



NEC HYDRAsstor – A Leap Forward in Online Backup and Archive Storage

Analyst: Michael Fisch

Management Summary

NEC's new HYDRAsstor grid storage system represents a leap forward in how enterprises can store data for backup and archive systems. Most notable is the efficiency and ease with which it stores large amounts of this secondary data online. A unique combination of technologies undergirds HYDRAsstor, making it a challenge to describe with a simple phrase or sound bite. However, with some imagination and artistic license, you might compare HYDRAsstor to:

- **An expandable cargo ship equipped with a shrink ray** – Not only can the ship stow a huge amount of cargo in miniature, but also it can grow as needed to fit more. HYDRAsstor is similarly sizable and expands by adding discrete nodes of either performance or capacity. It “shrinks” data by eliminating redundancy and applying compression, attaining typically a 20:1 de-duplication ratio after a month.
- **An inflatable rubber balloon filled with compressed gas molecules, no two alike** – Not only are the gas molecules packed tightly in a stretchable container, but also they are different from one another, so there is no redundancy. A swelling balloon is like HYDRAsstor's ability to scale up or down non-disruptively by adding or removing nodes. This process is automatic and does not interfere with data access. The gas molecules are like the compressed, de-duplicated data stored inside HYDRAsstor.
- **An armored airplane that can withstand multiple hits** – Unlike ground vehicles, armor protecting an airplane must be light and minimalist or it will not fly. Yet it must be effective because a failed airplane does not merely roll to a stop like a ground vehicle. Similarly, HYDRAsstor's data protection mechanism can withstand multiple disk or node failures (the number is user adjustable) without data loss, yet only requires a 25% capacity overhead for the default protection level against three node failures. It also continues to fly after a hit because drive rebuilds do not detract from access performance. So, HYDRAsstor's data protection is designed for robustness and efficiency.

While these analogies are admittedly far-fetched, they highlight HYDRAsstor's unique combination of capabilities. It connects to host servers over Gigabit Ethernet and supports a growing list of backup and archiving applications. The system itself is built from standard Intel server and SATA drives, and the intelligence is in the software. Enterprises can start relatively small with an entry platform that holds about 150 TBs and scale to multiple PBs over time.

If your enterprise needs disk storage for backup and/or archive data, read on for details about HYDRAsstor's new approach.

IN THIS ISSUE

➤ Need for Secondary Storage	2
➤ NEC HYDRAsstor	2
➤ Conclusion	3

Need for Secondary Storage

The emergence of HYDRAsstor follows in the wake two important data management trends:

- Backup to disk
- Active archiving of application data

Both employ so-called secondary disk storage, which is less expensive than primary disk but is not quite as fast or robust. These systems are designed for low-cost, high-volume storage. They typically use SATA drive technology instead of Fibre Channel or SCSI drives.

Backing up to disk, as opposed to tape or as a first stage prior to tape, is increasingly popular because its advantages. Backup to disk is faster, so jobs can finish in time and stay current. This is a serious issue, because data growth and lower tolerance for planned downtime have cramped backup windows. Backup jobs do not always complete in time, which puts data at risk. Recovery from disk is also faster than tape, especially if tapes are archived offsite. Therefore, business operations can resume more quickly in the event of a system failure or disaster. If an enterprise wants to use tape for long-term storage, backing up to disk first allows it to send significant less data on to tape, which conserves media, transportation, and storage costs.

Active archiving of application data, including e-mail, files and database records, is another growing phenomenon. Pruning data from primary storage systems keeps them lean and mean, so applications are faster, backups and restores complete more quickly. Overall storage costs are reduced when inactive data is placed in secondary storage. Archiving is a way to comply with industry regulations, such as HIPAA requirements for patient record retention requirements, as well as to support legal discovery requests.

NEC HYDRAsstor

NEC HYDRAsstor is a specialized platform designed to store large amounts of secondary data. It offers a unique set of capabilities.

Non-disruptive Scalability to Multiple Petabytes

HYDRAsstor employs a grid architecture that scales non-disruptively from small to large configurations. The grid consists of *Accelerator Nodes* for performance and *Storage Nodes* for capacity. (See *HYDRAsstor Nodes*, at top of next page.) Customers can scale performance and

capacity independently. The nodes are built from standard Intel servers and SATA drives and run NEC's *DynamicStor* software. HYDRAsstor distributes workloads and data among the nodes to maximize both performance and resiliency. Nodes can be added or removed without disrupting data access by users and applications.

