely.

During a scheduled stop for periodical maintenance or to change the system configuration, the work can be done without affecting operating resources.

A cluster system where the PRIMEQUEST 510A acts as one node can be implemented.

For internode communication of management information in a PRIMEQUEST cluster system, the LAN is duplicated and used for error monitoring in each node. TCP/IP is used as the communication protocol to prevent data delay and loss when the amount of information on the communication path between an active node and a standby node increases temporarily.

Cluster operation modes are described below.

1:1 standby

A standby node is prepared for the active node. This configuration features a high redundancy where operation can continue even if the active node fails.

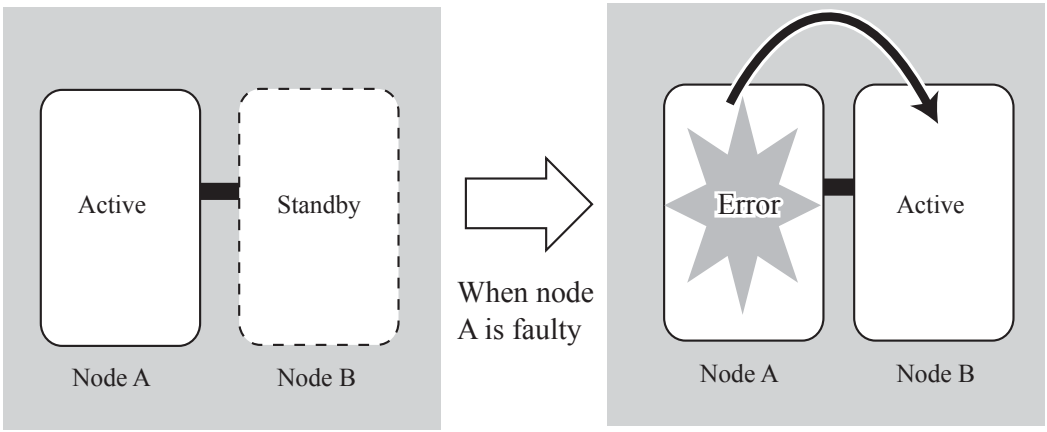


Figure 7.1 Example for 1:1 standby

N:1 standby

One standby node is prepared for multiple active nodes. Since only one standby node is required for multiple active nodes, a cluster system with good cost-performance can be constructed.

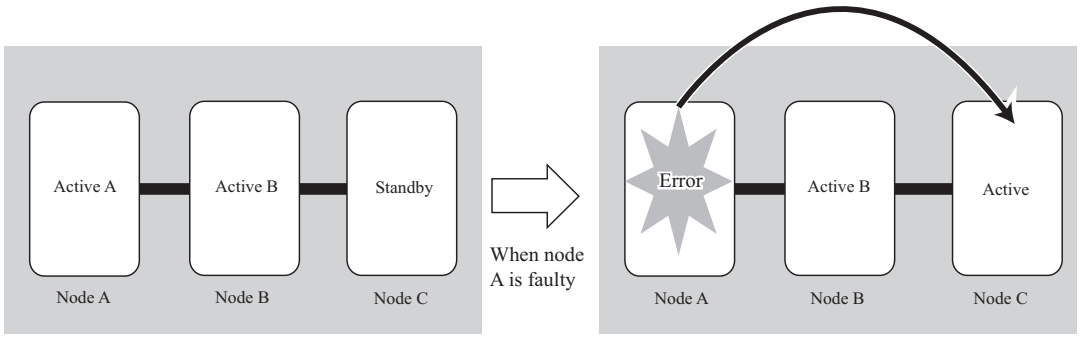


Figure 7.2 Example for N:1 standby

Cascade

Multiple standby nodes are prepared for one active node. When one node stops, the remaining nodes provide a redundant configuration, thereby maintaining availability during periodical maintenance tasks, etc.

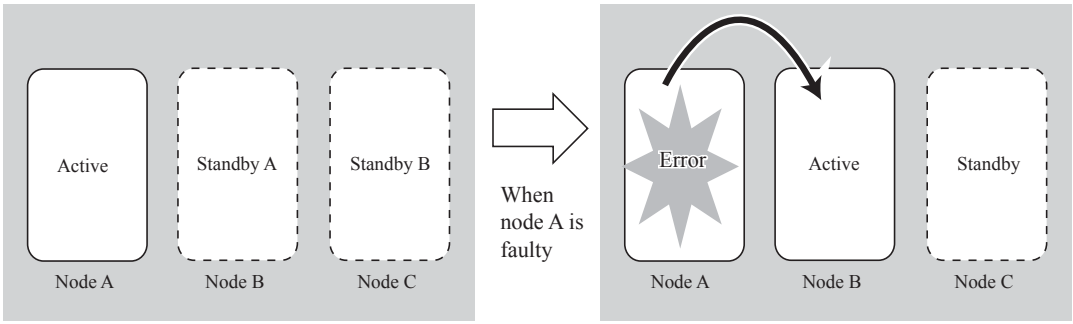


Figure 7.3 Example for a cascade

7.2 Cluster Functions Provided in Linux Environments

7.2.1 Cluster Configurations

For the PRIMEQUEST 510A cluster configuration between cabinets, the cabinet is treated as one node, which are connected to the business LAN to configure a cluster system.

Because the same operation is performed in another cabinet, the degree of redundancy is high, thereby increasing reliability.

The following example shows a connection of shared disk units by fibre channel. The system disk mirroring is applied using PRIMECLUSTER Global Disk Services (GDS).

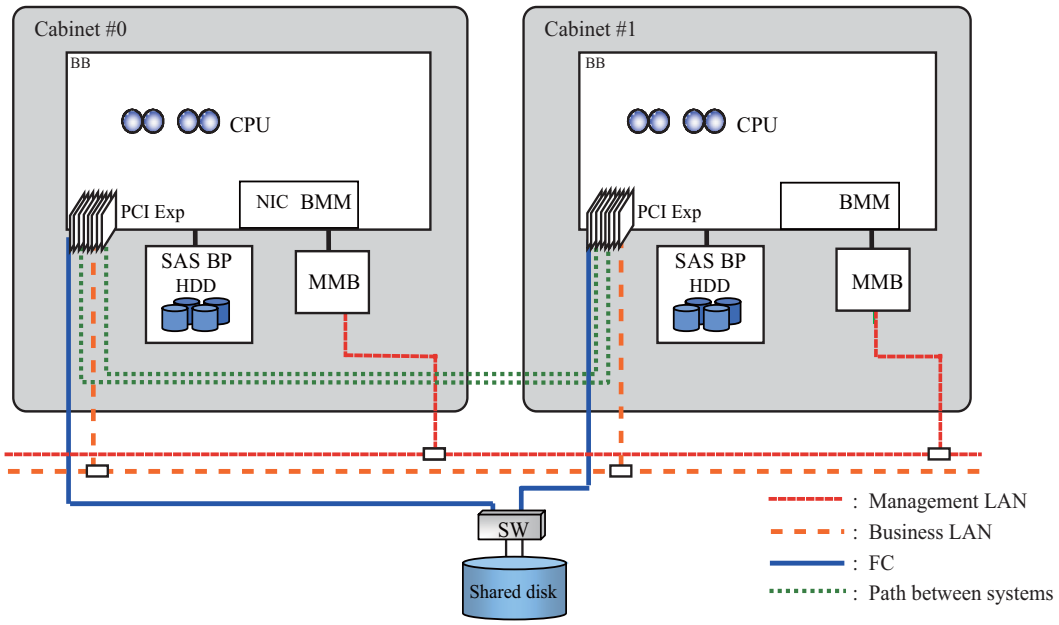


Figure 7.4 Inter-cabinet cluster system configuration

7.2.2 PRIMECLUSTER

PRIMECLUSTER is a software product for constructing a cluster system.

