C122-B009-07EN

PRIMEQUEST
520A/520/420

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

The contents of this manual shall not be disclosed in any way or reproduced in any media without the express written permission of Fujitsu Limited.

All Rights Reserved, Copyright © FUJITSU LIMITED 2006-2008

C122-B009-07EN

Revision History

(1/1)

Edition	Date	Revised section (Added/ Deleted/ Altered) (Note)	Details
01	2006-02-20	—	—
02	2006-04-17	Section 2.3.2, 2.3.4, 4.1, 4.3 (addition, correction) Section 2.11 (addition)	Addition and correction of descriptions and notes Section added
03	2006-08-11	All chapters (correction) Chapter 1, 2, and 7	Addition of descriptions for PRIMEQUEST 500 series and Linux SUSE Correction of description for partition configuration with XPAR installed
04	2007-03-05	Section 2.3.4 (correction) Chapter 9 (correction)	Correction of description for mounting condition of PCI cards Addition of description for supported backup software
05	2007-08-31	2.11 (correction) Chapter 10 (correction)	Addition of description for PEXU Addition of Linux dump function kdump
06	2008-03-10	Section 2.6 (correction)	Addition of information of PRIMEQUEST 520A Addition of notes on using the REMCS port
07	2008-04-10	Section 4.2.7 (correction)	Correction of description for NTP operation

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
- Other Reference Manuals
- Abbreviations
- Text Conventions
- Syntax of the Command Line Interface (CLI)
- Notes Regarding Notations Used in This Manual
- Conventions for Alert Messages
- Environmental Requirements for Using This Product
- Reader Feedback

Structure and Contents of This Manual

This manual is organized as described below.

CHAPTER 1 PRIMEQUEST Overview

Describes features of the main unit and software.

CHAPTER 2 General Hardware Configuration

Describes the hardware mounted in the main unit.

CHAPTER 3 General Software Configuration

Describes the OS, firmware, supplied software, and software that can be linked.

CHAPTER 4 Hardware System Management

Describes the basic mechanism related to hardware system management, functions of the Management Board (MMB) that manages the main unit as a whole, and functions of the PRIMEQUEST Server Agent (PSA) managed by the MMB.

CHAPTER 5 Redundant Configuration

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

CHAPTER 6 Hot Swapping

Describes hot swapping of hardware components of the main unit.

CHAPTER 7 Partition

Describes how to set up and operate partitions of the main unit and describes the necessary items for determining the partition configuration.

CHAPTER 8 Clustering

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

CHAPTER 9 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 10 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 520A/520/420 System Design Guide	Explains requirements, considerations, and notes on the system operation design of the PRIMEQUEST 520A/520/420.	C122-B009EN
PRIMEQUEST 520A/520/420 Installation Planning Manual	Explains specifications and requirements for installation sites that are applicable to the installation of the PRIMEQUEST 520A/520/420.	C122-H002 EN
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 580A/540A/520A/500/400 Series 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-E003EN
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)	Red Hat (*1)
Red Hat® Enterprise Linux® 5 (for Intel Itanium)	
Red Hat® Enterprise Linux® AS (v.4 for Itanium)	RHEL-AS4(IPF)
Red Hat® Enterprise Linux® 5 (for Intel Itanium)	RHEL5(IPF) (*2)
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: Version-independent abbreviation

*2: 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 520A/ 520/420 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.
- The IO Unit is indicated as "IOU" in the MMB Web-UI and in the figures shown in this manual.

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, rear, or side 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 520A/520/420 Installation Planning Manual* (C122-H002EN).

Reader Feedback

- This manual assumes two BMMs (optional products) are connected to a single IO Unit or IOX. This is reflected in the explanations and figures appearing in the manual. At present, however, the PRIMEQUEST 420 series supports connection to only one BMM (BMM#0) per IO Unit or IOX.
- 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

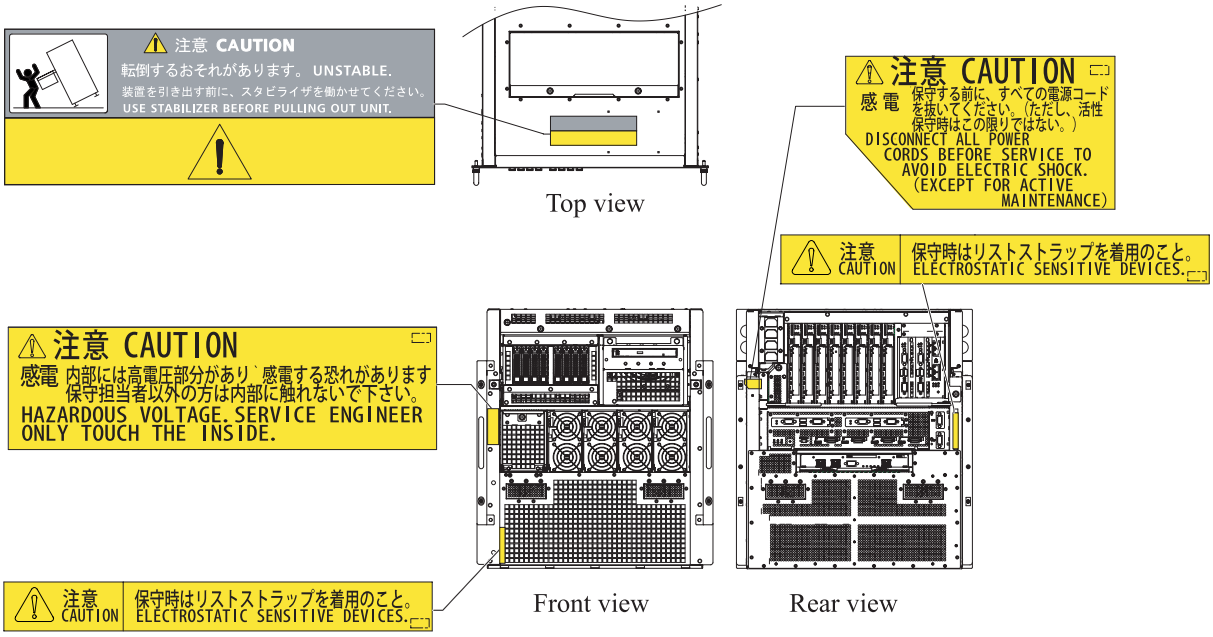
The following labels are attached to this product:

These labels provide information to the users of this product.

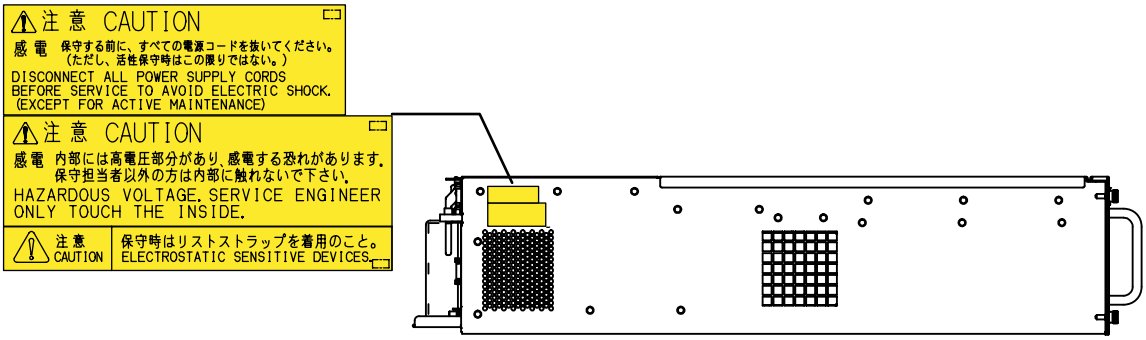
CAUTION

Do not peel off the labels.

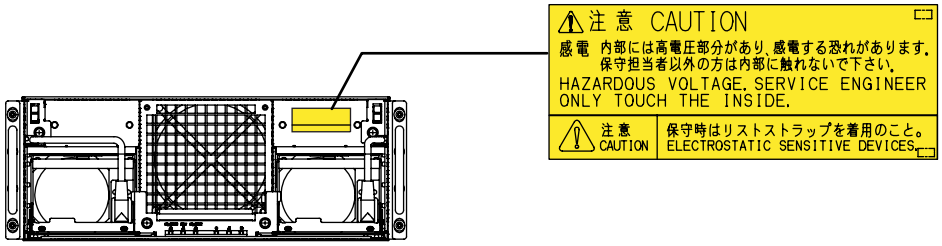
- Main unit



- PCI_Box



Left side view



Front view (without cover)

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, rear, or side 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

Preface i
Structure and Contents of This Manual i
Other Reference Manuals iii
Abbreviations iv
Text Conventions iv
Syntax of the Command Line Interface (CLI) v
Notes Regarding Notations Used in This Manual v
Conventions for Alert Messages vi
Environmental Requirements for Using This Product vi

NOTE ON SAFETY ix
Alert Labels ix

Product Handling xi
Maintenance xi
Remodeling/Rebuilding xi

CHAPTER 1 PRIMEQUEST Overview 1-1

1.1 Basics of PRIMEQUEST 1-1
1.1.1 PRIMEQUEST conceptual diagram 1-3
1.1.1.1 Hardware components 1-4
1.1.2 Product lineup 1-5
1.2 PRIMEQUEST Features 1-7
1.2.1 Hardware technologies 1-8
1.2.1.1 PRIMEQUEST-dedicated chipset 1-8
1.2.1.2 IPF 1-9
1.2.1.3 Symmetric multi-processing (SMP) technology 1-9
1.2.1.4 System Mirror Mode 1-10
1.2.1.5 Partition function 1-13
1.2.1.6 Flexible I/O mechanism 1-16
1.2.1.7 Hardware redundancy and hot maintenance function 1-20
1.2.1.8 Management board (MMB) 1-22
1.2.2 Software technologies 1-23
1.2.2.1 Firmware 1-23
1.2.2.2 Linux 1-24
1.2.2.3 Windows 1-24
1.2.2.4 Clustering software 1-25
1.2.2.5 Fujitsu middleware products 1-26

CHAPTER 2 General Hardware Configuration. 2-1

2.1 Cabinet 2-4
2.2 System Board (SB) 2-5
2.2.1 CPU 2-6
2.2.2 Memory 2-7
2.2.2.1 Memory interleave 2-8
2.3 IO Unit 2-9
2.3.1 BMM 2-10
2.3.2 PCI slot 2-10
2.3.3 Compatible PCI cards 2-12
2.3.4 Mounting conditions of PCI cards and usable built-in I/Os 2-13
2.3.4.1 Mounting conditions of PCI cards 2-13
2.4 HDD 2-14
2.5 IO Unit for PCI_Box Connection (IOX) 2-15
2.6 Management Board (MMB) 2-17
2.7 Crossbar 2-20
2.8 Backplane (BP) 2-20
2.9 Operator Panel (Board) (OPL) 2-21
2.10 Power Supply Unit (PSU) 2-22
2.11 PCI_Box 2-23
2.11.1 PCIU 2-27
2.11.1.1 PCI slot 2-27
2.11.1.2 PCI card cassette 2-28
2.11.1.3 Compatible PCI cards 2-28
2.11.2 PEXU 2-29
2.11.2.1 PCI slots in the PEXU 2-29
2.11.2.2 Card cassettes in the PEXU 2-30
2.11.2.3 Compatible PCI cards 2-30
2.11.3 Requirements on PCIU and PEXU connections and combinations 2-31
2.11.3.1 Requirements on connecting PCIUs and PEXUs to one IOX 2-31
2.11.3.2 IOX and PEXU connection examples 2-31
2.11.4 Connection requirements when XPAR partitioning is used 2-33
2.11.5 Requirements on a combination of PCIUs and PEXUs in one PCI_Box 2-33
2.11.6 Requirements on a combination of PCIUs and PEXUs in one partition 2-33
2.12 Expansion File Unit 2-35

CHAPTER 3 General Software Configuration 3-1

3.1 Firmware 3-2
3.1.1 Processor Abstraction Layer (PAL) 3-2
3.1.2 System Abstraction Layer (SAL) 3-2
3.1.3 Extensible Firmware Interface (EFI) 3-2
3.1.4 BMC firmware 3-2
3.1.5 MMB firmware 3-3
3.2 OS (Linux or Windows) 3-3
3.2.1 Linux 3-4
3.2.2 Windows 3-5
3.3 Supplied Software 3-5
3.3.1 PSA (system management software) 3-5
3.3.2 Drivers 3-5
3.3.3 SystemcastWizard Lite 3-5
3.3.4 System data output tool (fjsnap) 3-6
3.4 Linked Software 3-6
3.4.1 PRIMECLUSTER 3-6
3.4.2 PRIMECLUSTER GDS 3-6
3.4.3 PRIMECLUSTER GLS 3-7
3.4.4 PRIMEQUEST System Disk Mirror for Windows (PSDM) 3-7
3.4.5 Systemwalker 3-7
3.4.6 Backup and restoration software 3-8

CHAPTER 4 Hardware System Management 4-1

4.1 Basic Mechanism 4-1
4.1.1 LAN configuration (management LAN/private LAN/business LAN) 4-3
4.1.1.1 Management LAN 4-4
4.1.1.2 Private LAN 4-9
4.1.1.3 Business LAN 4-10
4.1.2 IP addresses of PRIMEQUEST management LAN 4-10
4.2 MMB Functions 4-12
4.2.1 User privilege management 4-13
4.2.2 Hardware configuration display 4-14
4.2.3 Hardware monitor 4-15
4.2.4 Partition settings and configuration display 4-17
4.2.5 Console redirection 4-18
4.2.6 MMB use environment setting 4-19
4.2.7 Time synchronization 4-21
4.2.8 Power control/scheduled operation 4-22
4.2.9 Firmware maintenance 4-22
4.2.10 Setting information save and restore 4-23

4.3 PSA Functions 4-25
4.3.1 Operation management GUI 4-26
4.3.2 Hardware configuration management 4-28
4.3.3 OS information display 4-28
4.3.4 Hardware error monitoring 4-29
4.3.5 Log collection, analysis, and display 4-31
4.3.6 Maintenance operation 4-32
4.3.7 REMCS linkage 4-32
4.3.8 PRIMECLUSTER linkage 4-33
4.3.9 Operation management software linkage 4-35
4.3.10 Control targets of MMB user interface 4-36
4.3.11 Information managed on partition 4-38

CHAPTER 5 Redundant Configuration 5-1

5.1 Redundancy of Components 5-2
5.2 System Mirror Mode 5-3
5.3 Redundancy of an HDD (Use of PRIMECLUSTER GDS) 5-6
5.3.1 Redundancy provided to system disk drives 5-6
5.3.2 Redundancy of external storage 5-8
5.4 Redundancy of the Business LAN (Use of PRIMECLUSTER GLS) 5-10
5.4.1 Duplication of a transmission line between servers (high-speed switch method) 5-11
5.4.2 Duplication between a server and a hub and between switches on the same network (NIC switch method) 5-12
5.5 Degradation Function 5-14

CHAPTER 6 Hot Swapping 6-1

6.1 Hot Swappable Components 6-2

CHAPTER 7 Partition 7-1

7.1 Partitioning Function Overview 7-1
7.2 PRIMEQUEST Partitioning Function 7-3
7.2.1 Mounting requirements 7-3
7.2.1.1 Mounting requirements of SBs 7-3
7.2.1.2 Mounting requirements of IO Units or IOX 7-3
7.2.2 PCI_Box allocation 7-4
7.3 Partitioning Function of the PRIMEQUEST 520A/520 with XPAR Installed 7-6
7.3.1 Splitting SBs, IO Units, or IOXs when XPAR has been installed . . . 7-6
7.3.2 Partition Granularities 7-7
7.3.2.1 SB partition granularity 7-7
7.3.2.2 IO Unit and IOX partition granularity 7-8

7.3.3 Mounting requirements (when XPAR has been installed) 7-9
7.3.3.1 SB mounting requirements 7-9
7.3.3.2 Mounting requirements of IO Units and IOX 7-11
7.3.4 Allocating a PCI_Box 7-18
7.4 Partition Configuration Definition 7-20
7.5 Points to Consider when Determining Configurations 7-21
7.5.1 SB and IO Unit maintenance policy 7-21
7.5.2 Flexible I/O component and reserved SB 7-21
7.5.3 Notes on System Mirror Mode operation 7-23
7.5.4 Notes on the memory configuration 7-23
7.5.5 Operating environment 7-23
7.6 Function for Changing Partition Configurations 7-24

CHAPTER 8 Clustering 8-1

8.1 The basics of clustering 8-1
8.2 Cluster Functions Provided in Linux Environments 8-5
8.2.1 Cluster Configurations 8-5
8.2.2 PRIMECLUSTER 8-6
8.3 Cluster Functions Provided in Windows Environments 8-7
8.3.1 Basic cluster configuration 8-7
8.3.2 Virtual server 8-8
8.3.3 Scope of remedies 8-9
8.3.4 Notes on construction 8-9
8.4 Points to be Considered at Cluster System Installation 8-9

CHAPTER 9 Backup and Restoration 9-1

9.1 Importance of Backing Up Data 9-1
9.2 General Methods of Backup and Restoration 9-1
9.2.1 OS-standard utility (Linux (Red Hat): dump/restore, Windows: NTBackup) 9-2
9.2.2 Cloning software (SystemcastWizard Lite) 9-5
9.2.3 Snapshot software (PRIMECLUSTER GDS Snapshot) 9-7
9.2.4 Backup software (ETERNUS SF ACM, VERITAS NetBackup, NetVault, NetWorker) 9-9
9.3 Comparison of Backup/Restoration Methods 9-15

CHAPTER 10 Considerations for Maintenance 10-1

10.1 Estimating the Dump-use Area 10-1
10.1.1 RHEL-AS4 (IPF) 10-1
10.1.2 RHEL5 (IPF) 10-3
10.2 Estimating the Dump-use Memory 10-4
10.2.1 RHEL5 (IPF) 10-4

Appendix A Software Supplied with PRIMEQUEST Hardware A-1

Glossary GL-1

Index IN-1

Figures

Figure 1.1 PRIMEQUEST 520A/520/420 1-2
Figure 1.2 Concept of PRIMEQUEST configuration 1-3
Figure 1.3 Configuration of all hardware 1-4
Figure 1.4 PRIMEQUEST technologies 1-7
Figure 1.5 System Mirror Mode 1-11
Figure 1.6 Outline of System Mirror Mode 1-11
Figure 1.7 Improved business continuity with System Mirror Mode 1-12
Figure 1.8 Partitions 1-13
Figure 1.9 Independent resource expansion in each partition 1-14
Figure 1.10 Operation on different platforms 1-15
Figure 1.11 Connection between SBs and IO Units in a conventional model 1-16
Figure 1.12 Example of the flexible I/O mechanism and partition configuration 1-17
Figure 1.13 Reduced MTTR 1-19
Figure 1.14 Duplication of the main components 1-20
Figure 1.15 MMB 1-22
Figure 1.16 Clustering 1-25
Figure 1.17 Fujitsu middleware products 1-26
Figure 2.1 PRIMEQUEST 520A/520/420 physical configuration (full package) 2-2
Figure 2.2 Main unit 2-4
Figure 2.3 SB 2-5
Figure 2.4 Memory (DIMM) slot numbers and locations 2-7
Figure 2.5 IO Unit (Full package) 2-9
Figure 2.6 PCI card cassette 2-10
Figure 2.7 Mounting conditions of SCSI cards 2-14
Figure 2.8 IOX (Full package) 2-15
Figure 2.9 MMB 2-18
Figure 2.10 OPL 2-21
Figure 2.11 PSU exterior view 2-22
Figure 2.12 PCI_Box 2-24
Figure 2.13 PCI_Box connections (only with PCIU) 2-25
Figure 2.14 PCI_Box connections (with a combination of PCIUs and PEXUs) 2-26
Figure 2.15 PCIU (with PCI card cassettes inserted) 2-27
Figure 2.16 PCI card cassette (PCI_Box) 2-28
Figure 2.17 PEXU (with PEXU card cassettes inserted) 2-29
Figure 2.18 PCI Express card cassette 2-30

Figure 2.19 IOX and PEXU connections 2-32
Figure 2.20 IOX and PEXU connections (prohibited patterns) 2-32
Figure 3.1 PRIMEQUEST software configuration 3-1
Figure 4.1 Basic mechanism 4-1
Figure 4.2 Management LAN 4-4
Figure 4.3 Private LAN 4-9
Figure 4.4 Management LAN network configuration and IP addresses 4-10
Figure 4.5 Main MMB functions 4-12
Figure 4.6 Example of the configuration display window 4-14
Figure 4.7 Example of the system event log display window 4-15
Figure 4.8 Example of the filter condition setting window 4-16
Figure 4.9 Example in which different OSs and OSs of different versions are mounted in respective partitions 4-17
Figure 4.10 Example of display of the partition configuration window 4-17
Figure 4.11 Flow of redirected information 4-18
Figure 4.12 [Console Redirection Switch] window 4-18
Figure 4.13 Example of setting the usage environment 4-19
Figure 4.14 Example of the access control settings window 4-20
Figure 4.15 Time synchronization image 4-21
Figure 4.16 Example of the scheduled operation display window 4-22
Figure 4.17 Schematic of configuration information save and restore operation 4-24
Figure 4.18 PSA functions 4-26
Figure 4.19 Web-UI function 4-27
Figure 4.20 Overview of error monitoring 4-30
Figure 4.21 REMCS linkage 4-33
Figure 4.22 PRIMECLUSTER linkage 4-34
Figure 4.23 Operation management software linkage 4-35
Figure 5.1 General redundant configuration 5-2
Figure 5.2 Extended Mirror Mode 5-4
Figure 5.3 Logical numbers of memory for which a redundant configuration has been adopted 5-5
Figure 5.4 Minimum redundancy configuration 5-6
Figure 5.5 Redundancy including SAS controllers 5-7
Figure 5.6 Disk device mirroring between cabinets 5-8
Figure 5.7 ETERNUS multipath connection 5-9
Figure 5.8 Example of implementing redundancy using the high-speed switching method 5-11
Figure 5.9 High-speed switching method 5-12
Figure 5.10 Example of implementing redundancy using the NIC switch method 5-12

Figure 5.11 NIC switching method 5-13
Figure 6.1 Layout of included components 6-2
Figure 7.1 Example of partition configuration 7-2
Figure 7.2 Allocation to a PCI_Box partition (only if a PCIU is mounted) . . . 7-4
Figure 7.3 Allocation to a PCI_Box partition (both a PCIU and a PEXU are mounted) 7-5
Figure 7.4 Sample partition configuration (logical splitting) 7-6
Figure 7.5 SB partition granularities 7-8
Figure 7.6 SB partition granularities and allocation rules 7-10
Figure 7.7 IO Unit granularities and allocation rules 7-12
Figure 7.8 Typical IO Unit partition granularity allocations (1/3) 7-13
Figure 7.9 Typical IO Unit partition granularity allocations (2/3) 7-14
Figure 7.10 Typical IO Unit partition granularity allocations (3/3) 7-14
Figure 7.11 IOX partition granularities and allocation rules 7-15
Figure 7.12 Typical IOX partition granularity allocations (1/3) 7-16
Figure 7.13 Typical IOX partition granularity allocations (2/3) 7-17
Figure 7.14 Typical IOX partition granularity allocations (3/3) 7-17
Figure 7.15 Allocation of a PCI_Box to partitions (only with PCIU) 7-18
Figure 7.16 Allocation of a PCI-Box to partitions (with a combination of PCIUs and PEXUs) 7-19
Figure 7.17 [Partition Configuration] window 7-20
Figure 7.18 Flexible I/O component 7-22
Figure 8.1 Example for 1:1 standby 8-2
Figure 8.2 Example for N:1 standby 8-3
Figure 8.3 Example for mutual standby 8-3
Figure 8.4 Example for a cascade 8-4
Figure 8.5 Inter-cabinet cluster system configuration 8-5
Figure 8.6 Basic cluster configuration (Windows) 8-7
Figure 8.7 Virtual server 8-8
Figure 8.8 Position of the interconnect 8-10
Figure 8.9 Extended Mirror Mode and cluster configuration 8-11
Figure 9.1 Local backup using a standard utility (when the KVM console is used) 9-3
Figure 9.2 Backup to the remote host by using a standard utility (when using the KVM console) 9-4
Figure 9.3 Backup using SystemcastWizard Lite 9-5
Figure 9.4 System volume backup using PRIMECLUSTER GDS Snapshot 9-7
Figure 9.5 Data area backup and restoration using PRIMECLUSTER GDS Snapshot 9-8
Figure 9.6 Backup of data on the ETERNUS by using ACM 9-9

Figure 9.7 Data backup using VERITAS NetBackup 9-11
Figure 9.8 Online backup of a DB using VERITAS NetBackup 9-11
Figure 9.9 Data backup using NetWorker 9-13

Tables

Table 1.1 Hardware specifications 1-5
Table 1.2 Relationship of the above partitions, SBs, and IO Units 1-18
Table 1.3 Linux OS features 1-24
Table 2.1 Types and maximum numbers of boards and components 2-3
Table 2.2 SB specifications 2-6
Table 2.3 IO Unit external interfaces 2-10
Table 2.4 PCI slot specifications (IO Unit). 2-11
Table 2.5 PCI-Express slot specifications (IO Unit) 2-11
Table 2.6 Compatible PCI-cards 2-12
Table 2.7 Compatible PCI-Express cards 2-12
Table 2.8 PCI card mounting conditions 2-13
Table 2.9 IOX unit external interfaces 2-16
Table 2.10 MMB external interfaces 2-18
Table 2.11 LED display on the OPL 2-21
Table 2.12 PSU specifications 2-22
Table 2.13 Number of SBs and IO Units mounted and PSU configuration . . . 2-22
Table 2.14 PCI_Box configuration for PRIMEQUEST. 2-24
Table 2.15 PCI slot specifications (PCI_Box) 2-27
Table 2.16 Compatible PCI Express cards 2-30
Table 2.17 Combination of PCIUs and PEXUs in one partition (pattern 1) . . . 2-34
Table 2.18 Combination of PCIUs and PEXUs in one partition (pattern 2) . . . 2-34
Table 2.19 Combination of PCIUs and PEXUs in one partition (pattern 3) . . . 2-34
Table 4.1 Inter-partition communication connection mode (No VLAN mode) 4-7
Table 4.2 Inter-partition communication disconnection mode (VLAN mode). 4-8
Table 4.3 All-partition communication disconnection mode (Port disable mode) 4-8
Table 4.4 Log file information 4-31
Table 4.5 Operations available from GUI on a partition 4-37
Table 4.6 Information managed on partition 4-38
Table 5.1 System Mirror Mode Specifications. 5-3
Table 6.1 Component hot swapping and/or redundant configuration availability 6-3
Table 7.1 Effects of LSB failures 7-8
Table 7.2 SB partition granularities and allocation rules 7-11
Table 7.3 IO Unit allocation rules 7-13
Table 7.4 IOX allocation rules 7-16
Table 9.1 Comparison of backup methods 9-15

Table 10.1 Areas used to collect dump data 10-2
Table 10.2 Devices that can be used as a dump device 10-2
Table A.1 Software programs supplied with PRIMEQUEST hardware A-1

CHAPTER 1 PRIMEQUEST Overview

This chapter provides an overview of PRIMEQUEST as follows:

- Basics of PRIMEQUEST
- PRIMEQUEST Features

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 520A/520/420.

The PRIMEQUEST 520A/520 houses up to 8 CPUs and can be equipped with up to 256 GB of memory, internal hard disks with a maximum capacity of 288 GB, PCI-X slots that can be expanded to up to 16 slots, and up to 8 PCI Express slots. A cabinet can normally be split into two partitions. When an XPAR (optional) is installed, one cabinet can be split in up to four partitions.

The PRIMEQUEST 420 houses up to 8 CPUs and can be equipped with up to 256 GB of memory, internal hard disks with a maximum capacity of 288 GB, PCI-X slots that can be expanded to up to 16 slots, and up to 2 PCI Express slots. A cabinet can normally be split into two partitions.

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 520A/520/420.

Figure 1.1 PRIMEQUEST 520A/520/420

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.

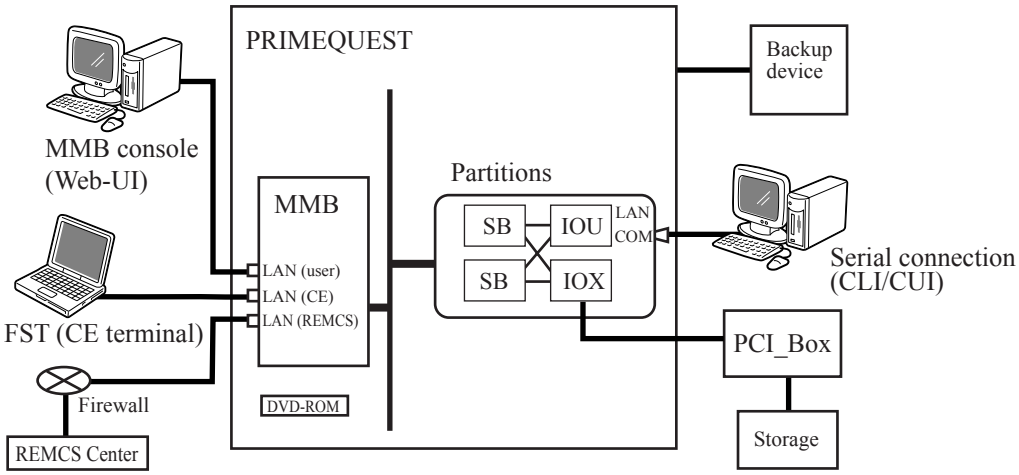


Figure 1.2 Concept of PRIMEQUEST configuration

A partition is a system that is equipped with the functions of a processing unit. Each partition contains software resources, such as an operating system and application programs, as well as hardware resources, such as system boards (SBs) and IO Units. Because partitions are independent of each other, they can be used as single servers that are assigned different tasks. Each partition can be connected to an external storage device via a PCI card (within the cabinet or through a PCI_Box). 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.

1.1.1.1 Hardware components

The main PRIMEQUEST components include the SB, IO Unit, crossbar, and MMB. Figure 1.3 shows the connection configuration of these main components. In Figure 1.3, a rectangle represents one SB, IO Unit or MMB board.

System boards (SBs) are boards on which CPUs, memory modules, PRIMEQUEST-dedicated chipsets, and other components are mounted. Up to two SBs can be mounted in the cabinet.

One SB can accommodate one to four CPUs. Therefore, the PRIMEQUEST 520A/520/420 can contain up to 8 CPUs in total. One SB can contain up to 128 GB of memory. This means that the PRIMEQUEST 520A/520/420 can contain up to 256 GB of memory.

IO Units are boards on which IO boards, PCI slots, and other components are mounted. Each unit accommodates four PCI-X slots. Moreover, the PRIMEQUEST 520A/520 accommodates four PCI-Express slots, while the PRIMEQUEST 420 accommodates two PCI-Express slots.

Also, an IO Unit for PCI_Box connection (IOX) can be optionally mounted, and each IOX contains four PCI_Box interface ports. The PCI_Box interface port is a port used to connect an optional PCI_Box. PCI slot expansion is enabled by connecting a PCI_Box.

The system interconnect is a bus with the crossbar function, which sends and receives addresses and data between SBs or between an SB and IO Unit.

The MMB is connected to each component in the cabinet and manages hardware as a whole. The MMB is also duplicated to increase its reliability.

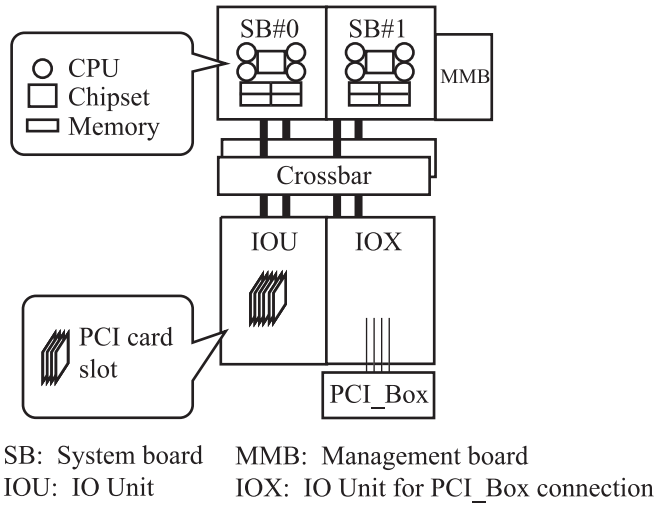


Figure 1.3 Configuration of all hardware

1.1.2 Product lineup

The PRIMEQUEST 520A/520/420 is a server that accommodates up to 8 CPUs.

Table 1.1 Hardware specifications

Item		Specification
CPU	Type	Itanium
	Freq./L3cache	PRIMEQUEST 520A: 1.60 GHz/24 MB, 1.42 GHz/12 MB PRIMEQUEST 520: 1.60 GHz/24 MB, 1.42 GHz/12 MB PRIMEQUEST 420: 1.60 GHz/9 MB, 1.50 GHz/4 MB
	Number of CPUs	Up to 8 CPUs
System board (SB)		Up to 2 boards
IO Unit	IO Unit	1
	IOX	1 (option)
Chipset		PRIMEQUEST-dedicated chipset
Memory (maximum capacity)(*1)		256 GB
Interconnect		Point-to-point crossbar
System Mirror Mode (option)		Standard Mirror Mode Extended Mirror Mode
Flexible I/O		Supported
Partition		PRIMEQUEST 520A/520: Up to 2 partitions * Up to four partitions if XPAR (option) is installed
		PRIMEQUEST 420: Up to 2 partitions
Built-in I/O	LAN	4 × Gigabit Ethernet per IO Unit (IO Unit)
	External port	Video, keyboard (USB), mouse (USB)
	DVD-ROM drive	1
HDD		Up to 8 drives (288 GB)
Number of PCI slots		Up to 20 slots • Built-in: - PRIMEQUEST 520A/520 8 slots (PCI-X slots: 4, PCI Express slots: 4) - PRIMEQUEST 420 6 slots (PCI-X slots: 4, PCI Express slots: 2) • PCI_Box: 12 slots
PCI_Box (option)		1
Server management		MMB
Redundant configuration		SB, PSU, HDD (Linux only), memory, PCI card (*3), IO_PSU, fan
Hot swapping		PSU, fan, HDD (Linux only), OPL, PCI card
External dimensions (WxDxH) (mm)		482 × 820 × 530

Item	Specification
Weight	150 kg
Power consumption (*2)	PRIMEQUEST 520A/520: 3.2 KVA PRIMEQUEST 420: 3.1 KVA

MMB: Management board

PSU: Power supply unit

*1 The maximum memory capacity of 4 GB/board is mounted. The effective memory size in Extended Mirror Mode is reduced by half.

*2 This indicates a power value across the equipment.

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

1.2 PRIMEQUEST Features

PRIMEQUEST has the following three features:

- High performance
  The high-speed multiprocessor and high-speed crossbar provide 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 such as those for partition-related settings and status monitoring can be centrally managed from the MMB Web-UI.

Figure 1.4 shows the technologies underlying the three features. The following sections describe the overview of these technologies.

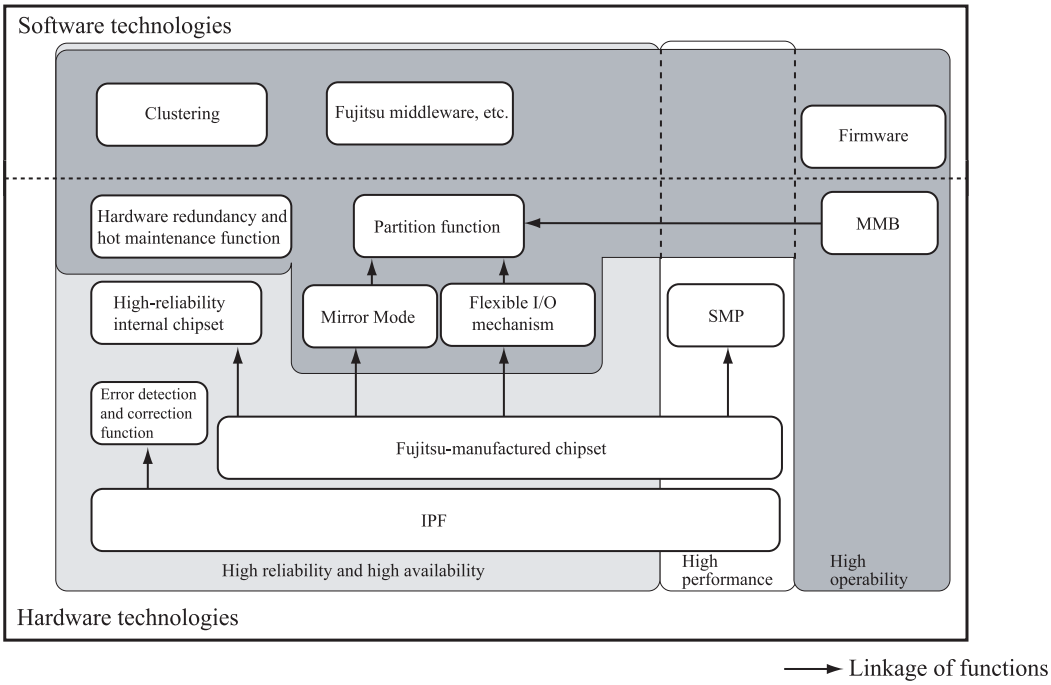


Figure 1.4 PRIMEQUEST technologies

1.2.1 Hardware technologies

This section describes the following essential hardware technologies shown in Figure 1.4.

- Technologies underlying high performance
  - PRIMEQUEST-dedicated chipset
  - IPF
  - Symmetric multi-processing (SMP) technology
- Technologies underlying high reliability and high availability
  - System Mirror Mode
  - Partition function
  - Flexible I/O mechanism
  - 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:

- Large-scale 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.

- System Mirror Mode

  The chipset supports duplication of CPU, dual reading/writing of memory, and synchronizes duplication of crossbar. System Mirror Mode, one of the new technologies of the PRIMEQUEST-series machines, is thus implemented.

- 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. High scalability can thus be obtained with applications unchanged.

With multiple SBs linked, the PRIMEQUEST machine supports an SMP system configuration that supports up to eight CPUs.

By using large-scale SMP technology originally developed for mainframes and UNIX servers, high performance suitable for a high-reliability large-scale mission-critical system is implemented.

1.2.1.4 System Mirror Mode

System Mirror Mode duplicates the memory, crossbar and the inside of the chip set. Thus, even if one of the duplicated components fails, the other component, which is its mirror, can continue the system processing.

The use of System Mirror Mode can thus increase reliability and improve continuity of business.

PRIMEQUEST has two types of System Mirror Mode, Standard Mirror Mode and Extended Mirror Mode.

