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

WHITE PAPER

The Business Value of Consolidating on Energy-Efficient Servers: Customer Findings

Sponsored by: Hewlett-Packard

Matthew Eastwood Jean S. Bozman
Joseph C. Pucciarelli Randy Perry
May 2009

OVERVIEW

Today's competitive business environment is putting IT managers under pressure to accomplish more each year with flat or even reduced budgets. Business managers are looking for more flexible business models, more competitive operations, and faster time to market. All of this means that IT organizations must find operational savings while staying within their budget. They must find new infrastructure solutions that are more efficient and more cost-effective than their current solutions. They must do all of this and, at the same time, accomplish the following: improve service levels, provide faster response times, cover more users, and ensure better-integrated applications that are always available.

Most organizations faced with such budget challenges put off capital expenditures (capex) and seek alternatives, such as extending server life cycles and extending software licenses. This paper demonstrates that such a buy and hold strategy actually adds costs to the datacenter. In fact, refreshing server infrastructure on pace with newer technology (e.g., every two years) can reduce six-year server costs by 33% compared with buying and holding servers for those six years. This occurs not only because today's servers, with the latest semiconductor technology, can do 16 times the work at almost one-half the power requirements but also because of the reduced maintenance overhead and IT labor costs associated with advanced technologies. As the servers age in place, maintenance costs; costs for power and cooling, managing, and monitoring servers; and staff costs can grow steeper, according to a demand-side, customer-based study conducted by IDC.

A server refresh financial decision that takes all these factors into account offers surprising findings that contradict and challenge traditional IT depreciation cycles.

SITUATION OVERVIEW

Server market dynamics are changing the form and function of a new generation of servers. Ten years ago, in 1999, a high percentage of servers were standalone, pedestal-based servers, including mainframes and scalable SMP servers. Then, with the sudden economic downturn of 2001 and 2002, the mix of servers being installed changed dramatically, with large numbers of small rack-optimized servers appearing in the datacenter, to be followed several years later by blade servers, managed in groups, within a blade chassis.

This proliferation of small servers brought about reduced capex, as shown in Figure 1, but did little to address operational expenses (opex) within the datacenter. In fact, the opex costs continue to rise — especially power/cooling costs (growing eight times as fast as the spending rate for the servers themselves) and maintenance/management costs associated with IT staff costs (growing at four times the rate of new server spending).

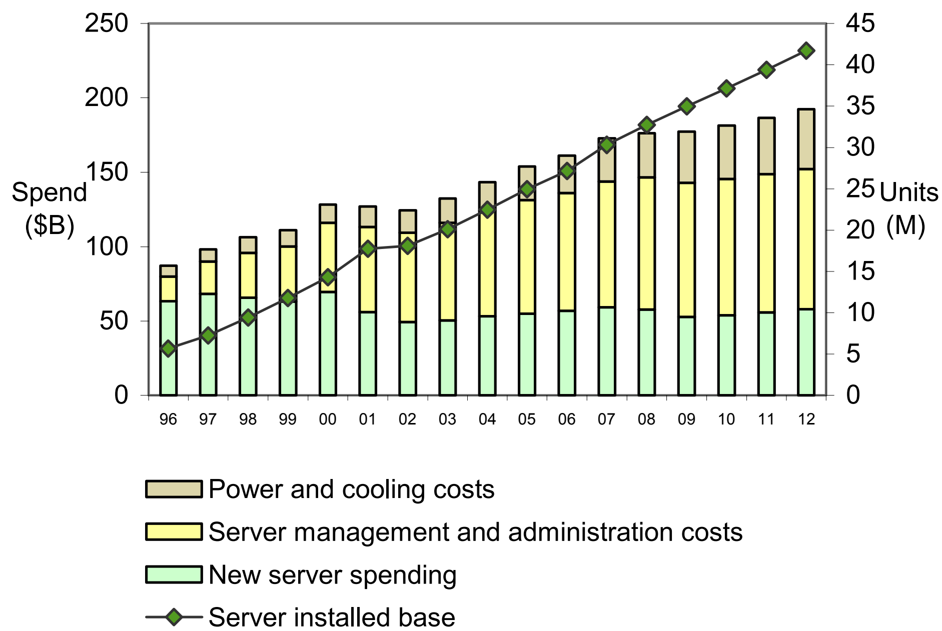
Capex continued to be controlled, however, by declining average sales prices for servers. At the same time, compute density and power density have increased dramatically, with the advent of multicore processors in the x86 server space (servers based on x86 processors made by Intel and AMD). Since 2004, the average system cost per processor has dropped 18% to $2,000 and the average core cost has dropped 70% to $715.

While servers have become much more powerful over time, acquisition costs and energy requirements have dropped. However, the older-generation servers have fewer cores per processor — and throw more heat — while generating increasing power and cooling, maintenance, and management costs over time, affecting IT operational costs and IT productivity.

The resulting growth in operating expenses severely limits IT's ability to meet the needs of the business for improvements in computing performance and capabilities while operating within a reasonable budget. Figure 1 shows this effect as growing operating costs — management and administration — continue to take a higher percentage (versus new server spending) of the IT budget.

