IDC
Analyze the Future

WHITE PAPER

Forecasting Total Cost of Ownership for Initial Deployments of Server Blades

Sponsored by: HP

Kelly Quinn
Jed Scaramella
Daniel Fleischer
John Humphreys

June 2006

Executive Summary

For organizations that are weighing the choice between deploying a large number of servers, total cost of ownership (TCO) analyses favor blade server systems over rack-optimized server systems for reductions in both capital and operating expenses. Blade server systems are all about exploiting the economies of scale when deploying servers in volume. Analyzing data provided by HP, IDC validated the nature of the expected cost savings:

- From a capital expenditure point of view, hundreds of independent rack servers require thousands of repetitive parts. This need is apparent when integrating servers with local area networks (LANs) and particularly with storage area networks (SANs). Without a shared backplane, each standalone rack server requires adapters, cables, and supporting switches, and there is no economy of scale as the number of servers grows. The blade server system, with its integrated backplane, consolidates LAN and SAN access and thus requires far fewer interconnecting cables and devices.

- From an operating expenditure point of view, further economies of scale emerge. Reductions in power, cooling, and datacenter space accumulate to lower facilities cost by an estimated 25%. Most importantly, the cost of initial system deployment drops dramatically as the software combines with the blade server system to automate system installation tasks. Moreover, consolidated networking for LANs and SANs reduces the number of cable pulls. Thus, installation tasks that once took a few hours per server can be accomplished in a few minutes per blade. In addition, costly tasks such as running hundreds of cables disappear entirely.

In this context, IDC believes that the HP BladeSystem c-Class is an important new platform for IT organizations to evaluate when weighing the relative TCO associated with building out server and infrastructure resources. With next-generation HP Thermal Logic technology to reduce energy consumption and improve cooling efficiency, the HP BladeSystem c-Class helps to alleviate power consumption limitations that are emerging in some datacenters. HP Virtual Connect architecture streamlines LAN and SAN connectivity and expedites the provisioning process. Built-in HP Insight Control management improves the ability of system administrators to deploy systems rapidly, catalog blade system resources, provision infrastructure, and monitor overall system health.

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

Situation Overview

In today's enterprise, control and reduction of expenses are top-of-mind issues. Both capital costs and operational costs are fundamental to any discussion about running and growing a business. Such costs extend to all lines of business within an organization as well as to the IT group.

However, when it comes to IT, identifying expenses may not always be as straightforward as with other business units. Primary expenses, such as the cost of hardware or software licenses, are relatively easy to see. However, secondary expenses, such as the hardware cabling costs, facilities charges, and personnel hours needed to configure the systems, are often not accounted for with precision. Tertiary expenses, which include the cost for service, maintenance, and retirement of old servers, are rarely allocated when estimating the TCO for a new IT investment.

The primary TCO metrics are relatively obvious and intuitive. They include up-front capital expenses that are accounted for when a company does a workup of the costs of deploying new IT hardware and services. Secondary TCO measures, which are primarily operating expenses, will be less apparent and will often require more research to quantify accurately. TCO analysis requires digging in to what system administrators do and how the datacenter is architected. Tertiary TCO expenses must be estimated by attempting to anticipate what future demands the installed hardware will place on the IT budget. These costs are all a part of the overall IT spend. The precise allocation of cost to inform investment decisions is the goal of TCO analysis.

Typically, IT customers will track the costs of infrastructure acquisition and operation over time so that they can make important decisions about capacity planning and computer acquisitions going forward. IT planners primarily use one-year, three-year, and five-year time horizons when conducting capacity planning for their computing systems.

How TCO Is Useful for IT Buildout

TCO analysis can be highly useful in relation to IT buildout in several ways. First and foremost, it can help customers identify the true cost of architectures for their datacenters before they invest in the infrastructure components. Second, it can assist customers in setting realistic expectations for how much the complete buildout of an IT solution will cost, in terms of hard costs, such as hardware, software, and services, as well as soft costs, such as personnel time, downtime, and maintenance.

