Data Center Management: The Key Ingredient for Reducing Server Power While Increasing Data Center Capacity

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for Cisco

June 2010

IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING

Table of Contents

Introduction

Enterprise Management Associates (EMA) devotes considerable time to discussions of current and desired IT management technologies with CIOs and other IT operations personnel. Many consider an ideal management stack to include the following features and capabilities:

- Top-down modeling of business applications and dependencies
- Holistic, policy-based, vendor-agnostic management of compute, storage, and network elements
- Virtualization-aware management, including network, compute, storage and virtual machine resources
- Definition of business- and element-level policies, including application prioritization, required performance levels, and desired optimal states
- "Stateless" management capabilities that abstract business applications from underlying hardware
- Single management console that incorporates all management capabilities while requiring a single multi-purpose agent on managed devices
- Orchestration capabilities that fluidly combine all management capabilities, increasing data center efficiencies while reducing overhead, saving money and resources

This wish list is a tall order to fill, but certain vendors are making excellent progress toward fulfilling these requirements. Effective IT management provides many or all of these benefits:

- Maximization of data center resource cost/benefit ratios
- Increased application availability and performance
- Decreased costs from automation of routine IT tasks and decreased hardware requirements through reduced spare requirements
- Increased compliance with corporate, IT and governmental regulations and policies
- Decreased power requirements; increased data center scalability and density
- Rapid adaptability to changing business demands, including provisioning of additional resources as needed to satisfy "burst" demands

This paper will focus on the application and benefits derived from this new breed of holistic IT management technologies, with an emphasis on increased efficiencies, reduced costs, and the environmental benefits that result from them. It then provides an in-depth comparison of the management and server offerings from two industry heavyweights, Cisco and HP, and concludes with an EMA analysis of these trends.

Business Drivers toward a New Management Paradigm

The drive to increase data center efficiencies has been unrelenting for a decade or more, and IT management is a key focus for many organizations seeking to reduce costs and minimize the environmental impact from IT operations. This trend shows no signs of slowing down as organizations grow ever more dependent on IT every day. Pressure to decrease power consumption and increase data center density are seemingly at odds with increasing demand for compute, storage and networking capabilities. At the center of the storm is the increase in server deployments, which EMA estimates will continue to increase at a minimum of 10% CAGR for the next few years, incrementally accelerating over time.

According to the U.S. Environmental Protection Agency (EPA), electricity consumed by servers in U.S. data centers in 2006 represented about 1.5% of national electricity use, or around 61 billion kilowatt-hours (kWh)—equivalent to approximately 5.8 million average U.S. households[1] and resulting in energy costs of approximately $5.4 billion while emitting 36 million tons of carbon dioxide into the atmosphere.[2]

Power consumption has emerged as a primary focus of many organizations that have undertaken "Green IT" initiatives. While reducing an organization's contributions to greenhouse gas emissions is compelling and necessary from a corporate responsibility standpoint, an even larger impact on the bottom line is created when a data center runs out of power. While every watt is important, the "last watt +1," which occurs when a data center runs out of power capacity and necessitates the build-out or relocation of the data center, costs tens or hundreds of millions of dollars. This scenario creates a massive stepping function and executive review, requiring data center managers to show that they have squeezed every watt out of the company's existing data center investment. Data center scalability has therefore emerged as a key driver in the "greening" of the data center. Organizations are discovering that data centers designed to provide the highest level of performance per square meter at the lowest power levels, can derive up to 30% additional capacity from existing facilities. They also reap the PR and cost reduction benefits of a green data center in the process. Proactive planning today can eliminate or delay the necessity of expanding or relocating the data centers of tomorrow, saving millions of dollars in the process.

> *"Increasing power density can lead to a situation in which companies are forced to build new data centers not because they are running out of floor space but because they need power and cooling beyond what can be provided in their existing data centers. This situation has driven much of the recent interest in energy-efficiency improvements for data centers. If the power consumed (and resulting heat generated) in data centers can be reduced through energy-efficiency measures, the existing infrastructure can continue to meet cooling and power needs, and costly investments in new data centers can be deferred."*
>
> Report to Congress on Server and Data Center Energy Efficiency—Public Law 109-431, United States Environmental Protection Agency ENERGY STAR PROGRAM, August 2, 2007

There are two fundamental ways to reduce data center power requirements: 1) purchase energy efficient equipment; and/or 2) employ "power-friendly" IT management technologies. The latter can reduce power consumption on a much larger scale than simple hardware replacements. This is true only if the data center management stack is not so complex that it causes management overhead to spiral out of control, increasing expenditures more (and faster) than the power savings realized.