FIGURE 1

Changing Datacenter Economics



Source: IDC, 2009

FUTURE OUTLOOK

The Current State

Today, a variety of key trends are combining to shape today's server cost dynamics, which are progressively changing the datacenter landscape. Among these trends are server virtualization, which improves resource utilization for each server on which it is deployed; consolidation of workloads, which reduces server management costs by reducing server footprints in the datacenter; and the move to "green IT," which addresses inefficient use of systems within the datacenter leading to excess heat generation and "hot spots" throughout the datacenter.

Server Replacement Cycles

IDC's 2008 Server Workloads study found that technology refresh helped to address some of these operational expenses. The Server Workloads study of 1,000+ IT sites found that 39% of new server purchases occurred as part of a routine server refresh. New application projects drove the purchase of another 33%, and 28% of new server purchases occurred to support additional compute capacity.

The usable lifetime for servers within the datacenter varies greatly, ranging from as little as two years for some x86 servers to seven years or more for large, scalable SMP server systems. The typical x86 server is in place for 4.5 years, according to the study. IDC surveys indicate that almost 40% of deployed servers have been operating in place for four years or longer. That represents over 12 million single core–based servers still in use.

Extended server life cycles are one of the leading drivers for growth both in terms of the size of the installed base of physical servers and in terms of the continued growth in operational expenses that are related to management and to power and cooling costs. As we see in this paper, aging server infrastructures can play a substantial cost-adding role in datacenter cost dynamics.

Virtualization Trends

Server virtualization — in simple terms, the capability to divide a single physical device into multiple logical devices thereby making a single physical server operate like multiple servers — remains one of the key infrastructure trends impacting the server market today. Virtualization allows IT to use a single server as if it were multiple servers, and because it involves logical and not physical implementations, it also allows for very rapid changes in application deployment practices.

IT initially focused on virtualization primarily to consolidate multiple workloads on a single machine. That focus has evolved as virtualization plays a more pronounced role in new applications and deployments. The 2008 IDC server virtualization survey found that organizations tend to implement virtualization on new servers rather than upgrading older servers with virtualization capabilities. The adoption rate is increasing, with more than 30% of all new servers shipped as virtualized servers — a number that is expected to top 50% in 2009. High numbers of virtual machines (VMs) were found, with the average virtualized x86 server hosting eight or more VMs compared with two to four VMs per physical server just three years ago.

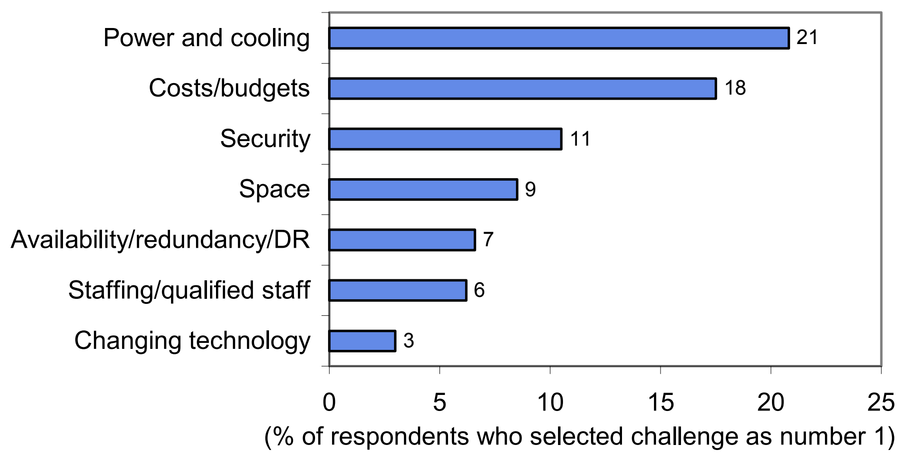
IT deployed virtualization on newly purchased servers (versus previously installed servers) two-thirds of the time. New application projects accounted for 52% of those new virtualized servers, and standard refresh cycles accounted for the other 48%. Although server consolidation remains a key benefit of server virtualization, many customers are implementing virtualization for other benefits such as business continuity, disaster recovery, IT test and development, and availability. This virtualization trend has produced cost savings primarily via reduced spending on new servers, lower power and cooling costs, and reduced space in the datacenter.

Power and cooling are two of the most pressing issues for IT managers today. IDC's Green Datacenter Study reveals that over 20% of datacenter managers feel that power and cooling capacity counts as their number 1 concern, while 18% of managers identified costs/budgets as their top concern (see Figure 2).

