
AUGUST 29, 2005

STORAGE BY THE CLUSTER

COMPUTERWORLD

An IDG company

By Lucas Mearian

Russ Miller runs a monster of a server cluster that eats storage at an incredible rate. The bandwidth requirements alone on his 22TFLOPS system force Miller to look outside the storage box, so to speak, for better throughput and scalability.

As the director of the Center for Computational Research at the University at Buffalo, Miller oversees a supercomputer comprising 6,600 processors that is used by the university and many businesses in western New York.

To support all that computational power, Miller turned to a clustered storage system that could alleviate bottlenecks and automatically load-balance and grow on the fly to accommodate user demand.

Like many IT managers who have seen the benefits of server clusters, Miller chose to try the relatively new technology of storage clusters as a means of attaining a fully redundant infrastructure that's highly scalable and easy to manage. Clustering provides massive throughput because of an increased port count that comes from cobbling many storage servers together into a single pool of disks and processors, all working on a similar task and all able to share the same data.

Management functions are distributed across the storage server farm. To an application server, the farm looks like a single, block-level storage system. Storage capacity can be added without disrupting applications running on the cluster.

There's lots of talk about storage clustering among vendors these days, but few market leaders have fully embraced the concept, according to analysts. Most of the development is still being led by start-up companies such as Ibrix Inc., Isilon Systems Inc. and Intransa Inc.

In April, Miller selected a system from Dell Inc. and Billerica, Mass.-based Ibrix that gave him storage read rates of 2.3GB/sec. and about half that rate for data writes -- far above what any monolithic storage array could produce, he says.

"We don't have any single points of failure. So if and when we need to make additional investments in storage, we can do that without any major downtime or major reconfiguration. We didn't see any downside to going this route," Miller says.

Tony Asaro, an analyst at Enterprise Strategy Group in Milford, Mass., agrees. "The beauty of some clustered architectures is you can start small and grow as much as you want," he says.

The University at Buffalo's storage cluster consists of three EMC CX700 storage arrays, each with 70 146GB drives that are managed with Ibrix's software.

"If one of these I/O nodes goes down, we won't lose anything except a little performance," says Miller.

Ibrix is a clustered file system that runs on hosts, but it can also run on storage arrays. For example, the internal disk drives on low-end Dell servers can be combined to create a storage pool. The result is a compute farm that also clusters its storage. "It adds no greater complexity by adding more servers to the cluster," Miller says.

In the future, he notes, the university will consider using commodity servers to create a storage cluster, "so long as the system meets the needs of our users and staff in terms of performance and reliability."

There's some confusion about the definition of clustered storage. Vendors describe several different

technologies as clustered storage -- from disk pooling to virtualization.

A true storage cluster should be able to scale linearly without bottlenecks or added management difficulty, according to Enterprise Strategy Group's Asaro.

"NetApp has been guilty of using the word clustering for many years. We used to use it as a fail-over term," says Jeff Hornung, vice president of enterprise file services and storage networking at Sunnyvale, Calif.-based Network Appliance Inc. "Clustering for scalability is what we're talking about."

NetApp offers a dual-node cluster, or two network-attached storage (NAS) boxes together in a pair that provide greater redundancy but don't fit what Asaro and others would consider a true cluster: a system with the ability to scale linearly without adding complexity.

IBM's SAN Volume Controller appliance and 3Par data Inc.'s Inserv arrays allow storage clusters to grow in eight-node increments, but data sharing is confined within those eight nodes. Other companies, such as Isilon, Ibrix and LeftHand Networks Inc., allow clusters to grow one node at a time, with data being shared throughout the cluster, no matter how large it becomes.

Two Categories of Clusters

Clustered storage falls into two categories: systems that combine block-based data on a storage-area network (SAN), and those that create a common file name space across NAS filers.

To date, most of the major storage vendors -- Hewlett-Packard Co., EMC Corp., Hitachi Data Systems Corp. and NetApp -- have released technology that can virtualize NAS systems by pooling disk capacity behind NAS engines. All claim to be developing or evaluating third-party clustering technology as well. Meanwhile, start-ups such as Isilon, PolyServe Inc. and Panasas Inc. are already offering clustering software that runs across Windows and Linux.