These modes can be switched using MMB Web-UI. A cabinet having more than one partition allows the combined use of Standard Mirror Mode and Extended Mirror Mode ; the appropriate mode can be used selectively according to the operation type.

Remarks: Split SBs or split IOUs and IOXs cannot be assembled into a partition running in Extended Mirror Mode. For more details, see CHAPTER 7, "Partition."

The features of each mode are described below.

- Standard Mirror Mode

  Reliability can be improved by duplicating the crossbar of the address path. The memory is not duplicated and up to the maximum amount of memory can be used.

- Extended Mirror Mode

  Substantial improvements in reliability can be achieved by duplicating the crossbar and the data part of the address path, the inside of the chip set, and the memory. For example, if a memory failure occurs, Extended Mirror Mode stops only the system that has the failure, and is able to continue the operation using the other system which is running normally. Therefore, the operation can be continued without stopping the entire system.

  For details of System Mirror Mode, see Section 5.2, "System Mirror Mode."

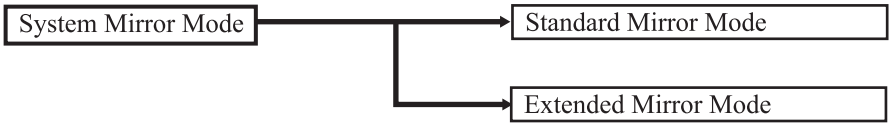


Figure 1.5 System Mirror Mode

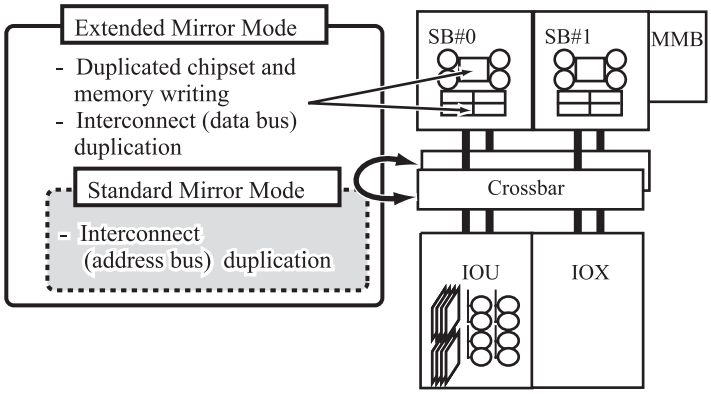


Figure 1.6 Outline of System Mirror Mode

Improved business continuity with System Mirror Mode

The PRIMEQUEST System Mirror Mode greatly improves business continuity.

Figure 1.7 shows a comparison of operation with and without System Mirror Mode. In operation without System Mirror Mode, if a hardware failure occurs while the system is operating, the system (task) stops. Then, the cause of the failure is investigated, which is followed by repair work and other processing. Finally, the system is recovered and restarted. The task remains stopped throughout this period.

In operation with Extended Mirror Mode, if a hardware failure occurs while the system is operating, the failure location (system) is disconnected, and the task is continued with the other system. The disconnection stops neither the system nor the task. After the disconnection, the task is performed only by one system, while a search for the cause of the failure is made so that the faulty component can be replaced quickly and operation can return to system mirrored operation quickly. Thus, Extended Mirror Mode can improve business continuity in the event of a hardware failure.

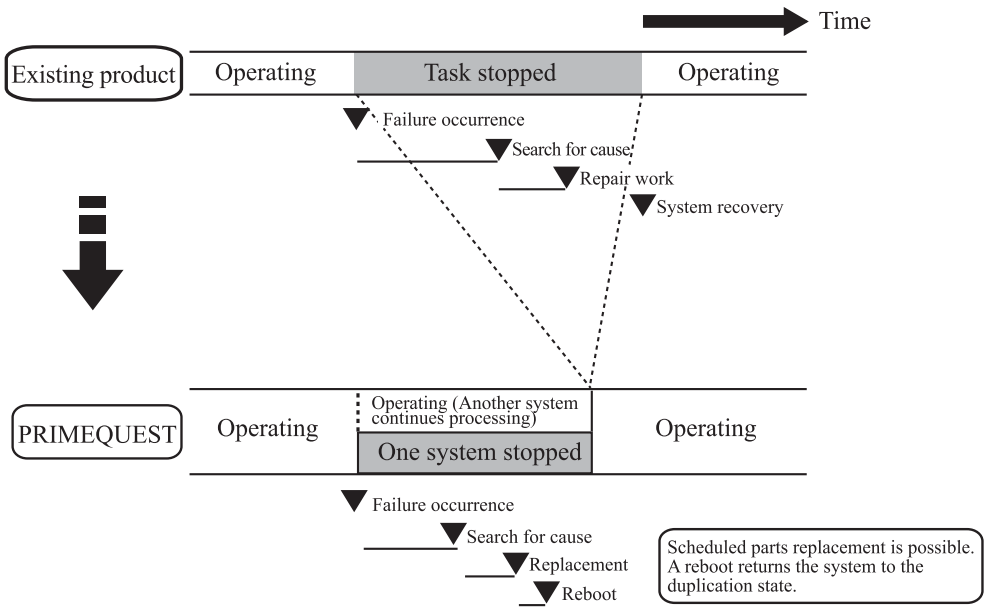


Figure 1.7 Improved business continuity with System Mirror Mode

Remarks: Even in Standard Mirror Mode, if either address bus should fail, the other address bus would allow service to continue without interruption.

1.2.1.5 Partition function

The partition function divides hardware resources in one cabinet into multiple systems, and each of the systems operates independently.

The PRIMEQUEST 520A/520 can be divided into up to two partitions or, when XPAR (option) has been installed, up to four partitions.

The PRIMEQUEST 420 can be divided into up to two partitions.

Remarks: In Extended Mirror Mode, the PRIMEQUEST 520A/520 can be configured into up to two partitions.

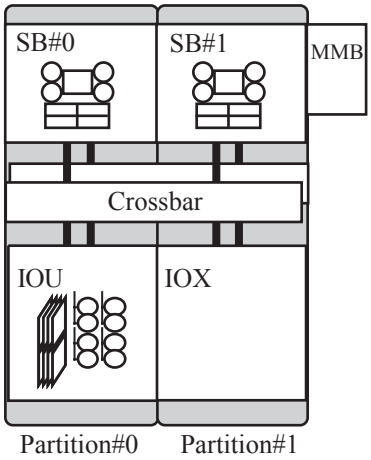


Figure 1.8 Partitions

The partition function provides the following advantages:

- XPAR (eXtended PARtitioning)

  XPAR (eXtended PARtitioning) enables each SB, IO Unit, and IOX as well as each split SB, IO Unit, and IOX to be configured in individual partitions. Split and unsplit SBs, IO Units, and IOXs can belong to the same partition.

  Split SBs, IO Units, or IOXs cannot be incorporated in a partition for which Extended Mirror Mode is set. XPAR is an option of the PRIMEQUEST 520A/520. An XPAR option is required for implementing XPAR.

- Flexible system construction

  The partition function can be used to construct multiple business systems in the same cabinet, thereby enabling flexible system construction.

- Robustness

  Hardware protects each partition from being affected by another partition. If a failure occurs in a partition, the other partitions are not affected by it.

Independent resource expansion in each partition

The PRIMEQUEST flexible I/O mechanism and partition function can be used to allocate resources effectively. For example, a system with the minimum configuration can be constructed first, and required hardware resources can then be added to the system at any time to improve its capabilities.

For details on the flexible I/O mechanism, see Section 1.2.1.6, "Flexible I/O mechanism."

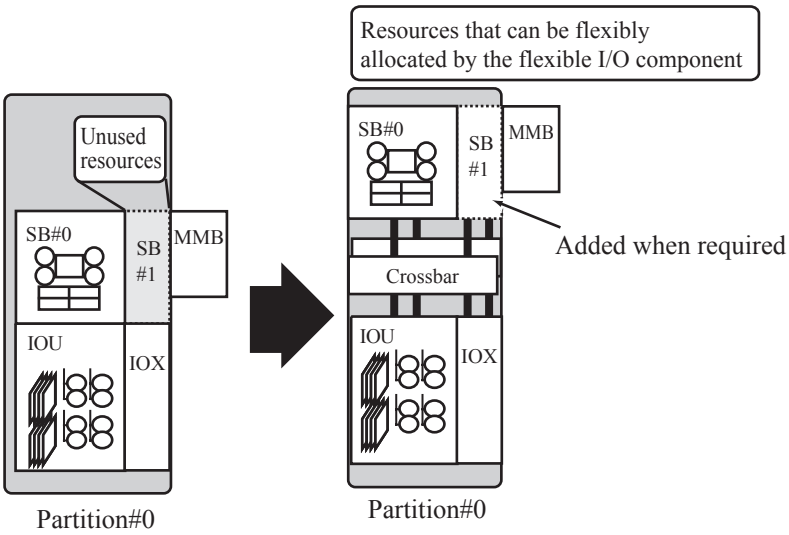
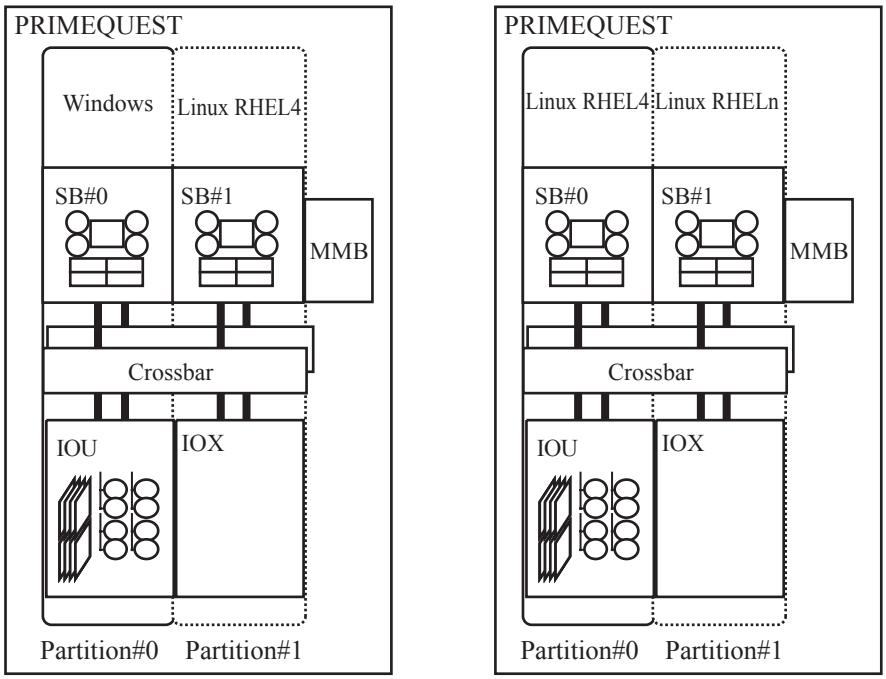


Figure 1.9 Independent resource expansion in each partition

Operation on different platforms

The partition function enables a mixture of systems in the same cabinet. Each partition in the cabinet can run a different OS or a different OS version.

This makes possible flexible development and capability improvements, so hardware resources can be used effectively.



RHEL: Red Hat Enterprise Linux

Figure 1.10 Operation on different platforms

1.2.1.6 Flexible I/O mechanism

The flexible I/O mechanism provides greater flexibility in connecting SBs and IO Units.

In operation without the flexible I/O mechanism, some resources remain unused because CPUs, memory, and I/O resources have fixed connections, as shown in Figure 1.11.

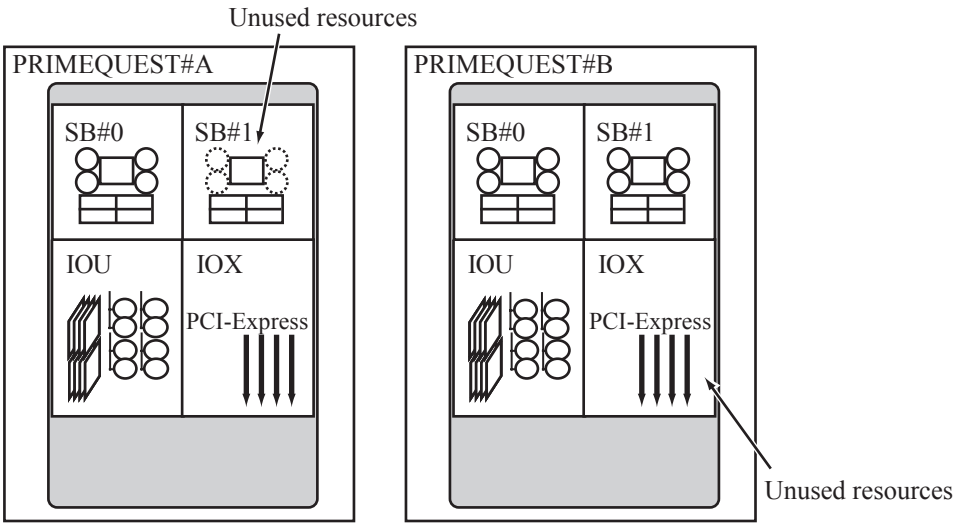


Figure 1.11 Connection between SBs and IO Units in a conventional model

In the above figure, Partition #A operation requires more I/O resources than CPU resources, and the CPU resources are underused. This state is typical in large-scale database processing and other processing that handles a large volume of data.

In the above figure, Partition #B operation requires more CPU resources than I/O resources, and the I/O resources are underused. This state is typical in processing that requires a high CPU processing capacity, such as scientific and technological computational processing.

As explained above, resources such as CPUs and memory cannot be freely allocated when SBs and IO Units have fixed connections, and this sometimes results in wasted resources.

The system configuration can be flexibly set and changed by using this mechanism and partitioning function.

The flexible I/O mechanism provides the following advantages:

- Free resource allocation

  If CPUs, memory, and I/O resources have fixed connections, resources may sometimes be used ineffectively. The flexible I/O mechanism enables CPUs, memory, and I/O resources to be allocated as required according to the operation mode to reduce costs.

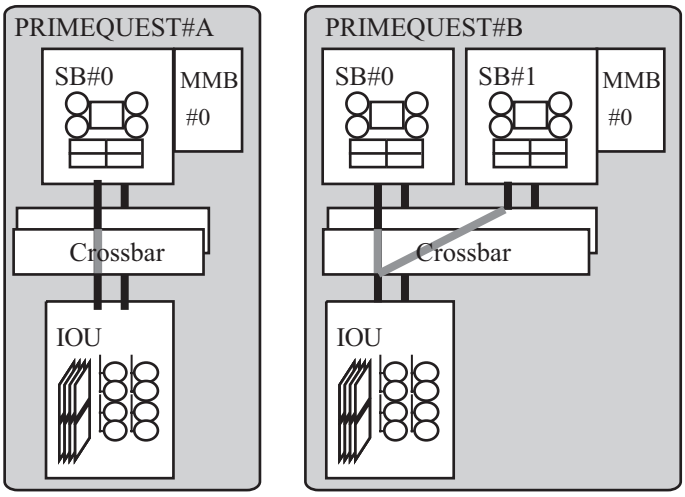
- Reduced mean time to repair (MTTR)

  If an SB fails, use of the flexible I/O mechanism and a reserved SB enables system recovery simply with a reboot, thereby significantly shortening the MTTR.

  The MTTR is the average time necessary for recovery of a faulty computer. It is the period in which the faulty computer does not operate before it is repaired. The shorter the MTTR, the higher the maintainability and the easier the repair work at the time of a failure.

Free resource allocation with the flexible I/O mechanism

The flexible I/O mechanism enables flexible use of hardware resources. Specifically, SBs and IO Units are freely connected by the flexible I/O mechanism in the crossbar, and the flexible I/O mechanism provides diversity and flexibility in such connections. Figure 1.12 shows an example of the flexible I/O mechanism and partition configuration.



Note: The lines in the crossbar represents connections between SBs and IO Units.

Figure 1.12 Example of the flexible I/O mechanism and partition configuration

Table 1.2 Relationship of the above partitions, SBs, and IO Units

Cabinet	SB	IO Unit
PRIMEQUEST#A	SB #0	IOU
PRIMEQUEST#B	SBs #0 and #1	IOU

In PRIMEQUEST#A, SB #0 and IOU are connected in a one-to-one basis. In PRIMEQUEST#B, SBs #0, and #1 are connected to IOU in a two-to-one basis by the flexible I/O mechanism. In PRIMEQUEST#B, the flexible I/O mechanism can establish connections in which the number of SBs does not match the number of IO Units.

As shown in this example, CPUs, memory, and I/O resources can be allocated independently because of the flexible I/O mechanism. Therefore, system resources can be used more effectively. That is, resources can be allocated as required according to the operation mode.

If an SB fails in a cabinet that has a reserved SB prepared, the current task can be continued with the reserved SB, which replaces the faulty SB. The reserved SB can be incorporated simply with a reboot of the partition. Therefore, the task can be restarted without waiting for the faulty SB to be repaired.

Reduced MTTR

Figure 1.13 shows an example where the MTTR is shortened by a reserved SB. Operation is switched to reserved SB #1 when a hardware failure occurs on SB #0. SB #0 is disconnected from Partition #0. The active IOU can be connected to the reserved SB #1 as is by the flexible I/O mechanism. This shortens the MTTR to the time necessary for rebooting only, resulting in a great improvement in system availability.

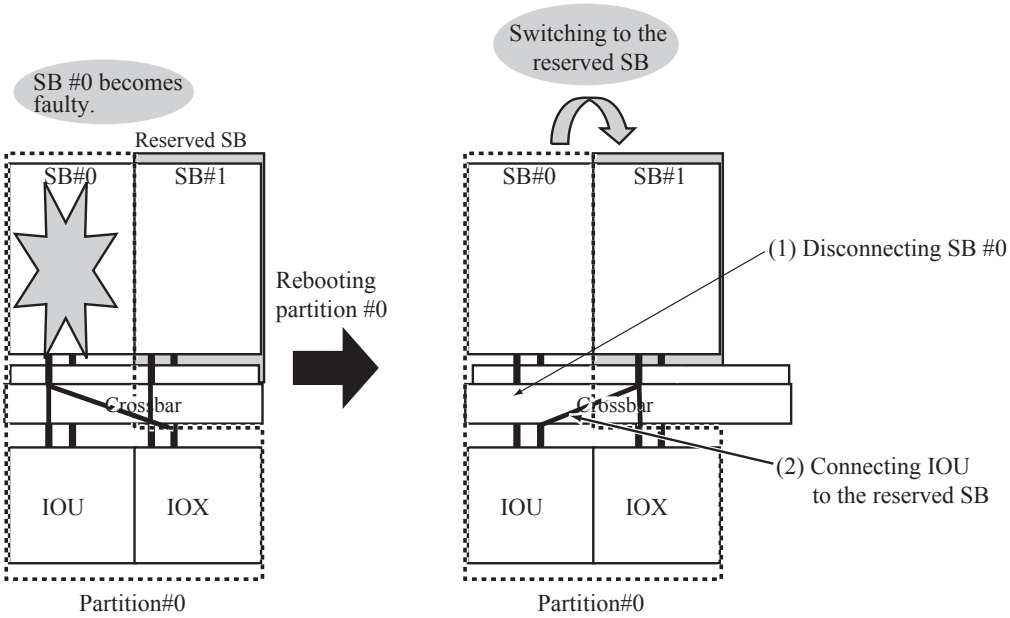


Figure 1.13 Reduced MTTR

Remarks:

- Fujitsu recommends specifying a reserved SB that has the same memory and CPU configurations as those of the SBs configured in the target partition. If the number of CPUs used increases during switching to the specified reserved SB, additional licenses are required.
- CPUs of the same type as those on the SBs composing the partition must be mounted on the reserved SB. If the SB has a different type of CPU mounted on it, it is not used as a reserved SB.
- An SB with 16 GB of memory mounted (4-GB DIMM x 4) cannot be mixed in the same partition with an SB having another memory configuration.

1.2.1.7 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.
- Board — For most board types, two or more boards can be mounted for a redundant configuration. Therefore, if one board fails, operation can continue with another board.
- 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, the crossbar, and other main components are duplicated (standard), and if one system fails, operation can continue with the other system.

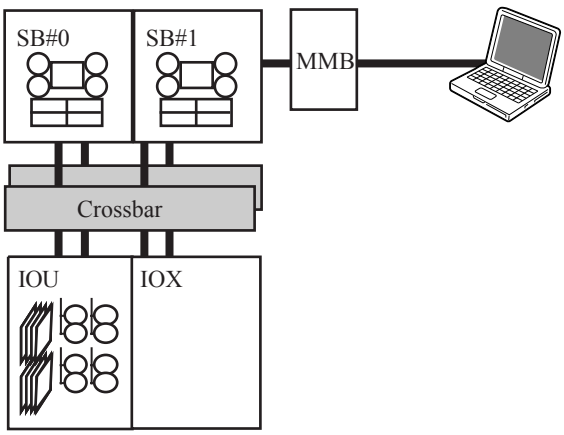


Figure 1.14 Duplication of the main components

Hot maintenance function

The hot maintenance function disconnects a unit from the system for maintenance work on the unit, without stopping the partition 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.
- Hot maintenance of a non-duplicated DVD is also possible.

Remarks: Maintenance on an SB, IO Unit, or IOX when the partition is stopped is possible.

1.2.1.8 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, partition management, network environment management, and power control.

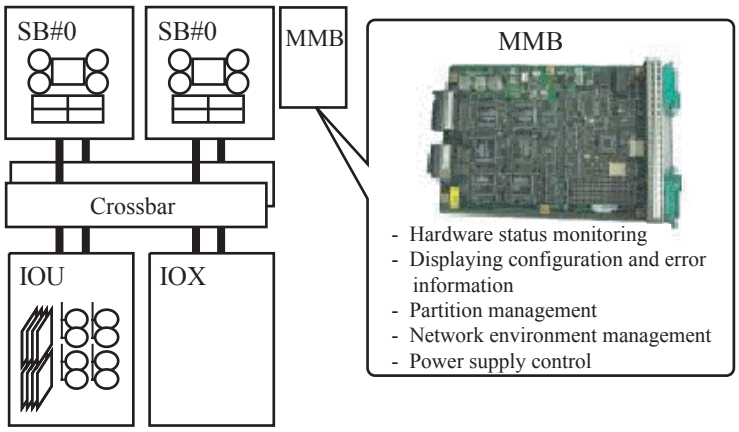


Figure 1.15 MMB

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 in individual partitions

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 1.3 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 partition or the cabinet, the current task can be handed to a standby partition 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.

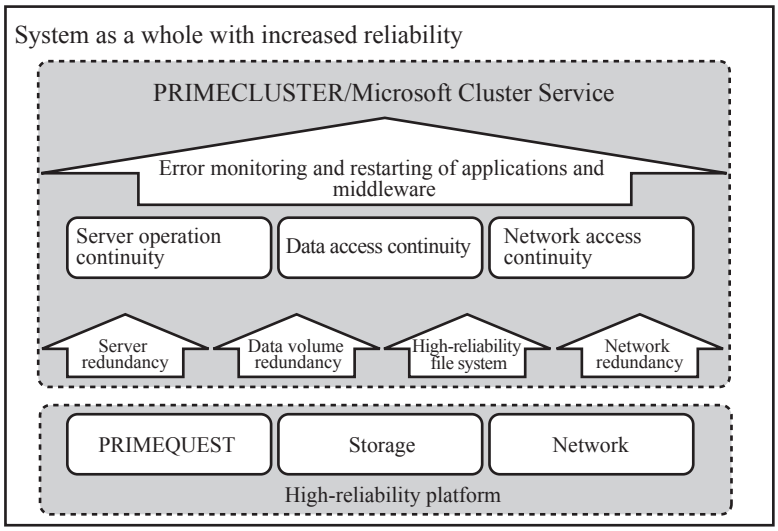


Figure 1.16 Clustering

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

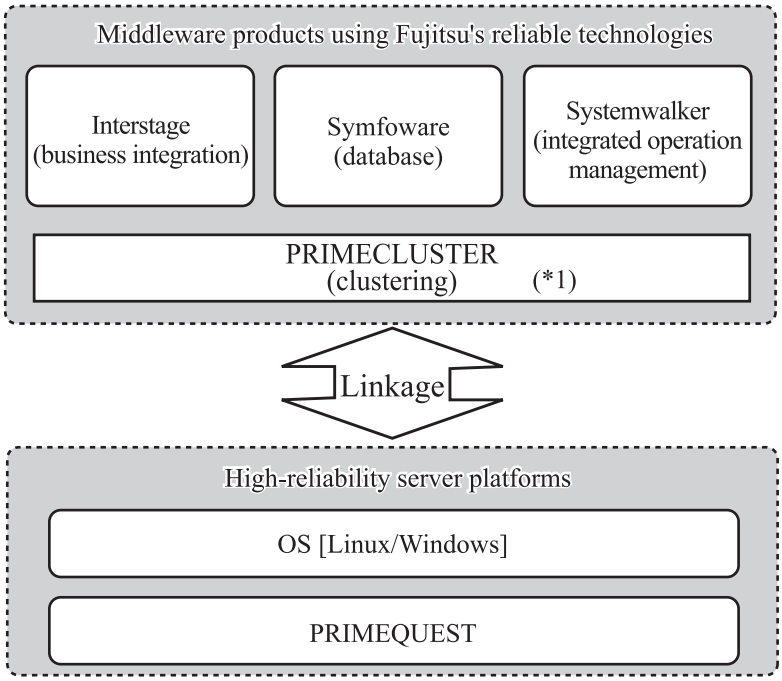


Figure 1.17 Fujitsu middleware products

*1 This is supported only for Linux.

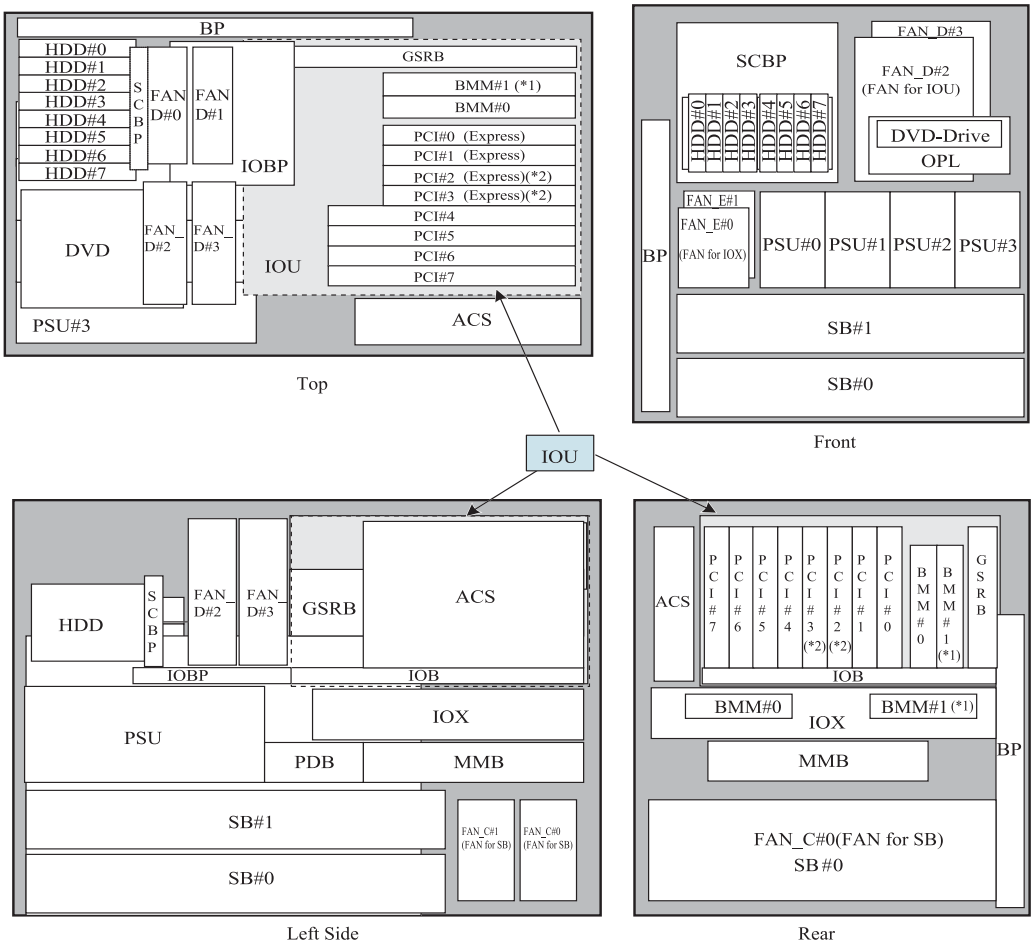
CHAPTER 2 General Hardware Configuration

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

- Main unit
  - Cabinet
  - System Board (SB)
  - IO Unit
  - HDD
  - IO Unit for PCI_Box connection (IOX)
  - Management Board (MMB)
  - BackPlane (BP)
  - Crossbar
  - Operator Panel (Board) (OPL)
  - Power Supply Unit (PSU)
  - PCI_Box
  - Expansion file unit

The configurations of PRIMEQUEST components are shown below.

Physical configuration (PRIMEQUEST 520A/520/420)



PDB :Power Distribution Board
PSU :Power Supply Unit
BP :Back Plane
SB :System Board
IOX :IO Unit for PCI_Box connection
MMB :Management Board
HDD :Hard Disk Drive

ACS :AC Section
OPL :OP-Panel
FAN :Fan
IOU :IO Unit
BMM :BMC Module
PCI :PCI slot

Figure 2.1 PRIMEQUEST 520A/520/420 physical configuration (full package)

Notes:

1. BMM #1 supports the PRIMEQUEST 520A/520 only.
2. PCI #2 and PCI #3 can be used with the PRIMEQUEST 520A/520 only.

Table 2.1 Types and maximum numbers of boards and components

Board/component		Maximum number	Remarks
SB		2	The minimum is one.
	CPU	8 (4 × 2 SB)	Each LSB requires at least one CPU.
	Memory (DIMM)	256 GB (32 × 2 SB) Maximum size	Each LSB requires at least four memory modules. The capacity is maximized when the maximum number of 4-GB memory modules is installed.
IO Unit		2 (including IOX)	The minimum is one.
	BMM	PRIMEQUEST 520A/520: 2 * Up to 4 partitions if XPAR (option) is installed	One BMM is required per partition.
		PRIMEQUEST 420: 2	
	PCI-X slot	4	-
	PCI Express slot	PRIMEQUEST 520A/520: 4 PRIMEQUEST 420: 2	
HDD		8	-
BP (including crossbar)		1	-
MMB		1	-
DVD drive		1	-
ACS		1	-
PSU		4	The minimum is three.
FAN	FAN_C	2	-
	FAN_D	4	-
	FAN_E	2	-
OPL		1	-
PCI_Box		1	Optional product
	PCIU	4/PCI_Box	-
	PEXU	2/PCI_Box	-

2.1 Cabinet

The main unit of the cabinet contains system boards (SB), IO Units, 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).

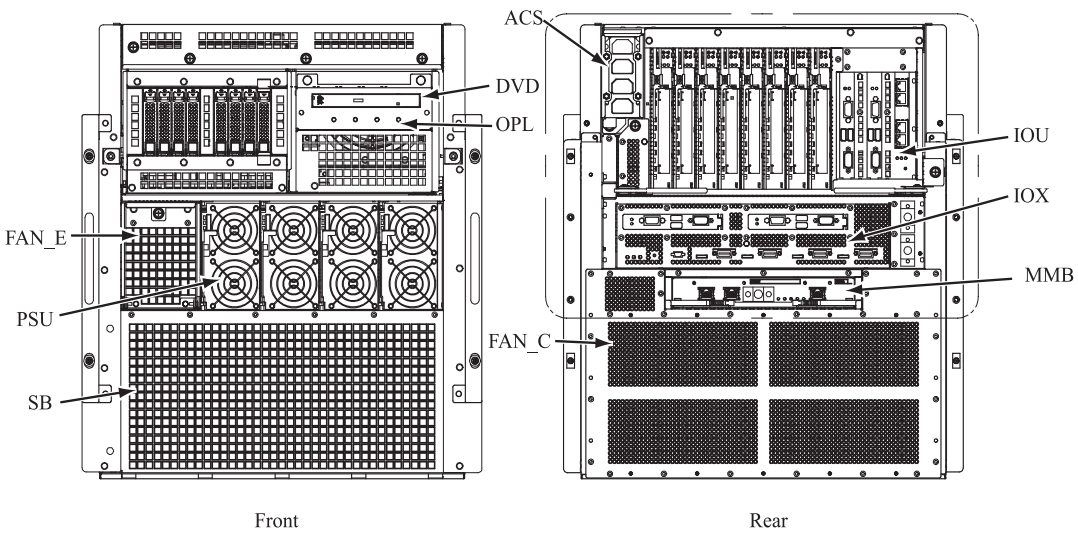


Figure 2.2 Main unit

2.2 System Board (SB)

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

- The main unit contains one to two SBs.
- The SB contains one to four CPUs and one set (four modules) to eight sets (32 modules) of memory.
- At least one CPU and one set of memory must be mounted on an SB that composes a partition.
- The PRIMEQUEST 520A/520 allows each SB to be split into two parts.

Remarks:

- To use split SBs, the XPAR option is required.
- For details on the partition function, see CHAPTER 7, "Partition."

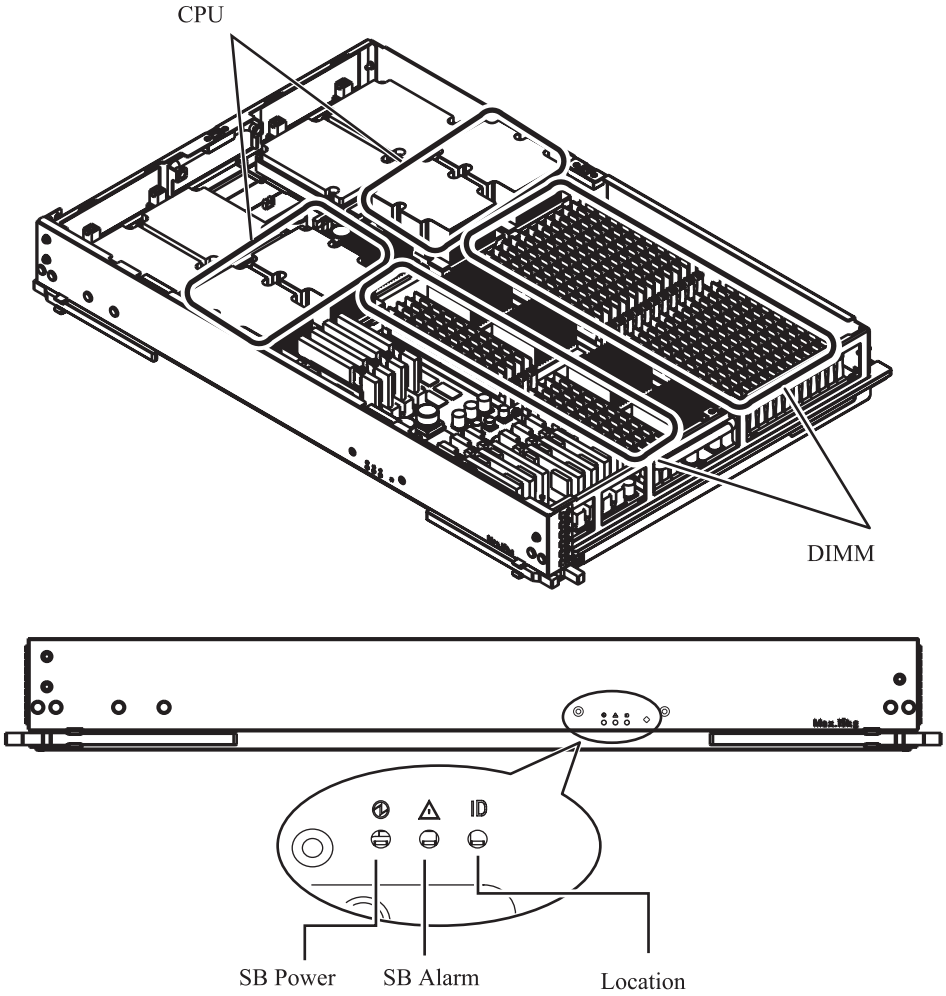


Figure 2.3 SB

Table 2.2 SB specifications

Item	Specifications
Number of CPUs	1 to 4
Number of memory modules	DDR2-DIMM × 4 to 32 (added in sets of four modules)

2.2.1 CPU

This section describes the CPU.

- The CPU is an IPF. 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 IPF 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 on one SB.

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 four 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 four DIMMs mounted in the same channel and the same slot on each LDX.
- One set (four modules) to eight sets (32 modules) can be mounted on one SB.

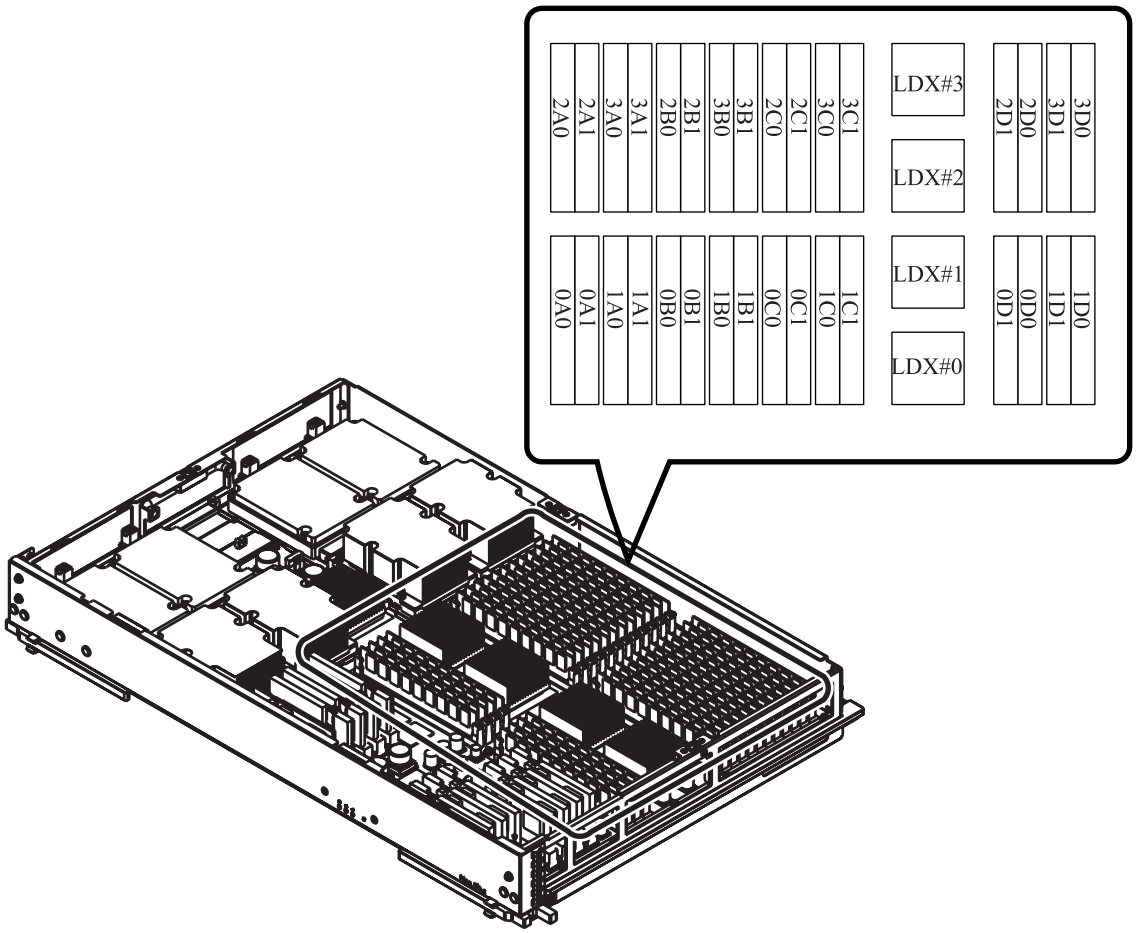


Figure 2.4 Memory (DIMM) slot numbers and locations

Notes on mounting DIMMs

Note the following points on mounting DIMMs:

- DIMMs must be added in sets of four DIMMs. One set is the minimum configuration unit.
- The four 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 (four modules). For this reason, it is recommended to mount two sets of DIMMs or more (eight modules or more).
- Different types of DIMMs can be mixed. For details, contact your Fujitsu certified service engineer.