FIGURE 2

Top Datacenter Challenges

Q. What is the number 1 challenge that your datacenter faces today?



n = 504

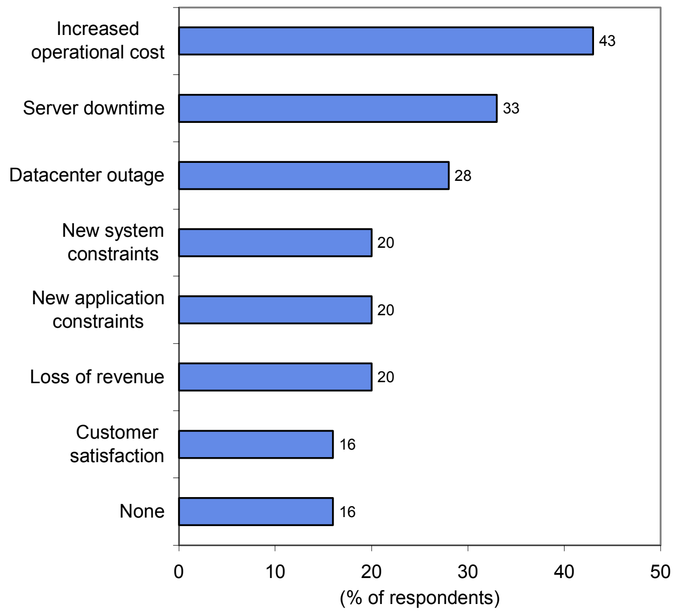
Source: IDC, 2009

IDC's study probed more deeply into why power and cooling topped datacenter managers' concern list, as depicted in Figure 3. Responding to how power and cooling impacts the business, 43% of managers highlighted "increased operational cost" as an issue and 33% identified "server downtime." The majority of datacenter managers had experienced some operational and business impacts from power and cooling issues. This remains a critical issue for organizations of all sizes and complexity and can severely inhibit IT's ability to support future business needs.

FIGURE 3

How Power and Cooling Issues Impact Business

Q. Has your organization experienced any of the following business impacts from issues related to power and cooling?



n = 504

Note: Multiple responses were allowed.

Source: IDC, 2009

Dealing with Datacenter Challenges

Datacenters must deliver operational savings to stay within budget. They must achieve better staff utilization and asset utilization while building business flexibility and resilience with lower capex and lower opex. They must achieve all of this within a predictable ROI and a lower TCO.

Today, businesses are looking for more flexible IT models to solve these challenges while improving service levels with faster response times. But they must look to innovative usage scenarios and new acquisition models to make this possible.

To effectively support the business during these challenging times, IT must deal directly with each of these critical issues:

- Increased server count
- Rising power and cooling costs and issues
- Increased business needs and requirements
- Budgetary constraints

IDC's field research indicates that IT can address each of these issues with a combination of server virtualization and consolidation — that is, the conversion from multiple aging servers to smaller numbers of energy-efficient servers, each of which delivers more performance than past generations of x86 server technology. These datacenter strategies address each of the issues as follows:

- Consolidating to fewer systems decreases server count and thereby reduces all of the associated expenses associated with each physical server.
- Consolidating to fewer, more energy-efficient servers reduces the rising power and cooling costs; this reduction in power and cooling demand can also impact the physical operations of the datacenter, reducing systems downtime and limiting the risk of potential datacenter outages.
- Newer generations of servers that incorporate throughput and performance capabilities that are significantly higher than those of legacy servers help address the accelerating demand to support increased business needs and requirements by providing better application throughput and scaling; the performance gains of new, current-generation servers lead to fewer servers providing more and better application performance for a larger number of legacy servers.
- Because IT can run on fewer, newer servers, acquisition costs drop (lower capex), and because these fewer, newer, and more energy-efficient servers consume less power and cooling resources, operating expenses fall (lower opex). Both factors help IT operate within stringent budgetary constraints.

Triaging the IT Infrastructure

Because each datacenter represents a collection of servers that reflect past IT decisions — and perhaps past IT management administrators — each organization must take stock of its server inventory and decide how to manage through this year's IT budget restrictions. In some cases, customers are electing to remove older systems in a selective way. This requires gathering up the workloads from those older machines and moving them to small numbers of new servers based on multicore, energy-efficient technology.

As mentioned previously, almost half of the server installed base lags the current available technology by at least one generation of processor technology. This means that organizations that have higher levels of aging servers within the datacenter are paying more for their server operations than they need to. Due to this phenomenon, they are missing two key cost benefits: one, the significantly better performance characteristics that the latest generation of semiconductors provides and two, the improved energy-efficiency and power management features designed into the newer generation of servers.