The Automated, Adaptive Data Center

Vendors created great hyperbole over the past decade, hailing the advent of a fully automated, adaptive, on-demand data center. In this brave new world, data center elements are grouped into generic resource pools while an all-knowing, all-seeing management genie, sometimes known as an orchestrator, manages the entire data center based on business policies. The "orchestrator" ensures that all loads are matched with the appropriate amount of with resources, delivering maximum ROI and business performance at the lowest cost.

[1] "Report to Congress on Server and Data Center Energy Efficiency—Public Law 109-431", United States Environmental Protection Agency ENERGY STAR PROGRAM, August 2, 2007

[2] UK National Energy Foundation Energy to Carbon Converter: http://www.nef.org.uk/greencompany/co2calculator.htm

Unfortunately, the grand vision outlined above never progressed beyond the vaporware stage due to a lack of control at the element level. Passive, unintelligent hardware controlled by an elegant orchestrator is akin to a grade school band led by a great director—except that dumb hardware never learns or improves! The primary gap between vision and reality has been hardware that lacked the ability to perform autonomous management actions based on business policies set by a centralized management engine. It is no longer acceptable to have a centralized orchestrator that **attempts** to manage dumb elements directly—the elements must themselves be intelligent and self-managing, dynamically adapting to changes without requiring the overhead of an orchestrator.

The advances of virtualization technologies is a key technical enabler toward these goals, allowing intelligent network, storage and compute resources to be pooled together and managed as a group. Provisioning technologies provide a level of automation, quickly assigning virtual resources to business applications when needed.

The biggest virtualization development to date is the Virtual Machine, or VM. VMs encapsulate an entire business application environment inside of a virtual, portable "bubble" that can be easily moved from server to server (or even into the Cloud) as needed. Combine a VM with virtual compute, network, and storage capabilities and an automated provisioning system, and truly portable business services started to become real (though still with strategic gaps). The VM is becoming the "atomic unit" of the data center, around which virtual resources are deployed as needed. It is important to realize, however, that VMs are not suitable for all purposes, and monolithic deployments like Oracle, with large memory footprints, will continue to exist for many years to come. Technology vendors need to be able to support fully virtualized, partially virtualized, and non-virtualized environments.

Until recently, flaws in virtualization management technologies prevented the fully automated data center vision from becoming a reality. First, many so-called virtualized resources still require a large amount of manual tweaking during the provisioning process in order to function correctly. For example, many "virtual" networks still require specific configuration settings be applied for certain applications or VMs, for certain types of "virtual" hardware. This is because some vendors have not delivered a hardware manager that fully completed the work required to package all of an application's configuration items, policies and dependencies into a fully portable, policy-based package. Without this fully portable package, there is still a manual intervention requirement to make all of the elements work together every time an application is moved or additional resources are provisioned. This capability is required across all server/network/storage environments, whether virtualized or not.

Once a stable hardware management base is attained (which includes deployment of intelligent, self-managing, policy-aware elements), an overarching, policy-based manager (the genie, or director) can be implemented. The bi-directional integration of the manager with the director is a critical component, and maximum flexibility/scalability by the hardware manager is key.

The director's job is to serve as a centralized repository for business and IT policies, ensuring that business-critical applications receive the resources they require in order to enable the business. IT and the business work together to define policies that govern all aspects of a business services, ranging from the type and quantity of resources required, to rules controlling the provisioning of extra capacity to satisfy bursts in business demand, to policies ensuring security.

The director, thanks to an automated, policy based hardware manager, can even mandate movement of VMs and their associated applications from server to server *in situ (via the hardware manager)*—without interrupting running business processes and security protocols and policies. This process of continuous resource optimization yields the highest levels of data center performance while maximizing resource utilization and security.

