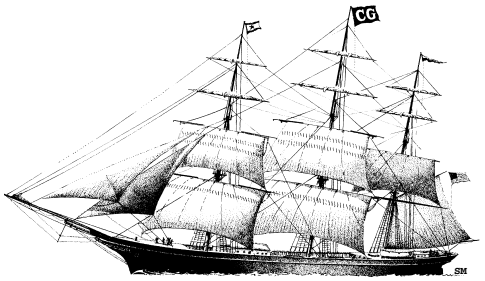
THE CLIPPER GROUP
Navigator™



Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2008034 June 19, 2008

HP Tames the *Mega Center* — Increasing Density while Lowering TCO

Analyst: David Reine

Management Summary

When we look back over time, everyone can select his or her own idea as to what was the greatest innovation of all time. For some, the invention of the wheel is number one – with all of its implications for use in labor saving efforts and transportation. Certainly, the wheel ranks high when considering inventions that improve man's life. When we look for innovation in a more cerebral sense, the printing press surely gets a few nods for its innovation in educating the populace. We take advantage of both the wheel and the printing press in our everyday life – between commuting to the office in our cars, buses, or trains and reading the morning newspaper. If the wheel and the printing press are numbers 1 and 2, where do you think the safety razor ranks? After all, men and women around the world use it on a daily basis. Originally invented in the late 18th century by Jean-Jacques Perret, the safety razor had a guard along the edge to prevent the razor from cutting the skin. This design was not very convenient, however, as the blade had to be removed to be sharpened. Gillette introduced the first safety razor with replaceable blades in 1909, providing them to the U.S. Armed Forces during WW1, thus converting an entire generation of men to the safety razor. There were occasional improvements to the blade over the years, using stainless steel and shifting from a bare blade to a cartridge, for example, but the next major innovation for the act of shaving occurred around 1971 with the innovation by Gillette of the *Trac II* razor, adding a second blade to the cartridge. This enabled a faster, smoother, and safer shave, and incidentally, was patented by Gillette to prevent others from copying their design.

In the same way that Gillette changed the shaving paradigm by putting two blades in a single cartridge in 1971, Hewlett-Packard (HP) has changed the compute paradigm for the largest Web 2.0 or HPC data center by packaging two servers into a single blade cartridge. Facing the dilemma of procuring thousands of new servers per month, the CIO of these enterprises must find a new path to improve data center efficiency. These "megacenters" have limited floor space and a fixed energy budget, limiting their options for additional performance. If infrastructure growth exceeds the available floor space, it could entail the construction of an additional facility, at a cost of millions of dollars. Public utilities ensure that the data center stays within an energy budget – **there is no more electricity, in fact, they offer credits or rebates if you reduce your electrical consumption**. The IT staff must increase server density in its domain by deploying more servers per square foot than ever before in order to support mission-critical applications. The enterprise must reduce energy consumption, not only to support the server environment, but also to cool it, especially when you are paying as much to air condition the data center as you do to power the servers. The CIO must reduce the total cost of ownership (TCO) for an integrated enterprise IT infrastructure yet still have enough processing power available to do the critical mission.

Many enterprises are currently deploying a complex interconnection of thousands of rack-mounted, pizza-box style X86 servers to ensure the availability of sufficient resources to satisfy the myriad mission- and business-critical applications needed to create a dense hosting solution. HP, through its new ProLiant BL2x220c G5 Server Blade, has now changed the deployment paradigm for provisioning of the megacenter, rewriting the rules with innovative packaging to ensure sufficient availability, standardized blade servers to facilitate the deployment process, and energy conservation to lower TCO. To learn more about how HP can shave enterprise TCO, please read on.

