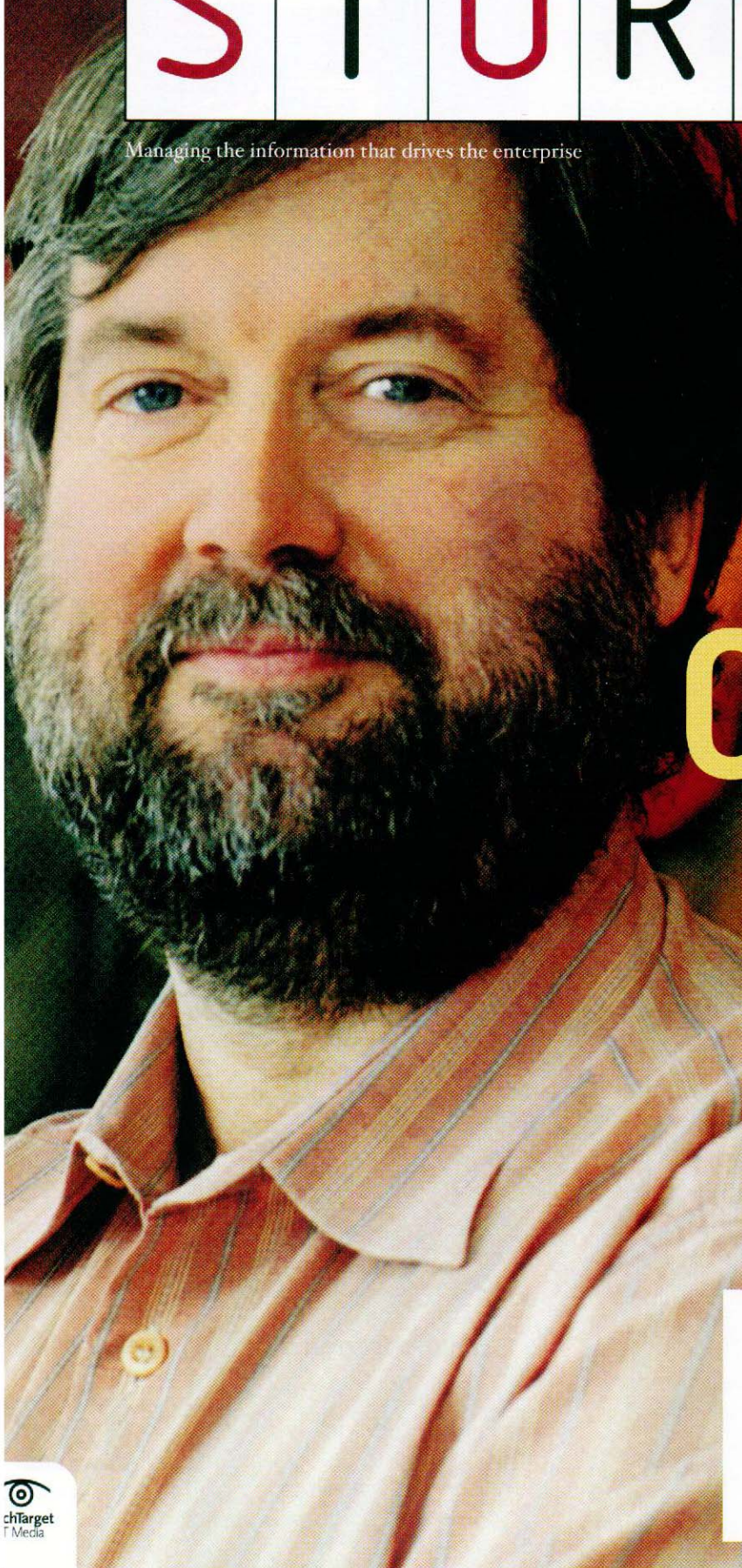


STORAGE

Managing the information that drives the enterprise

Vol. 4 No. 2 April 2005



scale up with storage clusters

Sports Illustrated's Phil Jache says clustered storage is so reliable and easy to manage that he shipped systems to the Olympics and the Super Bowl 20

Also inside:

The lowdown on virtual tape 32

Pumped up SATA 41

Coping with change 46



“The **clustered system** is pretty handy—every

Clustering storage resources can yield many benefits, such as easy capacity and performance scaling, as well as high availability

BY RICH CASTAGNA

clustering storage

The storage side of the data center is waking up to what the server side has known for some time: Clustering is cool.

