IDC
Analyze the Future

WHITE PAPER

Scalable Windows Servers in the Datacenter

Sponsored by: Hewlett-Packard

Jean S. Bozman Matthew Eastwood

January 2009

EXECUTIVE SUMMARY

Enterprise workloads running under Microsoft Windows have been playing more important roles in recent years — running more demanding applications and mission-critical databases. With the shipment of the Microsoft Windows Server 2008 product in early 2008, and the shipment of Microsoft SQL Server 2008 in mid-2008, customers have had an opportunity to evaluate how they can leverage these versions of software within large enterprise networks consisting of hundreds, or even thousands, of small Windows servers.

IDC uses the term "scalable Windows" to refer to the ability of users to scale up their Windows computing environment to run on larger servers that support high availability for applications and data. This provides a greater ability to handle larger and multiple workloads via the provisioning of more server capacity and system resources. Generally, these types of workloads are considered to be mission-critical (e.g., aspects of the business would stop if the computing stopped) or business-critical (e.g., highly important to continuing business operations). As the workloads scale up, they require more protection against downtime and the presence of hardware features (e.g., reliability, availability, and serviceability [RAS] features) and software features (e.g., high-availability software) that can ensure near-continuous business operations.

IDC notes that workload consolidation projects at customer sites are providing the impetus to evaluate the current inventory of business-critical and mission-critical workloads running under Windows and to consider how to deploy scalable Windows workloads that can take advantage of scalable server platforms, such as line-of-business (LOB) applications and scalable databases.

This paper focuses on HP Integrity servers running the Microsoft Windows server operating environment and mission-critical workloads (e.g., ERP, CRM, and databases). Often the result of consolidation initiatives within the datacenter, the deployment of HP Integrity servers running Microsoft Windows results in a computing platform that supports mission-critical Windows workloads for the enterprise. Consolidation projects, which gather workloads from many servers, can improve overall efficiency within the datacenter, reducing both IT staffing costs and downtime. As those Windows workloads scale up, reaching broader communities of end users and end customers, continuing access via improved availability of applications and data will become more important to both IT managers and business managers in enterprises worldwide.

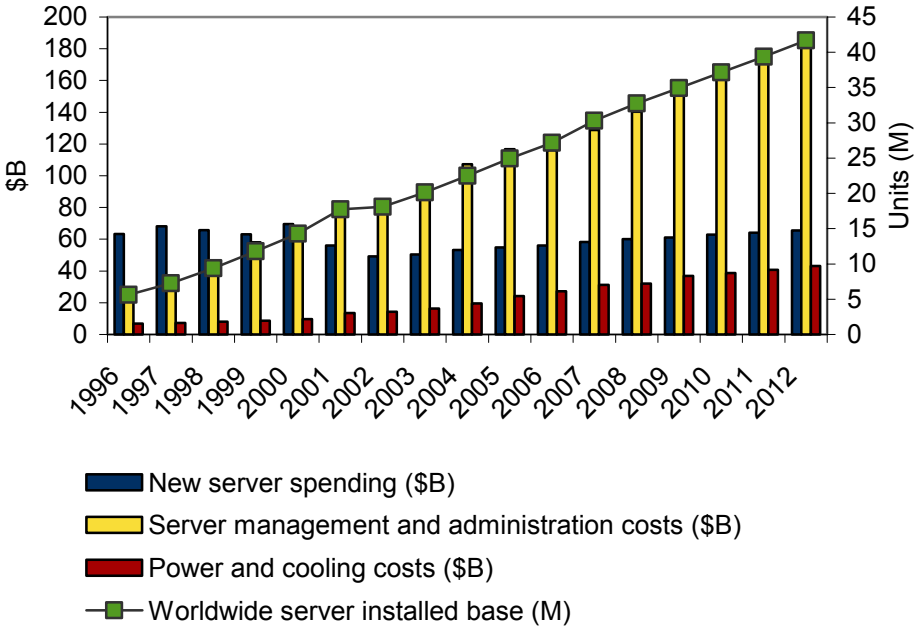
Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

TRENDS IN THE DATACENTER

IT organizations, in concert with business units, are working to reduce operational costs, driven by the cost of managing large numbers of small servers and the need to pay to power and cool them and to pay IT staff to apply updates and changes to them. As shown in Figure 1, these operational costs have increased dramatically over the past 10 years due to the rapid growth of small form-factor servers, primarily rack-optimized servers, in the datacenter.

