IDC
Analyze the Future

WHITE PAPER

Mission-Critical Workloads for Windows Servers: How NEC Supports IT Best Practices

Sponsored by: NEC Corporation of America

Jean S. Bozman Al Gillen

August 2009

IDC OPINION

Within the typical enterprise-class datacenter, delivering mission-critical workloads with predictable levels of scale and reliability and supporting high levels of application performance that meet service level agreements (SLAs) are all in a day's work for many IT professionals. In the past, from a server platform perspective, the most typical and safest way to meet these requirements was to buy a mainframe or scalable Unix-based system that could deliver these attributes.

As the volume of industry-standard x86 server platforms has increased dramatically in recent years, accounting for 95% of all server unit shipments and for more than 50% of factory revenue worldwide, many organizations have been deploying one or more of their mission-critical applications on x86 servers. However, mission-critical workloads have had to be deployed differently on x86 servers because many models lack the scalability and datacenter-class resiliency that are built into other types of datacenter systems. In some cases, IT managers have had to take extraordinary measures to ensure that their organizations have enough capacity and availability to run those applications smoothly. Importantly, there has been another change: More types of applications, including email Web hosting and workgroup management, are now considered to be mission critical. Why? Simply stated, if those workloads stop running, business would come to a halt.

Today's business environment, with its expectations of always-on availability, demands 24 x 7 x 365 access, and corporations have upped the ante in terms of the acceptable levels of overall performance that are required for a business' survival in this economic climate. IDC notes that these changing requirements have created stress for IT organizations that need to meet more demanding business requirements. The current datacenter environment includes the following factors:

- Constant downward price pressure for capital expenditures (capex) in the datacenter has caused many customers to gravitate toward x86 platforms for their IT solutions. This, in turn, has raised the level of pressure on x86 server vendors, independent software vendors (ISVs), and their respective IT development staffs to find ways to optimize performance — and to maximize business benefit — from these systems.
- Customers deploying x86 server solutions are increasingly finding out what customers with scalable Unix server systems (midrange and high-end systems) have known for decades: A well-established set of best practices, aimed at improving consistency and predictability in business results, can dramatically improve the availability and performance of the IT systems and the ability to meet the service-level requirements of the business.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

- It used to be possible to segment mission-critical workloads from less critical workloads, but in recent years, the lines have blurred considerably between these workload categories, given that business often depends on all applications working well and being available to end users. The need to improve availability has increased as consolidation projects, enabled by virtualization, have enabled these different applications to migrate from previously isolated systems to a smaller number of shared x86 servers.
- Server vendors and their software partners are in a strong position to help IT customers improve their best practice adoption and, in the process, improve the overall business value that these customers gain from their IT investments.
- High-end x86 server solutions, used with high-end Microsoft products supporting scalable workloads and high availability, can enable IT organizations to manage and deliver mission-critical services to internal and external end users.

SITUATION OVERVIEW

The first server versions of the Windows family emerged in the early 1990s, but Microsoft Windows has been considered to be capable of solving true enterprise-class challenges, such as those that traditionally ran on mainframes and Unix servers in the datacenter, only in recent years.

When Microsoft replaced Windows NT 4.0 with Windows 2000, the company changed its development process, engaged with end users to help set requirements and quality standards, and bought into the concept of applying best practices to Windows deployments. Best practices are those approaches to computing that produce desired results consistently and are based on proven deployments in actual datacenter production conditions. Microsoft applied the best practices concept to develop and utilize its internal Common Engineering Criteria, which all products must meet prior to being certified for shipment.

In parallel, separate development teams within Microsoft were working to grow the scalability and reliability of layered server products, including the Microsoft SQL Server database and the Microsoft Exchange collaborative email software platform. Complementing these efforts was the work being done by Microsoft's server hardware partners. The server vendors worked closely with processor vendors, such as AMD and Intel, as well as with Microsoft to provide the hardware resources needed to support the growing needs of today's x86 application portfolio.

