Future Trends of BladeServer: Virtualization and Optimization

By Tadashi OKANO*

ABSTRACT BladeServer is a server mounting form as well as pedestal type and rack mounting type. BladeServer is classified high-density type and high-performance type as usage. The requirements for each type are different, and strong points are also different compared with the conventional type server. Each type of BladeServer produces the new added value as well as physical consolidation. As for the high density type, cost reduction of operation management is a strong point, and the high performance type gets on the high usability. As enforcing these strong points, BladeServer has improved in its potential to evolve the system architecture to virtualization and decentralization. BladeServer and VALUMOware will contribute the evolution of IT platform and support the Dynamic Collaboration.

KEYWORDS BladeServer, Virtualization, TCO reduction, High usability

1. INTRODUCTION

The information technology basis which supports Dynamic Collaboration is claimed to adapt flexibly and rapidly to the continuously changing environment. BladeServer and VALUMOware will realize TCO (Total Cost of Ownership) reduction, and will flexibly cope with changes rapidly as a platform which supports Dynamic Collaboration.

The BladeServer definition which IDC recommends shoots at a target. "BladeServer is a computing system which has a processor, a memory and a network connection interface on a single mother board. The board which the processor is carried on is called "Blade," and a piece of Blade is counted as one server. BladeServer with modular architecture, which looks forward to high-density packaging, can add the necessary number of blades easily. High flexibility and high expandability are the advantages of BladeServer. It is inserted into an enclosure (Sub rack) which can accommodate several Blades, and made to work by a single-board computer. Power supplies and cooling fans are shared in the enclosure."

The popular name of BladeServer is a server mounting form as well as pedestal type and rack mounting type. Furthermore, this new product has the potential and ideas to change the concept to configure the system with new added value.

RLX324, which RLX Technologies shipped in 2001, was the beginning of BladeServer history. However, more than one removal circuit board mounted on chassis and sharing a power supplies and cooling fans existed as a Compact PCI standard before that, but they were merely the extension of the standardization of CPU or I/O module. They were a little different from today's BladeServer, since more than one server was connected via the network, and expanded over the scale out architecture.

2. ELEMENT TECHNOLOGIES OF BLADESERVER

2.1 CPU and HDD

The strong point of the mounting density is easy to understand, and this is often quoted as a strong point of BladeServer. As the RLX324 which was the first BladeServer, it was believed that the mounting density of the server was the core value of BladeServer at that time. The core technology to raise the mounting density of the server is the adoption of a power saving device. CPU and HDD consume important share of electric power at a conventional type of server.

RLX324 solves this problem in the CPU by adopting Crusoe designed by Transmeta Corporation, and 2.5-inch IDE HDD for the note PC. RLX324, which accordance with that name, 24 servers can be carried in 3U size chassis, the total amount of electric power per chassis was only 360W, and did full loading in the 42U rack, the electric power consumption was only 5KW. That is a part of the reason that appeals for saving electric power, an electric power crisis occurred in California in those days, and RLX Technologies advertised the necessity of reducing the electric power consumption of the server. Though it is considered that electric power crisis had been only an imaginary fear, the power consumption per rack so far

*Client And Server Division

restricts the density of BladeServer. In the general datacenter, 9~12KW per rack is the limitation of the cooling equipment and the supply of electric power equipment. This electric power restriction influences the mounting density of BladeServer substantially.

The point is that performance is sacrificed to raise mounting density alternatively. As RLX324 was almost limited to use only for the Web server, the scale-out approach was adopted to enhance the performance by increasing the number of blades. On the other hand, the performance per server is important in an AP server and a DB server or a server of the HPC area. The same performance as a general-purpose server is looked for in these servers. The latest Intel CPU consumes electric power of about 100W whereas Crusoe designed by Transmeta Corporation adopted in RLX324 consumes only some W of electric power. Adopting the latest CPU (assumedly 300W per server), electric power consumption exceeds 12KW per rack, assuming that 42 numbers of 1U density servers are mounted on the 42U rack. With a server of this use, we must give up the advantage of mounting higher density than a rack mounting type server from the viewpoint of electric power consumption.