Three Generations of HP x86 Servers

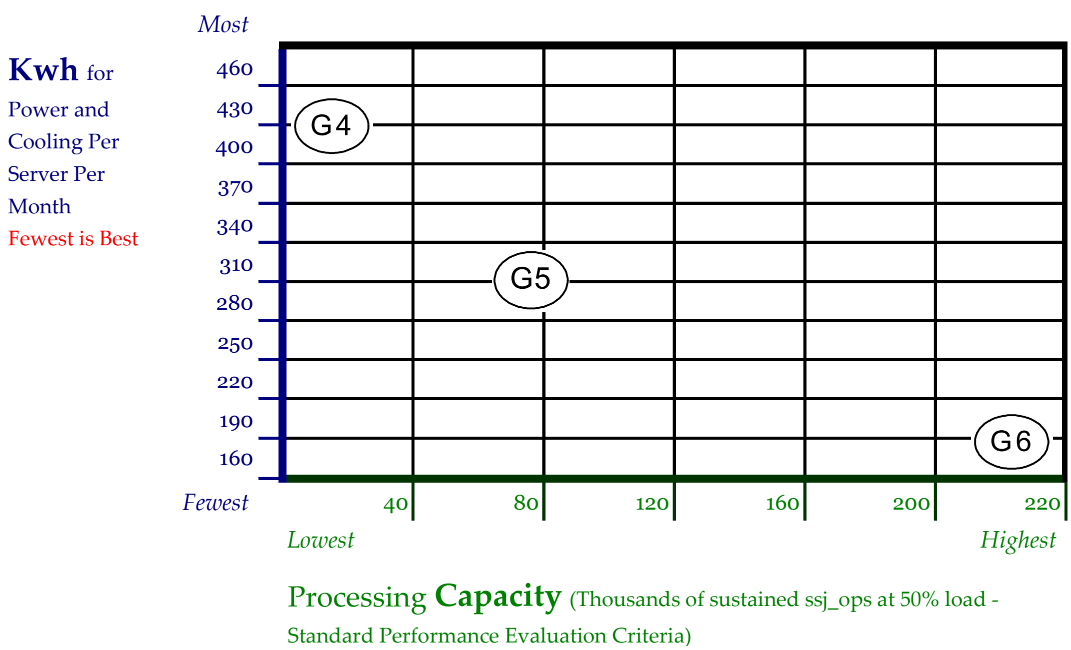
A closer look at HP's last three generations of servers and their Intel-based technology demonstrates the way that these technology trends have played out in terms of the form factors and capabilities of new servers offered by HP — all of which were designed to address the datacenter "pain points" around inefficiency, overheating, and aging technology.

To illustrate the generational change in performance and power usage efficiency, HP benchmarked three succeeding generations of two-socket servers using an industry benchmark, SPECpower_ssj2008, which reflects Web-enabled workload performance. The results show dramatic improvements for each newer generation — from the fourth generation (G4), to the fifth generation (G5), and finally to the sixth generation (G6) — using Intel's latest-generation microarchitectures such as the i7.

In addition, the newer system supports four times the memory and includes higher-bandwidth I/O and other newer technologies that make the HP ProLiant DL380 G6 server a much higher-performing and better balanced system than its predecessors. In addition to being better performers, these systems are much more capable of handling larger and more complex workloads than the previous generation because of additional capabilities in memory, I/O, etc. As a result, they are better able to handle additional workloads not only in terms of their general compute operations but also in terms of the added burden of complex I/O requirements typical of consolidated and virtualized operations. Figure 4 shows the increase in system performance and the decrease in the amount of power required while running the system.

FIGURE 4

The Effect of Moore's Law on Datacenter Costs



Source: Standard Performance Evaluation Corporation and IDC analysis, 2009

Note: Generational improvements above are based on testing submitted to the Standard Performance Evaluation Corporation (SPEC). The results quotes are based on HP's DL380 G4 (G5 and G6) platforms. The tests were run using workloads that varied from 0% to 100% in terms of server utilization to characterize the spectrum that may run on these systems. The numbers used for comparison are based on a 50% load. The benchmark compares performance and power usage. The G6 generation puts out over 19 times the performance (in ssj_ops) while using less than half the electricity of the G4 generation.

Options for Datacenter Technology Refresh

Depending on their current workload and legacy server maturity, datacenters may want to look at options for refreshing servers from a number of angles. Three such analytical views may cover most organizations' needs. They include measuring the effect of:

- Replacing one older server with a newer server: **One to one**
- Replacing many older servers with one server by consolidating/virtualizing: **Many to one**
- Combining a variety of one-to-one and many-to-one refreshes: **Many to many**

One to One

A straight one-to-one replacement involves replacing a particular machine with a machine using the latest technology to provide a significant boost in performance. Datacenters will do this to meet required performance levels at a lower cost per operation, better performance per watt, and reduced additional power and cooling requirements.

Many to One

Organizations may opt to replace several older systems and consolidate their combined workload onto a single more powerful server. This approach will likely include virtualization as part of the consolidation platform. The resulting platform will prove both more cost-effective and more flexible for the datacenter. Managers can justify the many-to-one refresh with cost savings on hardware and reduced power and cooling demands. The resulting decrease in the number of physical systems and their associated cables and wiring will have the effect of reducing datacenter complexity, along with its associated maintenance and servicing costs.

Many to Many

Large datacenter installations involved in large, complex refresh projects will likely carry out a variety of both many-to-one and one-to-one refreshes. Enterprises may refresh one to one, many to one, or many to many, using servers that have the same socket count as the older systems being replaced. They may also replace smaller systems with larger-capacity systems, especially when consolidating multiple older systems onto fewer, newer systems. For example, refreshing an older two-socket system with a newer two-socket system can result in consolidation ratios ranging from 4 to 1 to 6 to 1. By contrast, replacing the existing two-socket system with a four-socket system can double the consolidation ratio to 12 to 1.

