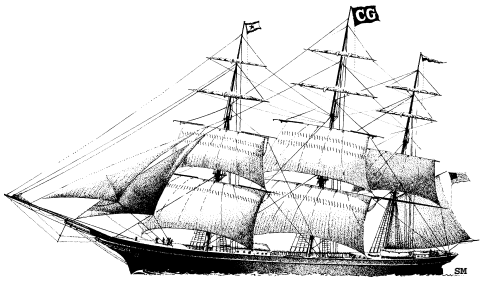
THE CLIPPER GROUP
Navigator™



Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2008043 September 11, 2008

NEC Brings Vast Scalability to Tier-2 Storage

Analyst: David Reine

Management Summary

Whenever you go shopping for a present for a young child, you are faced with a myriad of options. You can go with a hot fad, such as a *Hula Hoop*, *Cabbage Patch Kid*, or *Beanie Baby*, or, if you have a creative child, you can invest in an imagination-building toy such as *Legos*. These little bricks can be snapped together in an instant to form whatever shape a child's mind can imagine, growing in size with the child. Available in starter kits for just a few dollars for the youngest hands, these bricks can be combined with other, more advanced sets to build more complicated structures as the child grows into his teenage years. The Lego brick became popular four decades ago, and those bricks sold in 1963 still interlock with kits made in 2008 – because of a common interlocking architecture. In addition to a variety of bricks of all sizes, there are motors, gears, sensors, and mini-figures, an entire infrastructure of components to combine, reshaping designs to meet an evolving imagination. Further, the Lego Company does not stand still; they continue to create new kits with new elements, with those elements compatible with bricks from their parent's youth. Other toy manufacturers have copied the lego concept, introducing scalable toys with a common infrastructure that fit together in a virtual plethora of ways, to provide hours of enjoyment. This concept can be adapted to the enterprise data center, in an attempt to create a scalable architecture to simplify and protect a complex IT environment, reducing the total cost of ownership (TCO) of the data center.

The enterprise data center has been engaged in a full-scale assault against the inefficiencies created by the massive scale-out of x86 open systems architecture platforms, with many servers utilizing only 15% to 20% of their processor capability, while trying to cope with an exploding storage environment that is doubling in size every 12 to 18 months. In order to improve the utilization of the server network, and reduce the TCO of the IT infrastructure, the data center is consolidating multiple thin servers onto fewer multi-core, multi-sockets platforms, virtualizing a number of applications on a single platform, improving the CPU utilization while at the same time reducing the amount of floor space required. Most importantly, this reduces the amount of energy needed to run the infrastructure and cool the environment. Unfortunately, this also increases the complexity of the enterprise storage architecture, as the IT staff has to consolidate an ever-increasing number of storage islands into a single pool of shared resources, complicating mission-critical application deployment and the adoption of a Web 2.0 operating environments, as well as creating a need to simplify business-critical applications such as backup and archive. The enterprise data center has to consolidate their storage environment in order to reduce the TCO of the IT infrastructure.

Last year NEC introduced us to *HYDRAstor*, a scalable grid storage solution designed to provide the data center with a high-performance, energy efficient Tier-2 storage platform aimed at consolidating backup and archive targets. Now NEC has improved the performance and scalability of grid storage to deliver a denser, more energy efficient platform. To learn more about HYDRAstor, please read on.

IN THIS ISSUE

Enterprise Data Center Issues

Two of the highest profile issues that have put enterprise data centers in a state of crisis today are server sprawl and the explosion of storage requirements in support of mission- and business-critical applications. Because of inefficiencies in the deployment of mission-critical servers in a scale-out environment, these servers often utilize only 15% to 20% of the systems' computing power. At the same time, however, they are consuming 100% of the energy required to run these distributed platforms, putting a strain on the TCO of the data center. In order to improve these statistics, the data center staff implements programs to both consolidate multiple servers onto a single, multi-core, multi-socket platform, designed with a scale-up architecture, and also to virtualize multiple applications within these systems in order to better utilize previously-wasted compute cycles.

