Fujitsu's PRIMERGY BX620 S4 Blade Server for Solving Server Consolidation Problems

● Hideaki Fujimaki ● Masakazu Matsubara ● Masatoshi Shinoda

(Manuscript received May 31, 2007)

Based on a technical analysis of about 9000 business talks given by Fujitsu for customers, we found that amid the ongoing IT system optimization at enterprises, blade servers had become an increasingly important IT platform for optimizing infrastructure. We also discovered that about 80% of our customers had introduced blade servers in order to consolidate their own servers. The customers expect this consolidation of servers to provide management with a global understanding of their IT systems, reduce the heavy workload that server operation imposes on user departments, and curtail IT investment costs in information system departments. However, server consolidation also entails such problems as a more complex network design and increased workload on an information system department. Fujitsu fully understood these server consolidation problems based on extensive consultations on optimizing infrastructure conducted in the past. To solve these problems, Fujitsu developed and marketed the PRIMERGY BX620 S4 blade server. The PRIMERGY BX620 S4 offers superior IT system flexibility by expanding network functions, simplifies operation and management by utilizing virtualization technology, and also reduces power consumption. This paper describes the technologies that provide these advantages.

1. Introduction

Amid the ongoing IT system optimization at enterprises, the market for blade servers has significantly expanded. Given the need to consolidate servers in various enterprise departments, coupled with advances made in server virtualization technology, Fujitsu systematized TRIOLE BladeServer[1] technology to provide the most advanced IT systems for customers, and has continued making combined hardware/software technical innovations, mainly based on technological advances in automatization, visualization, and server virtualization. These technologies are also expected to reduce IT maintenance costs in information system departments, enhance security among the various departments, and concentrate IT visualization and resources in original services (all of which are advantageous for management and service departments). Fujitsu's application results suggest that consolidating 50 existing servers into 20 blade servers that adopt server virtualization technology could reduce the total cost of ownership (TCO) to about 55%.

Such server consolidation, however, entails the following three new problems:

1) Complicated network design

In a conventional blade server system, each server actually supports four ports as Gigabit Ethernet ports. Conversely, the new system employs a redundant LAN configuration where one LAN is shared by multiple services and thus necessitates a very complicated design using a tag virtual LAN (Tag VLAN) in considering interference with network performance.

2) Increased workload on administrators of

information system departments

Changing the connection of blades to a Storage Area Network (SAN) accounts for about 50% of the total server consolidation work. Since changing the server settings entails changing the SAN-side settings as well, administrators must handle an increased workload.

3) Increased power consumption at datacenters

Conventional blade servers consume more power in proportion to performance. This increases the total power consumption at datacenters.

To address the problems above, Fujitsu attempted to enhance blade servers and server management software by developing and marketing the PRIMERGY BX620 S4 blade server.[2)] The PRIMERGY BX620 S4 offers improved operability and availability as expected in server consolidation, and reduces the customer's TCO, including management cost and such running costs as that for power consumption.

