

# PRIMEQUEST 1000 Series: High-Reliability, Mission-Critical Intel Architecture Server Supporting Social Infrastructure Systems

● Motoyoshi Hirose   ● Keiji Sato   ● Masahiro Kurita   ● Yasufumi Honda

As we enter the age of Cloud computing, server platforms must provide not only high reliability and high availability but also virtualization as a base technology for achieving effective use and efficient operation of information and communications technology (ICT) resources. In 2005, Fujitsu applied its know-how and experience accumulated over many years of mainframe development to develop the Mission-Critical Intel Architecture (IA) Server PRIMEQUEST series of servers based on open IA technology and Linux/Windows operating systems. This series is being used in a variety of social infrastructure systems that are helping to open up the market for open mission-critical systems. Now, with the aim of furthering the evolution of this series, Fujitsu has developed the PRIMEQUEST 1000 series of servers that enhances virtualization support, provides long-term support, and strengthens green ICT support while inheriting the features of the existing PRIMEQUEST series. It has been marketing this new series since March 2010. This paper describes the objectives and features of the PRIMEQUEST 1000 series and associated technologies.

## 1. Introduction

Fujitsu's server strategy is to continuously provide mission-critical servers applying high-reliability, high-quality, high-performance, and leading-edge technologies to optimize social infrastructure systems and enterprise data centers. Targeting the world of open mission-critical systems, the Mission-Critical Intel Architecture (IA) Server PRIMEQUEST series of servers has been developed to incorporate both the reliability and robustness of mainframe/UNIX servers and the versatility and economy of Linux/Windows servers. Since first going on sale, the PRIMEQUEST series has received the IT Japan Award 2010 Award from the Minister of Economy, Trade, and Industry (Grand Prix) and has been adopted in a number of social infrastructure systems and large-scale and mission-critical enterprise systems such as the "arrowhead" trading system of the Tokyo Stock

Exchange (IT: information technology).

More recently, the marketing of Fujitsu's new PRIMEQUEST 1000 series of servers has begun in conjunction with Intel's announcement of its Xeon processor 7500 series on March 31, 2010.

This paper begins by describing the development objectives and features of the PRIMEQUEST 1000 series. It then describes functions for achieving high-reliability and high availability, functions supporting green information and communications technology (ICT), operation and management functions, long-term support, virtualization support, and an OSIV/XSP (name of Fujitsu operating system; XSP: extended dimension system products) operation mechanism.

## **2. PRIMEQUEST 1000 series development objectives and features**

The main application fields of the existing PRIMEQUEST series have been large-scale, mission-critical systems with a focus on financial systems and social infrastructure systems.<sup>1)</sup> However, recent advances in virtualization technology have created a desire to apply this technology to even mission-critical systems for more efficient use of computer resources. This desire has appeared alongside the trend in applying Cloud technology to corporate in-house systems to set up private Clouds.

If virtualization technology is used to integrate servers in a system constructed with several (10–100) typical industry standard servers, the failure of one physical server can have a far-ranging effect. Therefore, high-reliability high-availability servers are needed to support the virtualization platform.

### **2.1 Objectives**

In developing the PRIMEQUEST 1000 series, Fujitsu set out to satisfy performance, functional, and availability requirements for the use of this series not only as a large-scale, mission-critical system as before but also as a virtualization platform and private-Cloud platform. At the same time, to produce a high-quality, highly competitive product, Fujitsu set up a system for managing the entire business from product planning to development and sales strategies from the perspective of manufacturing and sales as a unified group and proceeded with development while reflecting market needs and trends on the basis of opinions voiced by sales staff and systems engineers (SEs.)

One result of this process was the requirement to use the Intel Xeon processor rather than the Itanium processor currently used in the existing series. This is because the Xeon excels not only in performance but also in the extensive lineup of products that it supports.

These products include, of course, operating system (OS) and virtualization software, but also middleware products like databases, packaged software for business use, and operation and management software.

The development department studied the Xeon's expected functions, while it was still under development, from various viewpoints in consultation with Intel. The end result was the PRIMEQUEST 1000 series using the Intel Xeon processor 7500 series that, in addition to excelling in performance and virtualization, features reliability and availability at least as good as the existing PRIMEQUEST series by combining new functions with existing PRIMEQUEST functions.

### **2.2 Features**

The PRIMEQUEST 1000 series provides the following features in addition to those it inherits and enhances from the existing series.

- 1) Scalability up to 8 central processing units (CPUs) (64 cores) resulting in twice the performance and at least 3 times cost performance of the existing series
- 2) Extended memory and input/output (I/O) to meet the needs of a virtualization platform and private-Cloud platform
- 3) High reliability through QuickPath Interconnect (QPI) technology of the Xeon processor 7500 series and a system mirror function<sup>2)</sup> based on memory mirroring, and the incorporation of reliability, availability, serviceability (RAS) functions equivalent to those of the Itanium processor
- 4) Provision of a partitioning function like the existing series, flexible I/O function, reserve system board function, and OSIV/XSP operation mechanism
- 5) Redundancy and maintainability at the unit level (power supply, cooling fan, hard disk drive [HDD], etc.)
- 6) Reliability secured at the component level while reducing the number and type of

components and lowering costs

- 7) Significant increase in the number of supported OSs (including virtualization software) and amount of middleware, and packaged software
- 8) Long-term support for hardware (up to ten years)

In addition, since the provision of products that take into account the need to reduce environmental load is the responsibility of the ICT vendor, the PRIMEQUEST 1000 series is offered to customers as a “Fujitsu Super Green” product with special attention paid to the following green ICT functions.

- Significant energy savings, space savings, and weight reductions
- Visualization of consumed power, temperature, and air flow
- Reduced noise and higher recycling rate
- Compliance with environmental regulations of various countries besides Japan

### 3. Functions for high reliability and high availability

Mission-critical systems must provide high reliability and high availability twenty-four hours a day, seven days a week (24/7 basis). In this section, we outline the hardware of the PRIMEQUEST 1000 series and the functions for achieving high reliability and high availability.