Maximizing Data Center Performance While Reducing Cost through Automation

Historically, lack of management and virtualization maturity forced IT organizations to over-provision resources to ensure that adequate spare capacity existed to satisfy cyclical and unforeseen demand spikes, provide disaster recovery and meet failover requirements. Common practices included allocating an extra backup server for every critical production server in the data center, some even going so far as to reserve two extra servers in order to ensure triple redundancy. Bare-metal server provisioning was impractical, requiring too much time and manual intervention to make it viable, which required fully provisioned backup servers to be running at all times. These practices are an incredible waste of human and system resources in addition to being very inefficient from a footprint and power perspective. Even though backup servers consume a fraction of the power, hundreds or thousands of idle servers collectively require a large amount of power and significant amounts of non-production oriented data center space.

In an environment with a highly effective hardware manager as described above, a highly automated data center (independent of virtualization) pools spare servers in a bare metal state, making them holistically available for provisioning to virtually any task within minutes. A large majority of spares are powered down, since they can be activated and provisioned from bare-metal state in minutes when needed. This reduces the total number of backup servers required since servers are now consumed on-demand, quickly allocated and de-allocated as needed, shared by multiple business services. This saves a large amount of the power formerly consumed by active backup servers. Bare-metal provisioners must work equally well for VMs as well as "classic" (monolithic or non-virtualized) server data center architectures.

The advent of blade computing several years ago is now reaching widespread acceptance, particularly due to significant advantages from power, efficiency and management perspectives. Blades pack a lot of compute power inside of a small, modular physical space, providing a generic compute resource that can be quickly and easily "hot swapped" if it fails. Couple these hardware advantages with the new breed of automated management technology that allows rapid bare-metal provisioning, policy-based virtualization and an orchestrator that dynamically and automatically moves workloads when a blade fails, or if business demand changes, and blade computing can provide a very strong business case.

Data Center Power Considerations

As mentioned previously, power capacity is a key data center and cost efficiency driver. In addition to the core power requirements for data center elements, every dollar spent powering the data center also incurs a dollar cost to power, heat and cool the hardware. As data center densities continue to increase, power constraints often limit scalability long before physical space runs out. Maximizing power utilization efficiencies not only increases data center capabilities, but it also decreases power/performance cost and environmental impact. As discussed above, the real business benefit is increased compute carrying capacity of existing data centers, deferring or completely avoiding additional data center build-outs.

To help understand data center power efficiency, the Power Usage Effectiveness (PUE) metric gauges the ratio of total power used by the data center against the power delivered to computing equipment. In other words, PUE evaluates the overall efficiencies of the data center from a power distribution and utilization perspective. A PUE of 1.0, for example, is a perfect score, indicating that each watt consumed by the data center translates into a watt delivered to the equipment (which is obviously unattainable). Another consideration is the newly unveiled Energy Star rating for data centers, introduced by the U.S. EPA in June, 2010. The new rating, which is entirely based on PUE, is awarded to data centers in the top 25% of their peers for PUE, and is audited by a third party. Organizations that may wish to attain the new Energy Star rating will definitely want to do everything possible to improve their PUE rating.

As one might expect, factors such as heating and cooling inefficiencies, as well as switches and vampiric loss to equipment like uninterruptible power supplies (UPS) and power distribution units (PDUs) directly affect the PUE. Today's state-of-the-art data centers may attain PUEs of 1.8 or lower, but most legacy data centers report PUEs between 2.0 and 2.4 (or even higher). PUE is important because it illustrates the hidden cost of data center power consumption. It can also incent departments to conduct thermal, cooling and power delivery studies, identifying problem areas and pinpointing candidates for increased power efficiencies.

Servers comprise a major percentage of overall data center power usage, and studies have shown that a watt generated at a power plant diminishes to 0.30 watts by the time it arrives at the data center, dropping to only 0.17 watt by the time a server translates it into business value. Reversing the math, every watt saved at the server saves ~1.8 watts of data center consumption[3] and obviating the requirement to generate ~5.9 watts in the first place.[4] Maximizing performance per watt should be a key goal.

One factor to consider when comparing server vendors is that today, virtually all major manufacturers utilize the same Intel Xeon processors that leverage the same Nehalem microarchitecture. This simplifies power comparisons between vendors, as servers that use the same processor type and speed levels the playing field, moving the conversation to other differentiators, including system architecture, other hardware design considerations, and management efficiencies.

In the next section, we perform a detailed comparison of blade servers from HP and Cisco, both of whom use Intel Xeon/Nehalem processors. First, the management capabilities of both vendors are examined, and then a quantitative analysis of the power requirements for both vendors is presented.