FIGURE 1

Worldwide IT Spending on Servers, Power and Cooling, and Management/Administration, 1996–2012

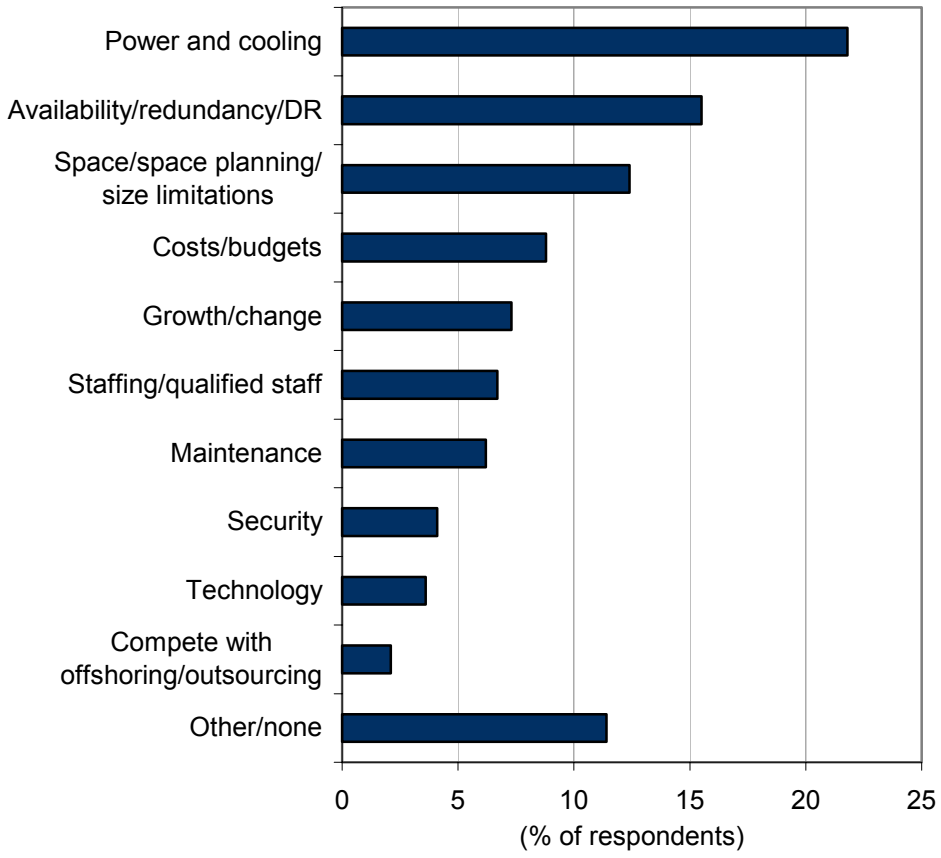


Source: IDC, 2008

As Figure 2 illustrates, top-of-mind datacenter challenges center on power/cooling, availability, and space utilization. By evaluating the overall workloads under management, IT can identify applications, many of them business-critical or mission-critical, that would benefit most from consolidation onto fewer server "footprints" in the datacenter.

FIGURE 2

Top Challenges in the Datacenter



n = 197

Source: IDC, 2008

Business Goals of Deploying Scalable Servers

In the process of workload consolidation, many touchstones of IT operations would be addressed. The primary goals of deploying scalable servers in the IT infrastructure for Windows workloads would include the following:

- **Replace aging servers in the datacenter.** IDC demand-side research, based on customer surveys of more than 1,000 IT sites, clearly shows that many scalable SMP servers in the datacenter are approaching — or have exceeded — 7 to 10 years of use. Importantly, many of these servers have lower-performance, yet higher power/cooling requirements than today's generation of servers, which are based on multicore processors.

- **Provide more capacity for growing server workloads.** Scalable servers allow for scale-up applications to support more end users, up to thousands of end users. The ability to directly support hundreds or thousands of end-user sessions — or large volumes of transactions — speaks to the ability to add capacity (e.g., processors, memory, I/O) as needed. In addition, the ability to provide capacity, in a balanced way, speaks to the ability to add memory and I/O to keep pace with the increasing "compute density" of multicore processors.
- **Balance performance across the system.** When IT scales up workloads on servers with four or more sockets, it needs to consider the balance between processing power, memory, and server I/O, which provides access to storage and the network. Throughput (or the total amount of processing work done) improves when processor-intensive workloads no longer battle with I/O-intensive workloads for system resources.
- **Consolidate the Windows workloads.** Scalable server architectures enable consolidation onto new servers so that multiple workloads can be combined to run within a single, scalable instance of the operating system, if needed (e.g., for a single, larger database instance). Alternatively, consolidation can mean bringing many workloads onto a single, scalable server. Scalable servers offer a high degree of control and security, which is essential for mission-critical workloads, including large databases and growing applications, supporting large numbers of end users.
- **Gain more RAS characteristics for the same workload.** In a mission-critical computing environment, RAS is key. RAS features include error correcting code (ECC), hot-swappable components, and protection for data transfer points throughout the system.
- **Support larger datasets.** Scalable servers support scale-up databases that can house more corporate data in data warehouses or data marts — which can support more queries from end users and end customers than was possible when the data was housed in multiple, smaller databases housed on multiple servers.
- **Obtain green IT benefits from running selected workloads more efficiently on fewer, more scalable servers.** Green IT initiatives cover a host of technologies — all of which act to reduce power/cooling costs and to improve the energy efficiency of servers, storage, and associated equipment within the datacenter. By consolidating workloads that otherwise would have run on a much larger number of servers and running them on fewer, more scalable server "footprints" within the datacenter, IT benefits from reduced power/cooling costs and more efficient management of workloads.