PRIMECLUSTER virtualizes system-configuring elements such as servers, storage, networks, and middleware, and enables job processing to be continued autonomously by error detection, failover (job takeover), and degradation functions to implement high reliability of the entire system.

Even if a failure occurs in the cabinet, high availability can be implemented by having a cabinet take over job processing.

Machines of the PRIMEQUEST series perform communication between PRIMECLUSTER and the MMB to provide the following functions:

- System state monitoring
  This function displays the state of a specified node and reports it to the MMB.
- System state notification
  This function reports changes in the operation state of the local node to another node.
  1. OS-detected events such as OS running and shutdown are reported to the MMB.
  2. When a change in the system state is detected, the MMB sends a trap to an associated node registered in advance.
- Event reception from the other system (in the pair active server ↔ standby server)
  This function receives information about changes in the state of the other node.
  This function receives information from the other MMB and filters duplicated trap data.
- Instruction to the other system (in the pair active server ↔ standby server)
  This function sends a Panic instruction and a Reset instruction to a specified other node.
  1. A control instruction from PRIMECLUSTER is reported to the MMB of the other system.
  2. The MMB controls the system through a private LAN.

7.3 Cluster Functions Provided in Windows Environments

Windows Server 2003 supports cluster functions as standard features of the operating system.

The cluster functions are services that provide redundant configurations in the system to increase business availability. If a software or hardware error occurs, such services continue business operations.

Installation of the cluster functions on the PRIMEQUEST machines can increase the availability of the entire system, including software.

This section explains Microsoft Cluster Service (MSCS), which is provided by Windows Server 2003.

7.3.1 Cluster configuration

Each server in a cluster is called a node. A cluster system consisting of up to eight nodes can be configured with MSCS.

Using Windows together with PRIMEQUEST, the most basic cluster system is a two-node inter-cabinet cluster using a shared disk.

The inter-cabinet cluster is configured with multiple cabinet systems, each of which operates in one cabinet. This offers the following benefits:

- Redundancy in the event of the occurrence of a cabinet failure
- Improvement of maintainability with separate cabinets
- Improvement of the maximum processing capability in one node

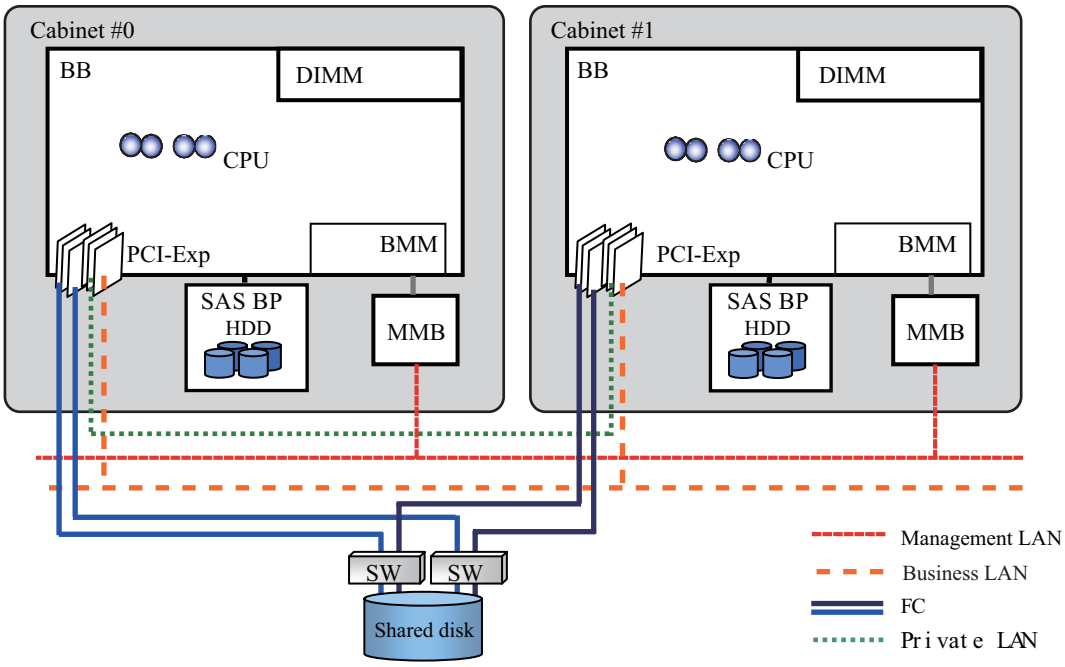


Figure 7.5 Basic inter-cabinet cluster configuration

Note: For a configuration other than the two-node cluster using a shared disk, contact your Fujitsu certified service engineer.

- Shared disk

  The data used by the business application software is stored on a shared disk unit so that it can be accessed and used from any node. However, note that the data cannot be accessed simultaneously from multiple nodes.

- Network configuration
  The nodes that belong to a cluster are connected via a dedicated LAN to ensure that they can monitor errors encountered by any of the other nodes and that data consistency can be maintained among the nodes. This network is called a private network (Private LAN). In contrast, the LAN used for business operations is called a public network (Business LAN).
- Business application software
  Business application software must be installed on each node used for business operations and on each standby node.

7.3.2 Virtual server

A virtual IP address instead of the IP address of a node is used as the IP address for access to business applications, because it is not changed by business application operations on any node. The server whose virtual network name is associated with the virtual IP address is called a virtual server. The virtual server can be referenced from clients as a Windows network computer.

As a rule, clients access the application software via a virtual server. During switching from a node on which business application software is running, the corresponding session is disconnected once. However, the session can be resumed by reestablishing the connection to the same virtual server after the business application software restarts.

Because of this mechanism, clients in a cluster system need not consider the particular node on which business application software is running.

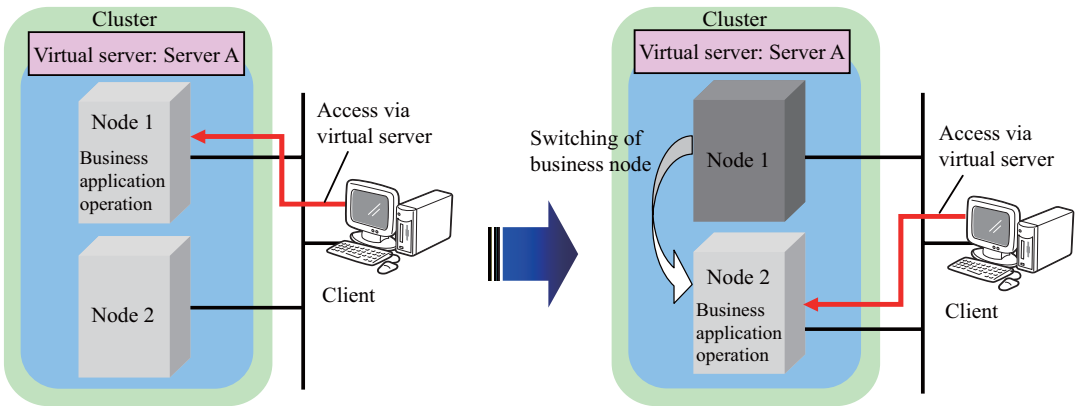


Figure 7.6 Virtual server

7.3.3 Scope of remedies in a Windows cluster

Remedies can be found for hardware (e.g., CPUs, motherboards, and LAN cards) failures and software failures (e.g., application errors). Note that a server cluster is not a fault tolerant (FT) system. If a business application failover occurs, operation stops for a short time.

7.3.4 Notes on construction

The dedicated startup service can be used for Active Directory (AD) and cluster system construction. For details, contact your Fujitsu certified service engineer.

CHAPTER 8 Backup and Restoration

This chapter provides instructions on how to back up and restore the system volume in which Linux (Red Hat) or Windows is installed. For information on SUSE backup and restoration information, refer to the relevant SUSE documentation.

8.1 Importance of Backing Up Data

PRIMEQUEST uses components and hardware with high reliability and duplicates many units to maintain high reliability. To be prepared for failure occurrences, however, be sure to back up data periodically.

