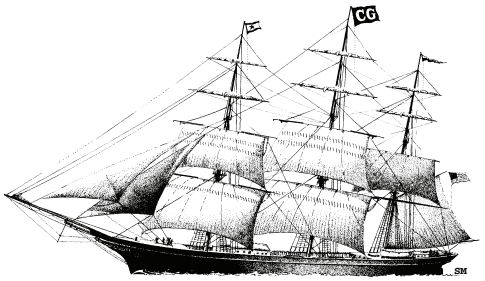
THE CLIPPER GROUP
Navigator™



Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2007056 April 26, 2007

Fault Tolerance for the Windows Environment — When It Positively, Absolutely Has To Be There

Analyst: David Reine

Management Summary

The internet has been the savior of many a harried spouse who has forgotten to buy that special gift for a birthday or anniversary. The late shopper can always find just what he needs at any time of the day or night. Buying the gift is not the problem, however. You also have to get it delivered, and frequently that means **"tomorrow!"** Forgetting a gift is not an option, but the method of shipment can be. The website offers standard USPS shipment estimated in days, 48-hour delivery, or, for those "special" gifts, next day delivery. It is times like this when we remember the slogan from Federal Express – "When it positively, absolutely has to be there!" When we know what is good for us, we are not reluctant play it safe and pay a little more – we go for next-day delivery.

Playing it safe and paying a little more is often the right choice when working in mission-critical application environments as well. When I dial "9-1-1" in an emergency, I "positively, absolutely" want the phone answered. I do not want to hear a busy signal or "Your call will be answered in eight minutes". If there is a fire, I do not want to read that the fire trucks could not respond because the fire station doors would not open because of a computer malfunction. There are certain applications in which immediate response is not an option. It is mandatory. Hurricane Katrina made perfectly clear that this applies to the public sector as well as the commercial world. Now, due to limited funds, restricted budgets and the cost of proprietary solutions, many enterprises and local communities have turned to open systems solutions to solve their mission-critical problems. While *Windows* on an x86 platform can lower acquisition and maintenance costs, it also introduces reliability and scalability issues into the purchasing equation. **Saving money is by no means a replacement for saving a life!**

NEC recognized the requirements for fault tolerance of mission-critical applications on x86 platforms several years ago with the introduction of its *Express5800* line of Windows servers. Over the past few years, NEC released second and third generation systems to take advantage of the improvements in Intel architecture. It has now transitioned the Express5800 line to a fourth generation, the *Express5800/320Fc*, based upon Intel's quad-core *Xeon* microprocessor. This is the first system to deliver *Lockstep* technology on a quad-core system. Along with the improvements in performance and virtualization available with quad-core, NEC has enhanced the I/O subsystem to improve the overall performance of the Express5800 platform. To see how the 320Fc can deliver fault-tolerant performance to your enterprise, and lower the TCO for your data center, please read on.

IN THIS ISSUE

Confronting Data Center Sprawl

The data center's total cost of ownership (TCO) is out of control. The CIO needs to make changes to the basic infrastructure in order to improve performance and reliability, reduce the amount of energy wasted every day, and to improve the profitability picture of enterprise IT. Whether you are the CIO of one of the largest enterprises or an SMB, you face the same problems when confronting the cost of managing a 21st century data center. The most significant issues include:

- Meeting the transactional processing needs of an expanding enterprise with reliable platforms to satisfy a 7x24 SLA;
- Improving the inefficient utilization of server resources;
- Rebuilding the data center infrastructure by taking advantage of both physical and logical consolidation;
- Attempting to reduce excessive operational costs due to energy consumption; and
- Reducing administrative staffing caused by server proliferation.

Data center problems become more complex if the mission-critical applications require a fault-tolerant environment.

The average data center consists of rack upon rack of mono- and dual-processor x86 *Windows* servers arrayed in a scale-out architecture. In the past, IT management rolled out a new server for each new application, attempting to simplify the environment, isolate access, and improve IT security. Unfortunately, this resulted in extremely poor utilization of system resources, with each server using only 15-25% of available CPU power.