#### 3.1 Partitioning function and flexible I/O function

Partitioning is one of the important functions inherited from the existing PRIMEQUEST series. It enables hardware resources within a cabinet to be divided into multiple systems that can run independently of each other. Since it divides resources on a hardware basis, partitioning excels in fault isolation between systems. It enables multiple systems differing in application, scale, and required reliability to be loaded with good efficiency while making maximum use of the high processing capacity of the PRIMEQUEST series.

The partitioning function of the PRIMEQUEST 1000 series enables multiple business applications to be implemented in the same cabinet and provides flexible system operation. It provides hardware-based protection between partitions so that a failure in any one system does not impact any other system. Moreover, it can be combined with the flexible I/O function of the PRIMEQUEST series to provide even greater system configuration flexibility.

The flexible I/O function enables the constituent elements of the PRIMEQUEST 1000 series to be combined as desired to configure partitions regardless of where they are mounted. As shown in **Figure 1**, the elements are system boards (SBs) equipped with CPUs and memory; I/O boards (IOBs) equipped with PCI Express (PCIe) slots and on-board I/O; Giga-LAN/SAS/PCI-box interface boards (GSPBs); SAS disk

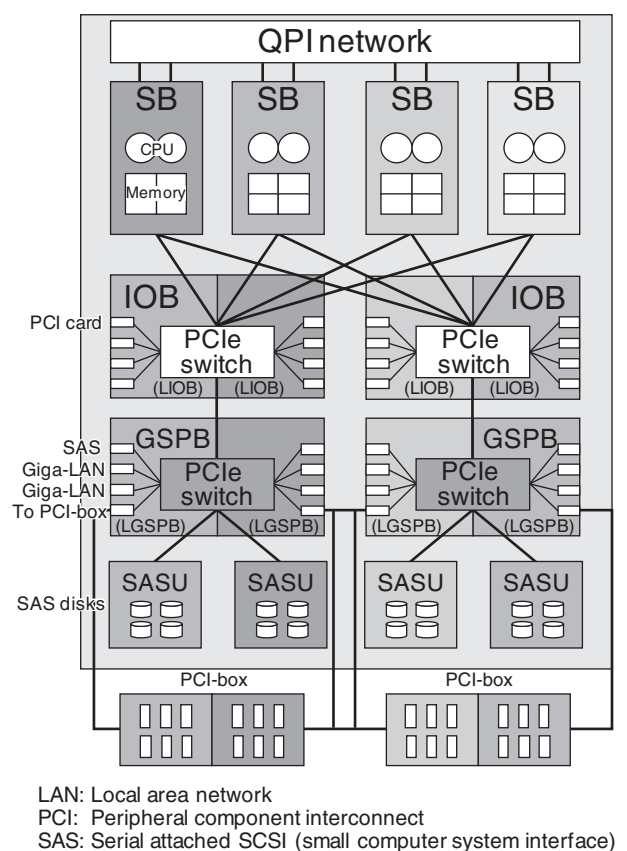


Figure 1  
Block diagram of PRIMEQUEST.

units (SASUs), and PCI-boxes. This function is achieved by using a switch to change how PCIe signals from SBs connect to I/O resources.

In terms of granularity, partitioning can be performed in units of SBs and, for I/O, in units of logical IOBs (LIOBs) and logical GSPBs (LGSPBs) (see **Figure 2**). An IOB and GSPB can be logically divided into two LIOBs and two LGSPBs, respectively. These elements can be combined to form up to four partitions in the cabinet. For example, all of the resources within a cabinet can be used to configure a large partition consisting of eight CPUs, all on-board I/O, and 40 PCIe slots, or the four SBs can be

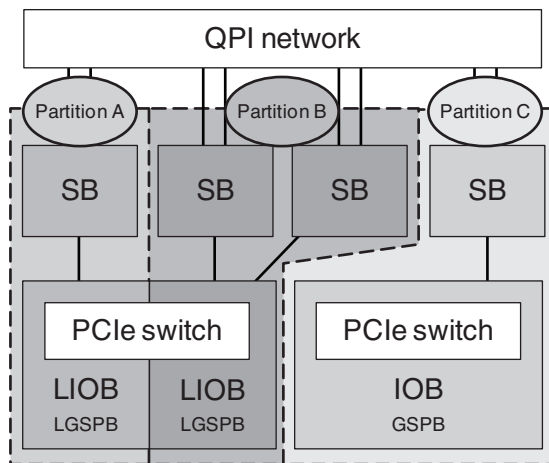


Figure 2  
Concept of partitioning.

used independently to create four partitions in which the amount of I/O resources used by each partition can be selected to suit business needs. Of course, two or three partitions may also be configured, in which case the number of SBs and I/O units in each partition can be set in terms of the granularity described above (Figure 2).

### 3.2 Rich I/O

Placing importance on I/O extendibility, PRIMEQUEST enables up to 40 PCIe slots to be used by connecting to PCI-box equipment. This extensive I/O capability can be used to configure multi-paths and raise system reliability. All 40 PCIe slots support the PCI Hot Plug specification and accommodate the cassette format. If a PCI card should fail, operations can continue while it is being replaced.