If a system is damaged due to some problem, or if data on a server is deleted by an operator error and the like, the data on the PRIMEQUEST system needs to have been backed up to recover the data in the server to its previous status.

As long as data is backed up, if data in a hard disk drive is damaged due to a hardware failure, an operator error, and the like, the user can restore the backup data and recover normal system operation. If the data is not backed up, it will be lost.

Be sure to back up data periodically to operate a system with confidence.

8.2 General Methods of Backup and Restoration

The following four methods of backing up and restoring a system are available:

- Using the standard utilities included in an OS
  Linux (Red Hat): dump/restore
  Windows: NTBackup
- Using cloning software
  Common to Linux (Red Hat) and Windows:SystemcastWizard Lite
- Using snapshot software
  Linux (Red Hat): PRIMECUSTER GDS Snapshot

- Using backup software
  Linux (Red Hat): ETERNUS SF ACM, VERITAS NetBackup,
  NetVault, NetWorker

Ask the sales representative in charge for details including the versions of the supported operating systems.

The methods above include a method that uses a backup server with a backup unit. Consult with the sales representative in charge, and prepare backup software and a backup server.

An overview of each backup and restoration method is provided below.

8.2.1 OS-standard utility (Linux (Red Hat): dump/restore, Windows: NTBackup)

There are two types of file backups and restoration. One is backup on a relatively small scale, such as backing up or restoring a specific file and directory, and the other is backup on a large scale, such as backing up the entire system volume.

This section explains the backup and restoration of the entire system volume by using the dump and restore commands, which are Linux standard utilities, and the backup and restoration of files by using NTBackup, which is a Windows standard utility.

Configuration example when using an OS-standard utility

- Local backup

  For Linux (Red Hat):

  Prepare an external unit to which backup data is to be saved for each partition. Enter the dump or restore command from the OS console of the partition (Partition#0 in Figure 8.1) to be backed up or restored. Either a console unit (a keyboard, a display such as a flat panel display, and a mouse) connected to the KVM switch or a serial console can be used as the OS console.

  To perform restoration, the system needs to be started in rescue mode by using the installation CD.

For Windows:

For the backup or restore of a data area of a Windows system, NTBackup is used. The backup can be performed either from the GUI screen when the console has been connected or with the backup command from the command prompt screen.

However, the restoration can be performed only from the GUI screen. Files cannot be restored with a backup command from the command prompt screen.

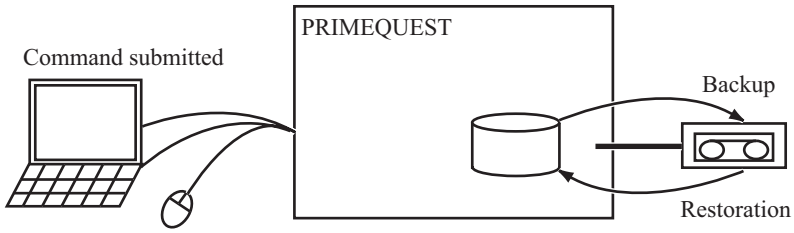


Figure 8.1 Local backup using a standard utility (when the KVM console is used)

- Backup to a remote host

  For Linux (Red Hat):

  Prepare a server with a backup unit, and back up or restore data using ssh.

  Using the LAN, back up or restore the backup/restoration target partition (PRIMEQUEST #0 in Figure 8.2) to a tape unit in a server (remote host) equipped with a backup unit.

  Perform the operation from the OS console of the partition to be backed up or restored (PRIMEQUEST #0). A serial console or KVM console can be used as the OS console.

  When this method is used, the network needs to have been activated. For the data transmission line, Fujitsu recommends using a LAN exclusively prepared for backup purposes. It is also possible to use the management LAN.

  Backup is performed by outputting data using the standard output as the output destination of the dump command, and copying (dd) the result to a tape unit of a remote host after logging in remotely using the ssh command.

  Restoration is performed by logging in remotely by using the ssh command, copying (dd) the tape unit data to standard output, and then expanding the contents of standard output by using the restore command. At this time, the system needs to be started in rescue mode by using the installation CD.

  A server outside the cabinet can also be used as the remote host.

For Windows:

If data in PRIMEQUEST #0 is backed up, it is necessary to use NTBackup on the remote host side (PRIMEQUEST #n) for the backup and restoration work.

First, perform remote-login to Partition #n from PRIMEQUEST #0 that is connected to the console by using remote desktop connection. Backup and restoration can be performed after assigning the Partition #0 as a network drive.

However, the server to be backed up and the remote host must be placed in the same network.

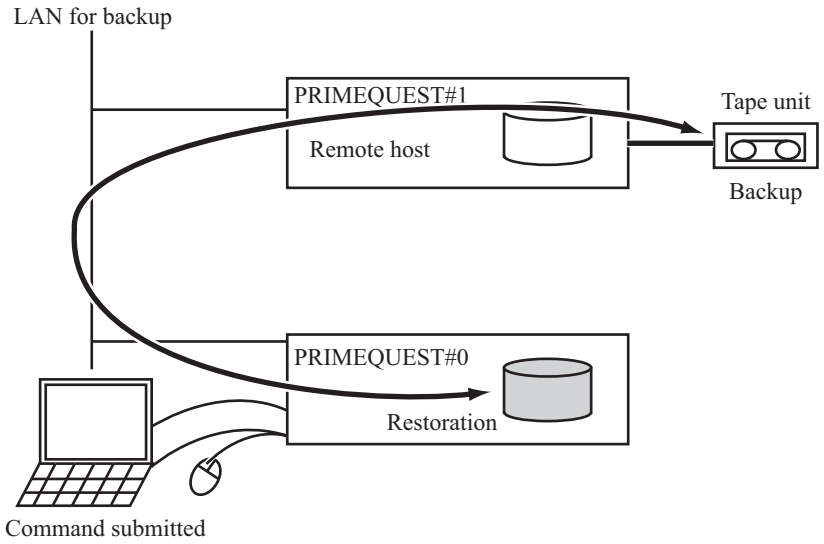


Figure 8.2 Backup to the remote host by using a standard utility (when using the KVM console)

8.2.2 Cloning software (SystemcastWizard Lite)

A cloning software program (SystemcastWizard Lite) is a tool for easily installing a system on multiple computers and performing maintenance in a short time.

SystemWizard Lite is a software program that performs OS installation, backs up hard disk data, and restores hard disk data via a network.

This software can be used for backup at the system construction or verification stage and also for recovery if a problem occurs. The PRIMEQUEST network activation function can be used to facilitate recovery of entire system volumes, even from a state in which the OS cannot start.

Configuration example when using SystemcastWizard Lite (common to Linux: Red Hat and Windows)

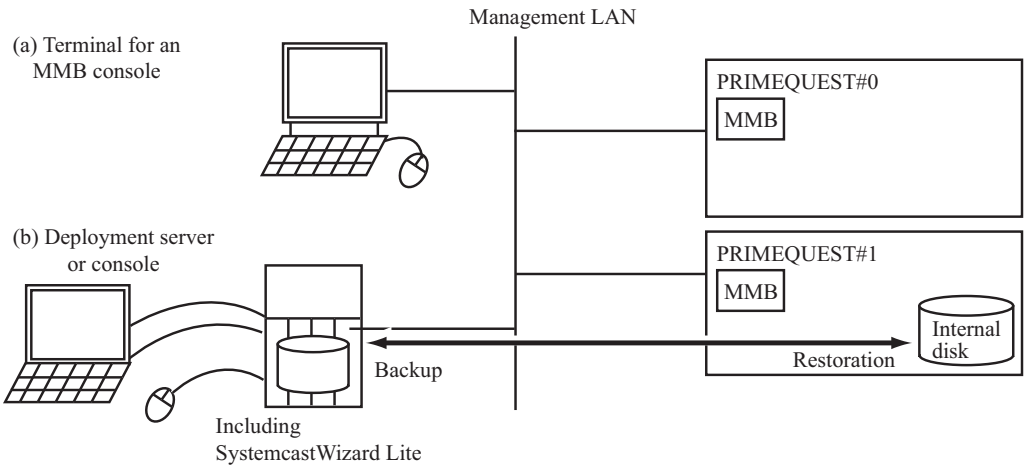


Figure 8.3 Backup using SystemcastWizard Lite

- Terminal for an MMB console
  A Web browser must be installed. It can be used on unit (b) deployment server console.
- Deployment server console
  This is a server or PC with Windows OS installed. SystemcastWizard Lite must be installed on the server, too. SystemcastWizard Lite is included in the supplied software. For details, see the *PRIMEQUEST SystemcastWizard Lite User's Guide* (C122-E010EN).