At the same time, the data center is experiencing tremendous growth in the quantity of information that the enterprise requires to remain competitive in a very tough economy. **Most data centers are experiencing a 50 to 70% annual growth while some are experiencing a doubling of mission- and business-critical information being saved on an annual basis, not to mention additional incremental growth resulting from mergers and acquisitions and new requirements needed to meet regulatory compliance.** Some of this growth is due to the deployment of new Web 2.0 applications, while backup and archive files make up a substantial amount. Consolidating and managing this data and protecting it from theft and/or accidental loss have become a full-time job in order to cope with the complexities of a virtualized data center. Capital costs are not the issue, as the acquisition of storage, on a per TB basis, is going down. TCO, however, is another matter.

As storage capacity rises, the number of devices rises proportionately, even considering increasing device capacity. The reliability and the availability of any physical device may be the same, or even improving, but the overall expectation for any disk farm is that the data center will experience more failures as the number of devices increases. **It is not a question of "if a disk device will fail," it is the realization that *multiple* disks will fail**. Today, most organizations are deploying disks with a capacity of up to 1TB with RAID protection. When a TB disk fails, the rebuild usually can take upwards of 10 to 12 hours, degrading performance on all of your business- and mission-critical applications. Moreover, we can realistically expect disk devices to continue to grow in capacity to 2TB, 4TB, and more. **Rebuild times are not going to get shorter anytime soon**. In order to protect this data, the IT staff has implemented any number of safety nets, including the implementation of RAID technology to ensure the integrity of your data. Unfortunately, the risk of a second device failure is so great now that data centers have been forced to take the extreme action of deploying a RAID-6 architecture to guard against that eventuality. With RAID-6, the IT staff must configure a second parity device, effectively *reducing* array capacity and performance, thus *increasing* the TCO at the same time. If your existing storage architecture can support these technologies, what will be the impact on your array environment?

In addition to the impact on reliability and availability, what is the effect of uncontrolled data growth on other TCO issues? The costs for floor space, administration, maintenance, and energy continue to rise at an alarming rate. At the current rate of growth, in five years, you could be managing six-to-eight times as much data. Does the enterprise have enough floor space in the data center to support that growth? Will the enterprise have to build a new data center at significant cost, i.e., millions of dollars? How many new administrators will you have to bring onboard to manage it? What will be the impact on the maintenance budget to support the added devices? Moreover, **will the local public utility even be able to supply the data center with enough energy to keep all of those drives spinning**, let alone trying to budget for it? Depending on where your data center is located, you could be paying anywhere from $.05 per KWh to $.18 per KWh, *today*. If the projected price of gasoline is any indicator, your energy costs might double or triple in the years to come. Even though capital expenses are going down, operational expenses are going up even more. The data center will have to do even more with even less. In five years, **how much storage will you have?**

Innovation in disk technology has not kept

up with innovation in other areas of the IT infrastructure. While there have been significant developments in server microprocessor technology, we have not seen a commensurate advance in disk technology. While processor performance increases dramatically year over year, with multiple cores and multiple threads, along with faster clock speeds, disk device capacity inches upward, up to 1TB in 2007 and 1.5TB later this year or early next year. However, the storage architecture has remained the same for several decades, with basic RAID technology remaining the same for the past 30 years. Today's data center demands a higher degree of availability to enable a faster recovery.

Another area of concern is scalability and the protection of the data center's storage investment. What is the scalability of the existing storage solution? If it cannot grow to meet tomorrow's needs, you must plan today to replace it with a solution that can. Moreover, as enterprises continue to evolve to a D2D architecture for applications such as backup and recovery, the IT staff will have to change the storage paradigm in the data center completely. The data center also needs to improve the reliability, performance, and ease of use of its storage with a scalable solution that will meet the long-term requirements of the enterprise. The data center needs to deploy an innovative solution in the realm of storage to the same extent that they have with their server technology.