IN THIS ISSUE

Enterprise Data Center Issues

Recently, there has been an emphasis among server venders to deliver systems geared to satisfy the requirements and desires of the single largest market, the small and medium business, the SMB. With an employee count of between 100 and 1000, there are literally thousands of SMBs with a need for anywhere from 10 to 50 servers to deploy in a scale-out, rack-mounted architecture, or packaged in blade chassis, to satisfy their mission- and business-critical application needs. At the other end of the spectrum is the next generation enterprise, with Web 2.0 and high-performance computing (HPC) requirements operating in massive data centers, with thousands of servers, looking to deploy *next generation data centers*, or *megacenters*, with potentially millions of customers seeking access to one of its web-based applications. This new megacenter has an entirely different set of issues than the more traditional, legacy data center due to the growth in digital information that is changing the very infrastructure of the data center.

Where the legacy data center is concerned with hardware resiliency, server availability, and price performance in a scale-up[1] OLTP environment, the next generation data center has a different set of priorities to satisfy. These are the needs of search engines, streaming video, online gaming, photo services, and social networks, where the application developers are responsible for the operating environment. This environment is best suited to a s*cale-out* architecture with thousands of X86 servers installed in a "green" computing environment. This new enterprise conceives of deployment in terms of "data center units", where each unit is a megacenter, as opposed to servers or racks. The new enterprise requires a solution built around a "cloud" consisting of a shared pool of homogeneous resources, built in a "stateless" environment where the user only sees a virtual portal to those resources. They **deploy servers in *giant steps***[2] with fully-populated, pre-configured racks of tested systems, **while maintaining the smallest carbon footprint possible to deploy the maximum number of servers**. The plan is for massive expansion in an extensible architecture. Chief among their issues are reducing the total cost of ownership (TCO) for IT, as measured by hardware acquisition cost, cost/Watt (or Watts/Sq. Ft.), through the optimization of capital costs and the reduction of operational expenses, including, but not limited to, reducing the amount of energy resources consumed in the data center. Another major issue is the availability of data center floor space – the enterprise must improve the density of servers within the data center or face the never-ending dilemma of building more data centers. For a more complete list of next generation data center concerns, see Exhibit 1, above.

Exhibit 1 – Next Generation Data Center Issues

- Lowering TCO to meet a limited budget;
- Reduce data center floor space requirements;
- Improving hardware efficiency to reduce the amount of energy consumed for power and cooling;
- Increasing the usable density of servers in any given rack in the data center;
- Configuring a single point of management for thousands of open systems servers in a scale-out environment;
- Deploying a flexible infrastructure with a few standard servers replicated thousands of times to ensure application availability;
- Enabling an architecture that permits rapid scalability to meet the needs of cyclical Internet workload; and
- Installing an architecture that can be deployed and maintained around the globe.

Controlling power consumption in the data center through more efficient server utilization is another important issue from both a TCO and environmental standpoint. When you consider that the enterprise is spending a dollar to cool the data center for every dollar spent on powering the server environment, you realize the tremendous savings that are available from improved power management. In addition, with thousands, or tens of thousands of servers per megacenter, needed to satisfy the mission- and business-critical application mix, there is a tremendous need to increase compute density, in order to make better utilization of the data center's floor space. Unfortunately, many of today's existing server architectures prohibit the IT staff from completely filling any standard 19" rack, as there simply is not enough power available to drive the replicated infrastructure in the rack, with the typical data center topping out at between 8KW and 15KW per rack full of 1U and 2U servers.

In addition, the enterprise needs to be able to simplify this ultra-scaled IT infrastructure through the implementation of an open, reproducible architecture, enabling the CIO to deploy it around the

[1] Check http://www.clipper.com/Clipper_Server_Index.htm for the forthcoming issue of Clipper Notes entitled *Perceiving the Dark Side of the Moon – Knowing When Scale-up Computing Makes Sense*.

[2] Say, in increments of 500 or 1000 servers at a time.

world. Make that, deploy it around the world **and** retain control of the IT infrastructure back to the data center.