An entry configuration has two Accelerator Nodes, four Storage Nodes, and stores 150 TBs of effective capacity. A large configuration has 48 Accelerator Nodes, 96 Storage Nodes, and 3.6 PBs of effective capacity. Customers can also create configurations in between these two and beyond – all the way to 10 PBs and 14,000 MB/s in aggregate performance.

Data De-duplication and Compression

In contrast to physical goods, having many instances of digital data is not necessarily good. If you had an abundance of aluminum ingots, diamonds, or automobiles, you would be considered wealthy! However, if you had an abundance of the *same* digital data (i.e., multiple copies), you would be wasting space and spending more than necessary on storage. This is easy to do since backup and archive data tends to be repetitive.

HYDRAsstor automatically handles this problem by de-duplicating and compressing data, a feature it calls *DataRedux*. It checks new streams of data for repetitive blocks and inserts pointers to identical blocks already stored on the system. Then it compresses the remaining unique data. In real-world backup environments, this results typically in 20:1 reduction or better after approximately a month. In other words, enterprises using HYDRAsstor only have to store effectively 5% of their original data.

Distributed Resilient Data

HYDRAsstor's *Distributed Resilient Data* (DRD) technology ensures de-duplicated data is resilient and safe. DRD breaks data into fragments, creates parity fragments, and places them on different disks and nodes. In this way, data can survive multiple, simultaneous disk or node failures. This is especially important in a de-duplicated environment because one data loss incident can have a multiplier effect that causes hundreds or thousands of files to be irretrievable. HYDRAsstor allows users to select the number of failures the system can tolerate and supports protection levels far beyond today's RAID 5 or 6 capabilities. The default of three delivers a protection level 300% greater than RAID 5 with

NEC HYDRAsTOR Nodes

Accelerator Nodes

Model	AN-10
Performance	100 MB/s
Connectivity	2 Gbit/s Ethernet
Access protocol	NFS, CIFS
Cache	6 GB
CPU	2 Intel Xeon 5150

Storage Nodes

Model	SN-250	SN-375
Disk drive	500 GB SATA	750 GB SATA
Capacity (raw)	2.5 TB	3.75 TB
Capacity (effective)	37.5 TB	56.25 TB
Cache	6 GB	6 GB
CPU	2 Intel Xeon 5160	2 Intel Xeon 5160

Source: NEC

only 25% capacity overhead. Since drive failures will occur at some point, HYDRAsTOR is designed so that performance does not slow during drive rebuilds. It reserves CPU overhead for these processes and distributes the re-build over the grid. In the first quarter of 2008, NEC also plans to release a replication feature for disaster recovery purposes.

Self-Management

HYDRAsTOR fundamentally manages itself. It automatically discovers new nodes, moves data, provisions capacity, and load balances. This minimizes the burden on IT administrators and helps save on management costs.

Application Support

The backup and archiving applications supported by HYDRAsTOR include:

- Symantec *NetBackup*
- Symantec *BackupExec*
- IBM *Tivoli Storage Manager*
- EMC *Networker*
- HP *Data Protector*
- Symantec *Enterprise Vault*
- EMC *DiskXtender*

Application servers connect to HYDRAsTOR over Gigabit Ethernet.

Conclusion

For enterprises that need online storage for backup and archiving data, consider your options carefully because this secondary storage element will be a major part of your infrastructure. It is worth finding a solution that will meet your long-term requirements smoothly and cost-effectively.

NEC's new HYDRAsTOR should be on your short list for consideration. NEC has clearly put significant thought and development into this purpose-built and streamlined platform for secondary storage. Its combination of easy scalability and efficient data storage and protection are what enterprises need for this category of data.



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

Michael Fisch is Director of Storage and Networking for The Clipper Group. He brings over ten years of experience in the computer industry working in sales, market analysis and positioning, and engineering. Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems, Inc. Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

- ***Reach Michael Fisch via e-mail at mike.fisch@clipper.com or at 781-235-0085 Ext. 211. (Please dial "211" 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.