

## WHITE PAPER

---

# Evaluation of Mission-Critical Computing Systems: Key Criteria for High-End Systems Customers

Sponsored by: HP

---

Jean S. Bozman

Matthew Eastwood

June 2011

## EXECUTIVE SUMMARY

Mission-critical computing has an important role to play in the modern enterprise, particularly given the importance of maintaining 24 x 7 x 365 operations for applications being accessed globally. In a world with supply chains that extend across continents, and with end-to-end applications that span computing tiers, mission-critical computing plays a vital role in supporting uninterrupted business processes.

Now a new generation of technologies is transforming the way that customers acquire and build out their enterprise computing systems. Modular components, multicore processors, and high-speed networking fabrics are giving system designers new building blocks for systems that support thousands of end users, run millions of transactions per day, and ensure 24 x 7 x 365 around-the-clock operations.

This IDC white paper provides a summary of top evaluation considerations for enterprise computing — including the top characteristics for mission-critical computing platforms — all of which impact acquisition costs and operations costs. IDC research has shown that the top concerns of datacenter managers include availability, security, use of datacenter floor space, and power/cooling. Today, all of these considerations must be addressed by scalable systems. These high-end servers must be efficient — and cost-effective — in order to be acquired and maintained over time.

Finally, the paper describes how HP Superdome 2 addresses these top IT concerns and how it addresses the evaluation considerations for scalable enterprise platforms. In step with its modular philosophy for building out computing environments, HP has launched the second generation of its HP Superdome mission-critical computing platform (HP Superdome 2). From a business perspective, HP Superdome 2 supports business continuity through its deep use of maximum reliability, availability, and serviceability (RAS) hardware and system software.

In summary, the scalable HP Superdome 2 server is leveraging the design principles of blade-based servers and flexible deployments, along with support for dozens of RAS hardware features at the processor and system levels, to:

- Scale up capacity, as needed, in a flexible and balanced way
- Support virtualization and isolation for workload security
- Provide the high levels of reliability and availability that mission-critical computing customers expect for their most demanding enterprise workloads

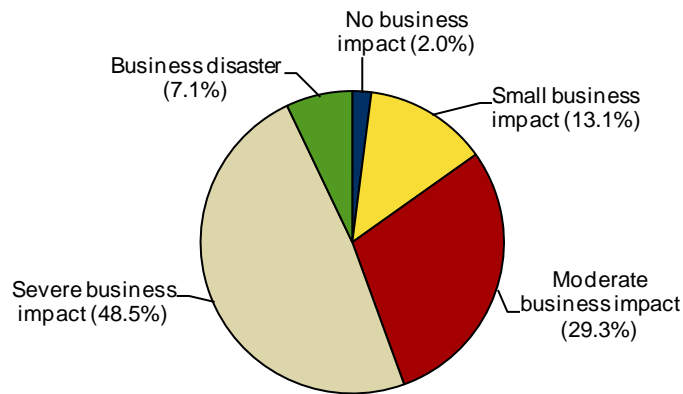
## SITUATION OVERVIEW

As organizations become increasingly global and business processes become more automated, organizations also become more reliant on their IT infrastructure to deliver consistent, highly available, and highly scalable services to users across the globe. This point is illustrated by looking at the importance of minimizing system downtime.

IDC's recent Mission-Critical Workloads study showed that the impact of even one hour of downtime can be far-reaching. Slightly more than 48% of respondents said that one hour of outage would have a "severe business impact," and another 7% rated it as a "business disaster." Taken together, more than 55% of respondents recognize that one hour of downtime would be excessive (see Figure 1).

**FIGURE 1**

**Business Impact of One-Hour Outage for Mission-Critical Applications**



Source: IDC, 2011

Organizations are considering more workloads to be mission critical than they have in the past. While a variety of approaches support mission-critical computing, the IDC Mission-Critical Workloads study determined that the majority of respondents (59.1%) prefer that high-availability solutions for their server systems reside at the hardware level rather than at the software or middleware layer.

Although software will continue to be key to all highly available deployments, the embedding of RAS features in the hardware platform results in less IT staff time devoted to one-off customization of scripts to manage a restart of a critical application on alternate server resources. To the extent that computing instructions can be automatically redirected to computing components, and that memory can be accessed, even in the event of transient errors, then processing will continue and business continuity will be preserved.

The next section presents a view of the evaluation process that IT organizations and business managers go through when considering new server platforms to run highly demanding, mission-critical workloads.