Clustering has improved the reliability, availability and manageability of data center servers while allowing bundles of inexpensive configurations like blades to replace costly, monolithic servers. The benefits of server clustering haven't escaped the notice of the storage industry, but clustering storage involves challenges other than just tying servers together. Vendors have taken diverse paths to address those challenges, but they fall into two main categories: clustered file systems and standalone hardware with a clustered architecture.

unit you have **just adds to the speed,**" says *Sports Illustrated's* Phil Jache.

As a clustered storage system grows to accommodate new requirements, administration of the system remains constant

“With traditional midrange storage systems, you can quickly run out of hardware resources,” says Tony Asaro, senior analyst at the Enterprise Strategy Group (ESG), Milford, MA. When more capacity or horsepower is needed, traditional systems offer few alternatives other than installing another storage device with all of its associated costs.

Implementing a clustered storage system doesn't require clustered servers. While the technologies are quite similar, they aren't interdependent.

Defining terms

The growing popularity of clustered storage has also spawned the usual industry buzzword mania. Storage vendors of all stripes are touting their hardware and software products as clustering technologies—products that may be implemented at nearly any point in a storage environment. While spiels tend toward hyperbole, most of these products are clustering applications, although many are point, rather than total, solutions.

Vendors have turned toward clustering technologies to address the four big issues facing most storage managers. These design goals aren't the exclusive province of clustering—nearly all storage systems strive for these—but they're the fundamental goals of clustered systems:

- **Capacity scaling.** Additional storage capacity should be easy to add in a non-disruptive manner.
- **Performance scaling.** As capacity is added and the number of supported hosts grows, performance should scale sufficiently to maintain an acceptable service level.
- **Availability.** Redundant components and trans-

parent failover should ensure data is always available.

- **Manageability.** Scaling, failover and capacity management should be as automated as possible.

These goals may be achieved in a variety of ways, but there are some basic precepts of clustered storage. For example, clustered systems pool their storage and present it as a single image to hosts as a global file system that's often referred to as “a single drive letter.” This makes better use of available capacity while easing storage management. It also enhances the ability of hosts to share data while avoiding multiple instances of the same files (see “Traditional vs. clustered NAS storage,” this page).

Cluster concepts

The simplest form of clustering involves two controller units paired so that one would provide failover for the other. In a two-way, active-passive paired configuration, one controller is essentially on standby. Because this scheme doesn't provide for scaling and the passive unit isn't sharing the primary's load, this is often referred to as “pseudo clustering.” An active-active controller arrangement is a step up from pseudo clustering, where the two controllers provide failover for each other and share the work.

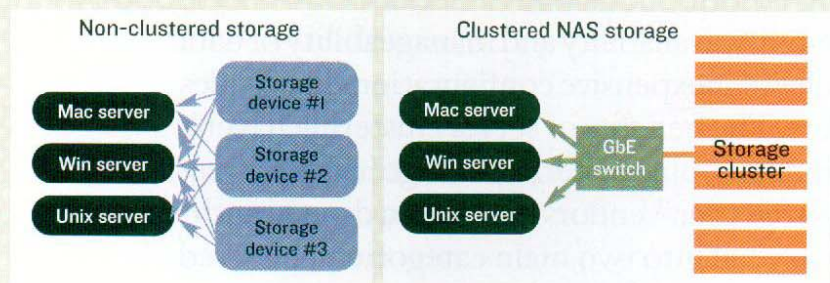
In a non-distributed, active-active cluster, cluster members share a file system and some other physical resources, but provisioning and LUN assignment for specific controllers are mostly manual chores. The distributed peer cluster is the most common architecture employed by vendors that have designed and built their clustered storage systems from the ground up. In a distributed cluster, physical resources are virtualized, so a storage administrator only needs to deal with how storage is associated with installed servers and the applications they host.

