STORAGE WIZARDRY.

TODAY'S STORAGE TECHNOLOGIES OFFER THE FLEXIBILITY AND CAPACITY TO DEAL WITH EXPLOSIVE CORPORATE DATA GROWTH.

»»»

One of the most complicated and demanding challenges that corporate IT teams face today is finding somewhere to stow all of their companies' data. This includes all the application files, disk images, backups and archives generated by Information Age businesses.

The digital storage needs at the typical company are increasing by 60 percent a year. At some companies, growth is expanding by as much as 90 percent to 100 percent, this according to research conducted by the data storage firm EMC.

Although storage costs have dropped drastically in the last decade, storage needs have increased at a much faster rate, points out John Sloan, senior research analyst for the Info-Tech Research Group. And the complexity of managing all that storage has increased even more dramatically.

Too little storage at any given moment can cause organizationwide system failures, data loss and wasted time. At the same time, too much ties up capital in unused equipment and infrastructure. Not only must today's IT manager have just enough storage capacity, it has to be the right kind of storage capacity — reserving faster and more expensive technologies for only those files that need to be accessed regularly and quickly.

» STRATEGIC BALANCE

Storage management is not quite magic, but it does take a bit of wizardry to get it right. Fortunately, an abundance of technologies and tools to automate it are available to help.

These range from storage area networks (SANs) and storage virtualization (which combine the total storage available and allow it to be allocated as needed without interrupting users' workflow) to data migration and hierarchical storage management (which move files to the most appropriate storage device based on rules).

In addition, data deduplication reduces demand by minimizing the amount of data stored without reducing its availability. And information lifecycle management (ILM) is a process for managing information throughout its lifecycle, from the time it is conceived until it's disposed.

There is wide variation in networked storage options — and in pricing, notes Greg Schulz, founder and senior analyst for the StorageIO consulting group. Therefore, business IT chiefs have to be savvy buyers.

"Businesses are realizing that the path to prosperity is not in saving your way to the future," Schulz says. "Instead, they are investing in IT as a means of becoming more efficient and more successful."

To do that smartly with any networked storage or a SAN project, it's up to the IT team to monetize the value of these technology purchases and show management a business return, he adds. It's important to look for return on investment and set plans for how to measure it.

» SOURCES OF DATA

To know how to handle the data an organization produces and determine the correct investment strategy, it is important to understand the different kinds of data generated.

"One of the things that we look at is the different kinds of data in the environment," says Ted Newman, chief technology officer for infrastructure at EMC. "They have different growth rates associated with them."

Newman outlines three different categories of data that corporate IT teams should take into consideration:

- **Structured:** Structured data includes information that lives in databases and data warehouses, where the size and format of each new record or transaction is determined by the application that generates it. For example, this might be a database that formats each new entry in the same way, containing the same fields.
- **Semi-Structured:** Semi-structured data is encapsulated within a larger container like structured data, but the form of each new transaction can vary widely. For example, e-mail inside Lotus Notes or Microsoft Exchange is structured by the e-mail program, but each message can range from a line or two up to a massive message with dozens of large attachments.
- **Unstructured:** The active files and objects that occupy a user's computer hard drive and network storage are unstructured data, including spreadsheets, presentations, downloaded audio and video, images and PDF files. New files in dozens or hundreds of different formats are created randomly and saved according to users' whims.

Of the three data types, structured data is generally the easiest to manage because it grows in fairly predictable ways. "Structured data tends to grow at a rate that is in line with the growth rate of the business and the economy in general," Sloan says.

"If business is booming, you are making more but also spending more," he adds. "You are buying and selling more. All the systems you use for managing that rely on structured data stores — from inventory to payroll to purchasing."

Unstructured and semi-structured data growth is much less predictable — and typically grows much faster. And because of the varied nature of the information represented within unstructured data, it can be trickier to determine how long to retain the data and how accessible it needs to be over time.

This is complicated even further by sweeping regulatory demands. These might include such things as the financial reporting requirements imposed by the Sarbanes-Oxley Act (SOX) or the privacy requirements of the Health Insurance Portability and Accountability Act (HIPAA).

The storage challenges presented by unstructured data are compounded by the changing nature of corporate work itself. As hierarchies flatten and the relationship between companies and their clients, vendors and the public evolve,

GLOSSARY

- **Data compression:** Algorithms are used to reduce the size of a file by reducing redundancy.
- **Data deduplication:** Redundant data is eliminated by replacing multiple instances of a block of data with metadata pointing to a single master copy of that data.
- **Data migration:** This process moves data from one device to another. During maintenance, it provides continuous access to data, but on a regular basis ensures efficient storage based on the value of specific data.
- **Storage virtualization:** Storage capacity can be administered and provisioned from a single, centralized interface for a variety of storage devices.
- **Storage area network:** The SAN architecture combines storage resources attached to multiple devices and treats them as a single large pool that can be shared by all devices attached to the network.
- **Thin provisioning:** Storage capacity is assigned across the network on a just-in-time basis, as needed by particular applications.

new applications and services — each with their attendant stores of new kinds of data — have multiplied.

"The thing that's really changed over the last couple of years is that everyone has become a content producer, whereas before, those workers were just collecting more data," says Chris McCall, product manager for HP's storage product group. "And that's really driven storage at an exponential pace and made it really difficult to manage."