### 3.3 Reserve SB function

The reserve SB function assumes that a reserve SB is installed in the cabinet beforehand. In the event of an SB failure during operations, the failed SB is disconnected autonomously and the reserve SB reboots and incorporates itself in a short time, which enables the failed SB to be replaced (**Figure 3**). As a result of this function, the occurrence of a hardware failure on an SB does not mean that SB resources within a

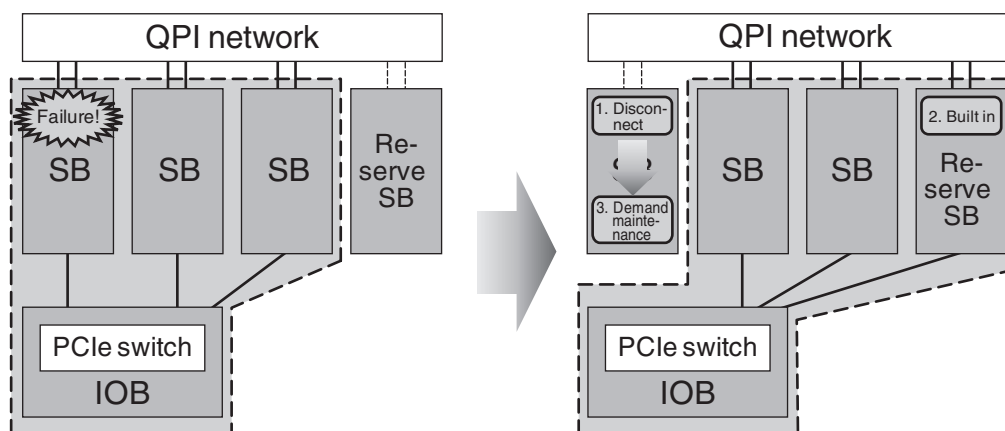


Figure 3  
Operation of reserve SB.

partition need to be reduced, but rather that the partition can be quickly restored in its entirety.

If SB switching by the reserve SB function should occur, the PCI bus tree structure as seen from the system must be maintained. Thus, for the case in which a partition is configured with more than one SB, one chooses a configuration in which I/O connections are made only by PCIe from a representative SB so that no change in the bus tree occurs whenever any one of the SBs switches.

The reserve SB function also enables an SB operating in another partition in the cabinet to be specified as a reserve SB. For example, if both a partition for a running system and a partition for a test system exist in the same cabinet, an SB in the test system can be specified as a reserve SB for the running system. If a failure should then occur in an SB in the running system, a shutdown command will be automatically issued to the test-system partition, and once shutdown is completed, the SB specified as a reserve SB will be incorporated into the running system and rebooted automatically. This capability raises system availability.

### 3.4 Hardware redundancy

The PRIMEQUEST series enables almost all components such as the power supply, cooling fan, and server management board (MMB) to be configured redundantly. Making hardware redundant in this way means that operations can continue even if some components fail by using the remaining components. There are five main redundancy measures in the hardware configuration of the PRIMEQUEST 1000 series.

#### 1) Memory mirroring

Memory mirroring provides duplicate units of memory and saves the same data to both units. If an uncorrectable error should occur in one of the memory units, data in the other memory unit can be used; this ensures continuous and stable system operations. Such memory mirroring achieves complete duplication of data

and consequently higher reliability than error correction mechanisms such as double device data correction.

#### 2) Power supply units

When receiving power from a single AC system, the power supply can be made redundant through the use of  $n + 1$  power supply units (PSUs). If one PSU should fail, business operations can still continue and the failed PSU can be replaced while the system is still running. When power is received from two AC systems, as many PSUs as are needed per system can be installed and the power supply can be made redundant by  $n + n$  units. In this way, the power supply from the AC power-supply source can be duplicated.

#### 3) Cooling fans

The cooling system is generally made redundant through the use of  $n + 1$  fans so that if one fan fails, cooling is maintained by the remaining fans. In other words, business operations can continue even if a cooling fan fails and that fan can be replaced while the system is still running.

#### 4) Hardware RAID

The PRIMEQUEST 1000 series supports RAID 0, 1, 5, 6, 10, and 1E (RAID: redundant array of independent disks).

#### 5) MMB duplication

The PRIMEQUEST series is equipped with an MMB for managing all equipment. This MMB consists of dedicated a CPU and memory and runs by firmware independent of partitions. Network ports and firmware code are duplicated within the MMB, and the MMB itself is duplicated. In the event of an MMB hardware failure, operation will automatically continue using the standby MMB. More details are given in section 5 Server Management Functions.

### 3.5 Degeneration functions

#### 1) QPI resend, degeneration, and detouring

The new Intel Xeon processor uses the QPI high-speed system bus for communications



between CPUs or between a CPU and a chipset. This protocol uses the cyclic redundancy check (CRC) method to check transmitted data and performs resends if errors occur. In the event of a permanent bus failure, the system degenerates to 1/2 or 1/4 of the bus width to avoid the fault location. Resending or bus-width degeneration can be performed while operations continue. If a specific QPI bus becomes unusable because of failure, the protocol establishes a detour route around that bus and automatically performs a reboot. If a detour cannot be established because of configuration conditions, the protocol performs an SB-based degeneration followed by a reboot.

## 2) PCIe retry and degeneration

Communications between PCIe, SB, IOB, and individual I/O cards are checked by CRC. They are retried if an error occurs. Since a bus is configured in units of “lanes,” the occurrence of a permanent bus error will result in degeneration of the bus width by up to 1/2, 1/4, or 1/8 toward essential Lane 0 to avoid the fault location. Retry or bus-width degeneration can be performed while operations continue. If a specific bus becomes unusable because of failure, devices under that bus are disconnected and a reboot is automatically performed. If a PCIe switch or I/O device should fail, the same technique can be applied to disconnect components affected by that failure and perform a reboot.

## 3.6 Hot maintenance

If a fault should occur in one of the partitions operating in the same cabinet necessitating the replacement of a component, that component can be replaced without impacting the other operating partitions as long as the unit in question is not shared. The PRIMEQUEST series is based on a cable-less design concept that enables each unit to be replaced easily without the need to remove or pull out drawers which contain equipment and internal cables. The server midplane that accepts each unit and connects inter-unit signals is located in the

center of the equipment. If it should fail for some reason, its replacement requires extensive work while the equipment is completely shut down. However, the PRIMEQUEST midplane does not contain any electronic components and consists of only connectors and wiring functions on printed circuit boards, so the failure probability is close to zero.