2.2.2.1 Memory interleave

Memory interleave is technology that improves data throughput by distributing memory access to different DIMMs.

- The PRIMEQUEST 520A/520 allows intra-SB memory to be interleaved.

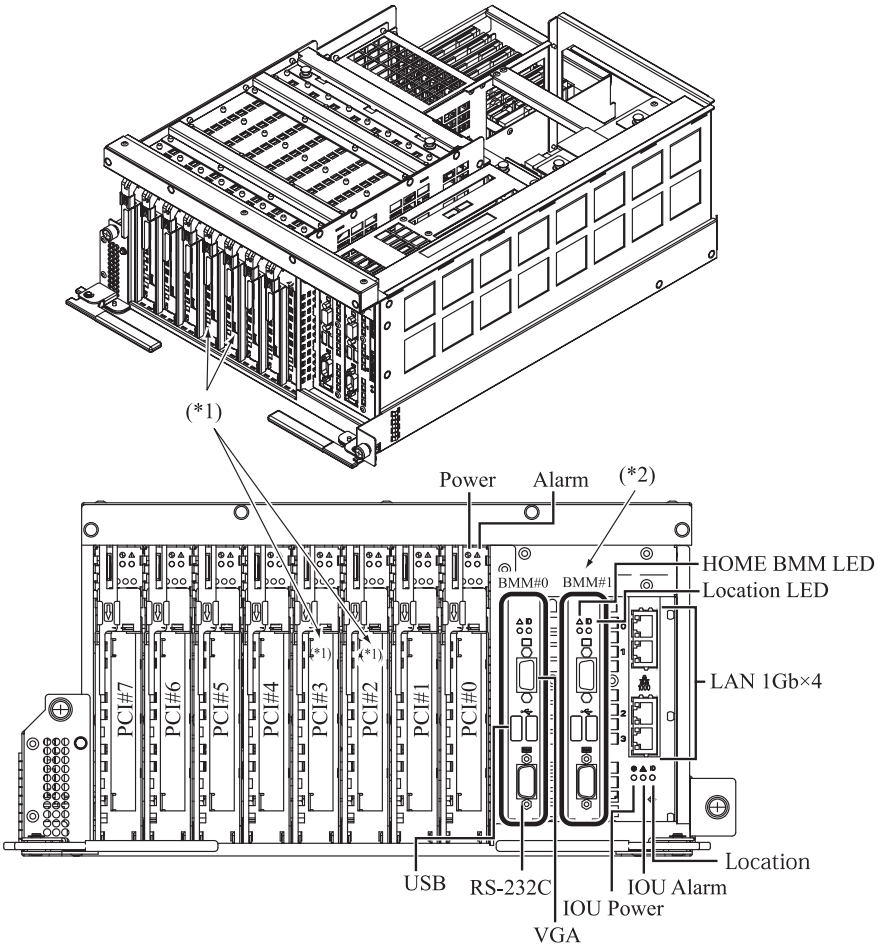
Note: To use the memory interleave technology, be sure to mount the maximum number of DIMMs using the same DIMM type and capacity.

2.3 IO Unit

This section provides an overview of the IO Unit. An I/O board, BMM (BMC module), HDD, PCI cassette, and PCI slot are mounted in the IO Unit, which provides instructions to devices and inputs and outputs data:

- The main unit contains one IO Units.
- The IO Unit contains two SAS controllers and two GbE controllers. The SAS controllers are connected to the HDDs mounted in the base cabinet.
- USB (USB 2.0) and COM ports are provided as the external interfaces.
- Up to two BMMs, which would be IO Unit control boards, can be mounted as options. However, only one BMM is valid at a time in a partition.
- The PRIMEQUEST 520A/520 allows each IO Unit to be split into two parts.

Remarks: To use split IO Units, the XPAR option is required.



*1：PCI#2 and PCI#3 can be used with the PRIMEQUEST 520A/520 only.
*2：BMM#1 can be used with the PRIMEQUEST 520A/520 only.

Figure 2.5 IO Unit (Full package)

Table 2.3 IO Unit external interfaces

External interface	Count	Remarks
USB	2/MMB	Mounted on optional BMM product
RS-232C	1/MMB	Mounted on optional BMM product
VGA	1/MMB	Mounted on optional BMM product
LAN	4	1 Gb

2.3.1 BMM

The BMM is a board that manages input and output of a partition. Each partition requires one BMM. Two BMMs can be mounted in an IO Unit and an IOX:

- Each partition requires one BMM.
- A VGA port, two USB ports, and a COM port are also mounted on the BMM.

2.3.2 PCI slot

The PCI slot is a slot in which a PCI-X or a PCI Express card is installed:

- Each IO Unit accommodates four PCI-X slots, with four PCI-Express slots on the PRIMEQUEST 520A/520 and two on the PRIMEQUEST 420.
- A PCI card is inserted into a PCI card cassette (adapter), and the cassette is inserted into the PCI slot.
- PCI cassettes are individually hot-swappable.

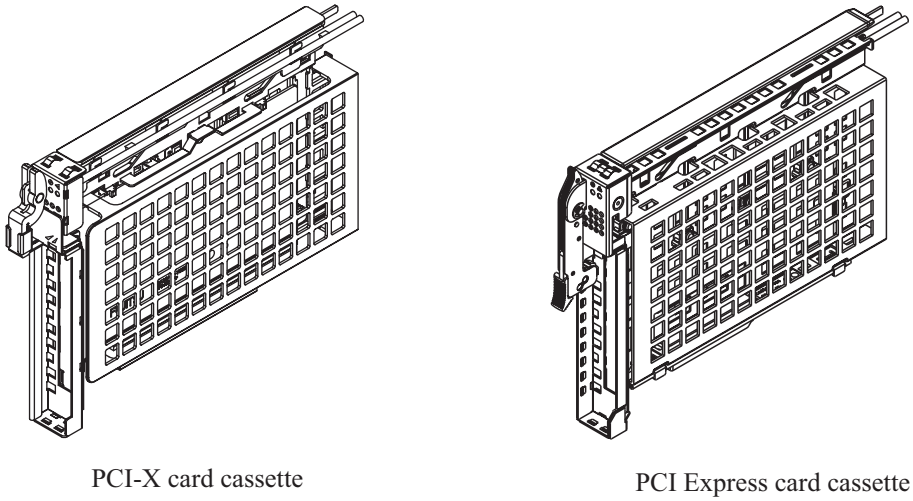


Figure 2.6 PCI card cassette

Table 2.4 PCI slot specifications (IO Unit)

PCI slot No.	Bus width	Operating clock	PCI bus operating voltage	Size
PCI Slot#4	64 bit	133 MHz	3.3 V	Long
PCI Slot#5	64 bit	133 MHz	3.3 V	Long
PCI Slot#6	64 bit	133 MHz	3.3 V	Long
PCI Slot#7	64 bit	133 MHz	3.3 V	Long

Remarks: The PCI slot is compatible with only the standard-type (full height) PCI card bracket.

Table 2.5 PCI-Express slot specifications (IO Unit)

PCI-Express slot No.	Number of lanes	Size
PCI-Express Slot #0	8	Short
PCI-Express Slot #1	8	Short
PCI-Express Slot #2 (*1)	8	Short
PCI-Express Slot #3 (*1)	8	Short

*1: PCI Express Slot #2 and PCI Express Slot #3 can be used with the PRIMEQUEST 520A/520 only.

2.3.3 Compatible PCI cards

PRIMEQUEST supports the following PCI cards.

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

Table 2.6 Compatible PCI-cards

Type	Function
SCSI card	Ultra320 SCSI (dual channel)
Fibre Channel card	Fibre Channel (2 Gbps × 1 port)
	Fibre Channel (4 Gbps × 1 port)
	Fibre Channel (4 Gbps × 2 ports)
Network-related PCI card	Gigabit Ethernet 1000Base-T (dual channel)
	Gigabit Ethernet 1000Base-SX (single channel)
	Gigabit Ethernet 1000Base-SX (dual channel)
	10Gigabit Ethernet 10GBase-SR (single channel) (*)

*: Supported only in the PRIMEQUEST 520A/520.

Table 2.7 Compatible PCI-Express cards

Type	Function
Fibre Channel card	PCI-Express Fibre Channel (4 Gbps × 1 port)
Fibre Channel card	PCI-Express Fibre Channel (4 Gbps × 2 port)

2.3.4 Mounting conditions of PCI cards and usable built-in I/Os

2.3.4.1 Mounting conditions of PCI cards

Table 2.8 contains the mounting conditions of PCI cards for the PRIMEQUEST series.

- A SCSI card can be mounted only in PCI Slot within an IO Unit and PCI slot#0 within a PCI Unit of a PCI_Box.
- SAN boot is enabled in all PCI slots in which FC cards are installed.
- Up to eight 10-gigabit LAN cards per partition can be mounted in arbitrary PCI slots in the PRIMEQUEST 520A/520.

Table 2.8 PCI card mounting conditions

	PCI slots in an IO Unit				PCI Express in an IO Unit				PCI slots in a PCI unit of a PCI_Box			PCI slots in a PCI Express unit of a PCI_Box (*3)	
Slot#	4	5	6	7	0	1	2(*2)	3(*2)	1(*1)	2	0(*1)	0	1
PCI Clock[MHz]	133	133	133	133					100	100	133		
Ultra320 SCSI Card (dual channel)	Y	Y	Y	Y							Y		
Fibre Channel Card (2-Gbps × 1 port)	Y	Y	Y	Y					Y	Y	Y		
[PCI-X Card-FC] Fibre Channel Card (4-Gbps × 1 port)	Y	Y	Y	Y					Y	Y	Y		
[PCI Express-FC] Fibre Channel Card (4-Gbps × 1 port)					Y	Y	Y	Y				Y	Y
[PCI-X Card-FC] Fibre Channel Card (4-Gbps × 2 port)	Y	Y	Y	Y					Y	Y	Y		
[PCI Express-FC] Fibre Channel Card (4-Gbps × 2 port)					Y	Y	Y	Y				Y	Y
1000Base-T LAN card (dual channel)	Y	Y(*3)	Y	Y(*3)					Y	Y			
1000Base-SX LAN card (single channel)	Y	Y(*3)	Y	Y(*3)					Y	Y			
1000Base-SX LAN card (dual channel)	Y	Y(*3)	Y	Y(*3)					Y	Y			
10GBase-SR LAN card (single channel)	Y	Y	Y	Y					Y	Y	Y		

Y: Slot in which the card can be mounted

*1 The PCI slot numbers of the PCI units in the PCI_Box are PCI Slot#1, PCI Slot#2, and PCI Slot#0 from the left when viewed from the front.

*2 PCI-Express slots #2 and #3 can be used with the PRIMEQUEST 520A/520 only.

*3 Only the PRIMEQUEST 520A/520 is supported.

Figure 2.7 shows the mounting conditions of SCSI cards.

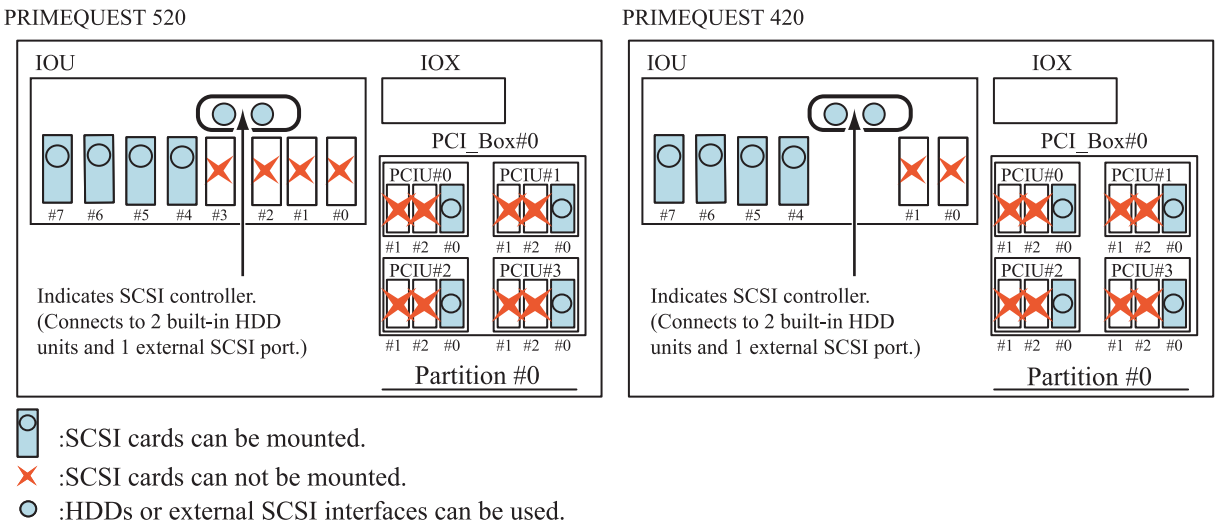


Figure 2.7 Mounting conditions of SCSI cards

Remarks: PCI Express Slot #2 and PCI Express Slot #3 can be used with the PRIMEQUEST 520A/520 only.

2.4 HDD

The HDD contains a 2.5-inch hot-pluggable serial attached Serial Attached SCSI (SAS) drive. The base cabinet can accommodate up to eight HDDs, which are connected to the SAS controllers in an IO Unit and controlled by the IO Unit.

2.5 IO Unit for PCI_Box Connection (IOX)

The IO Unit for PCI_Box connection (IOX) is a PRIMEQUEST 520A/520/420 dedicated board for I/O interface expansion that can be connected to an optional PCI_Box. If more slots are required than the number available, an additional IOX can be mounted. (Note that an IO Unit has four PCI-X slots, and four PCI-Express slots on the PRIMEQUEST 520A/520 and two PCI-Express slots on the PRIMEQUEST 420 mounted.) Once an additional IOX is mounted, a PCI_Box can be connected to the PCI_Box interface mounted on the IOX, thereby making additional slots available. The BMM can also be mounted to create partitions.

- Up to two BMMs, which would be IOX control boards, can be optionally mounted.
- Four serial bus PCI_Box interface ports are mounted.
- The PRIMEQUEST 520A/520 allows each IOX to be split into two parts.

Remarks: To use split IO Units, the XPAR option is required.

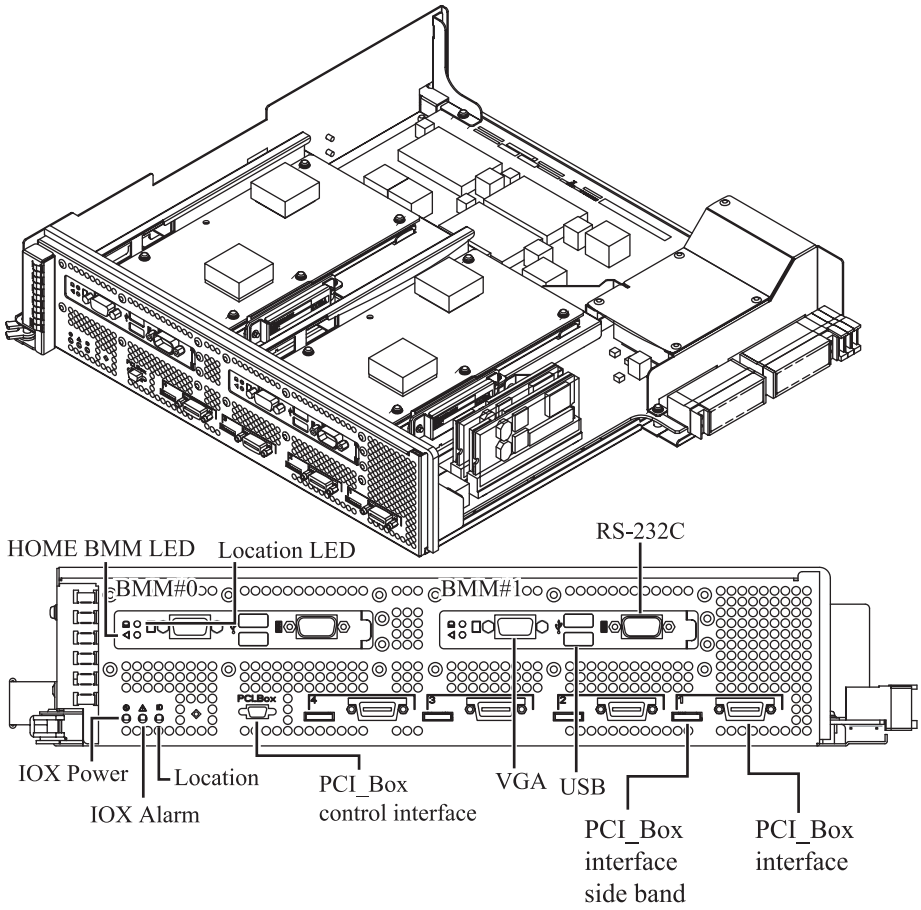


Figure 2.8 IOX (Full package)

Table 2.9 IOX unit external interfaces

External interface	Quantity	Remarks
USB	2/BMM	Mounted on an optional BMM
RS-232C	1/BMM	Mounted on an optional BMM
VGA	1/BMM	Mounted on an optional BMM
PCI_Box interface	4	Connected to an optional PCI_Box
PCI_Box interface side band	4	Connected to an optional PCI_Box

2.6 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 SBs. 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
- Remote operations including configuring settings, displaying the status, and various other operations from the MMB Web-UI
- Remote control of unit power and partition resets
- Monitoring sensors and other devices
- Partition configuration setup
- Setting the mirror mode
- Reset processing
- Scheduling operations
- REMCS operations
- Setting the Alarm E-mail

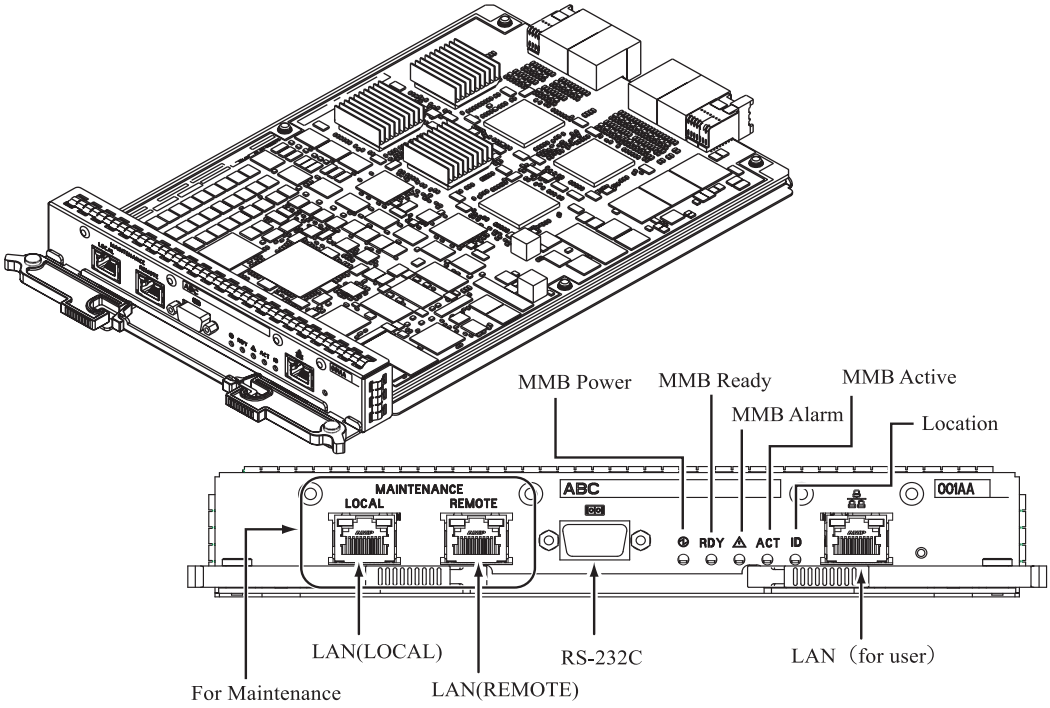


Figure 2.9 MMB

Table 2.10 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 - [LOCAL]: 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 partitions and 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.
However, when using an MMB with a part number (*) that ends in 004AB or lower, the following notes apply.

* Confirm the part number of the MMB with the Part Number item in the [MMB] window (Click [System] → [MMB].).

- If the Speed/Duplex setting of the REMCS port is set to other than "AUTO," use a LAN cable.
- If the Speed/Duplex setting of the device to connect to the REMCS port is set to other than "AUTO," set the Speed/Duplex setting of the REMCS port of the MMB to the same setting as that of the device.

2.7 Crossbar

The crossbar provides sharing of memory in up to two SBs and the symmetric multi-processing (SMP) system that supports 8 CPUs. It also provides the flexible I/O mechanism, which enables up to two SBs and up to two IO Units (including IOX) to be flexibly connected.

The crossbar is mounted on the backplane (BP).

2.8 Backplane (BP)

Connectors for SBs, IO Units, and other devices are mounted on the BP. A crossbar-control chip is also mounted.

2.9 Operator Panel (Board) (OPL)

The OPL contained in this unit has a four-LED display function.
Each LED status indicates the unit status and whether a failure has occurred.

A DVD drive is also mounted.

The OPL has the following functions:

- 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)

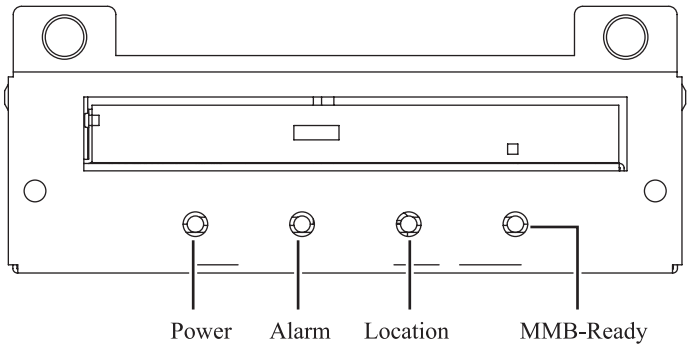


Figure 2.10 OPL

Table 2.11 LED display on the OPL

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
MMB-Ready-LED	Green	Off	Power to unit is off
		On	MMB is active
		Blinking	MMB is being initialized

2.10 Power Supply Unit (PSU)

The PSU converts AC input to 48 VDC. This unit implements a single-power feed 2+1 redundancy as standard. It supports a dual power feed option.

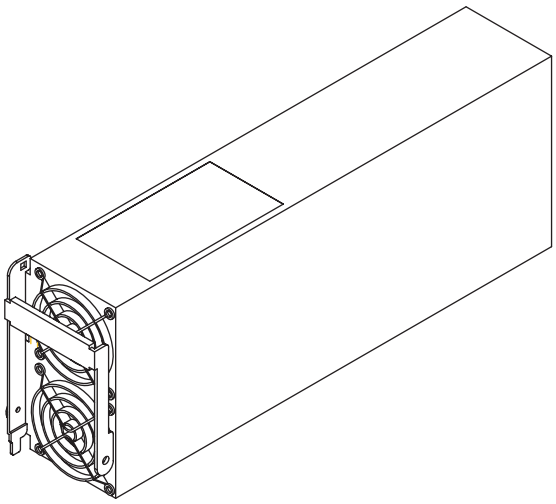


Figure 2.11 PSU exterior view

Table 2.12 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 2.13 Number of SBs and IO Units mounted and PSU configuration

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

2.11 PCI_Box

The PCI_Box is an optional unit for PCI slot expansion, and it is mounted in an Extended I/O cabinet:

- One PCI_Box can be mounted in one PRIMEQUEST 520A/520/420 main unit.
- Up to four PCI Slot Units (PCIUs) or up to four PCI Express Units (PEXUs) can be mounted in one PCI_Box. However, only the PRIMEQUEST 520A/520 can be connected to a PCI_Box in which a PEXU is mounted. Up to two PEXUs can be mounted in one PCI_Box.
  Up to three PCI-X cards can be mounted in one PCIU, and up to two PCI Express cards can be mounted in one PEXU.
- The PCI_Box is connected to the IOX and controlled by the IOX.
- A PCI-X card and PCI Express card are mounted in a PCI card cassette, which is then mounted in a PCI slot of a PCIU or PEXU.
- The PCI_Box is used while connected to a PCI_Box interface of an IOX and PCI_Box interface side band.
- Power supply units (IO_PSU) and fans use redundant configurations, increasing their reliability.

Remarks: The PEXU and PCI Express card are supported only for the PRIMEQUEST 520A/520.

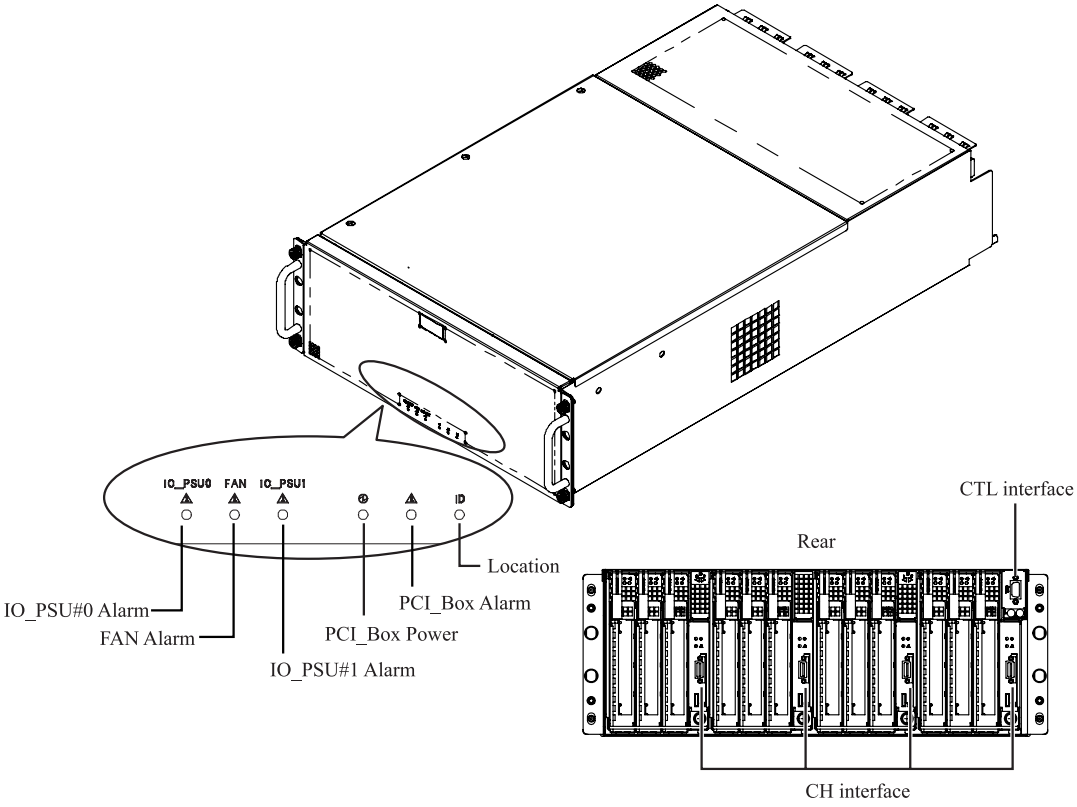


Figure 2.12 PCI_Box

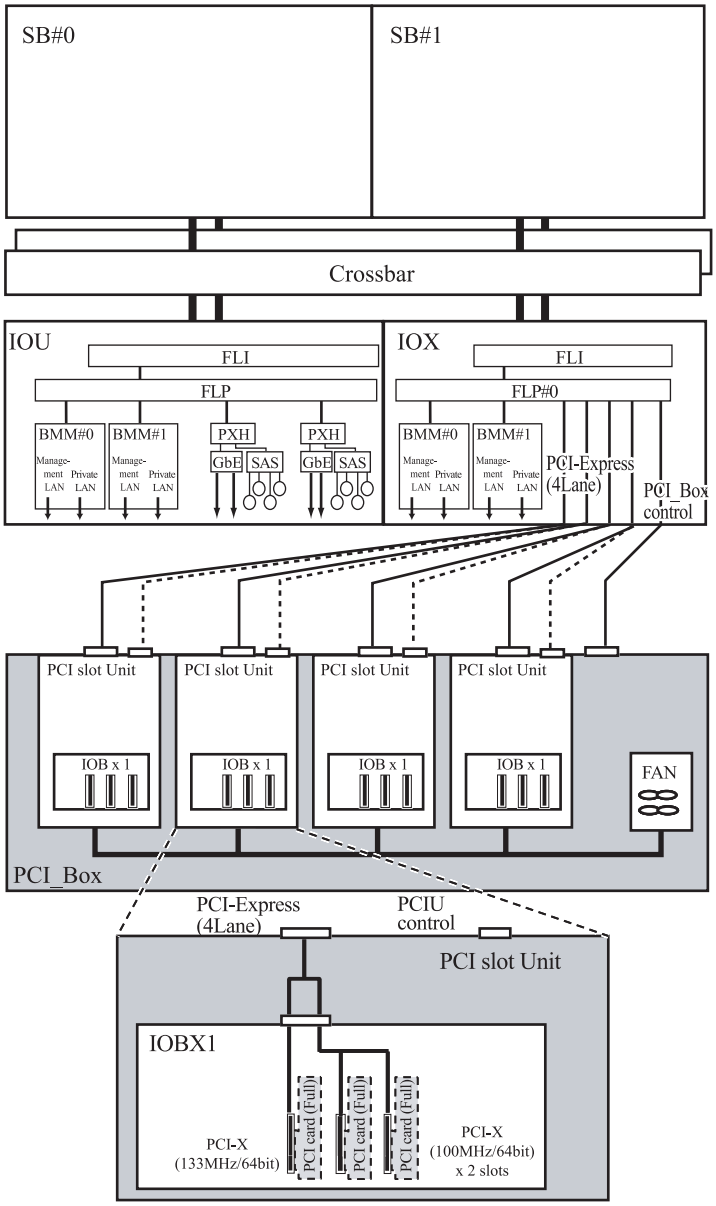
Table 2.14 PCI_Box configuration for PRIMEQUEST

Unit	Number of mounted units
PCI slot unit (PCIU)	4/PCI_Box
PCI slot (PCIU)	3/PCIU
PCI Express Unit (PEXU) (*1)	2/PCI_Box (*2)
PCI slot (PEXU) (*1)	2/PEXU

*1 The PEXU is supported only for the PRIMEQUEST 520A/520.

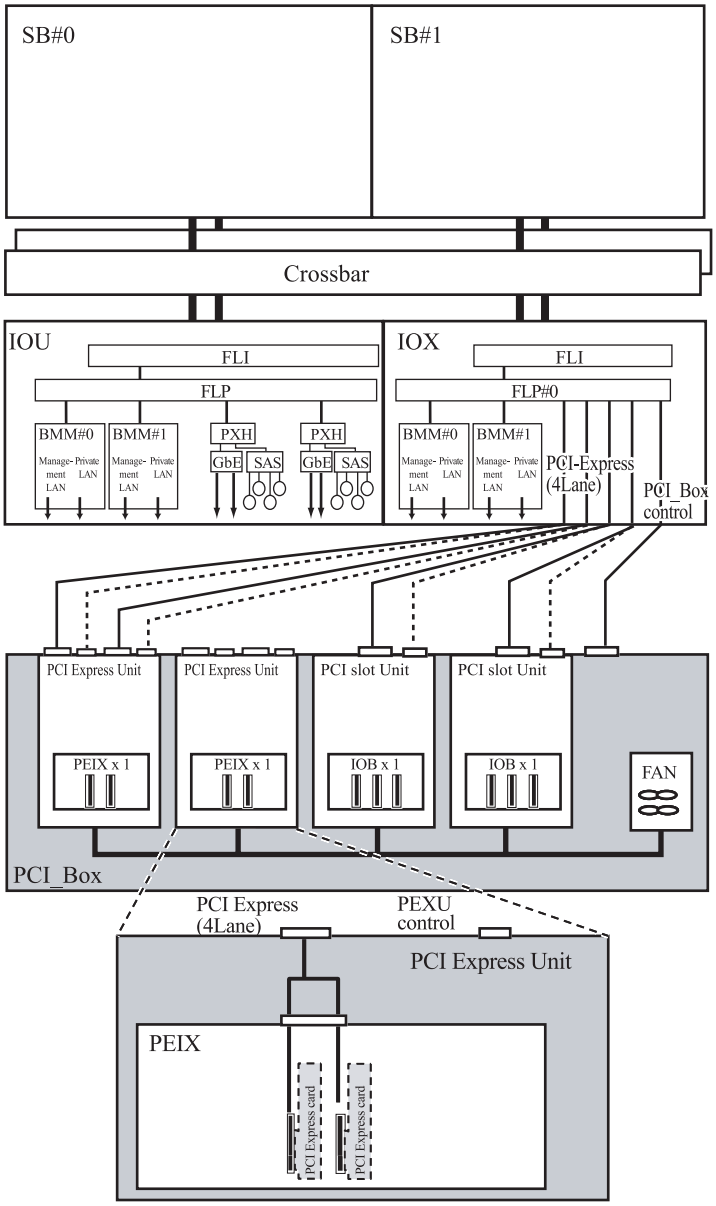
*2 Up to two units can be connected to the PRIMEQUEST 520A/520.

The following figure is a conceptual diagram of PCI_Box connections.



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

Figure 2.13 PCI_Box connections (only with PCIU)



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

Figure 2.14 PCI_Box connections (with a combination of PCIUs and PEXUs)

2.11.1 PCIU

Three PCI cards can be inserted in the PCIU. The PCIU contains three PCI card cassettes as standard.
The PCIU is connected to an IO Unit with a PCI_Box interface cable and a PCI_Box side band interface cable.

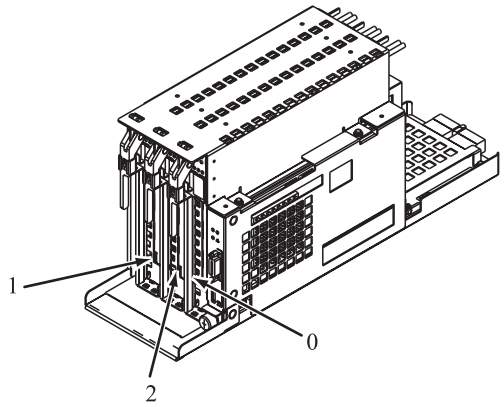


Figure 2.15 PCIU (with PCI card cassettes inserted)

2.11.1.1 PCI slot

The following table lists the types of PCI slots in the PCIU.

Only the standard-type (full height) PCI card bracket is supported.

Table 2.15 PCI slot specifications (PCI_Box)

PCI slot No.	Operating clock	Operating voltage	Bus width	Size
PCI slot#1	100 MHz	3.3 V	64 bit	Long-length
PCI slot#2	100 MHz	3.3 V	64 bit	Long-length
PCI slot#0	133 MHz	3.3 V	64 bit	Long-length

2.11.1.2 PCI card cassette

To use a PCI card in the PCIU, a PCI card cassette (adapter) is required. A PCI card is mounted in a PCI card cassette, which is then mounted in the PCIU.

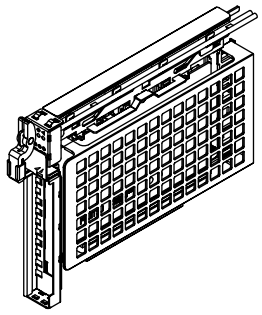


Figure 2.16 PCI card cassette (PCI_Box)

2.11.1.3 Compatible PCI cards

The PCI cards compatible with the PRIMEQUEST PCI_Box are the same as the PCI cards compatible with IO Units.

For details, see Section 2.3.3, "Compatible PCI cards."

2.11.2 PEXU

Up to two PCI Express cards can be mounted in one PEXU, in which two PCI card cassettes are inserted as standard equipment.

The PEXU is connected to an IOX in the PRIMEQUEST 520A/520 through two PCI Express cables and two PEXU control cables.

Notes:

- When mounting a PEXU, use PSA version 1.10 or later.
- The Linux OS versions that can be used in a partition containing a PEXU are as follows:
  - Red Hat® Enterprise Linux® AS (4.5 for Itanium)
  - Red Hat® Enterprise Linux® 5 (for Intel Itanium)
  - SUSE™ Linux Enterprise Server 10 for Itanium Processor Family Service Pack
  - Microsoft® Windows Server® 2003, Enterprise Edition for Itanium-based Systems Service Pack 1 or later
  - Microsoft® Windows Server® 2003, Datacenter Edition for Itanium-based Systems Service Pack 1 or later

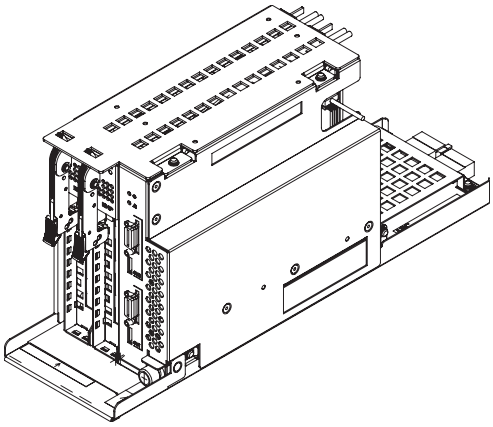


Figure 2.17 PEXU (with PEXU card cassettes inserted)

2.11.2.1 PCI slots in the PEXU

PCI Express cards are inserted in PCI slots in the PEXU.

- The PEXU has two PCI Express slots.
- A PCI Express card is mounted in a PCI card cassette (adapter), which is then inserted in a PCI Express slot.