It has long been the case that customers have deployed Microsoft SQL Server database, Microsoft Exchange, and many brands of line-of-business (LOB) applications on Windows servers, stretching back to when Windows NT 4.0 shipped in the summer of 1996. IDC notes that these important workloads have matured in terms of their features and capabilities over time, with improvements in their scalability, reliability, and availability. Although Microsoft SQL Server is widely shipped on x86 servers, the dedicated servers on which many SQL Server instances are running have limited "headroom" for rapidly growing database images. The ability to consolidate these database workloads onto more scalable x86 servers, combined with the software-based scalability improvements in SQL Server 2008, will result in fewer SQL Server "footprints" in the datacenter — and improved management by central site IT staff.

Finally, it is important to note that the trend toward virtualization/consolidation means that more workloads are running per physical server than ever before and that this pattern of consolidation demands more attention to the resiliency and scalability of the hardware than earlier generations of x86 server configurations could offer. Because a typical consolidation server may host an average of eight or more virtual machines, any outage affecting the physical server would be disruptive to the business, potentially affecting hundreds of end users and taking many applications offline.

DEPLOYING MISSION-CRITICAL WORKLOADS ON X86 SERVER INFRASTRUCTURES

Given the need for business continuity when deploying mission-critical Windows or virtualization workloads, IT managers should take steps to ensure that the applications, databases, and data are as highly available as possible — even in the event of unplanned downtime due to network outages, power outages, and natural disasters. These considerations include the use of scalable and resilient hardware platforms that have reliability, availability, and serviceability (RAS) features that IT managers have come to expect for mission-critical workloads. These features include memory mirroring, hot-swappable components, and CPU (processor) partitioning, as well as high ratings for mean time between failures (MTBF).

Customer Challenges with Deploying Mission-Critical Environments in Scale-Out x86 Infrastructures

As the total number of end users accessing the workloads rises, over time, typical customer concerns include, but are not limited to, the following set of challenges:

- Provisioning enough memory and storage in the server system, which if not addressed, compromises system performance along with the ability to run production workloads and to ensure efficient backup/restoration operations
- Taking steps to ensure high availability of the overall solution in the event of hardware or software component failure or due to external events (System downtime impacts user productivity and business continuity of processes that support revenue generation and customer interactions.)
- Making sure that rapid growth in the end-user population, with many more users logging on to the same systems, will not jeopardize reasonable response times for applications running on the server

For IT managers who are supporting large numbers of small, dedicated servers, consolidating some of those workloads onto a smaller number of more scalable server footprints will likely lead to more efficient management, better server utilization, and lower operational costs. It is likely that virtualized consolidation will also improve uptime and reduce the datacenter energy requirements of those workloads.

But the process of consolidation alone does not ensure that mission-critical workloads will receive all of the underlying hardware support they need to keep running without interruptions in service to end users. The high levels of demand for business continuity and availability of mission-critical data services are causing datacenters to rethink the ways in which RAS can be provided.

Forward-looking IT professionals already know that best practices developed in datacenter environments also can work well in the x86 server world. Reliability, availability, and serviceability for computer systems are as relevant to x86 server customers who are running mission-critical workloads as they were to customers running traditional datacenter systems.

Mission-Critical Windows Server Platforms

Evolution of the Windows Platform

Microsoft continued to pour investments into the Windows Server product, starting with the first version of its enterprise-focused product, Windows NT 4.0 Enterprise Edition, in September 1997. Since that time, the company has come to market with numerous high-end releases, culminating with the current portfolio, which includes Windows Server 2008 Enterprise and Windows Server 2008 Datacenter. These two products represent the high-end versions of Microsoft's portfolio. For those considering a Windows Server operating environment, Windows Server 2008 Enterprise or Windows Server 2008 Datacenter Edition would be a first choice Windows solution for customers supporting critical workloads.

As the Windows Server family continued to become more scalable and highly available than ever before, Microsoft began talking about taking on workloads running on high-end datacenter systems, including mainframes and Unix servers. Historically, these types of hardware systems have been priced higher than the scalable systems that were built on the x86 platform. In fact, the entire technology stack comprising hardware and system software on Unix servers is more costly than the stack for x86 platforms, although IDC notes that hardware pricing and software pricing have been eroding due to competitive pressure in recent years.