HP's Challenges

HP has made an overt decision to focus on the business goals of the largest data centers to insure success in a Web 2.0 or high-performance computing environment, including maintaining a "green" environment in an era of constant acquisition and IT globalization. Companies such as Google, Yahoo, and Amazon have been building their own massive *scale-out* infrastructures because of the lack of an available, simplified solution. However, there are other enterprises involved in financial services, oil and gas exploration, photo services, and online gaming, to name a few, who need to reduce cost and energy usage, and improve productivity, but lack the resources to do it. *HP has accepted this challenge.*

These enterprises have an insatiable demand for an ever-increasing supply of processor power, constrained not only by a limited budget, but also limited by a fixed amount of electrical energy available to the megacenter. Controlling the TCO of the IT environment is of paramount importance to them, not only by limiting capital expenditures, but also by implementing rigid controls on operating expenses. High availability is not an issue as there are sufficient available scale-out nodes to activate in the event of a node failure. Improving productivity efficiently is the key. Therefore, the data center can reduce some redundancy within each node in order to save energy and cost. Building an efficient platform that reduces the amount of energy required to power and cool the environment, as well as reducing the cost per watt, is vital. *HP has accepted this challenge.*

Another key contributing factor toward reducing the TCO of the megacenter is controlling the amount of floor space required to house the application servers. The data center needs to increase the density of their server deployment. However, as many CIOs have learned, increasing server density is dependent upon the ability to bring enough electrical power to the rack to drive this increased workload. As most data centers do not have an option to increase the amount of power to the rack, the data center staff must find a way to reduce the power required by each server within the rack. *HP has accepted this challenge.*

When deploying thousands of servers at a time, it is critical to the enterprise to implement a server environment that is easy to configure, buy, and own, with a rapid service delivery plan. The enterprise must simplify the data center infrastructure with preconfigured racks that plug and play with no negative impact on the data center environment. The enterprise needs a new and intelligent data center model to satisfy the need for explosive computing growth while lowering the TCO structure. *HP has accepted this challenge.*

In order to satisfy these requirements HP has integrated resources from around their company into a business unit dedicated to simplifying the mission of these enterprises – the *Scalable Computing and Infrastructure Business Unit*. With a stated philosophy of enabling the enterprise to blade everything, HP is taking a more holistic approach to data center management, from the design of the data center outward, including hardware, software, services, operational management, and power and cooling. This unit has not only changed the blade paradigm with the introduction of the *ProLiant BL2x220c G5* blade, but by integrating highly-scalable storage devices such as the *StorageWorks 9100 Extreme Data Storage System* (*ExDS9100*) and HP technologies, such as *Dynamic Smart Cooling*, to improve thermal management, and *Virtual Connect*, they can simplify deployment and management of the megacenter. HP has improved their server design by doubling the power capability and improving space utilization, while at the same time tripling the server density as compared to a typical 1U rack-mounted configuration.

HP's BL2x220c G5 Solution

Rather than reinventing the wheel, HP has relied upon technologies from across HP divisions, and the same principles that Gillette used in designing the Trac II cartridge; they combined two independent server blades together in a single blade enclosure. In doing so, they have helped the megacenter to lower scale-out data center infrastructure costs by doubling the performance in half the space of rack-mount solutions, enabling far greater energy efficiency. The new BL2x220c G5, based upon the architecture of HP's *c-Class* chassis, is ideal for scale-out applications taking advantage of the resources provided by thousands of servers that are deployed in Web 2.0 configurations and "clouds", as well as grids, compute farms, and HPC computing.

Using the Intel *Xeon 5400 Series Quad Core* CPU as a building block, or the *Dual-Core Intel Xeon 5200*, HP can package two half-height c-Class servers per blade, as a duplex, within each slot of an HP *ProLiant BL2x220c G5* chassis enclosure. This enables HP to package 32 two-socket server nodes within a 10U c7000 enclosure, or 16 servers within a 6U c3000 chassis. With this density, HP can configure up to 64 blades (128 servers), equal to 1024 Xeon 5400 CPU cores, within a single 42U rack, **occupying only eight square feet of valuable data**

center floor space. With the *E5450* running at 3.0 GHz and with 12MB of L2 cache, each CPU consumes up to 80W, a total of 5.12KW per hour[3] per enclosure at 100% utilization, or up to 20.48KW per hour per fully-configured rack. This configuration can deliver up to 12.3 teraflops of processing power, according to HP. HP can also configure the BL2x220c with a 50W *L5420* CPU running at 2.50 GHz. This solution will only consume 3.2KW per hour per chassis at 100% utilization, or 12.8KW per hour per rack, reducing power consumption significantly.