Cisco and HP: Comparing Blade Server Power and Management Capabilities

It is interesting to note that the price of server hardware, as a percentage of overall TCO (total cost of ownership), has decreased steadily over the past four years. Server hardware acquisition cost as a percentage of total TCO has been dropping as technology has advanced, and is now approaching only 20% of total three-year TCO.

Moreover, processor price as a function of TCO is now relatively constant across all vendors, with the processor price itself comprising only a very small, insignificant percentage of the total solution cost. With CPU costs basically equal and overall hardware costs steadily declining, the major TCO differ-

[3] 0.30 watts / 0.17 watts = 1.76 watts
[4] 1.00 watt / 0.17 watts = 5.88 watts

ences between vendors boils down to overall server performance, other hardware capabilities, power consumption, and the strength and scalability of the management stacks, particularly at scale.

HP has a long history of providing data center hardware and software solutions. HP's blade server lineup dates back to RLX, which patented and shipped the first blade server in 2001. After HP acquired RLX in 2005, the company leveraged RLX blade technology to produce HP's ProLiant blade server lineup. From a management perspective, the HP stack is comprised of a set of agent-based technologies either acquired from other sources or developed internally, all of which have been more or less integrated.

Cisco is a new entrant to the blade server market, first introducing the UCS blade server line in March 2009, although the company has had a dominant position in the data center networking market for years. While the UCS servers have been on the market only about a year, Cisco has already caused significant market disruption by introducing innovative blade server technologies that provide high performance with the advantage of organically developed management capabilities that are, by and large, built-in to the UCS hardware and firmware with a single interface.

CPU and Memory

As one might expect, CPU comparisons between the two vendors are virtually identical. Both vendors utilize the same Intel Xeon 5500 and 5600 series processors, both offer two-socket blades, and both currently offer (Cisco) or plan to offer (HP) four-socket Intel Nehalem EX-based blades. Both vendors use the same type of DDR3 DIMM memory, supporting 2GB, 4GB and 8GB modules.

There is one significant difference between Cisco and HP in terms of memory architecture, however. HP's conventional memory scheme supports either 12 or 18 DIMM slots, which translates to a maximum of 192GB of RAM per blade. Cisco's "Extended Memory" architecture, developed in partnership with Intel, supports up to 48 DIMMs for a maximum of 384GB per blade. In addition to providing large amounts of memory for RAM-intensive applications such as VMs, Cisco's memory architecture also enables the use of faster 1066 MHz chips across all 48 DIMMs, whereas HP servers currently drop to 800 MHz due to memory architecture.

Networking

When comparing blade vendors from a networking perspective, it is important to consider the overall supporting hardware required. The days of discrete networks for management, compute, and storage are fading fast thanks to the emergence of Fibre Channel over Ethernet (FCoE), which allows the convergence of all network protocols across a single medium.

HP's blade server architecture, while making some steps towards convergence, is still firmly steeped in the old world. HP requires one or two separate network interface cards (NICs) per blade (one Ethernet and one Fibre Channel), a minimum of 4 "in-chassis" switches, plus associated cabling and downstream switches, to support each blade's connectivity requirements. In addition to a large amount of supporting hardware and switches, this architecture requires a large quantity of network cables—up to ten cables per chassis (four Ethernet, four Fibre Channel, and two management ports per chassis) to reach a network aggregation density similar to Cisco UCS. HP's new converged networking offering is delivered and managed "old world," without true simultaneous convergence of network protocols with server, network and storage management.

Cisco brings converged Ethernet and FCoE directly to the blade chassis—a key advantage. This requires only four cables per chassis to support the same number of blades as HP while eliminating a large percentage of the downstream switching and physical cabling required. Converged fabrics carry the traffic for all networks, including compute, storage and management, through a 10 Gbps Ethernet connection.

Another major advance with Cisco UCS blades is their use of virtual interface card (VIC) architecture, allowing dynamic I/O configuration based on policies defined in UCS Manager. This allows the creation of up to 128 PCIe devices, configured as a vNIC (virtual NIC) or vHBA (virtual HBA). This is a significant concept, as it completely abstracts network configuration from the hardware, allowing it to travel with the Service Profile—something that virtually no vendor other than Cisco can do. This substantially reduces TCO and yields added benefits from increased network throughput and reliability, and major decrease in required management intervention when migrating VMs to other hosts; not to mention power savings from the reduced number of network components required.

