

WHITE PAPER

Highly-Reliable Server Required in Datacenter: Fault-Tolerant Server Is a Key Solution

Sponsored by: NEC

Kazuhiko Hayashi

October 2014

IDC OPINION

There's an increasing number of user companies that are considering cloud services when upgrading or installing new information systems. Results of the 2013 user survey showed high interest in cloud services, with more than half of all respondent companies indicating that they have already used cloud services or they are planning to use cloud services in the future.

Cloud services are categorized as "public cloud," in which a third party provides the cloud equipment and services, and "private cloud," in which the services and equipment are provided in-house or by an affiliate. By far, between these two categories, private cloud is the larger market in Japan. Compared with their international counterparts, user companies in Japan appear to have a much stronger preference for private clouds operated in-house or by affiliates, and for onsite environments. One of the primary reasons why so many Japanese users want private clouds and on-premise environments is that they are assured of system reliability and availability. IDC believes that most companies consider it important for their employees or affiliates to be able to identify the source of failure and control it when something goes wrong with their private cloud or on-premise environment. Many IT systems make use of the advanced virtualization technology and parallelization to ensure availability and provide a measure of reliability, but greater reliability will be required before core systems that must operate 24 hours a day, 365 days a year can move to private cloud. One of the crucial challenges facing IT departments is how to improve overall system reliability and ensure speedy maintenance and management.

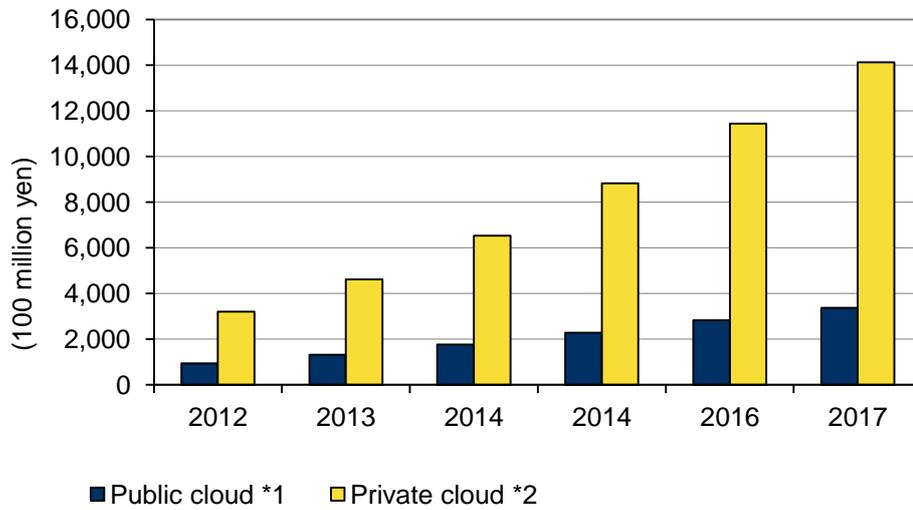
SITUATION OVERVIEW

Figure 1 divides the Japan cloud services market into the "public cloud market" and "private cloud market," and forecasts their sizes for the 2012–2017 period. The Japan private cloud market is expected to be more than triple the size of the public cloud market. During the 2012–2017 period, private cloud will have a compound annual growth rate (CAGR) of 34.5% as opposed to the public cloud's 28.9%.

In the cloud services market, there is strong preference for private clouds and onsite systems that can be managed by in-house or affiliate IT staff, instead of public clouds that are managed and operated by third parties. IDC believes that private cloud will be used for the vast majority of core information systems that handle crucial data when those systems are moved to the cloud.

FIGURE 1

Japan Public Cloud Service/Private Cloud Market Spending
Forecast, 2012–2017



Notes:

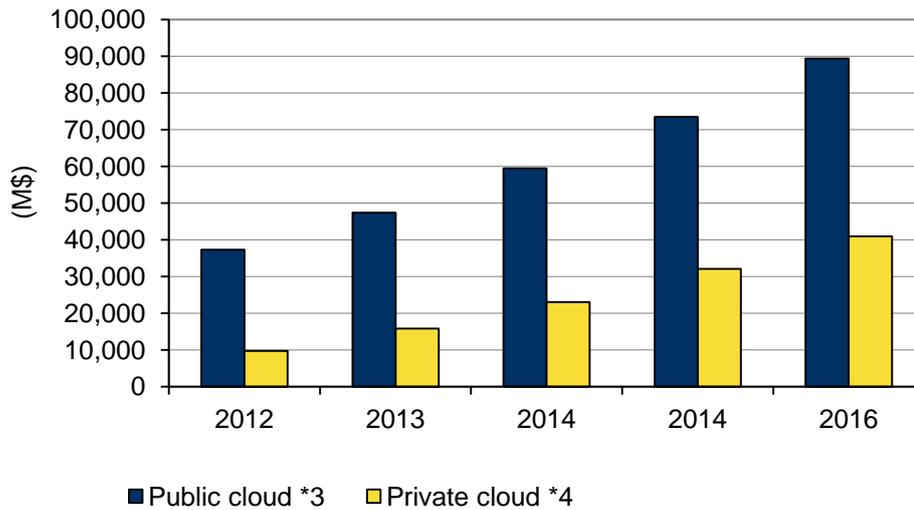
- * 1 Based on data from *Japan Public IT Cloud Services 2013–2017 Forecast Update* (IDC #JP1972808V, January 2013).
- * 2 Based on data from *Japan Private Cloud Services 2013–2017 Forecast and 2012 Review* (IDC #J13280105, October 2013).
- Calculated from vendor sales

Source: IDC Japan, August 2014

Figure 2 compares the sizes of the global public cloud service and private cloud service markets. In the global market, private cloud is less than half the size of public cloud. This is an indication that most user companies in Japan have strong resistance to placing the management and operation of their cloud systems in the hands of outside operators and third parties. Compared with the global market, most user companies in the Japan cloud services market consider it extremely important to ensure the security and reliability of their data with onsite IT environments that they themselves or their affiliates operate.

FIGURE 2

**Global Public Cloud Service/Private Cloud Service Markets
Sales Forecasts, 2012–2016**



Notes:

- *3 Based on data from *Worldwide and Regional Public IT Cloud Services 2013–2017 Forecast* (IDC #242464, August 2013).
- *4 Based on data from *Worldwide Hosted Private Cloud Services 2012–2016 Forecast: New Models for Delivering Infrastructure Services* (IDC #238689, December 2012).
- "Private cloud" refers to the sum of hosted private cloud and virtual private cloud (VPC) infrastructure as a service (IaaS).

Source: IDC Japan, August 2014

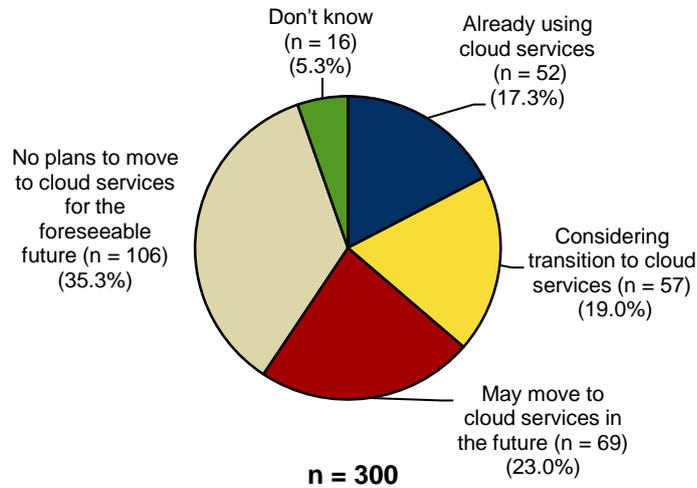
Figure 3 shows the results from a survey on cloud services of user companies — with at least 10 employees — that are planning on upgrading or installing servers.

Companies already using cloud services account for 17.3% of the total respondents; companies considering or planning to move to cloud services accounted for 42.0%. This shows a high degree of interest in cloud services among user companies. It also shows that most user companies are considering either to create their own systems or to make use of cloud services when bringing new systems online.

The survey asked companies that had transitioned to cloud services about the challenges they encountered. Among the answers, "concerned about failure response" got the second highest percentage of response rate, with 13.5%, which is on a par with "concerned about security" and "lack of cost-savings" (Figure 4). User companies apparently consider the reliability of cloud services to be an equally important issue as security and cost.

FIGURE 3

Use or Consideration of Cloud Services

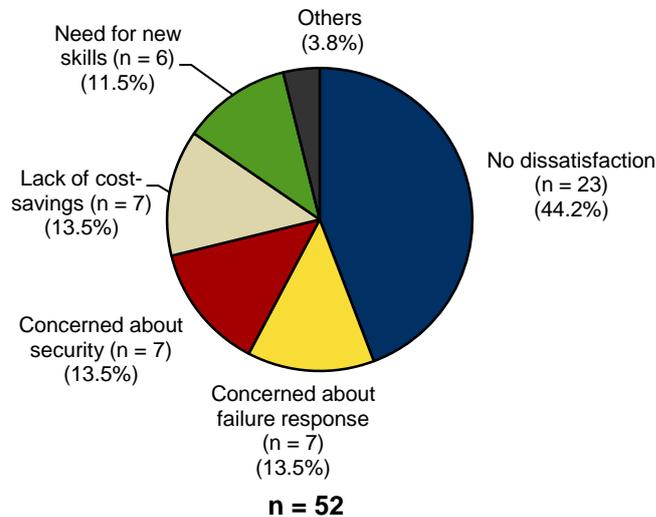


Note: Based on data from *Japan Server Systems Integration User Needs 2013* (IDC #J13230103, December 2013).