The NEC HYDRAstor Solution

Last year, NEC changed the paradigm for the management of secondary storage with the announcement[1] of a new grid storage system in April, and the subsequent introduction of the *HYDRAstor HS Series*[2] in September. HYDRAstor is more than just the integration of a bunch of high-capacity SATA drives with easy-to-use management software. If that is what you are looking for, there are any number of storage start-ups out there who can provide a standard commodity solution for your backup/-recovery and archiving needs. HYDRAstor is an innovative platform designed by NEC to provide highly reliable, non-disruptive scalability to the enterprise data center. With NEC, you get a mature Fortune 200 technology innovator with more than 100 years of experience and over 48,000 patents to its credit. HYDRAstor is the result of more than four years of research, development, and validation within real data center environments, solving real enterprise problems. NEC designed HYDRAstor to ease infrastructure complexity and provide virtually unlimited capacity, relieving IT staff of an excess of management overhead. Futhermore, HYDRAstor was designed as a self-evolving platform that continues to ride the technology curve and evolve over time, while protecting the investment in earlier infrastructure.

First, HYDRAstor is a specialized system designed to store very large quantities of Tier-2 data in environments experiencing high levels of data growth, in the order of two-times growth every 12 to 18 months. It has a two-tier grid architecture consisting of two types of standard Intel server nodes, *Accelerator Nodes (AN)*, for scalable performance, and *Storage Nodes (SN)*, with SATA drives, for scalable capacity. This enables NEC to scale performance and capacity independently, enabling the IT staff to configure an enterprise solution specific to any data center's requirements. NEC tested the initial release of HYDRAstor with up to 420 nodes, 140 ANs and 280 SNs, achieving an effective throughput of 14,000 MB/second and a capacity of 10PBs using data deduplication. Obviously, a different mix of nodes would achieve entirely different results, enabling the IT staff to create a HYDRAstor configuration tailored to changing enterprise requirements.

NEC introduced an advanced set of storage management technologies with the initial HYDRAstor release, including *DynamicStor*, *DataRedux*, and *Distributed Resilient Data* (*DRD*). DynamicStor enables the IT staff to add or remove nodes, either ANs or SNs, without disrupting user access to the data and to distribute the workload over all of the nodes to maximize performance and resiliency. Even though some companies would have you believe that *data deduplication* is the newest technology to hit the data center, NEC has been

[1] See **The Clipper Group Navigator** dated April 18, 2007, entitled *NEC's HYDRAstor Tames the Multi-headed Monster*, which is available at http://www.clipper.com/research/TCG2007052.pdf.

[2] See **The Clipper Group Navigator** dated September 14, 2007, entitled *NEC HYDRAstor – A Leap Forward in Online Backup and Archive*, which is available at http://www.clipper.com/research/TCG2007088.pdf.

using DataRedux to eliminate duplicate data globally, at a sub-file level, across all data from all nodes for over a year. In order to achieve maximum storage efficiency, DataRedux examines data inline, before it is stored on disk, to eliminate duplicate data and then compress it to reduce storage capacity by as much as 95%, or more, assuming a 20:1 space reduction ratio[3]. One of the unique advantages of HYDRAstor's grid architecture is that it enables global data deduplication, ensuring only unique data is stored once across the entire datacenter. Regardless which AN or which file system the data comes from, it gets deduplicated across the entire common pool of storage on the backend to ensure only unique data is stored. In addition, the deduplication processing is automatically distributed among all the SNs, so each additional SN continues to scale the deduplication-processing engine to maintain high performance for larger grid configurations. This not only reduces acquisition costs, but it lowers the entire TCO of storage by reducing power, cooling, space, and environmental costs. NEC's DRD technology enables greater data resiliency and higher availability. DRD provides default protection to support a three-disk failure, enabling 300% more protection than for RAID5 with similar storage overhead.