## THE EVALUATION PROCESS: A GUIDE FOR LOOKING AT MISSION-CRITICAL COMPUTING PLATFORMS

For enterprises with mission-critical computing requirements, the risk of suffering any downtime is unacceptable. In an unforgiving business environment, the loss of core services to customers, partners, or employees can be a make-or-break event. For this reason, enterprises with mission-critical demands must ensure their server infrastructure is optimized to deliver the maximum RAS. The degree to which enterprise systems provide support for highly demanding mission-critical workloads can distinguish the efficacy of that system solution in ensuring business continuity for mission-critical workloads.

Generally, when customers consider their mission-critical computing needs, they enter a period of evaluation in which they:

- ☒ Study the systems they may acquire
- ☒ Compare the systems' capabilities and price/performance
- ☒ Discuss those systems with other IT sites, system integrators, and consultants prior to testing their own applications on the system that is being evaluated

Throughout these stages of the process, organizations are evaluating their choices according to a number of criteria. The major criteria are discussed in the following section.

---

### Key Considerations for the Evaluation Process

Key considerations for customers who are evaluating scalable servers for their datacenters include the following:

- ☒ **Scalability.** In most organizations, databases expand and applications are accessed by larger numbers of end users over time. Capacity to support growing workloads must be built into platforms that support highly demanding enterprise computing. The ability to scale up involves more than merely adding processors; it includes the ability to balance the scale-up with enough memory and I/O resources to ensure continued performance at high levels. The process of workload consolidation — which allows customers to rehost applications onto fewer, more scalable servers for efficient processing and management — must be supported on next-generation enterprise computing systems.
- ☒ **Reliability.** The presence of RAS features improves the resilience of the hardware platform and allows processing to continue, even if that involves a retry of a given computing task or isolation of a segment of corrupted memory. All systems face the possibility of transient errors, but RAS features preserve process capabilities, prevent data corruption, and support business continuity, even in the event of a transient error.
- ☒ **Availability.** This is the ability of the system to continue processing production workloads, even in the presence of faults. It ensures the attainment of agreed-upon service levels, such as uptime levels over a given time period, or availability of a critical application that is being accessed by hundreds, or thousands, of end users.

- ☒ **Flexibility in deployments.** Bladed technologies have transformed the server world, allowing capacity to be added easily and quickly. Modular system designs support fast replacements or upgrades and minimize IT staff time devoted to maintenance. Older SMP platform designs often did not provide this type of flexibility. Today, with the lessons learned about modularity and flexibility from the use of bladed systems technology, organizations are looking for that same type of agility to be provided by their enterprise systems. With system boards that can be added individually as needed, modular system designs allow an organization to build up total system capacity in step with the business' growing demand for system resources.
- ☒ **Virtualization within the system.** Virtualization technology creates a more dynamic and fluid environment in which compute capacity can be treated as a resource pool, protecting application service levels and improving agility and responsiveness to business needs. Importantly, virtualization isolates workloads so that they do not interfere with one another, which improves overall system uptime. Having virtualization in the system also increases the importance of having RAS features in the hardware, as more virtualized workloads are housed within a single physical server — because with virtualized servers, organizations have "all their eggs in one basket." This is especially important for scalable servers supporting large numbers of end users who are accessing applications that are critical to the entire business.
- ☒ **Life cycle (longevity of deployments).** When customers acquire a scalable server system, it is usually deployed for five to seven years or more. The length of this deployment means that customers must have the ability to upgrade or to add to the resources that are initially installed. The era of "forklift upgrades" is long past — and now customers want to be able to utilize modular building blocks to support multiple generations of technology within the same infrastructure, as new technologies come onto the market. To minimize disruptions, IT organizations should consider future proofing new investments — and that means purchasing and installing servers that will grow and adapt as their needs change, extending the lifetime of their infrastructure. Systems that are based on flexible architectures, in which capacity can be added as needed and processing can be moved to available resources, not only allow customers to stay current on the technology with minimal disruption but also can provide a source of competitive advantage by supporting a more agile business.
- ☒ **Serviceability.** This is the ability of a system to be repaired or upgraded quickly; the goal of serviceability is to provide online service so that no downtime is incurred. Examples of serviceability features include system software that allows continual monitoring and hot-pluggable components in the server enclosure, which allow for rapid replacement without interrupting data processing by the rest of the system.
- ☒ **Integration.** For mission-critical workloads, integration of functionality and the ability of the hardware to work closely with the system software, through the process of optimization, are key to smooth performance. Support for automation will avoid IT staff time that would be spent scripting or doing custom work or system integration in order to enable the overall hardware/software solution. This impacts operational expenses, saving both time and money for the IT organization and for the business.