- A PCI Express card can be inserted or removed while remaining mounted in a PCI card cassette.

2.11.2.2 Card cassettes in the PEXU

To use a PCI card in the PEXU, a PCI Express card cassette (adapter) is required.

A PCI Express card is mounted in a PCI Express card cassette, which is then inserted in one of the slots.

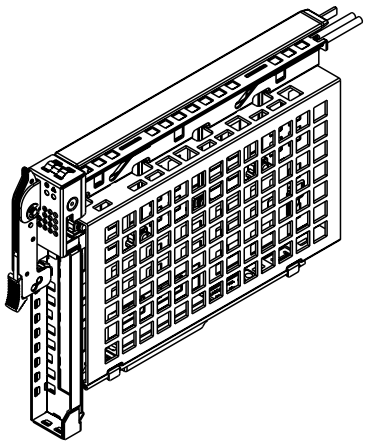


Figure 2.18 PCI Express card cassette

2.11.2.3 Compatible PCI cards

The following PCI Express cards are compatible with the PRIMEQUEST PCI_Box.

Table 2.16 Compatible PCI Express cards

Type	Function
Fibre Channel card	PCI Express Fibre channel (4Gbps × 1 port)
Fibre Channel card	PCI Express Fibre channel (4Gbps × 2 port)

2.11.3 Requirements on PCIU and PEXU connections and combinations

2.11.3.1 Requirements on connecting PCIUs and PEXUs to one IOX

One PCI-Box interface port and one PCI-Box interface side band port compose a set of ports for connecting a PCIU. An IOX is equipped with a total of four ports (Ports #0 to #3) for PCIU connections.

PCIU

A PCIU has one set of ports for connecting an IOX, and the set consists of one PCI Express port and one PCIU Control port. These ports that connect an IOX are connected with PCI_Box interface ports on the IOX.

PEXU

A PEXU has two sets of ports for connecting IOX, and each set consists of one PCI Express port and one PCIU Control port.

One PEXU must be connected to two PCI_Box interface ports on an IOX.

Make sure that PCI_Box interface ports #0 and #1 or PCI_Box interface ports #2 and #3 are paired for a connection to one PEXU. PCI_Box interface ports #1 and #2, #0 and #3, #0 and #2, or #1 and #3 cannot be paired for a connection to one PEXU.

2.11.3.2 IOX and PEXU connection examples

Examples of permitted and prohibited IOX and PEXU connection patterns are shown below.

- One PEXU is connected (pattern 1)
  Port #0 on the IOX is connected to Port #0 on the PEXU, and Port #1 on the IOX is connected to Port #1 on the PEXU.
- One PEXU is connected (pattern 2)
  Port #2 on the IOX is connected to Port #0 on the PEXU, and Port #3 on the IOX is connected to Port #1 on the PEXU.
- Two PEXUs are connected
  The PEXUs is connected as follows:
  Port #0 on the IOX to Port #0 on the first PEXU
  Port #1 on the IOX to Port #1 on the first PEXU
  Port #2 on the IOX to Port #0 on the second PEXU
  Port #3 on the IOX to Port #1 on the second PEXU

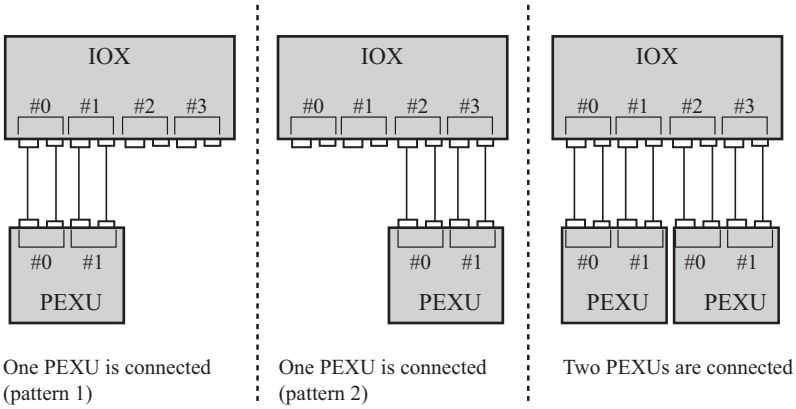


Figure 2.19 IOX and PEXU connections

Connections such as those shown below are prohibited.

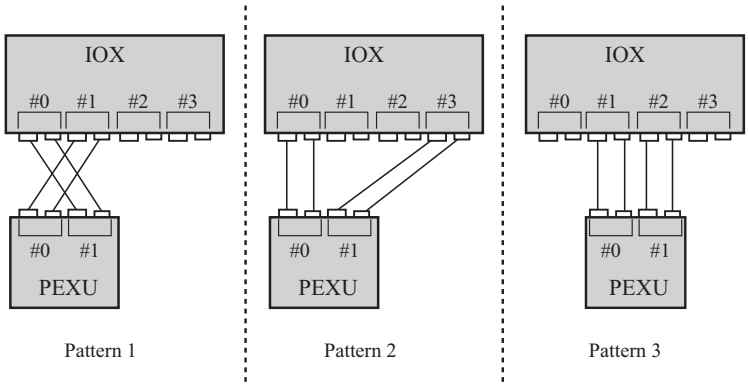


Figure 2.20 IOX and PEXU connections (prohibited patterns)

2.11.4 Connection requirements when XPAR partitioning is used

On a partitioned IOX, the ports for connecting PCIUs (Ports #0 to #3) can be freely assigned to LIOU_A and LIOU_B.

To connect a PEXU, however, the connection must be made in units of groups (group A consisting of Ports #0 and #1, or group B consisting of Ports #2 and #3), and the ports in one group (Ports #0 and #1, or Ports #2 and #3) must be assigned to one LIOU.

Also, the ports (two ports) connected with a PEXU cannot be assigned to a different partition.

Note: To assign ports connected with a PEXU to a different partition, the partition configuration must be changed after their cables are disconnected from the PEXU or after the PEXU is disconnected.

2.11.5 Requirements on a combination of PCIUs and PEXUs in one PCI_Box

PCIUs and PEXUs can be mounted together in one PCI_Box.

2.11.6 Requirements on a combination of PCIUs and PEXUs in one partition

There are some restrictions on mounting PCIUs and PEXUs together in one partition.

The MMB automatically recognizes the connection patterns of IOX, PCIUs, and PEXUs. If a recognized pattern does not correspond to any of the patterns in Tables 2.17 to 2.19, the MMB turns off the power to the relevant partition.

If the connection state of a partition corresponds to multiple connection patterns, the MMB chooses a pattern according to the following order of priority: pattern 1, pattern 2, and pattern 3.

For example, if PCIU ports #0 and #1 on IOU #0 are the only ports in a partition and these ports are connected to a PCIU, the partition corresponds to connection patterns 1, and 3. In this case, the MMB chooses pattern 1, which has the highest priority.

Remarks: If there were no PCIU or PEXU connected to the partition, the MMB would choose pattern 1.

Table 2.17 Combination of PCIUs and PEXUs in one partition (pattern 1)

PCIU port on IOX	IOU#0
#0	PCIU
#1	PCIU
#2	PCIU
#3	PCIU

Table 2.18 Combination of PCIUs and PEXUs in one partition (pattern 2)

PCIU port on IOX	IOU#0
#0	PEXU
#1	
#2	PEXU
#3	

Table 2.19 Combination of PCIUs and PEXUs in one partition (pattern 3)

PCIU port on IOX	IOU#0
#0	PCIU
#1	PCIU
#2	PEXU
#3	

2.12 Expansion File Unit

The expansion file unit is an external hardware option that is used to add hard disk drives. The expansion file unit, to which up to 14 hard disk drives (3.5-inch drives) can be connected, is used mounted in a rack.

- This expansion file unit is an optional device specially designed for the PRIMEQUEST 520A/520/420.
- The Ultra320 SCSI card is required for connecting the expansion file unit to PRIMEQUEST 520A/520/420 series machines.
- The file expansion unit is used after being mounted in any of the racks below by using the mounting rails that come with the unit as standard equipment. In such a case, a power distribution box (200 V) is required.
  - 19-inch global/36U base rack
  - 19-inch global/36U expansion rack
  - 19-inch global/40U base rack
  - 19-inch global/40U expansion rack

This expansion file unit has the following features:

- High speed data transfer
- Maximum number of disks to be mounted
- Redundant configuration
- Support for hot swapping
- Remote power control

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 software installed in each partition, 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 installed in a partition, 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.

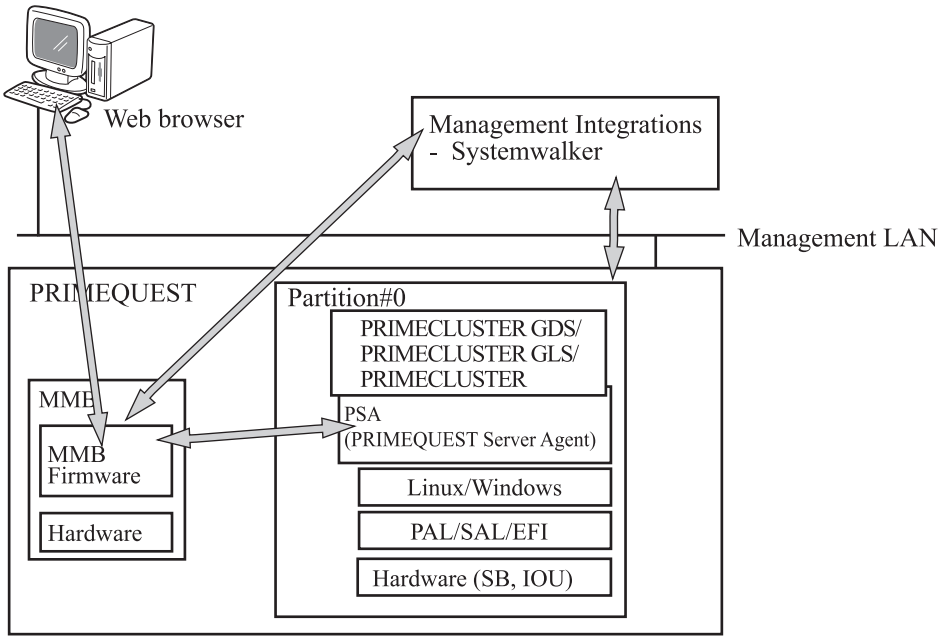


Figure 3.1 PRIMEQUEST software configuration

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 each IO Unit. BMC firmware manages the IO Unit 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 entire PRIMEQUEST hardware configuration, managing partition configurations, 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. The OSs are installed in individual partitions.

- 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. The software is installed and used in individual partitions.

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 in a partition.

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 in a partition as its operating system, install these programs as necessary. Cluster operations can be performed between cabinets and between partitions in a cabinet in the PRIMEQUEST system.

For details on cluster operations, see CHAPTER 8, "Clustering."

For details on PRIMECLUSTER, see the PRIMECLUSTER manuals.

3.4.2 PRIMECLUSTER GDS

When Linux is installed in a partition 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 in a partition 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 9, "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 each partition to be managed.

- 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.

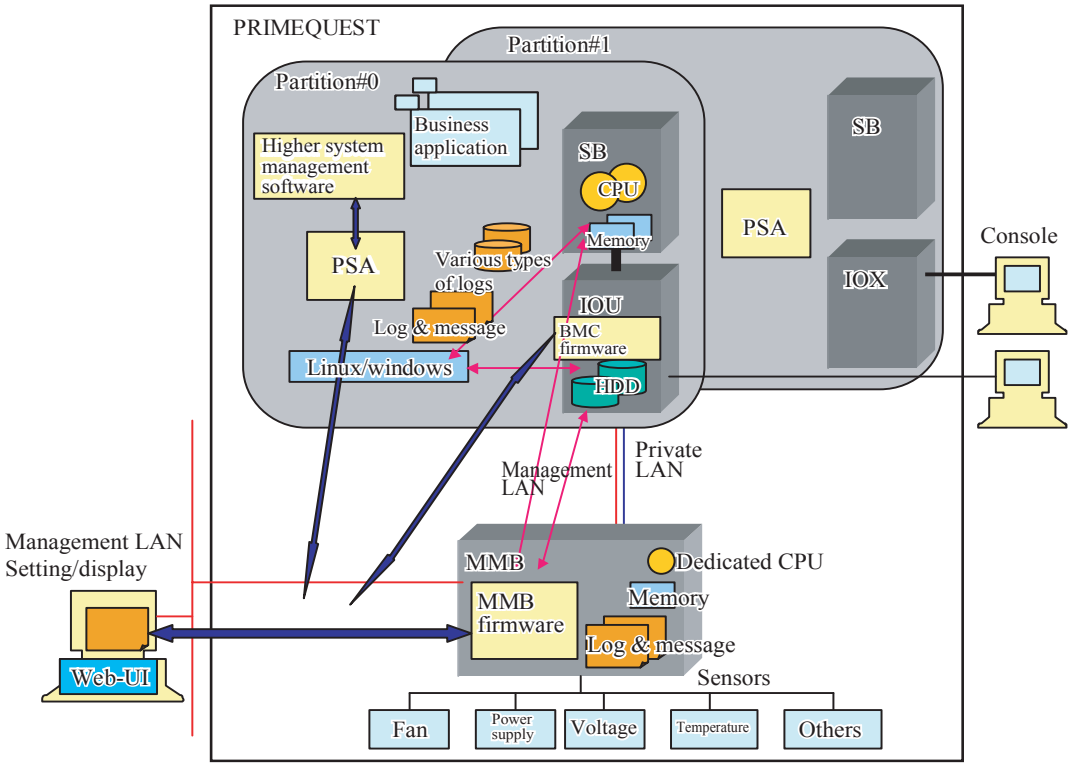


Figure 4.1 Basic mechanism

The MMB performs total management of the PRIMEQUEST hardware system.

The MMB manages the mounted components (SBs, IO Unit, chip sets, power supply units, fans, LANs, etc.) and partitions, and controls the system clock, management LAN and private LAN, and PCI_Box. The MMB sets various types of operations such as console redirection, and mirror mode. 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 mounted in the IO Unit for each partition and the PSA installed in each partition, by using the networks in the cabinet.

The PSA, software that runs on Linux or Windows, manages the partitions.

The PSA monitors the state of the SB and IO Unit forming the partition, and the configuration and state of the connected hardware units.

Monitoring targets include all of the hardware units incorporated in the partition such as the CPUs and memory of the SB, the HDDs and PCI slots of the IO Unit and, if the PCI_Box is connected to the IO Unit, the PCI card to which the PCI_Box is connected.

The PSA of each partition, operating independently of the PSAs in the other partitions, links with the MMB. The PSA 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.

Each partition has console information. This console information is input from and output to the COM port of the IO Unit incorporated in the partition.

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 of each partition to be operated from a general-purpose PC connected to the management LAN.

For details of the console redirection function, see Section 4.2.5, "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.

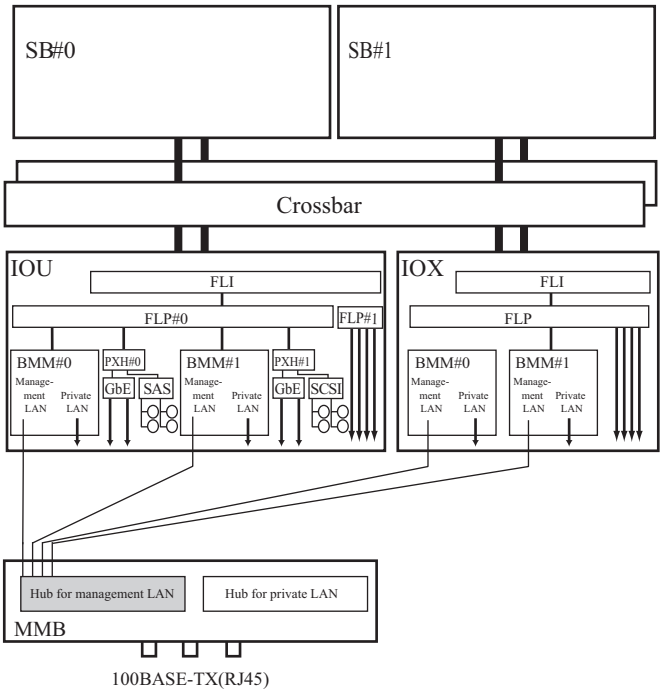
The private LAN is provided for internal control of the system.

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

Note: The management LAN and the business LAN must be configured on different subnets.

4.1.1.1 Management LAN

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



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

Figure 4.2 Management LAN

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

Applications of the management LAN are described below.

- Communication between the MMB and each partition
  Sends and receives information about Web linkage, SNMP linkage, and REMCS linkage between the MMB and the PSA managing each partition.
- Operator console (Web-UI, SSH, etc.)
  Connected as the console of the MMB and each partition
- 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:

- For the PRIMEQUEST 520/420
  The REMCS port also has the Auto MDI-X function. If Speed/Duplex is set to a value other than "AUTO," the appropriate cable for use is the same as that for the USER port.
- For the PRIMEQUEST 520A
  - 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.)

VLAN function of the management LAN

The hub for the management LAN can be found on the MMB. On this hub, cables for each partition network, user port, Fujitsu certified service engineer's port, and REMCS port are concentrated and the Virtual LAN (VLAN) function is mounted.

In a VLAN, the respective ports of one switching hub are logically divided into groups, each functioning as an independent LAN.

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

- VLAN mode of the management LAN

(1) Inter-partition communication connection mode (No VLAN mode)
This mode enables communication between ports of all partitions. 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 partition ports are not supported. Y in Table 4.1 indicates that communication between the corresponding ports is available.

Table 4.1 Inter-partition communication connection mode (No VLAN mode)

No VLAN mode	User port	CE maintenance port	REMCS port	MMB	Partition 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
Partition ports	Y	N	N	Y	Y

Y: Communication possible, N: Communication impossible

(2) Inter-partition communication disconnect mode (VLAN mode)
This mode disconnects communication with partitions. However, communication between the user port and partition ports and between the MMB and partition ports are enabled in the cabinet. Y in Table 4.2 indicates that communication between the corresponding ports is available.

Table 4.2 Inter-partition communication disconnection mode (VLAN mode)

VLAN mode	User port	Fujitsu certified service engineer's port	REMCS port	MMB	Partition 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
Partition ports	Y	N	N	Y	N

Y: Communication possible, N: Communication impossible

(3) All-partition communication disconnection mode (Port disable mode)

This mode is provided for enhancing security.

This mode disconnects communication of the management LAN with ports of all partitions. Intra-cabinet communication between an MMB and port in the same partition is not possible. Y in Table 4.3 indicates that communication between the corresponding ports is available.

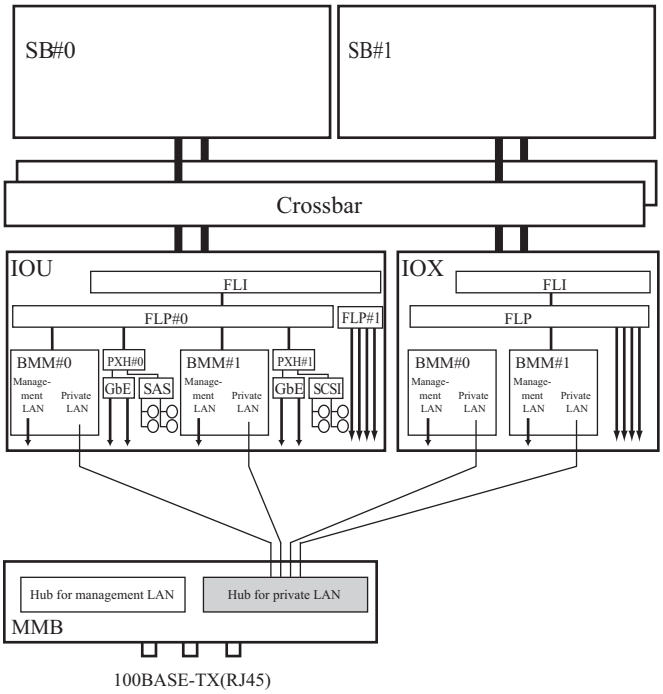
Table 4.3 All-partition communication disconnection mode (Port disable mode)

Port disable mode	User port	Fujitsu certified service engineer's port	REMCS port	MMB	Partition 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
Partition ports	N	N	N	N	N

Y: Communication possible, N: Communication impossible

4.1.1.2 Private LAN

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



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

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 an IO Unit 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 an IO Unit.
- The 1000BASE-SX card and other cards are mounted in PCI slots in the IO Unit or a PCI_Box, 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.

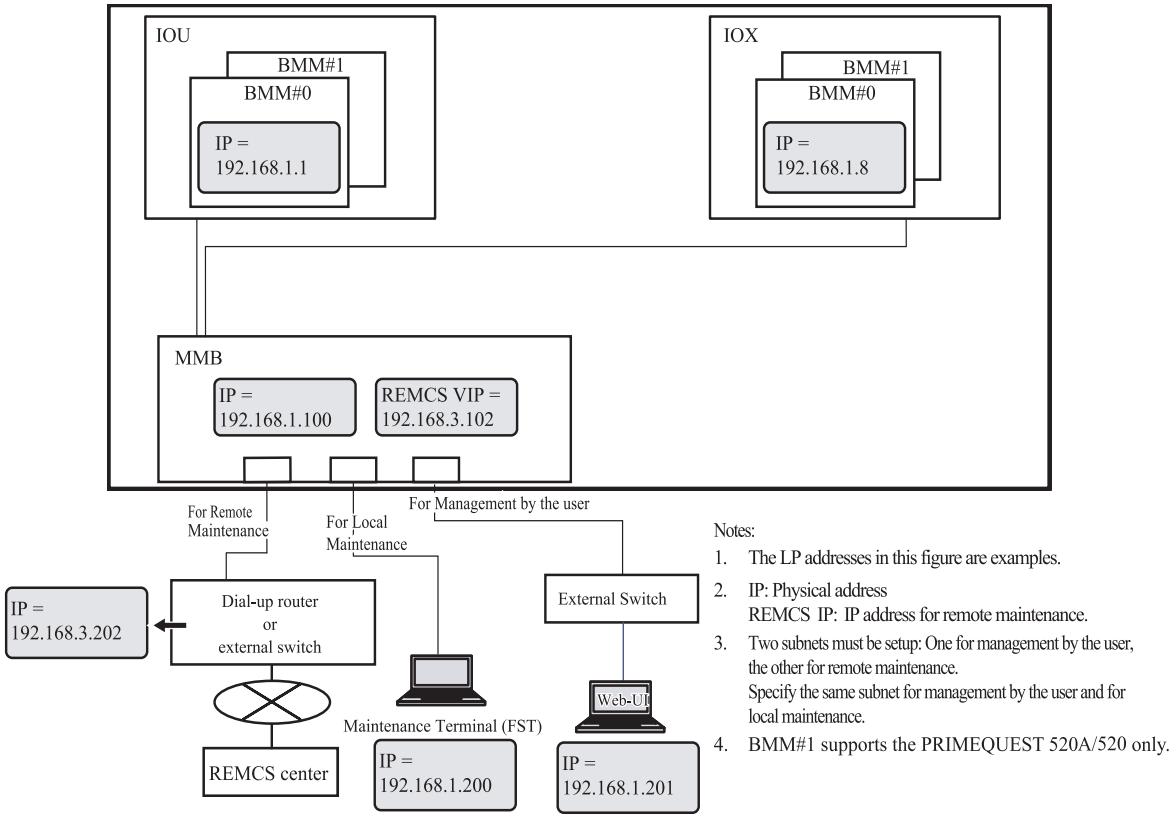


Figure 4.4 Management LAN network configuration and IP addresses

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 on a partition

  An IP address must be specified for the operating system in each partition. In physical terms, the IP address corresponds to a NIC in an IOU or IOX.

  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 with the PRIMEQUEST 520A 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.

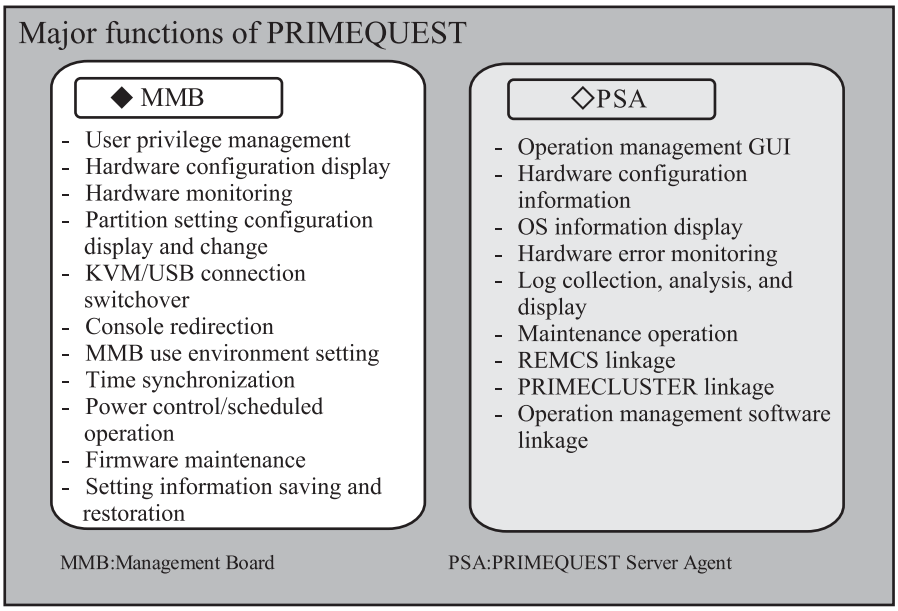


Figure 4.5 Main MMB functions

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 partitions.
- 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 such as SBs and IO Units including fans and power supply units mounted in the PRIMEQUEST system and units in the components such as CPUs, memory, and HDDs.

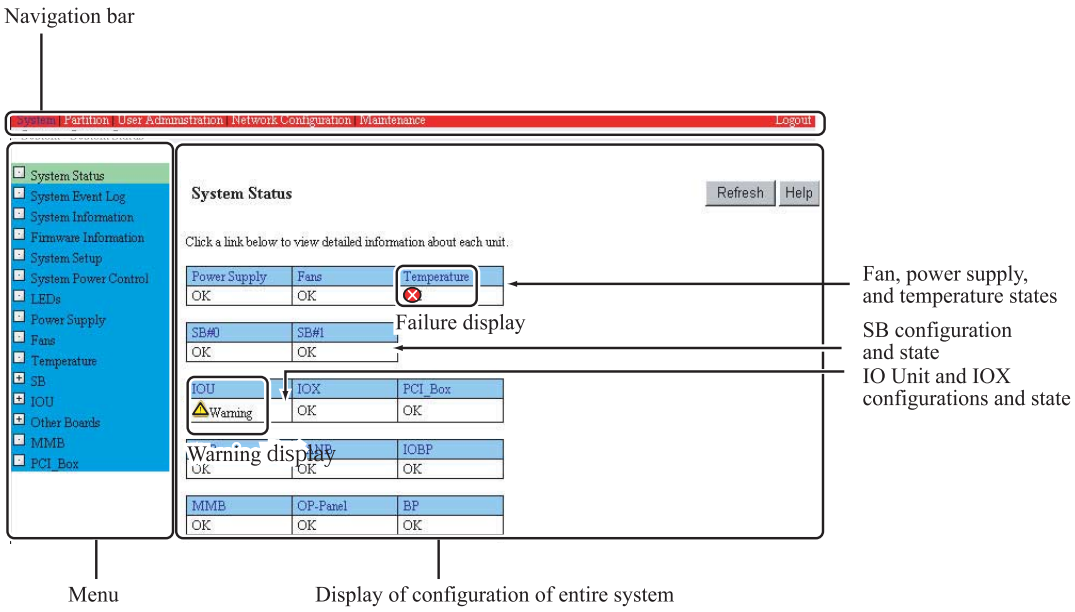


Figure 4.6 Example of the configuration display window

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 tracing the link in the display item field.

Also, the configuration can be displayed in units of partitions. Detailed information on the configuration and states of the SBs and IO Units allocated to each partition are displayed in this case.

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. If a failure or an error is detected, the required corrective action must be taken according to the location where the failure or error was detected, and the elements of the system configuration in place, such as mirror mode, redundant configuration, and the like. 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.

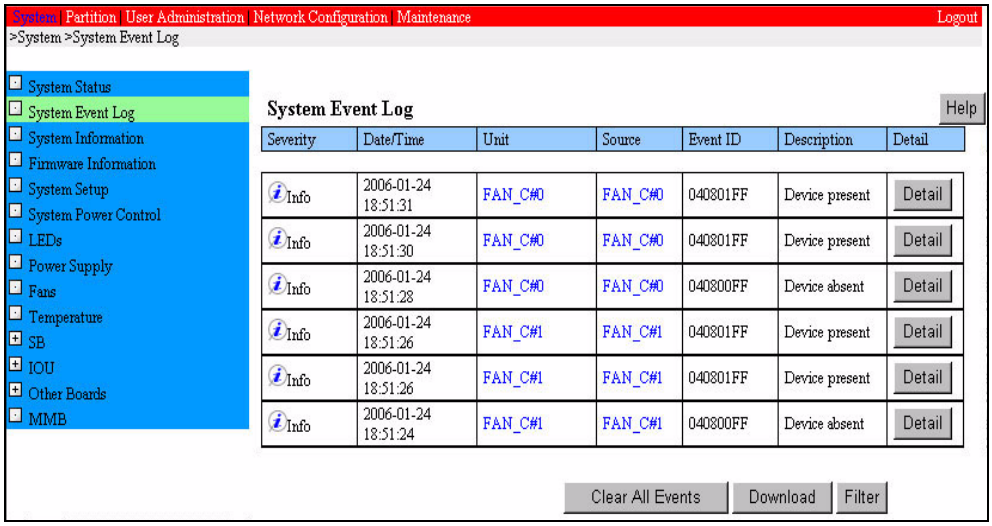


Figure 4.7 Example of the system event log display window

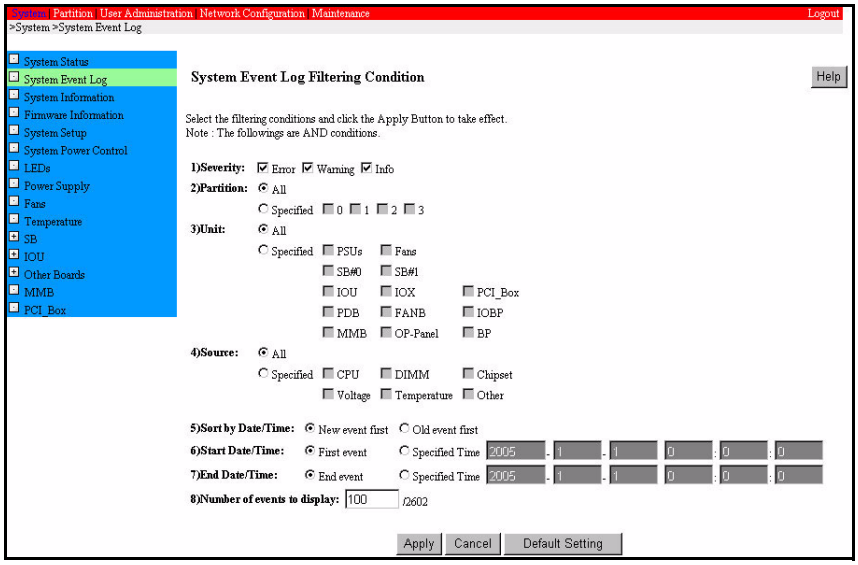


Figure 4.8 Example of the filter condition setting window

4.2.4 Partition settings and configuration display

This function is used to set a partition and display its configuration. Partition settings are used to divide mounted hardware resources and to determine the units of control and management.

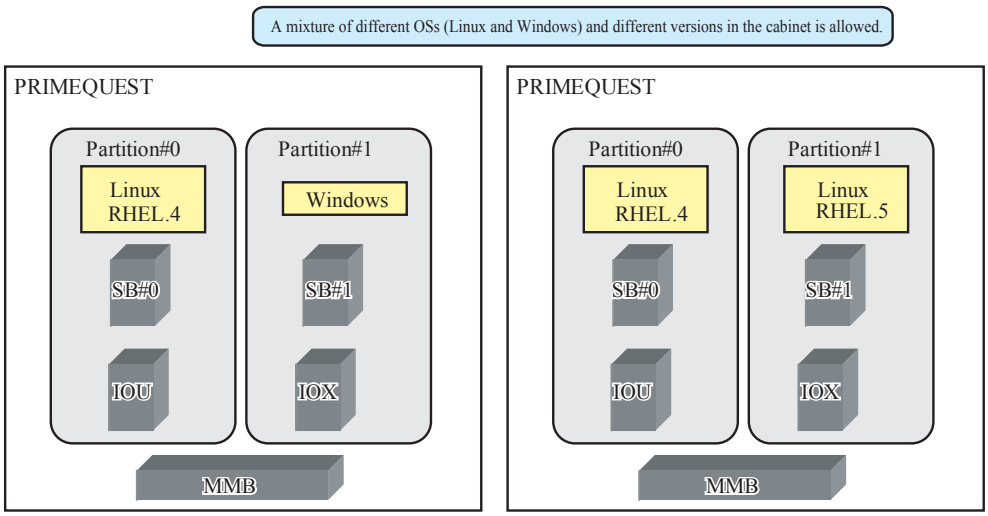


Figure 4.9 Example in which different OSs and OSs of different versions are mounted in respective partitions

A partition is set by specifying a combination of SBs and IO Units. The partition configuration information can be displayed in detail, showing incorporated SBs and IO Units and their configuration elements. As with the settings, the partition can be changed by changing the combination of the SBs and IO Units.

An example of the display of the partition configuration screen is shown in Figure 4.10.

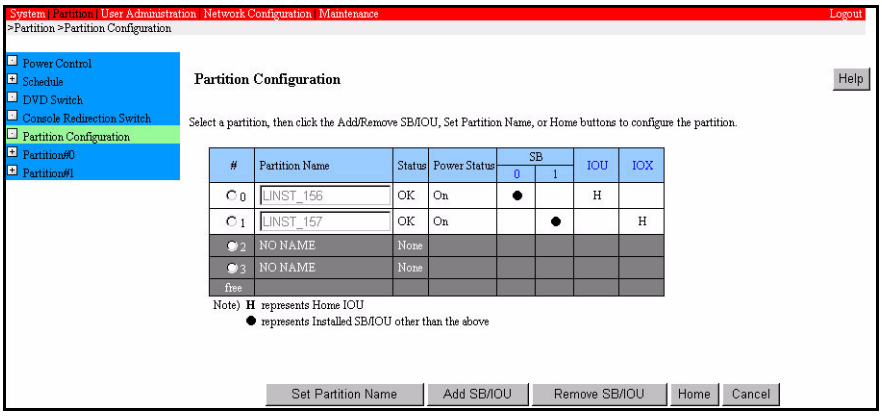


Figure 4.10 Example of display of the partition configuration window

4.2.5 Console redirection

This function sets I/O destinations of each partition's console information through a management LAN.

The two methods below are available for OS input and output in each partition. These methods can be set independently in each partition:

- Using the COM port of the IO Unit incorporated in each partition
- Using the management LAN through the MMB

Figure 4.11 shows the flow of console information sent when the information is redirected to the management LAN.

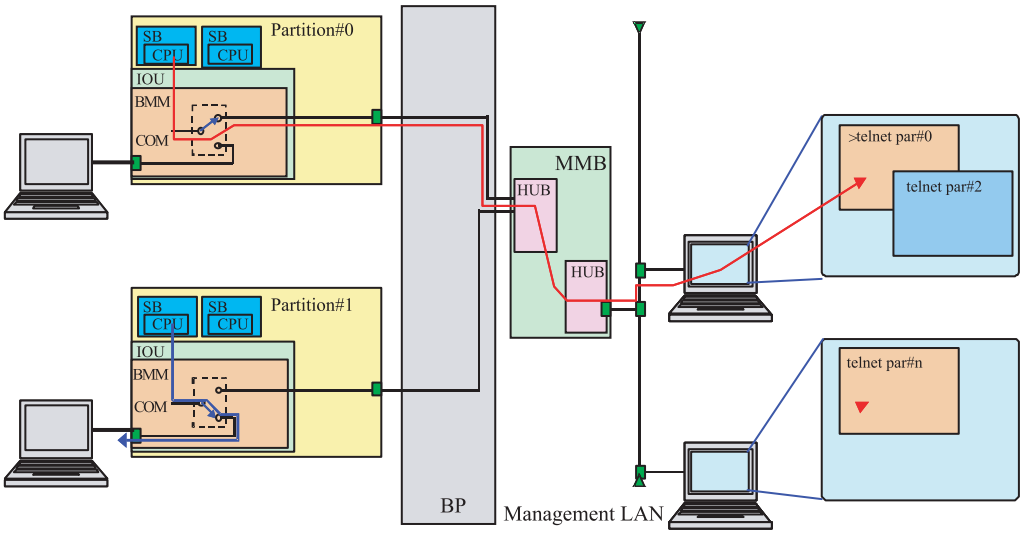


Figure 4.11 Flow of redirected information

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

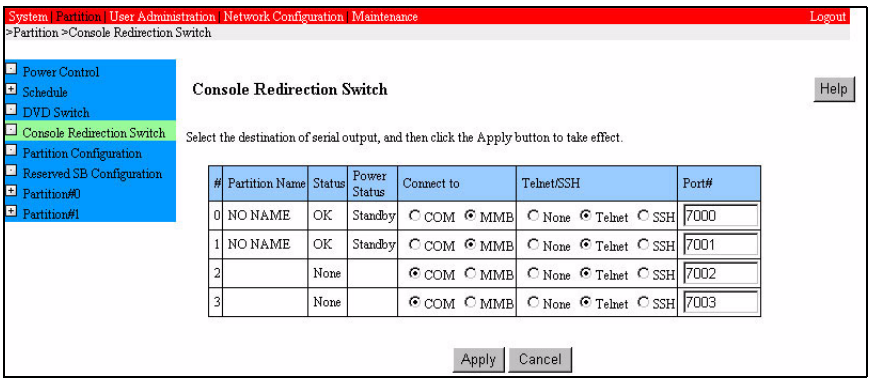


Figure 4.12 [Console Redirection Switch] window

4.2.6 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.