SCALABLE SERVER MARKET FOR MICROSOFT WINDOWS

Several broad categories of customers adopt scalable servers. These servers offer a high degree of control and security, which is essential for mission-critical workloads, including large databases and growing applications, that support large numbers of end users. Although both RISC servers and Itanium-based servers account for the majority of all scalable servers, RISC-based servers do not run the Microsoft

Windows operating system. Using scalable Windows servers, where applicable, is one way that datacenters can move to new equipment while leveraging IT staffers' familiarity with the Windows computing environment.

The following are examples of why customers deploy scalable Windows servers:

- **Scaling up from deployments on smaller servers, often including x86 servers.** Users with rapidly growing applications or databases often reach a point in the evolution of their overall computing environments when some, or many, of their servers supporting mission-critical and business-critical workloads are maxed out. For example, business intelligence (BI), database, or online transaction processing (OLTP) environments no longer perform well because of their growth and complexity.
- **Consolidation.** Server sprawl in the datacenter, as was referenced in Figure 1, is causing a sharp rise in operational costs (e.g., license costs, power and cooling costs, management costs, maintenance costs, IT staff costs, and the cost of datacenter "real estate"). This is diverting IT spending away from more productive and innovative tasks that could support business processes more effectively.
- **Modernization.** Customers are seriously investigating Windows servers as platforms that can take on workloads from aging mainframe and RISC server environments, many of them 7–10 years old. Some sites are offloading some of the mission-critical workloads running on older mainframe servers to scalable Windows servers, where it is possible to do so.
- **Building a BI resource within the enterprise.** An increasing number of organizations are making BI functionality in a Windows environment available to a broad community of stakeholders who need to analyze the enterprise's transactional data. This community of end users includes executives, LOB managers, customers, partners, suppliers, and customer-facing employees. IDC notes that BI workloads are growing due to the increasing amounts of data that systems must collect and analyze and also due to the increased numbers of users who need access to the BI data.

IDC Server Market Data

IDC supply-side data shows that Itanium-based servers saw rapid growth, generating more than $4.4 billion in factory revenue worldwide in 2007. Unit shipments of Itanium-based servers worldwide increased by 36% in 2007, compared with 2006, and factory revenue for these servers increased 31% over the same time period. From 2002 to 2007, Itanium-based server shipments increased at a compound annual growth rate (CAGR) of 30% and revenue increased at a 73% CAGR. IDC notes that Itanium-based systems grew almost three times as fast as x86 systems from 2002 to 2007, while RISC and CISC systems both showed some decline over the same time period.

Extension of Windows technology into the upper reaches of the midrange enterprise market (which IDC defines as servers priced from $25,000 to $499,999) and the high-end enterprise server market (which IDC defines as servers priced at $500,000 or more) is ongoing. By contrast, servers priced less than $25,000 are classified as volume servers. According to IDC's worldwide server market forecast, Itanium-based

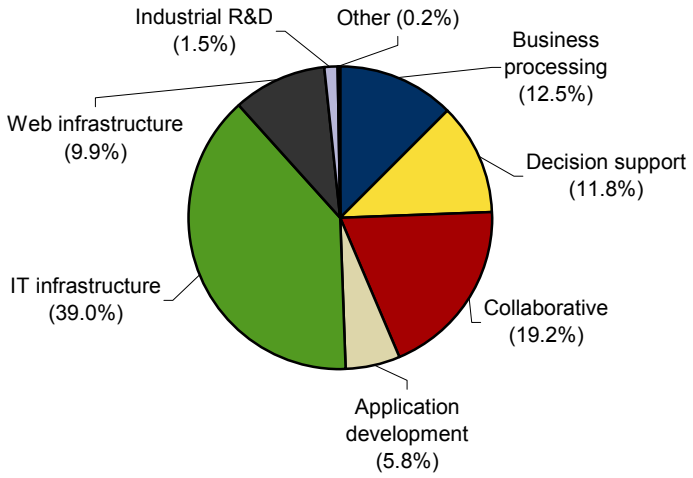
servers running Windows will realize a five-year CAGR of 24.9% from 2007 to 2012. The growth rate for Windows on Itanium-based server platforms worldwide is the fastest growth rate by operating system and is expected to outpace the growth rates for Unix and Linux on the same platforms.

