Technology Insight Paper

Are x86 Servers the Right Platform for Running Mission Critical Applications?

By Eric Slack, Sr. Analyst

February 1, 2016

Enterprise data centers are charged with supporting the most critical applications in the company, typically the ones that can't go down. To achieve this, IT organizations have traditionally used mainframes (and many still do), along with enterprise-grade server platforms that typically ran UNIX on proprietary RISC hardware. These systems were designed to run most of the important applications in the data center and keep them running through component and subsystem failures, human error and maintenance/upgrade windows. These servers were also designed with the headroom to scale non-disruptively as the enterprise grew, without 'box swaps' or 'forklift upgrades'.

Servers based on the Intel x86 architecture gained initial acceptance in IT for use as file and print servers and other less critical applications; however, recently the use of x86-based servers has steadily gained popularity in the technology industry due to their relatively low cost. These standard-architecture platforms are now being used for everything from specialty software appliances to the ultra-low-cost storage infrastructures supporting most of the 'hyper-scale' cloud data centers. But can x86 server platforms be used to create truly enterprise-grade compute infrastructures and support mission critical applications? In this paper we will discuss what "enterprise-grade" really means and how different x86-based server architectures stack up against the requirements of an enterprise data center.

Enterprise Expectations

The enterprise IT infrastructure has a big job. The compute, storage and network components that make up that infrastructure are responsible for the revenue engines of a corporation, handling tasks from on-line transaction processing to the organization's real-time, 24x7 interface to the world (including customers). They also provide information to support finance, operations and business analytics, the data-driven 'eyes and ears' of the people running the company, as well as regulatory compliance. Given the importance of these functions to the company's prosperity, an enterprise's information systems have to work first time, every time, all the time.

The IT professionals charged with maintaining this compute infrastructure must be confident it will deliver best-in-class performance, resiliency and operating system support. This confidence usually comes from sound design and quality construction, plus effective implementation and support. For this reason, enterprise-grade compute systems have historically been manufactured, installed and supported by tier-1 suppliers. To make sure their systems could handle the job, enterprise IT managers have some serious expectations.

Enterprise-grade Servers

Performance at scale is one of the starting points for true enterprise-grade server systems. These boxes must be able to do the work. And rather than network large numbers of smaller servers together, they must deliver high levels of compute power in a single chassis.

These systems must be built like tanks, able to endure equipment failures, environmental events and human errors. Since a company's very existence may depend on these systems they have to be up and running and providing service with no unplanned downtime.

Reliability, availability and serviceability are 'table stakes' for a compute platform that a corporation relies on to run its business. And, these characteristics are typically built-in, not 'bolted on', rooted in hardware, not software. IT needs the ability to hot-swap essentially any component that could fail anytime, with confidence, and to scale and upgrade capacity and performance non-disruptively.

Flexibility as it scales is another fundamental characteristic of enterprise-grade compute systems. They must support consolidation through server virtualization, providing the ability to run a wide range of applications and workloads and connect to a variety of data center resources. At the same time an enterprise-grade server must be designed for maximum resource utilization and cost containment by balancing workloads and fine tuning capacity planning.

These servers also need advanced management tools that provide critical access while reducing administration time and enterprise-grade implementation and support services from established vendors. The job these compute systems have to do is one of the most important in the company and the list of functional requirements is a long one. Again, these have historically been large UNIX-based systems, often with proprietary architectures and usually big price tags.

IT Costs and x86-based Hardware

Cost pressures are a fact of life in all businesses. Even in times of prosperity, well-run organizations continually look for ways to save money. One concept that has taken hold in the high tech world is the use of standardized server chassis built on the Intel x86 processor, often called "white boxes".

The volumes at which industry-standard processors are manufactured offer some real cost savings, even in servers with much more sophisticated designs than the typical "white box". This presents a cost-saving opportunity for the market, but only if enterprise IT is comfortable buying servers that use these "commodity" processor architectures.

However, there's a common sentiment in corporate data centers that x86 servers aren't really meant for mission-critical applications. This is probably the result of a number of factors. The term "commodity", for example, can imply low quality and diminish the importance of hardware in providing critical characteristics like system reliability and uptime. More importantly, many of the commodity server products on the market actually aren't enterprise-grade, but were originally designed for hyper-scale architectures, where low cost was paramount.

Hyper-scale Companies

These large internet-based businesses and cloud service providers, were instrumental in the emergence of low-cost, standardized servers as the basic IT platform for compute, storage and networking. In order to accommodate the enormous amounts of data they stored, "hyper-scalers" replaced proprietary storage controller hardware with standard x86 servers and then filled these chassis with disk drives, instead of using proprietary drive shelves.

