

# Developing a Data Protection Strategy

[eWEEK, January, 2002](#) by [Henry Baltazar](#)

In the end, the fate of businesses rests on data. More important than the processing of new data or its movement through the network is the task of keeping data available and uncorrupted.

There are vast differences in cost and technology among data protection solutions, but each of them makes some trade-off between resource costs—the costs of bandwidth, capacity and various forms of redundancy—and the business value of assured recoverability.

When developing or updating a data protection strategy, there's no better evaluation tool than solid knowledge of your organization's mission dependence on data. The level of importance of that data, and of its availability on various time scales, will be the ultimate guide for choosing what technology or combinations of technologies you need to implement. In this time of ever-tightening budgets, the importance of this one rule cannot be overstated.

## Offsite Storage

Every company, regardless of size, should have a removable media backup system with offsite storage for media. By setting up an offsite storage repository, IT managers can ensure that they will be able to recover most data in the event of a geographic-level disaster (such as fire or flood), where all co-located servers and storage facilities are rendered unusable.

It is important to note, however, that there will be a gap in data—possibly crippling—if backups are the only form of data protection employed. Backups are really nothing more than snapshots of data taken at scheduled intervals; data created or modifications made between the time of the last backup and the disaster is lost forever unless it can be reconstructed from transaction or other activity records that might be maintained by other parties.

For most companies, magnetic tape is still the removable storage medium of choice.

Magnetic tape has a long shelf life—usually a couple of decades—which makes it a good medium for archiving data. And newer tape formats, such as SuperDLT (digital linear tape) and LTO (Linear Tape Open), have larger capacities than previous generations—allowing storage of more than 100GB of data on a single tape.

Tape's data transfer rates are relatively slow, with the fastest midrange drives transferring data at about 11MB to 30MB per second. Depending on the size of the data set, a full data restore from tape can take anywhere from a few minutes for a small server to a couple of days for large data centers.

The three main technologies to watch in the enterprise-class range are Quantum's Corp.'s SuperDLT, the new LTO standard (from a vendor consortium composed of heavyweights such as IBM and Seagate Technology LLC) and Storage Technology Corp. (StorageTek), with its T9840B Fibre Channel tape drive.

High-end magnetic tape drives such as the T9840B can manage data transfer rates of roughly 19MB to 70MB per second, depending on the compressibility of the data being backed up: Traditional business records tend to be highly compressible, but multimedia files with their own compression schemes leave little room for additional compaction. However, these drives have somewhat limited capacity (20GB native and 40GB compressed capacity per tape).

In the workgroup/midrange market, there is a wide range of tape products, including Sony Electronics Inc.'s AIT tape drive and offerings from smaller vendors such as Ecix, OnStream Data B.V. and Benchmark Storage Innovations Inc. These products are less expensive than enterprise-class offerings but by design are slower and have smaller capacities.

Magnetic tape drives are expected to hit capacities of 1 terabyte and speeds of 100MB per second in 2007, according to projections listed on Quantum's SuperDLT's road map (see Web resources list, next page).

However, with the capacity of IDE hard drives doubling every year or so, IT managers will soon have another cost-effective alternative to magnetic tape as a backup medium. Because IDE hard drives can provide direct data (unlike tape solutions, which must be restored to a hard drive before users can readily access data), IDE hard drive solutions will be able to provide data on demand, even when primary storage solutions go down.

Although tape has a longer life span, hard drives can be protected with RAID to ensure hardware fault tolerance.

Network Appliance Inc.'s forthcoming NearStore line combines storage management software with an inexpensive IDE drive to provide consolidated backups and rapid data recovery. NearStore will begin shipping in the first half of this year, with the ability to scale from 12 terabytes to 100 terabytes at a cost of roughly 2 cents per megabyte.

The software you invest in will also have a significant impact on the success of your backup solution. A key thing to look for is a system that will allow the management and backup of multiple platforms from a single console, which will significantly reduce management overhead.

eWeek Corporate Partner Robert Rosen, CIO of the National Institute of Arthritis and Musculoskeletal and Skin Diseases, uses Computer Associates Inc.'s ArcServe and scripts for backups, but he said the NIAMS' long-term plan is to set up a consolidated backup system.

Another feature to look for when evaluating backup software is snapshot backups, a technology that allows the backup of data while an application is still running. Usually associated with database backups, snapshot backups are a necessity for companies that have applications that must run 24-by-7 and are especially effective for protecting e-mail servers. Backup snapshot capabilities are available as an add-on to Veritas' NetBackup system, for example.

Fibre Channel-based backup systems, which off-load backup traffic from IP networks and onto Fibre Channel SANs (storage area networks), have grown in popularity during the last few years. In addition to off-loading the IP network, Fibre Channel SAN backups—when properly designed—allow the sharing of expensive resources such as tape libraries and Fibre Channel networking equipment. Before investing (switches can cost from \$10,000 to \$100,000, depending on the number of ports, and host bus adapters cost around \$1,000 per server) make sure that your backup system can support all of the platforms in your network.

eWeek Corporate Partner Gary Bronson said the move to Fibre Channel will allow his organization to consolidate all backups. "With the recent implementation of our Compaq SAN, we are planning to implement centralized backup for all systems in our data center, including Windows NT and Solaris systems," said Bronson, enterprise operations manager, information technology, at construction and engineering firm Washington Group International Inc., in Boise, Idaho. "We focused on backing up the Solaris systems first and NT within the next month."

One of the main benefits of Fibre Channel is its ability to support network links several kilometers long. This allows backup sites to be set up far from the data center, which provides fault tolerance in the event that the servers and storage units in the data center are damaged.

### **Business Continuity**

Although backups are sufficient for most sites, enterprise sites that cannot afford to lose any data or suffer any downtime need to implement business continuity solutions, such as remote mirroring or electronic vaulting.

To protect key databases, Bronson uses multiple data mirrors to limit downtime. "On the system on which we run our production instance of Oracle Financials, we mirror three disk arrays," he said. "We disassociate one array at night, then perform an offline backup of that array and reassociate after the backup is complete. Our downtime is about 15 minutes during the 24-hour period while we shut down the database."

Mirroring solutions are extremely expensive to implement because a high-throughput network link with low latency must be set up between the mirrored pair to maintain synchronous communication. Because of these stringent guidelines, leased lines are almost always required between non-co-located mirrored pairs.

A successful disaster recovery system goes beyond products, installation and maintenance to include regular testing and verification.

To bulletproof the system, test to see how long it will take to recover data and if data will be corrupted during the backup process. Recovery tests should be run fairly often—quarterly, at least.

In addition, an effective disaster recovery system is the result of thoughtful threat assessment. The terrorist attacks of Sept. 11 were an all-too-extreme example of the importance of a data protection and recovery scheme, but IT managers must neither overreact nor underreact to threat.

This task has been made more difficult by organizations that have used the tragedies to market products, but keeping sight of what your data means to your company will allow you to take the necessary precautions to safeguard it.