Single file system

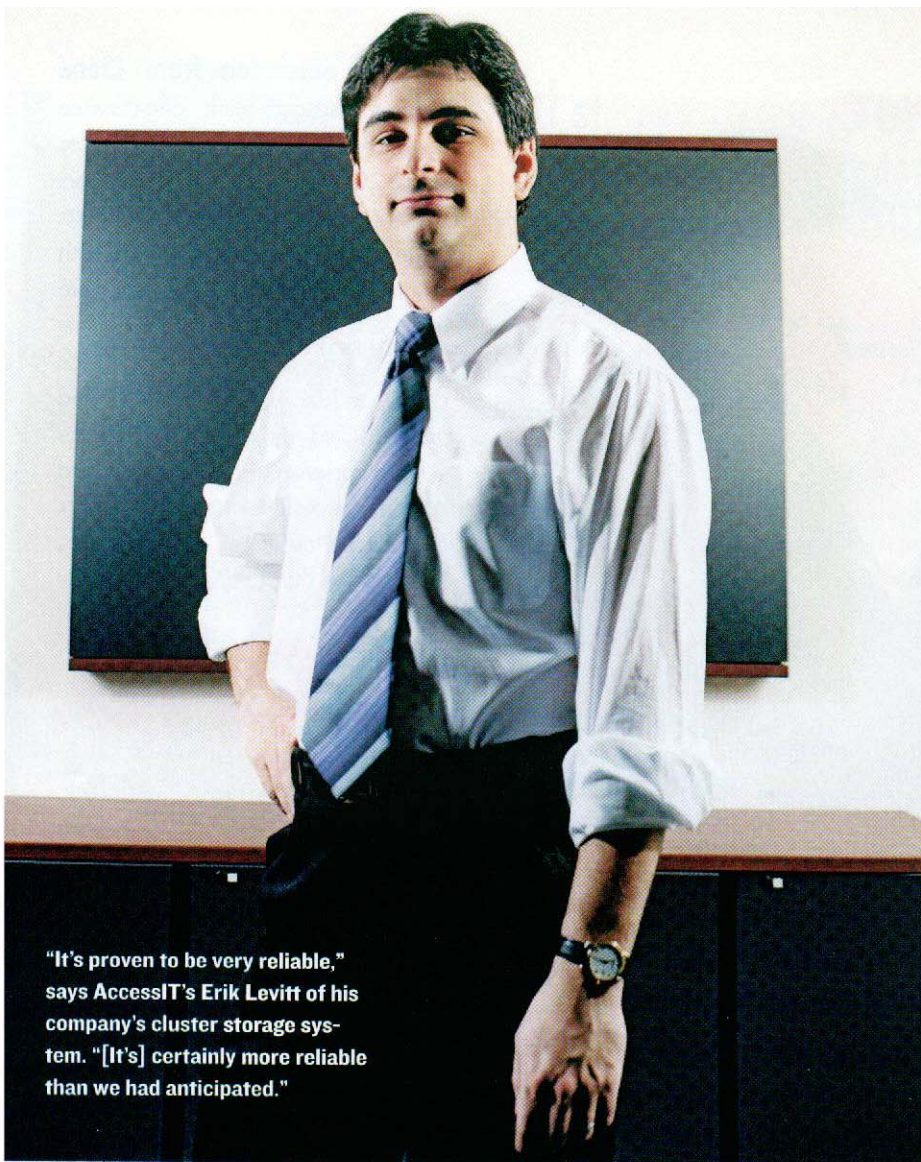
There are products that provide global file system capabilities for aggregated storage systems like IBM Corp.'s SAN File System (SAN FS). These applications typically run on an appliance or an intelligent switch with client software on supported hosts to deliver one of clustering's key requisites, the global file system.