This standardization of server architectures that originated in large cloud and social media companies was an innovation that drove down the cost of servers. But the way these hyper-scale companies used this "commodity" hardware was different as well.

Enterprise Functions from Software

The hyper-scalers' economic model mandated the use of the lowest-cost components available, in many cases consumer-grade components. This fostered a very different paradigm than the one traditional IT was used to. Overall system resiliency, the ability to keep compute services running and data protected, was more a function of software than hardware.

In this model hardware is somewhat expendable, instead of robust and resilient. It is the responsibility of software to spread the workloads and data around in order to maintain availability when nodes failed and to sustain the often more frequent hardware replacement that occurs in the hyper-scale environment. This, in turn, requires a sizable IT staff - one with some advanced skill sets, since at least a portion of most hyper-scale software systems are 'home grown'.

The Hyper-scale Paradigm and Enterprise Applications

While this model of cheap hardware and software-defined architectures has worked for hyper-scalers that need primarily an economical storage solution, one can't simply assume it will work the same way in the enterprise IT organization that needs a more robust compute infrastructure.

Most enterprise IT organizations have a different topology, different personnel and essentially, a different mission. Where hyper-scale companies were fine assuming much of their infrastructure development, and staffed their IT teams accordingly, traditional IT is focused on operations. And, the notion that low priced hardware always delivers lower total cost of operation (TCO) may not be accurate either.

For enterprise IT professionals who have historically taken confidence from robust hardware designs, relying on a complex software layer to ensure overall system uptime is a leap of faith. Moreover, most enterprise datacenters will deploy a few dozen servers, not hundreds of server-based storage and compute nodes that the typical hyper-scale data center does. Even if they had the scale-out topology required, most IT organizations are focused on operations, not in-house, "on-the-fly" system development, as the hyper-scale companies tend to be.

The Reality of Low-cost Hardware

The logic (and popularity) of the hyper-scale model can sometimes lead to the belief that cheap hardware and software-defined architectures will cut costs for a typical IT organization, compared to "traditional" systems whereas in practice they may actually require *more* hardware, when high node counts and the cost of networking infrastructure are considered. In addition, the complex software designed to provide system-level reliability, being less mature, typically requires more hands-on operation and new skills that must be acquired.

For IT organizations not staffed with developers, running this complex software can be an expensive proposition, if it's possible at all. This infrastructure that hyper-scalers designed to save money may end up costing the typical enterprise IT department a lot more.

An Alternative Solution?

Scale-out architectures are good for certain use cases, like large cloud data centers and certain workloads, like web-based services. They're not ideal for all use cases, such as traditional enterprise IT environments running transaction-heavy databases or other mission critical applications.

Rather than turning to disposable hardware and complex software to save costs, could enterprise-grade servers be built with x86 architectures to preserver the high-end enterprise-class features, and still offer a lower cost alternative for users currently operating legacy environments?

Hitachi Compute Blade

Hitachi first entered the server market in the late 1950s with some of the earliest mainframe systems (the "HIPAK mk1"). The company's reputation for quality and reliability in servers continued in Japan, even after the mainframe business declined, and it became best known worldwide as a supplier of enterprise storage systems. The same focus on quality and enterprise features was applied to the first x86 blade servers Hitachi made, starting in 2006, and today Hitachi Compute Blade servers are the foundation of many solutions and appliances sold by Hitachi Data Systems.

Performance at Scale

In terms of power and performance, Hitachi Compute Blade (CB) servers can support up to 504 processor cores based on the Intel® Xeon® processor E5-2600 v3 product family in fourteen standard server blades or 288 cores in eight full-width scalable blades using the Intel® Xeon® processor E7-8800 v2 product family, or a combination. The larger 12U chassis (CB 2500) can support up to two SMP blade configurations with up to 12TB of system memory each (using 64GB DIMMs) and an I/O subsystem with up to 28 PCIe Interface cards and 32 embedded switch ports.

High Availability

To ensure the highest levels of reliability and uptime, all critical components in Hitachi Compute Blade servers are redundant and hot-swappable, including switch, management and power supply modules. Also, standard is Hitachi's "N+M Cold Standby", a feature that enables cascading failovers, providing "M" backup blades for every "N" active server blades. This allows one spare server blade to provide automated failover for a pool of active blades, even those running different OSs.

Flexibility in CB 500 and CB 2500 servers is provided by an architecture that runs 32-bit and 64-bit applications on Microsoft® Windows® or Linux, supporting multiple blade configurations to enable scaling without clustering. I/O flexibility, another important characteristic, is provided by Hitachi's hybrid I/O feature.

