



THE ONLINE STORAGE NETWORKING MAGAZINE FOR IT LEADERS

What's New



ARTICLE POSTED May 30th, 2005

"Grid" storage is in the eye of the beholder (and vendor)

By Robert L. Scheier

Greg Bosworth, Jess Carruthers and Mike Luter are all testing or deploying grid storage to hold down costs, streamline backups and restores or make storage administration easier.

But none of them went shopping for grid storage per se. To them, grid storage is a useful technology for solving specific problems but not a new product or technology in its own right.

"It doesn't matter to me if you hook it together with glass or copper, or if it's red or blue," says Carruthers, who last year consolidated 10 Network Appliance 800 Series filers into a single FAS960 cluster running NetApps' Data ONTAP 7G storage software. "The functionality is what you need."

As vendors promote differing visions of "grid," current customers and analysts urge users to evaluate products based on the benefits they deliver today. Among the benefits customers should look for is how much a grid architecture will reduce the upfront costs of buying storage, how it will ease backups and restores, how it will make it easier to reallocate storage as needed, and whether the grid can support both block- and file-level access.

The storage grid

Grid storage is an architecture in which independent storage nodes are linked and governed by common control software. That control layer provides a single management interface and fault tolerance among the nodes, as well as the ability to access either file- or block-level storage. It also makes it possible to easily or even automatically reassign nodes to different functions, such as from online to archival storage, as needs change.

"As you need capacity, you add a node to the grid, and they self-configure and become part of an entire pool of storage," which greatly reduces storage management costs, says Stephanie Balaouras, a senior analyst at the Yankee Group.

That's the ideal. In the real world, different vendors offer different grid capabilities, sometimes packaged as "clusters" or "virtualized" storage. Virtualization means managing separate physical storage pools as one virtual pool, which is often one of the capabilities provided by grid storage.

In a storage cluster, the nodes are networked as in a grid, but they all generally support the same application or function. This is different from a grid, where each node can ideally be reconfigured for different purposes. Then there is storage for grid computing, in which the storage is not necessarily linked in a grid but is optimized for processors linked in grids for high-end scientific and technical applications.

Various vendors use combinations of these technologies to deliver the flexibility, manageability, performance and data protection promised by grid storage.

Different vendors, different grids

If storage nodes linked by a network sounds an awful lot like a SAN to you, you're right, as SAN vendor EMC is quick to point out. "You could argue that Celera, which is a cluster of file servers, is a grid," says Rick Strom, director of NAS product marketing for EMC. While EMC is paying close attention to evolving grid architectures, Strom says the company is trying to figure out what problems the architecture solves rather than market grid technology as a benefit in and of itself.

But others argue that the grid-like capabilities in a SAN are usually built using proprietary hardware and software, which makes the resulting products more expensive than they need to be. "Our view of grid is something that is much less monolithic than, say, an EMC Symmetrix," says Jeff Hornung, vice president and general manager of the Gateway Business Unit of NetApp.

NetApp moved into grids with the acquisition of Spinnaker Networks last year and now sells Spinnaker's storage grids as a standalone product. In November, NetApp brought grid-like capabilities to its storage operating system with the release of Data ONTAP 7G. Late this year or early next year, says Hornung, NetApp will merge the Spinnaker grid capabilities with ONTAP 7G into a single converged product line.

NetApp is well-positioned to move into grid storage because its technology supports both file- and block-level access and uses a distributed file system, which will provide a common virtualization layer for all storage in the grid, says Balaouras. Other vendors taking a similar distributed file system approach include Isilon Systems and Exanet, with 3PARdata the strongest player on the pure block access side, Balaouras says. Dell is reselling IBRIX's IBRIX Fusion software suite, a highly-scalable parallel file system and logical volume manager, for its HPCC, Scalable and Cluster Computing products.

IBM, EMC and Hitachi are not stressing grids as much as the virtualization a grid-like architecture can deliver. IBM sees grid storage as only one of many capabilities needed to provide the many-to-many relationship customers want so that any application or user can have access to any computing or storage resource they need, says Tom Hawk, general manager of enterprise storage at IBM.

According to Ken Wood, senior director of data and storage technologies at Hitachi, "We don't market the TagmaStore USP (Universal Storage Platform) as a storage grid or grid storage to avoid user confusion. What matters is what customers need, not what's inside."

Senior analyst Tony Asaro at the Enterprise Strategy Group in Milford, Mass., says NetApp has done a better job than IBM, EMC, or HP of articulating and executing a grid storage strategy. However, adds Asaro, vendors including 3Par, LeftHand Networks, Compellent Technologies, Isilon, EqualLogic, **Intransa**, Panasas, Exanet and ONStor "all have clustered network architectures, which is a step towards grid storage but not the entire journey."