## 3.7 Fault analysis and logging

The PRIMEQUEST series runs three main types of firmware: basic input/output system (BIOS), baseboard management controller (BMC) firmware, and MMB firmware. If a hardware anomaly occurs, information collected by each type of firmware is consolidated at the MMB to comprehensively determine a suspected fault location. In addition, event information associated with operational and environmental changes and error information as described above are recorded in the MMB system-event log and used in the analysis process whenever an anomaly occurs.

## 3.8 Intel Xeon processor

In addition to the QPI fault-recovery functions described above, the new Intel Xeon processor features significantly enhanced RAS functions as in Machine Check Architecture (MCA) Recovery. This processor has reliability on par with that of the Itanium processor used in the PRIMEQUEST 500A series and earlier models. During the development of the Itanium processor, Fujitsu worked with Intel to enhance RAS functions and perform evaluation tests. The results of that joint effort continue to be used today. Fujitsu plans to continue working with Intel toward even higher levels of reliability in the next-generation processor.

The functions and system configuration described above provide the PRIMEQUEST 1000 series with high reliability and high availability as a mission-critical line of servers.

## 4. Functions for green ICT

The PRIMEQUEST 1000 series has significantly better green ICT functions than the existing series as a product that gives due consideration to the global environment. The main improvements are summarized below.

- 1) Maximum CPU power consumption reduced to 1040 W (500 A series: 3328 W)
- 2) First use of an 80Plus-Gold-certified power supply unit using high-switching-frequency and high-efficiency diodes in servers of this

class. Use of high-conversion-efficiency DC-DC converters (**Figure 4**).

- 3) Improved cooling efficiency within the cabinet and reduced load on cooling fans achieved by using a CPU heat sink with a heat pipe, a duct structure that directly guides intake air to high-temperature components, and a shutter mechanism that prevents the disruption of cooling air flow during hot maintenance (**Figure 5**).
- 4) Optimized cooling by dynamic control of

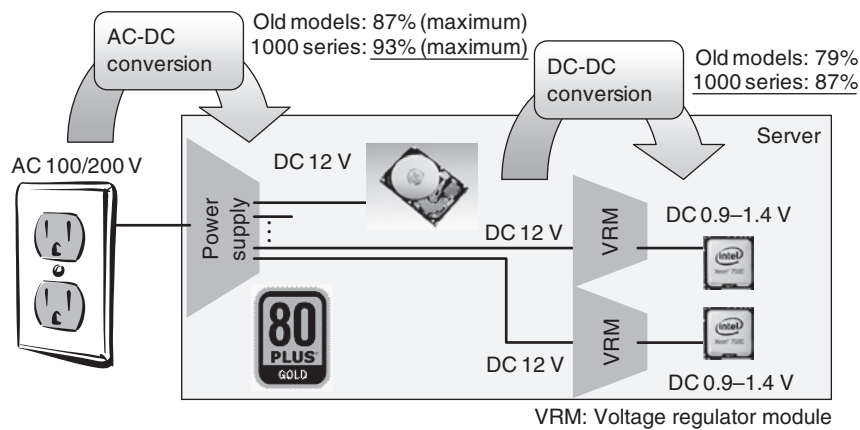


Figure 4  
Efficiency of voltage conversion.

Duct structure concentrates cooling air on high-heat-generating components

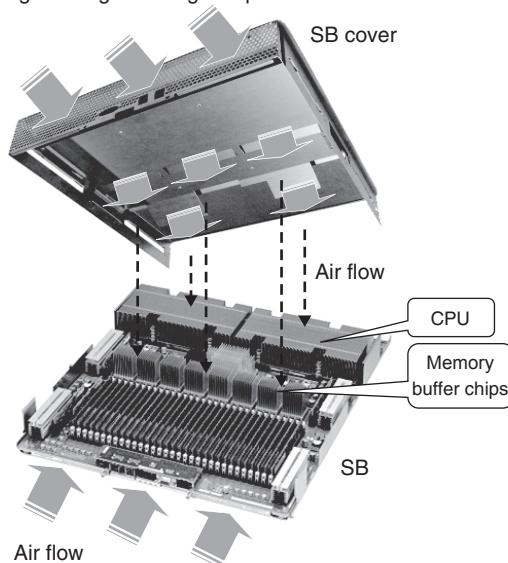
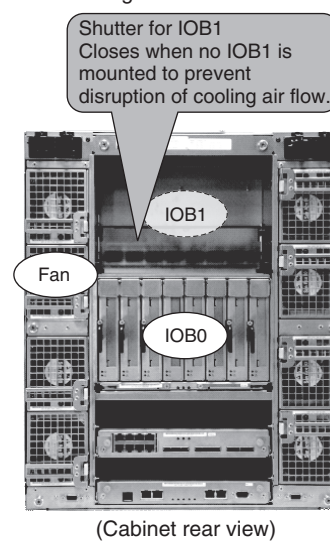


Figure 5  
Efficiency of cooling.

Structure preventing inflow of cooling air



fan speed in 1% steps, reduced fan power consumption, and reduced fan noise during operation.

- 5) Reduced space (height reduced by about 1/3 and volume by about 1/7) and reduced weight (by about 1/5) owing to high CPU-core density, energy-saving components, and high power-supply efficiency. All models are rack-mounted, which enables peripheral equipment to be accommodated in the same rack whereas in existing top-level models it must be accommodated in a separate rack from the main unit.
- 6) Visualization not only for cabinet power consumption and temperature but also for air flow to optimize facilities.
- 7) Reduced noise level (66 dB → 59 dB).
- 8) Improved recyclability rate (99%).
- 9) Compliance with the EU's Restriction of Hazardous Substances (RoHS) directive and other environmental laws and regulations around the world.
- 10) Use of electronic manuals on DVDs (digital versatile discs) to save resources.
- 11) Remote operation by a video-redirection function to cut down on energy used by the movement of system administrators.

These improvements have achieved an annual reduction in CO<sub>2</sub> emissions per server, which is equivalent to about 2000 cedar trees.

