

VERSION 2
Word Count: 2049

The Future of Enterprise Storage Management is High Mix, High Intelligence

By Gary Wright
Compaq Computer Corporation

It's no secret that the data storage needs of the typical Fortune 1000 company are doubling — some say tripling—every year. In the Year 2001 alone, disk drives capable of storing 4,642 petabytes of data will be produced¹. This explosive growth is fueled in large part by the pervasive role of the Internet in the corporate way of life. The world demands that companies make themselves, their products and their public data available on the Internet, and the smart companies comply.

Dealing with this massive increase has become a mission-critical focus of the IT department: failure to manage storage capacity, accessibility, security and performance would be a death knell to the organization.

The concomitant to massively increased storage capacity is complexity. As an organization grows, it tends to add different storage devices and assimilate new networks through acquisition. Implementing Intranets, e-commerce or order-entry, CRM or other services over the Internet creates additional levels of complexity. If the organization is to avoid being brought to its knees in a welter of devices that don't speak to one another, network failures, backup nightmares and unacceptably slow performance rates, it must continually seek out new means of managing data storage.

In our years of managing data storage issues, Compaq has determined that the keys to managing rapid growth of data complexity are:

- Ability to implement and manage heterogeneous environments
- Greater intelligence in the storage network
- Centralization of storage management
- Increasing data accessibility
- Efficiency in storage utilization

¹ "How Much Information?", Peter Lyman and Hal. R. Varian, School of Information Management and Systems, University of California at Berkeley, 2000.

As an organization grows, creation of heterogeneous environments is almost unavoidable. Whether because of acquisitions, changes in policy or management or a dozen different reasons, growth companies find themselves dealing with many different data storage devices and systems that must be integrated. Few organizations can afford to junk their existing investment in servers, storage devices and software.

According to a report published by the Enterprise Storage Group², most IT groups are currently implementing homogeneous storage area networks (SANs), often as “islands” in a Direct Attached Storage or Network Attached Storage environment, with only one operating system and one type of storage array. However, this group said that the cost of implementing SANs was not significant (versus the cost of implementing other types of storage), and planned to tie together SAN islands, or implement more elaborate networked systems in the future.

For these more complex, heterogeneous, networked environments to be successful, the industry must evolve toward an open SAN environment, and the truth is that the industry is not there yet. At present vendors are starting to enable device communication at the API level. Future developments look promising: recently, the Storage Network Industry Association, including Compaq, Brocade Communications Systems, Inc., EMC Corporation, Hitachi Data Systems Corporation, International Business Machines Corporation and McDATA Corporation announced several initiatives intended to provide the first qualified cross-vendor, interoperable storage networking solutions. As a starting point, the six companies have completed joint qualification of two open storage area network solutions that enable the coexistence of data zones containing Compaq, EMC, Hitachi Data Systems and IBM storage system products on a single, shared Fibre Channel fabric. The four storage system vendors have signed bilateral cooperative support agreements intended to simplify joint customer support in multi-vendor environments.

With the increasing complexity and heterogeneity of the storage environment comes the problem of how IT personnel can manage multiple operating systems, languages, and hardware devices, and make them all work together. According to Carolyn DiCenzo, chief analyst for Gartner Dataquest’s Storage Management Software and SAN Appliances research group, storage hardware budgets are growing faster than budgets to hire trained IT people because “It’s cheaper to buy more software than hire trained people,” and it is increasingly difficult to find qualified people to manage the storage.³ Many IT organizations lack the skills and knowledge, according to CNET Enterprise, and “people with the needed knowledge and skills are hard to find.”⁴ And no wonder. IT personnel have to deal with an array of technologies that mushrooms every year. Storage Networking World Online offers the following list of technologies “now directly involved in or significantly touching upon data storage”:

- Fibre Channel Arbitrated Loop and fabric architectures
- Ethernet and IP

² Enterprise Storage Report, Vol. 8, © The Enterprise Storage Group, Inc.

³ “Gartner Dataquest: Storage Management to Hit \$16.5 Billion in 2005,” Joseph F. Kovar, CRN, May 21, 2001

⁴ “Store This,” CNET Enterprise, CNET.com, May 9, 2001

- Server clustering
- Data backup and restoration
- SCSI
- Wide- and metropolitan-area networking protocols
- Bridging, routing and switching
- Host bus adapters and drivers
- Broadband communications
- Enterprise management applications
- Database and file system architectures
- Operating systems⁵

Although fully interoperable storage networks may not yet be realized, nonetheless many companies have been successful in integrating multi-platform storage network solutions. Canada's pre-eminent brewer, Molson Breweries, found it had outstripped its data storage capabilities. The company was placing priority on systems to support strategic operations and improved IT productivity, including the creation of a marketing data warehouse and other multi-platform computing projects. Of critical importance was Molson's on-line transactions processing (OLTP) system, which processes orders for more than 25,000 customers. Molson also wanted to extend the life of the company's investment in existing hardware and software.

Molson's solution incorporated the deployment of production systems on seven Hewlett Packard HP/9000 systems running the HP-UX operating system. Compaq servers running Microsoft's Windows NT Server network operating system throughout the organization host various departmental applications. Compaq's StorageWorks open system design supports this multi-platform environment, and will be able to support Molson's future plans to support Microsoft Windows NT applications as well.