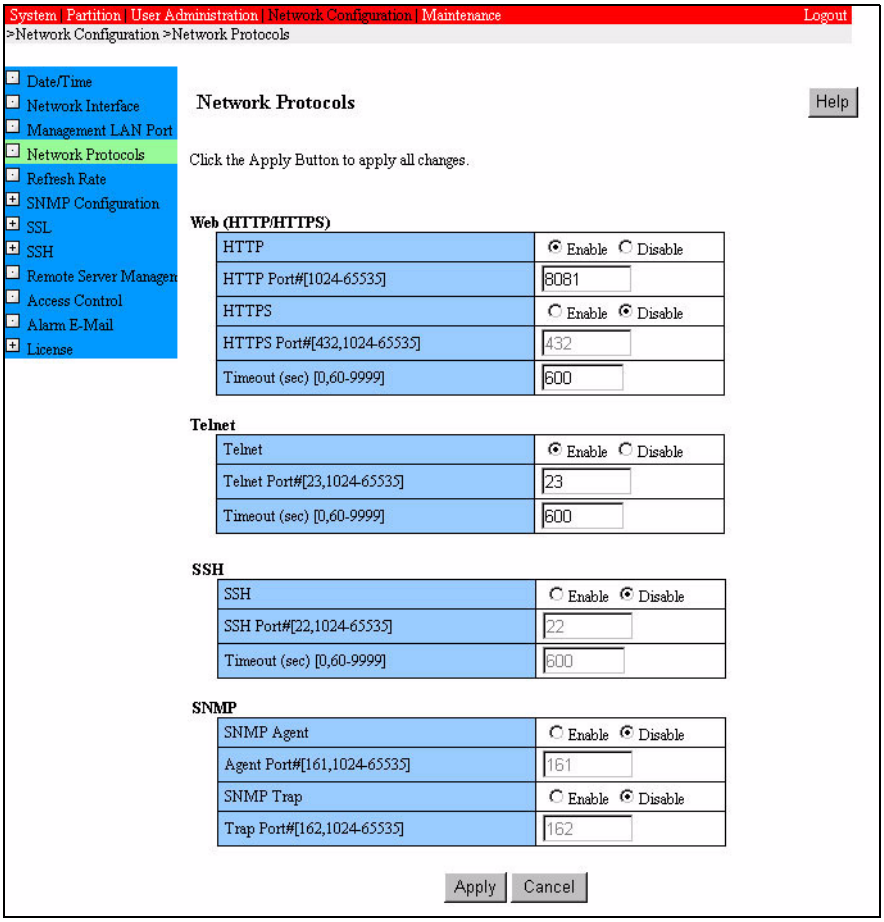


Figure 4.13 Example of setting the usage environment

- 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.

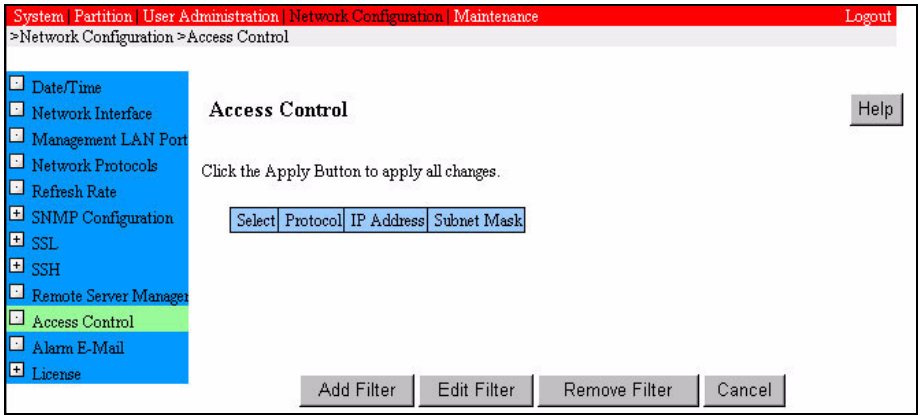


Figure 4.14 Example of the access control settings window

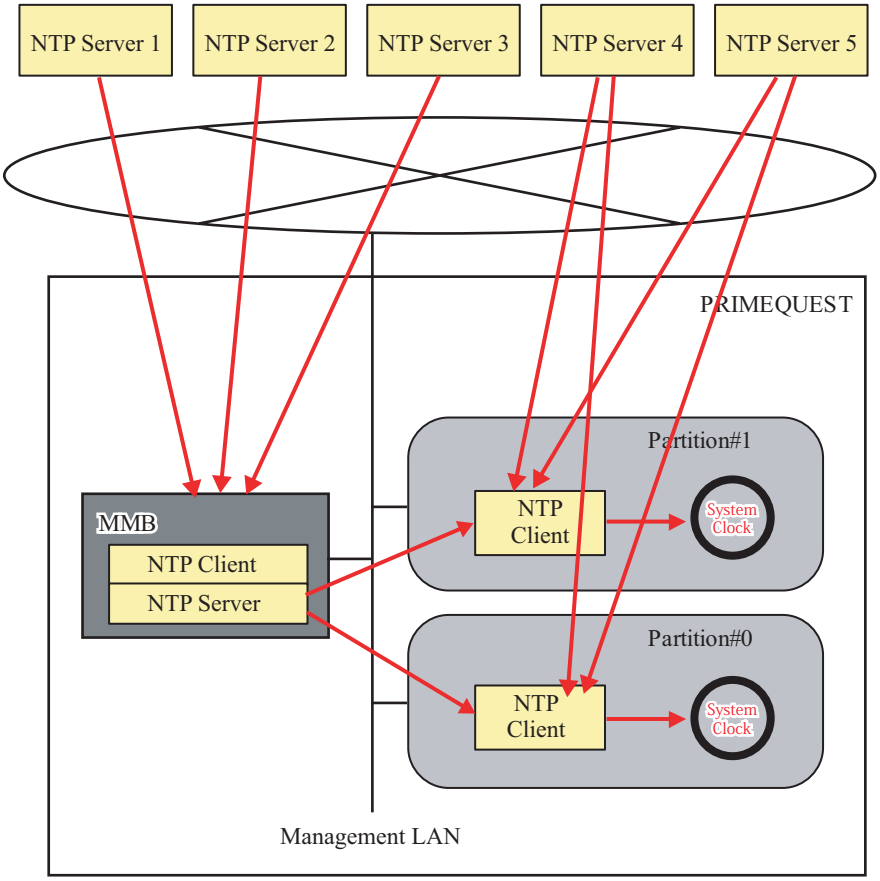
- Secure Sockets Layer (SSL) support

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

4.2.7 Time synchronization

To assure synchronization among partitions, this function synchronizes the partition times with that of the MMB.

The NTP server function and the NTP client function are used to synchronize partition times with the MMB time. As shown in Figure 4.15, synchronization with the NTP server time is possible by accessing the other NTP server. However, the MMB need not be specified for the NTP servers in any of the partitions. 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 4.15 Time synchronization image

4.2.8 Power control/scheduled operation

This function sets a schedule for turning on and off the power supply of the partitions and automatically controls the operation of the partitions in accordance with the set schedule. Figure 4.16 shows an example of the display for scheduled operation.

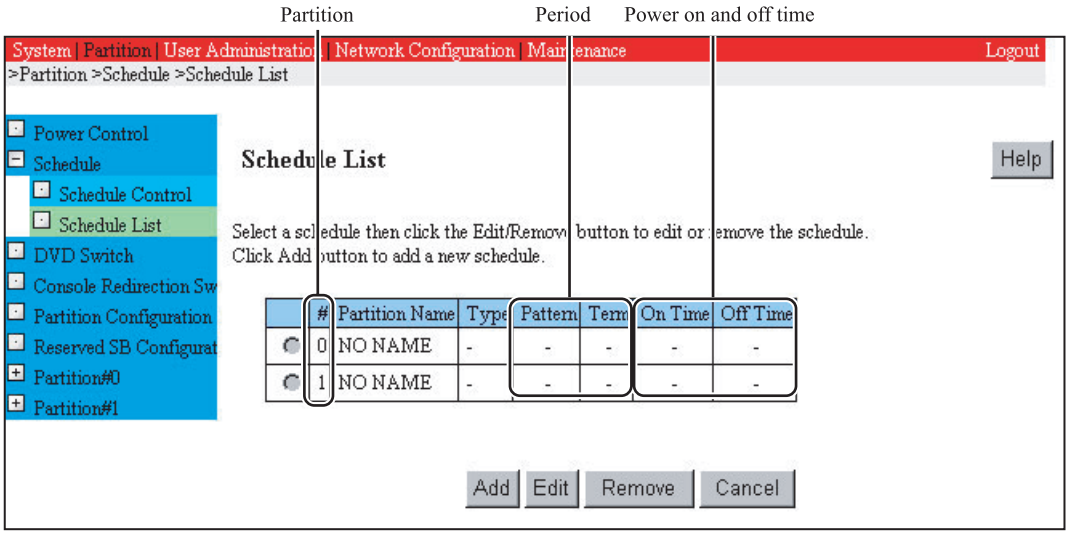


Figure 4.16 Example of the scheduled operation display window

4.2.9 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 (installed in each partition)
- PAL/SAL/EFI firmware (installed in each partition)

4.2.10 Setting information save and restore

This function saves and restores PAL/SAL/EFI setting information for each partition and MMB setting information.

The MMB provides a save and restore function for PAL/SAL/EFI setting information of each partition as shown in Figure 4.17.

This function enables the following operations:

- After PAL/SAL/EFI settings are made from the PAL/SAL/EFI SETUP screen in one partition, the PAL/SAL/EFI setting information for that partition can be saved and applied to another partition from the MMB.
- If a failure occurs in an SB and the SB is swapped, the saved PAL/SAL/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 PAL/SAL/EFI configuration information, a save and restore function for saving and restoring MMB configuration information is provided as well.

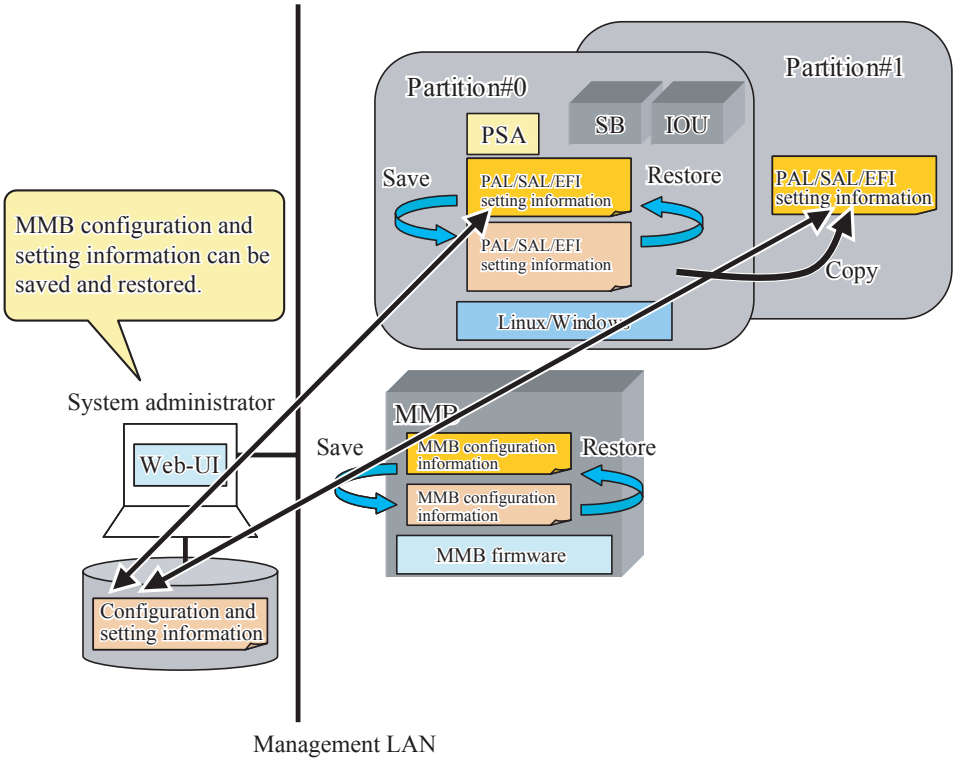


Figure 4.17 Schematic of configuration information save and restore operation

4.3 PSA Functions

The PSA functions are indicated by ◆ PSA in Figure 4.18 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.
- Partition 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 from the partition side.
- When a REMCS contract is entered, software errors are not reported.
- HDD cannot be hot-maintained. The partition 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.

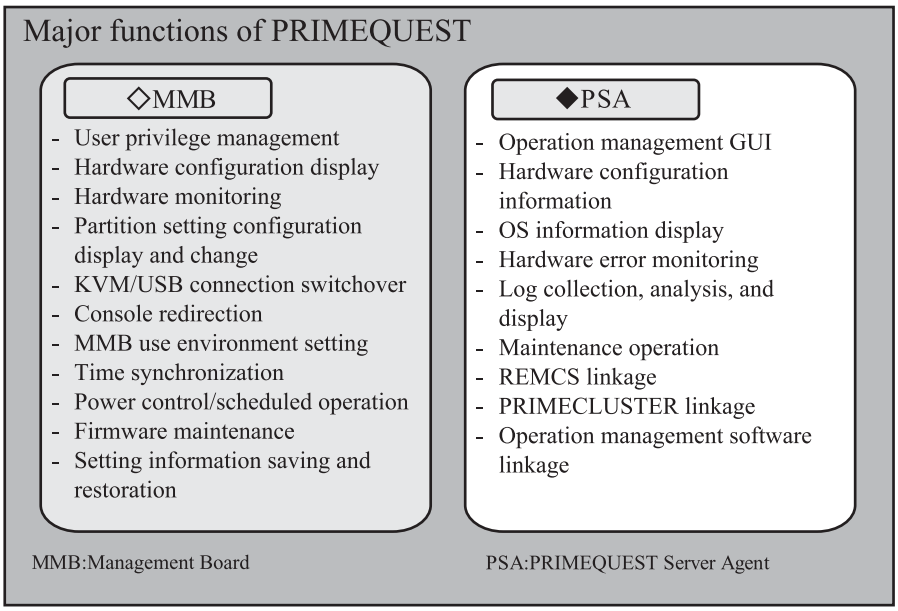


Figure 4.18 PSA functions

4.3.1 Operation management GUI

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

By linking the PSA in each partition to MMB firmware, the partition can be displayed and operated via the MMB Web-UI without a Web server function at the partition end.

As shown in Figure 4.19, the MMB firmware has the Common Gateway Interface (CGI-WebGateCGI) which is used for communication with the Web server function. The PSA in each partition 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 partition operation is provided.

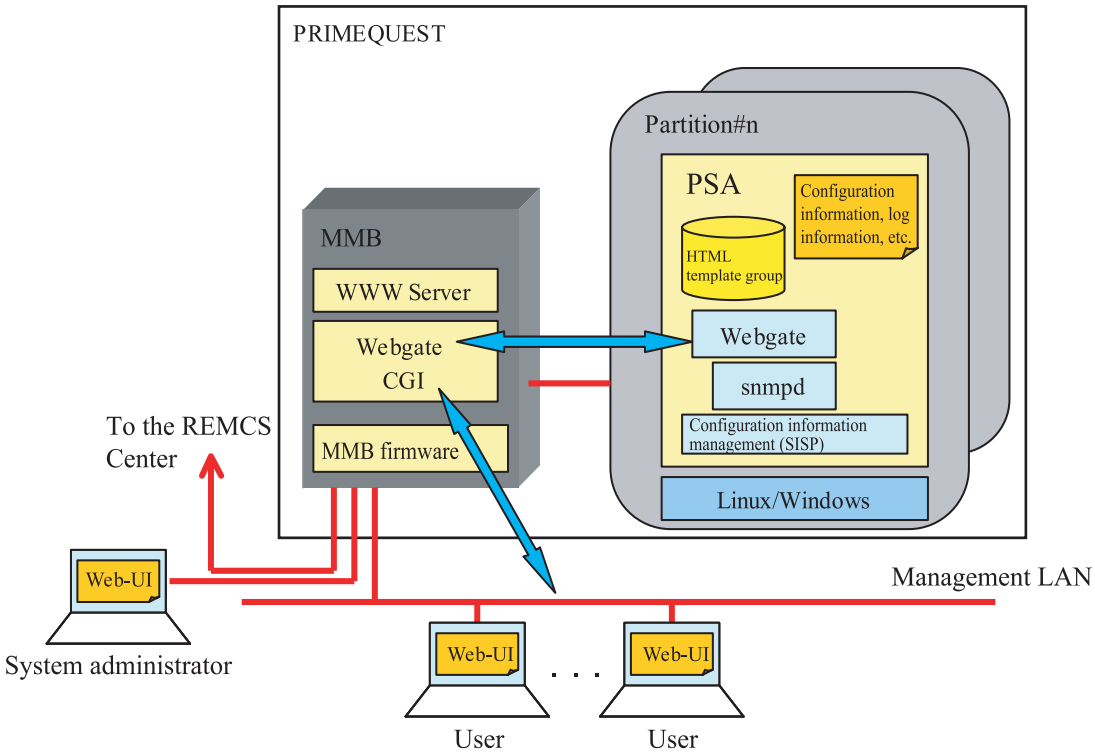


Figure 4.19 Web-UI function

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 580A/540A/520A/500/400 Series Reference Manual: Basic Operation/GUI/Commands* (C122-E003EN).

- 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 that form a partition.

This function displays the following configuration information:

- SB configuration, IO Unit/IOX configuration
- CPU configuration (maximum number of CPUs which can be mounted, CPU mounting locations, identification information such as the CPU type, error status information (PSA 1.13.0 or earlier))
- Memory configuration (mounting locations, detailed information such as the memory type, error status information (PSA 1.13.0 or earlier))
- 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 in a partition.

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 in a partition 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.

As shown in Figure 4.20, PSA records the counts of various errors detected in the CPU, memory, chip sets, and interchip buses by the OS and firmware.

The PSA performs an error analysis to identify a component or unit that had more correctable errors than defined in the threshold for a given period. The PSA then records this information as log information and reports the error information to the MMB and upper-level system management software. This is for determining that the frequent reoccurrence of an error is assumed to be a sign of an impending failure even if the error is a correctable error. The PSA instructs the MMB to disconnect the CPU and memory automatically at next reboot. This is called degradation reservation.

Remarks: PSA versions 1.11 and earlier support the functions of error count analysis and degradation reservation.

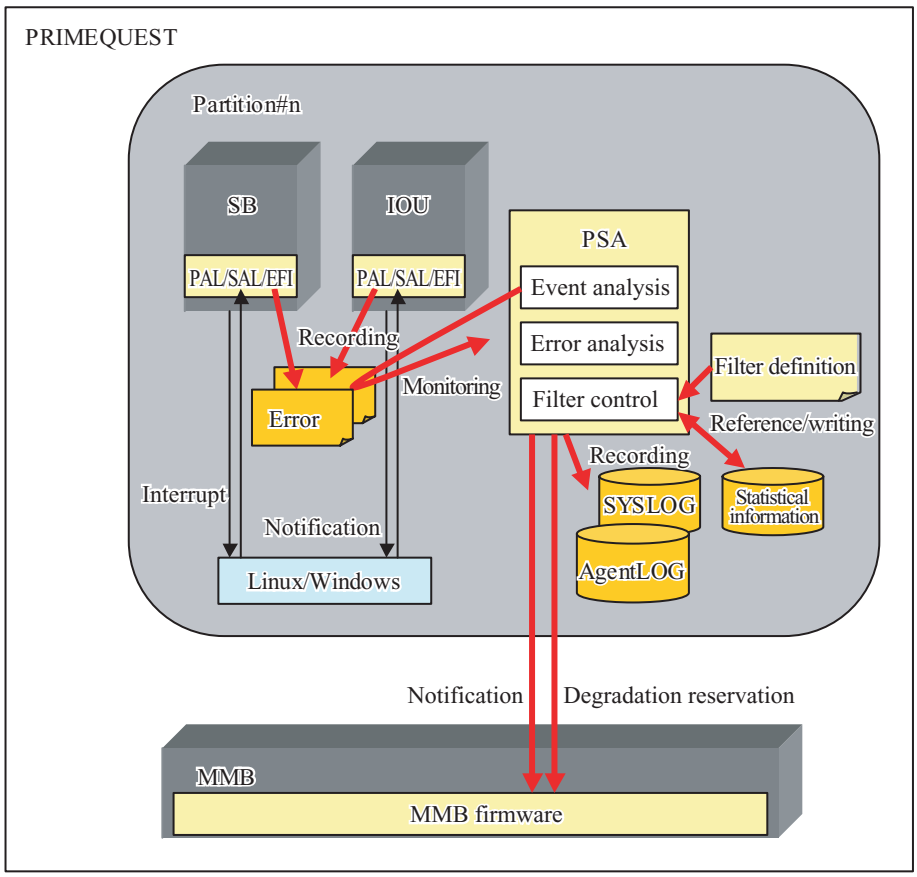


Figure 4.20 Overview of error monitoring

Remarks: The hardware can detect errors of the following two types:

- Uncorrectable error (UE)
- Correctable error (CE)

If an UE occurs, the hardware stops all partitions affected by the error, disconnects the faulty component, and attempts restart. Or the hardware keeps the partition stopped and waits for maintenance.

CEs are corrected by the hardware function. Therefore, the partition need not be stopped or the faulty component need not be disconnected immediately. However, if a CE occurs frequently, the component may be degraded, making it likely that a fatal failure will occur in the future.

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 4.4 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. From a partition, information related to the partition can be collected. 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 in partitions.

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.

4.3.7 REMCS linkage

This function works with the MMB to report resource information and errors of the partition 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 each partition. As shown in Figure 4.21, 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 on the partition 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 6, "REMCS" in the *PRIMEQUEST 500A/ 500/400 Series Reference Manual: Tools/Operation Information* (C122-E074EN).

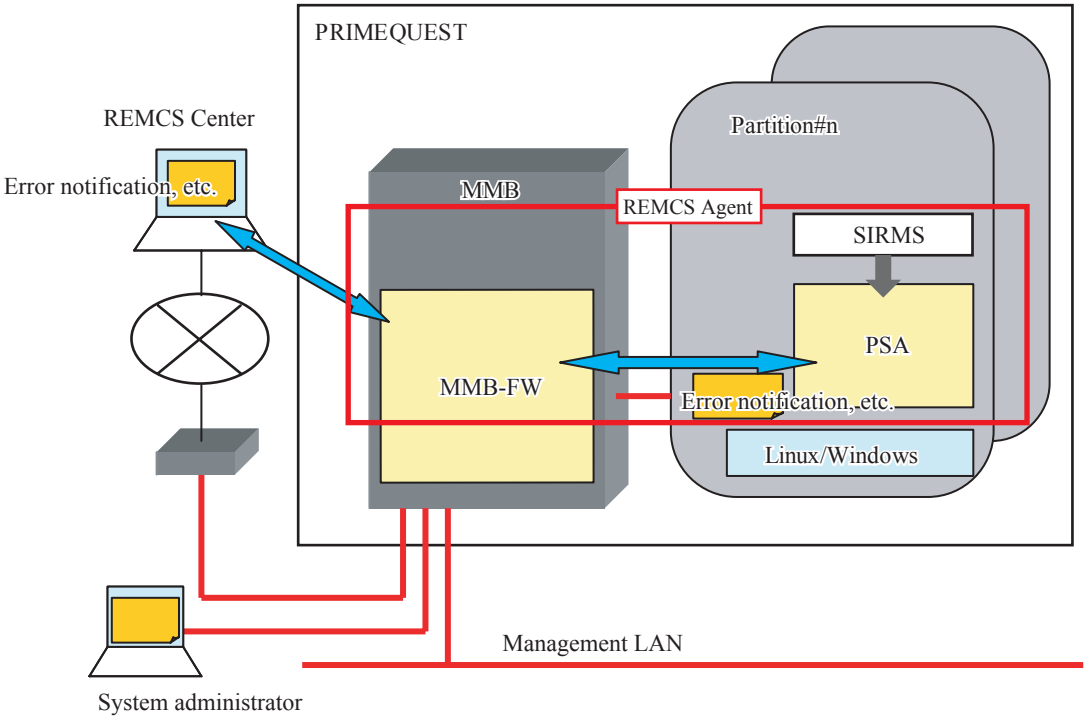


Figure 4.21 REMCS linkage

4.3.8 PRIMECLUSTER linkage

The function allows for a clustering configuration that works with PRIMECLUSTER machines. A cluster consisting of multiple nodes (a node can be considered one server, and for PRIMEQUEST-series machines, a node can be considered one partition) 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 (partition).
- 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 4.22, control information is taken over from the active system to the standby system to continue operation.

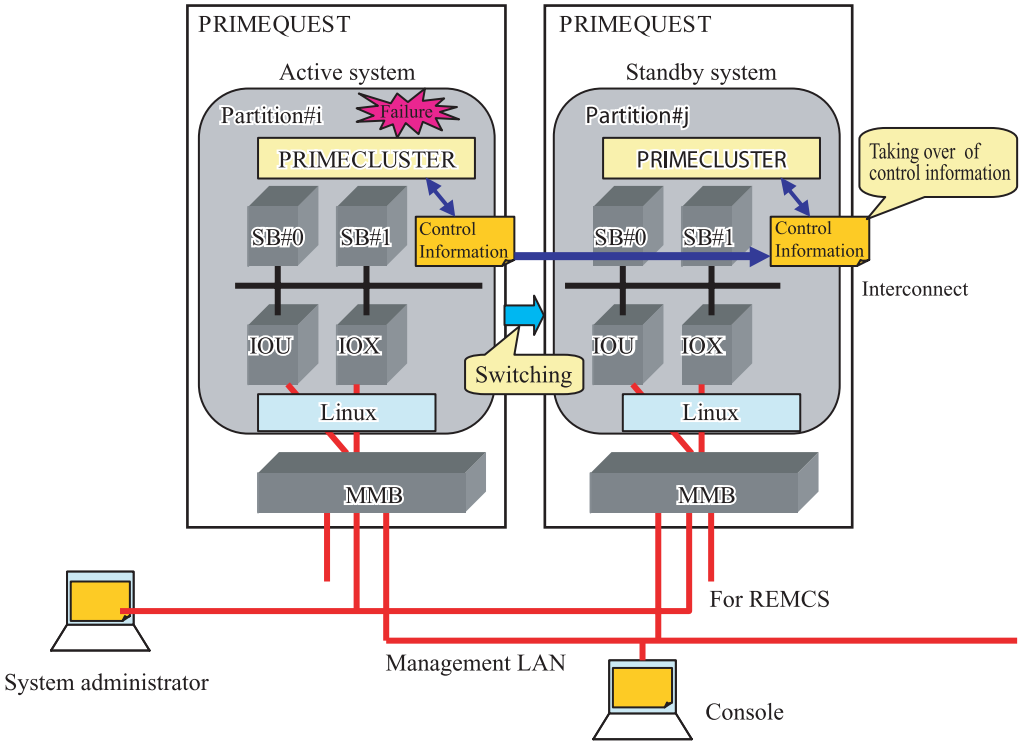


Figure 4.22 PRIMECLUSTER linkage

4.3.9 Operation management software linkage

This function enables a linkage with operation management software.

Figure 4.23 shows software that is installed on a partition and that can work with the PSA.

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

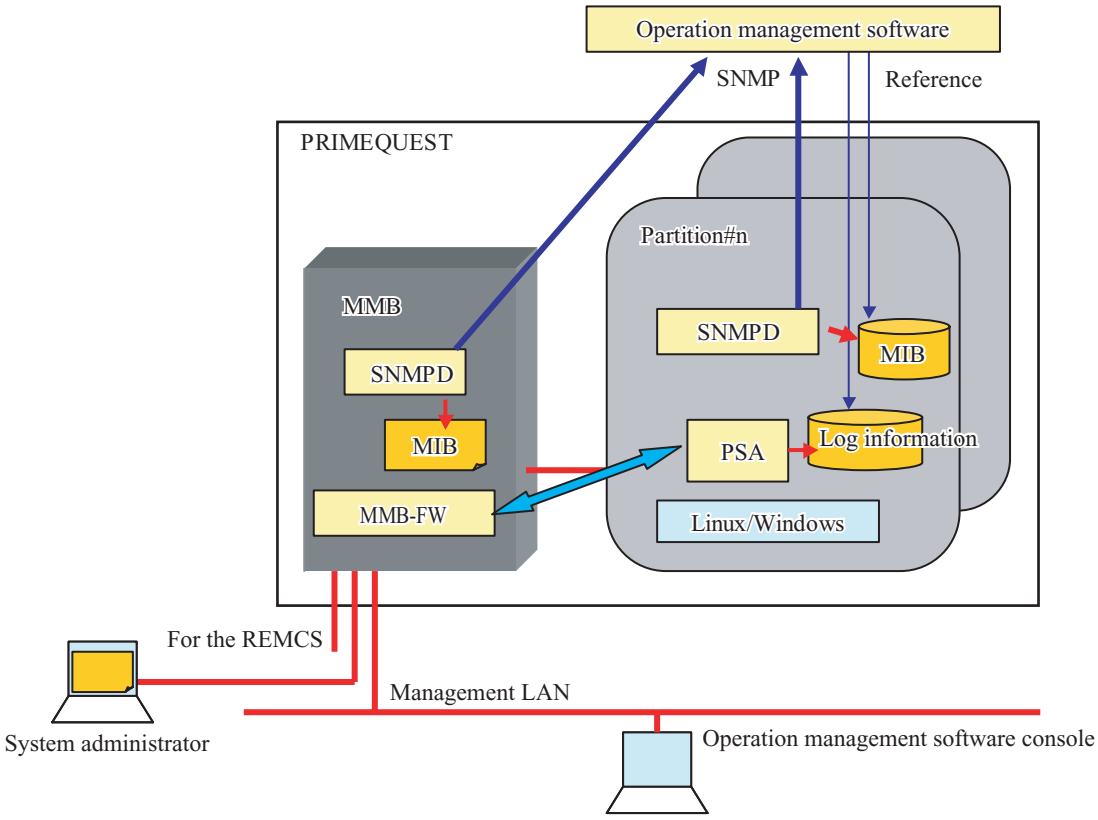


Figure 4.23 Operation management software linkage

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
  (SB, IO Unit/IOX, PCI_Box, FAN TRAY, PSU, MMB, etc.)
- System information display and settings (including cabinet setting, MMB, etc.)
- Partition configuration management and settings
- Maintenance operation (hot swapping of PCI cards, collection and display of logs for partitions)

PSA management targets

- Information management and operation within a partition
  (PCI card, connection I/O, OS information display, and OS resource performance)
- Maintenance operation (hot swapping of disks, collection and display of logs for partitions)

Available functions

The following operations are available from GUI on a partition.

Table 4.5 Operations available from GUI on a partition

Available operation	Overview
Displaying partition configuration information	SB configuration and IO Unit/IOX 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, error status)
	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 in a partition

4.3.11 Information managed on partition

Table 4.6 Information managed on partition

Information type	Description	
Hardware information	Model information	• CPU information: mounting information, status, type, version, and frequency
		• Memory information: mounting information, status and type (size)
		• SB/IO Unit/IOX/PCI_Box 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 most elements of the hardware configuration, such as memory, I/O devices, chip sets, 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).

- Redundancy of Components
- System Mirror Mode
- Redundancy of an HDD (Use of PRIMECLUSTER GDS)
- Redundancy of the Business LAN (Use of PRIMECLUSTER GLS)

5.1 Redundancy of Components

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

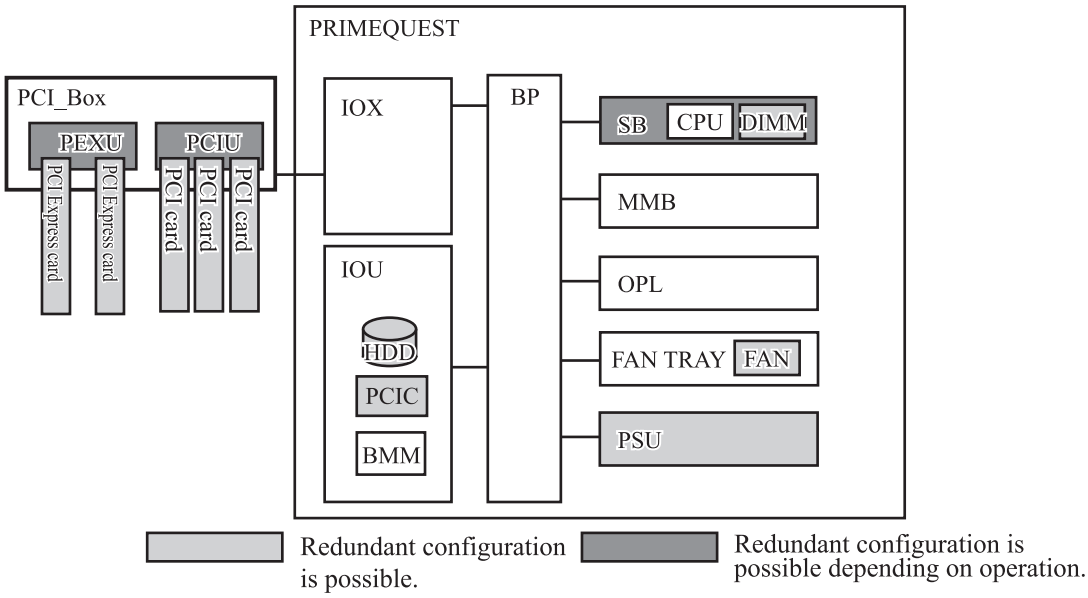


Figure 5.1 General redundant configuration

This part describes redundant configuration of the following components.

- System Board (SB)

  From one to up to two SBs can be mounted. Redundant configuration can be adopted depending on the operation methods. For details, see CHAPTER 7, "Partition," and CHAPTER 8, "Clustering." An SB can have one to four IPF CPUs and up to 8 set (32) memory boards (DIMMs).

- Power Supply Unit (PSU)

  A PSU converts AC input to 48 VDC. This unit adopts redundant configuration as basic configuration.

- FAN Unit (FAN)

  Redundant configuration can be adopted as basic configuration for each cooling system in units of individual fans, with the fan units.

5.2 System Mirror Mode

The PRIMEQUEST has two types of System Mirror Mode, Standard Mirror Mode and Extended Mirror Mode. Standard Mirror Mode duplicates the crossbar (address bus). Extended Mirror Mode duplicates the essential server elements of the memory, crossbar (address bus), crossbar (data bus) and chipset to allow data to be written to and read from them concurrently. Since hardware components and read/write are duplicated, even if a hardware component fails, operation can be continued without stopping the system.

Table 5.1 contains the specifications of System Mirror Mode.

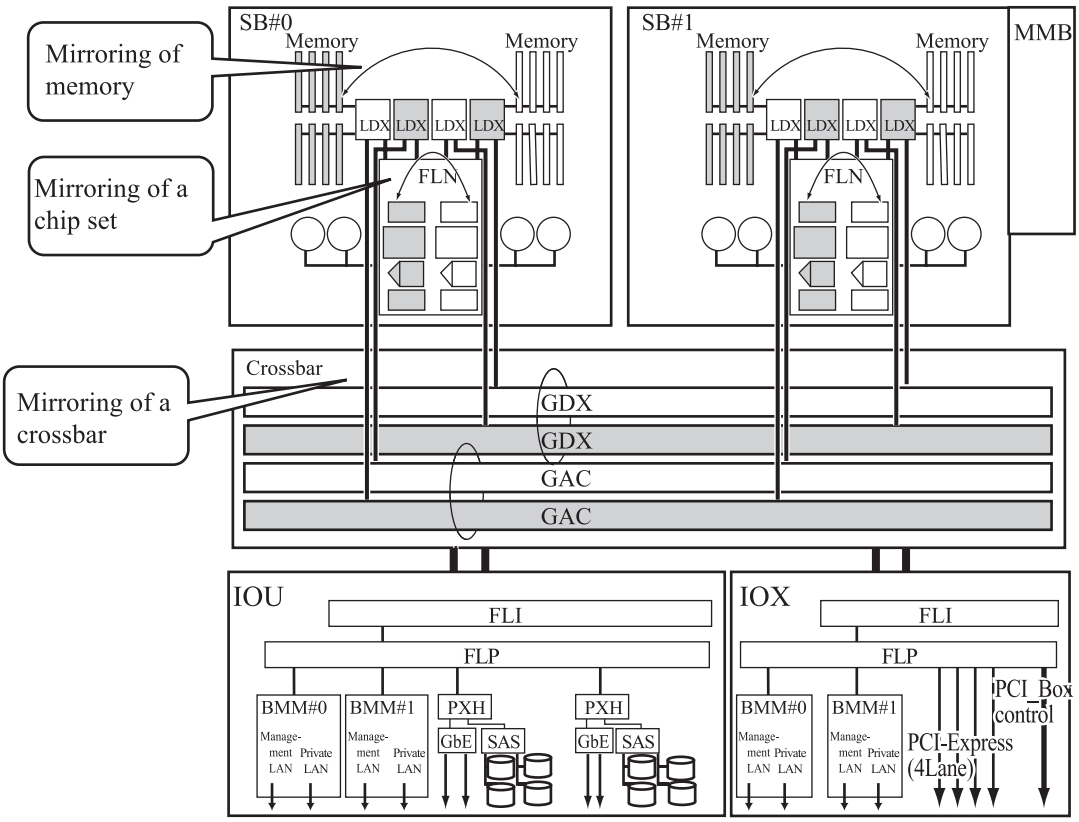
Table 5.1 System Mirror Mode Specifications

Mode / Item	Mode	
	Standard Mirror Mode	Extended Mirror Mode
Crossbar (address bus)	Duplicated	Duplicated
Crossbar (data bus)	Not duplicated	Duplicated
Memory	Not duplicated	Duplicated
Inside of a chip set	Not duplicated	Duplicated

Remarks: Split SBs, IO Units or IOXs cannot be allocated to a partition that is configured for Extended Mirror Mode.

Figure 5.2 outlines the system configuration of Extended Mirror Mode.

Extended Mirror Mode duplicates important hardware in the system, a crossbar (address part), a crossbar (data part), memory, and the inside of the chip set. Data is also simultaneously processed by using both of duplicated components. Thus, even if one of two duplicated components fails, the other component can continue operation without stopping the system.



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

Figure 5.2 Extended Mirror Mode

Mirroring of memory

In Extended Mirror Mode, memory is divided into two groups for mirroring. Therefore, the logical memory capacity becomes half of the physical memory capacity installed. However, data read and write operation is duplicated and performed simultaneously. Thus, even in the case where a part of memory fails, the operation continues using the rest of the memory. Reliability increases with the mirroring of memory. Figure 5.3 shows the logical numbers of memory that are mirrored.