The PRIMEQUEST 1000 series delivers even more energy-saving effects by integrating servers and optimizing server usage efficiency through the ServerView Resource Coordinator VE (Fujitsu's middleware).

## 5. Server management functions

Server management functions are important elements in achieving stable operation in a server system. The PRIMEQUEST 1000 series has an MMB inherited from existing models to manage the entire cabinet. Moreover, for each partition, the PRIMEQUEST 1000 series uses the ServerView Suite (SVS) server management

software (common to Fujitsu IA servers including the PRIMERGY series) and PRIMEQUEST Server Agent (PSA) dedicated agent software to provide functions for meeting mission-critical requirements such as monitoring for anomalies in PCI cards and performing high-speed cluster switching. The server management function configuration is shown in **Figure 6**.

### 5.1 Overview of server management functions

#### 5.1.1. MMB

The MMB manages partitions and monitors hardware throughout the PRIMEQUEST cabinet. This hardware includes SB, IOB, SASU/SASRU (SAS array disk unit), GSPB, power supply, fans, and PCI-box.

The MMB connects to each hardware component via an internal local area network (LAN) and to each sensor via internal buses to continuously monitor and manage the entire system. As an option, the MMB can be given a redundant configuration. In this case, a failure of the MMB itself or in the associated network results in automatic switching to the backup MMB to allow system monitoring to continue. When it detects an anomaly, the MMB can report it in various ways (report to REMCS center, snmp trap, and E-mail message [REMCS: remote customer support system, snmp: simple network management protocol]).

The MMB also provides functions for linking with the BMC and performing remote partitioning operations (video redirection, remote storage, and text-console redirection). The video-redirection function enables video graphics array (VGA) output to be displayed via the Web on the user's terminal.

The input from a mouse or keyboard connected to the user terminal is connected to the partition side via the LAN. When the text-console redirection function is launched, the serial output from the target partition is displayed via the Web on the user terminal. The remote-



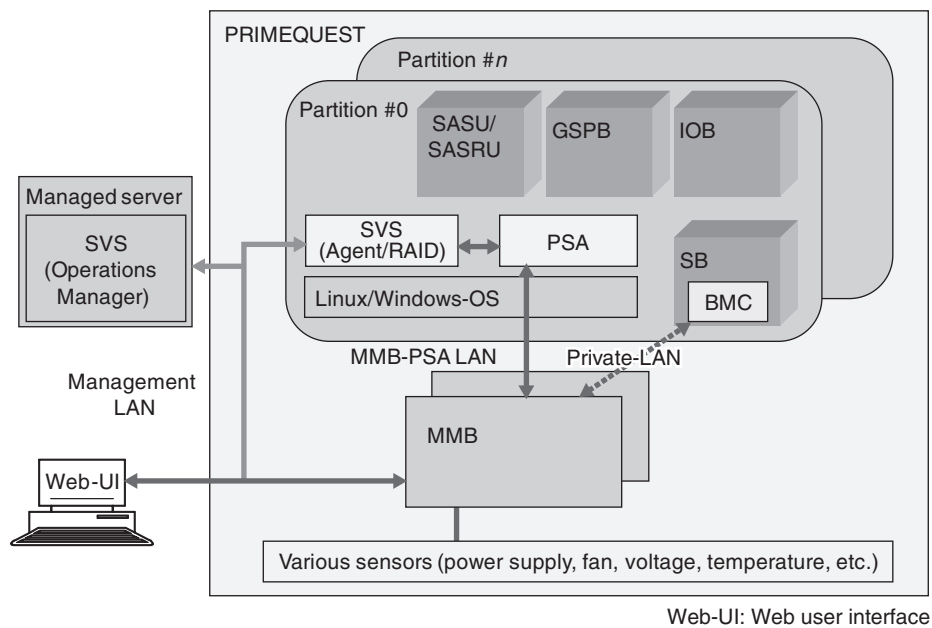


Figure 6  
Configuration of PRIMEQUEST server management.

storage function can be used in conjunction with the video-redirection function. A CD/DVD drive (CD: compact disc), ISO image (archive file or disc image, ISO: International Organization for Standardization), floppy disk, or USB (universal serial bus) device on the user terminal can be shared with the partition as a storage device. In the case of an ISO image file, the terminal-side disk drive is seen as an emulated drive from the partition side.

### 5.1.2 SVS

SVS is a group of integrated system management tools that are also used in the PRIMERGY series. These tools enable simple and accurate deployment, 24/7 stable operation, and straightforward system management in which the process from system deployment to operation is defined in terms of categories, namely, deployment, operation, diagnosis & analysis, and integration.

The ServerView Installation Manager supports initial deployment and reconstruction

work. It achieves stable operations by simplifying deployment work and supporting the installation of recommended drivers and security patches.

The ServerView Operations Manager supports operations monitoring and diagnosis, analysis, and integration. A ServerView Agent installed in each partition manages the configuration of CPUs, memories, and HDDs that can be detected in each partition and monitors system resources (such as CPU utilization rate). It also determines performance conditions and sends out notifications in the event of system anomalies. The ServerView Operations Manager installed in managed servers or managed partitions on a PRIMEQUEST server enables integrated management via the graphical user interface of each partition or other servers including those of the PRIMERGY series. It can also notify the system manager of events received from partitions in various ways (E-mail, snmp traps, alarm lights, pop-up menus, etc.)

The ServerView RAID Manager supports array-controller monitoring. It is installed in

each partition to monitor the configuration of the array controller, the state of HDDs, and battery conditions and sends out notifications when anomalies are detected. It also provides disk array configuration and battery calibration functions.

### 5.1.3 PSA

The PSA installed in each partition is a PRIMEQUEST exclusive monitoring agent. It interfaces closely with the MMB through the internal LAN, and as software running on Linux and Windows, it monitors the configuration and state of PCI cards and HDDs. The PSA also checks for anomalies in multipath drivers and if it detects an anomaly, it notifies the MMB, which can then take a variety of report actions. In addition, the PSA can issue instant notifications in conjunction with the MMB in response to OS state changes to the PRIMECLUSTER clustering software used in mission-critical systems. The PSA provides a function for forcibly terminating an OS and can perform immediate and accurate switching.

