THE CLIPPER GROUP
Navigator™

Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2012029 November 30, 2012

Location, Location, Location — HP Brings the Data Closer

Analyst: David Reine

Management Summary

Before the prospective home buyer goes hunting for a new abode, he or she makes a list of the most important features that they are seeking. That could include the number of bedrooms or the number of bathrooms; it could include cathedral ceilings in the living room; it might even include a secluded man-cave in which to enjoy some relaxing moments with friends, or perhaps, just some quiet time. When you talk to a trusted real estate broker, however, you find out that the three most important factors in real estate are*: location*, ***location,*** and ***LOCATION.*** This location could refer to the proximity of a school district, your place of work, public places like parks and libraries, or any one of a dozen other reasons. The value of a home is often directly connected to the "distance factor" – *how far are you from [fill in the blank]?*

Location is no stranger to the data center. For the past five decades, the IT staff of every data center has been in constant motion to determine the best location for servers and storage. When mainframes roamed the land, storage was directly connected to the processing power (DAS). Then processing power became distributed as we tried to get both servers and storage closer to the user, connecting multiple servers to each other via a networking and each server had its own silo of local storage. We then saw the value in connecting the storage closer to a set of servers, connected on a LAN, via a storage area network (SAN), creating an island, or multiple islands of storage. Inefficiencies in server support saw us move to physically consolidate these servers back into the data center, and then to logically consolidate multiple applications back into a single, open systems mainframe. All of these architectures were developed for the same goal – to improve the utilization of the IT infrastructure and to move the storage closer to the processing power, if not physically then in ways that speeded the transit of data. Now with the Internet and The Cloud, we often see storage moving farther and farther from the point of processing. Also, with the exponential growth of data for analytics and data preservation, we see "Big Data" consuming more and more of this crucial resource. Together, they represent a problem – the need to move more data, more frequently.

In order to overcome this "burden of distance" and, thus, to bring the Big Data closer to the processing unit, we are now seeing a new breed of server appearing on the data center radar – converged Storage. Converged storage is, in effect, a server purpose-built for analyzing Big Data. This new family of servers needs to be optimized for multiple workloads, with a balanced architecture to enable the IT staff to grow in two dimensions – nodes (filled with cores) for processing power and drives for capacity. One company that is paying attention to this growing need is HP, which has just announced their *HP ProLiant SL4500 family*, the first servers purpose-built for the processing of Big Data. To learn more about HP's ProLiant SL4500, please read on.

IN THIS ISSUE

Infrastructure Matters

We are now in the midst of an era of *data explosion*. More data gets created in one day, today, than had been created in years when mainframes ruled the earth. Enterprises of all sizes are yearning to turn all of this Big Data, much of it unstructured, into useful information that can be used to gain a competitive advantage or intellectual advances. All of that data must be collected and preserved, in many cases for decades, if not longer. As a result, the data center must find space to store all of the information vital to the operation of a modern enterprise, including floor space as well as the terabyte disk drives required to house this key resource. In many cases, the data center has not been efficient in the storage of this Big Data, in terms of both cost and capacity utilization. In order to rectify this, and restore some sanity to the enterprise IT budget, many IT staffs are attempting to break the decades-old paradigm of storage being assigned to a DAS silo or a SAN or NAS island. The design of the application architecture, or infrastructure, is critical to the execution of your mission- and business-critical application servers.

One alternate solution that has been proposed for the preservation of mission- and business-critical data is the Internet and *The Cloud*. Some major enterprises have established private clouds to act as a library of sorts for the collected information that is vital to enterprise operations. Other enterprises, many in the SMB category, are choosing to go the route of the public cloud, utilizing the services of providers, such as Amazon Web Services (AWS) with their *S3* and *Glacier* data storage alternatives. While perhaps inexpensive to deploy, these services may not meet the SLA required by many enterprises, in terms of timely retrieval and use. Response time can be critical to the enterprise depending upon that information for critical operations. The distance between the data and the application server is significant as *distance matters*[1] in establishing the cost and/or effectiveness of operations.