Windows Workloads in the Business Processing Space

Windows workloads have been moving more heavily into the business processing space, as IDC's annual Workloads research shows. These types of workloads, including OLTP and LOB applications such as ERP and CRM, have greater requirements for reliability and availability than many other types of workloads. The combined business processing, decision support, and collaborative workloads amount to 43.5% of all Windows workloads (see Figure 3). These types of workloads often require scalability, as demanding applications require more capacity and as larger groups of end users begin to access the system. As such, the use of a platform that has RAS features built into the hardware — along with Windows support for large memory features and error handling — would be useful when supporting these mission-critical workloads.

FIGURE 3

Windows Workloads Worldwide Customer Revenue Share



Source: IDC, 2008

Workloads on x86 Server Platforms

IDC notes that Microsoft Windows on x86 servers (servers based on Intel Xeon or AMD Opteron processors) offers another type of platform for scaling up Windows workloads. This type of deployment occurs most often on midrange Windows servers. Due to their widespread use in datacenters, x86 servers are increasingly being used to run a range of enterprise workloads, even though they were not originally designed

with an emphasis on RAS features. Accordingly, customers must take steps to improve the availability and reliability of x86 servers through the addition of memory protection features, redundant hardware components, and the use of ECC and high-availability failover software.

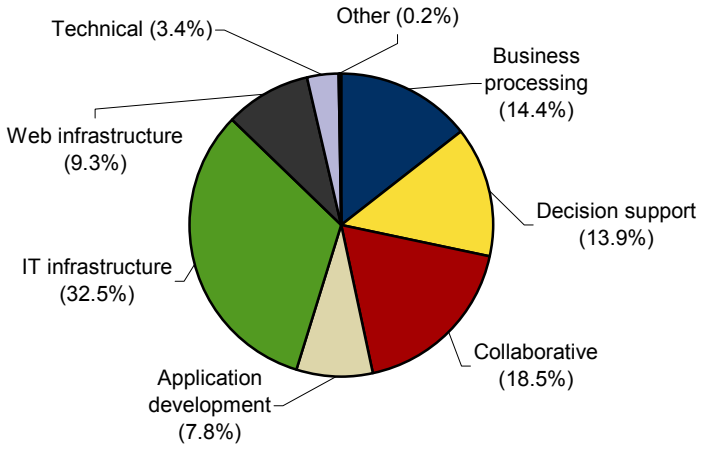
IDC notes that some x86 server models, especially midrange x86 servers made by HP and its competitors (e.g., IBM, Unisys), are gaining more RAS features over time and they are offering increasing support for more processing capability through the use of multicore x86 processors. In some cases, multiple server "nodes," each with 4–8 sockets, will be assembled in industry-standard racks, or clustered together, along with storage modules and cooling modules to support those nodes. However, IDC believes that ultimately, customer preferences for deployment types (e.g., scale-out versus scale-up) and workload requirements determine when, and whether, to deploy x86 servers to support these workloads.

Workloads on Itanium Server Platforms

Figure 4 shows the "mix" of workloads for Itanium-based servers running Windows. This mix heavily favors business processing, decision support, and IT infrastructure workloads (e.g., workloads pertaining to network-enabled data processing). A comparison of the pattern for overall Windows servers, including both x86 servers and Itanium-based servers, shows that the percentage of business processing and decision support workloads was greater on Itanium-based servers than on x86 servers. This IDC Workloads data is based on customer surveys conducted in more than 1,000 sites in North America in the 2006–2007 time period.

FIGURE 4

Share of Windows Workloads on Itanium Servers, 2007



Source: IDC, 2007

HP INTEGRITY SERVERS FOR SCALABLE WINDOWS

HP Integrity servers running Microsoft Windows Server 2008 provide a scalable platform for business-critical and mission-critical workloads that ranges from 2-socket blade and entry-level rack-optimized servers up to the 64-socket (128-core) HP Integrity Superdome. Based on Intel Itanium processors and HP systems design, the midrange and high-end Integrity systems are based on a series of hardware "cells" (with each cell housing 4 sockets) that can be added to expand capacity; smaller Integrity server models, with 1–4 sockets, are not cell-based.

HP Integrity servers have been designed with advanced RAS features, which many customers now require for their most demanding mission-critical databases and LOB applications. Because they are highly scalable systems, HP Integrity servers allow customers to add processors and memory to handle growing capacity requirements as needed. Further, they provide a path for customers who wish to retain Windows as they expand their systems' processing capacity while avoiding migration to another operating system.

RAS Features on a Range of Integrity Servers

RAS features are available throughout the HP Integrity product line, extending from the small form-factor servers to the very largest Integrity Superdome servers. All of these servers are based on dual-core Intel Itanium processors. IDC notes that HP also sells 4-socket and 8-socket systems based on x86 processors in the midrange server market, which can also support scalable Windows workloads. In the current version of Microsoft Windows Server 2008, up to 32 sockets (and 64 cores) can be supported in a single-system image (SSI). Later this year, HP plans to ship an HP Integrity Superdome model with 64 Itanium processors and 128 cores (2 cores per processor) that supports a total of 256 threads. This system will run Windows Server 2008 Release 2, which is currently in beta release and scheduled to ship in 2009.