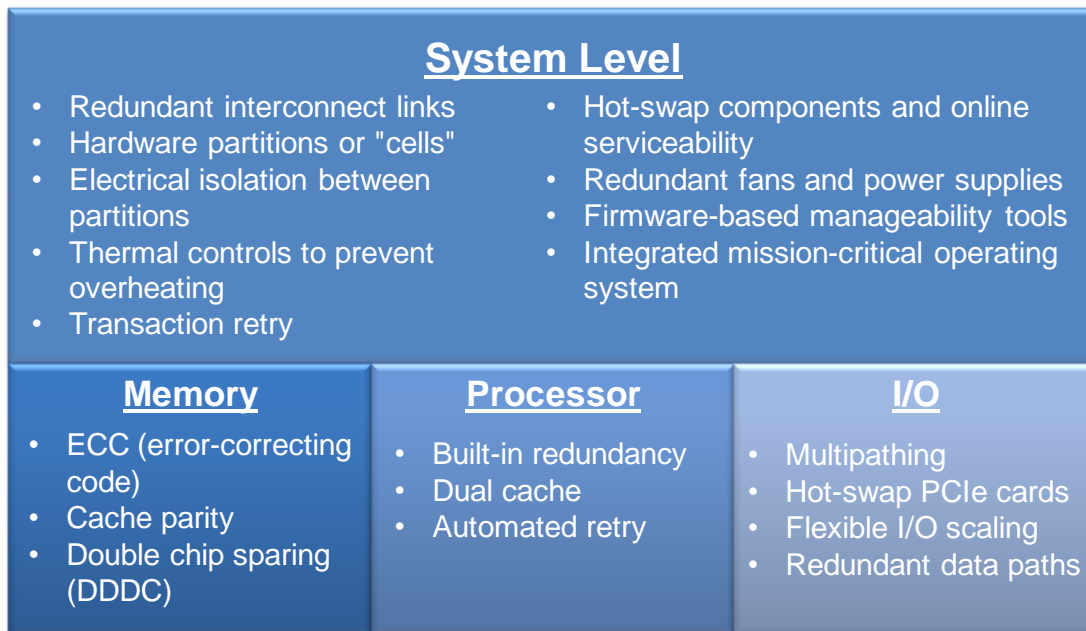
## Technical and Business Considerations

Businesses considering mission-critical computing platforms must evaluate not only important technical requirements but also business requirements that address economic considerations, such as the ability to deliver a good return on investment (ROI). Key technical and business requirements include:

- ☒ **Technical requirements.** Technical requirements include multipathing, which provides alternate paths for all data traffic, additional memory to surround the multicore processors, advanced virtualization that includes partitioning capability (for workload isolation), and the ability to shift workloads to alternate hardware resources, as needed, so that production workloads can continue to run smoothly (see Figure 2).

**FIGURE 2**

### RAS Features in Servers



Source: IDC, 2011

- ☒ **Business requirements.** IT staff must be as productive as possible. IT staff costs should be reduced, wherever and whenever possible, through automation and reduced periods for planned downtime to perform routine maintenance. Operational costs including unplanned downtime and power/cooling/datacenter efficiency should be addressed. The productivity of both IT staff and end users should be optimized by keeping systems online and improving business continuity. Last but not least, IT must have the flexibility to support rapidly changing business requirements. By enabling greater business agility, IT can help businesses take better advantage of opportunities, more rapidly meet shifting market demands, and respond better to competitive threats.

## KEY COMPONENTS OF ROI

With IT organizations facing severe budget constraints, many CIOs are required to demonstrate the bottom-line business value for any and all new strategic investments. The most common metric to project business value is ROI.

ROI analysis seeks to understand and compare all economic returns and economic costs associated with the investment. Costs include not only the acquisition price of the servers themselves but also the operational expenses associated with IT personnel management time, power and cooling, and software and upgrades.

The economic benefits of a new server deployment may vary from organization to organization and workload to workload, but broadly speaking, they tend to fall into one of several categories:

☒ **Improved system availability.** Improved system availability can reduce the risk of direct costs associated with system downtime as well as the damage to the organization's reputation and resources required to find and fix incidents.