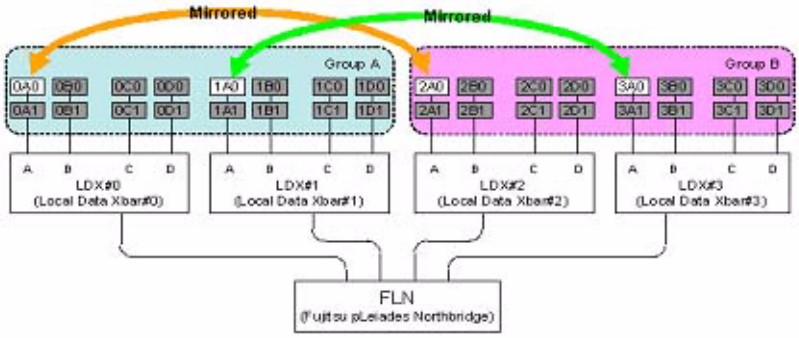


Figure 5.3 Logical numbers of memory for which a redundant configuration has been adopted

5.3 Redundancy of an HDD (Use of PRIMECLUSTER GDS)

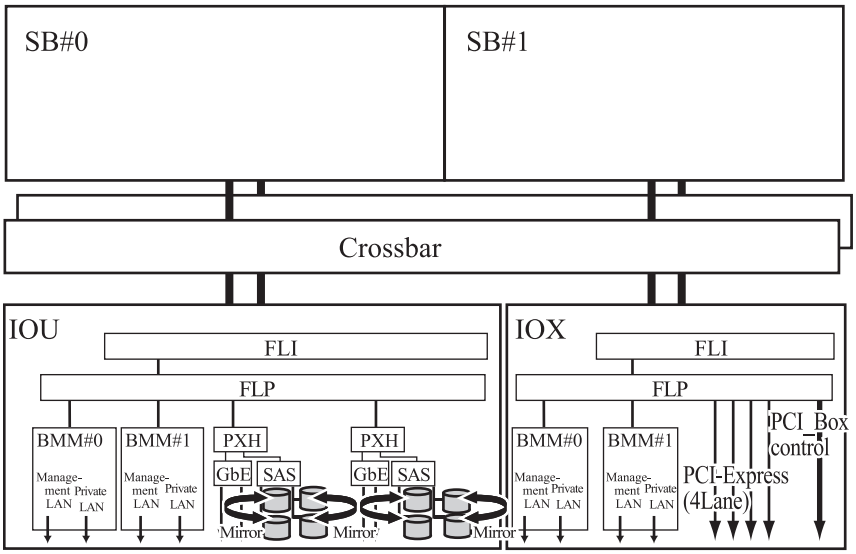
5.3.1 Redundancy provided to system disk drives

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 external RAID.

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

Minimum configuration for HDD redundancy

Figure 5.4 shows the minimum configuration for HDD redundancy. In this configuration, a redundant configuration of HDDs is implemented by connecting four HDDs to one SAS. Redundancy, however, is not applied to SAS controllers, and other I/O devices other than HDDs.

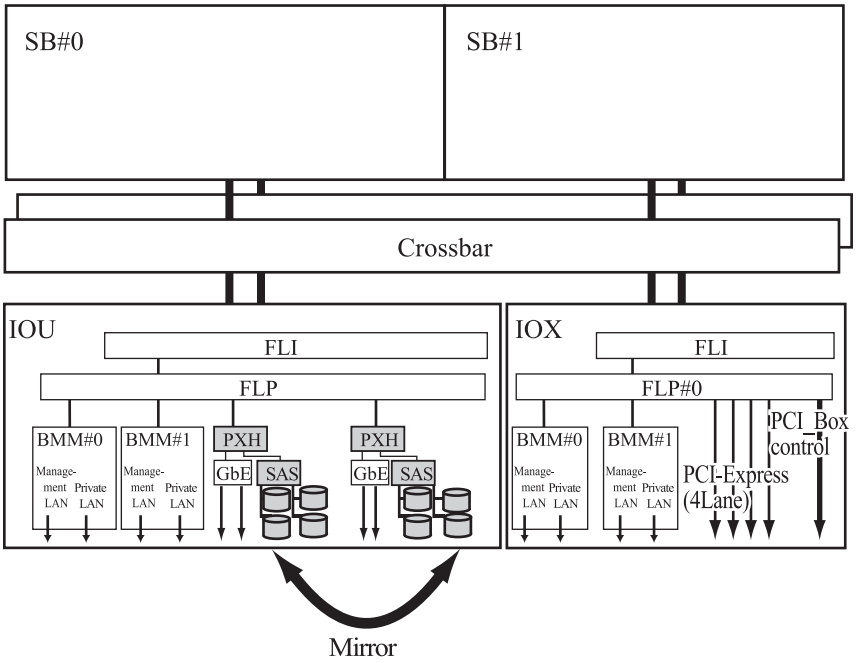


Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

Figure 5.4 Minimum redundancy configuration

Configuration providing redundancy of an HDD and a SAS controller

Figure 5.5 shows a configuration providing redundancy of an HDD and a SAS controller. In this configuration, HDD-SAS controller PXHs are mirrored. I/O paths including FLP and FLI are not mirrored.



Note: BMM #1 supports the PRIMEQUEST 520A/520 only.

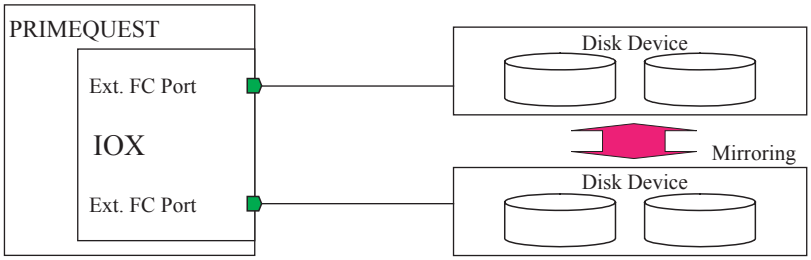
Figure 5.5 Redundancy including SAS controllers

5.3.2 Redundancy of external storage

Mirroring between cabinets

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

Figure 5.6 shows an example where mirroring between cabinets is enabled using external SCSI ports incorporated in an IO Unit. In this example, external SCSI ports of the same IO Unit are used. To improve system reliability, use external SCSI ports of different IO Units.



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

Figure 5.6 Disk device mirroring between cabinets

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.

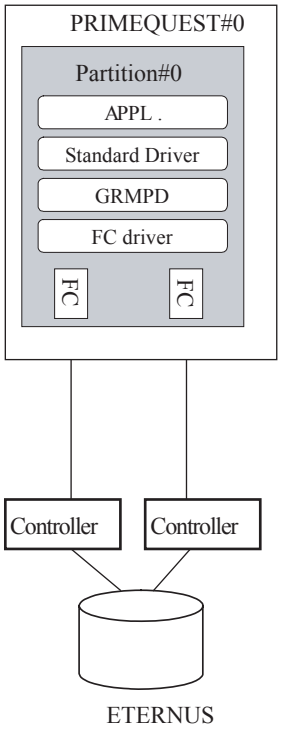


Figure 5.7 ETERNUS multipath connection

5.4 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.

5.4.1 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.

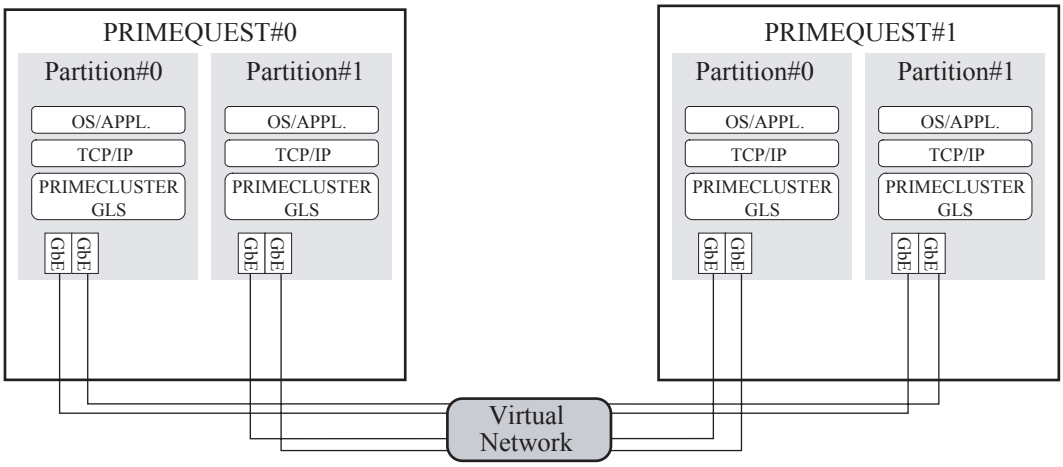


Figure 5.8 Example of implementing redundancy using the high-speed switching method

With this method, as shown in Figure 5.9, 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.

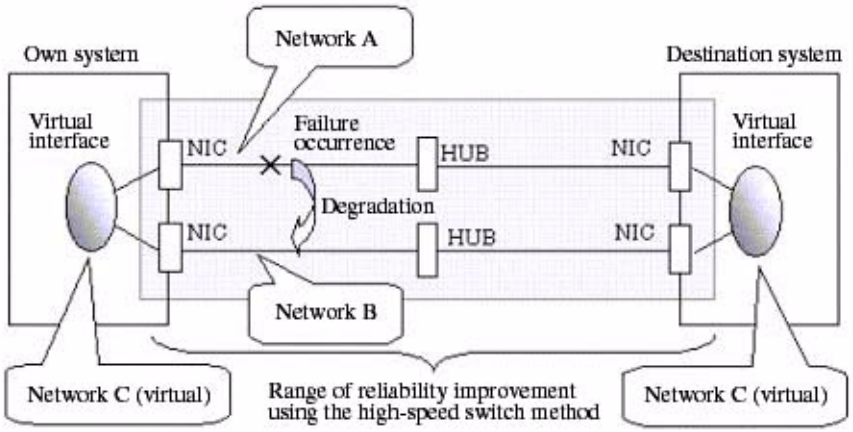


Figure 5.9 High-speed switching method

5.4.2 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 5.10, 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.

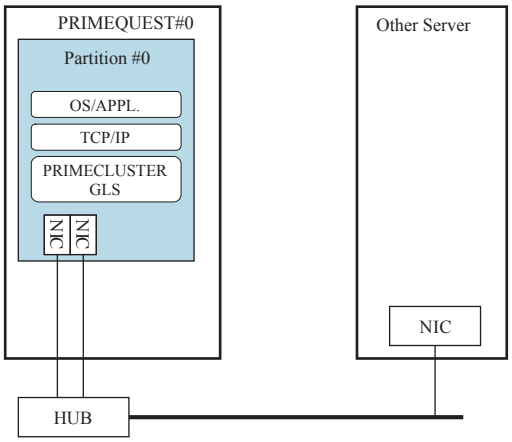


Figure 5.10 Example of implementing redundancy using the NIC switch method

With this method, as shown in Figure 5.11, 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.

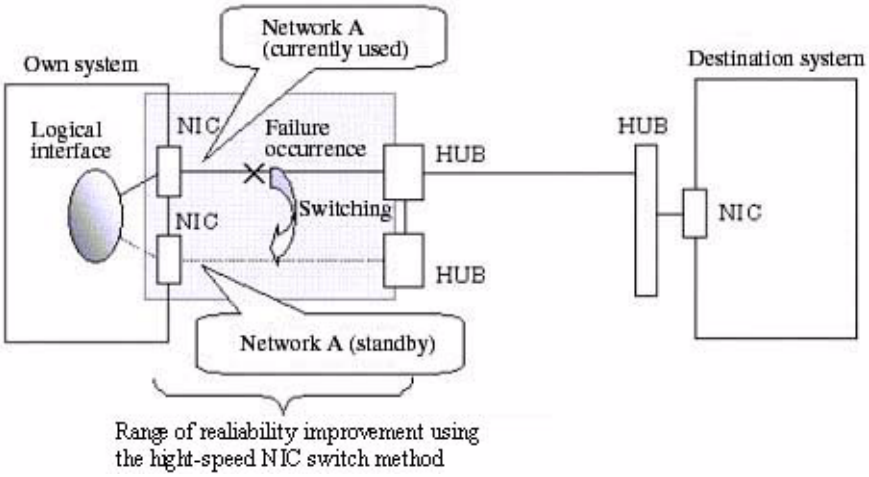


Figure 5.11 NIC switching method

5.5 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, the redundant configuration of a component where a failure has occurred, and the mirror mode status. 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.

If an SB or an IO Unit fails, and reserved SB set to be switched to in case of an SB failure exists, and an IO Unit to which redundant configuration is applied exists, reserve degradation and stop the partition. Incorporate the reserved SB into the partition at the reboot.

Details on partitions are described in CHAPTER 7, "Partition."

Case where the location of a failure cannot be identified

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

If, after the partition 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 partition 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 partition 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 partition, 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.

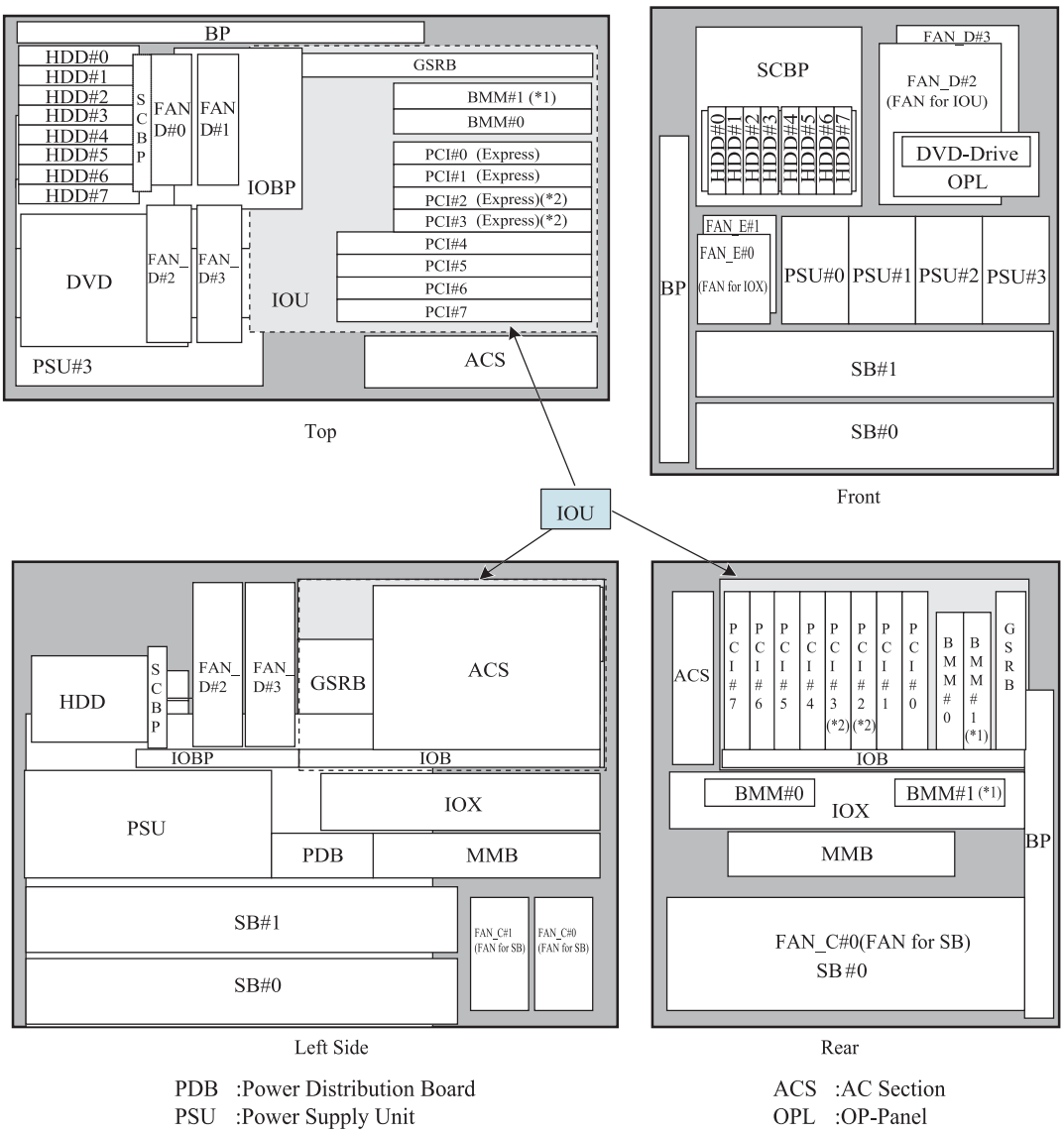


Figure 6.1 Layout of included components

Notes:

1. BMM #1 supports the PRIMEQUEST 520A/520 only.
2. PCI #2 and PCI #3 can be used with the PRIMEQUEST 520A/520 only.

Table 6.1 Component hot swapping and/or redundant configuration availability

Item	Component	Hot swapping	Redundant configuration	Notes on redundant configuration
1	SB	N	Y (optional)	
2	CPU	N	N	
3	Memory (DIMM)	N	Y	Redundancy is ensured when Extended Mirror Mode is set.
4	IO Unit/IOX	N	N	A redundancy function that is ensured by hardware alone is not supported.
5	BMM	N	N	
6	HDD	Y (*1)	Y	Linux: Redundancy is ensured by PRIMECLUSTER GDS (software mirror) See Section 5.3, "Redundancy of an HDD (Use of PRIMECLUSTER GDS)." Windows: System disks only. Redundancy is ensured by a PSDN.
7	PCI Card	Y	Y	Windows Server 2008: See the PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide (C122-E087EN). Windows Server 2003: Hot swapping of PCI cards is not supported.
8	BP	N	N	
9	MMB	N	N	
10	DVD drive	Y	N	
11	OP Panel	Y	N	
12	PSU	Y	Y	
13	FAN Unit	Y	Y	
14	FAN	Y	Y	
15	PCI_Box	N	N	
16	PCIU	N	-	Redundancy is ensured depending on the operation.
17	PCI card	Y	Y	Windows Server 2008: See the PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide (C122-E087EN). Windows Server 2003: Hot swapping of PCI cards is not supported.

Item	Component	Hot swapping	Redundant configuration	Notes on redundant configuration
18	PEXU	N	-	Redundancy is ensured depending on the operation.
19	PCI Express card	Y	Y	Windows Server 2008: See the PRIMEQUEST 500A/500 Series Microsoft Windows Server 2008 User's Guide (C122-E087EN). Windows Server 2003: Hot swapping of PCI cards is not supported.
20	IO_PSU	Y	Y	
21	Expansion file unit	N	N	

Y = Available, N = Unavailable

*1 This is supported only for Linux.

Redundant configuration and hot maintenance of each unit are described elsewhere in this manual.

Remarks: In a state where System Mirror Mode is not set, neither memory nor crossbars are mirrored. In this state, the System Mirror Mode setting as displayed in the [System Setup] window is "Disable."

(1) SB

A reserved SB can be used for a redundant configuration, but hot swapping is not available.

If an SB fails in a partition for which a reserved SB is set, the reserved SB is incorporated into the partition to replace the failed SB. The partition must be rebooted to incorporate the reserved SB into the partition.

(2) CPUs

Hot swapping is not available.

To swap the failed CPU, stop the partition that includes the SB containing the failed CPU, extract the SB containing the failed CPU, and maintain (swap) the CPU. Insert the SB containing a new CPU into the partition, and then restart the partition.

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

Isolating the failed CPU may cause no normal CPU to be available in the SB. In this case, the entire SB is subjected to isolation even if a normal DIMM exists. If no normal SB is available in the partition, the partition stops.

(3) Memory (DIMMs)

Memory (DIMM) cannot be subjected to hot swapping. Stop the partition that includes the SB containing the failed DIMM, extract the SB containing the failed DIMM, and replace the failed DIMM.

- When the system is operated in Extended Mirror Mode:
  Only a redundant configuration is available. If a DIMM fails, the system containing the failed DIMM stops, and operation continues with the other normal system alone.
- When the system is operated without Standard Mirror Mode or System Mirror Mode Set:
  If a DIMM fails while System Mirror is not set, the partition stops. The failed DIMM is automatically isolated from the partition when the partition restarts. The unit of isolation is a set of four DIMMs including the failed DIMM (same as the unit of expansion). The partition automatically restarts after the DIMM failure.
  Isolating the failed DIMM may cause no normal DIMM to be available in the SB. In this case, the entire SB is subjected to isolation even if a normal CPU exists. If no normal SB is available in the partition, the partition stops.

(4) IO Units/IOX

A redundant configuration and hot swapping are not available.

The partition may stop due to a failure.

After the partition stops, the failed component can be swapped. To add a new IO Unit to the partition, the relevant partition must be restarted.

(5) BMMs

A redundant configuration and hot swapping are not available.

The partition stops if a BMM fails. Turn off the power to the IO Unit containing the failed BMM, and replace the failed unit.

(6) 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.
- In a partition with Windows installed:
  Redundancy is ensured for system disks only by a PSDN. Hot swapping is not supported.

(7) PCI cards (in IO Units)

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 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 card replacement using PCI Hot Plug, consult with a Fujitsu certified service engineer.

(8) BPs

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

(9) MMBs

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

(10) DVD drive

Redundant configuration is not adopted, but hot swapping is available. However, the DVD drive cannot solely be swapped since it is the built-in drive of the OP Panel.

(11) PSU

Redundant configuration is adopted, and hot swapping is available.

(12) OP Panels

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

(13) FAN Unit

Redundant configuration is adopted, and hot swapping is available.

(14) FANs

The fan has a redundant configuration to make it hot-swappable. It can be swapped on a fan unit basis.

(15) PCI_Boxes

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

(16) PCIUs (in PCI_Boxes)

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

(17) PCI cards (in PCIU)

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 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 card replacement using PCI Hot Plug, consult with a Fujitsu certified service engineer.

(18) PEXU (in PCI_Boxes)

Hot swapping is not available.

Redundant configurations are available for some operation patterns.

(19) PCI Express cards (in PEXU)

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 card can be hot-swapped with the PCI Express 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 card replacement using PCI Hot Plug, consult with a Fujitsu certified service engineer.

(20) IO_PSUs (in PCI_Boxes)

Redundant configuration is adopted, and hot swapping is available.

(21) Expansion file unit

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

CHAPTER 7 Partition

This chapter describes the partition function the PRIMEQUEST provides and the following items that are required to determine a system configuration:

- Partitioning Function Overview
- PRIMEQUEST Partitioning Function
- Partitioning Function of the PRIMEQUEST 520A/520 with XPAR Installed
- Partition Configuration Definition
- Points to Consider when Determining Configurations
- Function for Changing Partition Configurations

7.1 Partitioning Function Overview

The PRIMEQUEST series allows one system to be divided into multiple systems, with each system operating as an independent system. Each of the individual systems resulting from such division is called a partition. An operating system (OS) can run independently on each partition. This enables running multiple servers, which operate independently from one another, as a single server.

Each partition is protected from the effects of other partitions. Therefore, a failure (such as an OS panic) occurring in one partition does not affect job processing in other partitions directly. The OS can be rebooted and shut down independently in each partition. These operations are performed by the system administrator of each partition.

Partitions can be built using any mix of SBs (system boards), IO Units and IOXs. The configuration of a partition can be defined by using the Web-UI of the management board (MMB).

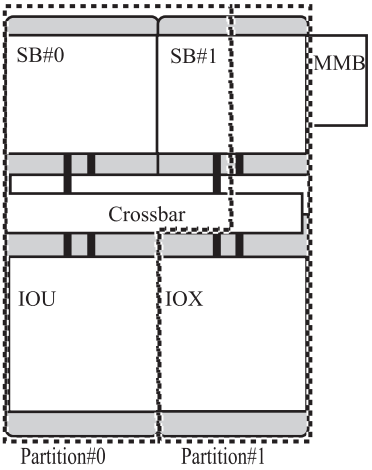


Figure 7.1 Example of partition configuration

Without XPAR installed, the PRIMEQUEST 420 or PRIMEQUEST 520A/520 can each be configured into up to two partitions. Each SB, IO Unit, and IOX in the PRIMEQUEST 520A/520 can be logically split, and they can configure up to four partitions. The process of logically splitting an SB and IO Unit to configure a partition is called “XPAR” (eXtended PARtitioning). An XPAR option is required for implementing XPAR. The XPAR feature is an option of the PRIMEQUEST 520A/520.

7.2 PRIMEQUEST Partitioning Function

This section describes the mounting requirements of SBs and IO Units and their allocation to the PCI Box.

7.2.1 Mounting requirements

When a partition is defined as a combination of SBs and IO Units/IOX, the individual mounting requirements of these units must be met.

7.2.1.1 Mounting requirements of SBs

When defining a partition, ensure that the following mounting requirements are fulfilled:

- The SB that has memory mapped in the area from 0 GB to 4 GB is called the home SB, and it is automatically configured.
- On each SB, at least one CPU and one set of memory modules (four modules) must be mounted.
- DIMMs on a bus must have the same capacity.
- CPUs on the same SB mounted on the same partition must be of the same type.
- An SB with 16 GB of memory mounted (4-GB DIMM x 4) cannot be mixed in the same partition with an SB having another memory configuration.

7.2.1.2 Mounting requirements of IO Units or IOX

When defining partitions, ensure that the following mounting requirements are fulfilled:

- The partition must include one IO Unit or IOX having a BMC module (BMM) mounted.
- When a partition contains IO Units and IOXs with a BMM mounted on each , only one BMM actually operates. The IO Unit or IOX that contains the operating BMM is called the home IO Unit or home IOX.

Remarks: The Home IOU and Home IOX can be set on the [Home of Partition] window, which is displayed by clicking the [Home] button on the [Partition Configuration] window of the MMB Web-UI that is displayed by selecting [Partition].

7.2.2 PCI_Box allocation

A PCI_Box is connected via an IOX. Each PCIU/PEXU is allocated to the partition in which the IOX belongs.

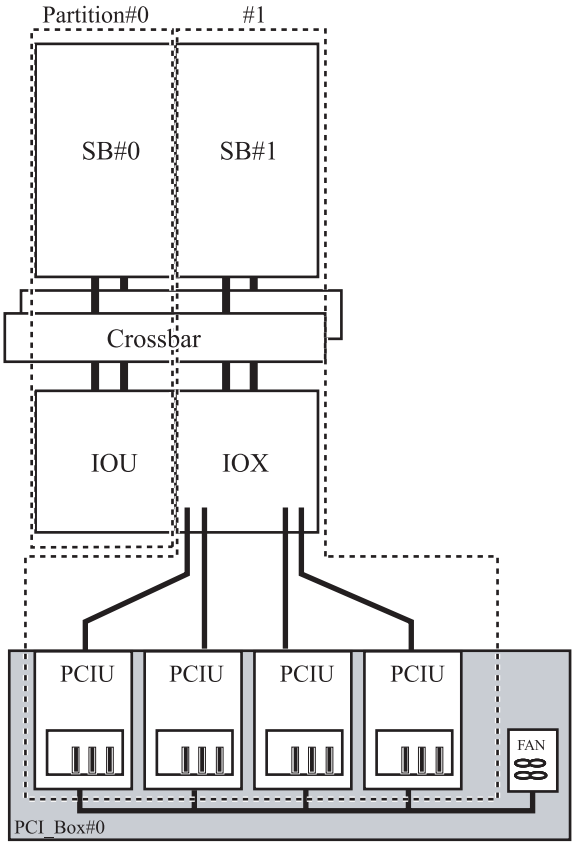


Figure 7.2 Allocation to a PCI_Box partition (only if a PCIU is mounted)

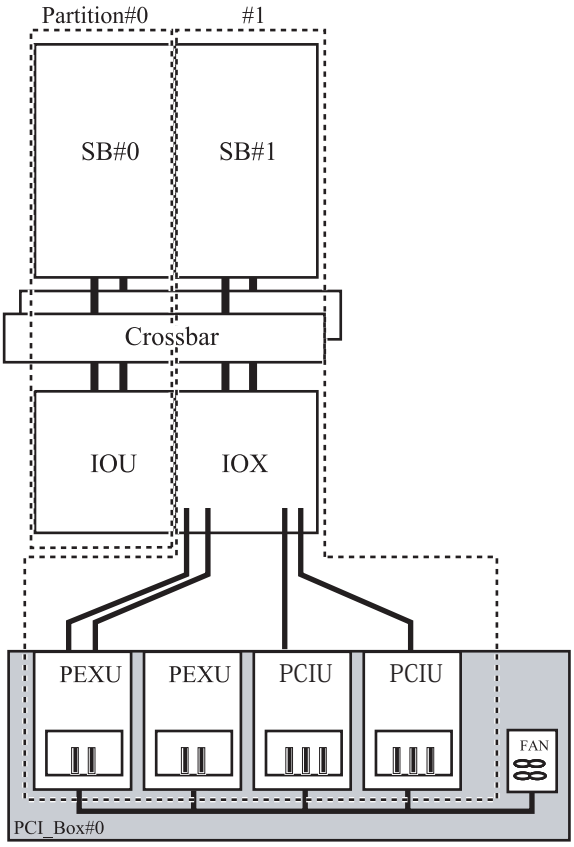


Figure 7.3 Allocation to a PCI_Box partition (both a PCIU and a PEXU are mounted)

7.3 Partitioning Function of the PRIMEQUEST 520A/520 with XPAR Installed

This section describes mounting requirements and allocation rules for SBs and IO Units in the PRIMEQUEST 520A/520 with XPAR installed.

7.3.1 Splitting SBs, IO Units, or IOXs when XPAR has been installed

The hardware resources of each SB, IO Unit, and IOX in the PRIMEQUEST 520A/520 can be logically split into halves internally. By splitting the hardware resources, the PRIMEQUEST 520A/520 with XPAR installed can be configured into up to four partitions.

Notes:

- Extended Mirror Mode cannot be set in a partition containing a split SB, IO Unit, or IOX. Split SBs, IO Units, or IOXs cannot be incorporated in a partition for which Extended Mirror Mode is set.
- A split SB cannot be allocated as a reserved SB to a partition for which Extended Mirror Mode is set. Similarly, Extended Mirror Mode cannot be set in a partition to which a split SB is allocated as a reserved SB.

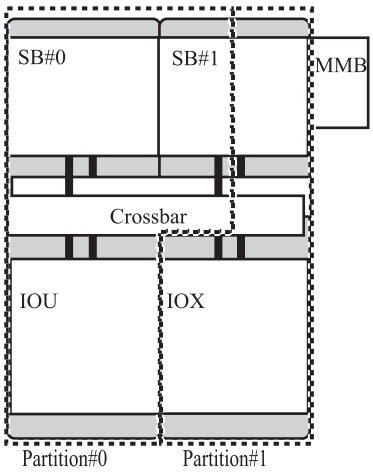


Figure 7.4 Sample partition configuration (logical splitting)

7.3.2 Partition Granularities

The smallest unit of an SB, IO Unit or IOX that makes up a partition is called a "partition granularity." The partition granularity of an SB is the "LSB (Logical System Board), and that of an IO Unit or IOX is the LIOU (Logical IO Unit).

7.3.2.1 SB partition granularity

The partition granularity (LSB) of an SB comes in two kinds: the SB itself and the SB logically split into two halves. The mode of SB operation in which a single SB is used as a partition granularity (LSB) is called the "1LSB mode," and that of SB operation in which a single SB is logically split into two halves is called the "2LSB mode."

Remarks:

1. When a partition runs in Extended Mirror Mode, the SBs associated with that partition are operable only in 1LSB mode. (Split IO Units cannot be assembled into a partition running in Extended Mirror Mode.)
2. When a partition runs in Standard Mirror Mode, the IO Units associated with that partition are operable in both 1LSB and 2LSB modes.
3. Every LSB must always hold a CPU and memory. An LSB without a CPU or memory would be inoperable.

A logically split SB together with a logically split IO Unit or IOX (details below) can configure a partition. The maximum number of partitions with such split components in the PRIMEQUEST 520A/520 depends on operation mode and XPAR (option) settings as follows:

- Operating in Standard Mirror Mode and System Mirror Mode not set
  - XPAR (option) not installed: Up to two partitions
  - XPAR (option) installed: Up to four partitions
- Operating in Extended Mirror Mode: Up to two partitions

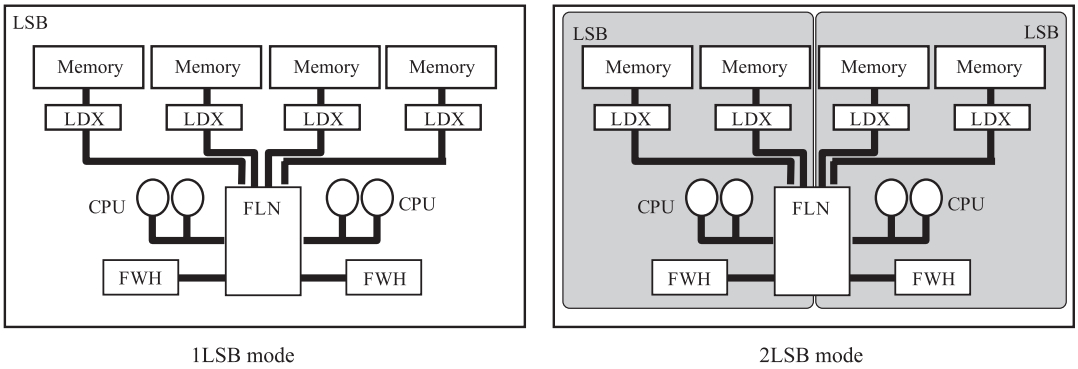


Figure 7.5 SB partition granularities

The table below summarizes the effects of partition granularity (LSB) failures.

Table 7.1 Effects of LSB failures

Partition granularity	Partition granularity component unit	Effect of partition granularity failure on other partitions
LSB (Logical SystemBoard)	1LSB mode	No
	2LSB mode	Where the SB is logically split, the partition with which each split of the SB is associated is affected.

7.3.2.2 IO Unit and IOX partition granularity

The IO Unit and IOX partition granularity (LIOU) is an IO Unit or IOX logically split into two halves.

A logically split SB together with a logically split IO Unit or IOX can configure a partition . The maximum number of partitions with such split components in the PRIMEQUEST 520A/520 depends on operation mode and XPAR (option) settings as follows:

- Operating in Standard Mirror Mode and System Mirror Mode not set
  - XPAR (option) not installed: Up to two partitions
  - XPAR (option) installed: Up to four partitions
- Operating in Extended Mirror Mode: Up to two partitions

When an IO Unit or IOX is logically split into two halves, some of its components are fixedly allocated to either granularity while others are allocatable to either granularity at the user's option. For more details, see 7.3.3.2, "Mounting requirements of IO Units and IOX."

Remarks:

1. The IO Units and IOXs belonging to a partition operating in Extended Mirror Mode can be used only in 1LIOU mode. (Split IO Units and IOXs cannot be incorporated in a partition for which Extended Mirror Mode is set.)
2. The IO Units and IOXs belonging to a partition operating in Standard Mirror Mode can be used in both 1LIOU mode and 2LIOU mode.

7.3.3 Mounting requirements (when XPAR has been installed)

When a partition is defined in a mix of SBs (LSBs) and IO Unit or IOX (LIOUs), the relevant individual mounting requirements must be met.

7.3.3.1 SB mounting requirements

When defining a partition, ensure that the following mounting requirements are fulfilled:

- The SB that has memory mapped in the lowest address location is called the "home SB," and it is automatically configured.
- Each SB must have at least one CPU and one set of memory modules (four modules) mounted on it.
- DIMMs mounted on a bus must have the same capacity.
- CPUs mounted on the same SB mounted or in the same partition must be of the same type.

CPU and memory allocations in 2LSB mode

When the SB is run in 2LSB mode, CPUs and memory are fixedly allocated to partition A and B as shown below.

- An SB with 16 GB of memory mounted (4-GB DIMM x 4) cannot be mixed in the same partition with an SB having another memory configuration.
  - Partition granularity LSB #A: Memory A (Channel #A and Channel #B) and CPU A
  - Partition granularity LSB #B: Memory B (Channel #C and Channel #D) and CPU B

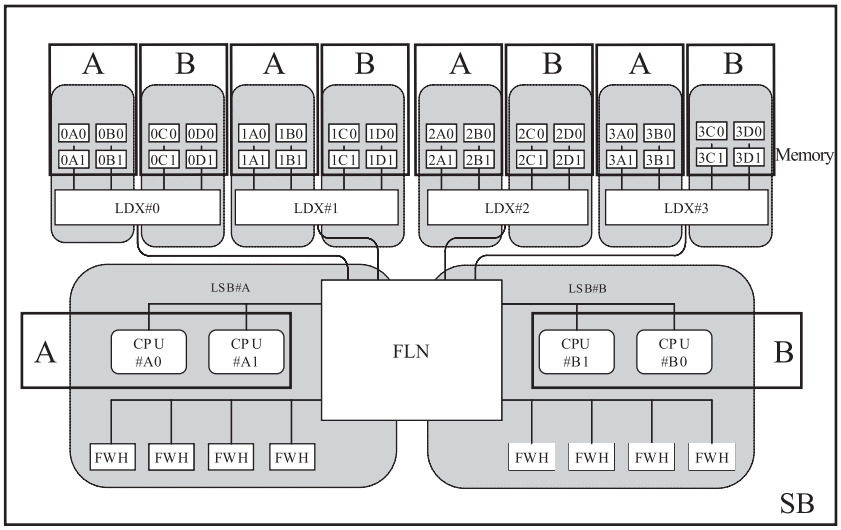


Figure 7.6 SB partition granularities and allocation rules

Table 7.2 SB partition granularities and allocation rules

Partition granularity	CPU and Memory		Partition granularity	CPU and Memory	
LSB#A	CPU#A	0	LSB#B	CPU#B	0
		1			1
	Channel#A	0a0		Channel#C	0c0
		0a1			0c1
		1a0			1c0
		1a1			1c1
		2a0			2c0
		2a1			2c1
		3a0			3c0
		3a1			3c1
	Channel#B	0b0		Channel#D	0d0
		0b1			0d1
		1b0			1d0
		1b1			1d1
		2b0			2d0
		2b1			2d1
		3b0			3d0
		3b1			3d1

7.3.3.2 Mounting requirements of IO Units and IOX

When defining partitions, ensure that the following mounting requirements are fulfilled:

- Each partition must include at least one IO Unit or IOX having a BMC module (BMM) mounted.
  Remarks: The BMM is an optional product.
- When a partition contains IO Units and IOXs with a BMM mounted on each, only one BMM actually operates. The IO Unit that contains the operating BMM is called the "home IO Unit." The IOX that contains the operating BMM is called the "home IOX."

Remarks:

The Home IOU and Home IOX can be set on the [Home of Partition] window, which is displayed by clicking the [Home] button on the [Partition Configuration] window of the MMB Web-UI that is displayed by selecting [Partition].

(1) Rules on IO Unit granularity allocation

This section details the procedures for allocating IO Unit granularities.

When an IO Unit is split into two granularities, some of its components are fixedly allocated to either granularity while others are allocatable to either granularity at the user's option.

IO Units are allocated to partition granularities A and B pursuant to the following rules:

- Blocks a0 and a1 that contain an HDD, and BMM0 and BMM1 that contain a BMM are allocated fixedly. Blocks a0 and a1 are allocated to partition granularity A, and blocks a1 and e1 are allocated to partition granularity B.
- Blocks b, c, d0 and d1 that contain PCI and PCI-Express slots are optionally allocatable to either of partition granularities A and B.