How the HP G6 Series Addresses These Refresh Approaches

Organizations working to refresh their servers can evaluate not only the types of scenarios described in this document but also a variety of flexible means to update compute resources. The refresh can also introduce newer HP systems and solutions that improve cost, efficiency, and operations. With respect to the HP ProLiant G6 servers, the generational improvements noted earlier illustrate the value of new HP platforms for this level of refresh. For example, a G6 running at 20% load would directly reduce energy usage by over 50% and still provide multiple times more performance than a fully loaded G4. The G6 would reduce operating costs while providing significant headroom for growth in existing workloads, new workloads, and future consolidation to fewer systems. As shown in Figure 4, the G6 offers 10 times the performance of the G4 series of HP ProLiant systems at roughly half the cost for power/cooling. With respect to performance, G6 servers can be shown to use fully 95% less power per transaction than G4 servers.

HP's G6 Generation of Servers

HP introduced a new generation of HP ProLiant servers in March 2009 that are designed to provide more performance and better energy efficiency than previous generations of ProLiant servers. These changes address many of the most pressing requirements in the datacenter: reducing power/cooling costs, ensuring reliability and availability, and reducing the amount of datacenter floor space required to house the servers. All of these improvements are aimed at improving operational costs while keeping acquisition costs competitive in the marketplace.

The new G6 line of ProLiant servers has 11 models, including rack-optimized servers, blade servers (housed within a chassis), and freestanding tower systems. New models with G6 technology include the ProLiant DL380, DL370, DL360, DL180, and DL160 rack-optimized servers; the BL490c, BL460c, and BL280c blade servers; and the ML370 and ML150 tower servers.

All of these systems are based on Intel Nehalem-generation Xeon 5500 quad-core processors, which Intel announced as capable of two to three times the performance of earlier Xeon processors. Key factors in the improved Xeon performance are an integrated memory controller and a new technology, QuickPath Interconnect (QPI), to replace the front-side bus (FSB) — the traditional connection to I/O to off-chip resources. At the same time, Xeon 5500 processors operate using up to 40% less power, which reduces heat dissipation, allowing multicore server systems to be densely packed in the datacenter while reducing cooling requirements.

Intel Nehalem Xeon 5500 Support for Virtualization

The Intel Xeon 5500 processors, based on Intel's Nehalem microarchitecture, offer improved support for virtualization capabilities along several dimensions. Overall, improved memory use and improved I/O speeds — along with overall improved energy efficiency, compared with earlier Xeon processor designs — will be leveraged to improve performance of workloads running on a virtualized x86 server.

First, support for enhanced VM memory management in hardware, together with the processor's integrated memory controller, allows users to increase the density of VMs running on each physical server while also reducing VM switching latency that would impact the overall performance of the virtualized computing environment. This is important because the average number of VMs per physical server is increasing — to an average of eight or more in 2009 — as adoption of virtualization in the x86 server space continues to rise. With average VM densities increasing, customers will need to ensure performance for the workloads running with VMs and managed by software hypervisors running on the hardware platform.

Further, Intel's Virtualization Technology (VT) features improve connections to the physical system's I/O and access to networked devices. Intel VT-c provides hardware assistance in network devices, which removes system bottlenecks and enables full network bandwidth to the server. This is an important step in allowing users to consolidate server network connections from multiple 1 Gigabit Ethernet (GbE) I/O ports onto the new 10GbE hardware. This approach to server I/O consolidation leads to tangible energy savings by reducing the number of network interface cards (NICs) as well as the number of switch ports. In addition, Intel VT-d assists VMMs (virtual machine monitoring software) by allowing VMs and I/O devices to be directly connected, saving the cycles required by the VMM to emulate those devices. Improved access to hardware resources is another key aspect of improving performance for VMs running on physical servers.

IDC also notes that there is improved support for live migration via the VT FlexMigration capability, providing compatibility from servers based on earlier generations of Xeon processors to those based on Xeon 5500 processors, which are only beginning to ship in the marketplace. This extends compatibility back to the Xeon 5100 series processors, allowing customers who have installed those earlier processor products to move today's VMs between servers of different vintages. This is especially important in sites with large pools of physical x86 server resources, as VMs are moved from one server to another, within the installed base, for proactive repair or software upgrade purposes.

HP G6 System-Level Features

Following are some of the top features of the G6 server product line:

- **Dynamic Power Capping:** HP Dynamic Power Capping dynamically sets, or caps, the power that is drawn by the servers. Customers can preset the power limits for each type of server usage (workload dependent), which supports reallocation of power and cooling resources in the server racks. This technology addresses the problem of overprovisioning energy to servers in the datacenter.

- **Sea of Sensors:** HP calls its monitoring of the G6 servers through the use of 32 "smart sensors" located throughout the server "the sea of sensors," which is designed to improve systemwide energy efficiency. This "HP Sea of Sensors" tracks thermal activity, as it occurs in real time, within the server enclosure. This system dynamically adjusts the fan speeds, as needed, to keep the enclosure cool — and to eliminate hot spots.

- **Common Power Supplies:** HP is using what it calls the Common Power Slot design, which leverages four power supplies to support servers running a wide range of workloads (e.g., types of applications and databases). By matching the power supply to the workload, power use is appropriately allocated, and wasting of power can be minimized.

- **HP Virtual Connect Flex-10 Ethernet Module:** The Flex-10 interconnect for G6 servers allocates the bandwidth of a 10GbE network port across four NIC connections. This supports virtualized I/O and reduces requirements for additional network hardware equipment.