Source: IDC Japan, August 2014

FIGURE 4

Challenges in the transition to cloud services



Note: Created from data in "*Japan Server Systems Integration User Needs 2013*" (IDC #J13230103, December 2013).

Source: IDC Japan, August 2014

RELIABILITY OF PRIVATE CLOUDS AND ON-PREMISE ENVIRONMENTS

IDC thinks that cloud services will drive the market as the main focus of future information systems, but as cloud services become popular, user companies will be increasingly interested in system reliability and security, not just costs. For cloud services to satisfy both cost and reliability requirements, IDC believes they will need to offer customers a wider service menu, with options that emphasize cost-savings and other options that emphasize reliability.

As previously stated, user companies that are considering a move to private cloud or to an onsite datacenter view reliability as a crucial factor, not just cost-savings. This is particularly the case with core systems that must be in operation 24 hours a day, 365 days a year with not even a second of downtime. Users must be able to choose hardware according to their required service levels before moving to a cloud environment. Current cloud services are set to achieve advanced virtualization technology to enable system redundancy and parallelization that respond to failures; however, there are cases in which software-based redundancy and parallelization are not enough to satisfy user requirements. Examples include user authentication systems and online transaction systems, which must not experience any single second of downtime. Fault-tolerant servers ("NEC FT servers") are useful when servers require high levels of hardware reliability.

NEC's FT servers, such as the "Express5800/ft series" servers, are designed as single servers to achieve system availability by employing unique, redundant hardware components in an individually modularized configuration. The redundant modules operate in sync. When the system detects a failure in one of the redundant modules, it is automatically isolated and the other module keeps the system running. To the network and OS, it appears to be a single ordinary server. Applications likewise require no specific redundancy settings. One of the things that set this configuration apart is that it can operate as a single server, which simplifies management and operation. In particular, it can operate like conventional systems in a private cloud or onsite environment, which makes the FT server extremely useful when reliability is essential.

NEC'S FAULT-TOLERANT SERVERS

Features and Advantages of the Fault-Tolerant "Express5800/ft Series" Servers

NEC has offered x86 server-based "Express5800/ft series" since 2001. The FT server incorporates a proprietary redundant chipset LSI for its control. This LSI is the heart of the FT server and the mechanism to control the redundant hardware is the hallmark of FT servers. It performs advanced synchronization, server management, failure detection, and switchover. The company modularizes all of the major components other than the backplane to enable quick replacement of a module should a failure occur, and modules can be replaced while the system remains in operation.

In March 2014, NEC announced the fault-tolerant server "Express5800/ft series" Windows-model "Express5800/R320d-M4" for the global market. It features a 20% improvement in processing performance (Figure 5). The servers have moved to Intel's new generation of Xeon E5-2670v2 (Ivy Bridge-EP) processors with 10 cores per processor from the previous 8 cores to achieve the 20% performance boost. The model can mount up to two processors, achieving a maximum of 40 threads, 256GB memory, and up to 9.6TB internal disk capacity, making it suited to large-scale

virtualized environments. On the network side, the unit has two 10 GBASE-T Ethernet ports. Additionally, new support for SSD (eMLC) accelerates I/O processing and improves enterprise-grade functions.

It supports the Windows Server 2012 OS to enable use of enhanced Hyper-V, the latest virtualized environment, which provides a huge boost in capacity. The OS specs have increased the number of virtual processors per virtual machine from 4 on Windows Server 2008 R2 to 64, and the memory per virtual machine from 64GB to a maximum of 1TB. Likewise, virtual disk capacity has expanded from 2TB to a maximum of 64TB, and the number of virtual machines from 384 to 1,024.

The "Express5800/ft series" has two modules corresponding to one physical server that operate in sync and constantly monitor each other's operation through the "GeminiEngine" redundant chipset. While hardware is redundant, software does not know which of the modules is operating, and from the perspective of the system, it is equivalent to operating on a single piece of hardware.

Because of this, under existing OSs and applications, the "Express5800/ft series" has the following features, which do not make the user aware of the FT server at all.