Cisco UCS Manager is not only embedded in the UCS 6100 Series Fabric Interconnects (FIs), but also fully enabled in each blade with a dynamic dedicated private bandwidth for "Out Of Band" management connectivity, totally separate from production network traffic. The FIs scale across multiple blade server chassis in a single domain (potentially up to 40 chassis per FI). This eliminates need for expensive switches in every chassis, duplicate management servers, and database servers—a major savings compared with domain proliferation that can happen with HP at scale. UCS Manager therefore can manage up to 320 servers in a single domain structure regardless and independent of the total VM population.

Power Requirements

From a power perspective, there are a number of items to consider when comparing blade vendors. In addition to the raw amount of power consumed by servers and other supporting hardware (such as networking as outlined above), also consider the extra power required by management servers and consoles. This is particularly interesting at scale, where power required by management and supporting components must also scale. Comparing HP and Cisco, Cisco again has an advantage in this area from a number of perspectives. First, UCS requires only one management server, compared with up to four required by HP. Also, consider that each management server also requires an associated database server, driving power requirements even higher, not to mention the management burden required to maintain the servers and extra licensing costs.

From an architectural and management perspective, Cisco UCS compares very favorably against HP. UCS Manager can power up a blade and provision it from bare-metal state within a few minutes using templates to define firmware on service profiles. This allows spare servers to be consolidated down to a fraction of the quantity required by other vendors—saving a lot of power while reducing overall hardware requirements and additional capital cost.

Blade Management Capabilities

Hardware is important when comparing blade server vendors, but management overhead plays an even larger role in TCO than the hardware. Incumbent blade server vendors, HP included, base their blade server management approach on existing management technologies, which are comprised of a number of independently developed management technologies, integrated and extended to support new technologies like blade servers and controlled by a top-down orchestration engine. The result is an unfortunate mish-mash of somewhat integrated components that typically require multiple agents and consoles or applets, not to mention a lot of training in order for staff to understand all of the moving parts. In addition, incumbent vendors also tend to price their management components separately, further adding to the solution TCO.

Cisco began with a blank slate, taking advantage of their deep hardware management heritage and strong partnerships with Intel, EMC and VMware to build management components directly into the blade, chassis and networking hardware. Advantages of this approach include a single management agent on each blade, coupled with a true single management console. HP, along with most other blade vendors, requires a dedicated management blade for each chassis and multiple software management agents on each blade. This also requires multiple consoles or a single console view with multiple applets to access the various consoles.

Cisco's architecture does not consume a blade for management, allowing one extra production server blade per 16 blades. They use a single, lean UCS agent, increasing overall blade capacity. Cisco includes the UCS management stack in the base price of its solutions with the exception of BMC BladeLogic, which is an OEM that provides bare-metal provisioning for Cisco and other blade vendor hardware. Cisco integrates UCS Manager with many other management software companies' products as well—not just BladeLogic. A key advantage of the Cisco architecture is that logic is pushed to the level of the action—when a blade is defined, it is given a set of policies that enable it to make intelligent, policy-based management decisions without requiring the latency and overhead of an orchestrator.

Digging into HP's blade management stack, one discovers a bewildering "alphabet soup" of components. HP requires the Virtual Connect Enterprise Manager (VCEM), Insight Control Environment (ICE) bundle, HP Systems Insight Manager (SIM), HP Integrated Lights-out (iLO) Web interface, HP On-board Administrator, and HP Insight Dynamics VSE Suite, etc., which is comprised of up to 11 agents and up to four different management consoles. This requires a staggering amount of memory and processor overhead on every server, not to mention the administrative overhead required to maintain all of the management components, as the numerous interdependencies and complexities of the HP schema can cause failures if firmware versions fall out of synchronization.

Cisco UCS, in contrast to HP, requires a single Cisco UCS Manager console and a single UCS management agent per blade. Cisco also designed UCS Manager to work with more than just UCS blades. Cisco is adding support for Cisco rack mount servers to UCS Manager, which can also manage UCS network switches and Fabric Interconnect devices. When you consider the amount of additional installation, staff training and management overhead required for the HP blade solution, it is easy to see how Cisco comes out ahead from this perspective.