2.2 Storage

When thinking about BladeServer architecture, the storage form is important. In general, pedestal type and rack mounting type servers have some high-performance disks locally (Direct Attached Storage). These disks are used for boot image storage of the OS, swap area, dump area and data storage area, and that composition influences the performance and reliability of the server. Therefore, conventional types of server generally adopt the disk array controller, and construct the array disk of Raid5 by using a plural number of disks. On the other hand, BladeServer has a space restriction; it is difficult to use the same array disk composition as that of conventional types of server. Because of this, BladeServer tends to have the minimum disks locally, and has a dedicated disk array device on the network (Network Attached Storage), or SAN (Storage Area Network). The advantage of having the dedicated disk array device are performance, reliability and easy capacity reservation. Furthermore, a variety of functions such as Dynamic Data Replication, difference data backup, etc. is attractive. However, it is difficult to move all storages to the dedicated disk array device in the current technical environment. Bottleneck is an OS function such as booting OS, swap, and dump functions. From now on, technical advance will allow these functions, and the combination of BladeServer and dedicated disk array device will become more general in the future.

2.3 Network

The network is an indispensable function for BladeServer as well; but we should note the restriction of network compared to conventional servers. BladeServer exchanges interface signals via backplane. Consequently, Backplane interface such as the number of ports, connection relation cannot be changed. Conventional servers easily add ports by PCI NIC, or change connection by cable antithetically. In addition, the physical number of external ports in dedicated switch restricts the network expandability. Tag VLAN technology eases restriction of the number of physical ports. Nevertheless, the physical port is necessary in every segment when BladeServer comes in the existing network. Also, the inter-chassis connection restricts the usual network bandwidth even if built-in network switch has sufficient bandwidth. Therefore, network traffic should box in the chassis for certain applications such as PC cluster (the interconnect network bandwidth is especially important).

Although we mentioned the restriction of the network in BladeServer, there is an advantage of adopting a built-in network switch. The first advantages are reduction of the construction time by cable physical integration and improvement in reliability. For example, each of 20 servers has two network ports, the cable from chassis overreach 40 cables. By adopting a built-in switch, an external cable can be integrated. Reduction of the number of cables facilitates drastically the installation work time. Moreover, server and switch are inter-connected via backplane, and cable mishandling which accounts for the large part in maintenance trouble, is reduced drastically. And built-in switch contributes to virtualization as described later.

3. CLASSIFICATION OF BLADESERVER BY USE

Next, we will introduce the classification by use and advantage of BladeServer in each class. State of the art network-centric system produces server hierarchy and provides overall cost reduction, easy solving of problems, flexible correspondence to the performance requirement and improvement in reliability. In general, servers are hierarchically classified into 2 - 4 class. The requirement in each class is different, so we discuss the following 4 classes separately.

3.1 Edge Server

An edge server arranged at the front end of the network terminates and/or changes the network protocol, and hand over the process of Tier1 server. Processing in this tier is stateless, and reliability is not required. (When a server breaks down, other servers take over processing.) The network performance is required to process a short packet in large quantities though CPU performance is not required by each server. For each server, less environment loads (the electric power, the volume and the weight) for establishment is important because the number of servers increases corresponding to the processing load.

3.2 Tier1 Server

Typical use of the Tier1 server is a Web server. Almost all requirements are equal to the edge server and tend to operate in the same server. It receives concentrated traffic, and hand over the process to Tier2 server.

3.3 Tier2 Server

Typical use of the Tier2 server is an application server. These applications require CPU performance and I/O performance, but some applications are dispersed to reduce the server load to other servers connected via network. Many applications have state record, and that requires reliability as well. Process is received from the Tier1 server and hands over to the Tier3 server, but Tier2 server tends to operate in the same server as Tier1 or Tier3 server when a system is actually constructed.