Note:

- Since deployment server console has an interface for direct exchanging information with an MMB, it needs to connect to the management LAN. Place deployment server console in the same segment as the PRIMEQUEST system. ((a) terminal for an MMB console can be used also when placed in another segment.)
- With a view toward reliability and performance, use a server machine as the deployment server for backup and restoration.

8.2.3 Snapshot software (PRIMECLUSTER GDS Snapshot)

By using PRIMECLUSTER GDS Snapshot, a snapshot (copy of data at a certain point in time) of a logical volume managed by PRIMECLUSTER GDS can be collected and stored in a logical backup volume.

By using the snapshot function, the amount of time that business is suspended for a backup or restore operation can be reduced.

Configuration example when PRIMECLUSTER GDS Snapshot is used (Linux: Red Hat)

- Backup of system volume (internal disk)
  If data on the system volume is destroyed, the backup destination disk can be used as a new system volume. It is also possible to restore data to the original system disk by starting the system from the backup destination disk.

Note: A backup copy can be output also to a disk on the ETERNUS system. However, since booting from a backup volume on the ETERNUS system is not supported, restoration cannot be performed from a backup volume on the ETERNUS system.

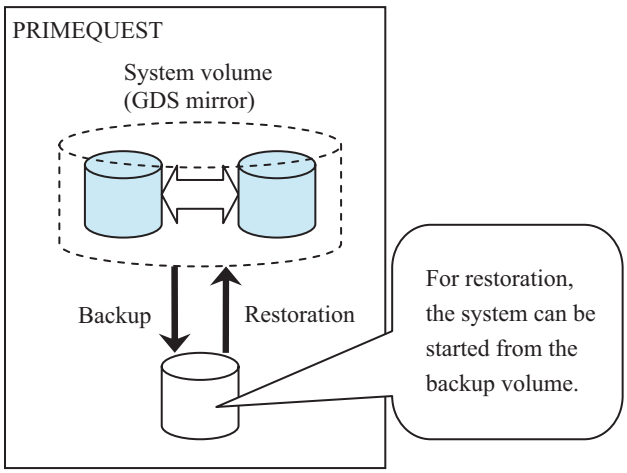


Figure 8.4 System volume backup using PRIMECLUSTER GDS Snapshot

- Backup of data area
  - Whether online backup is enabled or not depends on the data management software such as database management software or the file system. Generally, business operation needs to be stopped temporarily when a snapshot is created.
  - Backup from the backup destination disk to tape is enabled during business operation. For this purpose, however, backup software that supports backup from the GDS volume to tape is required separately.
  - The restoration destination volume can be used immediately after the start of restoration.

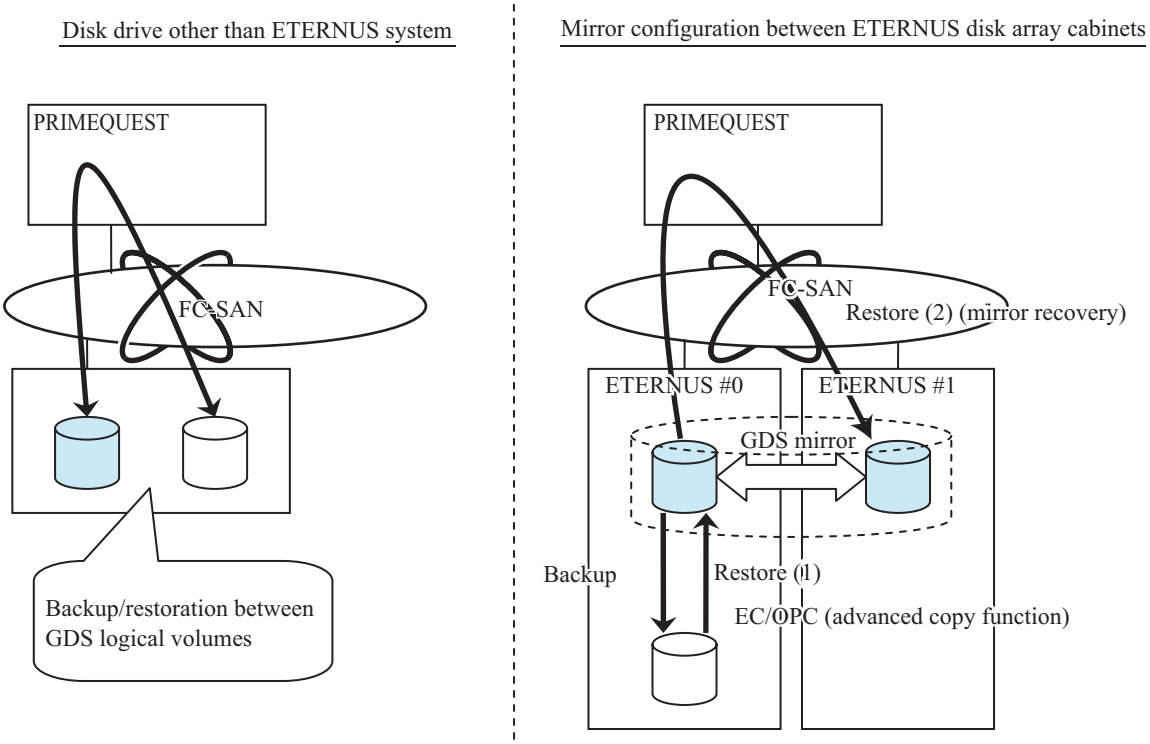


Figure 8.5 Data area backup and restoration using PRIMECLUSTER GDS Snapshot

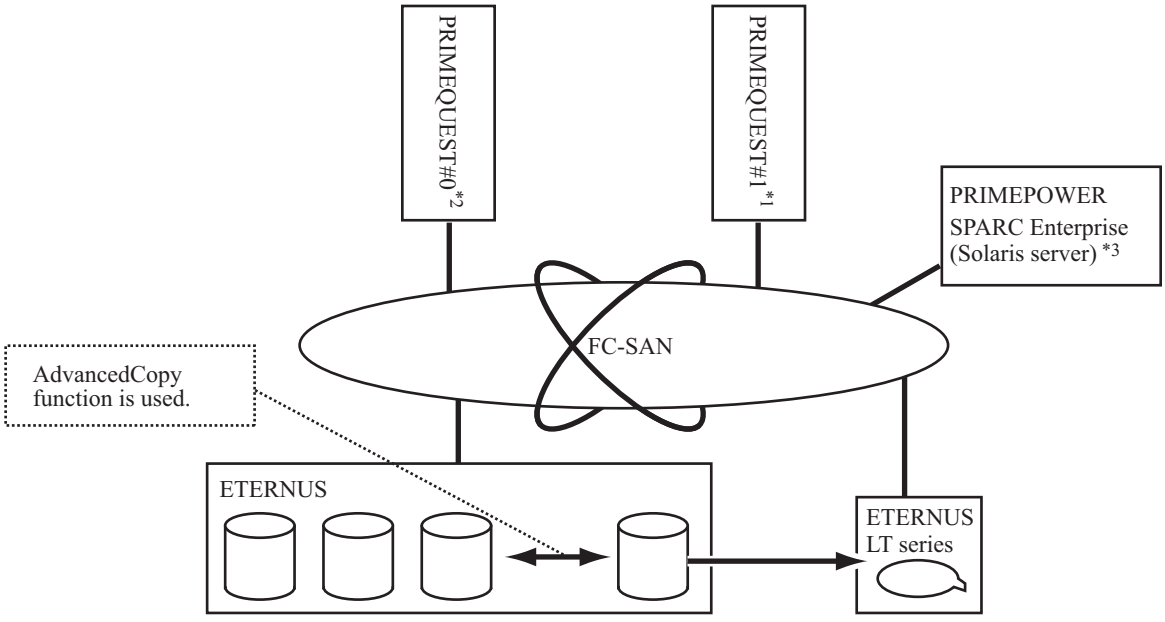
8.2.4 Backup software (ETERNUS SF ACM, VERITAS NetBackup, NetVault, NetWorker)

To design a schedule for backup in conformity with basic business operations, and to back up data in a database operation in cooperation with database software, use a backup-dedicated software program.