Running Mission-Critical Workloads on Windows

Windows Server 2008 Enterprise and Windows Server 2008 Datacenter can be leveraged as the base platform for a Microsoft mission-critical solution. Important features provided by these two Windows 2008 products include the following:

- **More memory support.** Microsoft's 32-bit versions of Windows Server 2008 Enterprise and Windows Server 2008 Datacenter now offer memory support for up to 64GB of RAM, while 64-bit versions of these Windows products can address up to 2TB of RAM. This capability allows alternate memory to be tapped for scalable application workloads and can support data mirroring for high availability. Other versions of Windows Server 2008 support up to 4GB of non-fault-tolerant RAM when deployed as a 32-bit operating system and up to 32GB of non-fault-tolerant RAM when deployed as a 64-bit operating system.

- **Hot add memory.** This feature is supported by the Windows Server 2008 Enterprise and Datacenter editions. It allows memory to be added to the system without requiring a reboot of the system and thus avoids interruption to business processes.

- **Hot replace memory.** Hot add processors and hot replace processors are supported by the Windows Server 2008 Datacenter product. This capability supports expansion of memory, which may be needed as database footprints expand. Adding memory, via hot replace, helps customers to expand their inventory of servers — speeding time to market for customers that need to quickly add memory and processors without impacting system availability.

- **Unlimited connection potential.** Windows Server 2008 Datacenter and Windows Server 2008 Enterprise offer unlimited connections for Terminal Services clients and for network access connections. Other editions of Windows are limited in connection volume.

- **More processor support.** Windows Server 2008 Enterprise supports up to eight x86 sockets or up to eight processor sockets. Each socket supports multicore processors, such as 6-core Intel Xeon processors discussed later in this paper. Windows Server 2008 Datacenter supports up to 32 x86 sockets (32 cores) or up to 64 x64 sockets (64 cores). These limits will be expanded to 256 cores in Windows Server 2008 R2 Datacenter. This additional scale-up capability makes this operating system more capable of handling modern workloads and consolidation activities that are becoming increasingly common today.

- **Virtualization support.** While Windows Server 2008 Standard includes a Server Core Virtualization Role, the product does not include replica copy deployment rights (other than the standard 1+1 deployment scenario, where a single replica copy guest managed by one instance of the Windows host can be deployed under a single license). By comparison, Windows Server 2008 Enterprise gives customers integral rights to create up to four replica copies of Windows; Windows Server 2008 Datacenter offers unlimited replica copy use rights. The integration of virtualization support into Windows Server 2008 also enables this platform to become a consolidation platform for multiple existing instances of Windows Server operating environments aboard a single physical server.

- **Failover clustering.** Support for up to 16 server nodes within a single cluster, through Microsoft Failover Cluster software, which is shipped with the Enterprise and Datacenter Editions of Windows Server 2008. Windows Server 2008 doubled the number of servers supported within a cluster, compared with earlier Windows versions, making today's Windows clustering more comparable with Unix clustering in the datacenter, given the maximum node count.

Best Practices for Mission-Critical Solutions on x86 Platforms

It is clear that Microsoft is investing in Windows Server 2008, but for customers, an operating system is only a single component of a total server-based solution. Other components include hardware and services in the form of the IT professional support

that helps customers to design, configure, and manage the day-to-day operations and life cycle of the entire solution.

IDC sees a number of prescriptive activities that IT professionals should use in managing a Windows Server 2008–based solution that will support mission-critical workloads. Such activities include all of the following:

- **Define expectations about mission-critical operations.** It is important to establish the criteria that define mission-critical operations and to establish measurement criteria that an organization will use to characterize the behavior of a mission-critical application. These criteria should include service-level definitions (response time, transactional rates, number of simultaneous supported users with a given response time, etc.), availability (maximum acceptable number of minutes or hours of unscheduled and scheduled downtime per year), and other similar metrics. It is important to note that not every application should be measured with the same criteria. In fact, it is likely that a variety of target performance levels should be established for different types of mission-critical applications. Furthermore, with an increasing number of end-to-end applications tapping several computing tiers (Web, application, database tiers) to complete their work, customers must study and analyze the resiliency and requirements of each tier to ensure that there is no single point of failure in the end-to-end solution.

- **Map out an uptime strategy for Windows.** Organizations need to have a discussion of best practices associated with maximizing uptime and system reliability on a Windows Server environment as well as improving availability for key applications (such as Microsoft Exchange mail serving and the Microsoft SQL Server database) and associated functions (such as the recovery of files). Following this discussion, an organization may conclude that it needs to select an OEM hardware partner with experience in mission-critical deployments or to expand its training and on-staff expertise.