Today, IDC estimates that the hardware acquisition costs associated with the datacenter contribute approximately 20% of TCO, which leaves 80% of the total cost for the setup, operation, and support of the hardware. This ratio of capital to operational expenses has grown tremendously as the amount of IT infrastructure needed to support a business has grown. It is not uncommon to find businesses having to support thousands of servers across an increasingly global organization. Because IT departments are growing in size and importance in today's businesses, a proper TCO analysis before deployment can help customers identify where costs can be streamlined and reduced.

Additionally, it is important to note that underutilization of computing resources is rampant in datacenters today. Many systems are overprovisioned to meet peak workloads that demand more computing resources than ongoing workloads. Accompanying this underutilization is a less-than-optimal deployment of personnel supporting these systems. Improving the utilization rates of these systems enables skilled administrators to spend more time on higher-value projects than simple server maintenance.

TCO analyses should include not just the hard and soft costs of the datacenter but also the impact on the levels of service and on average system utilization of computing resources in the datacenter. Many organizations would benefit from TCO analyses that demonstrate where consolidation can improve utilization rates while maintaining and potentially improving the service levels of the applications necessary for user productivity.

Estimating Total Cost of Ownership

IT personnel must gather the financial metrics associated with datacenter deployments. While members of the financial or accounting department can be helpful, only IT personnel can understand fully the total amount of time and materials that must be accounted for. As a result, IT personnel need to have a strong understanding of TCO and all that goes into its analysis.

Overall, IT organizations must pay attention to both the technical and business implications of their deployment plans. The business perspective includes identifying and quantifying the strict IT budgets that require stretching dollars to the snapping point, planning for the system life cycle, and monitoring IT staffing costs and the ever-decreasing supply of resources to pay for additional system administrators.

IDC Methodology

IDC and HP worked together to construct and compare estimates of TCO components for a typical deployment of blade servers and rack-optimized servers. The estimates do not include all possible TCO components but rather focus on three key areas where data is available:

- **Hardware and management software costs.** HP provided IDC with the costs of the components needed for a typical deployment of a new server infrastructure. Specific HP products required to build the target infrastructure, a 320-server system, are included.
- **Facilities-related operating expenses.** HP test data provided IDC with measures of power consumption, cooling cost estimates, and datacenter footprint size for both blade servers and rack-optimized servers. IDC validated these estimates based on knowledge of similar deployments gained from conversations with IT organizations.
- **Initial deployment costs.** HP's data takes the form of time spent installing and configuring the two server systems and pulling the necessary interconnect cables. IDC confirmed that these estimates and assumptions were reasonable.

The result is an analysis of the expected customer experience. The blade server cost model for hardware and software uses pricing and capabilities for a new product that has just reached the marketplace.

Capital Expenses

Capital expenses related to the TCO assessment of a datacenter rollout encompass not just the computing hardware but all related infrastructure components, such as the servers themselves, which include microprocessors, memory, hard drives, network adapters, power supplies, and fans, as well as the associated storage hardware. Components also include system management tools, Ethernet and Fibre Channel cabling, as well as LAN and SAN switches — all of which add up to significant costs for the datacenter.

No matter what form factor of x86 server is deployed — whether blade or rack-optimized — the same number of processors will be required. However, when it comes to fans, cooling, power supplies, management tools, Ethernet and Fibre Channel networking solutions, as well as all the associated cabling, the blade architecture offers organizations a strong and compelling option for reduced capital expenses.

For a point of comparison, HP provided IDC with estimated costs for the purchase, deployment, and operation of two 320-server systems. One system uses ProLiant DL360 G5 servers; the other system uses HP BladeSystem BL460c with 320 server blades. The systems were provisioned to be as similar as possible and full-featured with two processors, hot-plug disk drives, 2GB of RAM, four network interface cards (NICs), one host bus adapter (HBA) per blade, and system management software. Tables 1–3 illustrate the cost savings related to capital expenses, facilities (i.e., power and space costs), and initial deployment.

Differences in initial capital investment for the two 320-server solutions are shown in Table 1. Without including the cost of Fibre Channel HBAs, we found that blade servers are slightly less expensive than rack-optimized servers. When the cost of HBAs is added (see line 2), the server blade solution becomes more cost-effective; this trend continues throughout the remainder of Table 1. The cost advantage of the blades approach is driven in large part by reductions in interconnect components, which is especially important when considering deploying servers in LAN and SAN environments. In this example, 560 fewer Fibre Channel ports are needed, which leads to a cost savings of $603,600. Simpler examples of cost avoidance include a substantial savings on power cables and power distribution units (PDUs) alone (640 in the rack architecture and 80 in the blade architecture) and a significant reduction in hardware associated with Ethernet ports and cables (from 1,600 components in the rack architecture to 100 in the blade architecture). Overall, implementing a blade architecture reduces capital expenses associated with these hardware and software components by 36%.