### 5.1.4 SVS/PSA linking

The PSA can detect anomalies from notifications sent out by the SVS and can report them to the REMCS center via the MMB.

## 5.2 Future activities

At present, SVS and PSA coexist as monitoring software installed in the same partition, but we intend to integrate PSA functions into SVS to unify these management functions on IA servers. Specifically, these PSA functions will be the monitoring for PCI-card and HDD anomalies detected by OS drivers, the monitoring for anomalies in multipath drivers, and the interfacing with PRIMECLUSTER. This integration will help to fortify SVS functions, unify the IA-server management platform, and fortify server management functions in support of mission-critical systems. All in all, it will improve management functions for all IA servers including those of the PRIMERGY series.

## 6. Long-term support activities

There has always been a demand for 24/7 availability in social infrastructure systems and finance-related systems, but now there is also a strong demand for stable operation over the long term. The desire for long-term use is also appearing for mission-critical systems in other fields. The results of a survey on customer wishes toward vendors (from the Japan Users Association of Information Systems [JUAS] FY2007 Hardware/Software Lifetime Research Project Report) revealed the desire for extended maintenance support periods for both hardware and software (**Figure 7**)<sup>3)</sup> and the desire for a

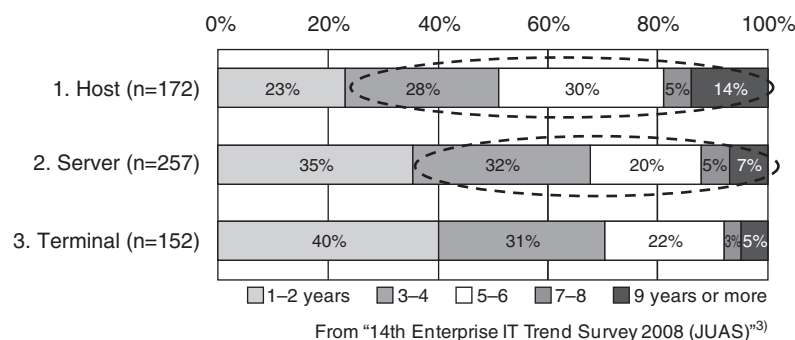


Figure 7  
User demand for extended period of hardware maintenance and guarantee.

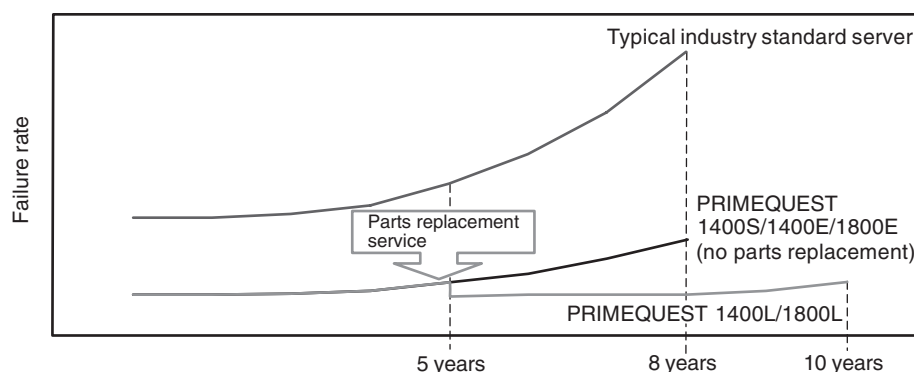


Figure 8  
Running time and failure rate.

study of countermeasures to the depletion of hardware parts and components. In response to these wishes, more products and services that can be used over the long term are now being provided for OSs, middleware, and packaged software.

Fujitsu sets rigorous quality and screening standards from the component level for the PRIMEQUEST series of servers. At the same time, cooling fans, HDDs, batteries, and capacitors all have a finite lifetime, and as such, it is difficult to maintain a failure rate within a fixed standard over the long term (**Figure 8**).

In response to the above desire for long-term use, Fujitsu has decided to provide a long-life model for the PRIMEQUEST 1000 series that guarantees a supply of maintenance parts for a 10-year period. This is a first for servers of this class. In addition to a guaranteed supply of maintenance parts, the hardware maintenance support service of this long-life model includes a replacement service at the 5-year mark for components like fans and HDDs to avoid wear that occurs over time. This model enables stable hardware use for a 10-year period.

## 7. Virtualization support policy

This section describes Fujitsu's virtualization support policy for the PRIMEQUEST 1000 series. Virtualization technology is becoming

increasingly important for making effective use of ICT resources and running them efficiently. In particular, virtualization that includes servers, networks, and storage is practically becoming a precondition for Cloud computing that needs to pool ICT resources and apportion them in a flexible manner. The PRIMEQUEST 1000 series combines hardware partitioning having high fault isolation properties with a software-based virtual machine (VM) function to provide a more efficient server virtualization environment. It also provides the following features as technologies supporting a virtualization environment.

### 7.1 Large-capacity memory

Running multiple VMs on a server can result in bottlenecks in memory and I/O more so than in the CPU. The amount of memory that can be mounted is therefore a key consideration when considering high-density virtual consolidation. The PRIMEQUEST 1000 series has 16 memory slots per CPU, which means that up to 128 slots are available and 1 TB of memory can be installed. If we assume an average memory capacity of 2–4 GB per VM, then about 256 servers can be consolidated.

### 7.2 Abundant I/O slots

The next source of bottlenecks after memory is I/O throughput. In addition to 16 on-board

ports for the 1-Gb/s LAN, the PRIMEQUEST 1000 series can have up to 40 PCIe extension slots (16 standard and 24 extended). Sufficient LAN performance per VM can be easily secured even for high-density virtual consolidation. Simply passing on the server's LAN configuration to VMs eliminates the need for complicated virtual LAN (VLAN) settings.