3.4 Tier3 server

Typical use of the Tier3 server is a database server. In case of a renewal type database, it deals with the process with state record in the limited number of high-performance servers. Multi-CPU beyond 4CPU and I/O expandability is necessary to enhance the performance and reliability.

4. THE VALUE OF SERVER CONSOLIDATION

To classify these types of servers with a viewpoint of the requirement of BladeServer, they are divided into the edge and Tier1, Tier2 and Tier3 roughly; in other words, the high density (CPU performance is not required), and high performance/reliability classes. They can realize the following advantages by substituting BladeServer for the system building with a combination of conventional servers.

4.1 The Advantage of High Density Type BladeServer

An edge and the Tier1 server are asked for flexibility corresponding to the concentration of process. Additionally, high density, saving electric power which does not become the load of the establishment environment, and easiness of the control of management is required. Each server does not required for high performance. Considering the work to establish and remove a conventional server so far, the effect is obvious and tremendous substitutes BladeServer for the edge and the server of Tier1. The conventional type of servers, even pedestal type or rack mounting type servers, needed to install each server, wire the power and the network cable for each server, connect console cable for each sever, install OS for each server, install application for each server and setting of each server. Enormous time is necessary for installation. Configuration changes corresponding to the amount of process needs several days. It is difficult to accommodate the short time traffic concentration.

After substituting BladeServer, work is very much simplified. At first, mounting a chassis on a rack is common when adding a new server, Blade only put it in chassis physically. Wiring of every server is completed inside the chassis without the need for additional wiring, and there is in no danger of omit wiring of the next server by accident, too. After that, the work of the rest is carried out automatically with the help of middleware. In the case that content of installation is registered in each slot of Blade, the setup work of each server is carried out in accordance with the scenario. Whole system configuration change completes in several hours, it is not difficult to carry out the optimization by the server reallocation several times a day along with the traffic load.

There is an additional advantage that the high density and saving electric power bring. The data center setting environment (electric power, cooling equipment, weight, network) for each rack is limited, and a server beyond the limit of setting environment cannot be installed. Moreover, each resource worth the cost, it is important to use limited resources effectively for cost reduction. High density type BladeServer can mount a larger number of servers in comparison with the conventional server in the limited space, it makes possible to use the limited resource of datacenter effectively.

4.2 The Advantage of High-Performance Type BladeServer

Next, we will examine the function of BladeServer used in Tier2 and Tier3. In these tiers, the

performance and usability of each server are important. High-performance CPU generate much heat; thus they need a larger space for cooling. Fortunately, these tiers of server are in small number, and density is not so important. Most traditional applications are programmed based on SMP, and many servers beyond 4 way are used for it. But some applications have the function to disperse the workload to other servers via network. In addition to the improvement of CPU performance, the needs of multi-way server decrease by degree.

Improvement in the I/O performance is a more serious subject for BladeServer. There was a case that network performance and disk I/O performance became the whole bottlenecks with conventional Tier2 and Tier3 servers, and this problem was solved by expanding an I/O device in many cases. In other words, I/O expandability of such as LAN port, disk array controller, FC HBA is connected directly with the whole I/O performance. As mentioned above, BladeServer is restricted its I/O expandability by backplane, it is difficult to expand I/Os by I/O performance upgrade demand. And it is unrealistic that backplane always has the maximum expandability. There is a hint to solve this problem in BladeFrame shipped by Egenera company, which separates dedicated I/O controller, and connect between control blade and I/O blade by high-speed inter-connection. But this architecture is more expensive in cost than a conventional server, so that it may take a long time to become a general solution.

These tiers of server should attach weight to high usability so as not to interrupt the service, not to give up a system. Cluster technology which raises usability of the server, makes redundant server composition, detects the server trouble by monitoring mutually, and responds to take over process automatically in other servers. Minimum hardware requirement is interconnect for the heartbeat to monitor trouble mutually, and host bus adapter to connect a shared disk array device, and it can realize Blade easily.