TABLE 1

Capital Expenses: Rack-Optimized Versus BladeSystem

Type of Cost	Rack-Optimized Quantity	BladeSystem Quantity	BladeSystem Savings ($)
Server systems without HBAs	320	320	135,020
Fibre Channel HBAs	320	320	464,000
Ethernet cables	1,600	100	37,500
Ethernet switch ports	1,600	100	75,200
Fibre Channel cables	640	0	209,200
Fibre Channel ports	640	80	603,600
Racks	32	20	17,868
Power cables and power distribution units (PDUs)	640	80	30,848
Management software	960	320	108,160
Total savings ($)			1,681,396
% savings			36

Note: The data was obtained directly from HP and refers only to HP products. IDC believes that the dollar value in savings, as well as the total percentage of capital cost reduction, can serve as a guideline to where opportunities to decrease capital expenses exist.

Source: HP, 2006

Operational Expenses

Operational expenses are the second part of the TCO formula. These costs include such critical elements as the direct datacenter expenses, such as utilities for power and cooling, floor space, and cabling. Again, using HP data for illustrative purposes, we found that the clear savings available with blades compared with rack-optimized servers can be substantial. Table 2 builds on Table 1, continues the assumption of 320 servers being deployed, and contrasts the power and cooling costs for rack-optimized servers and a BladeSystem.

Power and Cooling Savings

HP and IDC forecast a 69% reduction in energy consumption over a three-year period for IT organizations that migrate to blade architectures. Power consumption and heat dissipation are two of the most controversial and top-of-mind issues addressing the server industry today. Both multicore processors and lower-voltage processors reduce power requirements. However, within the overall design of the server itself, blades have established a clear power advantage compared with rack-optimized servers.

In Table 2, HP data is used to illustrate the difference in power consumption by rack-optimized servers and blade servers. On average, rack-optimized servers draw 280 watts. Deploying with the same computing power on blade systems results in an average draw of 216 watts. While the wattage may not seem dramatically different, HP and IDC forecast a cost savings of $53,701 over a period of three years for 320 servers.

Lower power requirements lead to reduced demand for cooling. Again, while the difference in cooling demands for rack servers and blade servers (140 watts versus 108 watts) is relatively small, the overall savings estimate shown in Table 2 is $26,810 over three years for 320 servers.

By reducing the average server power draw, IT organizations will be able to extend the life of their existing datacenters because they can continue to add servers that fit within the power budgets of already-plumbed facilities. This move avoids or at least delays the need to build a new IT datacenter.

TABLE 2

Three-Year Power, Cooling, and Space Expenses: Rack-Optimized Versus BladeSystem

Type of Cost	Rack-Optimized Quantity	BladeSystem Quantity	BladeSystem Savings ($)
Power	280 watts/system	216 watts/system	53,701
Cooling	140 watts/system	108 watts/system	26,810
Datacenter space	384 square feet	240 square feet	9,000
Total savings ($)			89,511
% savings			25

Assumptions:

- 320 servers are operated 8,736 hours a year for three years.
- Electrical power costs $0.10/kW-hour.
- Cooling costs are approximately half of power costs.
- Datacenter space costs $62.50 per square foot for three years.

Source: HP and IDC, 2006

One of the first clear benefits of blades that was recognized by the industry following their introduction several years ago was that they required less space to deploy than rack-optimized servers. Within one 10U space, for example, an organization can potentially deploy up to 16 dual-core servers. That is a significant increase in the amount of computing power available within that space, particularly when compared with the maximum of dual-core rack servers that could occupy the same real estate.

Doubling the number of servers available within the same amount of rack space may seem insignificant when dealing with a few servers in a standalone deployment. However, the cost avoidance grows larger for a deployment of 320 servers. If 320 blade servers are deployed, the required square footage shrinks by 144 square feet, leading to $9,000 in cost savings over three years.