The one blade management technology that Cisco did not build in-house was bare-metal provisioning, which they chose to provide via tight integration with BMC BladeLogic, a highly regarded, agnostic, multi-vendor blade provisioning solution. Cisco fully integrated UCS with BladeLogic using open interfaces, performing all management tasks from the same UCS Manager console. In addition to simplifying provisioning management, BladeLogic integration also allows UCS to manage non-Cisco blades. Cisco UCS Manager is also fully integrated with other management software suppliers as well, such as CA, Symantec/Altiris, Microsoft, EMC , VMware and even HP and IBM. Interesting enough, Cisco provides full access to developers and uses standard XML API, so custom tools can be integrated with UCS Manager via a Cisco API SDK.

Distributed, Policy-based Management

As indicated in the beginning of this paper, a key aspect of today's data center includes moving to a service-oriented management model via a centralized policy management engine, coupled with intelligent self-managing hardware elements. The management engine pools data center resources, interfaces with lower-level managers (such as provisioners and element managers), and ties it all together with policy-based management capabilities. Data center assets are combined into pools, and policies define management "units of work," bundling all components required to deliver an IT service into an easily managed profile that works for both virtual and non-virtual architectures.

Cisco took the lead in this area by creating UCS "Service Profiles" as a core part of UCS Manager. A UCS Service Profile is a policy-based grouping of related hardware, including compute, storage and networking profiles, plus VMs, all in a self-contained package that travels with the business service as UCS Manager moves loads around the data center to maximize resource utilization and performance. In a VM environment, Cisco has a unique ability to deliver network QoS and security policies that adhere to, and move with, every virtual machine using their VIC (Virtual Infrastructure Card). This is inherent in their VIC and delivered by the hardware, removing this workload from the processor, further increasing performance and therefore performance/watt.

This is a key UCS advantage, as it allows complete abstraction of a service (work) from the hardware on which it runs, enabling fully automated, policy-based provisioning and de-provisioning. Service Profiles significantly reduce the excess capacity burden described above by enabling pooled burst and high availability (HA) capacity to be rapidly provisioned based on business need. It can also assist with dynamically balancing data center workload, ensuring maximum performance from each hardware component.

In this EMA head-to-head comparison, Cisco UCS wins the architecture and management comparison hands-down.

Cisco and HP: A Quantitative Power Consumption Analysis

It is always useful to have deep quantitative data when comparing two vendors, particularly from a power perspective. As indicated earlier in the paper, power savings at the server level translate not only into direct OPEX savings, but also produce even greater benefits in terms of reduction of power plant demand and impact on the environment.

As indicated earlier, CPU power requirements are equal between vendors, since all blade server vendors are using the same Intel chips. This is helpful, since it removes one of the largest power consumers as a variable when comparing vendors. The differences in power consumption therefore boil down to items like relative airflow (and the fans required for cooling), network hardware, and so on.

Cisco commissioned a quantitative comparison of UCS and HP blade servers in March 2010, conducted by Principled Technologies (PT), [5] based in Durham, North Carolina. PT is a highly respected testing laboratory used by many industry leaders, including Cisco, Dell, IBM, Intel and many others.

PT conducted an exhaustive power comparison between Cisco UCS B200 M1 blade servers in a Cisco UCS 5108 Blade Server Chassis and HP ProLiant BL460c G6 servers in an HP BladeSystem c7000 Enclosure, and the entire test is available on the PT website.[6] The PT test measured overall power

[5] http://www.principledtech.com
[6] http://www.principledtechnologies.com/clients/reports/Cisco/UCSPower0310.pdf

consumption of the two blade solutions, both under load as well as at idle, using industry-standard tests including SPECjjb2005[7] and Prime95[8]. SPECjjb2005 is a Java server benchmark used to evaluate the performance of server-side Java by emulating a three-tier client/server system, measuring the performance of the CPU, cache, and memory. Prime95 is a well-known computer "torture test" intended to run a computer at top compute capacity by repeatedly calculating prime numbers for as long as needed.

PT tested both power usage and performance per watt for both vendors. While reporting all of the test results is beyond the scope of this paper, PT did report the following key findings:

1. The Cisco blade solution achieved up to 10.7% more SPECjjb2005 bops/Watt than the HP blade solution;
2. Cisco blades used 10.2% less power per blade than the HP blade solution when running the Prime95 torture tests;
3. The Cisco blade solution used 3.3% less power per blade when idle, compared with the HP blade solution.