Delays in the availability of data, either for active operations, backup, archive, or disaster recovery, can be crucial to the success or failure of the enterprise. Any delay in data retrieval can be expensive, especially when that data is required to do some specific business or statistical analysis, or "analytics". The costs of transporting data to the application server can grow, depending upon the **stability** of the data, the **volume** of the data, and the **frequency** of retrieval requests.

- ***Frequently changing data*** can result in the creation of, and transportation of, multiple copies, or snapshots, of rapidly changing data, i.e. a moving target.
- ***The transportation of Big Data***, think terabytes, can have a serious impact on the enterprise network, especially one that is being used by many applications and even more users. Moving data for a long distance is never easy and often expensive, based upon the cost of your bandwidth. The faster that you have to move it and the bigger it is, the more expensive it gets. It may simply take too long to move the data to where it is needed.
- In addition to the speed is the ***frequency of use***. Every time that you send the data over your network costs you money, and potentially more. Delays in the receipt of that data may force you to use an "out-of-date" version of the data, even if that version is just minutes old. Sometimes, it simply takes too long to move the data through the network.

If in fact *distance matters*, then putting data nearer to the server cores has become a prerequisite in the processing or analyzing of Big Data because distance does matter. Putting this data somewhere on the Internet where it could be thousands of miles away could be counter-productive in terms of both time and cost. However, moving data over lesser distances matters as well. Consider the difference between moving data within a data center between appliances, measured in feet or meters, as opposed to moving it within an appliance, say measured in inches or millimeters. The length of time and the cost of that time matters, especially if it is you who is doing the waiting or footing the bill, or worse, making decisions based on outdated data.

Data Center Convergence

The question that we all must answer at this point is: *Why are we moving all of this Big Data at all, let alone across the country or around the world?* It is now, and always has been,

[1] For more on why distance matters, see the November 6, 2012, issue of Clipper Notes entitled *Bringing "the Power" Closer to Home - Introspections for Future Storage Architectures*, which is available at http://www.clipper.com/research/TCG2012026.pdf.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

intuitively obvious that the closer the data is to the processing, the faster you will get your response, be it for transaction processing, ERP, or analytics.

Former Congressman Tip O'Neill once said: *All politics is local*. Well, the same should be said for data processing, or at least, could be said, if the data in question is close to the processing source. This is especially true for any processor and I/O intensive applications where multiple passes through the data may be required, even more so for Big Data. If you need to **process** or **analyze** your data locally, *it only makes sense if your data is closer to your servers*. Why move the data? **Process it near where it resides.**

The consolidation of the data center has led to the facilitation of both storage arrays and highly virtualized, multi-processor servers connected within the same data center by a high-speed Fibre Channel or Ethernet network. This network usually is shared by many applications, placing a heavy burden on the bandwidth and other data center resources. Why isn't the next step in this consolidation process the convergence of storage and processing within the same appliance to further reduce the distance between them? This would isolate the communication paths in order to hasten the transportation of the data to the processing cores. No matter how many cores you have to process the data into information, they still have to wait for the data to be transported to the server through a network controller, and then through the bus to the cores. Even though this is a very small distance, it often forces the CPU to wait. If the IT staff can keep the servers busy by increasing the speed of access to the data, the applications, and therefore, the decision makers, transaction processing, analytics, and ERP applications can be that more effective. The big question then becomes: *How do we do that? How do we remove the bottlenecks and get the data closer to the processors?*

HP is one company that may have the answer to our questions with their *ProLiant SL4500 Server*, a server designed specifically to handle the Big Data proliferating and being analyzed throughout the enterprise.