- **Establish a process for consistent and efficient system design.** Merely looking for tactical bargains on hardware and software will lead to an inconsistent hardware/software environment, without the needed levels of interoperability and availability. This approach is one of the first violations of best practices. Consistency of implementation of system configurations across the infrastructure allows for consistent management and deeper expertise on a smaller number of configurations. In many cases, customers will select one vendor for volume x86 servers and another vendor for scalable, high-end x86 systems.

- **Limit the number of operating system images in use.** IDC research has found that total cost of ownership (TCO) can be lowered and the overall uptime of the systems can be improved by reducing the number of operating system images in use. In research conducted by IDC in 2005, we found that IT labor costs associated with supporting Windows Server 2003 operating environments could be reduced by up to 37% through the use of a standardized operating system image. We believe that the same trend would hold true for Windows Server 2008 today.

- **Establish strategies for virtualization, high availability, and disaster recovery.** These considerations are neither secondary decisions nor discrete decisions. An outage within a computing tier can jeopardize business processing across the entire IT infrastructure. Today, it makes great sense to combine all three considerations in the same discussion. The use of virtualization technology on x86 servers may offer suitable high availability for some types of applications (e.g., live migration for planned downtime), while disaster recovery (DR) solutions restart virtual machines (VMs) that have already been moved to alternate server systems. However, many mission-critical workloads require the use of high-availability software, such as clustering/failover software, to ensure a rapid application restart, even for applications running inside VMs. Finally, it is now possible to codeploy high-availability software and virtualization software so that applications can fail over from physical servers to virtual servers (P to V) or from virtual servers to virtual servers (V to V), thus avoiding the cost of maintaining "standby" hardware servers just for the purpose of failover scenarios.

- **Take an inventory of all workloads.** Determine what types of protection will be needed to ensure high availability. Different types of workloads may require more high-availability protection than others. Some small amounts of downtime would be tolerable for some applications while unacceptable for others. This is particularly true where access to applications by end users or by customers is involved.

- **Establish management best practices.** Management software is a critical component to apply in conjunction with best practices. An organization that does not currently have a comprehensive management solution in place is asking for a haphazard approach to change and configuration management and, in the process, loses the opportunity to increase its knowledge base around successful management practices versus unsuccessful management practices.

- **Carefully select system components.** IT professionals should carefully consider their deployment needs, the expected life cycle of the solution in terms of overall years of expected service, and the expected life cycle (period of usefulness) of individual components. The questions to be asked are as follows: Will a single hardware solution and a single operating system solution be maintained and upgraded through the entire life cycle? Or will one or more of those components be retired and replaced while the remainder of the solution survives the full life cycle? Understanding product road maps, upgrade programs, and associated costs will be critical to the business results.

- **Consider the full range of layered solutions for the platform.** Ultimately, any mission-critical solution will include a multitude of hardware and software products, but the face of that solution is usually an application or an application service that end users and/or customers and partners will touch and experience. Therefore, part of the planning process would need to include a review of layered products (e.g., Microsoft SQL Server, Microsoft Exchange, Oracle Database 11g, and ERP or CRM applications from SAP and Oracle and other providers).

- **Upgrade, secure, or supplement appropriate IT skill sets.** Understanding the technology platform, and how to deploy it, is critical. Having the appropriate IT skills ensures a savings in setup time and speeds deployment. In today's economic environment, many organizations have constrained budgets, and many projects are competing for additional investment. However, existing IT skill sets can be supplemented by technology. For example, in the area of clustered server deployments, IDC notes that Microsoft has addressed many of these skill set issues. Introducing automation into the Failover Cluster software in Windows Server 2008 reduces the need for custom scripting and provides autodiscovery of servers, storage, and networking software.

- **Understand business benefits.** Having IT staff who are familiar with Windows and Windows systems management means they can leverage previous training on Windows products when they use a scalable Windows server (reducing training costs). Consolidation of workloads may mean fewer individual server "footprints" in the datacenter, reducing overall maintenance costs associated with large numbers of small servers.