- **Insight Management:** HP's Insight Control Environment (ICE) management console is now being "bundled" with HP ProLiant server hardware. The console can be used to monitor both physical servers and virtual servers from a console that is located within the datacenter — or remotely. This type of intensive monitoring has the potential to significantly reduce operational expenses, given its support for visualization of the current health, or status, of all of the server resources.

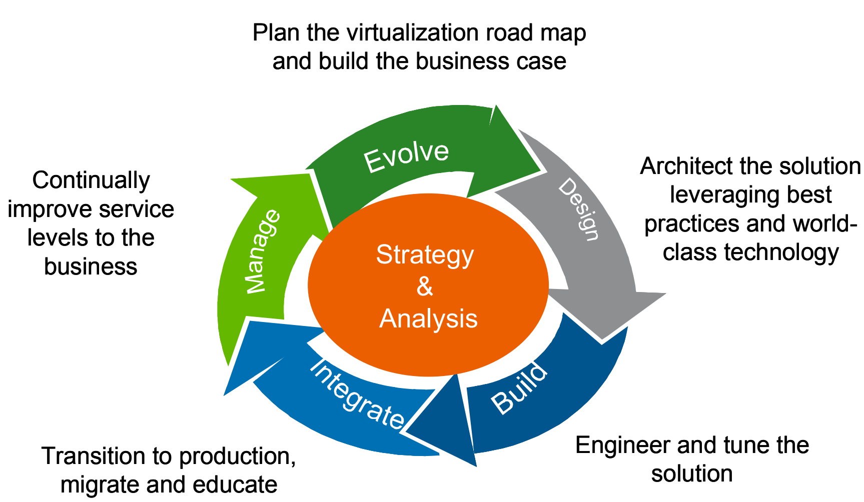
- **Technology Refresh Programs:** HP offers a number of programs that are designed to reduce acquisition costs — including the HP Server Migration Pack, which provides a "wizard" to help move applications from one server to another as new hardware is added into the computing environment; HP ProActive Select, a point-based system for discount purchases of HP server systems; and HP Financial plans that provide flexible financial options (including zero percent financing) and leaseback offerings to reduce acquisition costs.

Server Life Cycles and Technology Refresh

As depicted in Figure 5, HP offers a range of platform capabilities that match a wide range of company sizes and system complexity — ranging from sites that use a single system to organizations using thousands of systems in multiple datacenters. Figure 5 represents HP's published server life-cycle plans, which not only provide new technology via technology refresh but also leverage that technology through an iterative process that fine-tunes, or optimizes, the overall server solution to run more efficiently and to enhance service levels to the business over time. This process is important because business requirements change over time in response to changing business conditions. IT sites can work with HP or with HP's channel partners as they deploy, optimize, and maintain their systems over time.

FIGURE 5

HP and Partners Support the Full Server Refresh Process



Source: HP, 2009

Saving Money by Leveraging Moore's Law: How Current Accounting Misses

Most organizations continue to purchase their servers and IT equipment and then use a "standard" financially derived amortization period, typically five years — most IT shops refresh their x86 systems between three and five years — as a "life-cycle" proxy. Once they acquire, capitalize, and initiate the amortization period, most IT managers avoid replacing the equipment before its normal depreciation cycle runs its course. This approach to server replacement/renewal cycles misses an assessment of the actual conditions and cost factors experienced and instead relies on only the calendar. Administrators work to repeatedly upgrade and reconfigure servers in support of workloads rather than consider a fully burdened cost assessment that would highlight cost reductions the organization could gain by replacing the servers sooner.

Additionally, IT technology planners, knowing that equipment will remain in place for five years or more, tend to specify new acquisitions with as much capacity as possible to maximize flexibility and reduce future in-place upgrades. IDC believes that these management patterns lead to unnecessarily expensive and potentially wasteful infrastructure costs.