☒ **Reduced IT operational management costs.** Newer servers often have improved management tools, automation, and other capabilities over the systems they replace, enabling IT administrators to be more efficient in their server management and free up their time for more productive tasks.

### Avoidance of Downtime

Organizations have employed a number of architectural and deployment choices to avoid downtime, including load balancing, failover, and disaster recovery methods. Strategies exist in both hardware and software, and combining solutions from these two layers can drive an even more robust solution, reducing the risk of unplanned downtime significantly. High-performance systems need to be designed with the appropriate architecture and technology to be highly resilient and self-healing and include hot-swappable components — all of which contribute to maximum availability with very low risk of user-facing downtime.

#### Planned and Unplanned Downtime

Downtime fundamentally can be divided into two categories: planned and unplanned. Planned downtime is usually associated with systems maintenance, and while this was a common requirement for previous generations of servers, the nature of today's global 24 x 7 x 365 business environment makes it difficult for organizations to justify even small windows of planned downtime.

Unplanned downtime can potentially affect any number of business-critical functions and can occur at any time of day. For this reason, unplanned downtime can have a number of economically costly ramifications for the business, including the following:

- ☒ **Revenue loss.** The most direct and immediately felt effect of unplanned downtime is when it occurs in revenue-producing, customer-facing systems. As businesses of all varieties automate their business functions, revenue loss of downtime is not limited merely to industries such as financial services and e-commerce; instead, it can be felt in enterprises in nearly any industry. For mission-critical workloads in many large enterprises, the real cost of downtime can often be measured in millions of dollars per hour.
- ☒ **Reduced user productivity.** Users across the organization, from employees to contractors and partners, rely on IT-delivered services and applications for their business productivity. Downtime can greatly reduce their productivity, and for many knowledge workers, downtime in just the right (or wrong) systems can grind their workday to a halt.
- ☒ **Customer disruption and reputational damage.** Even when downtime occurs in systems that are not directly revenue producing (for example, if it happens in customer service or support systems), customers can still suffer from the disruption and the organization's reputation can be damaged. Sufficiently provoked, dissatisfied customers can take their business elsewhere, and poor word of mouth can cause future sales to suffer.
- ☒ **Resources required to isolate and repair issues.** The costs of downtime also extend to the personnel and resources required to find, troubleshoot, and fix the issue. Postmortem analysis must also be completed, siphoning off dozens of hours of productive IT staff time that could be spent on more strategic projects.
- ☒ **Overprovisioning of resources to compensate.** Many organizations guard against unplanned downtime by overprovisioning server, network, and storage resources; building redundancy into their systems; and keeping pools of hot-swappable hardware on hand. This is expensive not only in terms of the additional equipment that must be purchased and maintained but also in terms of the additional staff hours required to set up and maintain the equipment and the extra network and storage resources that must be kept online.

#### Operational Efficiency: IT Staff Time, Power and Cooling Considerations

As the installed base of servers has grown over the past several decades, the operational costs of maintaining servers in terms of IT staff time and power and cooling have grown to the point where they have eclipsed the server acquisition costs themselves. This has placed a squeeze on IT budgets, which are severely constrained as it is, and as a result, it has become an imperative for IT organizations to wring every bit of efficiency possible out of their operations.

To address this challenge, IT organizations need to consider computing platforms that provide the greatest amount of operational efficiency and enable their IT staff to be as productive as possible. Such systems drive efficiencies by, for example, providing the ability to closely monitor system components and to flag components that may need to be replaced while avoiding unplanned downtime.

- ☒ **Reduced security and compliance risk.** Security and compliance are top-level concerns for enterprises around the world, and observing the highest levels of systems security can guard against the potentially costly loss of sensitive data.
- ☒ **Reduced power and cooling.** Many systems today incorporate innovations in power and cooling. Replacing previous-generation servers can yield very real benefits in operating expenses for power and cooling.
- ☒ **Reduced facilities and overhead costs.** By reducing their server footprint — for example, by implementing blade servers — organizations can free up space in their datacenters and reduce associated facilities and overhead costs.

## HP SUPERDOME 2 TECHNOLOGY

In 2000, HP introduced the HP Superdome line of high-end servers, targeted toward the most demanding mission-critical environments. HP Superdome, the flagship model in the HP Integrity line of Unix servers, was designed to address the needs of business-critical computing and to address each of the key considerations in the evaluation process described earlier: scalability, reliability, availability, flexibility, virtualization, longevity, and serviceability. Further, in its design, HP considered not just the technical requirements but also the business requirements to justify its customers' investment in the system.