» MAXIMIZING EFFICIENCY

A number of new technologies have evolved that address the growing storage needs of today's enterprises. Although the technologies vary widely, ultimately storage management boils down to a few basic strategies:

- Maximize utilization of the storage media at your disposal
- Reduce the size of the data that needs to be stored
- Store each piece of data in the most efficient and cost-effective way possible

Historically, each application — be it a server, a shared network drive or an end-user system — had its own storage associated with it. This forced IT departments to manage each piece of storage directly and separately. Plus, most of this storage went unused for great lengths of time.

Storage Area Networks — This architecture, for example the HP LeftHand P4000 SAN Solution, can be used to tackle both the management and utilization issues by centralizing disparate storage devices in a single pool of data. The storage space then can be apportioned to particular applications as needed.

With Fibre Channel and now iSCSI technology making access over the network as fast as local access, a SAN can present just the right amount of storage to each application. And it can do it in a way that is completely transparent to users.

Storage Virtualization — These benefits are multiplied with the implementation of storage virtualization over the SAN. It abstracts the use of storage from the actual physical media on which data is stored.

"Storage virtualization is a technology that allows you to separate the hardware constraints of your storage system from the actual capacity that's being used," McCall says. "If you think about storage as a resource, what storage virtualization does is it separates that resource of capacity from the actual physical implementation."

Instead of assigning physical storage space to each app, storage virtualization creates virtual drives that appear real to the application. Still, they can be moved from drive to drive or server to server, spread across multiple devices, or reduced or increased in size.

Using tools built into SANs, such as the HP StorageWorks SAN Virtualization Services Platform or IBM's TotalStorage SAN Volume Controller for storage virtualization, the same physical resource can be assigned to multiple applications that at any given moment are not using the full capacity available to them.

The storage gains are cumulative. "If you go from direct-attached storage or storage attached to a server to a SAN, you can increase the utilization of your storage from 15 to 20 percent to about 35 or 40 percent," McCall says. "Storage virtualization takes what a SAN would do and increases utilization even further."

Storage virtualization also allows tiering of data, moving files to the most appropriate form of storage without interrupting work processes that rely on accessing that data. "What you're hoping to do is move growth off of your most expensive top-end storage to archive for cheaper storage," advises Newman.

"The idea is to move it off of very fast Fibre- Channel or solid-state disks that are geared toward performance and availability and onto an archive storage tier," he adds. "So you have a good balance of total cost of ownership on the environment."

Data Deduplication — This technology makes archival and backup storage even more efficient by minimizing the storage needed to house data that exists in numerous copies across the enterprise. For example, consider an image of the corporate logo.

It might be used in thousands of files — on letterhead, in e-mail signature blocks, in PowerPoint templates and elsewhere. Traditional compression can reduce the size of each of those files independently, but you're still saving thousands of copies of the compressed file.

Data deduplication reduces the data that makes up a file — or a part of a file — to a single copy that is then referred back to in all the places that data originally appeared. "Data deduplication is making sure that you're making good use of information and not having multiple copies," Newman says.

This smart storage use becomes especially important in a virtualized server environment, where there may be thousands of identical virtual machines, each generating data. "You don't want to back up the data associated with the operating system," he adds. "You really want to make sure that you're only backing up the operational data associated with those virtual machines."

» FUTURE PLANNING

The amount of data companies need to manage is going to continue to grow, and grow erratically. For IT departments, coping with that growth means grappling with the behavior of the workers they serve, as well as the technological demands of their storage needs.

"When I talk to clients about storage management," Sloan says, "I talk about the one-two punch of storage management."

First, a company must efficiently manage its data creation, use, archiving and destruction. "This is about policy and working with the owners of the data to assess business value," he says. The business value assigned to data will drive policies around backup, archiving and deletion.

Second, the company must gauge the application of various technologies to make sure that no dollar spent on storage is wasted. "SANs and storage virtualization are about pooling storage resources so that management, as well as backup, is not fragmented and replicated across an archipelago of server-attached storage islands," Sloan adds.

Because the choice of technology must follow from the behavior of users, it is difficult to predict the amount, value and nature of the data that users will be producing in years to come. For that reason, storage planning needs to remain as flexible as possible.

»» Managing storage costs seven times as much as the physical storage media itself.

Source: EqualLogic

Systems such as SANs that allow new storage to be attached on the fly and that can incorporate new and legacy devices, regardless of their manufacturer, offer a lifeline to IT teams charged with keeping up with their organizations' changing data storage needs.

Manufacturers design current storage tools with growth in mind. Even so, HP's McCall says, there is no one-size-fits-all solution. Taking a careful look at your environment and how your users create and use information is a must-do exercise. ◊

EFFECTIVE STORAGE MANAGEMENT IS PART OF INFRASTRUCTURE OPTIMIZATION

Consider storage management a part of best practice technology approaches for optimizing a firm's infrastructure and providing efficiencies and flexibility. The components include:

Storage Management
- Efficiencies in storage management systems — storage virtualization, data deduplication and SAN environments

Server Virtualization
- Efficiencies as well as business innovations enabled through a virtual environment

Client Virtualization
- Best practices for virtualizing client systems and applications

Data Center Optimization
- Improve efficiencies and flexibility through blade technologies and power and cooling management