Take rising complexity and add a shortage of skilled people, and the equation equals a need for more intelligent storage management software. Every major enterprise storage vendor has its own "flavor" of storage management software, but they all depend on implementation across a network. Networking the storage system enables several critical capabilities:

- Monitoring of storage devices remotely
- Remote upgrade, maintenance and reallocation of storage resources
- Implementation of a choice of backup schemes
- Centralization of management control

Taken together, these capabilities offload much of the management of the storage system from IT staff, and offer significant improvements in terms of data availability, security, and the time and cost of administering the system.

When storage management becomes centralized, users can benefit from increased data accessibility. Centralization significantly reduces the complexity of the task, allowing

⁵ "The Overwhelming Importance of Storage," John Webster, Storage Networking World Online, April 9, 2001

fewer people to manage more data. When storage usage is centrally monitored, it is also possible to re-allocate storage from the central console (remotely) based on that usage, instead of by predictive algorithms (never well accepted by users who believe they are overcharged or under-served!). Heavy users can be allocated appropriate storage capacity instead of being forced to rely on backup and archiving to make room for more data. Central control allows for a complete “storage-centric” view of enterprise storage, instead of trying to make sense of a fragmented picture.

EarthLink, an Internet Service Provider, managed its explosive growth (from 0 to 3 million members in five years) in part by centralizing its data storage management. As EarthLink rapidly added servers and storage devices to keep pace with users’ demands for service, the company found itself trying to manage many isolated storage islands and many single points of failure. The LAN services group decided to move from a server-centric model to a storage-centric model and implemented a new, more intelligent storage architecture that could be monitored and controlled from a single console. One of the greatest short-term benefits of the change, according to Greg Friedman, director of EarthLink’s LAN Services, was the ability to add or remove LAN servers without downtime. He reports that a process that used to take several hours has been reduced to about 15 minutes.

E-commerce over the Internet continues to be a major driving force behind the demand for 24/7 data availability. It is no longer acceptable to take servers off-line for backup, maintenance or changes. According to industry analyst Meta Group, companies that are the most dependent on automated systems (banking and telecommunications, for instance), accrue an average of nearly \$3 million in losses for every hour of downtime. Businesses like hospitality or travel, less dependent on IT infrastructure, suffer revenue losses of between \$330,000 and \$636,000 per hour of downtime.⁶

Businesses engaged in 24-hour business operation need to be able to capture “snapshots” of data at any given moment in time. The organization’s management can process or analyze these snapshots or clones without bringing the network down for data backup, as backup is done from the clone instead of the primary volume. This process is best accomplished in a SAN with centralized management tools that can be used to create the clones, using agents to retrieve point-in-time data.

The South Financial Group is an example of a business where data cloning is essential not only because of business requirements, but because the weather demands it. The South Financial Group, offering mortgage, brokerage and investment services, operates 108 branch offices in some of the most hurricane-vulnerable states in the union: North and South Carolina and Florida. Hart Raley, vice president of Client Services for the company, explained that data replication and data protection were top priority: “With Fibre Channel technology from Compaq, we designed a completely redundant infrastructure to improve performance, protect against system malfunction and optimize our backup capabilities.”

⁶ “Performance Engineering and Measurement Strategies: Quantifying Performance Loss,” Meta Group, October 2000

With data replication software at the heart of a SAN, The South Financial Group implemented a cloning and snapshot scheme that minimizes downtime during backup and data migrations, while fully protecting the integrity and timeliness of the data. The firm tested its data protection scheme by placing a sample database in Lexington, S.C., replicating the database through Greenville, N.C., breaking the connection and running the application off the Greenville site. “All data was up-to-date and correct,” reported Haley.

Another key to managing the increasingly complex storage environment is more efficient utilization of storage. Storage hardware already claims at least half of the IT hardware budget.⁷ Just as it makes sense to have fewer people managing more data storage, it’s common sense to run fewer devices that can accommodate more data. Today, adding more discs to a logical unit number (LUN) means that you must manually re-allocate the storage, which is time-consuming and imprecise, and limits data accessibility during the process. Automating this process is less disruptive, takes greater advantage of storage capacity, and reduces both the time and IT personnel for the process.

The concept the storage industry is currently pursuing to optimize storage utilization is called virtualization. Virtualization separates the representation of data storage capacity from the physical devices — a technique that allows a pooled view of total storage capacity, instead of viewing storage device by device. Automated management and programming tools simplify storage management and allocation, making this process transparent and dynamic. Virtualization increases storage utilization and availability, while reducing significantly the space required for storage (some estimate space reductions of 28%).

That being said, virtualization has not yet been fully realized. Key vendors are working on virtualization and have delivered some of the required tools. In the near future, IT managers will be able to deploy virtualization to manage data storage at the highest strategic level.

In the longer term, tools and technologies will evolve that will make the management of SANs increasingly transparent and easy. For instance, some vendors have a vision of self-healing systems that detect and repair data pathway failures before human technicians could even become aware of the problem. The tremendous pressures imposed by the globalization of business will drive the industry to create technologies that will come as close as possible to the goal of providing the user with delivery of data, at any time of the day or night, every minute of the year.

⁷ “Store This,” CNET Enterprise, CNET.com, May 9, 2001