SAN FS and similar products take a two-pronged approach: They virtualize the storage they sit in front of into a single file system and interface with the hosts' OSes to present that file system as if it were native to the hosts. In this manner, these sys-

Traditional vs. clustered NAS storage



Storage systems typically become siloed as more capacity is required. Multiple connections are required to allow hosts to access all installed storage. In a clustered storage system, the storage controllers communicate with each other internally and present a single file system to the hosts. Multiplatform hosts can connect to the cluster through a single connection to a switch which, in turn, is attached to the cluster.



"It's proven to be very reliable," says AccessIT's Erik Levitt of his company's cluster storage system. **"[It's] certainly more reliable than we had anticipated."**

tems improve capacity management by providing policy-based data migration across all connected storage. That enables more effective storage tiering, a basic step toward an information lifecycle management implementation. SANFS works with a variety of Windows, Linux and Unix hosts, but requires an IBM storage system for its meta data store. It supports numerous back-end storage systems and can be used in conjunction with IBM's SAN Volume Controller (SVC) to support a range of storage arrays.

Examples of other clustered file system products include Ibrix Inc.'s Fusion, PolyServe Inc.'s Matrix Cluster, Red Hat Inc.'s Global File System (formerly Sistina GFS), SGI's InfiniteStorage Shared Filesystem CXFS and Veritas Software Corp.'s Cluster Server. These are all host-based apps that cluster servers and provide a single image of the storage available attached on SANs.

Clustered file systems are attractive because they can work with installed storage. On the other hand, hardware clustering systems require the purchase of new storage (see "Software vs. hardware clustering," p. 26).

But virtualization and a global file system don't necessarily add up to a fully clustered storage system. Randy Kerns, a senior partner at Evaluator Group Inc., Green-

wood Village, CO, describes SAN FS as "a meta data server approach to storage virtualization." It provides a key element, but it's only part of the clustering picture. "It's one way to provide a global namespace," notes Kerns, "but a global namespace and clustered storage are not necessarily connected."

Beyond the file system

A fully clustered storage system goes beyond what the servers and applications see; it provides the underpinnings and infrastructure of the storage system itself. Among available products, the best examples are those that have been built from the ground up to deliver clustered storage. These hardware-based systems address the scalability of physical resources, not just that of the file system. According to Kerns, these systems have an advantage over some of the software-only approaches to clustering. "You're going to put on another layer of software and yet you're still probably going to manage those devices independently," says Kerns.

Some examples of purpose-built clustered storage systems include EqualLogic Inc.'s PS Series, Isilon Systems Inc.'s IQ series arrays, LeftHand Networks Inc.'s SAN/iQ IP SAN and Xiotech Corp.'s Magnitude 3D (see "Clustered storage system sampler," p. 29).

While most midrange storage systems offer a modular approach to growing capacity, clustered systems take the concept a step further. Typically, in a non-clustered midrange array, a module (or expansion unit) is added to increase disk capacity; in some cases, another controller can be added to increase the horsepower of the array. For the most part, these modular midrange arrays can scale capacity, but not performance. "If you're just adding disk, but aren't doing anything about performance," says Kerns, "obviously you'll see some degradation."

In a clustered storage architecture, modules are typically packages that include not only additional disks, but a controller assembly with its own set of interfaces. Building out a clustered array also increases performance and connectivity. Because a full complement of processors, memory, ports and so forth is added with each set of new disks, the performance of a clustered storage system will often scale linearly as it expands. This is in stark contrast to non-clustered modular systems where performance is likely to suffer as disk expansion units are added.

Software vs. hardware clustering

CLUSTERING APPROACH	PROS	CONS
Clustered file system (software-based clustering)	<ul style="list-style-type: none">• Works with installed storage• Works with heterogeneous storage• Provides server failover	<ul style="list-style-type: none">• Doesn't scale capacity or performance of supported storage systems• Doesn't add any redundancy or failover to storage systems
Clustered storage system (hardware-based clustering)	<ul style="list-style-type: none">• Provides capacity and performance scaling• Can manage all clustered storage from a single console• No server software required	<ul style="list-style-type: none">• Self-contained systems don't cluster other storage• Typically not appropriate for enterprise-class applications

When a module is added to the cluster, the other members of the cluster automatically recognize the new module. The cluster then reorganizes itself to accommodate the added capacity by re-striping data across all disks, sharing data management policies and balancing the workload among all members. Usually, cluster modules interconnect with each other using a Fibre Channel (FC) or Gigabit Ethernet (GbE) interface, although Isilon recently announced it will offer clustered storage systems that use InfiniBand connections, which are approximately 10 times faster than GbE.