HP Superdome 2 is HP's first mission-critical platform that combines the RAS features of its HP Superdome platform with common modular components that are consistent across the entire HP Converged Infrastructure (CI) portfolio. In many ways, this provides a unified approach for IT installations, combining the modularity and flexibility of blades with the mission-critical computing features of the scalable HP Superdome servers. Importantly, the entire hardware/software stack for HP Superdome 2 is binary compatible with the HP Integrity series of Unix server systems. This approach enables organizations to preserve previous investments in system software and in the IT skill sets that will deploy and maintain HP Superdome systems.

Over the past decade, Fortune 1000 organizations in a variety of industries have used HP Superdome to run their most mission-critical applications. With the launch of HP Superdome 2 in 2010, HP has built on the modular, cell-based architecture of the original system and introduced a blade-like form factor to improve the overall platform's modularity, flexibility, and scalability. In addition, with HP Superdome 2, HP has introduced a number of new RAS features designed to provide even greater levels of uptime than were available with the previous HP Superdome offering.

The HP Superdome 2 system has been completely renewed, from its components to its system boards and from its system interconnects to its support for I/O devices. These changes build upon the innovative HP Superdome mission-critical computing technology, aligning it even more fully with HP's Converged Infrastructure strategy — combining the server/storage and networking elements of the datacenter into a tightly integrated system-level product to support mission-critical workloads.

## Integrated Systems Capabilities

The HP Superdome 2 design features enhanced RAS capabilities, as described in this paper, and integrated functionality for availability, security, system management, performance management, storage management, power utilization, and power management. This combination of functionality, which works in a holistic way, allows the system to scale up resources, as needed, to support growing applications and database workloads. It also supports flexibility, in terms of the ability to add capacity, as needed, without rehosting workloads on other servers.

Chief among these system components is the HP-UX 11i v3 operating system itself, which is completely optimized for use with HP Superdome 2. By leveraging HP-UX 11i v3, customers can purchase, install, and maintain a package of up to 160 software components — and manage them as a single system. This built-in integration reduces risks because it reduces the potential for human error in the installation of dozens of software products — and it results in simpler life-cycle management.

### ***Modular Design***

The HP Superdome 2 has a highly modular design, to which system boards can be added as demand for processing grows over time. Resiliency — also of paramount importance in a server — is designed with component redundancy, electrical isolation of workloads, and a service advisor (Analysis Engine) in firmware that identifies system issues before they become operational problems. Together, these features make the HP Superdome 2 server a flexible, reliable platform designed to meet customers' current needs while scaling to provide a growth path for the future.

HP Superdome 2 has numerous self-healing, error detection, and error correction features to provide the high level of reliability and availability expected from HP Integrity servers. Meeting customer preferences for hardware-based RAS features, HP Superdome 2 includes many new innovations, and features are included to keep critical applications up and running, including:

- ☒ **Modular architecture.** With HP Superdome 2, HP incorporated many best practices and technologies from its HP BladeSystem offerings, including hot-swappable components, modular design, and flexible deployment scenarios. This greatly improves the ability to deploy and to manage the systems and allows customers to use interchangeable, standardized parts throughout the datacenter, driving down inventory costs and driving improvements in component quality.
- ☒ **Fault-tolerant fabric.** HP Superdome 2 Crossbar Fabric is an innovative design that provides a fault-tolerant approach to maintaining blade-to-blade and I/O traffic while enabling scaling capabilities throughout the system. With this fabric, transactions are tracked to successful completion, and, even in the event of a link failure, traffic can be rerouted down a secondary path with zero data loss and no interruption of service. This allows failed fabric components to be serviced with no incurred downtime and prevents numerous errors without the system ever knowing. The fabric allows I/O scalability independent of processors and memory.

- ☒ **Online serviceability.** Online serviceability is provided through tool-free, hot-pluggable components. Hot-swappable components include utilities modules, HP Onboard Administrator (which allows for local and remote administration of the enclosure), I/O switches, power supplies, and cooling fans. One-click online firmware updates can be accomplished without bringing down the enclosure.
  
- ☒ **Analysis Engine.** The HP Superdome 2 Analysis Engine proactively monitors a broad set of hardware attributes. It enables predictive, real-time failure analysis and self-healing by driving automated responses to failures. For example, if a processor is showing a high number of errors, the Analysis Engine can deactivate the processor and inform the administrator that a replacement is needed. In this engine, HP moved diagnostics capabilities from the operating system level to the firmware level, which allows offline analysis without the need for an additional, external management system. The Analysis Engine utilizes these features to prevent outages and significantly reduce planned and unplanned downtime.
  