The space savings associated with a move to blades allows a customer to pack more infrastructure into an existing datacenter. Increased compute density helps to ensure that the enterprise is maximizing its datacenter investments. The datacenter is a sunk cost, however. Greater potential capacity avoids future costs but does not lower current costs. For IT organizations that colocate servers with third-party providers, however, there should be a more direct and immediate cost reduction.

Initial Deployment of Server Infrastructure

The largest cost reductions are associated with initial deployment. Table 3 shows HP's estimates for time spent installing and initializing servers as well as installing network cables. These estimates are the result of economies of scale. Installing a server is much like installing a server blade enclosure. Once that task is completed, adding blades takes far less time.

TABLE 3

Deployment Expenses: Rack-Optimized Versus BladeSystem

Type of Cost	Rack-Optimized Task Quantity	Rack-Optimized Task Time	BladeSystem Task Quantity	BladeSystem Task Time	BladeSystem Savings ($)
Initial system and software deployment	320 installations	4.5 hours/ installation	320 installations	0.8 hours/ installation	59,200
Pulling cables	2,304 pulls	$100/pull	180 pulls	$100/pull	212,400
Total savings ($)					271,600
% savings					90

Assumption:

- $50/hour for technician time.

Source: HP and IDC, 2006

While connecting server and storage equipment is a rudimentary function, the cost associated with this task grows quickly when deploying large numbers of servers. With an estimated cost of $100 per cable pull, HP and IDC estimate the operational savings for moving to bladed infrastructures to be $212,400 for a 320-server solution.

Additional Cost Savings

The preceding tables illustrate quantifiable cost-saving estimates that are attributed to the blade server approach. Additional savings not modeled in these TCO analyses, while harder to quantify, are equally important to consider, especially savings related to the cost of downtime. IDC believes that downtime can easily contribute between 18% and 35% of TCO and should be a key factor in a full TCO analysis.

Server architectures including blade server systems can address redundancy schemes in a more cost-effective matter. A 1 + 1 design (i.e., a primary system and a hot backup) provides failover, but at twice the cost. An N + 1 design (i.e., multiple primary servers with a spare blade available for failover) can be more cost-effective. Blade servers are capable of supporting N + 1 designs.

Downtime matters on several different levels. First, the administrator time required to resolve issues can be quantified in hard numbers based on the rate of pay. Second, the cost of purchasing failover infrastructure to ensure availability, at a minimum, can double capital costs. Finally, because the data services provided to end users are interrupted during downtime, there is a clear compromise of end-user productivity, which drives up the TCO for the analysis period.

While clearly very important, estimated cost of hardware for high availability and estimated cost of downtime are not included in these examples because these costs vary widely by industry and by application. That said, server platforms that include redundant hardware and software components, alternate I/O paths, highly available systems software, and robust systems management and automation software tools can improve TCO by keeping downtime to a minimum.

HP BladeSystem c-Class

HP's next-generation BladeSystem c-Class portfolio has been designed to address the key TCO issues faced within today's datacenter, in particular issues associated with server management, utilization, and power and cooling. HP also has introduced three new technologies, which will be native features of the new BladeSystem and which will ultimately help users reduce overall datacenter operating expenses. These technologies will allow HP to differentiate its product from both competitor blade offerings and the rack-optimized form factor at the TCO level. The following sections outline these new HP technologies.

HP Thermal Logic Technology

Many IT and datacenter facility managers do not understand the power and cooling dynamics of their datacenter facilities. Shrinking form factors have enabled users to extend the longevity of datacenter real estate where physical capacity has become an issue. Blades have played a major role in increasing overall server densities over the past two years. Essentially, the number of processors per U has increased significantly, placing strain on power and cooling infrastructure, which predates the advent of blade servers. Understanding and managing power consumption across the datacenter are key factors when reducing TCO over time. IT and facility managers also need to ensure that deployed infrastructure can operate within the confines of power and cooling capacity, as defined by existing power and cooling infrastructure, to negate the need for expensive upgrades.