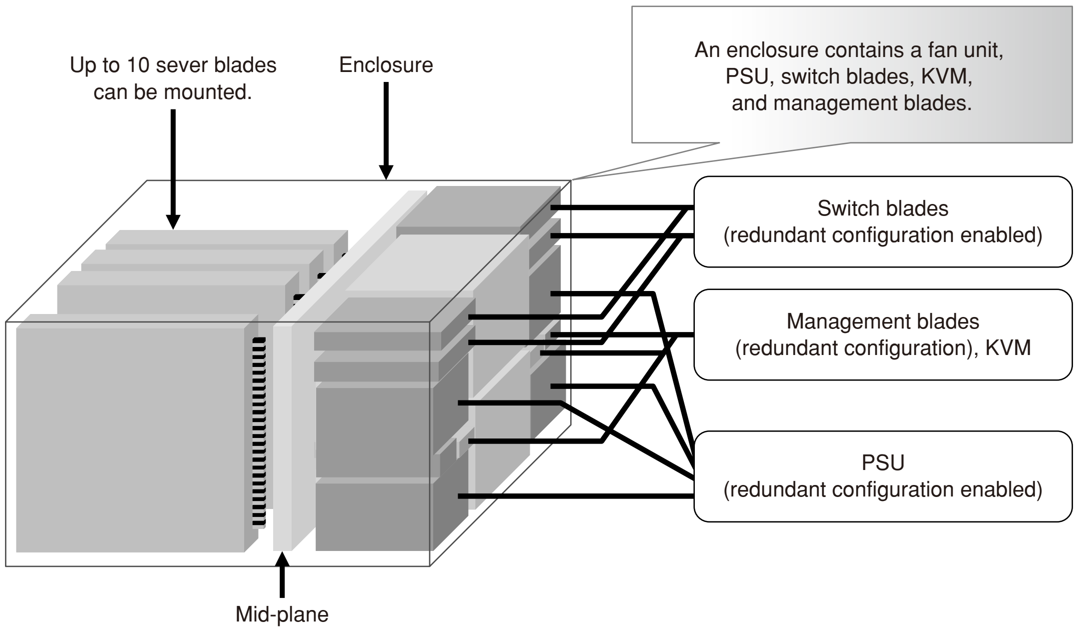
This paper introduces the following advantages of the PRIMERGY BX620 S4:

1) Improved flexibility by extending network functions.
2) Improved operation management and service consistency facilitated by virtualization technology.
3) Reduced power consumption.

2. Improved flexibility by extending network functions

Adding and changing services in a consolidated server system poses the problem of complicated network design. To address this problem, we enhanced the network functions for the new system. The following describes the technology adopted by the PRIMERGY BX620 S4 to enhance network functions.

In describing the structure of a blade server, it mainly consists of server blades, an enclosure, and switch blades (**Figure 1**). The server blade contains a CPU, memory, hard disk drive, and I/O



KVM: Keyboard/Video/Mouse switch
PSU: Power supply unit

Figure 1
Structure of blade server.

ports, and operates as a server. The enclosure is the chassis that houses the sever blades, switch blades, power supply unit, cooling fan unit, and management blades for system management. These components inside the enclosure are interconnected by a cable-less design, via a mid-plane positioned at the center of the enclosure. The I/O ports of the server blades are connected via the switch blade to a network outside the enclosure.

These blade servers offer levels of serviceability, power consumption, operability, installation, and ease of adding servers that are superior to those of rack-mounted servers.

The PRIMERGY BX620 S4 is equipped with more server-blade I/O ports, enclosure transmission lines, and switch-blade I/O ports, so that one server blade can have up to ten I/O ports.

2.1 Enhancement of server blade I/O ports

One server blade can have up to ten I/O ports. Therefore, even when adopting the LAN redundant configuration, physically independent ports can be assigned to up to three different purposes or services. For example, when a blade server is used as an application server in a Web 3-layer model, two ports for the Gigabit Ethernet interface can be assigned to a Web server, database server, and management server, respectively. This prevents the sharing of multiple purposes and services on one LAN, as well as interference with performance and complicated network settings (**Figure 2**).

Adding a Gigabit Ethernet, 4-Gb/s fibre channel (FC), or 10-Gigabit Ethernet LAN (to be supported in the future) enables the addition of two ports, depending on the expansion board to be mounted on the server blade. This facilitates network and storage connection by using the SAN boot function.