### 7.3 High reliability and high availability

Next, the extent to which a failure in a server can affect other elements in the server is an issue that must be addressed in high-density virtual consolidation. In the worst case, a server failure affects all VMs. To address this problem, the PRIMEQUEST 1000 series incorporates three measures: it ensures reliability at the component level, limits a failure's range of impact through partitioning functions, and achieves rapid recoveries (within about 1 hour) through a reserve SB and flexible I/O.

### 7.4 Software-license savings

For typical virtualization software like VMware, software license fees are often set according to the number of CPU cores. Savings can therefore be achieved by consolidating many VMs in the CPU cores.

### 7.5 Software selectivity

The PRIMEQUEST 1000 series supports the following software for achieving virtualization operations.

- 1) Virtualization software
  - VMware, Hyper-V, and Xen (Red Hat Enterprise Linux 5)
- 2) Cloud management software
  - ServerView Resource Orchestrator
  - Systemwalker Service Catalog Manager
  - Systemwalker Software Configuration Manager
  - Systemwalker Runbook Automation

### 7.6 Simplified server-resource upgrading

In contrast to 2-CPU scale-out servers, the PRIMEQUEST 1000 series can scale up to 8 CPUs. A system can be extended by simply adding more CPUs or memories without design changes having to be made. There is also no need to perform a system reconfiguration or operation test when a scale-out server is upgraded. In short, server resource addition is easy and upgrade costs are relatively low.

### 7.7 Application example

The virtual consolidation of the existing mainframe-based mission-critical system at Tokyo's Toshima City Office through the deployment of PRIMEQUEST 1000 series servers and VMware provides a good case study of PRIMEQUEST 1000 series application.<sup>4)</sup> In addition, the Fujitsu, Microsoft, SAP Japan Three-company Joint Verification Project PRIME3<sup>5)</sup> has demonstrated stable operation in a next-generation SAP system through virtual consolidation using a PRIMEQUEST 1000 series server, ETERNUS disk storage systems, and Hyper-V software (**Figure 9**).

### 7.8 Future activities

Fujitsu intends to intensify its virtualization support through the provision of tested models using the PRIMEQUEST 1000 series as a server product group supporting private-Cloud systems in which Cloud technology is applied to customers' in-house systems.

## 8. OSIV/XSP operation mechanism and deployment benefits

The OSIV/XSP OS used in Fujitsu Global Server (GS)<sup>6)</sup> and PRIMEFORCE mainframes and middleware such as the Advanced Information Manager (AIM) and business applications running on that OS can run as such on the PRIMEQUEST series of servers. The hardware option needed for this purpose

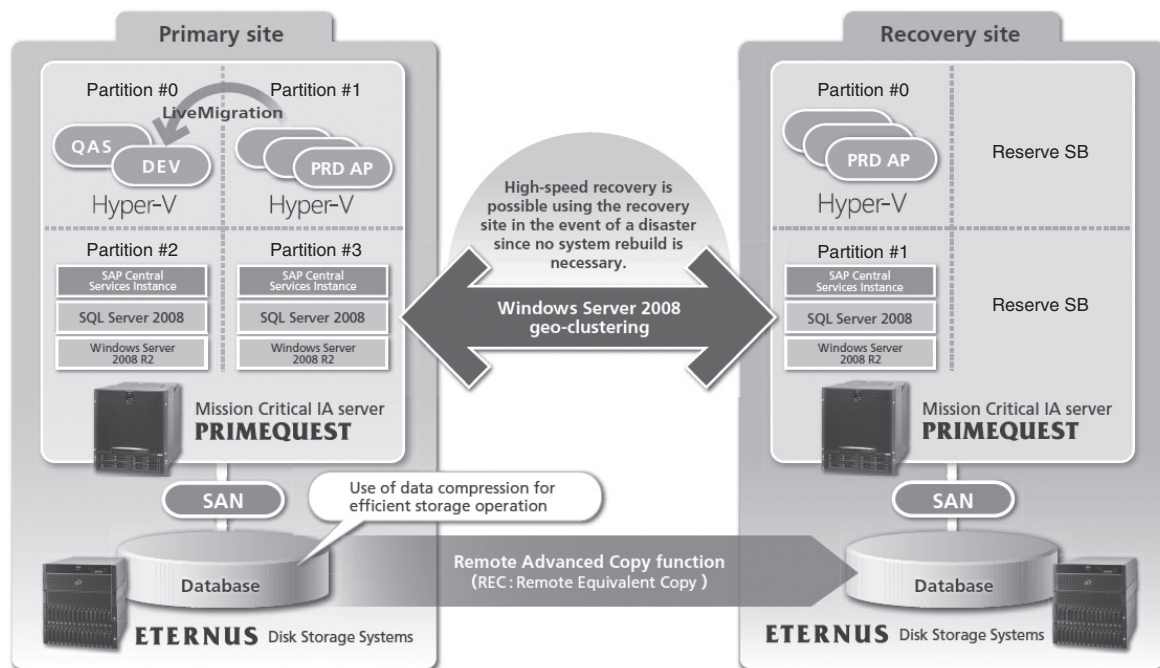


Figure 9  
Verification by PRIME3 project.<sup>5)</sup>

is the OSIV/XSP operation mechanism.<sup>7)</sup> This mechanism is placed in the firmware layer between the CPU and OS. It virtualizes the Intel Xeon processor as a GS CPU handled by OSIV/XSP and converts OSIV/XSP instruction code into Intel-Xeon-processor instruction code for execution (**Figure 10**). The OSIV/XSP operation mechanism is provided as an OSIV/XSP operation mechanism kit combined with hardware including an SB, CPU, memory, and HDD. Furthermore, to simplify migration from GS, a PRIMEQUEST server that includes the OSIV/XSP operation mechanism maintains high compatibility with GS. The main compatible features are introduced below.