Sonja Erickson, vice president of technical operations at Kodak Easy Share Gallery, a service of Kodak Imaging Network Inc. in Emeryville, Calif., says NAS clustering technology has already saved her company hundreds of thousands of dollars in personnel costs alone.

The Kodak unit uses clustering technology from both Seattle-based Isilon and Beaverton, Ore.-based PolyServe to connect hundreds of Intel servers at Kodak that host its online digital photo image service. Erickson has a staff of five people to manage more than a petabyte of data on the servers.

"In terms of staffing, since we installed [a NAS cluster from Isilon] a year and a half ago, we've hired no additional staff," she says. "That's hundreds of thousands of dollars saved. In terms of efficiency, it takes only a day to get the systems up and running."

Prior to installing NAS clusters, Erickson used direct-attached SCSI arrays that lacked scalability and could take up to a month and a half to get online. In contrast, the PolyServe boxes require about a week, and Isilon's take about a day, she says.

EMC offers clustering technology in its Centra content-addressed storage array, as does HP through its Remote Installation Service platform. Both products, however, are targeted at online archival uses. "These are clusters, but the object is not performance," says Bob Passmore, an analyst at Gartner Inc. in Stamford, Conn.

Most of the large vendors are interested in using Intel-powered servers for storage clusters because of the enormous growth in the use of Linux clusters in server farms.

"The requirements for availability are a long way from what a bank would need. Typically, banks are looking for tons of computing and storage capacity. And while you get lots of performance, the trade-off is low availability. IBM has a group focused on this problem," Passmore says. "But we're going to see more and more of this stuff, especially the cheap nodes used to create storage clusters."

Most of the adoption of clustered storage is being spurred by the technology's relatively low cost, adequate performance and high levels of redundancy and throughput.

Reluctant User

Ron Minnich, team leader of the Cluster Research Lab at Los Alamos National Laboratory in New Mexico, is a reluctant fan of clustered storage because it scales with his needs and gives him the throughput to support a 1,024-node, \$6 million Linux server cluster. The cluster, called Tank, works as a supercomputer for crunching scientific equations.

Tank is backed by a 64-node storage cluster from Fremont, Calif.-based Panasas.

"In terms of pricing, it cost more than a NetApp [NAS array] per terabyte," Minnich says. "I have it mainly because of its ability to sustain throughput from our server cluster, which you can't achieve with a standard file servers, which have one network connection."

Minnich says that each of the storage blades on the Panasas system has an individual Gigabit Ethernet connection, giving him 64 individual network connections.

"It's not as reliable as a mature product. I had computers when [the Network File System protocol] first came out. There was a certain failure rate. This system is like the early failure rate that NFS had years ago," he says.

Ron Godine Jr., manager of IT operations at Royal Appliance Manufacturing Co. in Glenwillow, Ohio, had been supporting his Oracle 11i ERP system with two EMC Clariion 4700 arrays prior to buying LeftHand SAN Filers, which are clustered using Microsoft Cluster Server for high availability.

Godine laments that each time he had to upgrade his EMC array, there was an arduous series of planning and implementation steps that had to be carried out with the help of the vendor. "It became a fairly drawn-out process to get the equipment working and figure out a number of problems," he says.

The three-node cluster from Boulder, Colo.-based LeftHand Networks gave Godine about 6TB of storage capacity for \$31,000 -- about half the cost of his EMC arrays. "And the maintenance is significantly cheaper," he says. "I was able to get the equipment up and running in-house. We thought we'd done something wrong because it was so simple."

Godine's greatest concern in moving away from time-tested EMC equipment was the performance on the new cluster technology, which he says surprised him.

"We found performance on this cluster faster than local disk and certainly much faster than using CIFS [Common Internet File System] on these monolithic file-sharing systems," says Godine, who adds that when used in the SAN configuration, the cluster is comparable to monolithic boxes.

Godine says the ability to scale as needed was also an enormous draw. "If we needed more bandwidth to hosts, we could do it by adding more building blocks," he says.

In theory, managing storage clusters should be no more difficult than managing a single array, but some users say their management interfaces could still use some tweaking.

But most, like Godine, say clustered storage's strengths outweigh its drawbacks.