Hybrid I/O

Hitachi Compute Blade servers offer two modes for connecting server compute blades to the external environment, through direct PCIe I/O interface cards or through embedded switch modules. CB 2500 provides up to 28 PCIe slots in each chassis, connected through mezzanine I/O expansion cards, and two switch modules per chassis that provide a shared switch network across all blades. (The smaller CB 500 does not have PCIe slots in its chassis but can support PCIe cards via expansion blades).

Each shared switch module can be independently configured as a 1/10GbE managed, layer-3 switch providing up to 28 internal ports, plus ten copper and four optical ports for external connection. Or, a 10/40GbE Data Center Bridging (DCB) embedded switch can be chosen, with 42 internal and 14 external 10GbE ports, plus 4 optional 40GbE ports. Each server blade can also be connected to redundant switch modules for high availability and fault tolerance.

Logical Partitioning

Logical partitioning (LPAR) allows a physical server's resources (CPU, memory and IO) to be divided into "partitions" which act as separate servers, each isolated from the others since hardware is not shared or virtualized. Another "big box" feature from Hitachi's mainframe roots, this technology is built into every blade's firmware, so it's a standard feature, and there is no virtualization software to buy or install.

LPARs offer a significant performance advantage over traditional software-based virtualization platforms since (like in the mainframe world) applications and OS requests can execute directly on the processor, without host intervention. And, LPARs leverage Intel Virtualization Technology (VT) to optimize CPU performance.

Security is also provided at the hardware level, as each logical partition is isolated from the others, with its own dedicated physical memory. Storage is accessed directly, an on any physical server so LUNs and can be configured with different security settings. Finally, LPARs are managed through a interface that resides on a separate physical network from data traffic, further isolating management from user traffic and adding another layer of security.

Logical partitioning makes it easy to create up to sixty independent hardware resource pools per blade on CB 2500, each appearing to its operating system as a separate physical server. Coupled with Hitachi hybrid I/O features, LPARs support server consolidation without buying or administering a hypervisor and enable multi-tenancy and quality of service by enhancing reliability and stable performance.

Efficiency and Management

Hitachi Compute Blade 2500 includes sophisticated energy consumption management including a power capping function that controls the frequency of CPU steps based on processor service rates. This feature can reduce consumption of server blades by roughly 16% and by allowing a power cap to be enforced, prevent over-provisioning, an important feature in environments where power is limited. The system can also provide sophisticated monitoring and historical reporting of power consumption data to optimize data center resources.

Enterprise-grade servers also need to support enterprise-grade management. Hitachi Compute Blade servers are controlled with the Hitachi Compute Systems Manager, a web-based utility that handles deployment, configuration, maintenance and performance tuning of thousands of server blades. This secure management suite allows seamless integration with Hitachi Command Suite, providing a single interface for server and storage management across the data center.

Summary

Users of legacy enterprise systems which support mission critical applications are under increasing pressure to move to systems and architectures based on low cost high-performance Intel x86 architecture processors to save capital and operational costs.

Much of this pressure is coming from C-Suite executives reacting to the recent high profile industry focus on a number of hyper-scale data centers which have pioneered the use of scale-out architectures using applications that are able to leverage low-cost, x86-based server hardware based on “commodity” components.

Given the universal imperative to reduce IT cost, any new paradigm that claims to deliver big savings is appealing to many companies, and has gained notoriety and executive interest in companies far outside of the hyper-scale world.

However, this often turns out to be a false hope for enterprise IT organizations since many mission-critical applications in traditional, large-company data centers simply aren’t suited for deployment in these types of scale-out infrastructures. In addition, the expected cost saving may be illusory, since, although the individual compute nodes are relatively inexpensive, many more nodes may be required for high availability. Plus, software and networking complexity increases operational costs, not to mention staff retraining and hiring costs.

Typical enterprise IT users are accustomed to server systems that are much more reliable and robust at the hardware level and don't require intricate software to make it all work. They're also highly skilled in running these mature, highly reliable solutions for their critical storage and compute systems and have come to depend on many advanced features, such as logical partitioning for efficiency.

But while the hyper-scale model may not be right for enterprise IT, it may still be possible to leverage x86-based server platforms to reduce cost and remove reliance on expensive proprietary architectures. With the right reliability and robustness features designed in and the right tier-1 manufacturer behind them, industry-standard platforms can indeed be used to create a true enterprise-grade compute system, at a lower cost than legacy hardware.

About Evaluator Group

*Evaluator Group Inc. is dedicated to helping **IT professionals** and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver **in-depth, unbiased analysis** on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and **education**. **www.evaluatorgroup.com** Follow us on Twitter @evaluator_group*