- 1) Runs AVM/EXS virtualization middleware for GS
- 2) Incorporates block multiplexer channels and optical channel link (OCLINK) channels and enables GS I/O equipment to be connected directly to PRIMEQUEST
- 3) Enables power-supply linking, anomaly alarm generation, etc. in relation

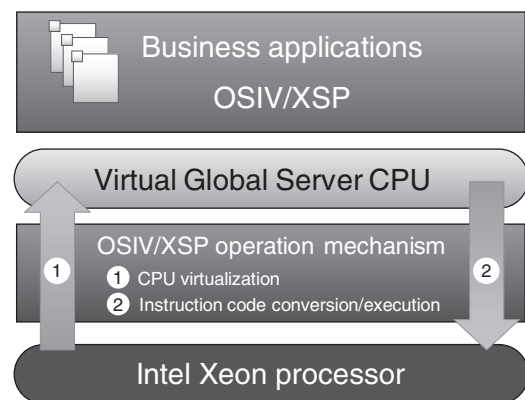


Figure 10  
Translation facility of OSIV/XSP operation mechanism.

to peripheral I/O equipment and air conditioning facilities to be operated the same as in GS

These compatibility-maintaining features enable the same operation as in GS to be achieved without the need to modify existing user application assets or change system operation.



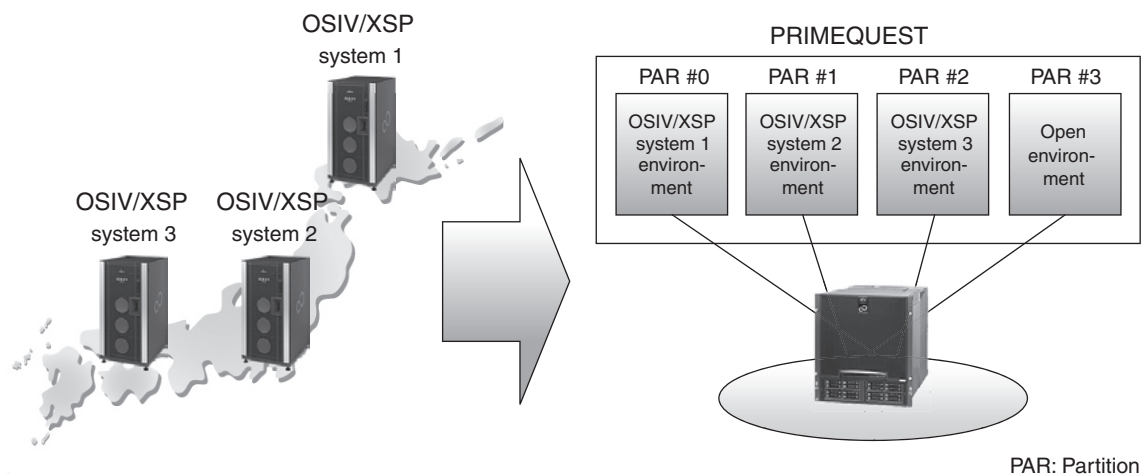


Figure 11  
Integration of OSIV/XSP and open systems.

## 8.1 Deployment benefits

Merging with PRIMEQUEST in this way provides benefits that could not be enjoyed in the past. The construction of multiple independent systems using the PRIMEQUEST partitioning function enables separately deployed OSIV/XSP systems to be integrated on one PRIMEQUEST server. An open system can also be integrated on the same PRIMEQUEST server (**Figure 11**). Besides mainframe OSs, we are also planning to support OSs for the PRIMERGY 6000 series of office computers.

## 9. Conclusion

This paper introduced the objectives, features, and associated technologies of Fujitsu's PRIMEQUEST 1000 series of servers, the latest series in the company's Mission-Critical IA Server PRIMEQUEST line of servers. Looking forward, Fujitsu will continue to develop mission-critical IA servers and drive the expansion of the open mission-critical market by utilizing leading-edge technologies and establishing strong and effective partnerships with leading companies.

## References

- 1) A. Kurokawa et al.: Value Proposition of Newly Developed Mission-Critical IA Server:

- PRIMEQUEST. *Fujitsu Sci. Tech. J.*, Vol. 41, No. 3, pp. 277–283 (2005).
- 2) O. Hamada: Highly Reliable System Mirror Function of Mission-Critical IA Server: PRIMEQUEST. *Fujitsu Sci. Tech. J.*, Vol. 41, No. 3, pp. 298–301 (2005).
- 3) Japan Users Association of Information Systems: 14th Enterprise IT Trend Survey 2008. (in Japanese).  
<http://www.juas.or.jp/servey/it08/index.html>
- 4) Fujitsu: Deployment Case Study, Toshima Ward City Office. (in Japanese).  
<http://jp.fujitsu.com/solutions/theme/infrastructure/virtual/casestudies/toshimaku/>
- 5) Fujitsu, Microsoft, SAP Japan Three-company Joint Verification Project PRIME3.  
[http://jp.fujitsu.com/solutions/sap/brochures/pdf/prime3project3\\_english.pdf](http://jp.fujitsu.com/solutions/sap/brochures/pdf/prime3project3_english.pdf)
- 6) H. Egawa et al.: Global Server GS21. (in Japanese), *FUJITSU*, Vol. 53, No. 6, pp. 471–476 (2002).
- 7) Y. Saito et al.: Case Examples of Migrating to Open Systems Using Mainframe Assets. (in Japanese), *FUJITSU*, Vol. 61, No. 5, pp. 451–457 (2010).



**Motoyoshi Hirose**

*Fujitsu Ltd.*

Mr. Hirose is engaged in technology support for mission-critical IA servers.



**Masahiro Kurita**

*Fujitsu Ltd.*

Mr. Kurita is engaged in the development of server management software.



**Keiji Sato**

*Fujitsu Ltd.*

Mr. Sato is engaged in the development of mission-critical IA servers.



**Yasufumi Honda**

*Fujitsu Ltd.*

Mr. Honda is engaged in the development of Cloud infrastructure models.