Specifically, HP Integrity servers in the midrange server market include the HP rx7640 (2–8 sockets) and the rx8640 (2–16 sockets) models. High-end enterprise models include the HP Integrity Superdome models with 16, 32, or 64 sockets (and deployments may also include HP Integrity volume server models, such as the rx2660 (1–2 sockets), rx3600 (1–2 sockets), and rx6600 (1–4 sockets). In addition, there are blade form-factor Integrity server models: the BL860c (with 1–2 sockets) and the BL870c (with 1–4 sockets).

The HP Integrity servers support expandable memory up to 2TB and numerous I/O connections to storage — which is extremely important for I/O-intensive workloads accessing large databases, including data warehouse and business intelligence applications. Their ability to aggregate workloads from multiple, smaller servers — and to redeploy those workloads on midrange or high-end servers — makes them an efficient element of an organization and enterprise consolidation system.

Engineers from HP and Microsoft work to tune and test Microsoft SQL Server performance on scalable HP Integrity servers in an ongoing approach to improve performance for HP Integrity/Microsoft SQL Server deployments. HP Integrity server engineers, many of whom work at Microsoft's Redmond, Washington, campus, work directly with Microsoft engineers on benchmarks, reference platforms, integration testing, and performance tuning.

Customers often deploy x86 servers and Itanium-based servers in the same datacenter, depending on customer preference and workload requirements. For example, in a three-tier SAP solution, HP ProLiant servers could be used at the application tier and Integrity servers could be used at the back-end database. Both ProLiant and Integrity servers can be managed through a common management interface with HP Systems Insight Manager (SIM) and HP Integrity Essentials.

Virtualization and Partitioning on Integrity Servers

HP Integrity servers running Windows provide the flexibility of a broad range of choices for utilizing the resources of the system via hardware-defined and software-defined partitions, including the following:

- **HP Integrity nPartitions (nPars).** These are hardware-defined partitions, with each partition built upon 4-socket "cells" within the hardware design. Electrical isolation is supported within the midrange models (rx7640 and rx8640) and high-end Superdome models. This electrical isolation within the nPartitions protects important workloads — and prevents workloads in any given partitions from interfering with workloads in other partitions.

- **Virtualization.** Virtualization is utilized to provide more granular levels of control within the system, isolating workloads, allowing them to be moved within the system (to other partitions), and supporting workload consolidation within the scalable servers themselves. In the Integrity server systems running Windows, there are two primary options to achieve virtualization: HP Integrity Virtual Machines (VMs) provide hardware virtualization and can support multiple operating systems images, running in different guests on the same VM. In addition, Parallels Virtuozzo Containers (PVC) provide operating system virtualization that requires less system overhead than that associated with system hypervisors and increases the "headroom" for applications to run in a virtualized environment.

- **Native consolidation using resource management tools.** This approach allows a customer to consolidate multiple instances of the Microsoft SQL Server database onto one instance of the operating system. Customers can use native Microsoft resource management tools to manage the resources between the SQL instances. The native resource management tools include Windows System Resource Manager (WSRM), which enables the allocation of processing resources to SQL Server instances based on business priorities. Included is the ability to dynamically assign memory and processors to specific SQL Server instances.

- **Ability to consolidate multiple SQL Server instances.** To further enhance performance, a customer can consolidate Microsoft SQL Server instances and can place multiple databases within one database instance. The SQL Server 2008 Resource Governor can be used to manage the resources between the databases within an instance.

Management of Integrity Servers

Importantly, HP SIM manages all type of HP servers, including HP ProLiant x86 servers and HP Integrity servers for customers who have deployed both types of servers. There is also a common remote management solution across x86 and RISC-class servers. Integrity iLO 2 is now available on volume servers and blades as well as Superdome servers, with advanced remote management features for a virtual graphical console.

The SIM approach to hardware systems management focuses on system administration, monitoring for hardware component failure and delivering operations alerts, support for lights-out management (iLO), preventive maintenance capabilities, based on parameters set by administrators — and agents that link HP SIM to HP OpenView and to other system management frameworks.

Regarding software updates for the Microsoft Windows operating system acquired with the Integrity server, HP provides a subscription service, through the HP service contract for the server and operating system, which allows customers to access all system software updates, driver updates, patch and security update information from Microsoft, and new documentation. This provides investment protection, as long as the operating system support contract for Windows is maintained.

If customers supply their own copy of the Microsoft Windows operating system, such as when a company has a site license for Windows, the HP Smart Setup CD supplied with every Integrity server can be used to set up the system for operating system installation.

Balanced Performance

Balanced performance is an important design goal of scalable Integrity server systems so that throughput increases as capacity is added, with as little system "overhead" as possible during the scale-up process. When IT scales up workloads to run on servers with 4 or more sockets, it needs to consider the balance between processing power, memory, and server I/O. Without that balance, overall throughput would suffer, as processor-intensive workloads compete with I/O-intensive workloads for system resources.