Backup software programs for the ETERNUS SF ACM, VERITAS NetBackup, NetVault, NetWorker, and others. The characteristics and configuration examples of each software programs are described below.

Configuration example when using ETERNUS SF AdvancedCopy Manager (Linux: Red Hat)

ETERNUS SF AdvancedCopy Manager (ACM), software for backup purposes, is used in combination with the ETERNUS8000/ETERNUS4000/ETERNUS2000/ ETERNUS6000/ETERNUS3000 disk array unit. Data of a disk array unit can be backed up or restored by using this software without passing through the business server or LAN.



(*1) ACM Manager and ACM Tape Server are mounted.
(*2) ACM-Agent is mounted.
(*3) ACM Tape Server is mounted.

Figure 8.6 Backup of data on the ETERNUS by using ACM

- ACM Manager
  Comprehensively manages the device information and policy of all the agents.

- ACM Agent
  Controls the AdvancedCopy function of the ETERNUS. ACM Agent is installed in the business server that is a backup restoration target.
- ACM Tape Server
  Controls the tape library and the backup to and restoration from tape media. ACM Tape Server is installed on the Solaris server and the tape library is connected to the server.

Configuration example when VERITAS NetBackup is used (Linux: Red Hat)

Note: Before using an Oracle product, confirm the operating system on which it runs.

- Normal backup of data

  Use the LAN or SAN, back up client data to a tape unit connected to a backup server. As a backup destination, the user can specify a disk area connected to the backup server.

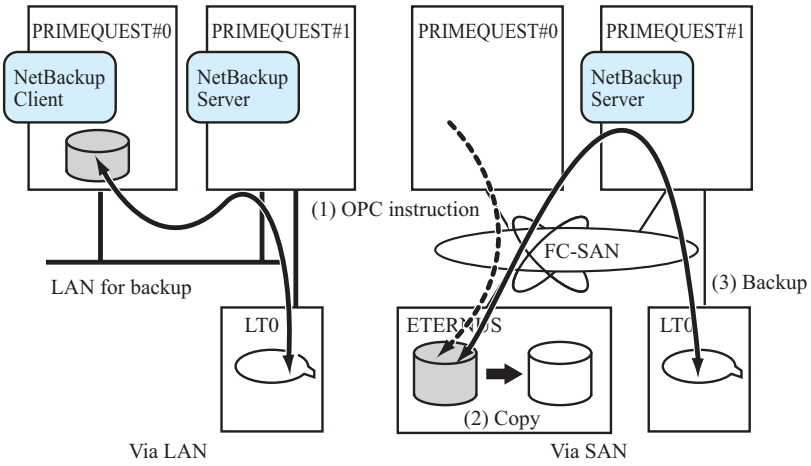


Figure 8.7 Data backup using VERITAS NetBackup

- Online backup of a DB (Oracle)
  For Oracle, Recovery Manager (RMAN) provided by Oracle becomes the interface to the DB. VERITAS NetBackup Oracle Agent is linked to RMAN, and data is passed from Oracle Agent to VERITAS NetBackup.

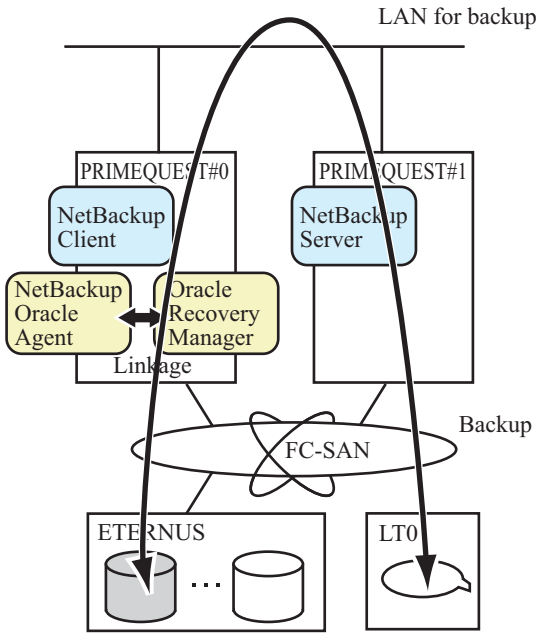


Figure 8.8 Online backup of a DB using VERITAS NetBackup

- Backup of a system volume
  This operation is not supported.

Configuration example when NetVault is used (Linux: Red Hat)

Note: Before using an Oracle product, confirm the operating system on which it runs.

- Normal backup of data
  The configuration is the same as the case where VERITAS NetBackup is used. The user can use the GUI on a NetVault server terminal or a NetVault remote management terminal via a LAN to back up or restore data. To perform backup of multiple terminals at the same time, a NetVault server disk, as a backup destination, should be a virtual tape library.
- Online backup of a DB (Oracle)
  The configuration is the same as the case where VERITAS NetBackup is used. To perform online backup of a database by NetVault GUI, use the NetVault application plug-in module (APM) to seamlessly integrate the online backup API provided by Oracle and NetVault.

- Backup of a system volume
  Not supported.

Configuration example when NetWorker is used (Linux: Red Hat)

- Normal backup of data

  The configuration is the same as when VERITAS NetBackup is used.
  Using the LAN or SAN, back up client data to a tape unit connected to a backup server. As a backup destination, the user can also specify a disk area connected to the backup server. Data backed up on a disk can be restored at high speed, enabling quick completion of staging, cloning, and other operations.
  If the user purchases an optional product (Dedicated Storage Node), the user can back up or restore data to tape without the intervention of a server in the SAN environment.