- ☒ **Electrically isolated hard partitions (nPartitions).** The HP Superdome 2 utilizes an innovative nPartition architecture that takes the hard partitioning features available in the original HP Superdome to the next level, enabling customers to consolidate multiple workloads on a single platform with greater confidence. HP has significantly improved the reliability of the crossbar switch that logically separates physical partitions. These advances provide improved isolation for independent coordination and servicing of workloads and partitions, which allows processing to continue even while system components are replaced. The design in HP Superdome 2, using electrically isolated hard partitions, contrasts with many industry-standard designs in which a shared backplane results in competition for a shared electrical bus with potential shared failure modes, queuing delays, and saturation of the backplane.

---

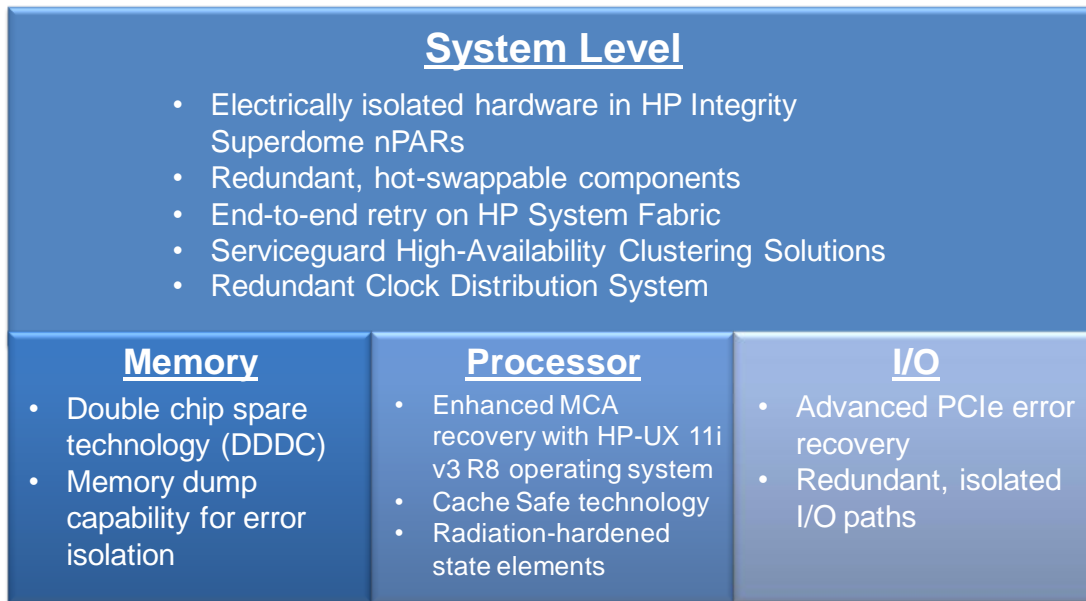
## **Addressing Key Market Considerations for Mission-Critical Computing with HP Superdome 2**

### ***RAS Features***

The innovative redesign of HP Superdome 2 builds on the RAS features found in the original HP Superdome to provide a new approach to scaling up enterprise computing resources. As stated earlier in this paper, RAS features enforce automatic retries for instructions that find a transient error — thus preserving uptime for the overall system. The HP Superdome 2 combines the RAS features built into the Intel 9300 Series of quad-core Itanium processors with the system-level RAS features designed and implemented by HP engineers. Together, these hardware features provide a solid foundation for high availability and support for demanding business processing and decision support workloads (see Figure 3).

**FIGURE 3**

Top HP Superdome 2 RAS Features



Source: HP, 2011

Key RAS features built into the HP Superdome 2 platform include:

- ☒ **A unified fabric leveraging crossbar switches** connects all computing components of the system and ties them together with fully redundant, high-speed links. This is called the HP System Fabric.
- ☒ **Quad-core Intel Itanium 9300 processors** were introduced in 2010 with a new wave of RAS features and onboard QPI system interconnects to increase system performance compared with earlier generations of Itanium processors.
- ☒ **Memory double chip spare technology (DDDC)**, as instantiated in the Itanium 9300 series processors, allows systems to continue operation with up to two failed DRAMs in a single rank and still have protection from single-bit errors at lower cost than memory mirroring or DIMM sparing techniques.
- ☒ **HP Superdome 2 Analysis Engine.** A feature supported on Intel Itanium 9300 series processors running HP-UX 11i v3, the HP Superdome 2 Analysis Engine allows administrators to isolate the specific root cause of a failure, gathering enough evidence to determine whether the failure was caused by a hardware component or a software component — and speeding the time to repair.