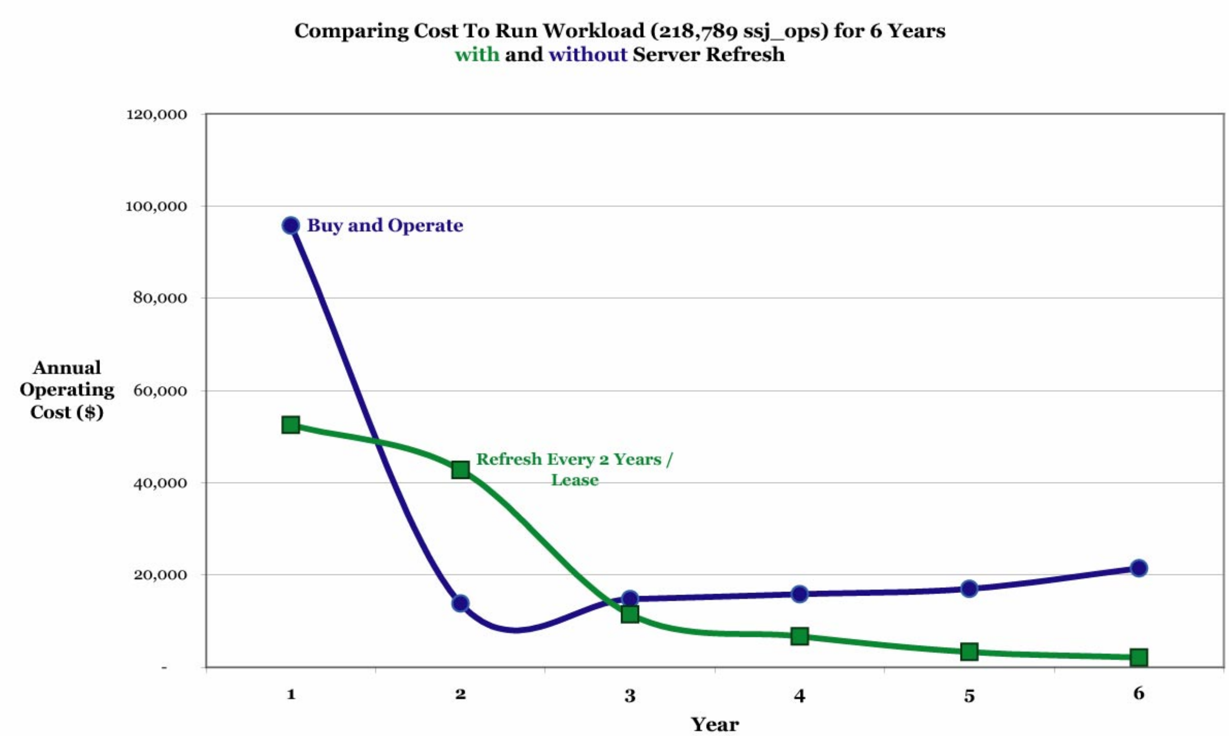
Previous sections of this document have highlighted the immense potential savings in power, cooling, and space costs available through new server technology upgrades. HP's current DL380 G6 performs almost 20 times the number of operations that an equivalent G4 performs per unit of electrical power. Tests indicate that a single G6 could do the work of 16 G4s while consuming less power than a single G4 system. Other cost savings benefits can be added to this list: administrative labor associated with physical equipment management and cabling, upgrades to firmware and the associated regression testing. Because of the high cost and disruptive nature of upgrades, IT organizations strive to avoid this activity. Nevertheless, as the deployment length increases, the cost of maintenance per server increases. Because newer servers require less of this type of maintenance and management, labor cost savings ensue. Operational expense factors materially reduce with newer servers.

Figure 6 demonstrates how more frequent server upgrades that leverage Moore's Law can actually reduce current capex and ongoing opex budgets. This analysis evaluates the relative cost of supporting a specific sized workload across six years with two server technology approaches. The first involves fully depreciating and sustaining a caste of servers for all six years. The second involves deploying new technology servers as available using a leasing option for capitalization. We estimated deployment, maintenance, and decommissioning costs from our business values database and estimated power usage and leasing purchase costs from HP data.

The results show a substantially reduced cost to the datacenter with more frequent server refreshes.

FIGURE 6

Refresh Frequency's Effect on Datacenter Budget



Year	1	2	3	4	5	6	Total
Buy and Operate							
# of servers	18	18	18	18	18	18	
Acquire	72,000						
Deploy	10,800						
Run (power)	8,989	9,304	9,629	9,966	10,315	10,676	
Maintain	3,960	4,508	5,133	5,844	6,653	7,575	
Decommission cost						3,207	
Buy and Operate	95,749	13,812	14,762	15,810	16,968	21,458	**178,559**
Refresh Every 2 Years / Lease							
	G4		G5		G6		
# of Servers	18	18	3	3	1	1	
Lease	28,800	28,800	4,800	4,800	1,600	1,600	
Deploy	10,800		1,928		689		
Run (power)	8,989	9,304	1,138	1,178	238	246	
Maintain	3,960	4,666	707	732	252	261	
Decommission			2,892		516		
Refresh Every 2 Years / Lease	52,549	42,770	11,466	6,710	3,295	2,108	**118,898**

Source: Server performance statistics from Standard Performance Evaluation Corporation and datacenter costs and analysis from IDC, 2009

Customer Snapshots

This section presents two "customer snapshots" based on interviews that IDC conducted with two sites that participated in an early adoption program for the new G6 servers. The snapshots describe the business organizations, their computing environments, and how they have leveraged technology refresh to address operational costs.

Agnes Scott College

Agnes Scott College is a private liberal arts college located in Decatur, Georgia. Operating within tight budgets, the college is working hard to provide quality applications and services to the campus' 16,000 students, faculty, and staff. The datacenter had become quite complex over the past few years, growing into a heterogeneous mix of 58 separate servers and switches that caused "silos" of technology that could not work together easily and increased operational costs. In 2006, the college's IT team started working on a complete overhaul of the server room infrastructure. In the IT team's own words, they were "looking for ways to cut costs, simplify management, and increase the quality and variety of services we could offer to the college."