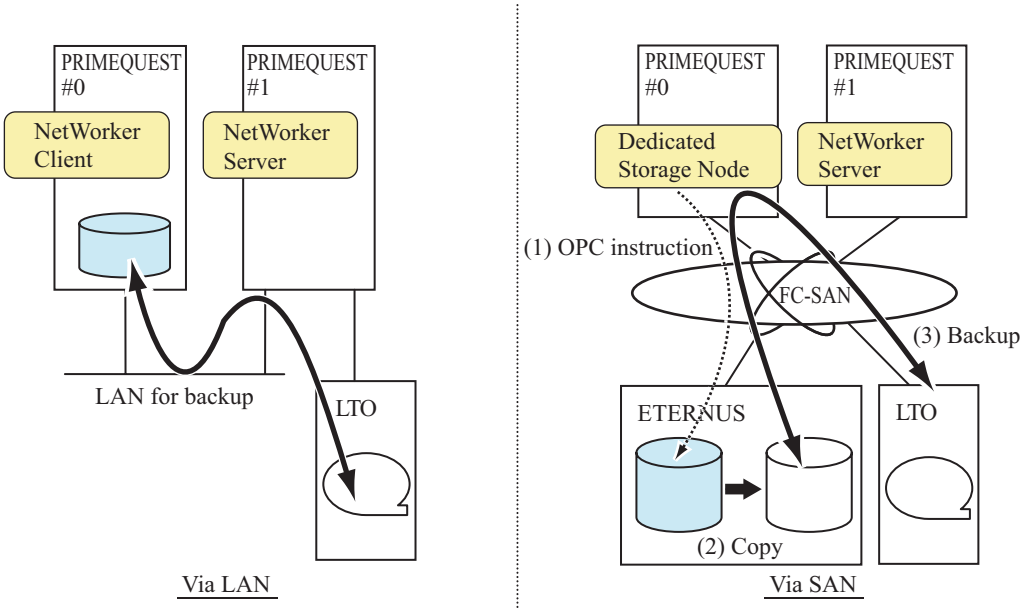
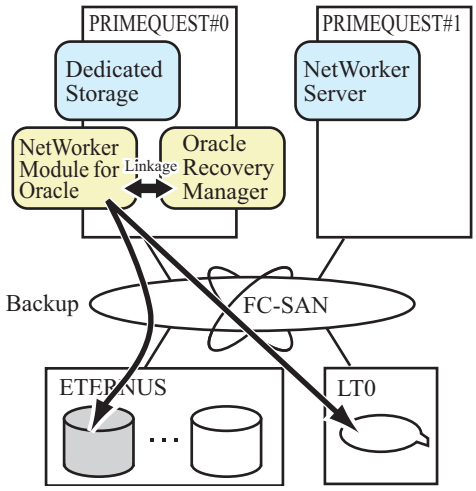


Figure 8.9 Data backup using NetWorker

- Online backup of a DB (Oracle)

  Oracle databases can be backed up with the dedicated storage node and NetWorker Module for Oracle to enable linkage between the online backup APIs and NetWorker Module for Oracle.



- Backup of system volume
  This operation is not supported.

8.3 Comparison of Backup/Restoration Methods

Table 8.1 lists the usage method and characteristics of various backup and restoration methods. The requirements for the backup methods vary with the operation. The methods appropriate to each requirement need to be combined to back up data.

Table 8.1 Comparison of backup methods

Method		Main usage	Characteristics
OS-standard	dump/restore (Linux: Red Hat) NTBackup (Windows)	Backup of a system volume (single)	• It is possible to back up and restore a system volume. • These are OS-standard functions. The user is normally accustomed to their operation.
Cloning software	Systemcast Wizard Lite (Linux: Red Hat/ Windows)	Backup of a system volume (batch)	• It is possible to back up and restore a system volume. • Automatic processing is possible in cooperation with a MMB via a network. • It is possible to perform concurrent parallel backup and restoration of multiple units.
Snapshot software	PRIMECLUSTER GDS Snapshot (Linux: Red Hat)	Backup of GDS logical volume (system volume and data area)	• It is possible to perform online backup and restoration of a system volume and data area. • For a data area on an ETERNUS disk array, it is possible to perform backup and restoration without applying a load to the PRIMEQUEST CPU and LAN. Remarks: Setting and operation for GDS mirror configuration are not necessary • No backup server needs to be prepared. • This software has high affinity with the PRIMECLUSTER and PRIMECLUSTER GDS.

Method		Main usage	Characteristics
Backup software	ETERNUS SF ACM (Linux: Red Hat)	General data stored in the ETERNUS disk array unit (backup of a system volume by booting SAN and backup of data areas)	• It is possible to perform backup without applying load to a LAN. • It is possible to perform backup during DB (Oracle/Symfoware) operation. • It is possible to back up and restore a system volume while booting from a SAN. • PRIMECLUSTER is supported.
	VERITAS NetBackup (Linux: Red Hat)	Backup of data areas (large scale DB)	• It is possible to perform backup or restoration during DB (Oracle) operation. • NetBackup has a track record for DB backup of large-scale systems. • It is possible to manage a mixed platform environment by using one server. • PRIMECLUSTER is supported.
	NetVault (Linux: Red Hat)	Backup of data areas	• It is possible to perform backup during DB (Oracle) operation. • Setup for a small-scale system is possible at low cost. • Operation with a Japanese GUI is possible. • PRIMECLUSTER is supported.
	NetWorker (Linux: Red Hat)	Backup of data areas	• It is possible to perform backup during DB (Oracle) operation. • It is possible to perform backup of a wide range of environments from a small-scale system to a large-scale system. • It is possible to perform concurrent parallel backup of multiple units. • It is possible to manage a mixed platform environment by using one server.

CHAPTER 9 Considerations for Maintenance

This chapter summarizes the system design considerations pertaining to the work of maintaining the system in which Linux (Red Hat) is installed. For information about the system in which SUSE is installed, refer to the relevant SUSE documentation.

9.1 Estimating the Dump-use Area

In a PRIMEQUEST system using Linux (Red Hat), the estimate of the dump-use area varies depending on the version of the operating system. This section explains the following estimation methods:

- RHEL-AS4 (IPF)
- RHEL5 (IPF)

9.1.1 RHEL-AS4 (IPF)

If using RHEL-AS4 (IPF) in PRIMEQUEST systems, the dump function provided by Linux distribution (diskdump) is available.

This section explains how to estimate the disk spaces necessary for the use of this dump function. The explanation is provided with respect to the following items in the order given:

- Area used to collect dump data
- Area used to save dump data

Area used to collect dump data

When a system becomes faulty, a disk partition (dump device) area exclusively used for dump collection is required.

A dump device is a device on which the dump function temporarily stores information on problem occurrences. Table 9.1, "Areas used to collect dump data" lists the partition sizes of the areas required for a dump device.

Table 9.1 Areas used to collect dump data

No.	Function name	Partition size	Remarks
1	diskdump	Amount of installed memory + 512 MB	

The capacity of the dump device needs to be expanded according to the amount of memory installed. Therefore, if the user plans to expand the amount of installed memory, estimate the size of the dump device based on the amount of installed memory after expansion.

When the amount of installed memory is large, the partial dump function or compression function can be used to reduce the disk space requirements for the dump function. The partial dump function can collect data from only the areas used for kernel control. Although the function can reduce the disk space requirements, it may take a little more time for troubleshooting upon the occurrence of a kernel error or system failure.

A dump device can be allocated either on an internal disk or on an external disk. Table 9.2, "Devices that can be used as a dump device" lists the devices that can be used as a dump device. Allocate a not duplicated dump device on a disk not managed by PRIMECLUSTER GDS.

Table 9.2 Devices that can be used as a dump device

No.	Type	Driver	Remarks
1	Internal disk	LSILogic Fusion MPT driver (SCSI)	Internal SAS disks are supported.
2	External disk	Emulex FC driver	
		LSI Logic Fusion MPT driver (SCSI)	SCSI disk connected via a SCSI card

As the capacity of internal disks is limited, the user may be unable to construct a system using only internal disks. In such a case, prepare a dump device on an external disk. Note that a dump device prepared on an external disk cannot be shared by multiple clusters.

Area used to save dump data

At the system restart time, the dump information collected by diskdump and sadump on a dump device is saved in /var/crash/ after automatically being converted to an appropriate dump file format so that it can be referenced by a dump tool. If a problem occurs, diskdump and sadump may create two dump files at the same time. Therefore, the file system must have enough free space for saving at least two dump files.

The size of this area that must be reserved in the file system is based on the number of saved dump files as follows:

Required area size per dump file = Amount of mounted memory + 1 GB

To confirm that the area required for the saved dump files has been reserved on the disk, execute the following command:

```
# df -h /var/crash
```

Note: When creating a dump file storage area, such as /var/crash/, as a dedicated disk partition, create the disk partition to have a size that is at least 10% larger than the above estimated value, in consideration of the file system management area.

9.1.2 RHEL5 (IPF)

If RHEL5 (IPF) is used in the PRIMEQUEST system, the dump function provided by the Linux distribution (kdump) is available.