HYDRAstor is a low-cost solution initially designed to provide a scalable resource for backup and archive processes, but it also provides an ideal environment for the development of a Web 2.0 environment in support of social networking, streaming media, or photo web sites, scaling performance and capacity, but not administration. With HYDRAstor, administrators can manage 10s of PBs of data, enabling the enterprise at least to start to control their storage budget woes. Now, only one year after introducing the *HYDRAstor HS8-1000* Series, NEC is back, this time with the *HS8-2000* with three-times node performance and 4-times node capacity.

Introducing HYDRAstor HS8-2000

Continuing their theme of "one system, forever," NEC has taken the next step towards evolving HYDRAstor to enable even more capacity and more performance than was available in the first generation, with up to 12TB of raw capacity in a single Storage Node and a performance of up to 300MB/second in a single Accelerator Node. These newer nodes can be added to expand existing older generation grids, immediately making them faster, denser, and greener while preserving the investment in first generation infrastructure. Similarly, the new nodes can be used to build entirely new HYDRAstor storage grids based on the latest technology. An entry-level system based entirely on the latest nodes, *HS8-2002S*, with one AN and two SNs has an effective capacity of 315TB, with a throughput of 300MB/second. Using NEC's standard modeling, this means that an *HS8-2110R*, configured for high performance for backup with 55 ANs and 110 SNs, can achieve an effective capacity of 17.3 PBs with a throughput of 16,500MB/second. Alternatively, using HYDRAstor's independent linear scalability, in high-capacity archive mode with 33 ANs and 132 SNs, HYDRAstor can support 20.8 PBs with a throughput of 9,900MB/second. Moreover, this new evolution of NEC's grid technology is not at the expense of existing data center HYDRAstor infrastructure: the IT staff simply can add new nodes to the existing architecture preserving the investment that the enterprise has made in NEC technology.

Where each HS8-1000 Storage Node could support an effective capacity of up 37.5TB, each HS8-2000 node can support up to 157.5 TB, more than a four-fold increase. NEC has also improved the energy efficiency of the SN, reducing the power consumption from 11.5 Watts/TB down to 3.8 Watts/TB, cutting energy consumption, and energy costs, by 67%. With the inclusion of 1TB SATA drives, NEC reduced the HYDRAstor footprint, enabling the deployment of 78.8 TB per rack unit (RU), up from 18.8 TB per RU. This is, once again, greater than a four-fold increase.

In terms of improvements to the Accelerator Node, NEC is using the latest 3GHz quad-core Intel Xeon technology, along with additional, faster memory, more Ethernet ports, and software enhancements, to help increase performance from 100MB/second to 300MB/second, a three times improvement, increasing the number of communication ports from two 1GbE to six 1GbE or two 10GbE. They have also driven down the energy consumption of the AN, although reducing the power requirement from

[3] As always, your mileage may vary!

4.3 Watts/MB/s to 1.3 Watts/MB/second may not have as significant an impact on the enterprise's energy budget, it still is a 70% reduction. The new AN also reduces its footprint in the data center with 150MB/second of throughput per RU as compared to 50MB/s for the HS8-1000. In keeping with NEC's philosophy to protect the enterprise investment in NEC systems, all HYDRAstor software applications continue to support all generations of both ANs and SNs to ease the transition for the IT staff.

Conclusion

Transitioning to the latest version of HYDRAstor is a no-brainer: there is no need to migrate or lose data, there is no application downtime, and there is no need for a forklift upgrade. HYDRAstor HS8-2000 provides the enterprise with maximum investment protection, extending the life of first generation technology, and increasing storage scalability from a data pool to a veritable ocean. Because of that scalability, there is no need for over-provisioning at deployment; the data center can add either capacity or performance, or both, on an on-demand basis.

HYDRAstor provides the enterprise with a greener, faster (deployment and performance), denser storage solution, while decreasing storage costs by greater than 40%. It improves capacity and management efficiencies while eliminating multiple silos of deduped data, presenting one system for the IT staff to manage, rather than multiple islands.

If your IT budget is being threatened by uncontrolled storage growth, look at NEC's HYDRAstor to achieve the level of scalability required by your enterprise, and regain control of your bottom line.

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.