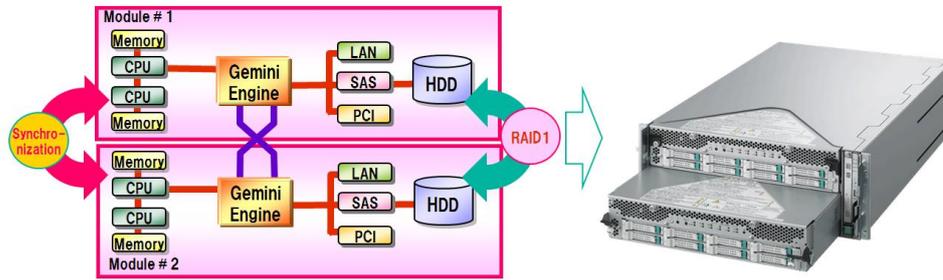
- Exists on networks as a single server with no perception of redundancy
- Perceived by the operating system and application software as a single server that requires no special processing or applications for redundant operation
- Can be treated like a single server for operations and management

However, it should be noted that there are some restrictions on compatible devices because the drivers that support the "Express5800/ft series" are required for networks, discs, Fibre Channels, USB connections, video display, and other I/O interfaces.

One distinctive feature of the "Express5800/ft series" is that it has a structure that indicates the location of hardware failures when they occur to enable quick replacement of the component that failed without having to shut down the system. Its biggest benefit, however, is that existing OSs and applications can be used without modification to provide a high-availability system that does not make users aware of a special hardware that has been installed. NEC explains that the "Express5800/ft series" has been designed to achieve a "five-nine" (99.999%) availability, so the installation of the server can be expected to achieve an extremely high hardware availability. The company's FT servers have a solid track record in plants, workplaces, and other applications that require high availability worldwide.

FIGURE 5

Fault-Tolerant Server "Express5800/R320d-M4" Specs



- CPU: Maximum of 10 core processors, processing of up to 40 threads
- LAN: Two 10 GBASE-T ports standard
- HDD: Max 9.6TB
- SSD: 200GB SAS SSD (eMLC) available
- OS: Supports Windows Server 2012. Support Hyper-V

Source: NEC, 2014

Examples of FT Servers in Operation

One example of the NEC FT server in operation is the "UNIVERGE SV9500" IP-PBX product, which offers an FT server on the "UNIVERGE SV9500 UC server model" in Japan. IP-PBX refers to a private telephone exchange installed within a company that brings both ordinary telephones and smartphones into the internal exchange while also allowing internal communications among multiple locations. They also support the full gamut of audio and visual communications over PCs and smartphones, including presence, instant messaging, and web conferencing to achieve what the company calls a "Unified Communication" (UC).

NEC uses the FT server as the hardware platform for the "UNIVERGE SV9500 UC server model", which requires the highest "Reliability, Availability and Serviceability" (RAS) in its IP-PBX product line. The redundant hardware and monitoring functions of the FT server mean that hardware can be instantly isolated in the event of failure, with operations automatically switching over to the functioning hardware. The modularized hardware enables replacement of failed components while the system remains in operation, which represents a significant boost in reliability and maintainability.

The fact that the NEC FT server is able to achieve the high levels of reliability required for exchange systems indicates that it is an effective option not only for cloud services and onsite environments, but also as a replacement for servers in other areas that require high levels of availability and service.

NEC's Business Opportunities and Challenges

The NEC "Express5800/ft series" is the only FT server offered by a Japanese vendor, but users who want high-availability servers have many other options to choose from, including HA clustering software. It will be necessary for the company to explain in a way that user companies can easily grasp the differences between hardware availability and software availability, and the pros and cons that each approach has in terms of operations and response to failure.

CONCLUSIONS

Summary

Cloud services are becoming increasingly common among user companies. In the Japan market, there is high interest particularly in private clouds in which the user company itself or an affiliate controls the IT system, and the market for these systems is potentially huge.

Companies that are already using cloud services are still concerned about response to failure, signifying that vendors have yet to fully address such concerns. There are many approaches to improving the reliability of cloud services, including the use of highly-reliable IT equipment, redundant systems, and backups. The choices that users make will be based on a general evaluation of the cost of investment, tolerable downtime, and required service. Within this, fault-tolerant (FT) servers, which offer hardware-level redundancy, will make their greatest mark in critical systems in the private cloud, where short downtime for hardware failures and ease of maintenance are key factors. IDC believes that it will be important for users to choose hardware according to the importance of the system and the cost of maintenance.

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact IDC Japan Sales at +81.3.3556.4760 (jp-sales@idcjapan.co.jp) for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2014 IDC. Reproduction is forbidden unless authorized. All rights reserved.