This section explains how to estimate the amounts of disk space necessary for the use of these dump functions. The explanation is provided with respect to the following items in the order given:

- Area used to save dump data

Area used to save dump data

Dump files in a format enabling them to be referenced by a dump tool are created directly in the dump storage area by kdump. At the system restart time, the dump information collected by sadump on a dump device is saved after automatically being converted to an appropriate dump file format so that it can be referenced by a dump tool.

Unlike diskdump, kdump requires that a dedicated disk partition be created for saving its dumps. The dedicated disk partition must not be managed by PRIMECLUSTER GDS.

The size of the dedicated disk partition must be at least the following:

Total size of saved dump files × 1.1

If a problem occurs, kdump and sadump may create two dump files at the same time. Therefore, the file system must have enough free space for saving at least two dump files.

The size of this area that must be reserved in the file system is as follows:

Required area size per dump file = Amount of mounted memory + 1 GB

To confirm that the area required for the saved dump files has been reserved in the file system (e.g., /dev/sda5) on the disk, execute the following command:

```
# mount /dev/sda5 /mnt
# df -h /mnt
# umount /mnt
```

9.2 Estimating the Dump-use Memory

In a PRIMEQUEST system using Linux (Red Hat), RHEL5 (IPF), if used, affects the design of memory used for dump functions.

9.2.1 RHEL5 (IPF)

kdump starts a kernel (OS) for kdump separately from the kernel that is run to collect a dump. For this reason, the memory required for running the kernel for kdump is reserved in advance. This reserved memory cannot be used for an ordinary operation. The size of this memory is fixed at 512 MB in PRIMEQUEST. Take this value into consideration in memory design.

Appendix A Software Supplied with PRIMEQUEST Hardware

The software programs supplied with PRIMEQUEST hardware are listed below.

Table A.1 Software programs supplied with PRIMEQUEST hardware

No.	Name	Function	EFI	Target OS		
				Linux (RedHat)	Linux (SUSE)	Windows
1	EFI/BIOS	Firmware	Y	Y	Y	Y
2	PSA	Hardware system management	-	Y	Y	Y
3	SIRMS	Collecting software configuration information	-	Y (*1)	-	Y (*1)
SCSI Related						
4	Driver	Driver	Y	Y	Y	Y
5	Fusion MPT EFI appl.	F/W update tool	Y	-	-	-
6	Fusion MPT management tool	CIMS Browser		Y	Y	-
Broadcom LAN Related						
7	Driver	Driver	Y	Y	-	Y
8	BCM EFI appl.		Y	-	-	-
9	BACS(Win)	VLAN, aggregation, etc.	-	-	-	Y
Intel LAN Related						
10	Driver	Driver	Y	Y	Y	Y
11	Intel PROSet (Win)	VLAN, AFT/ALB, etc.	-	-	-	Y
Neterion LAN Related						
12	Driver	Driver	-	Y	-	Y
13	Xframe Control Panel (Win)	VLAN, Team, etc.	-	-	-	Y
Emulex FC Related						
14	Driver	Driver	Y	Y	Y	Y
15	lputil		-	Y	Y	Y
16	HBAnyware		-	Y	Y	Y
Others						
17	HBA block driver for PCL		-	Y	-	-
18	Log trace capture	Collecting logs (compatible with extended drivers)	-	Y (*1)	-	-

No.	Name	Function	EFI	Target OS		
				Linux (RedHat)	Linux (SUSE)	Windows
19	Installation support tool	Runs on an external PC (Windows) to create a configuration file	-	Y	Y	-
20	Installation Support Tool for Windows	Creating an installation floppy disk used to let Windows install automatically.	-	-	-	Y
21	Bundled-Software Package Installer	Installation tool (bundled software)	-	Y	Y	-
22	DSNAP	Executing a command as part of an OS operation to collect basic information about Windows OS.	-	-	-	Y
23	High-reliability tool package installer	Installation tool (bundled software)	-	-	-	Y
24	System Parameter Check Tool	Checking the environment variable (argument necessary for OS operation) that the user is going to set	-	Y	-	-
25	System Data Output Tool (fjsnap)	Executing the command for collecting the information required for support under OS operation.	-	Y	Y	-
26	SystemcastWizard Lite	• Remote installation • Backup or restoration of a disk	-	Y	Y	Y
27	SNMP MIB	MIB	-	Y	Y	Y
28	Software Support Guide	Executing a command as part of an OS operation to collect relevant support information.	-	-	-	Y
29	HRM/server	Maintenance support tool	-	Y	-	Y

Y: Supported

*1 This is supported only in Japan.

Glossary

ACS (AC Section)

AC power input section

ASIC (Application Specific Integrated Circuit)

Integrated circuit (IC) designed and manufactured for specific applications

API (Application Program Interface)

A set of instructions and functions used for developing operating systems and middleware

BB (Baseboard)

Unit on which CPUs, memories, and various chipsets are mounted

BIOS (Basic Input Output System)

Part of the operating system (OS) function. The BIOS is the system that controls input/output to devices. For the PRIMEQUEST-series machine, BIOS is a general term for PAL, SAL, and EFI.

BMC (Baseboard Management Controller)

The BMC is a system management controller that continuously monitors the system for serious hardware errors and notifies the OS of such errors.

BMM (BMC Module)

Board on which legacy I/O ports such as BMC, VGA, USB, and COM ports are mounted

Business LAN

LAN used to configure a user business system

CLI (Command Line Interface)

This interface with UNIX or DOS allows the user to enter commands and optional arguments to communicate with the OS.

CoA

Abbreviation for Certificate of Authenticity. This is a visual identifier that helps identify genuine Microsoft software and components.

COM Port (Communication Port)

Abbreviation for communication port. This is an RS-232C serial port for PC/AT-compatible machines. The COM port is also called an "RS-232C port." Most PC/AT-compatible machines each have two COM ports at the rear, and the ports are often used to connect modems, terminal adapters, or scanners. Most of these ports use D-Sub 25-pin or D-Sub 9-pin connectors.

DDR2 (Double Data Rate 2)

Standards for the next generation of memory that operates at higher speeds and consumes less power than conventional DDR memory

DIMM (Dual Inline Memory Module)

This compact memory module has pins on both sides and is mainly used in notebook PCs.

DVD-ROM (Digital Versatile Disc-Read-Only Memory)

Digital format for high-volume storage of data on optical disks

ECC (Error Checking Correction)

Error correction code or a method of using the error correction code to check for and correct errors

EFI (Extensible Firmware Interface)

Specifications for an interface between an OS and firmware. Instead of the BIOS, EFI is used for hardware control.

FC (Fibre Channel)

One of the serial interface standards. The Fibre Channel standard uses fiber cables as the transmission medium.

Firmware

Built-in software for basic hardware control

FWH (Firmware Hub)

LSI device from Intel Corporation. FWH is flash memory that stores SAL (BIOS). The PRIMEQUEST-series machine uses two types of FWH: one type is mounted on an SB and the other is mounted in an IO Unit.

GAC (Global Address Controller)

One of the ASICs developed by Fujitsu for the PRIMEQUEST-series machine

GbE (Gigabit Ethernet)

Ethernet standards for high-speed communication of up to 1 Gbps

GDS

Abbreviation for PRIMECLUSTER GDS

GDX (Global Data Xbar)

One of the ASICs developed by Fujitsu for the PRIMEQUEST-series machine

GLS

Abbreviation for PRIMECLUSTER GLS

HBA

Abbreviation for a host bus adapter

HDD (Hard Disk Drive)

Device that reads a hard disk. HDD may also be an abbreviation for the hard disk itself.

Hot Plug

Method of replacing components while power is on

HTTP (Hypertext Transfer Protocol)

Protocol used by Web servers and clients for data transmission

I2C (Inter Integrated Circuit)

Protocol used for high-speed communication between integrated circuits (ICs)

IA (Intel Architecture)

Generic term for the basic design (architecture) of Intel's microprocessors

IFT (Instruction Fetch)

Mechanism for reading instructions stored in memory

IHV (Independent Hardware Vendor)

This hardware provider has no special relationship with a particular hardware or OS maker.