The major trends for energy consumption are as follows:

- Power costs continue to increase in share when considering total datacenter operating expenses through a combination of increasing energy prices, server densities, and a focus on reducing other components of datacenter expenses.
- Increasing server densities have seen power consumption soar to the level that power capacity dictates the maximum number of servers that can be deployed within a datacenter.
- Heat output is directly related to power consumption, with estimates suggesting that for every two kilowatts of power consumed by servers, an additional kilowatt is required to cool the emitted heat.

HP Solution

Once customers deploy BladeSystem c-Class blades, they will be able to take advantage of HP Thermal Logic technology, which combines monitoring, reporting, and adaptive management functionality of power and cooling resources within the BladeSystem c7000 enclosure. This capability is enabled through built-in thermal instrumentation that affords BladeSystem manual or policy-based control of power consumption and cooling to match demand without decreasing processor performance. HP will also enable customers to quantify power consumption and cooling needs across the datacenter and therefore help IT and facility managers make informed decisions on how to maximize the ROI associated with energy investment and reduce the overall TCO associated with the infrastructure. Importantly, this management can be achieved at system and rack levels or through the creation of *zones* within the datacenter.

TCO associated with power and cooling is reduced via the following approaches:

- Increased efficiency of power and cooling provisioning through shared resources that allow delivery to be adapted to the most efficient state according to requirements at a given time
- Use of HP fan and cooling architecture for more power efficiency that reduces the actual power consumed and lowers the power requirement to cool the blades
- Elimination of overprovisioning of power and cooling and associated energy costs and extension of the life cycle of existing infrastructure
- Enablement of users to deploy a standard, scalable solution for power and cooling across multiple server form factors

HP Virtual Connect Architecture

Bandwidth bottlenecks and scalability pose significant problems for IT managers today as data-transfer volumes increase exponentially. Efficient coordination and planning are needed among the server system, the SAN, and the LANs so that servers can be added, moved, or replaced as needed.

Currently deployed technology inhibits efficient utilization of network fabric as connectivity remains attached to the server or, in the case of blades, the enclosure backplane. Furthermore, previous generations of blade servers suffer from Fibre Channel rates that limit usefulness in working environments. This situation has resulted in overprovisioned, and underutilized, connectivity fabrics to accommodate for peak-traffic flow. As such, TCO is negatively impacted by overprovisioned, complex fabric infrastructure that relies on close management of server, SAN, and LAN components to maintain seamless operations.

HP Solution

Customers deploying BladeSystem c-Class can run up to four simultaneous redundant fabrics with eight identical, high-performance interconnect slots that can house any chosen interconnect. The backplane has the potential to support aggregated bandwidth of 5Tb sec^{-1} or 320Gb sec^{-1} per server bay, providing scalability and investment protection. Fibre Channel rates of c-Class blade infrastructure are twice those of p-Class systems. Connectivity becomes enclosure-dependent irrespective of processing platform, providing investment protection. Support for 10Gb Ethernet, when available, is also built in.

With the option of HP Virtual Connect modules, LAN and SAN connectivity can be aggregated into a pooled virtualized resource that can be shared physically or virtually across c-Class blades. Viewing connectivity resources in this way enables customers to consolidate switches and thus increase utilization. Estimates show that up to 90% of the time spent provisioning connections between server and LAN/SAN can be removed through provisioning up to 64 server blades plus additional virtual machines on each blade, all in parallel. Cable costs can be reduced by up to 98% because c-Class enclosure requires 1+1 Ethernet cables to connect to LANs and 1+1 Fibre Channel cables to connect to SANs. An added benefit is improved airflow, which leads to more efficient cooling.

Ultimately, TCO reduction is achieved as a compounded result of all these factors allowing IT managers to reduce the number of administrators required to manage LAN and SAN connectivity, increase flexibility, reduce time spent managing switch infrastructure, remove expensive travel costs associated with switch management, increase switch fabric resilience through redundancy, and remove costly error margin by removing human intervention.

HP Insight Control Management

Management and administration costs are the primary costs in datacenter economic models. Users are looking for solutions that can reduce the associated cost by both simplifying IT infrastructure and streamlining management processes through automation. IT manager spend today comes under close scrutiny, and favorable ROI characteristics are a prerequisite, in part because recent changes in compliance requirements have driven CFOs to become more closely involved in the IT procurement process.