2.2 Enhancement of the mid-plane inside the enclosure

The mid-plane inside the enclosure was improved to enhance transmission lines that connect the server blades to switch blades. Specifically, the number of transmission lines was increased based on the ports added. As a result, the bandwidth between the server blades and switch blades was widened up to 520 Gb/s, or about 4.3 times wider than the conventional one. This enables a flexible network configuration

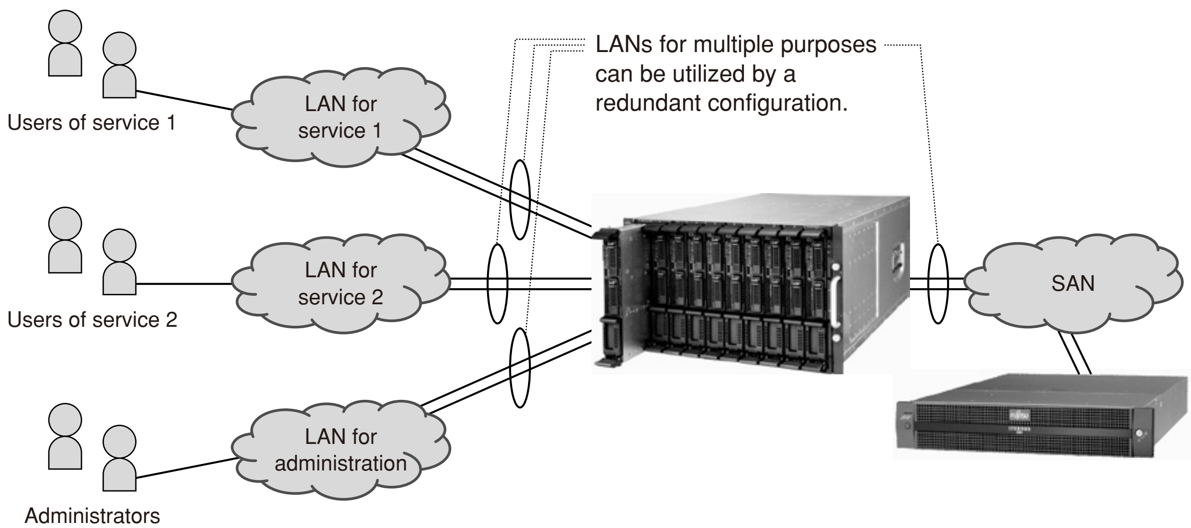


Figure 2
Example of enhanced network.

and assures a wider bandwidth for consolidating servers.

2.3 Enhancement of switch blades

More ports were added to the 1-Gb/s switch blade side. That is, the number of downlink ports was increased to 30, given the increased number of standard LAN ports for the server blade. The number of uplink ports was increased to 12 to assure the necessary bandwidth relative to downlink port enhancement. Moreover, the user interface of these switch blades and that of Fujitsu's LAN switch "SR-S Series"[3] (highly evaluated for its high reliability and security) were unified for easier operation.

3. Improved operation management and service consistency facilitated by virtualization technology

Connecting a blade server to SAN in a conventional configuration poses an operation problem, such as taking much time for the server administrator and SAN storage administrator to make arrangements at server maintenance or replacement. This problem was resolved by enhancing SAN connection virtualization technology that provides easier operation management, and improves service consistency. The following describes the conventional method and problems in interconnecting a blade server to SAN, and then describes the new technology — SAN connection virtualization technology.

3.1 Conventional method of interconnecting blade servers to SAN

A blade server is interconnected to the SAN environment via an FC expansion board and an FC switch blade. The FC expansion board can be mounted in the server blade; the FC switch blade can be mounted in the enclosure. The FC expansion board that can be mounted in each server blade has two FC interface ports. Up to two FC switch blades can be mounted per enclosure. The FC expansion board and FC switch blades configure a multi-path SAN connection environment. The blade server can be connected to a disk array unit by either of two methods: 1) directly connecting a disk array unit supporting SAN connection to the FC switch blade in the blade server, or 2) connecting the disk array unit via an external FC switch cascade-connected to the blade server.

Manually performing server maintenance or replacement in such a SAN connection environment entails many complicated operations, and minute mistakes could cause serious problems. To address such problems, we developed automatic operation management software known as "Systemwalker Resource Coordinator".[4] This product supports the SAN boot function and N:1 automatic recovery function. The SAN boot function executes system bootstrap operation from the SAN disk. The N:1 automatic recovery function recovers the system in case of server failure by automatically switching the faulty server blade to a standby server blade. These functions reduce administrator workload by shortening the system recovery time, and automating server maintenance and replacement.