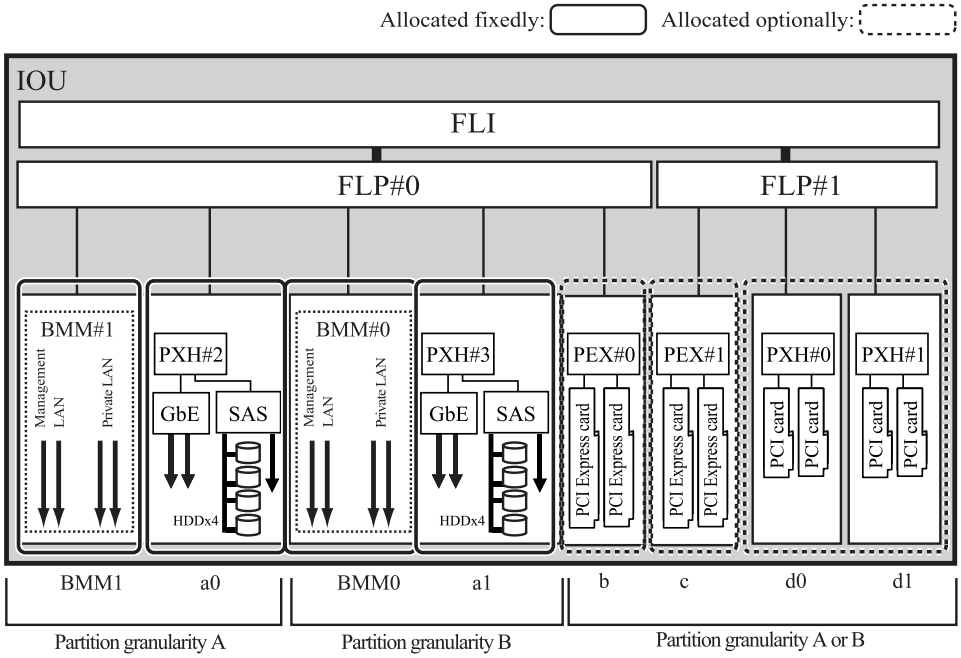


Figure 7.7 IO Unit granularities and allocation rules

Table 7.3 IO Unit allocation rules

	a0	a1	b	c	d0	d1	BMM0	BMM1
			#0, #1	#2, #3	#4, #5	#6, #7		
If IO Units are not split	A	A	A	A	A	A	A	A
If IO Units are split	A	B	A or B	A or B	A or B	A or B	B	A

Typical allocations of an IO Unit as logically split are shown below.

- Allocate all optionally allocatable blocks b, c, d0 and d1 to partition granularity A.

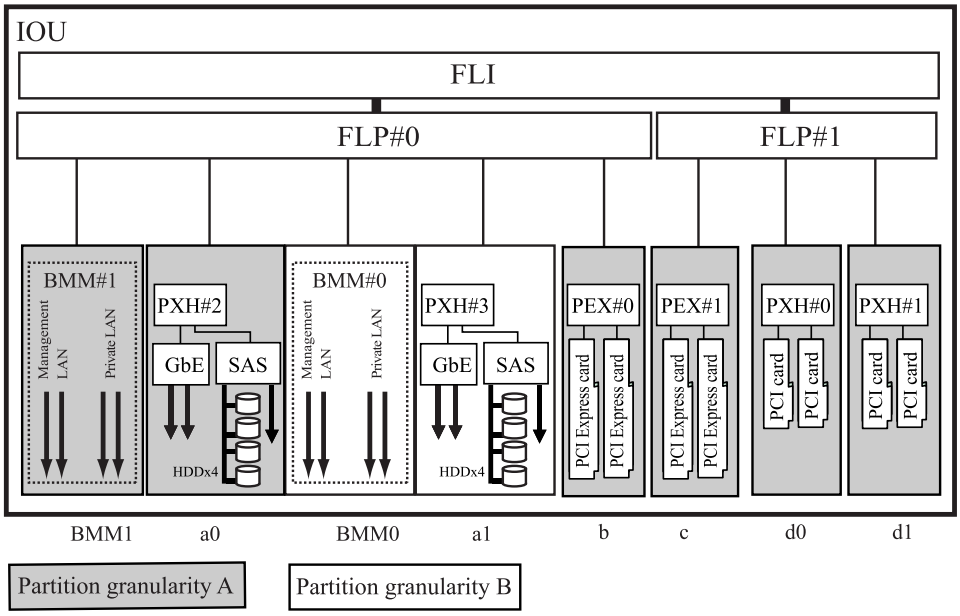


Figure 7.8 Typical IO Unit partition granularity allocations (1/3)

- Allocate one half of optionally allocatable blocks b, c, d0 and d1 to partition granularity A and allocate the other half to partition granularity B.

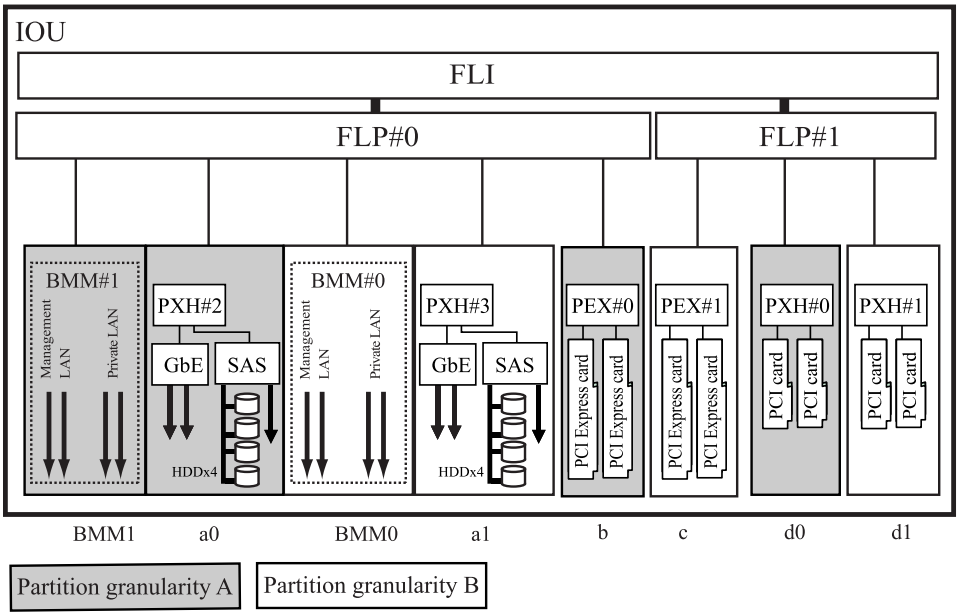


Figure 7.9 Typical IO Unit partition granularity allocations (2/3)

- Allocate optionally allocatable blocks b, c, d0 and d1 to partition granularities A and B at a ratio of 3 to 1.

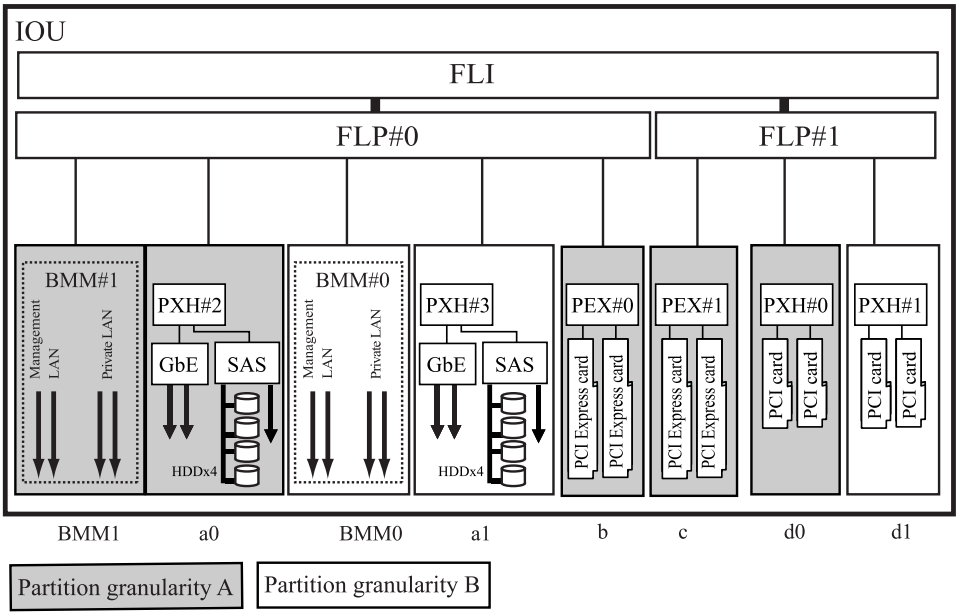


Figure 7.10 Typical IO Unit partition granularity allocations (3/3)

(2) Rules of IOX granularity allocation

This section details the procedures for allocating IOX granularities.

When an IOX is split into two granularities, some of its components are fixedly allocated to either granularity while others are allocatable to either granularity at the user's option.

IOXs are allocated to partition granularities A and B pursuant to the following rules:

- BMM0 and BMM1 that contain a BMM are allocated fixedly. BMM1 is allocated to partition granularity A, and BMM2 is allocated to partition granularity B.
- Blocks b0, b1, BMM0 and BMM1 that contain a PCI_Box connector are optionally allocatable to either of partition granularities A and B.

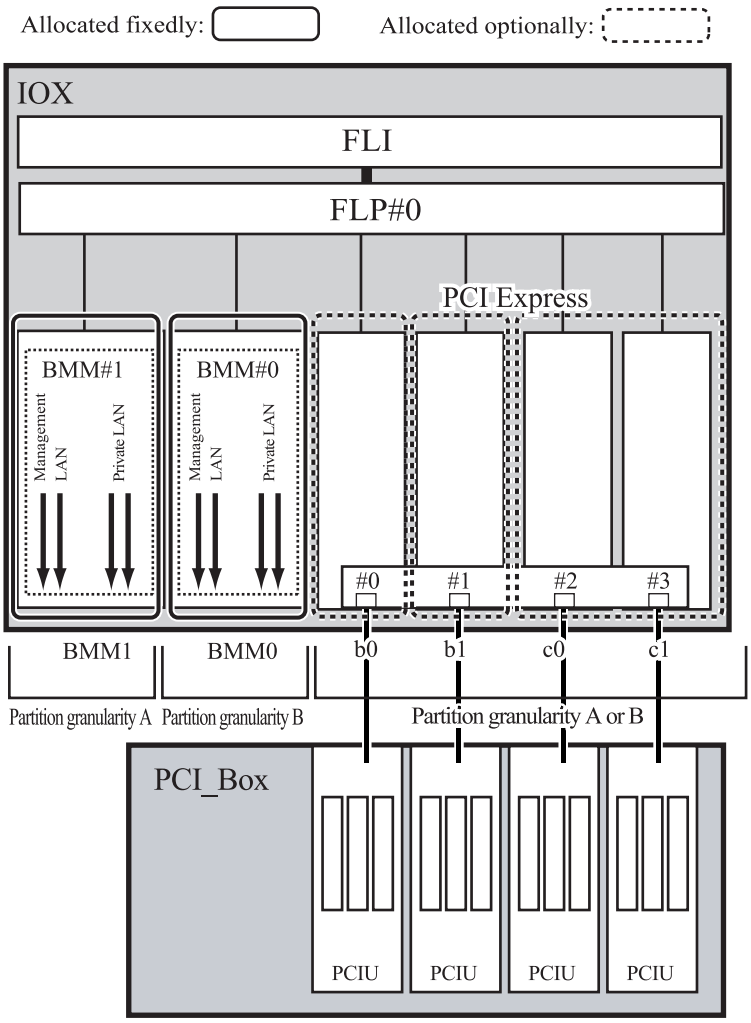


Figure 7.11 IOX partition granularities and allocation rules

Table 7.4 IOX allocation rules

	b0	b1	c0	c1	BMM0	BMM1
	slot#0	slot#1	slot#2	slot#3		
If IOXs are not split	A	A	A	A	A	A
If IOXs are split	A or B	A or B	A or B	A or B	B	A

Typical allocations of an IOX as logically split are shown below.

- Allocate all optionally allocatable blocks b0, b1, BMM0 and BMM1 to partition granularity A.

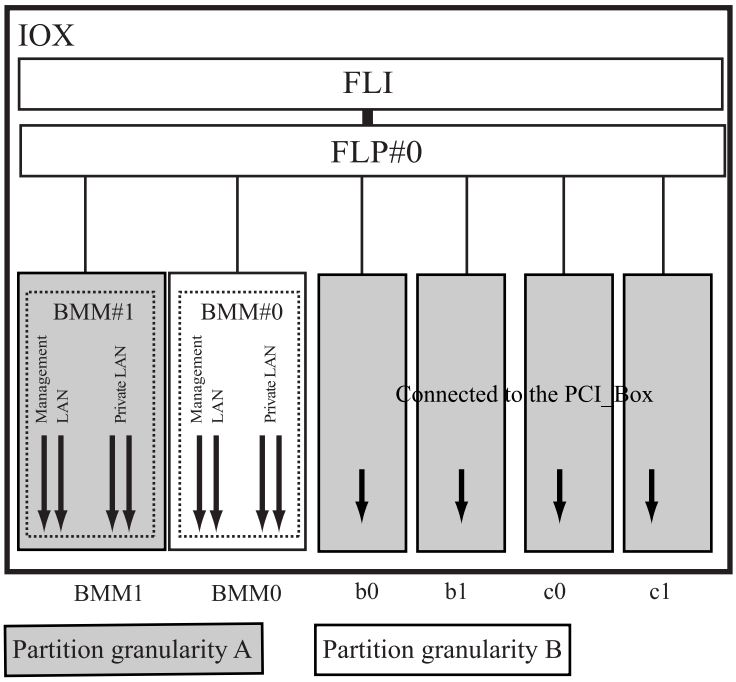


Figure 7.12 Typical IOX partition granularity allocations (1/3)

- Allocate one half of optionally allocatable blocks b, c, d0 and d1 to partition granularity A and allocate the other half to partition granularity B.

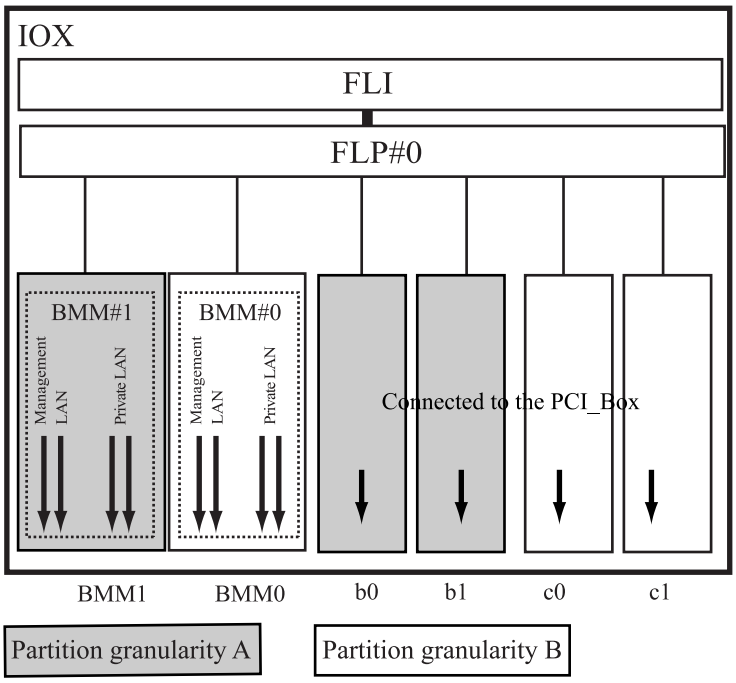


Figure 7.13 Typical IOX partition granularity allocations (2/3)

- Allocate optionally allocatable blocks b, c, d0 and d1 to partition granularities A and B at a ratio of 3 to 1.

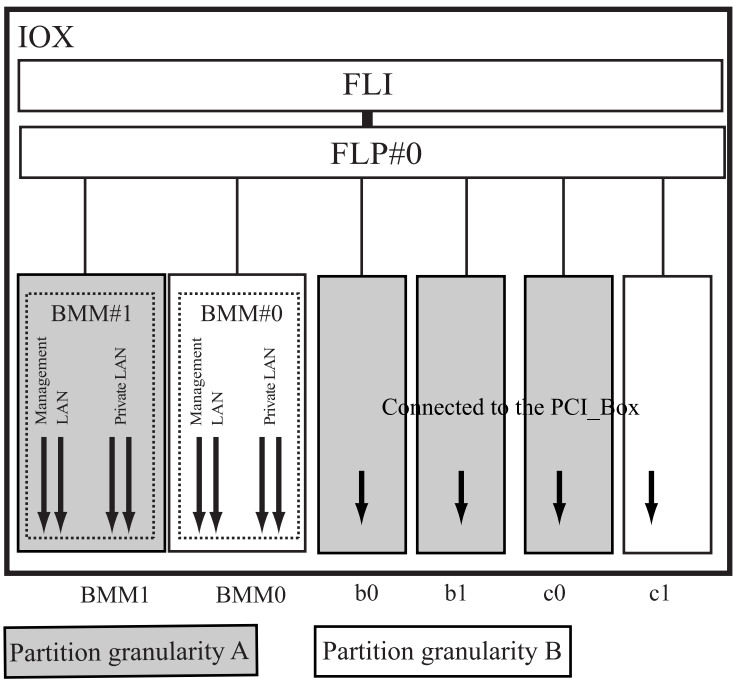


Figure 7.14 Typical IOX partition granularity allocations (3/3)

7.3.4 Allocating a PCI_Box

The PCI_Box is connected by way of an IOX. The PCIU/PEXU is allocated to the partition with which the IOX is associated.

A PCI_Box can be allocated on a PCIU basis. The PRIMEQUEST 520A/520 allows one PCI_Box to be split in up into two partitions and allocated.

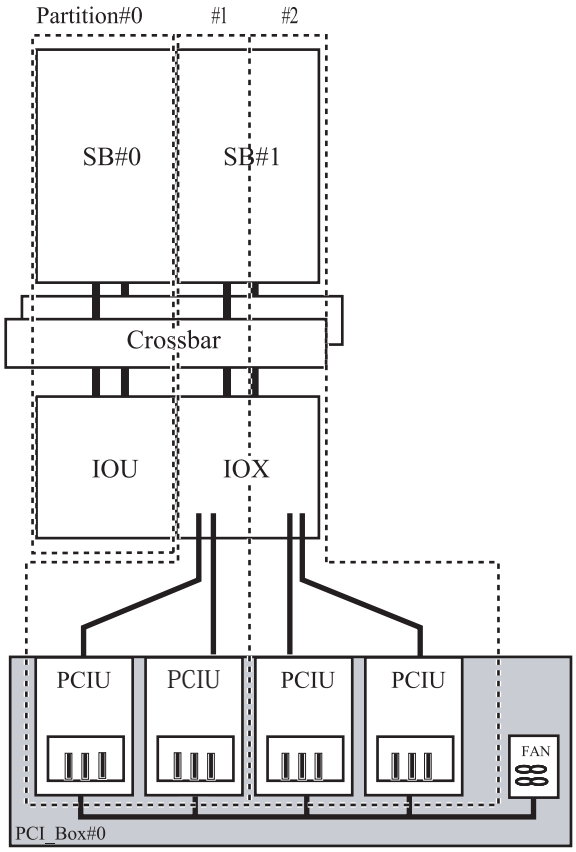


Figure 7.15 Allocation of a PCI_Box to partitions (only with PCIU)

Remarks: Split IOXs cannot be incorporated in a partition for which Extended Mirror Mode is set.

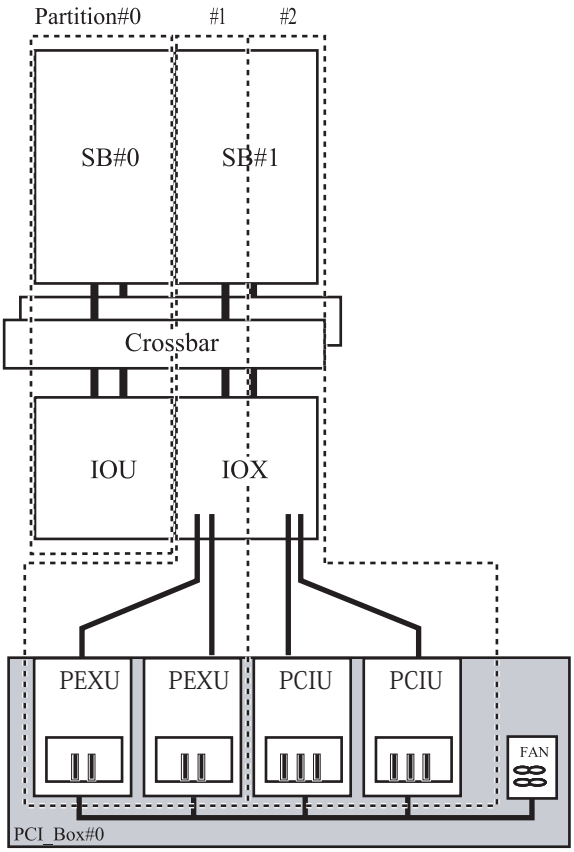


Figure 7.16 Allocation of a PCI-Box to partitions (with a combination of PCIUs and PEXUs)

Remarks: Split IOX cannot be incorporated in a partition for which Extended Mirror Mode is set.

7.4 Partition Configuration Definition

To change the configuration of the partition, define the partition by using the MMB Web-UI.

To do this, first, create a partition having a specific partition ID (PartitionName) , and then register SBs and IO Units belonging to the partition on the [Partition Configuration] window of the MMB Web-UI that is displayed by selecting [Partition].

For information on the setting procedure, see the *PRIMEQUEST 580A/540A/520A/500/400 Series Reference Manual: Basic Operation/GUI/Commands* (C122-E003EN).

Notes:

- Extended Mirror Mode cannot be set in a partition containing a split SB, IO Unit, or IOX. Split SBs, IO Units, and IOXs cannot be allocated to a partition for which Extended Mirror Mode is set.
- A split SB cannot be allocated as a reserved SB to a partition for which Extended Mirror Mode is set. Similarly, Extended Mirror Mode cannot be set in a partition to which a split SB is allocated as a reserved SB.

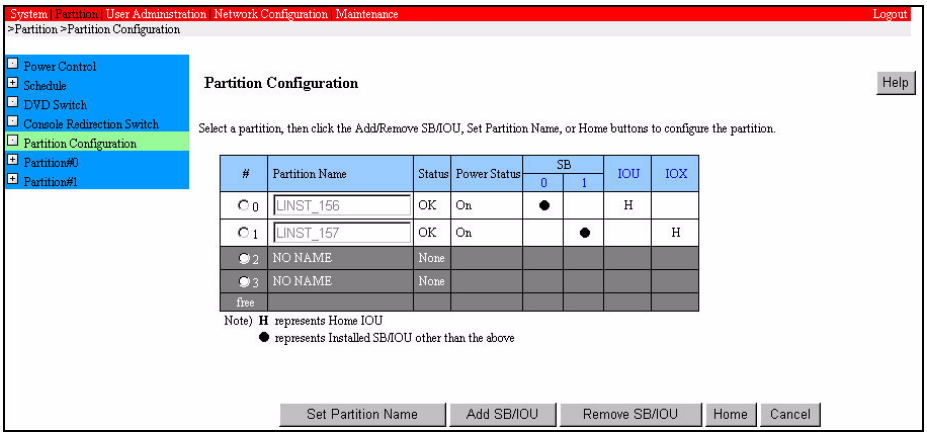


Figure 7.17 [Partition Configuration] window

7.5 Points to Consider when Determining Configurations

This section describes the points to consider when determining the configuration and operation of partitions. The partition configuration must be determined by considering the requirements of each system.

7.5.1 SB and IO Unit maintenance policy

Replacement of an SB or IO Unit involves work with partitions stopped.

7.5.2 Flexible I/O component and reserved SB

A flexible I/O component is mounted in PRIMEQUEST series machines. The flexible I/O component provides a function which separates SBs and IO Units and allows the SBs and IO Units to be combined freely. This component can connect SBs and IO Units freely in comparison with a method that allocates SBs and IO Units permanently.

Although a reserved SB does not belong to any partition, it can be incorporated in a partition to replace a faulty SB. To replace a faulty SB with a reserved SB, the reserved SB can be set as the reserved SB of the partition on the [Partition]→[Reserved SB Configuration] window of the MMB Web-UI.

If an SB in a partition becomes faulty, the system switches from the faulty SB to the reserved SB. At the same time, the connections between the faulty SB and IO Units are switched to those of the reserved SB to continue job processing. The flexible I/O component enables the system to restore only by reboot.

Mounting the flexible I/O component and reserved SBs can widely reduce the restore time necessary if an SB becomes faulty.

Notes:

- If an SB failure causes the reboot of a partition, a switch is made to the reserved SB indicated as the Reserved SB in the partition.
  If a failure occurs in one of the components (such as memory) in a duplicated component configuration in Extended Mirror Mode and the other component continues operating without interruption, switching to a reserved SB is not made.
- A split SB cannot be allocated as a reserved SB to a partition for which Extended Mirror Mode is set. Similarly, Extended Mirror Mode cannot be set in a partition to which a split SB is allocated as a reserved SB.
- An SB with 16 GB of memory mounted (4-GB DIMM x 4) cannot be mixed in the same partition with an SB having another memory configuration.
  Therefore, an SB with 16 GB of memory mounted (4-GB DIMM x 4) cannot be set as a reserved SB in a partition consisting of SBs having another memory configuration.

Remarks:
- Fujitsu recommends specifying a reserved SB that has the same memory and CPU configurations as those of the SBs configured in the target partition. If the number of CPUs used increases during switching to the specified reserved SB, additional licenses are required.
- CPUs of the same type as those on the SBs composing the partition must be mounted on the reserved SB. If the SB has a different type of CPU mounted on it, it is not used as a reserved SB.

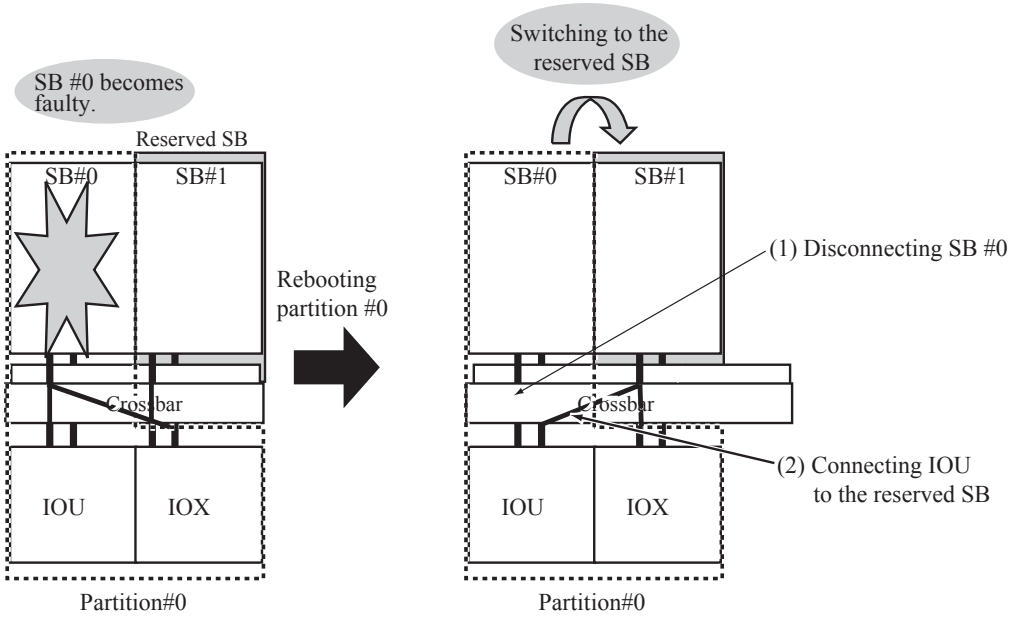


Figure 7.18 Flexible I/O component

7.5.3 Notes on System Mirror Mode operation

The PRIMEQUEST series supports Extended Mirror Mode, which duplicates the memory, crossbar, and internal chipset and gets them working in an identical way, and Standard Mirror Mode, which duplicates the crossbar (address bus) to get it working in an identical way.

Accordingly, if either duplicate copy of a component fails, the other takes over service from its failing counterpart without interruption. Operation in System Mirror Mode provides higher reliability. For more details on System Mirror Mode (Extended Mirror Mode, Standard Mirror Mode), see Section 5.2, "System Mirror Mode."

Note: Extended Mirror Mode cannot be set in a partition containing a split SB, IO Unit, or IOX. Split SB, IO Units, and IOXs cannot be allocated to a partition for which Extended Mirror Mode is set.

7.5.4 Notes on the memory configuration

The SB requires at least one CPU operating normally and one set of memory (DIMM × 4) operating normally. Consequently, if the memory degrades due to detection of an error in the memory and no set of memory operating normally is available, the entire SB is degraded. To prevent the CPU from being degraded concurrently with the SB degradation, two or more sets of memory should be mounted in each SB.

Remarks: The term "degradation" means that if a hardware error is detected by the initial diagnosis hardware executed when the power is turned on or during system operation, the system runs after disconnecting the respective hardware elements.

7.5.5 Operating environment

Two types of OSs (Linux and Windows) can be installed on PRIMEQUEST series machines. Because the OS can be selected for each partition, determine the partition configuration by considering the operation method.

7.6 Function for Changing Partition Configurations

The partition configuration of PRIMEQUEST series machines is changed using the [Partition Configuration] window of the MMB Web-UI that is displayed by selecting [Partition]. For details of the operation, see the *PRIMEQUEST 580A/540A/520A/500/400 Series Reference Manual: Basic Operation/GUI/Commands* (C122-E003EN).

Remarks: A change to the partition configuration is applied by power-off and subsequent power-on in units of partitions. The change is not applied by a partition reboot. Be sure to power-off the partition and power it on again.

CHAPTER 8 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.

- The basics of clustering
- Cluster Functions Provided in Linux Environments
- Cluster Functions Provided in Windows Environments
- Points to be Considered at Cluster System Installation

8.1 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.

In PRIMEQUEST series machines, one cabinet can be divided into multiple logical servers by the partition function so as to implement a cluster system, where each partition acts as one node.

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.

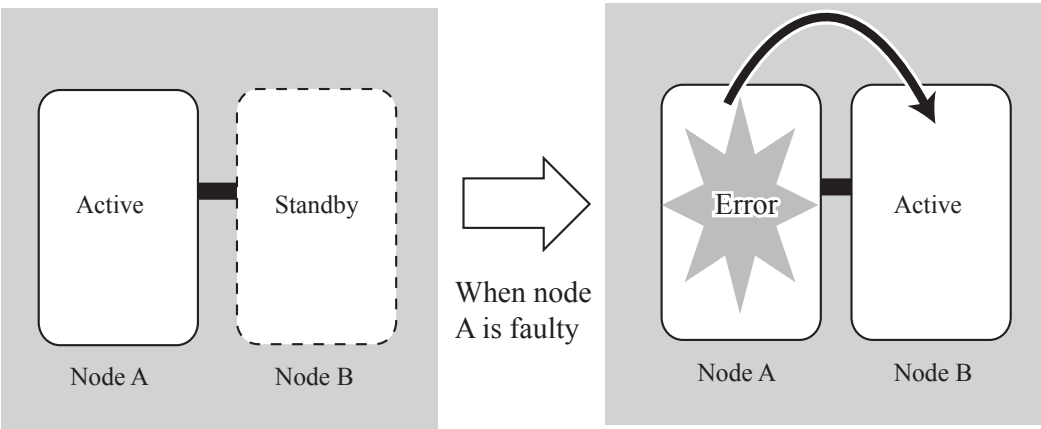


Figure 8.1 Example for 1:1 standby

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.

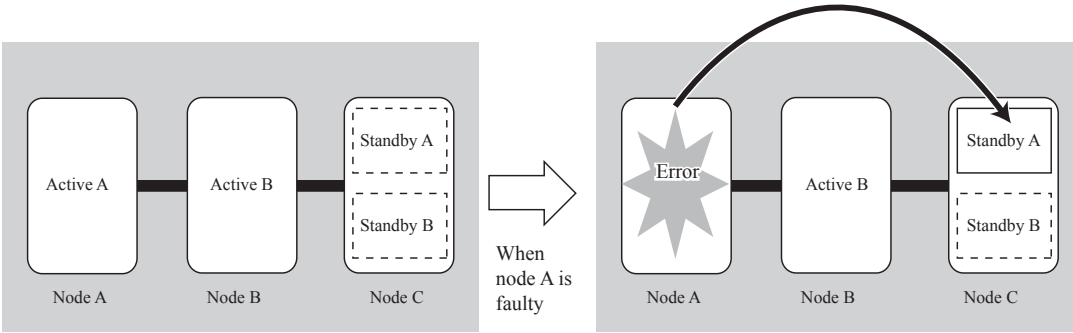


Figure 8.2 Example for N:1 standby

Mutual standby

Each active node also operates as a standby node for another node. If an error occurs in one node, the other node takes over job processing. During normal operation, all nodes can be used, but when a malfunction occurs, the job processing for two nodes needs to be performed on a single node. Therefore, the node needs to be constructed in such a way that a single node can perform the processing for the jobs on both nodes.

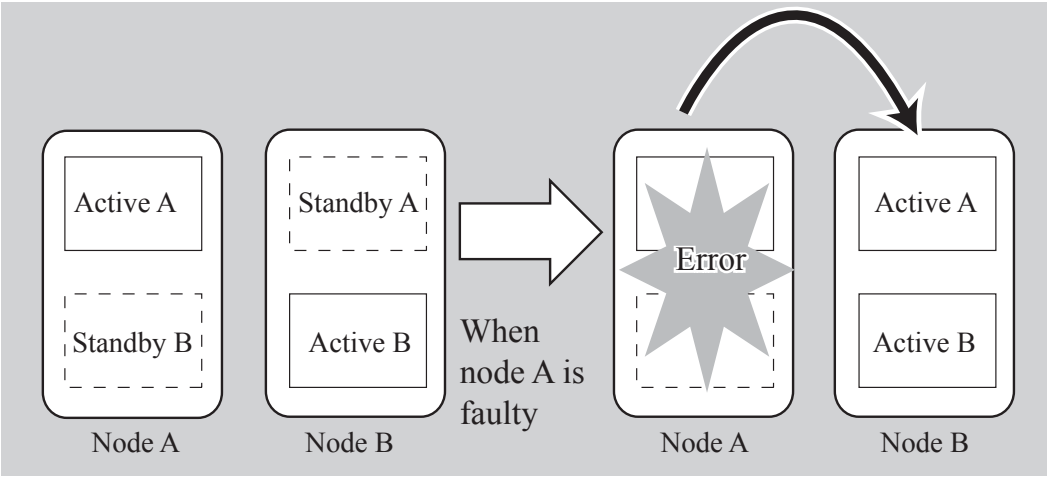


Figure 8.3 Example for mutual standby

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.

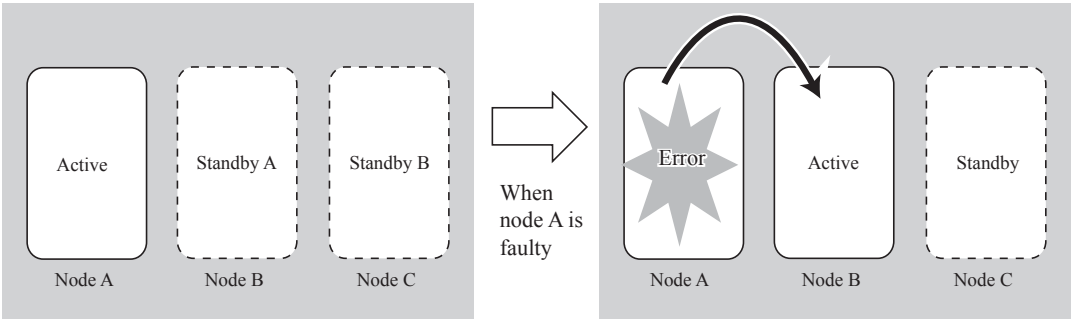


Figure 8.4 Example for a cascade

8.2 Cluster Functions Provided in Linux Environments

8.2.1 Cluster Configurations

For the PRIMEQUEST 520A/520/420 cluster configuration between cabinets, the cabinet is treated as one node or multiple nodes, 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 compared with the intra-cluster configuration, thereby increasing reliability.

The following example shows a connection of shared disk units by fibre channel. The system disk is connected to two IO Units and mirroring is applied by the PRIMECLUSTER Global Disk Services (GDS). Two separately mounted IO Units increase the resilience against failures.

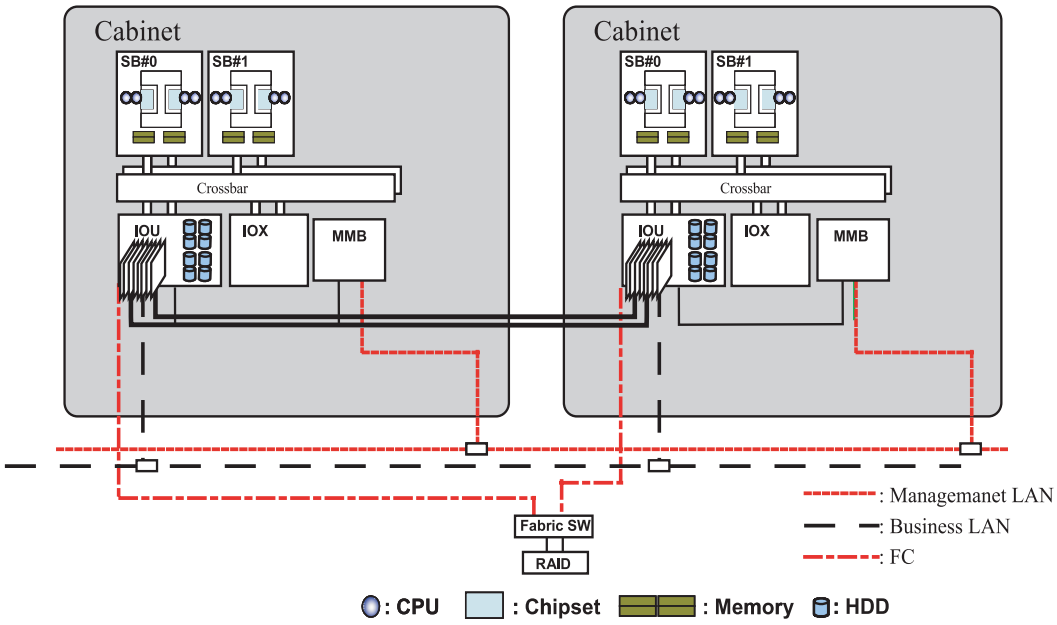


Figure 8.5 Inter-cabinet cluster system configuration

8.2.2 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 an operating partition or in the cabinet, high availability can be implemented by having a standby partition and –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 partition 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 partition through a private LAN.