Servers connected to the clustered array are unaffected. Typically, there's no need for client software on the host servers, and they can continue to access storage from the pool even as new capacity is added. Within the storage cluster, the specific controller that a host connects to is almost irrelevant, as cluster modules can hand off responsibility for those interfaces to one another to adjust to failures or varying loads and bandwidth requirements.

For cluster modules to interact effectively, their operating systems must be in constant communication. If a unit fails—or shows signs of an impending failure—its processing workload is picked up by other cluster modules and data is transferred from its disks to others, if necessary. This arrangement provides effective failover to ensure availability and, as more modules are added, data protection and availability increases as well.

Most importantly, as modules are added to accommodate new requirements, administration remains constant. Even as the cluster grows, "I can administer it as a single system and don't have to change anything," says Kerns. "I don't have to administer another box."

Sports Illustrated in New York City opted for clustered storage to support its onsite digital photography operations. Phil Jache, deputy director of technology for the magazine, says their three Isilon IQ arrays have been air-shipped to the Olympics, Super Bowl and other major events. The Isilon systems cut two or three hours from the magazine's photo processing time. "It enabled us to do some things that just weren't possible [before]," says Jache.

AccessIT installed a Xiotech Magnitude 3D clustered storage system at its Managed Services Division in New York City and another at its Media Services headquarters in Los Angeles. Erik Levitt, president and COO of the Managed Services Division, says AccessIT installed one of the Xiotech boxes to support its IT

services business, which supports clients in 35 countries from 10 data centers. The Los Angeles-based Xiotech system is used primarily for the distribution of digital films, such as *I, Robot* and *Shark Tale*, to nearly 30 movie theaters equipped with digital projection systems.

The Xiotech cluster in New York replaced a traditional monolithic SAN. Levitt says the price of the Magnitudes was a major selling point. "We're adding about 5TB at a clip, so scalability is extremely important to us," says Levitt.

Basic foundation

Clustered storage systems owe their intelligence to the sophisticated software that controls operations. The underlying hardware is generally unremarkable—off-the-shelf parts such as Intel's Xeon processors and other standard components used in ordinary servers and server blades. In many cases, the cluster OS is built on top of a Linux kernel. This adds up to a low-cost architecture that provides the necessary economies for modularity.

Vendors of clustered storage hardware systems have managed to hide the inherent complexities and advanced features so effectively that most users first cite ease of use when describing what impresses them most about the clustered systems they've implemented. Users report that scaling with additional modules is as close to plug and play as storage gets. With no client software to install on hosts and streamlined, Web-based user interfaces that reduce many configuration and admin chores to point-and-click operations, implementation seemingly couldn't be easier.

"Most of the clustered storage systems ESG has analyzed are extremely easy to work with," says the firm's Asaro. *Sports Illustrated's* Jache concurs: "Honestly, the setup is 30 minutes," he says. "It's drop-dead easy."

Ron Godine, manager of IT operations at Glenwillow, OH-based Royal Appliances Manufacturing Co., the maker of Dirt Devil vacuums and other floor-care products, moved some data from an EMC Corp. Clariion to a LeftHand clustered array. In his search for a successor

Clustered storage essentials

GLOBAL FILE SYSTEM All storage is pooled and made available to hosts through a single, scalable file system that hosts see as an ordinary drive letter.

CAPACITY SCALING Disk capacity can be added incrementally and is automatically detected and merged into the storage pool; existing data is re-striped to take advantage of the new capacity.

PERFORMANCE SCALING Performance scales with each added controller.

LOAD BALANCING The workload is shared among cluster members and redistributed as necessary.

FAILOVER Redundant components with no single point of failure; transparent recovery from failure.