- ☒ **Third-generation Machine Check Architecture (MCA) Recovery.** Each Itanium processor supports MCA, which protects against unplanned downtime through the use of online error detection and correction features. This provides HP Integrity systems based on Itanium 9300 processors with the capability to recover from common memory errors without downtime, either by refreshing memory pages or by halting affected processes, preserving overall system continuity.
- ☒ **Isolated I/O paths** protect I/O and allow processing to continue, even if other I/O paths experience transient errors. Radiation-hardened state elements in the Itanium 9300 processor and chipset shield the entire system from cosmic radiation, such as gamma rays, which can disrupt data bits in transit.
- ☒ **Electrically isolated hardware** prevents errors in one hardware-defined partition from affecting other partitions. On a system level, this feature avoids downtime that could otherwise occur due to interference from workloads running elsewhere in the system.
- ☒ **A front-to-back cooling system** is aligned with the front-to-back cooling designs seen in HP's rack-optimized and bladed servers — allowing HP Superdome 2 to be housed in the same hot/cold aisle installations in today's datacenters as other HP servers. This supports a unified approach to hardware maintenance and is housed in the same 19in. racks as HP's ProLiant x86 servers.
- ☒ **Additional RAS features.** Other RAS features include fully redundant, hot-swappable components such as the system clock tree, custom PCIe Gen2 I/O hub with advanced I/O RAS features, and support for easy serviceability of major components and field replaceable units (FRUs), which can be kept at the customer site.

---

## HP Software Contributes to Availability

HP Superdome 2 also benefits from a number of HP system software solutions that are designed to further increase the organization's reliability, availability, and serviceability, including:

- ☒ **HP-UX 11i v3.** HP-UX 11i v3 Unix is the flagship operating system for HP Superdome — as it has been since HP Superdome's launch in 2000. HP-UX 11i v3 (Update 8) is the newest version of the HP-UX Unix operating environment, released in March 2011. Targeted to support mission-critical environments, it is designed to deliver always-on resiliency, dynamic optimization of resources, and stability for mission-critical computing. It features HP innovations in high availability, security, workload management, and instant-capacity-on-demand features to improve IT flexibility and stability in the datacenter infrastructure. Among these features are virtual partitioning capability, resource containers for software-enforced workload isolation, and support for moving workloads between software-defined virtual machines (VMs), which are managed by HP's hypervisor for HP Integrity platforms.

- ☒ **HP-UX 11i Options for Deployment.** HP-UX provides several options for customers, depending on the types of workloads they intend to deploy on the system. The options are HP-UX 11i v3 Foundation Operating Environment, HP-UX 11i v3 Mission-Critical Operating Environment, and HP-UX 11i v3 Enterprise Operating Environment. Each option presents a combination of functionality aimed at supporting specific workload requirements.
- ☒ **HP-UX 11i Virtualization.** HP Virtualization capabilities are present in all of the Operating Environment deliverables, supporting the creation of VMs and allowing them to be adapted to changing business needs.
  - ☐ HP hardware-defined nPartitions ensure electrical isolation and security isolation, preventing any given workload from interfering with another one. This is highly important in a datacenter where multiple workloads, supporting multiple departments, are colocated.
  - ☐ Highly granular software-defined virtualization provides highly granular control for workloads, as follows: HP-UX Virtual Partitions allow customers to carve nPartitions into finer-grained software partitions. In addition, Integrity Virtual Machines provide hypervisor-based virtualization with shared resourcing — supporting dynamic resource allocation that is based on demand for resources and role-based user entitlement. Finally, HP-UX Containers (formerly known as HP-UX Secure Resource Partitions) support a lightweight workload deployment that allows IT organizations to consolidate multiple workloads within a single instance of the HP-UX 11i operating system.
- ☒ **HP Serviceguard Solutions.** Serviceguard Solutions provide automated availability and clustering software (ACS) that monitors system availability while enabling access to important enterprise workloads, including custom applications, ISV applications, and databases. It supports workload failover, in the event of an outage, restarting applications and databases on other resources, as needed. When a failure or threshold violation is detected, the software restarts the service in a controlled manner to resume normal operations on alternate hardware resources. This means that it supports clustering between multiple server nodes, linking servers to storage and triggering a failover. It can also be used to move workloads proactively, in the case of planned downtime, when a given server or server resources are being maintained or repaired. HP Serviceguard is shipped with the Serviceguard Manager, which gives system administrators a "single pane of glass" view of the entire infrastructure through a single console.
- ☒ **Enhanced MCA Recovery.** HP's HP-UX 11i v3 operating system has been optimized through joint engineering work between Intel and HP to maximize the benefits associated with the MCA that is built into every Itanium processor. These software enhancements provide an even greater level of protection against unplanned downtime through the use of error-correcting code (ECC), along with retry for incoming instructions if an error is detected.