HP's ProLiant SL4500 Server

In order to optimize Big Data and manage all of this information more efficiently, HP has designed the newest member of the ProLiant family, the ProLiant SL4500 server, with a converged and balanced architecture to change the paradigm of the IT infrastructure. **It was designed to be a best-in-class server to optimize multiple, heterogeneous workloads to drive new business opportunities.**

Designed to support scalable performance while reducing the total cost of ownership (TCO) of the IT infrastructure through a very dense design, the ProLiant SL4500 reduces floor space requirements, consumes less power, and utilizes far fewer cables. HP's *Smart Array* technology provides the flexibility to balance cost and performance across a variety of drive types, including SAS, SATA, and SSD. Each *ProLiant SL4540* node comes with up to two 2.4 GHz Intel *Xeon E5-2400* processors; each *ProLiant SL4545* node comes with up to two 3.3 GHz AMD *Opteron 4200* processors. Both can be configured with up to eight cores in each CPU and 192GB of memory per node. Both SL4500 models can support up to 240TBs in a 4.3RU chassis or 2.16PBs with nine servers in an industry-standard 42U rack. HP provides a one-year on-site warranty for both models.

The modular design of the SL4500 provides a multitude of compute and storage configurations to enable the IT staff to optimize the infrastructure for each workload-specific application. This eliminates the need for additional racks, switches, cable, etc. in a supporting infrastructure. With one architecture, the ProLiant SL4500 supports *Hadoop* vendors such as Cloudera and Hortonworks, as well *as additional software including OpenStack Cloud Software* and *MongoDB*. This results in the ability of the SL4500 to reduce the rack count of a comparable previous generation configuration from seven to about three, reduce the switch count from 14 to seven, and reduce the cable count from 448 to 168. In an example cited by HP, this helped to reduce the TCO from $3M to $2M over three years, including a reduction in energy consumption from 79KW to 31KW.

Using the HP ProLiant SL4500 server with an SSD, HP can deliver seven times the performance of a previous server with a 15K SAS drive. Combined with the intelligent analytics provided by HP's *SmartCache*, the SL4500 can optimize storage traffic to provide the lowest latency response and investment. HP's *ProActive Insight Architecture* provides embedded intelligence and automation, and the following benefits.

- ***Eliminates downtime and preserves valuable data*** with automated data protection through

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

Exhibit 1 — Three Models of HP ProLiant SL4500 Gen8

Large-storage-capacity one-node configuration with 60 drives

Balanced core-to-spindle ratio configuration with two nodes, each with 25 drives

Balanced core-to-spindle ratio configuration, with three nodes for high-performance applications and 15 drives for each node

Source: HP

HP's *Advanced Data Mirroring* and controller-based encryption, with the capability to move data to an alternate device through HP's *Predictive Spare Activation*.

- ***Ensures maximum server productivity*** through HP's Active Health, and automate firmware updates with HP's *Smart Update*.
- **Leverages HP's most comprehensive services, support, and warranty offering** with HP's *Insight Online*.
- ***Lowers data center power costs and improve compute per watt*** by up to 70%, when compared to the previous generation, with HP's Intelligent Infrastructure.

The ProLiant SL4500 comes in three basic configurations of one-, two-, or three-nodes, as shown in Exhibit 1 to the left.

- ***The single node comes*** with 60 drives and is ideal for applications requiring a large amount of storage.
- ***The dual-node configuration*** is best used for solutions such as email and data analytics, applications that require a more balanced CPU to storage ratio. This configuration comes with 25 drives for each node.
- ***The three-node configuration*** has been designed for parallel processing applications such as *Hadoop*, also requiring a balanced CPU to storage ratio. This configuration comes with 15 drives allocated per node.

Conclusion

The HP ProLiant SL4500 provides the enterprise data center with a solid foundation for future innovation involving Big Data. With workload-optimized solutions, a converged and balanced architecture and integrated applications management, the SL4500 helps to remove the guesswork from deploying Big Data solutions.

Moving data through ever-increasing pipes with higher bandwidth carries a high cost; waiting for terabytes to be moved might even carry a higher cost, in terms of inaccurate analyses or incomplete decision-making. Placing your enterprise data in a ProLiant SL4500 storage server simply makes sense from both a performance aspect and a concern for the TCO of your IT infrastructure, as your applications' appetite for consuming data increases.

When you look into your crystal ball to predict the future, do you see more and more data required to give your enterprise a competitive edge? If the answer is yes, then you need to review the advantages of an HP SL4500 for your enterprise. Do it now, before it is too late!

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twentieth year, is an independent publishing and 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 a Senior Contributing Analyst for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, 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.

Disclosures

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.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

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.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.