In order to correct this, CIOs are changing their infrastructure paradigm. They are now consolidating complementary applications onto fewer, more powerful platforms, sharing system resources. Unfortunately, this has a downside also. It makes your infrastructure highly vulnerable to a single outage. One crash could disable your network until a replacement part is located and installed. This scenario makes a highly available system not only desirable, but also mandatory.

Last year, the majority of x86 servers were configured with the latest dual-core microprocessor technology from one of two companies: Intel with its *Xeon* processor or AMD with its *Opteron* CPU. Due to their reduced energy requirements[1] and the inclusion of virtualization technology into the chip design, data centers have improved their basic performance per watt consumed in running enterprise servers and cooling the data center environment.

The availability of energy-efficient, dual-core servers[2] enables the CIO to lower data center operating costs, consolidating multiple applications from obsolete servers, with expired warranties, onto newer technology. This also frees the budget from maintenance charges. Reducing the energy needed to operate the data center efficiently is a major contributor to lowering recurring costs and the TCO for IT.

With one eye on enterprise profitability and the other on future performance requirements, the CIO must protect the enterprise's investment with a reliable system, capable of scaling processors and memory without the need to replace the entire platform. By addressing these issues, the CIO can regain control of the data center's operating expenses.

In an effort to assist all enterprises in achieving that goal, Intel has made a commitment to a 300% increase in performance per watt by the end of the decade. They have taken a big step toward that goal with the introduction of a quad-core Xeon processor for the high-volume, or two-socket, server, improving both performance and performance per watt for x86 processors. Intel now reassumes the x86 leadership mantel, enabling server vendors to develop platforms that are more efficient.

Consolidating the Mission-Critical Data Center

With newer, higher performing CPUs, the data center is now better equipped to con-

[1] The typical dual-core CPU consumes the same or less energy that its single-core counterpart.

[2] See the March 31, 2007, issue of Clipper Notes entitled *Reducing Cost and Improving Performance – Consolidating the Smaller Data Center*, which is available at http://www.clipper.com/research/TCG2007049.pdf.

solidate a myriad of applications onto single, multi-socket, multi-core processing systems. However, there is more to consolidating a mission-critical environment than simply throwing compute cycles at it. Sharing resources on one system requires a scalable platform with sufficient memory and I/O capability to run multiple applications on multiple operating systems, while managing the allocation of those resources between the operating environments. This requires a *hypervisor*, defined as *a virtualization platform that allows multiple operating systems to run on a host computer at the same time*.[3] The most common examples of virtualization platforms are VMware's *ESX* for Windows and *Xen*, an open source virtualization layer for Linux applications.

With multiple applications active at any given time, the consolidation platform must have sufficient memory to provide an appropriate response time for the increased number of users. As the user count increases, the system demand for I/O traffic can also expect to increase, necessitating increased bandwidth for both the front side bus and expansion slots.

With additional application capacity, more users, and more I/O consolidated on one server, the data center can now retire those older servers that are off warranty. In addition, with the new servers running within the same power envelope as the platforms they replace, the data center can realize significant savings in their energy budget. These steps reduce the TCO for the environment.

One new problem is introduced with this new environment: the consolidated server must be available 7x24x365. **When you put all of your processing "eggs" in one transactional "basket", reliability is no longer optional.**

How do we measure reliability? In baseball, a hitter with a .300 batting average is a reliable performer because he is successful 30% of the time (three hits in every ten at-bats). If he gets four hits per ten at-bats, he is a **superstar**. When we evaluate a mission-critical server, however, one that is available 99% of the time is considered **highly available,** but it will be down for 89.5 hours per year, **almost four full days**. That may be satisfactory for test and development environments, but it is not reliable enough for a mission-critical server that must be continuously available. To be the "superstar", our server must be *fault tolerant* (F/T). With a mission-critical application running on multiple virtual servers, a fault-tolerant system can ensure continuous service with no loss of data. With the ability to continue processing through any outage, the enterprise will realize an improved TCO.