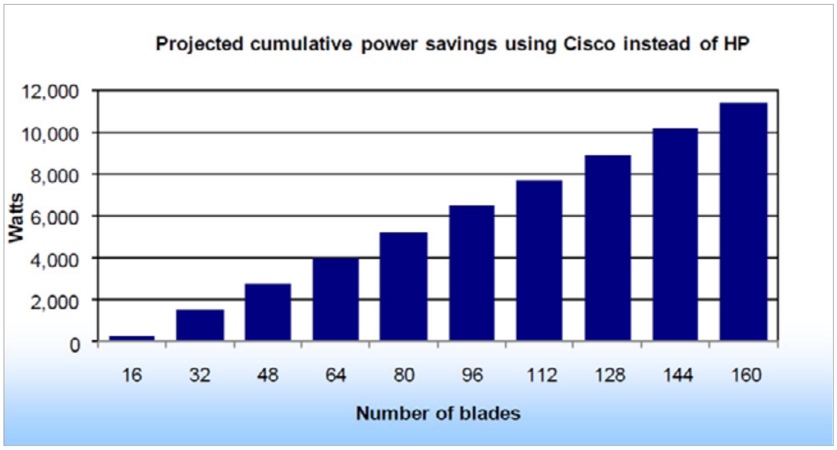

Figure 1: Project Power Savings for the Cisco UCS Blade Solution Compared with HP Blades (source: Principled Technologies)

These findings are significant. As an example, consider the Prime95 test results. The Cisco blades used 38.6 watts less, per blade, than HP. Extrapolating these savings out to a 500 server data center, these savings add up to 169,165 kilowatt hours per year ((38.6 watts * 500 servers * 8766 hours/year) / 1000). Also consider that for every watt saved at the server, another watt in heating and cooling is saved, and the total becomes 338,330 kWh.

Using the January 2010 commercial electricity rate of 9.58 cents per kWh calculated by the United States Department of Energy[9], this calculates out to $32,412 in annual energy savings—and that is only comparing blades. Keep in mind that these figures are for fully loaded blades, so these figures are clearly "best case" calculations. Also consider the additional savings from the significantly reduced network hardware with Cisco due to the utilization of converged I/O architecture and the savings can even be higher.

[7] http://www.spec.org/jbb2005/
[8] http://www.mersenne.org/
[9] http://www.eia.doe.gov/electricity/epm/table5_6_a.html

Conclusion

Cisco has created an impressive "next generation" data center architecture with UCS. Combining a dramatically simplified, intelligent hardware solution that abstracts servers, network and storage from the applications that depend on it with a centralized, policy-based engine drastically reduces management complexity.

From Enterprise Management Associates' perspective, Cisco wins the comparison with HP hands-down, for a number of reasons:

- Cisco UCS blades use ~10% less power than HP under load, and ~3% less power than HP when idle, potentially saving tens of thousands of dollars per year in energy costs.
- The UCS architecture allows many server spares to be turned off when they are not used, saving 100% more energy than HP, and UCS allows fewer total server spares to be used than other vendors, since blades can be quickly and automatically provisioned from a bare-metal state, regardless of the application. This allows many more production servers per square meter of data center space than comparable blade servers.
- Bringing converged networking (including FCoE) directly to the blade chassis is a stroke of genius. The combination of eliminating the requirement for up to eight network cables per chassis, coupled with embedding the UCS manager in the chassis, blade and networking hardware, drastically reduces the TCO, power requirements and management complexity of the UCS solution.
- The UCS Manager ties all of this together from a single console, requiring only a single lightweight agent on each blade, compared with up to 11 agents and four consoles in the HP solution. Power savings and extra processing capacity derived from the elimination of management blades, extra management consoles and other supporting servers is significant.
- UCS hardware, which includes the ability to self-manage based on policies set by UCS Manager, allows management to be moved to the element level. This increases data center adaptability while decreasing management overhead and latency.

EMA believes that organizations that are concerned with reducing environmental impact through decreased energy consumption and saving money on energy costs, reducing management overhead and increasing overall service levels, should take a hard look at Cisco's Unified Computing System. While UCS represents evolutionary advances in hardware technologies, combining UCS hardware with the innovative UCS management stack shows great promise that could very well make a revolutionary impact on the data centers of today and in the future.

About Enterprise Management Associates, Inc.

Corporate Headquarters:
5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

2103.062810