- **Make sure the datacenter is ready for consolidation requirements.** Consolidation projects require systems that are capable of supporting the consolidated workloads. If the appropriate skill sets are not available within the organization, consider working with third-party channel partners or IT consultants during the consolidation process itself, prior to deployment on the new systems.

- **Develop a long-range plan for datacenter capacity.** Make sure that the site has enough capacity today and will have a way to expand capacity tomorrow. Capacity planning, long a necessary element of mainframe computing, is now applicable to x86 systems in the datacenter. This style of planning will protect uptime and ensure smooth operations as demands for processing expand, and change, over time.

NEC'S SCALABLE X86 SERVER SOLUTION

NEC introduced its NEC Express5800/A1160 enterprise server (which is also known as the NEC Monster Xeon Server) in the United States during the fall of 2008. The NEC Express5800/A1160 is scalable up to four server "nodes," each of which is based on a 4-socket module. All nodes can be managed as one single-system image, allowing applications to scale up seamlessly when more capacity is added. In turn, each socket houses a 4-core or 6-core Intel Xeon 7400 Dunnington processor (and the configurations may include either type of processor). Intel Xeon processors, available in the 2.4GHz to 2.6GHz range, are designed to be energy efficient while still providing substantial computing power for physical and virtual workloads. Similar to mainframe environments, NEC's scalable server can also be deployed in a partitioned configuration enabling support of multiple OS or database instances when necessary.

In this way, NEC supports a pay-as-you-grow path to gaining more scalability for Windows and Linux applications. NEC Express5800/A1160 enterprise server installations can start with a 4-socket/24-core model. They can then be expanded,

over time, resulting in 8-, 12-, or 16-socket server models. Each server node, or building block, supports up to 32 DIMM slots and therefore up to 256GB of memory utilizing 8GB DIMMs. A 4-node NEC Express5800/A1160 system has a total of up to 96 processor cores and up to 128 DIMM slots (supporting up to 1TB of memory). It provides rapid expandability, as needed, to support growing workloads — virtualization workloads in particular — as well as increasingly large groups of end users. Additionally, each server node contains two memory array structures (16 DIMM slots each), enabling higher system resiliency through a memory mirroring option if desired.

Further, the ability to add I/O bandwidth, through the use of hot-swappable PCIe slots, gives IT managers added flexibility to rapidly and easily expand system network capacity. Overall, the NEC Express5800/A1160 enterprise server is designed to enable a dynamic IT infrastructure solution that is capable of supporting changing and demanding business requirements.

For the purpose of replacing components (processors, memory DIMMs, I/O, SAS drives, and internal RAID storage), NEC Express5800/A1160 systems are accessible from the front or rear of the enclosure. The ability for IT staffers to reach system components easily reduces maintenance time and improves productivity.

As mentioned earlier, one of the key design points for the NEC server is providing support for virtualization software and for virtualization-enabled workload consolidation. NEC provides direct support for a variety of hypervisor platforms, including those from VMware (ESX Server, Virtual Infrastructure 3, and vSphere products), from Microsoft (Hyper-V, as shipping now and in its next release), and from Citrix (with its Xen-enabled solutions, including support for virtualized x86 desktops).

As support for more virtual machines increases over time, the RAS features of the NEC Express5800/A1160 scalable high-availability symmetric multiprocessing (SMP) server help to protect solution availability for purposes of business continuity. In combination with NEC's high-availability software, NEC ExpressCluster, or Microsoft's Failover Cluster software, NEC's hardware platform provides a holistic solution that addresses the types of concerns about business continuity that typically accompany high-density virtualization planning and deployment. This capability becomes increasingly important as the average virtual machine count on physical servers rises, because of workload consolidation, and as the number of enterprise applications in virtualized computing environments increases.

CHALLENGES AND OPPORTUNITIES

The worldwide IT market is a competitive one, with many companies producing products and services that supply customer demand for servers, storage, and software. The recent economic downturn has caused IT organizations to think carefully about IT spending — and how to allocate the funds they have. Companies that address this pressing reality and offer business value for the capex spending will be in the best position to gain new business or repeat business. The following sections highlight the top challenges and opportunities facing IT platform providers and customers.