Even though the IT budget was limited, the college's Network Services group is accustomed to finding new ways to cope — and to make everything work. "Our budgets are not huge — my boss actually calls it *budget dust*," explained H. Duke Miller, Network Services Manager for the college. "We have to work efficiently and get really creative to stay within our budget on all infrastructure projects. I don't think it shows across the rest of the campus [that we have a limited IT budget] because we work smart and [we] put a high priority on key hardware and software replacement projects."

By consolidating the server room, the IT team was able to reduce the number of servers from 58 to 42. The newer HP ProLiant systems use much less power than the older servers, which is having a cumulative effect on power usage in the datacenter. "Whenever we unplug one of the older tower systems or rackmount systems, our power consumption drops by 2% to 3%," Miller said. By removing the extra monitors and a few of the tower systems, the college was able to reduce datacenter power consumption by 23%. The Network Services team estimates that when the six systems are consolidated to one system, power and cooling costs will be reduced by an additional 18%.

Opus Interactive

Opus Interactive is a business hosting company created on 1995. It built its first true datacenter in 1999–2000. The company set itself apart as a datacenter facility, in terms of managed services and hosting, by adopting a micro-datacenter model before "green" was "in." This IT site is green not only from an environmental standpoint but also from a cost perspective. Most datacenters build thousands and thousands of square feet, with all of the required power and cooling. In contrast, Opus Interactive owns the building in which its primary datacenter is located. It has the potential of a 40,000 sq ft facility that is housed within 1,000 sq ft. In addition, Opus maintains several locations, with access to multiple datacenters, supporting disaster recovery on behalf of its hosting and managed services end customers.

Inside the Opus datacenter, everything is modular, allowing server capacity to grow as the business continues to grow. As technologies advance, the company has been able to consolidate workloads more efficiently onto existing server footprints. As the company refreshes its servers, it hasn't had to expand the datacenter in size because it has been able to grow its server capacity in terms of computing density. With each new generation of server technology, this hosting and managed services company has been able to meet its space and power needs by using systems that support more computing density and greater energy efficiency.

CHALLENGES AND OPPORTUNITIES

The worldwide server marketplace is highly competitive, with several large competitors and many smaller competitors across geographic regions. The challenges in this marketplace are clear: Customers expect average sales prices to be reduced over time, leading to declining price tags and relatively low margins in the high-volume x86 server business. The top producers must maintain a high-volume business to keep component and system costs low, and the quality must be high in order to meet customers' business requirements.

To meet these challenges, companies need to continually invest in improving the performance, price/performance, and overall efficiency of the servers they produce. They also must provide globally available service and support, either directly or through their channel partners. Companies such as HP and its top competitors in the x86 server space, Dell and IBM, are well-engaged in this competitive marketplace. With this announcement, HP is demonstrating its R&D investments aimed at improving power/cooling efficiency and reducing customers' ongoing maintenance, system management, and operational costs. This is especially important during the economic downturn, which is sharply limiting IT budgets and IT staff resources and increasing the need for efficiency and ease of use.

CONCLUSION

Refreshing servers this way on an ongoing basis can reduce costs to the datacenter. Rather than put off capital expenditures and extend server life cycles (buy and hold strategy), organizations faced with sharp budget challenges should consider aggressively seeking financial deals that allow them to upgrade their servers with today's technology. Refreshing server infrastructure on pace with newer technology, (e.g., every two years) can reduce a full view of six-year server costs — that is, a view that includes not only acquisition costs but also power and cooling, space, and IT labor costs — by a third, as we have seen in this paper.

Holding Assets Longer May Increase Opex Costs

Most budget managers, especially those with an accounting and finance orientation, tend to assume that the accepted notion that avoiding frequent equipment upgrades means lower cost. As we have seen, Moore's Law, which applies across the computing IT industry, with processor power steeply increasing in cycles faster than typical three- to five-year server life cycles, turns this long-held idea about holding assets over time on its head. As with any established mindset, changing people's perceptions takes time, and effort. This new generation of technology exceeds the normal generational advance typically seen by IT, making the IT benefits substantial and quantifiable but the educational challenge steep.

Economic Conditions

Current economic slowness, caused by the downturn that began in late 2007, will continue to dampen enthusiasm for acquiring new equipment. However, the pace of Internet and compute growth continues for firms, increasing demand for computer capabilities. The challenge for both vendors and customers in this environment involves structuring system upgrade financial terms that encourage achieving the acknowledged cost savings benefits of implementing new technology as available.

Green IT and Business Value

Vendors face the challenge of translating the energy use reductions available via newer technology into compelling business benefits, which will include not only energy cost reductions but also reductions in the broader costs associated with building, running, and maintaining datacenter facilities. The new technology "green IT" value message needs translation into real cost savings and business value.

As we have seen, the Moore's Law continual progression of ever-increasing computing power, as delivered by the processor companies, counters the conclusion that avoiding new equipment and capital expense is the best way to reduce capex acquisition costs. As organizations consider server refresh in their datacenters, they also should consider incorporating a full accounting of all these cost factors — not only capital costs but also labor and electricity costs. This type of analysis, including avoidance of opex costs, may provide surprising conclusions, showing cost projections over the server life cycle that challenge traditional IT depreciation cycles.