IT managers can have a positive impact on TCO by reducing this cost if given the tools to do so. They can improve the productivity and satisfaction of staff by reducing the amount of time that staff members spend on mundane maintenance operations so that they can focus on revenue-generating activities. System management is key to enabling customers of HP solutions to deploy dynamic IT compute models and thus achieve the maximum possible ROI on datacenter infrastructure.

HP Solution

HP Insight Control management comprises two core components: intelligent infrastructure and Insight Control software. It forms the management link between HP Thermal Logic technology and HP Virtual Connect technology and integrates with ProLiant Essentials and HP Systems Insight Manager applications to enable single standardized management of datacenter infrastructure.

Intelligent management is built in to the HP BladeSystem enclosure, and integrated Lights-Out Management 2 (iLO2) is a standard component in each server blade, with chassis enclosures housing the new Onboard Administrator module that enables simplified setup, diagnosis, and maintenance of the blade infrastructure — either at the rack or remotely. Further management efficiencies are gained through the ability of HP Insight Control to catalog resources, automate deployment and workload reprovisioning, and monitor the health of HP BladeSystem, providing recommendations to IT managers on how to alleviate performance bottlenecks.

By incorporating HP Insight Control into future infrastructure deployments, customers will be able to deliver automation for key management processes. As a result, customers will be able to remove HR from mundane processes with the resultant improved TCO metrics.

IDC Analysis

Key barriers exist, however, despite the broad market adoption blades have experienced to date. One of the top barriers is proving the general applicability of blades across IT organizational needs and mission-critical applications. The market is just reaching the beginning of the mainstream adopter phase, and IDC believes that HP, along with other blade vendors in the x86 market, must overcome the market perception of blades being the exclusive domain of specialty solutions for large infrastructures.

Density and Cooling

With the adoption of blades, IDC finds customers are concerned about the disruption of the datacenter from a facilities perspective. Concerns about having to upgrade power and cooling infrastructures as well as rearchitect the datacenter layout cause users to hesitate when making the move to blades. The opportunity for HP is to leverage the chip, system, rack, and datacenter technology advancements to enable customers to deploy blades without having to make significant investments in a new datacenter or having to overhaul an existing environment.

Choices with Networking

Historically, the integrated networking of a blade solution has been positioned as a benefit to customers. Increasingly, IDC is finding that the inclusion of switch and server in the same chassis is causing conflict in that networking and server administrators are challenged over ownership issues. The opportunity for HP is to offer choice with respect to networking options. With the c-Class product, customers can now choose from a pass-through, a switch, or the Virtual Connect option for either Ethernet or Fibre Channel networking.

Confirmed Proof Points

IDC and HP prepared a TCO analysis in the same way that an IT organization would, that is, as an estimate to inform decision making. For up-front capital expenditures and especially for downstream operating expenses, data from actual customer experiences is needed. Moreover, the estimates in this white paper are for a 320-server infrastructure. Over time, actual TCO measures for larger and, in particular, smaller infrastructure deployments will be crucial to building customer trust.

Messaging to the Midmarket

As part of its offering, HP will need to clearly articulate how small and midsize organizations can map their needs to a complete blade portfolio. The company should create a succinct value proposition that appeals to the market majority, continue to drive down the cost of computing and networking, as well as highlight the importance of IT infrastructure reliability, availability, and serviceability (RAS) to all customers. Key to all these challenges is the development of a future road map that is both pragmatically tangible in the near term and strategically comprehensive in the longer term.

Perceptions Concerning the Lack of Standards

Another barrier that must be overcome is that of the perceived lack of standards in these servers. IDC research shows that standards are important to mainstream technology users because standardization ensures price competition and reduces choice in multivendor environments. On the pricing side, the dynamics of the blade marketplace are already addressing the drive for commoditization. In terms of choice, HP crafted its BladeSystem chassis to support multiple generations of blades so that chassis investments could be leveraged over a five- to seven-year time frame.

Conclusion

TCO analyses conducted by IDC and HP indicate that significant costs are expected for IT organizations that deploy server blade architectures. Immediate savings are forecast at the onset, and the blade advantage is particularly salient for servers that work in conjunction with SANs. HP and IDC estimate power, cooling, and space savings of 25%. Moreover, HP and IDC forecast dramatic reductions in initial deployment expenses as a result of built-in, optional automated management tools and simplified cabling design.