To achieve this balance, HP Integrity servers are designed so that the processing, memory, and I/O in the Integrity system boards can be adjusted to the workload requirements. In production, these servers scale up with appropriate system resources, based on whether the workload is processor-intensive, memory-intensive, or I/O-intensive.

RAS and High Availability

The avoidance of downtime is essential to ensuring business continuity. Importantly, it is a key consideration for data warehouse and business intelligence workloads, ensuring that user productivity is not impacted by outages related to the server systems. HP Integrity systems have reliability, availability, and serviceability (RAS) features that are built into the server hardware itself to minimize disruptions to processing due to small errors or failure of hardware components.

These RAS features are designed into the hardware from the Itanium processor's Machine Check Architecture (MCA) to the I/O data pathways that connect the servers to outbound data storage devices. At the hardware level, this includes extensive support for error correcting code (ECC), use of code that ensures accurate data transfer within hardware components throughout the system, and protection of memory (and server I/O components).

For the scalable servers (those with 8 or more sockets), isolation of workloads within cell-based server partitions (via electrical isolation) means that any given workload can run unaffected by processing in other cells. This ensures that processing in the system will continue, even if one partition within the system encounters a hardware or software problem. It also provides an opportunity to adjust workloads, moving them to alternate resources, while some hardware components are repaired and brought back online.

Specific RAS features include the following:

- Double chip sparing, which restores chip-spare protection in the event that single-bit or multibit errors occur, which means that memory mirroring is not needed. This feature is supported by both the HP zx2 and HP sx2000 chipsets, which are included on HP Integrity systems boards.
- Dynamic memory resiliency through memory scrubbing. This feature removes single-bit errors that may arise within the memory.
- Dynamic page deallocation. This feature removes segments of memory in which some data errors persist.
- Support for ECC throughout the processor boards and throughout the data transfer paths within the system.
- Error detection and error handling for the PCI-X and PCIe I/O ports on the HP Integrity systems.

The presence of many RAS features improves overall availability for applications and data running on the systems. This support for high availability is a key aspect of running important workloads on scalable servers because there is very little tolerance for downtime when mission-critical workloads are being run.

Therefore, it is important for IT staffers to correctly "map" the applications to run within the appropriate amount of computing resources to which they are assigned. That is, it should be easy for them to assess which workloads are able to run within specific partitions on the servers — and which should be moved to run within other partitions. Equally important is the ability to partition parts of the server to "isolate" workloads so that interference between workloads is avoided — and processing problems with one workload do not slow processing for another workload. Finally, the ability to add resources, as needed, in a nondisruptive way, is another passport to scalability.

Microsoft Windows Server 2008 Support

By working with Microsoft, systems vendors that build scalable Itanium-based servers, including HP, will be able to deliver hot addition and hot replacement of processing resources without the loss of operating system or application state. These capabilities are supported by Windows Server 2008's support for dynamic hardware partitioning. This feature of Windows Server 2008 supports high availability of workloads and business continuity for mission-critical enterprise applications. HP plans to support this feature in future hardware products.

The release of Microsoft Windows Server 2008, with its enhanced high availability (Microsoft Failover Cluster) for clustering up to eight Windows servers, enhanced security (Network Access Protection, or NAP) that ensures that end-user access is secure), faster Web performance (via IIS7), and support for interoperability with other datacenter servers, has brought new levels of functionality in all of these areas, affecting both Itanium-based and x86-based servers. However, many longtime Windows customers will likely retain copies of Windows 2003 on some of their servers for some time to come while moving to Windows Server 2008 for new servers or for servers that would benefit most from the new feature/functionality provided by Windows Server 2008.

For Itanium-based servers, Windows Server 2008 Release 2 (currently in beta release) will support up to 64 Itanium processors, up to 128 processor cores — and up to 2TB of random access memory (RAM). Given Release 2's ability to support 2 threads per Itanium core, a single Superdome server will support up to 256 threads, meaning that up to 256 software-defined logical images of Windows workloads can be supported, expanding the system's overall scalability for workloads.

The Microsoft Failover Cluster software included in Windows Server 2008 supports up to 8 server nodes, each of which can house from 2 to 32 sockets (from 4 to 64 Itanium processor cores). IDC notes that a number of packaged software products, including the Microsoft SQL Server database, are supported by the Failover Cluster product, without any need to modify SQL server code or to create custom scripts before deployment.

Windows Server 2008 supports enhanced high availability via dynamic partitioning for hot add and hot replace of processors and memory. Importantly, Microsoft Windows Server 2008 for Itanium-based systems provides unlimited virtual instance rights — meaning that an unlimited number of VMs are supported on those systems at no additional charge. Similarly, the unlimited number of VMs is also supported by SQL Server 2008.