Grid in action

Demand is highest for grid-like features in the high-performance computing market, such as seismic research in the oil and gas industry, media companies creating video and animation, and drug companies doing biomedical research. More mainstream demand, however, comes from customers looking to:

- Speed backup and restore.
- More easily move less vital data to less-expensive storage.
- Store more data without constantly buying more disk.

"Before adopting NetApps' grid technology, we were buying about 700 gigabytes of disk every two months for ERP, clinical, payroll, imaging and other data," says Carruthers. A Storagetek Powderhorn tape library with four drives was busy all day backing up data from the 800 Series filers one volume at a time. The hospital has now consolidated that data on the FAS 960 in volumes that can grow or shrink as needed.

Rather than copy each volume to tape and take the tape off-site every day, Carruthers stores only the changes to the data in a NearStore R200 series filer that is used to back up the data to tape. Among the benefits, he says, "We're not using high-cost, high-performance storage to do backup." Carruthers also wanted something that was easy to administer. "We didn't want to take six meetings to do a SAN design and setup," he says. Carruthers was able to design and set up his new consolidated architecture in two days and migrate his data to it in a week.

Bosworth, the director of IT operations at the Watertown, Mass., headquarters of civil engineering firm Vanasse Hangen Brustlin, provided the main office with the same level of data protection as its 14 field offices, which were doing real-time replication to an EMC CX400 in the main office. That main office, though, was burning backup data to DVDs that had to be manually loaded in the server if a restore was required.

Bosworth considered real-time SAN-based replication but rejected it as too expensive. He then turned to the Advanstor storage system from ExaGrid. Advanstor uses a Windows Storage Server 2003-based filer for primary NAS storage, and a grid of Linux-based storage "bricks" as a virtualized disk pool for longer-term backup. Bosworth uses a set of Advanstor's to replicate data between two locations, automatically moving less active data to lower-cost repositories.

Mike Luter, chief technology officer of the Cancer Therapy and Research Center in San Antonio, Texas, sees grids as "plumbing" even as he tests a grid-like storage system from Waltham, Mass.-based Archivas. The company's Archivas Cluster (ArC) consists of multiple Linux-based storage nodes connected by an Ethernet network. Each node manages its own data, as well as metadata about the data stored on other nodes, so that applications can access any data on the system even if one node fails.

Luter is testing the ArC to store patient records that are accessed for periodic follow-up tests and treatment but that aren't needed often enough to be stored on the center's higher-performing but more expensive storage. That's important because growing use of medical imagery has boosted his storage needs from two terabytes three years ago to seven currently, with another seven expected to come on-line this year.

He also likes the Archivas failover capabilities, which ensure that "if we lose a node we don't lose the data." Such data protection on a conventional array would require backups to tape, an often cumbersome process that requires ongoing investment in equipment and staff.

Questions to ask

With all the confusion and multiple approaches to grid, customers should ask vendors what they mean by grid storage specifically — what they have today and what the road map is, suggests Asaro. "Customers should ask what services, applications and storage protocols are supported, as well as how the grid architecture will be managed," he says. The ability to easily adapt the grid to changing needs is also important, Asaro adds.

A database application that has a large number of transactions and is mission-critical will demand a certain level of storage performance, capacity and protection, he says, whereas another application may generate large documents but have low performance requirements.

Customers should further inquire about how to migrate from their current environment to a vendor's, Balaouras declares, adding, "If you invest in somebody's virtualization technology today, how do you migrate to nirvana three years from now?" That nirvana should include common management tools for grid and other storage, support for both file-and block-level data access, and day-to-day benefits such as faster backup or restoration of data, she says.

Balaouras says grid-based storage should also cost 20% to 30% less than monolithic arrays, a price difference that (combined with savings on storage management) makes it worthwhile for some customers to buy storage from a startup even when they factor in the risk that the vendor may only be around for two to three years.

Over the next three to four years, predicts Hornung, virtually all storage will be grid-like in that it will be based on smaller, easily integrated building blocks that can scale horizontally in performance and capacity in seamless pools of storage. Until then, though, look carefully under the hood at purported grid storage and base your purchase decision on the real-world problems it solves.

About the author

Robert L. Scheier is a freelance writer who covers storage from Boylston, Mass. He can be reached at rscheier@charter.net.

[Feedback: feedback@snwonline.com](mailto:feedback@snwonline.com)

[Questions: questions@snwonline.com](mailto:questions@snwonline.com)

[Technical difficulties: WebMaster@snwonline.com](mailto:WebMaster@snwonline.com)