3.2 Problem with the conventional interconnection method

The conventional interconnection method described above poses the problem of having to change the SAN-side settings at automatic recovery.

When SAN areas are assigned at blade server installation, the World Wide Name (WWN) — a fixed address set in the FC expansion board of the server blade — must also be set on the SAN side to establish a one-to-one relation between each server blade and each SAN-side volume. Therefore, when server conditions must be changed, the SAN-side settings must also be changed. Should the server for service A fail in the server N:1 automatic recovery configuration, for example, service A must be processed by the standby server. For this purpose, the address of the standby server's FC expansion board must be recognized as the address of the server for service A. Therefore, the address of the volume corresponding to service A on the SAN side was switched to the WWN of the standby server (**Figure 3**).

In the example given of a user system, the server administrator is different from the SAN storage administrator. Therefore, when the server administrator changes the server conditions, the SAN storage administrator must also change the SAN-storage setting conditions. This increases workload on the server administrator because it takes a considerable amount of time to make arrangements with the SAN storage administrator.

3.3 Method and advantages of SAN connection virtualization

The PRIMERGY BX620 S4 can switch server conditions without having to change SAN-side settings because the interconnection between blade servers and the SAN environment is virtualized. Specifically, the uniqueness of each port of an FC expansion board mounted in a server blade is assured, and virtual WWNs that can be freely assigned to devices are written to volatile RAM of the FC expansion board in each server blade to enable recognition of those virtual WWNs from the SAN side. We also developed operation management software known as "Systemwalker Resource Coordinator Virtual server Edition"[5] to manage, automatically set, and control the virtual WWN to be set in each server blade.

The following describes the concrete operation of server switching by using the virtual SAN connection. **Figure 4** shows an example of automatic server switching at error occurrence in the current server used for service A. The current server is automatically switched to the standby server as follows: First, the current server is stopped, then the virtual WWN assigned to that server is set to the standby server, followed by the standby server being restarted to complete server switching.

Since the settings on the SAN side need not be changed in this new system, no SAN-side adjustments are required when switching servers, even when the server administrator is different from the SAN storage administrator. Changing the server configuration in a conventional system takes several hours to several days because arrangements must be made between the server administrator and SAN storage administrator. In the new system, however, the time needed to make a configuration change was reduced to mere minutes.

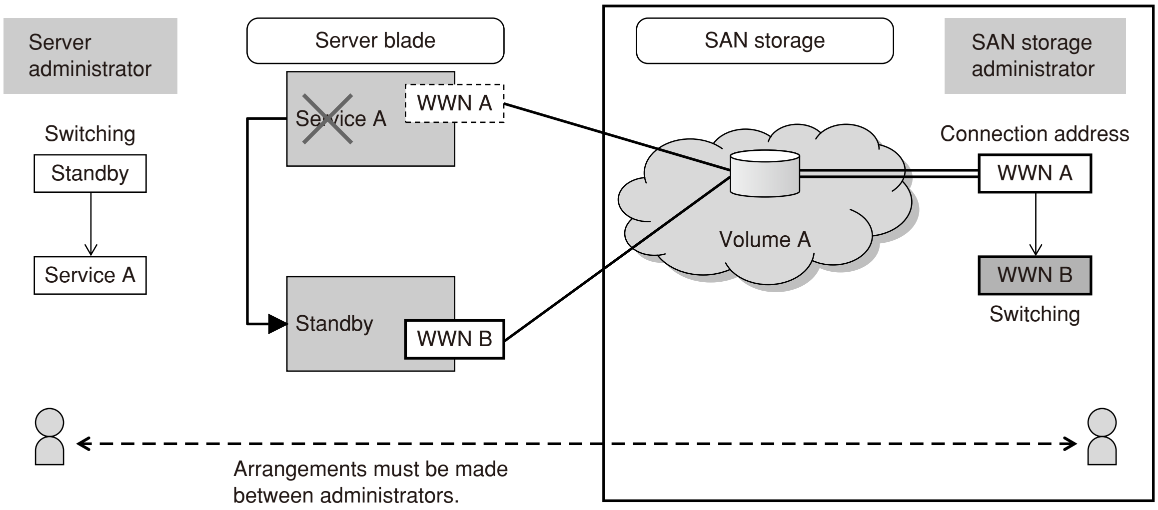
WWN is saved on the FC expansion board in the new system as follows: Each time a virtual WWN is written to volatile RAM of the FC expansion board, it can be recognized from the SAN side, while retaining the WWN fixed for the FC expansion board. When power is turned off, the virtual WWN is cleared and only the fixed address remains valid. Therefore, when a server blade is pulled out from a blade system and inserted into another blade system, the server blade can be used without any trouble.