The virtualization on Itanium-based systems is provided by systems vendors, or by ISVs, such as Parallels Virtuozzo Containers virtualization software. In the case of HP Integrity systems, HP supports Windows virtual machines as software-defined partitioning for the Integrity servers.

Important features of Microsoft SQL Server 2008 Enterprise that support scalable Windows workloads include support for up to 64 processor cores, data and backup compression, enhanced auditing, transparent data encryption, enhanced database mirroring, Resource Governor (to better manage resources depending on workloads), and policy-based management capabilities. SQL Server 2008 supports a broad range of data types, including geospatial data, images, videos, and sensor information, in addition to standard types of data.

Microsoft SQL Server 2008 (as well as its predecessors) offers organizations data integration and BI functionality through SQL Server Integration Services, SQL Server Analysis Services, and SQL Server Reporting Services. Other Microsoft BI-related products include PerformancePoint Server, Microsoft Excel and Excel Services, and SharePoint Portal.

IDC notes that HP Integrity server customers also have the option of running other database management systems on their Itanium-based server hardware, including Oracle Database 10g, Oracle Database 11g, and other database products from ISVs. In addition, a number of business processing workloads, including many in the ERP and CRM categories, are available for use on Itanium-based servers running Windows. Readers should refer to the Itanium Solutions Alliance Web site for a partial listing of applications available for Windows/Itanium servers.

IT Benefits of Scalable Windows Servers

For sites that have deployed Windows servers in large numbers, acquiring scalable servers is an option that provides headroom for existing applications and additional capacity to run new applications. It also provides the following technical benefits to the IT organization:

- **Support for consolidation.** Consolidation — whether server consolidation, workload consolidation, or even datacenter consolidation — generally leads to reduced operational costs. IDC demand-side, customer-based ROI research has shown that operational costs related to power/cooling, downtime, and IT staff time are reduced through consolidation because it is simpler and easier to manage fewer systems — and to apply changes to those systems.

- **Virtualization for isolation.** Scalable Windows servers support multiple partitions, or separate virtual machines on scalable servers (preserving application isolation), each of which runs a separate Windows workload, for purposes of application isolation, ensuring that workloads do not interfere with one another and enhancing security. This ability to isolate, using Parallels Virtuozzo Containers or Integrity Virtual Machines on the Integrity platform, maps well to business initiatives that require compliance with government regulations, enhanced security for enterprise workloads, and protection against interference from competing workloads, which is a potential cause of system downtime.

- **Ease of management** (e.g., system administration, change management, provisioning, application of security updates). HP provides multiple and integrated management tools, including HP SIM for hardware management and HP OpenView for enterprisewide systems management. SIM and OpenView exchange information, allowing system administrators to manage more servers with less effort — and in less time. Software changes can be applied to fewer servers when workloads move to fewer server "footprints" — and updates will happen more quickly and efficiently, for the same reasons.
- **Nearby memory.** OLTP, data warehouse, and large single-system images benefit from being able to access more memory that is directly connected (proximate to processors). HP Integrity servers support large memory features that aid in data warehouse and BI analysis.
- **Improved availability characteristics.** Integrity servers have built-in RAS features to improve availability of applications. Their uptime characteristics reflect the presence of those hardware features. When these characteristics are combined with high-availability software, overall uptime is greatly enhanced. If two or more servers are used, then high-availability software will allow workloads to shift to alternate resources within the datacenter — or to remote sites, for disaster recovery purposes, if needed.
- **High availability — at a distance.** Geoclusters — once difficult to implement — are benefiting from improved networking feeds/speeds — and better SAN connectivity for rapid prototyping and entry to this market segment. Given compliance auditing and governmental regulations — and the need for business continuity — the business drivers are in place now to support the move to disaster recovery planning and testing and the use of geoclusters to enable continued processing at alternate sites. HP has added software that complements the Microsoft Failover Cluster high-availability software. In particular, these Microsoft-certified extensions, written by HP for use with Integrity servers, support geographically separate servers and storage for disaster recovery and disaster tolerance purposes. This HP solution includes Integrity servers, HP StorageWorks EVA/XP storage, and StorageWorks Cluster Extension software.

Business Benefits of Scalable Windows Servers

Businesses are under extreme pressure today to "do more with less," and this impacts business unit costs as well as IT costs. Because IT infrastructure needs to support business processes, IT infrastructure that is not efficient can be a drag on revenue and profitability. It is possible to optimize the IT infrastructure in a number of dimensions through the use of scalable servers for the most demanding IT workloads. Operational benefits stemming from the use of scalable servers for selected workloads include the following:

- **Saving operational expenses (opex) from consolidation of Windows workloads.** Historically, many Windows workloads were deployed on small, dedicated systems, which has provided an opportunity for customers to achieve operational cost savings through consolidation of those workloads. Fewer server footprints translate to more efficient management of workloads, in many cases.