- ☒ **Security in HP-UX 11i v3.** The HP-UX 11i v3 operating environment integrates security tools that are certified and independently validated by third-party security organizations. The validity of these certifications is important to both government and enterprise customers. Many governments, including the U.S. government, require certification as part of the IT procurement process. For the same reasons, enterprise customers value independent security certification because it ensures that their global networks of systems will be able to resist tampering or hacking due to Internet-based or internally based security threats. HP-UX 11i v3 complies with federal standards for highly granular access control lists (ACLs) and role-based access — all of which is built into the HP-UX 11i v2 Base Operating Environment, which underlies all three operating system "builds": Foundation, Mission-Critical, and Enterprise. Finally, HP-UX 11i v3 has Security Container technology, which isolates workloads, prevents security breaches, and, by doing so, avoids downtime.

## CHALLENGES/OPPORTUNITIES

Like all vendors, HP must compete in a large, global marketplace for servers, in which there are many platform solutions that can run the spectrum of enterprise workloads. In this environment, much attention has been placed on scalable servers and less costly alternatives. However, the importance of mission-critical workloads still demands highly available platforms, especially when hundreds, thousands, or even millions of end users depend on the resilience of an enterprise server.

As it goes to market, HP demonstrates some key points to prospective customers: the cost-effectiveness of its modular design; its reliability in support of the most demanding workloads within the enterprise; its flexibility in terms of deployment options; and its highly resilient hardware, which avoids costly downtime.

At a time when many customers are looking to update their central-site IT infrastructure, HP has an opportunity to present the HP Superdome 2 as a platform for workload consolidation. This consolidation can reduce operational costs over the system's useful life cycle when compared with older Unix servers from HP and other vendors that were carrying similar workloads — many of which had been installed at divisional locations within the larger enterprise.

Now, with networking improvements, it is possible to recentralize many of these geographically dispersed workloads for central-site IT staff management. Once deployed with the HP Superdome 2 system, the HP Analysis Engine will provide preventive maintenance, improving overall uptime for a highly available resource designed to provide near-continuous computing capabilities to the entire organization.

## CONCLUSION

Many customers are evaluating technology refresh projects or outright replacements of older technologies. For many IT planners, such projects had been delayed or deferred due to the economic downturn.

Now, computing demand is on the rise, fed by new applications, support for mobile devices, and the rapid growth of data — and the need to analyze it. In some regions of the world, there is rapid growth in infrastructure — telecommunications, transportation, financial services, and government — that must support large numbers of end users and that requires resilient and highly capable, scalable systems.

As users and businesses become increasingly reliant on IT-delivered services and applications for their everyday needs, a greater number of workloads are being classified as mission critical. Organizations must meet this challenge by scaling their mission-critical computing infrastructure appropriately. In 2011, many longtime customers of Unix servers are looking to consolidate workloads onto fewer, more scalable footprints in the datacenter — and they are actively evaluating high-end systems as destinations for their organization's most demanding workloads. The advanced virtualization capabilities, high levels of RAS features, and overall support for business continuity make scalable servers important platforms to host mission-critical applications.

HP Superdome 2 builds on the mission-critical performance of the original HP Superdome, redesigning it with new innovations from the ground up. The entire system leverages high-speed interconnects and supports balanced processing by providing I/O scalability independent of processor and memory scaling. For those who have used HP's Unix servers before, the new HP Superdome 2 is a next step toward evolving the architecture to be more flexible — and more modular — than before. For new customers, it will provide an enterprise computing resource that can be applied to immediately address real-world IT challenges, locally or globally, across the organization.

---

## **Copyright Notice**

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2011 IDC. Reproduction without written permission is completely forbidden.