Arranging this technology; there is a new technology of provisioning which the role of the server is changed automatically by the schedule or in accordance with the concentration of traffic as for autonomy. For example, when the load of the AP server rises, another server that a load is comparatively light is reconfigured for the AP server, and cooperates the process as for autonomy. Through this reconfiguration, not only a CPU but also network and storage must be rearranged for the AP server. The basic technology for this process is the virtualization of the resources.

Basically, defining association function between the physical resources and logical resources allows the process transfer to another server by only rewriting the link table. And progress of resource communization makes it possible to minimize the influence of switching. For example, virtualized common storage minimizes the influence of switching rather than DAS (Directly Attached Storage to each server). In a similar manner, virtualized common I/O controller minimizes the influence of switching rather than server-fixed I/O controller. Resource communization becomes easier by server consolidation in Blade.

5. ADDITIONAL VALUE OF BLADESERVER

5.1 TCO Reduction

When discussing the advantage of BladeServer, "TCO reduction" is the keyword. Though BladeServer is comparatively expensive than conventional server, it can reduce the whole lifecycle cost including not only purchase cost but also operation cost. At the early time, the standard for the TCO reduction is saving electric fee and utilization fee of datacenter. It is the reason to reduce these costs for the whole server life cycle by substituting high density, low electric power server. It is right in calculation, but an operation manager in charge is different from the purchase decision person, and the ratio of these costs is relatively low that it cannot become a conclusive factor; it has been gradually going out of use as an appraisal standard.

Conversely, operation and management cost become the standard for the cost reduction. After installing the server, it always changes configuration in search of the most suitable composition. Total cost becomes larger if the configuration of servers cannot change quickly and flexibly. In addition, following operation is necessary; when server breaks down, switching to the substitutive machine, and always watch the condition of the server to avoid a trouble. The bottom line of these costs is personnel expenses and difficult to estimate effect on a reduction. But the advantage to reduce the operation and management costs is the important reason to purchase BladeServer in comparison with the conventional server so far as mentioned above.

5.2 The Additional Value of Blade Chassis Unit

What is the additional value of BladeServer? As mentioned above, the function that a conventional server did not have gives BladeServer additional value so far. The first value comes from the physical form of BladeServer. The main part of BladeServer is

constructed by a module, and is connected by backplane going through. Almost all modules are hot-pluggable, and that makes a substantial contribution to reduce the cost of initial settings and operations. Additional advantage is that physical mounting density rises more than the conventional server by sharing power supplies and cooling fans.

The second value comes from the built-in CMM (Chassis Management Module). Almost all of function that CMM realize was adopted by the conventional server; though an optional function from the cost side or which was provided only for the expensive server like 4 way. BladeServer shares CMM with several servers inside chassis, and realizes an equal function by the lower cost.

The third value comes from built-in switch. A switch not only consolidates physical cables but also plays an important part in the virtualization. Even the external switch realizes a similar function; limited configuration brings to ease the design, evaluation of the middleware for virtualization. Built-in network and storage switch bring more advanced virtualization.

5.3 Restriction of the Chassis Unit

While the chassis unit provides the new function so far, the chassis unit also restricts the evolution of BladeServer. The function of chassis unit will dictate the strong point of BladeServer through the future. BladeServer is connected with the outer world and each module inside the chassis via backplane going through. In BladeServer, there is no de facto standard so far, each company decides the backplane interface to devise it, but the conclusive interface is not decided. Even if a standard interface is decided, expansion is restricted to that backplane interface.

For example, as for the LAN interface that is a typical I/O interface, 2 - 4 port Giga Ether interface per server is standard at present. It seems to be sufficient compared with the common sense of the conventional server so far. The point is that BladeServer cannot expand interface anymore. In the conventional server, the number of ports can be easily increased by using the PCI expansion card when it is necessary. However, BladeServer is connected with the outer world via the backplane going through; the interface that is not wired on the backplane cannot connect the signal. And even if 10 Giga Ether becomes general, an interface cannot be changed easily. The board of chassis becomes a restriction, too. A connection between chassis becomes a usual network connection and a bandwidth is restricted drastically even if module of chassis inside is connected with high-speed switch.