Moreover, through consolidation, the business can achieve savings as a result of fewer licenses, lower support and maintenance costs, reduced power and cooling, and less floor space required. A variety of approaches may be used as part of the consolidation project, including virtualization of system hardware to support more workloads, improving compute density — and importing data from Microsoft SQL Server databases to run on scalable servers.

- **Improving uptime via enhanced high-availability features and software as well as management software.** This happens through failover clustering that allows applications to be restarted on alternate servers (including geoclusters via WAN connections), through workload balancing that allows work to move to servers with the capacity to run new workloads, and through the use of software that more efficiently manages all workloads, whether running on physical or virtual servers. Integrity servers support multiple levels of system management via Microsoft MOM, HP SIM, and HP Integrity Essentials, as a component of HP OpenView.

- **Reducing IT staff time associated with managing Windows workloads.** Consolidation and more efficient management reduce IT staff time that would otherwise have been devoted to maintenance of servers that have gone offline — or to maintaining large numbers of smaller servers. By bringing the most demanding workloads to servers with advanced RAS features and highly controllable high-availability software, IT staff time associated with outages and downtime can be reduced, in turn reducing a business' overall operational expenses.

- **Scaling up applications.** Scaling up applications on scalable servers results in the ability to support more end users, with hundreds to thousands of end users supported per server. The capability of Windows servers has increased over time due to a new generation of multicore Itanium processors, larger memory components in the Integrity systems, improved server I/O technologies, and new software products that support improved availability.

- **Obtaining green IT benefits.** These benefits can be gained by running selected workloads more efficiently on fewer, more scalable servers. Green IT covers a host of technologies — all of which act to reduce power/cooling costs and to improve the energy efficiency of servers, storage, and associated equipment within the datacenter.

CHALLENGES AND OPPORTUNITIES

Bridging the IT and Business Cultures Within an Organization

The scenarios for deployment of Itanium-based servers need to be made more accessible to business unit managers. The two "cultures" within the enterprise — IT and business managers — are quite different. IT managers look toward technical specifications (regarding RAS and performance) and their own in-house evaluations to inform purchasing decisions. In contrast, business managers want to understand

the changes in their business operations that the new technology will provide. Accordingly, there is a need to translate the technical aspects of Itanium-based servers into a series of business cases for deploying Itanium-based servers. IT and business managers must understand the Itanium-based servers' value proposition before both groups agree to acquire, and deploy, those servers across the organization.

One of the ways to do this is to show how the technical changes being made in the infrastructure will translate into operational cost reductions and operational efficiencies that will go to the bottom line of business results. IDC demand-side research has consistently shown that concerns about power/cooling, availability, and use of datacenter space are the "top-of-mind" challenges within the datacenter. This means that IT organizations need to demonstrate how the servers, storage, and networking equipment in which they are investing will pay off for the business. The ongoing process of workload consolidation can address these opex issues and show real results, but the IT organization will need to quantify and to explain those savings for management.

Improving Business Agility

HP is finding ways to "map" the technical benefits into a series of business benefits. The principles show how business goals — such as achieving greater flexibility in IT systems or more business agility — are supported by specific IT infrastructure technologies. One example is HP's support for both Itanium server blades and x86 server blades within its HP BladeSystem blade-server chassis offerings. Another is the way HP provides choice in form factors in its server offerings (blade, rack-optimized, and pedestal/free-standing), depending on customer requirements and capacity needs.

CONCLUSION

Mission-critical workloads — those workloads that would negatively impact business results if interrupted — require RAS features to ensure continued operations on an ongoing basis. RAS features in hardware and software features that support high availability and disaster recovery are two components of an overall high-availability strategy for all enterprise customers. For Microsoft Windows workloads, the ability to leverage RAS features is becoming increasingly important, especially for enterprise workloads, such as LOB applications (e.g., ERP, CRM), database workloads, and BI workloads. As these applications grow in size, and the number of end users accessing them grows, the availability of these applications is becoming increasingly important to business unit managers — and to senior management in company headquarters.

Meanwhile, increasing compute density, through the use of multicore processors and virtualization software, is packing more processing power into smaller and smaller form factors, including server blades and rack-optimized servers. Projects that consolidate workloads onto fewer server footprints only magnify the business impact of any interruption in processing on these platforms. So, as virtualization and

consolidation projects go forward, with the aim of reducing operational expenses, the importance of deploying highly reliable servers increases, over time.

For HP Integrity servers, RAS features are built into the hardware design, protecting memory and processors and data transfer points throughout the Integrity systems. Microsoft Windows Server 2008 and the next release of the operating environment (Release 2) are being adopted for scalable workloads on Itanium-based servers, such as the HP Integrity server products. As those Windows workloads scale up, reaching broader communities of end users and end customers, continuing access to those demanding workloads will become increasingly important to both IT managers and business managers in enterprises worldwide.