SINGLE-PANE MANAGEMENT All controllers and storage installed in the cluster can be managed via one interface.

to the Clariion, he considered traditional and clustered storage systems. "Instead of a monolithic array, we wanted something that was more scalable," says Godine. For Royal, the LeftHand system has proven to be economical and easy to manage. "You can create and destroy LUNs in a fraction of the time it takes with other systems," says Godine.

report that their customer rosters have grown beyond the entertainment and scientific industries to include financial, government, education and healthcare sectors. For example, Mark Rivard, network systems specialist at Johnson Memorial Hospital in Stafford Springs, CT, uses a 3TB EqualLogic PS array for file serving, e-mail and as a virtual disk for backup.

Best fit for clusters

The vendors in the vanguard of clustered storage largely found their way into data centers on the strength of how well their systems handle digital imaging applications, such as video editing and pre-production. For *Sports Illustrated*, Isilon's systems filled the bill. "This is digital data that only gets read a few times," notes Jache. "It's not like a database where you're doing a lot of incremental reads."

Some clustered storage vendors



IS YOUR DATA WHERE IT SHOULD BE?

Introducing the TagmaStore™ Universal Storage Platform. You know how critical your data is. Even more critical is making sure your applications have fast and easy access to it. That's where TagmaStore can help. The TagmaStore Universal Storage Platform manages all your storage systems to work more efficiently for your business. As one of our

Clustered storage system sampler

HARDWARE

- 3PAR InServ Storage Server
- EqualLogic PS Series
- Exanet ExaStore
- Isilon IQ
- LeftHand Networks SAN/iQ
- Panasas ActiveScale Storage Cluster
- Xiotech Magnitude 3D

SOFTWARE

- HP StorageWorks Scalable File Share (HP SFS)
- IBM SAN File System (SAN FS)
- Ibrix Fusion
- NetApp SpinServer cluster
- PolyServe Matrix Server
- Red Hat Global File System
- SGI InfiniteStorage Shared Filesystem CXFS
- Veritas Cluster Server

"We back up the entire domain to the cluster first and then to tape," says Rivard. And he plans to expand the number of applications using the system. "Every application that comes into the organization that has any volume of storage will be attached to [the EqualLogic cluster]," he says.

But despite its undeniable appeal, clustered storage is not necessarily a good fit for mission-critical database and online transaction processing applications. "Clustered architectures have to coordinate every request over the fabric to communicate with other nodes," says Sujal Patel, Isilon's CTO, chairman and founder, "and that coordination takes time because there's latency in GbE."

But Patel sees technologies such as Infiniband reducing that latency and increasing bandwidth between nodes. "The speed of the networking interface between the

nodes is going to approach the speed of the computer bus; as that occurs, clustered architectures won't have any disadvantage vs. monolithic architectures," he says.

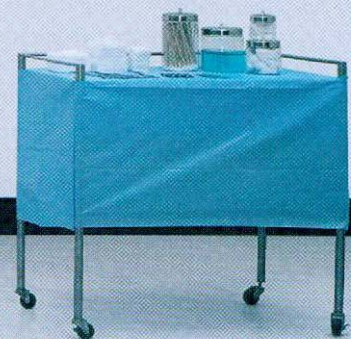
Clustered storage has a lot going for it. It effectively—and often elegantly—addresses some of the key drawbacks and bottlenecks associated with traditional

storage systems. "It can greatly simplify the management of storage networks," says ESG's Asaro, and "when combined with intelligence such as moving data between tiers, data de-duplication, retention policies and meta data, it can create new [storage] applications."

Evaluator Group's Kerns is bullish on clustering, too: "It's the logical evolution," he says. "As more people catch on, it's going to become more and more popular." ☉

RICH CASTAGNA is the Editor of Storage magazine.

©2004 Hitachi Data Systems. All Rights Reserved.



lication Optimized Storage™ Solutions, it moves data between itself and lower tiers and brands of stor- as if they were one. It's the latest example of how Hitachi Data Systems remains committed to being your ner Beyond Technology. To learn more about our Universal Storage Platform, visit us at www.hds.com

HITACHI
DATA SYSTEMS