Separated I/O function architecture can clear way for the restriction of I/O expansion. Separated CPU and I/O function is connected with high-speed virtual interface each other. These interfaces must be concealed from the application visibility to transparent as of these virtual interfaces. For example following operations become possible by separating an I/O function from the CPU and connecting high-speed virtual interface; adding I/O interface in accordance with the performance requirement, or substitute high-performance CPU module without I/O module change.

6. THE FUTURE OF BLADESERVER

The history of the server started from the client-server model. That is to say, an office computer installed in the office is connected to several numbers of terminals, and data from the terminal is accumulated and processed in the server. The center of processing changed to the client side caused by the improvement in the CPU performance and the advance of the network technology after that and the role of the server changed in the file server and the mail server. The form to use a server via the network is going on until today. It became general that the server put together physically, and manage in the machine room safely and efficiently in about the end in the 1990's. Housing service business is an extension of this stream and consigns machines in the enterprise and operates in the professional datacenter. At that time, pedestal type might be general and consume big space physically, and take a lot of work in management.

In the same time, hosting service business of rental server is occurred. In the form to lend a special prepared server to the user; one server was lent to one user - several users. The datacenter recognized that the need to manage many servers efficiently due to the appearance of hosting service. It is because managing many servers efficiently and reducing the cost become the source of the cost competitiveness. Physical putting together that is stated pursuance of efficiency changed to the new motivation to enhance efficiency.

On the other hand, massive servers of Tier2 and Tier3 are asked for logical putting together of the server function. A typical example is putting together of Storage. While each HDD becomes big capacity, reliability does not advance, important data are collected in the high-performance disk array device, and managed, and server concentrates on the calculation function. It is the flow to increase the efficiency of

investment. This flow has been reaching the server of Tier1 which common Storage was not necessary so far, and it is anticipated to become a trend.

The server virtualization is being derived from this flow. The entire server connected to the common storage means that the data which each server has can be shared, and makes it easy to substitute other servers. Virtualization of servers allows the new function and added value such as; disperses heavy process to other servers. And, the server that an I/O function was virtualized becomes possible to improve I/O performance only to put I/O function to server. If I/O function is fixed to the server, a server has to be replaced entirely. But absolutely achieving these function needs the virtualization of sever environment such as Storage and Network. It is being expected to become general technology by the advance of the future technology.

A thing to come in the next of the virtualization is the flow of the dispersion. Virtualized resource loses the necessity to fix physically, and can be rearranged freely. Grid is being watched as a technology which the dispersed resource can be used effectively. When this technology advances, for example Storage data are dispersed around several points and prevent from being lost due to the disaster, or rearrange logically flat data to the most suitable place in accordance with the use form, becomes practical.

7. CONCLUSION

Concept of BladeServer starts from one form of the server, and grows to be a product which keeps the possibility to change the general idea of the server firm. There will be technical subjects to get over toward the goal in the future. It is certain that BladeServer will become the main stream of the server in the future. BladeServer supports Dynamic Collaboration with the VALUMOware and realizes each function of autonomy, dispersion, virtualization, cooperation, and contributes to the social development.

ACKNOWLEDGMENTS

On the occasion of developing the BladeServer system, we deeply appreciate the cooperation and advice of following sections: Computers Software Division, NEC Software Hokkaido, NEC Software Hokuriku, NEC Computer Techno, and NEC System Technologies.

REFERENCE

[1] "Japan BladeServer market trends," IDC Japan.

*Names of companies and products in this paper are trademarks or registered trademarks of each company.

Received April 30, 2004

* * * * * * * * * * * * * * * *

Tadashi OKANO received his master's degree in electronic engineering from Tokyo University of Agriculture and Technology in 1984. He joined NEC Corporation in 1984, and is now Engineering Manager of the 1st Engineering Department, Client And Server Division. He is engaged in the processor design of main-frame and supercomputers. He is also engaged in the product planning of Blade Server.

* * * * * * * * * * * * * * * *