No special hardware is required to provide SAN connection virtualization. In other words, the FC expansion boards and FC switch blades can be used without modification since both conform to FC vendor specifications as the industry standard. This improves compatibility with the existing external SAN environment.

4. Reduced power consumption

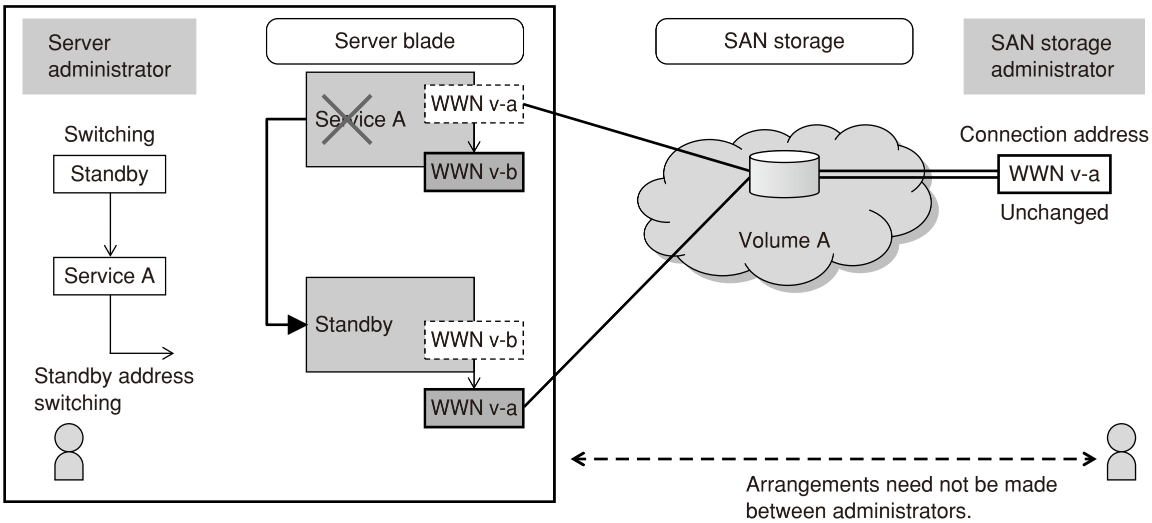
In response to recent growing needs for compliance, security to prevent illegal access and information leaks, business continuity in case of recovery from a disaster, and diversification of risks, system installation sites have been rapidly concentrated at datacenters. Datacenters often adopt a rack-mounted type of system or a blade-type system in considering system integration and collective settings.

In line with the growing concentration



note) WWN A and WWN B are fixed World Wide Names.

Figure 3
Conventional server switching method.



note) WWN v-a and WWN v-b are virtual World Wide Names.

Figure 4
Server switching method by SAN connection virtualization.

of equipment at datacenters, ways to reduce datacenter running costs, such as expenses for electricity, investments in air-conditioning equipment, and management cost are attracting more attention.

The increased power consumption of a single server system is one of the main reasons for greater attention being paid to reducing such running costs. Server power consumption has increased proportionally with higher levels of server performance. For example, server power consumption in 2006 was about 2.5 times as high as that in 2002. This increased power consumption also entails a higher power cost and greater heat output, which makes the installation of high-density equipment more difficult.