IP (Internet Protocol Address)

Identification number assigned to each computer connected to an IP network, such as the Internet and intranets

IPMI (Intelligent Platform Management Interface)

Standardized interface specifications established so that SNMP and server management software can monitor server hardware independently of specific hardware systems and OSs

ISV (Independent Software Vendor)

This application software provider has no special relationship with a particular hardware or OS maker.

LAN (Local Area Network)

Using optical fiber, for example, this network allows data to be transferred among computers and printers connected in a facility.

LDAP (Lightweight Directory Access Protocol)

Protocol used to access directory databases in a TCP/IP network, such as the Internet and intranets

LDX (Local Data Xbar)

One of the ASICs developed by Fujitsu for the PRIMEQUEST-series machine

LED

Abbreviation for a light emitting diode

MAC address (Media Access Control Address)

Unique address assigned to each network interface device, switch, or router mounted on a network interface card (NIC) or motherboard

Management LAN

This LAN connects the MMB to the system in the BB and to LANs outside the cabinet so that the PRIMEQUEST system can be managed.

MIB (Management Information Base)

Information released by a network device managed by SNMP in order to post the device status to an external destination

Middleware

Software that runs under an OS and provides application software with more advanced and detailed functions than the OS. It is positioned between the OS and application software in terms of its characteristics.

MMB (Management Board)

This server management board is a system control unit whose tasks include control and monitoring of cabinet hardware, and system initialization.

NIC (Network Interface Controller)

Hardware that supports network functions

NTP (Network Time Protocol)

Standard time information protocol used on the Internet. Highly precise time information with consideration of line speeds and load changes in paths can be obtained with this protocol.

PAL (Physical Abstract Layer)

Firmware that provides platform initialization and operating system boot functions

PCI (Peripheral Component Interconnect)

Bus architecture established by the PCI SIG for connecting PC components

PCI Hot Plug

Technology that enables PCI cards to be mounted and removed while the system is operating

PCI (Peripheral Component Interconnect)

Bus architecture established by PCI SIG for connecting PC components

PCIU (PCI Unit)

PCI-X card slot expansion unit that is mounted in a PCI_Box

PEXU

PCI Express card slot expansion unit that is mounted in a PCI_Box

Platform

OS type or environment that is the basis for operation of application software

POST (Power-On Self Test)

Hardware test that is automatically run when the computer is powered on

Private LAN

LAN used for internal control, under which firmware programs installed on hardware components communicate with one another. BMC firmware and MMB firmware installed on BB can use a private LAN for communication with one another. OSs and applications cannot use a private LAN.

PSA (PRIMEQUEST Server Agent)

Software that performs hardware error monitoring and configuration management over PRIMEQUEST

PSU (Power Supply Unit)

Component that converts AC voltage to DC voltage as a DC power supply

PXE

PXE (Preboot eXecution Environment) Network boot standard based on BIOS technology that enables remote operation of management tasks such as system start and OS installation/update

RAID (Redundant Array of Independent Disks)

Technology that increases reliability and processing speeds by using multiple hard disks as a single disk

REMCS (Remote Customer Support System)

Fujitsu's remote customer support service

RHEL (Red Hat Enterprise Linux)

Linux distribution released by Red Hat, Inc.

SAF-TE

Abbreviation for a SCSI accessed fault-tolerant enclosure

SAL (System Abstraction Layer)

Firmware that supports processor initialization and error recovery functions

SAN (Storage Area Network)

Dedicated network for connections between a server and storage devices

SAS (Serial Attached SCSI)

One of SCSI standards. SAS is a serial transmission interface for connecting devices such as hard disks to a computer.

SASBP (SAS Back Plane)

SASBP is used for HDD monitoring and LED on/off control.

SCSI (Small Computer System Interface)

Standards for connections between PCs and peripherals. SCSI was established by the American Standards Association.

SDRAM (Synchronous DRAM)

Memory standard for access speeds that are higher than those of DRAM

SEL (System Event Log)

Information on the processing parameters, processing, and processing results logged during hardware and software operations

SIRMS (Software Product Information Collection for Remote Maintenance Support)

Software that collects configuration information on software installed in PRIMEQUEST partitions

S.M.A.R.T. (Self-Monitoring Analysis Reporting Technology)

Function that enables a hard disk to monitor its own conditions and notify the BIOS of any error detected

SMP (Symmetric Multiple Processor)

Parallel processing system in which all processors work together through common memory resources

SNMP (Simple Network Management Protocol)

TCP/IP-compliant protocol for managing devices in a network

SSL (Secure Sockets Layer)

Protocol under which information is encrypted for transmission. SSL was developed by Netscape Communications Corp.

Systemwalker

One of Fujitsu's middleware products. Systemwalker is integrated operation management software.

Telnet

Protocol or standard method for remote control of computers connected to a TCP/IP network, such as the Internet and intranets

UPS (Uninterruptible Power Supply)

Power supply unit that stores power and protects against possible damage and loss of computer data from a momentary voltage drop or unexpected power failure

USB (Universal Serial Bus)

One of the standards on connecting peripheral devices such as keyboards and mice

VLAN (Virtual LAN)

Function that logically groups the ports of one switching hub so each group works as an independent LAN

Web UI (Web User Interface)

Interface that uses a Web browser for displaying information to users and for user operations

Index

I

L

M

N

O

P

R

S

T

U

V

W

Reader's Comment Form

We would appreciate your comments and suggestions for improving this publication.

Date:_____
Your Name:_____
Company: _____
Address: _____
City/State/Zip:_____
Phone/Email address: _____

Publication No.: C122-B018-01EN
Publication Name:PRIMEQUEST 510A
SYSTEM DESIGN GUIDE

Your Comments:

Page	Line	Comments
Reply requested: ❑ Yes ❑ No		

Please evaluate the overall quality of this manual by checking (✓)the appropriate boxes

	Good	Fair	Poor		Good	Fair	Poor		Good	Fair	Poor
Organization:	❑	❑	❑	Use of examples:	❑	❑	❑	Legibility:	❑	❑	❑
Accuracy:	❑	❑	❑	Index coverage:	❑	❑	❑	Binding:	❑	❑	❑
Clarity:	❑	❑	❑	Cross				Figures and tables:	❑	❑	❑
Overall rating of				referencing:	❑	❑	❑	General appearance:	❑	❑	❑
this publication:	❑	❑	❑								

Technical level: ❑ Too detailed ❑ Appropriate ❑ Not enough detail

All comments and suggestions become the property of Fujitsu Limited.

For Users in U.S.A., Canada, and Mexico

Fold and fasten as shown on back
No postage necessary if mailed in U.S.A.

Fujitsu Computer Systems
Attention: Engineering Ops M/S 249
1250 East Arques Avenue
Sunnyvale, CA 94085-5401
FAX: (408) 746-6676

For Users in Other Countries

Fax this form to the number below or send this form to the address below.

Fujitsu Learning Media Limited
FAX: 81-3-3730-3702
37-10 Nishi-Kamata 7-chome
Oota-Ku
Tokyo 144-0051
JAPAN

FUJITSU LIMITED

FOLD AND TAPE

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

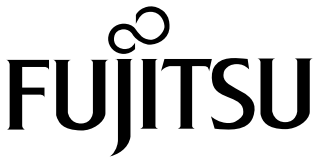
FIRST-CLASS MAIL PERMIT NO 741 SUNNYVALE CA

POSTAGE WILL BE PAID BY ADDRESSEE

FUJITSU

FUJITSU COMPUTER SYSTEMS
ATTENTION ENGINEERING OPS M/S 249
1250 EAST ARQUES AVENUE
SUNNYVALE CA 94085-5401

FOLD AND TAPE


FUJITSU