We judge a fault-tolerant solution under a set of spotlights focused upon four major criteria.

- First, how reliable is the platform? Does it have five "9"s availability (a batting average of .99999)?
- Does it meet the application performance requirements? How much overhead does F/T add to standard processing?
- How quickly can the staff restore availability, after the failure of a critical component? and
- What is its TCO? Can I afford it?[4]

Reliability

The data center can achieve one of several different levels of reliability, depending on the hardware architecture employed. A server with three "9"s availability (99.9%) is out of service for almost nine hours per year. That's better than 99% but still not acceptable for a truly mission-critical environment. On this server, the goal is not to try to prevent a crash but to recover as quickly as possible from one. Unfortunately, an enterprise could lose hundreds of thousands of dollars for every hour that its server is down. A fire department could lose a life.

[3] For more on server virtualization, see two issues of Clipper Notes: *Server Virtualization Made Real (Part 1 of a Multi-Part Series on Server Virtualization)* dated February 27, 2007, and available at http://www.clipper.com/research/TCG2007028.pdf and *Virtual Machines - Three Things to Consider + Three Ways to Use (Part 2 of a Multi-Part Series on Server Virtualization)* dated February 28, 2007, and available at http://www.clipper.com/research/TCG2007029.pdf.

[4] Cost was positioned last in this scenario because reliability and performance issues have greater importance in a fault tolerant environment.

To be considered a true mission-critical server, it needs to be F/T, to provide at least *five nines* of availability, equal to only 5 minutes of downtime per year. In the past, achieving 99.999% reliability required an expensive implementation of proprietary hardware and software. Today, open systems enable relatively low-cost commodity hardware and software combined with innovative design to deliver five nines in an affordable package with no single-point-of-failure.

Performance

A wealth of application software is available to run on mission-critical application servers based on an x86 processor, rather than a proprietary chip. There is also substantial comparative performance data from independent organizations, such as the *TPPC*[5] and *SPEC*[6] for those operating environments. We know how the microprocessor will perform. We do not know, however, how the F/T infrastructure will affect that performance.

In order to ensure continuous availability, servers clustered together to provide failover support must communicate. This can place an undue burden on the Ethernet LAN connecting the platforms, creating a performance bottleneck. The "master" server can create additional overhead as it issues messages to ensure synchronization with its "slave" or failover platform. If the two systems connect in an active-active mode, a failure will place a heavier load on the surviving system causing a longer delay in response time. A failure will also initiate the execution of a failover script, taking the functional server out of commission until the script has run its course.

Maintainability

If the first notification that you receive that a component has failed is a system crash, it is already too late; the serviceability of that platform is not F/T. A fault-tolerant system must have self-diagnosing software that alerts the data center or vendor field personnel, locally or remotely, of a *pending* failure. If the customer service personnel can replace a failing component *in an active platform* before it causes a system crash, the system maintains continuous availability. This also applies to FRU upgrades. You must be able to do field retrofits or system upgrades while the system remains active.

Cost

There is more to the cost of a fault-tolerant system than simply the acquisition price of the hardware platform. First, typically, the data center is acquiring two systems, a *primary* and a *failover*, so the CFO has to allow for double the expenditure for the hardware. Next, you must acquire a redundant LAN interface so that the systems can communicate. Third, there are software costs – multiple copies of the operating system and application, as well as custom scripts to enable the backup system to initiate recovery of the users in the event of a failure. This custom software can often be the most expensive component of the F/T configuration, as it requires the attention of an increasingly expensive resource – IT staff with the programming and administrative talents to keep the data center operational. To be complete, you should also factor in the potential cost of any planned or unplanned downtime that puts your mission-critical applications at risk.