To solve these problems, the authors developed a blade server that consumes low power.

This blade server is the PRIMERGY BX620 S4[2)] equipped with server blades containing high-performance, low-voltage, quad-core, Intel Xeon 5300-series processors and 2.5-inch, serial attached, SCSI (SAS) hard disk drives.

At the stage of designing the server blade motherboard for the PRIMERGY BX620 S4, both electrical design and thermal design were conducted by considering the low-voltage Xeon processors in addition to regular Xeon processors. As a result, CPU power consumption was reduced by a maximum of about 60% compared with conventional processors of equivalent performance (**Figure 5**). This reduces the power cost by about 15 000 yen per server blade and 150 000 yen per enclosure when running the system 24 hours a day, 365 days a year.

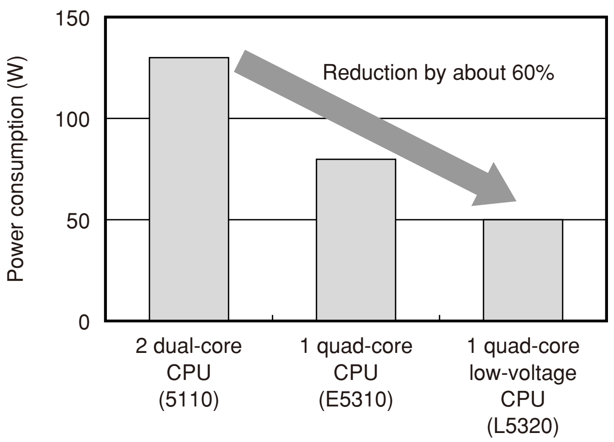


Figure 5
Comparison of power consumption among processors of equivalent performance.

The hard-disk form factor was also changed from 3.5 inches to 2.5 inches. Therefore, the higher integration of LSI of the new hard disk drive reduced the power consumption of electric circuits to about half that of the 3.5-inch SAS hard disk drive having equivalent capacity and rotational frequency.

5. Conclusion

This paper introduced the new problems posed by blade servers and the three technologies adopted to resolve these problems, based on our extensive experience in server consolidation. This paper mainly described new network functions and virtualization technology. Recent years have witnessed growing demands for lower power consumption and less heat output at datacenters, as well as for lighter, higher-density equipment. To satisfy these demands, Fujitsu will continue advancing power saving technology, high-density packaging technology, and virtualization technology based on the results of blade server development, in order to develop the next generation of superior servers.

References

1) Fujitsu: TRIOLE BladeServer. (in Japanese). *http://primeserver.fujitsu.com/primergy/blade/*
2) Fujitsu: Fujitsu Launches PRIMERGY BX620 S4 Blade Server. *http://www.fujitsu.com/ph/news/pr/20070514.html*
3) Fujitsu: Secure Switch SR-S Series. (in Japanese). *http://fenics.fujitsu.com/products/sr-s/*
4) Fujitsu: Systemwalker Resource Coordinator. (in Japanese). *http://systemwalker.fujitsu.com/jp/rc/*
5) Fujitsu: Systemwalker Resource Coordinator Virtual server Edition *http://www.fujitsu.com/global/services/software/systemwalker/products/rcve/*

Hideaki Fujimaki, *Fujitsu Ltd.*
Mr. Fujimaki joined Fujitsu Ltd., Kawasaki, Japan in 1979. Since 1996, he has been engaged in product planning of PC servers.

Masatoshi Shinoda, *Fujitsu Ltd.*
Mr. Shinoda joined Fujitsu Ltd., Kawasaki, Japan in 1987. Since 2006, he has been engaged in product planning of PC servers.

Masakazu Matsubara, *Fujitsu Ltd.*
Mr. Matsubara joined Fujitsu Ltd., Kawasaki, Japan in 2003, where he has been engaged in product planning of PC servers.