This configurability enables HP to deliver the highest density and performance per watt available in a blade architecture. With more compute capacity per square foot of data center than any standard blade or rack-mount solution, HP enables the Web 2.0 megacenter to provide their clients and partners with the performance they need, and with new low-energy quad-core Xeon CPUs, the reduced power consumption that they can afford.

In addition to this massive compute capability, HP has also configured each server node with a single 120GB, non-hot-pluggable SFF[4] SATA drive and Intel's low-power *5100MCH* chipset, with four DIMM sockets, enabling up to 16GB of low-power DDR2-667 DRAM per node, using 4GB DIMMs, with up to two terabytes within the rack[5]. This is sufficient memory capacity for the most demanding compute applications. Each server node also supports one X8 (PCIe) mezzanine slot for high-performance I/O supporting a 4Gb dual-port Fibre Channel interface, full bandwidth 4x DDR, InfiniBand, or 10Gb Ethernet.

The BL2x220c supports *Windows*, *Linux*, *Netware*, and *Solaris* operating systems and includes an embedded copy of HP's *Integrated Lights-Out 2* (*iLO 2*) for remote management for each server node. iLO also includes embedded power management and power capping to help control energy consumption. HP's *Thermal Logic Technology* provides an instant view of power usage and temperature for the server, the chassis, and the rack. It automatically adjusts power and thermal controls to minimize power and cooling requirements while maintaining adequate cooling for all devices to ensure availability. See Exhibit 2, in the next column, for Thermal Logic capabilities.

HP's *Insight Control Environment for BladeSystem*, an integrated suite of software that simplifies the provisioning and management of the BladeSystem infrastructure, is also available. It provides comprehensive system health, remote control, operating system deployment, vulnerability scanning, and patch management for the megacenter.

Exhibit 2 –
Thermal Logic Capabilities

- Active cooling fans;
- Parallel Redundant Scalable Enclosure Cooling (PARSEC) design for optimum cooling efficiency and lowered fan power;
- Instant power and thermal monitoring;
- Pooled power for a variety of power redundancy modes;
- Dynamic Power Saver mode;
- Power regulator; and
- Power workloads balancing.

Source: HP

Conclusion

The requirements of "cloud" computing are changing the way the next generation megacenter will operate. Limitations in the availability of both floor space and energy require significant improvements in scale-out server density and power efficiency, as well as storage management and data center services. HP has gone about this task with gusto to improve upon the power efficiency of existing HP racks and blade platforms, as well as those from leading competitors such as Dell and IBM.

HP has integrated technologies from across their enterprise to enable them to assume the mantle as a leader in massive scale-out solutions. They have implemented an architecture that enables the enterprise megacenter to deploy a configuration that requires no compromise between performance and configurability.

The BL2x220c G5 blade provides the performance and power efficiency you need, wherever you need it, whenever you need it, and however you need it. With preconfigured platforms that roll out and turn-on, HP blades shave the time for deployment, enabling improved productivity for all Web 2.0 or high-performance computing applications. If your enterprise is installing hundreds or thousands of servers at a time, you need to consider HP's technological innovation for megacenter deployment, the BL2x220c.

[3] Compared to 7.4KW per hour for competitive products.
[4] Small Form Factor.
[5] Four GB per server, using 2GB DIMMS, is standard.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

David Reine **is Director, Enterprise Systems for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. He joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group Captain's Log**, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.