The NEC Express5800/320Fc Solution

NEC has recently announced a fourth generation to their *Express5800*[7] line of fault tolerant servers, based on Intel's Xeon processor. The *320Fc* scores well in each category, qualifying as a fault-tolerant system. Using the open systems *Xeon 5300 series* (*Clovertown*) architecture, the Windows operating system, and a team of innovative engineers, NEC designed the *Express5800/320Fc* as a fault-tolerant system with 5 "9"s reliability that excels in all four categories. The 320Fc is the first implementation of fault tolerance to use NEC's *Lockstep* technology in a quad-core environment to provide complete hardware redundancy.

Reliability

NEC designed the Express5800 platform

[5] Transaction Processing Performance Council

[6] Standard Performance Evaluation Corporation

[7] See **The Clipper Group Navigator** dated October 22, 2003, entitled *HA Servers from NEC – Under the Radar, Above the Bar*, available at http://www.clipper.com/research/TCG2003056.pdf.

to perform logic self-checking and to replicate physically all of the major subsystems: CPU, PCI bus, memory, hard drives, and power supply. Fault tolerance is transparent to the application set with all self-checking logic resident on the main circuit boards to detect and isolate any failure immediately. By replicating all of the major components, the Express5800 enables the application to continue with no performance degradation, even if a component fails.

Performance

Starting with its third generation platform, the *Express5800/320Ma* based upon Intel's dual-core Xeon processor (*Paxville*), NEC has enhanced all of the major sub-systems in order to scale to the capability of quad-core. It has expanded memory capacity and an enhanced memory bus with 24 GB of DDR2 memory running at 667 MHz. This improves capacity 50% and throughput almost 70%. Next, the 320Fc has 8MB of cache, 2MB per core, to take advantage of the quad-core design. It has a 1333MHz front side bus, 66% faster than the 320Ma. In addition, the 320Fc has a pair of PCI Express slots, to take advantage of improved performance and open systems compatibility, and an option to use SAS[8] disk devices at up to 15K RPM for additional performance and reliability.

The fault-tolerant design of the Express5800 eliminates the development of many of the bottlenecks that wreak havoc with the performance of other F/T systems. In fact, NEC maintains that the 320Fc delivers more than 200% added performance over current dual-core models[9].

Maintainability

The redundant design of the Express5800 enables the data center staff to perform basic maintenance functions; they do not have to wait for vendor service engineers to arrive to restore operations. All Express5800 servers have circuitry to monitor their own operation. The IT staff is notified of any pending or actual failures, enabling them to install a replacement part while the system is still active. Similarly, they can install any part scheduled for upgrade on-line.

Cost

While there is more to the cost of a fault-tolerant system than simply the acquisition price, innovations NEC implemented in the Express5800 eliminate the need for an expensive and performance impeding communications connection between two independent servers. **The Express5800 is managed as a single logical entity**. In addition, because the 320Fc is based upon the low-cost x86 architecture, this F/T environment costs far less than proprietary solutions.

Software costs normally affect the total cost of ownership of an open IT solution far more than the hardware. Because the Express5800 is a single entity, the enterprise only needs to acquire a single copy of the operating system and a single copy of the application, reducing licensing costs significantly. With Windows Server 2003 as the operating environment, the Express5800 does not require expensive customized applications or equally expensive modifications to existing software.

Conclusion

Express5800 fault tolerant servers provide the only platform for Windows Server 2003 applications designed for continuous availability. Along with outstanding performance, the Express5800/320Fc delivers the availability and reliability that your mission-critical applications demand in a cost-effective package that minimizes TCO. If your enterprise is looking for a fault-tolerant system, you will find that the NEC-Express5800 is fast to implement, easy to maintain, and fits within your budget.

[8] See the March 1, 2007, issue of Clipper Notes entitled *Breaking the I/O Paradigm – SAS Enters the Nearline Storage Race*, which is available at http://www.clipper.com/research/TCG2007033.pdf.

[9] Unfortunately, at this time NEC has not released any commodity benchmarking data to substantiate this claim. Any prospective client should contact NEC directly for data concerning the benchmarking of your application.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

David Reine **is Director, Enterprise Systems for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. He joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group Captain's Log**, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.