8.3 Cluster Functions Provided in Windows Environments

Windows Server 2003 supports cluster functions as standard features of the operating system. This section explains Microsoft Cluster Service (MSCS), which is provided by Windows Server 2003.

A cluster system consisting of up to eight partitions (*1) can be configured with Windows Server 2003. Each server (partition) in a cluster is called a node. When Windows is used with PRIMEQUEST, the most basic cluster system is a two-node inter-partition cluster using a shared disk. This section explains the inter-partition cluster as the basic configuration.

*1: The partitions referred to in this section may be contained in a single or multiple PRIMEQUEST cabinets.

Note:

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

8.3.1 Basic cluster configuration

The following figure shows a typical two-node cluster configuration.

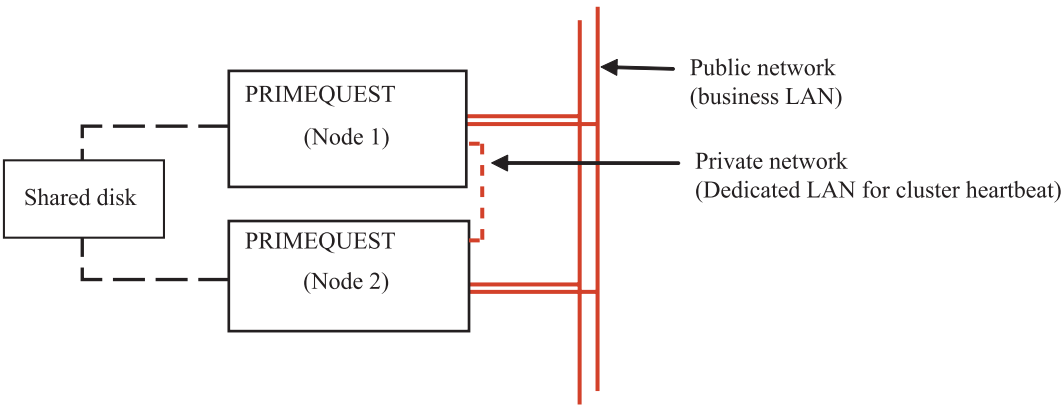


Figure 8.6 Basic cluster configuration (Windows)

- Business application software
  Business application software must be installed on each node used for business operations and on each standby node.

- 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. In contrast, the LAN used for business operations is called a public network.

8.3.2 Virtual server

A virtual IP address instead of the IP address of a node is used as the IP address for access to business application software, because it is not affected 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.

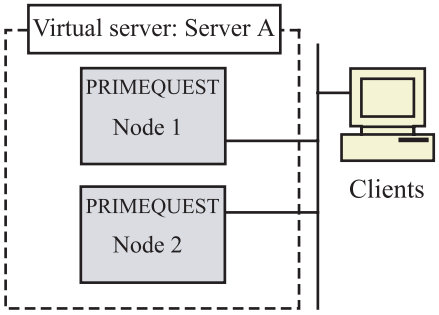


Figure 8.7 Virtual server

8.3.3 Scope of remedies

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.

8.3.4 Notes on construction

MSCS construction requires a separate environment for the Windows Server 2003 directory service (Active Directory). You can use the cluster construction service for the Active Directory and MSCS construction. For details, contact your Fujitsu certified service engineer.

8.4 Points to be Considered at Cluster System Installation

This section discusses points to be considered before determining a cluster configuration. The cluster configuration should be determined after the requirements of the individual systems are considered.

Interconnect (network connecting nodes) disposition

In the PRIMEQUEST cluster configuration, all nodes must be connected by a dedicated LAN. The network connecting nodes is called interconnect. Multiplexing the interconnect increases the availability of the system. The business LAN must be installed separately from the cluster interconnect.
The PRIMEQUEST interconnect should be connected to either of the following types of NIC card:

- NIC card added to the PCI slot in an IO Unit
- NIC card added to the PCI slot in the PCI_Box

Remarks: Never use the management LAN for interconnect communication because it is used only for controlling the PRIMEQUEST system.

The following figure shows a configuration in which NIC cards are added to the PCI slot in the IO Unit.

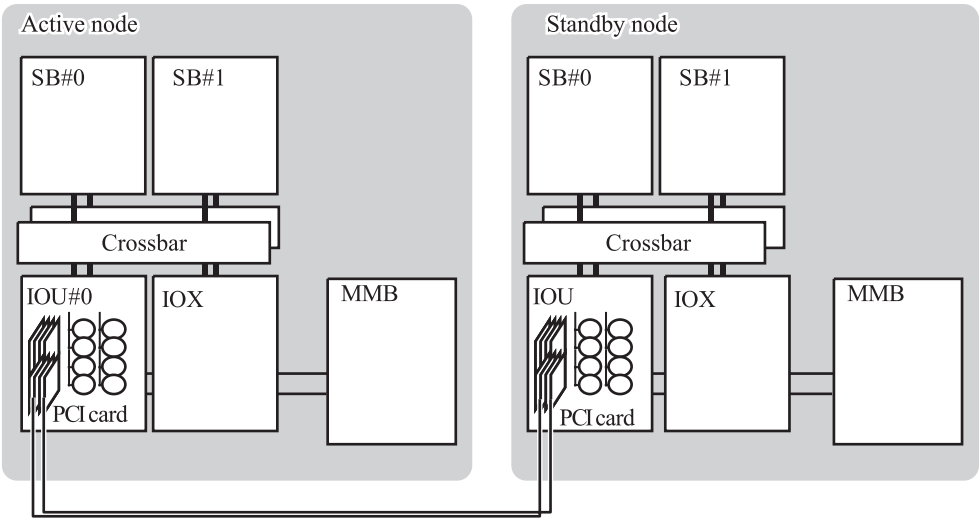


Figure 8.8 Position of the interconnect

Remarks: Do not use the spanning tree protocol (STP) for a switch to be used as interconnect. Otherwise, the system will enter the non-communication state at reconfiguration of a link, thereby disabling interconnect communication.

System Mirror Mode and clustering

The PRIMEQUEST series supports Extended Mirror Mode and Standard Mirror Mode. Extended Mirror Mode splits the available memory, crossbar, and internal chipset into two halves each and gets them working in an identical way. Accordingly, if either duplicate copy of a component fails, the other takes over the service from its failing counterpart without interruption. Standard Mirror Mode duplicates the crossbar (address bus) to get it working in an identical way. Operation in System Mirror Mode provides higher reliability.

Thus the PRIMEQUEST series increases the availability of hardware by operation in System Mirror Mode. However, System Mirror Mode does not support software redundancy.

Installation of a cluster system can increase the availability of the entire system also as far as software is concerned.

By constructing a system using both System Mirror Mode and clustering functions, the availability of the system can be improved even further.

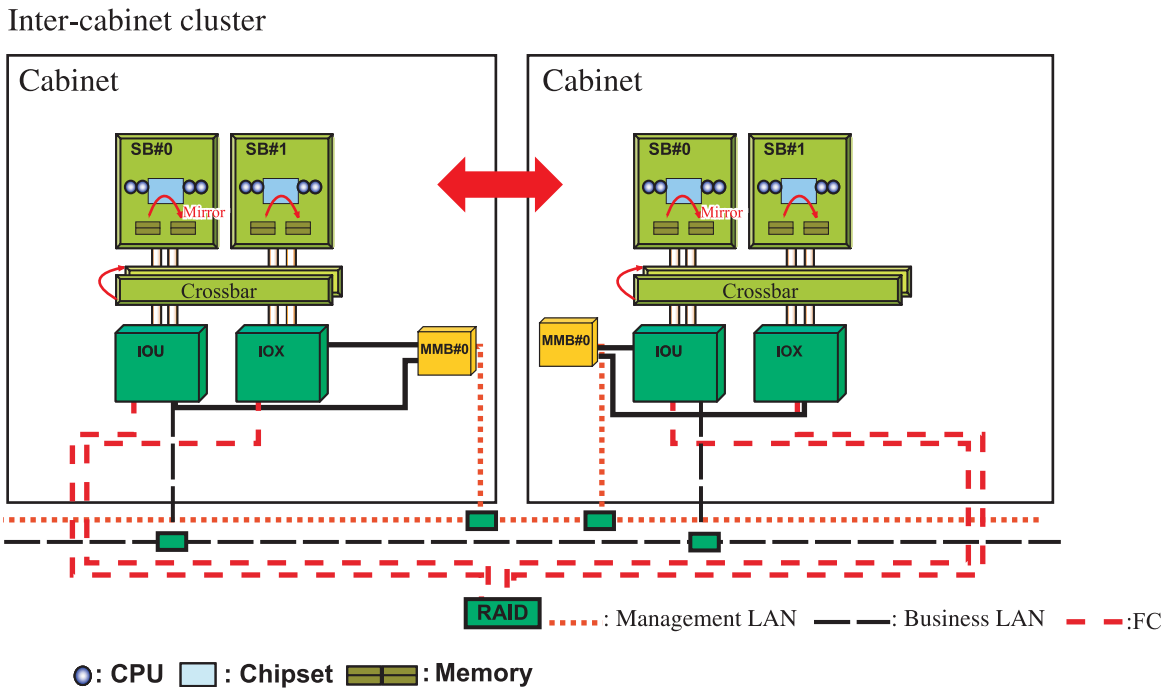


Figure 8.9 Extended Mirror Mode and cluster configuration

CHAPTER 9 Backup and Restoration

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

9.1 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.

9.2 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.

9.2.1 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 9.1) 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.

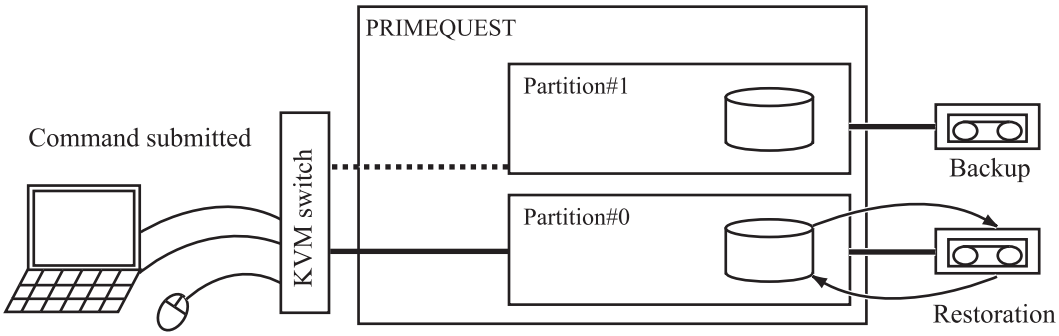


Figure 9.1 Local backup using a standard utility
(when the KVM console is used)

- 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 (Partition#0 in Figure 9.2) 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 (Partition #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 Partition #0 is backed up, it is necessary to use NTBackup on the remote host side (Partition #n) for the backup and restoration work.

First, perform remote-login to Partition #n from Partition #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.

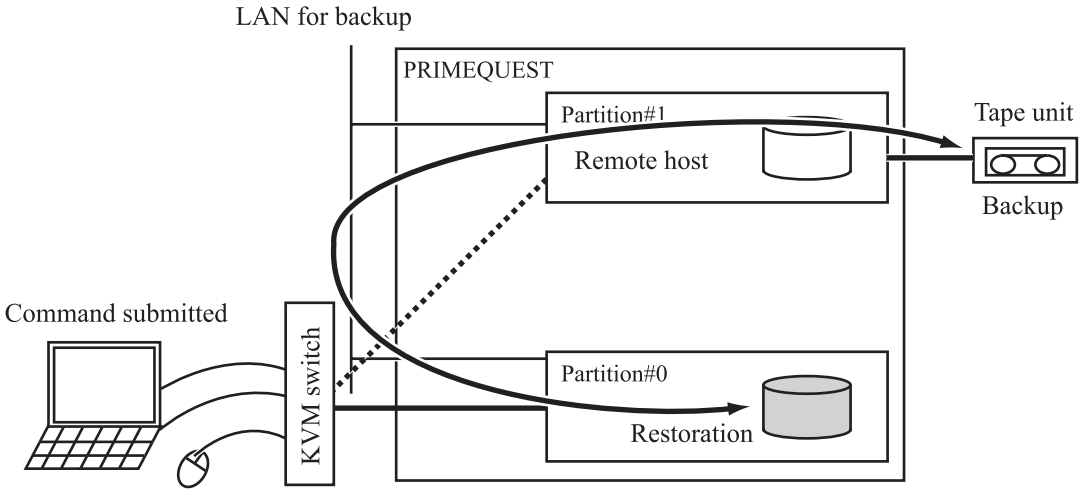


Figure 9.2 Backup to the remote host by using a standard utility (when using the KVM console)

9.2.2 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)

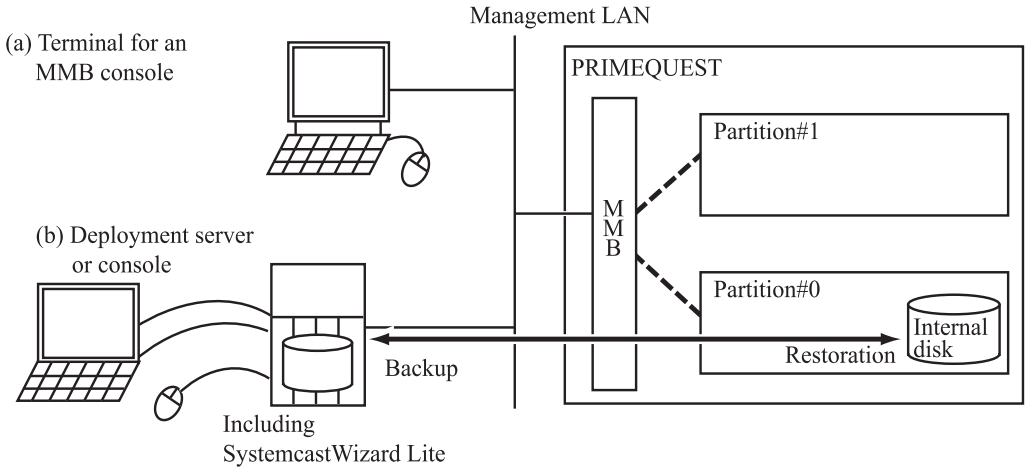


Figure 9.3 Backup using SystemcastWizard Lite

- 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.)
- If SystemcastWizard Lite is used for backup and restoration, up to only four partitions can be used in parallel.

9.2.3 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.

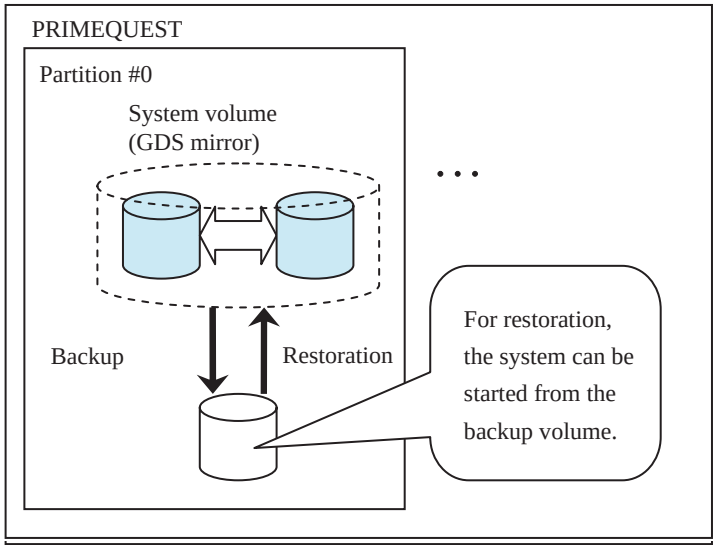


Figure 9.4 System volume backup using PRIMECLUSTER GDS Snapshot

- 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.

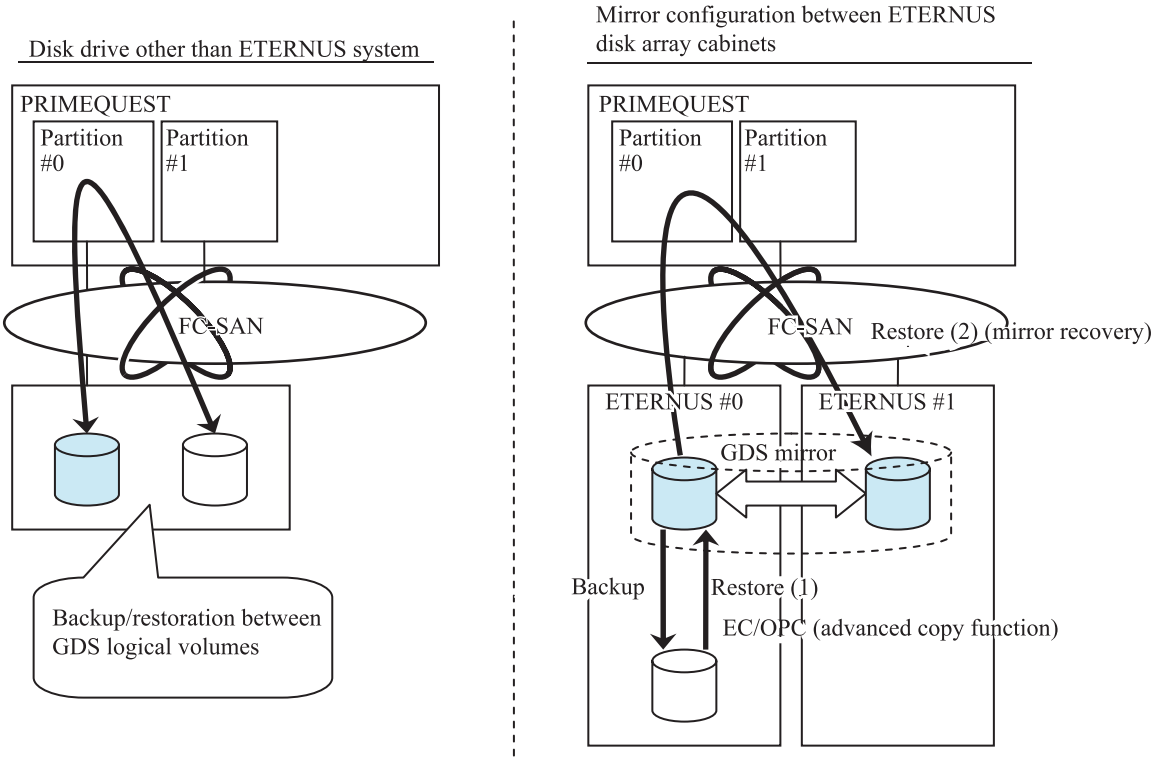


Figure 9.5 Data area backup and restoration using PRIMECLUSTER GDS Snapshot

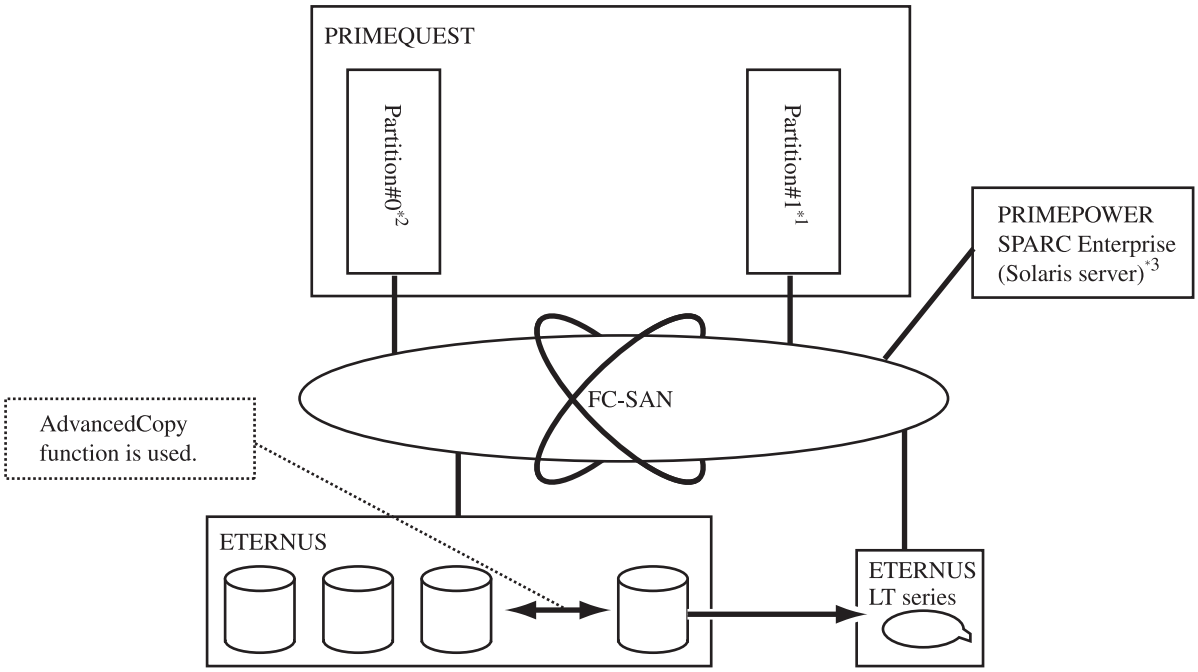
9.2.4 Backup software (ETERNUS SF ACM, VERITAS NetBackup, NetVault, NetWorker)

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 9.6 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)

- 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.

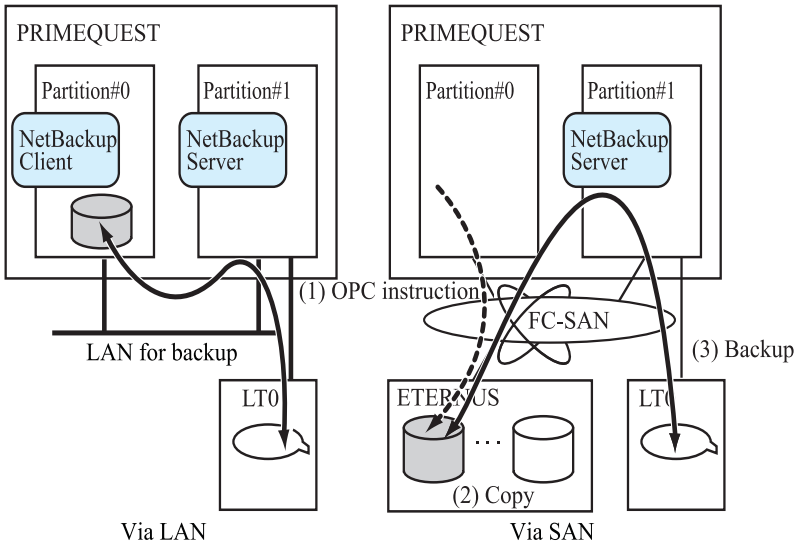


Figure 9.7 Data backup using VERITAS NetBackup

- 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.

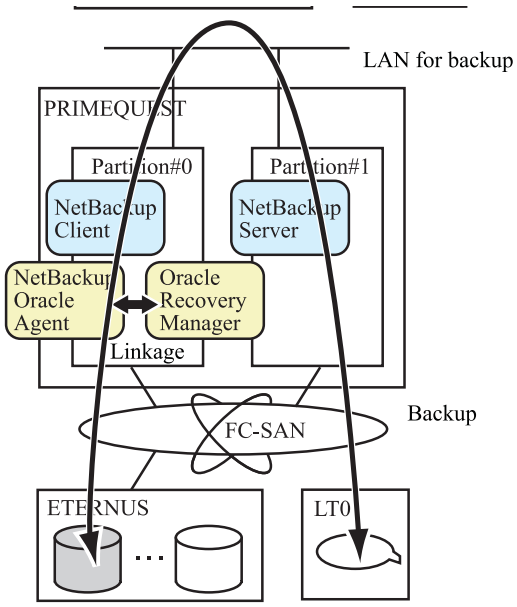


Figure 9.8 Online backup of a DB using VERITAS NetBackup

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

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

- 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.

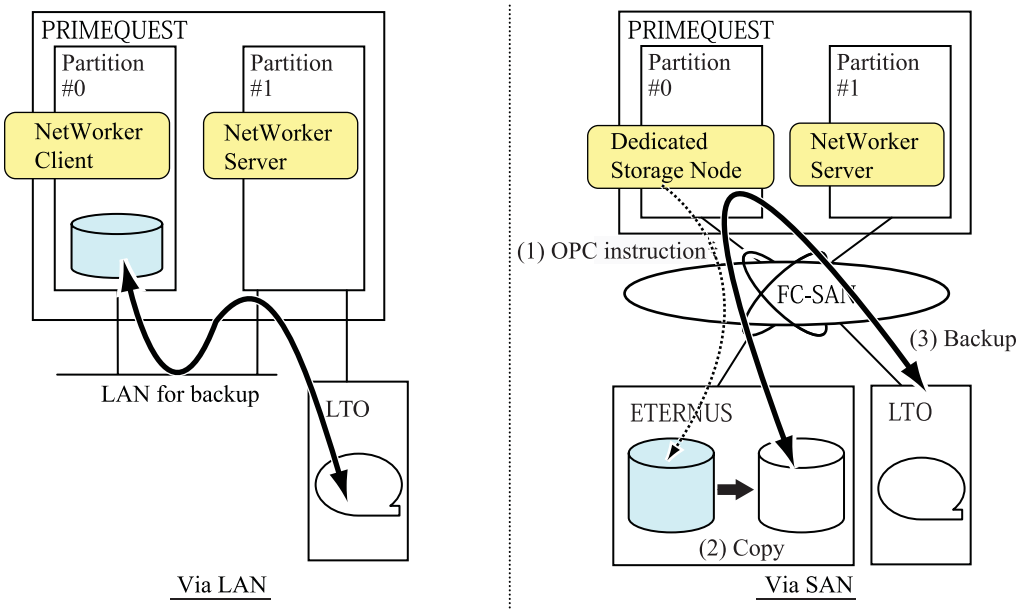
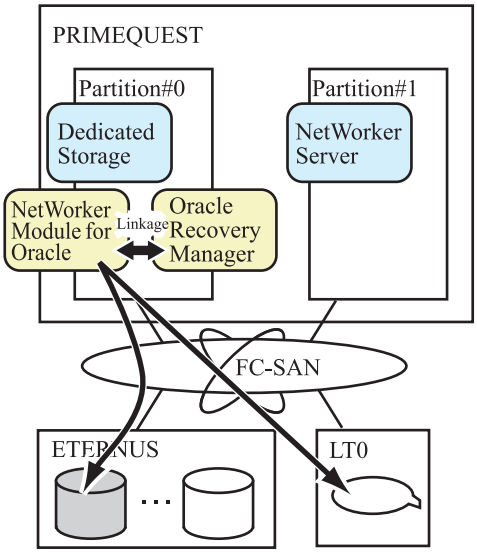


Figure 9.9 Data backup using NetWorker

- 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.

9.3 Comparison of Backup/Restoration Methods

Table 9.1 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 9.1 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 partitions.
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 partitions. • It is possible to manage a mixed platform environment by using one server.

CHAPTER 10 Considerations for Maintenance

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

- Estimating the Dump-use Area
- Estimating the Dump-use Memory

10.1 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)

10.1.1 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 10.1, "Areas used to collect dump data" lists the sizes of the areas required for a dump device.

Table 10.1 Areas used to collect dump data

No.	Function name	Size	Remarks
1	diskdump	Amount of installed memory (*1) + 512 MB	

*1 When Extended Mirror Mode is used, the Linux system can only recognize one half of the amount of the actually installed memory. The amount of memory above indicates the amount of memory that Linux can recognize.

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 10.2, "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 10.2 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

When the system is restarted, the dump information that has been collected by diskdump on a dump device is saved after it is automatically converted to a format that a dump tool can reference. Therefore, an area with the size indicated below must be created for each type of dump information to be saved.

The dump saving areas must be in /var/crash.

Area size required per saved dump = Amount of memory installed + 1 GB

When creating a dump saving area, such as /var/crash, as a dedicated disk partition, estimate the area size and increase the estimated value by 10% in consideration of the file system management area.

10.1.2 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

At the time the system becomes faulty, dump information is directly saved in this area by kdump. Therefore, an area with the size indicated below must be created for each type of dump information to be saved.

Area size required per saved dump = Amount of memory installed + 1 GB

The areas must be created on a disk not managed by PRIMECLUSTER GDS.

When creating a dump saving area as a dedicated disk partition, estimate the area size and increase the estimated value by 10% in consideration of the file system management area.

10.2 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.

10.2.1 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.

*2 This is supported only in Japan. Overseas support is planned for Red Hat Enterprise Linux AS (v.4 for Itanium) Update3 and later versions.

*3 This is supported only in Japan and the Asian region.

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

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

BP (Backplane)

Abbreviation for backplane. Constituting a memory and I/O interconnect (crossbar), the backplane is connected to SBs, IO Units, and other devices. Chip sets are mounted on the backplane.

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.

Crossbar

This concept covers the address crossbar and data crossbar that transfer data and control the data transfer between SBs and IO Units. Memory and I/O interconnect has the same meaning as crossbar.

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.

IO Unit

I/O control unit that contains PCI card slot, HDD, SCSI controller, GbE controller, and other I/O interfaces

IOX (IO Extension Unit)

Board that is used for I/O interface expansion and contains PCI_Express ports for PCI_Box and BMM connections

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 partitions 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, partition management, 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

Partition

System equipped with the functions of a processing unit. Each partition contains software resources such as an OS and applications as well as hardware resources such as SBs and IO Units.

PCI_Box

Device used for PCI slot expansion

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 I/O Units 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 partitions

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

Reserved SB

Standby SB reserved for possible failures

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

SB (System Board)

Board on which a CPU and memory are mounted

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

SERDES (Serializer Deserializer)

Parallel-to-serial converter (from low speeds to high speeds)

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

A

access control 4-20
agent message log. 4-31
Alarm-LED 2-21
alert message, convention for vi
allocation PCI_Box to partition 7-18, 7-19
all-partition communication
  disconnection mode 4-8

B

backup 3-8
  general method of backup and
  restoration 9-1
  importance of backing up data 9-1
backup software
  NetBackup. 9-9
  NetVault. 9-9
  Softek ACM 9-9
  Softek TSM 9-9
backup/restoration method,
  comparison of. 9-15
BMC firmware 3-2
BP 2-20
business LAN. 4-10
  redundancy of business LAN 5-10

C

cabinet 2-4
card cassette in PEXU 2-30
chipset 1-8
cloning software
  SystemcastWizard 9-5
cluster configuration. 8-5
clustering software 1-25
clustering, basic of 8-1
COM port 2-10
compatible PCI card. 2-12, 2-30
connection requirement 2-33
console redirection. 4-3, 4-18
CPU 2-6
crossbar 2-20

D

data area backup using PRIMECLUSTER
  GDS Snapshot. 9-8
data backup using NetWorker 9-13
DDR2 SDRAM 2-7
degradation function 5-14
degradation reservation. 5-14
driver 3-5
dump 9-2
  area used to collect dump data. 10-2
  area used to save dump data 10-3
dump-use area, estimating 10-1
dump-use memory, estimating 10-4

E

EFI: Extensible Firmware Interface 3-2
environmental requirement for
  using product. vi
error
  error correction 2-7
  record log 4-31
ETERNUS multipath driver 5-8
ETERNUS SF AdvancedCopy
  Manager 9-9
Expansion File Unit 2-35
Extended Mirror Mode and cluster
  configuration 8-11

F

firmware. 1-23, 3-2
  BMC firmware. 3-2
  MMB firmware 3-3
fjsnap. 3-6
flexible I/O component. 7-21
flexible I/O mechanism 1-16
function
  MMB function 4-12
  PSA function. 4-25

G

GUI
  operation management. 4-26

H

hardware
configuration display 4-14
configuration management 4-28
hardware error monitoring 4-29
hardware technology 1-8
monitor. 4-15
hardware component 1-4
BP 2-20
crossbar. 2-20
IO Unit 2-9
MMB 2-17
OPL 2-21
PCI_Box. 2-23
SB 2-5
HDD
redundancy of HDD 5-6
high-speed switch method
duplication of transmission line
between servers 5-11
hot maintenance function 1-20
hot swappable component 6-2

I

independent resource expansion 1-14
interface
IO Unit external interface 2-10
MMB external interface 2-18
inter-partition communication
connection mode 4-7
inter-partition communication
disconnect mode. 4-7
IO Unit 2-9
IP address set by OS on partition . . . 4-11
IO Unit and IOX partition granularity. 7-8
IO Unit granularity and allocation rule. . . 7-12
IOX and PEXU connection 2-32
IOX and PEXU connection
(prohibited) 2-32
IOX and PEXU connection example. . . . 2-31
IOX partition granularity and
allocation rule 7-15
IP address 4-10
IP address set by OS on partition. 4-11
IPF 1-9

L

LAN
business LAN 4-10
configuration 4-3
management LAN. 4-4
private LAN 4-9
large-scale SMP 1-8
LED display 2-21
linked software
backup/restoration. 3-8
PRIMECLUSTER 3-6
PRIMECLUSTER GDS. 3-6
PRIMECLUSTER GLS 3-7
restoration 3-8
Systemwalker 3-7
Linux 1-24, 3-4
load balance 5-8
load distribution 5-8
Location-LED. 2-21
log
analysis 4-31
collection 4-31
display. 4-31

M

maintenance
firmware maintenance. 4-22
maintenance operation 4-32
maintenance policy
SB and IO Unit 7-21
management board 1-3, 2-17
management LAN 4-4
memory 2-7
memory configuration, note on 7-23
memory interleave 2-8
middleware product 1-26
minimum redundancy 5-6
mirror mode operation, note on 7-23
MMB 1-3, 1-22, 2-17
function 4-12
IP address. 4-11
MMB external interface 2-18
MMB firmware 3-3
MMB use environment setting 4-19
MMB-Ready-LED. 2-21
mounting DIMMs, note on 2-8

mounting requirement 7-3
IO Unit and IOX. 7-11
IO Unit or IOX 7-3
SB 7-9
SBs 7-3
mounting requirement
(when XPAR installed) 7-9
MTTR. 1-17
reduced MTTR 1-19

N

NetVault 9-12
network environment setting 4-19
NetWorker 9-13
NIC switch method
duplication between server and hub and between switches on same network 5-12
node. 4-33, 8-1
NTBackup 9-2
NTP
NTP server 4-21

O

operating environment 7-23
operation management GUI. 4-26
operation management software
linkage 4-35
OPL 2-21
OS 3-3
Linux 3-4
Windows 3-5
OS console 4-27
OS information display. 4-28
OS input and output. 4-18
OS-standard utility 9-2

P

PAL: Processor Abstraction Layer 3-2
partition 1-3
configuration definition 7-20
configuration display 4-17
function for changing configuration . . 7-24
point to consider when determining configuration 7-21
setting 4-17
partition function. 1-13
partition granularity 7-7
partitioning function of the PRIMEQUEST 520A/520 with XPAR installed 7-6
partitioning function overview 7-1
PCI card
compatible PCI card 2-12, 2-28
mounting condition of PCI card. 2-13
PCI card mounting condition 2-13
PCI card cassette 2-28
PCI slot 2-10, 2-27
PCI slot in PEXU 2-29
PCI_Box 2-23
allocating 7-18
PCI_Box allocation 7-4
PCI-Express 2-11
PCIU 2-27
PCIU and PEXU
connection/combination. 2-31
PCIU and PEXU in one partition 2-34
PEXU 2-29
physical configuration
PRIMEQUEST 520A/520/420 2-2
platform, operation on different 1-15
power control. 4-22
Power-LED 2-21
PRIMECLUSTER 3-6, 8-6
PRIMECLUSTER GDS 3-6, 5-6
redundancy of HDD (use of PRIMECLUSTER GDS) 5-6
PRIMECLUSTER GDS Snapshot. 9-7
PRIMECLUSTER GLS 3-7, 5-10
redundancy of business LAN (use of PRIMECLUSTER GLS) 5-10
PRIMEQUEST
feature 1-7
PRIMEQUEST conceptual diagram 1-3
PRIMEQUEST management LAN,
IP address of 4-10
PRIMEQUEST partitioning function 7-3
PRIMEQUEST System Disk Mirror for Windows 3-7
PRIMEQUEST-dedicated chipset 1-8
private LAN 4-9
product handling xi
product lineup 1-5
PSA
function. 4-25
system management software 3-5

PSDM 3-7
PSU 2-22, 6-7
PSU exterior view 2-22
PSU specification 2-22

R

reduced mean time to repair 1-17
redundancy
hardware redundancy 1-20
redundancy of component 5-2
REMCS linkage 4-32
requirement on combination of
PCIU and PEXU 2-33
reserved SB 7-21
restoration 3-8
restore 9-2
setting information restore 4-23
RHEL5 (IPF) 10-3, 10-4
RHEL-AS4 (IPF) 10-1

S

safety, note on ix
SAF-TE function 4-32
SAL: System Abstraction Layer 3-2
sample partition configuration
(logical splitting) 7-6
SAS 2-9
save
setting information save 4-23
SB 1-3, 2-5
SB and IO Unit maintenance policy 7-21
SB and IO Unit mounted and PSU
configuration number 2-22
SB partition granularity 7-7
SB partition granularity and
allocation rule 7-10
SBs: system boards 1-4
scheduled operation 4-22
secure socket layer 4-20
setting
MMB use environment setting 4-19
setting information restore 4-23
setting information save 4-23
simple network management protocol . . 4-35
SMP 1-8, 1-9
snapshot software 9-7
SNMP 4-35
software
linked software 3-6
software supplied with
PRIMEQUEST A-1
supplied software 3-5
software configuration 3-1
software mirroring 5-6
software technology 1-23
splitting SBs, IO Units, or IOXs when
XPAR installed 7-6
SSL 4-20
support 4-20
Standard Mirror Mode 1-10
storage redundancy
external storage redundancy 5-8
STP 8-10
structure and contents of manual i
supplied software 3-5
symmetric multi-processing
technology 1-9
system board 2-5
system event log 4-31
system interconnect 1-4
system management
basic mechanism 4-1
System Mirror Mode 1-8, 5-3
improved business continuity with 1-12
notes on operation 7-23
System Mirror Mode and clustering 8-10
system volume backup using
PRIMECLUSTER GDS Snapshot 9-7
SystemcastWizard Lite 3-5, 9-5
Systemwalker 3-7

T

text convention iv
time synchronization 4-21
typical IO Unit partition granularity
allocation 7-13, 7-14
typical IOX partition granularity
allocation 7-16, 7-17

U

update
firmware update 4-22
USB port 2-10
user privilege management 4-13

V

VERITAS NetBackup 9-11
VGA port 2-10
VLAN
    VLAN function of management LAN . . 4-7

W

Windows 1-24, 3-5

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-B009-07EN
Publication Name: PRIMEQUEST 520A/520/420 SYSTEM DESIGN GUIDE

Your Comments:

Page	Line	Comments
Reply requested: o Yes o No		

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

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

Technical level: o Too detailed o Appropriate o 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