Challenges in the Marketplace

- **Lack of IT budget.** One of the first business spending areas negatively impacted by the economic downturn was IT budgets in North America. For many organizations, the thought of devoting precious IT dollars to high-end hardware and software seems like a nonstarter. This is where IT professionals will want to seek executive sponsorship and to develop a business plan to justify investment in systems that, in turn, will better prepare the business to compete as the company emerges from the economic downturn. The operational benefits of workload consolidation, and improved management, can play an important role in reducing ongoing IT staff costs.
- **Taking a broad view of business value for the enterprise.** Often, taking a broad view of the benefits that a given IT investment will bring to the wider organization will shed light on its business value. Improved availability of IT systems means improved user productivity, and higher uptime means fewer interruptions to customers or to business partners in the supply chain. Better uptime on a customer's key systems may result in competitive advantage versus that customer's business competitors.

For NEC, these challenges will mean competition with other providers of x86 scalable systems. However, the combination of product differentiation, system engineering and packaging, and system services/support will be an important consideration for IT customers. NEC brings a long history of supporting and servicing mission-critical workload environments on both x86 servers and non-x86 servers — experience that can be leveraged to meet customers' business requirements.

Opportunities

- **Leveraging successes from other markets.** NEC has the opportunity to demonstrate its product capabilities in the scalable x86 server space and to highlight the way in which its systems align with customers' processes, providing a platform for the next-generation dynamic datacenter. The system's performance capabilities, its support for advanced virtualization, and price/performance, all of which have already been demonstrated in Asia/Pacific and Japan, can be brought to the Americas and to EMEA to make the benefits of this technology more widely known to a broader set of customers. For more details regarding the technical capabilities and performance of the NEC scalable x86 servers, refer to **www.necam.com/servers/enterprise/A1160.cfm**.
- **Savings coming from IT investment.** IDC research has found that many IT investments, if selected carefully and well planned, can lead to a return on investment (ROI) and result in lower TCO through reduced operational costs (e.g., cost of maintenance, reduction in IT staff time, reduced power consumption through workload consolidation, improved user productivity in the organization). For customers, the key to analyzing these metrics is not only to understand how the investment will lower hardware and software costs, but also to recognize that IT labor costs typically are the single biggest cost area for any IT organization. So when planning an IT investment, customers must be sure to focus their analysis on how the investment will lower management and IT labor costs.

For its part, NEC can highlight the scalability and RAS features of the NEC Express5800/A1160 system that reduce TCO and improve ROI as it markets the scalable server system worldwide. Total capacity for the system far exceeds the capacity of 2-socket and 4-socket x86 servers, due to its design (with up to 16 sockets and up to 96 cores on board). This means that the NEC system allows customers to add capacity, as needed, in 4-socket increments to meet growing computing needs, as user demand grows, over time. Importantly, the RAS features of the system support scalability for applications and databases running on the machine. Examples of use cases include supporting workload consolidation, supporting growing databases, supporting larger numbers of users on the same system, supporting more data-intensive workloads, such as business intelligence (BI) — and supporting highly virtualized workloads, with many virtual machines per physical server. By providing the ability to add capacity, on a modular basis, if needed, up to 96 cores (in a total of 16 sockets), NEC is well positioned to compete in the scalable x86 server space.

CONCLUSION

Delivering mission-critical workloads with predictable levels of scale and reliability and providing application performance that meets service-level agreement (SLA) commitments are all in a day's work for many IT professionals, as stated at the beginning of this paper. In recent years, the increasingly heavy penetration of x86 servers in enterprises has led to a widespread need to support mission-critical workloads — whether it be operating systems such as Microsoft Windows or Linux, workloads such as Microsoft SQL Server databases, virtualization software solutions, or consolidation of workloads on virtualized hardware — running aboard scalable x86 servers.

Today's 24 x 7 x 365 world has raised the bar for the continuous availability of systems, elevating the status of many business applications to mission critical. If those workloads go offline, the downtime results in lost revenue, and business processes come to a halt.

Customers deploying x86 server solutions are increasingly finding out what large system customers have known for decades: A well-established set of best practices can dramatically improve availability, performance, and the ability to meet service-level requirements. Best practices, leveraging real-world experiences from IT professionals, have the potential to dramatically increase the performance levels of scalable x86 servers. Importantly, these best practices will aid in the process of datacenter transformation, as an increasing number of scalable x86 servers support next-generation mission-critical applications and databases.