

Redo The Data Vault

Deduplication technology lightens the data burden.

BBUSINESSES DON'T NEED the annoyance of increasingly costly data storage to go along with a flagging economy. They're asking their IT departments to find more efficient ways to retain and back up their data. Many have turned to data deduplication technology for the task, finding that it continuously sorts and trims unnecessary or redundant information, saving time and resources.

Deduplication is not a replacement for data backup, says Noemi Greyzdorf, research manager at IDC. Rather, it helps decrease both the data footprint and the amount of media used to store data, Greyzdorf explains.

Ultimately, the technology moves data offsite from backup tapes to disk media, which is more compact and can dramatically cut down the volume of backup tape used, Greyzdorf says.

"Data deduplication technology replicates only unique data and ignores repetitive data," she says. "It replaces repetitive data with a marker to the original pertinent occurrence."

In other words, an e-mail attachment won't be stored 4,000 times when it's sent to 4,000 recipients. With data deduplication, only one copy of the attachment is stored. If that attachment is 1 megabyte, that's a data savings of 3,999MB — or about 31 gigabytes — of storage capacity.



Derek Neidermayer (left) and Bryan Nash of McHenry Savings Bank use deduplication to increase storage, streamline data management and free the company from its old tape system.

For businesses paying for magnetic tape storage, eliminating the space required to store all that tape and the time it takes to transmit all that redundant data can add up to substantial savings.

Time Is Money

McHenry Savings Bank in McHenry, Ill., had been backing up image data every day from all of its remote locations. It had been using tape backup from those locations, employing couriers to move those tapes to various locations for storage, says Bryan Nash, senior vice president of information

technology at the bank. "We wanted to get rid of tape," he says. And with the decision boiling down to efficiency and time, the bank chose to switch to data deduplication.

"Using tape to back up thousands of images per day from branch offices at the institution's data center was a horror story," says Nash.

The company chose a Data Domain DD510 deduplication storage system. The system uses inline deduplication, which monitors data before it's stored on disk. Another popular deduplication technique, called post-processing, analyzes data after it's stored.

HARMONY MEANS EFFICIENCY

Coordinating all the moving parts when switching over to a data deduplication system is a challenge, say those who have implemented the technology.

For instance, McHenry Savings Bank had to tweak its data flow among branches to wring the most out of its new system. "If we had a problem, it was properly setting up how the data would multipath and stream all of our data efficiently" from the bank's central data center to other branches, says Derek Neidermayer, network support supervisor for the bank.

"Properly using all of our available bandwidth and meeting our data backup window, Data Domain worked with us extensively on this problem, and with the help of their support staff they set the device to effectively use all of its capacity, giving us more ability than we had first projected," he says.

It wasn't all smooth sailing for Pro-Dex either. The developer and manufacturer of technology-based solutions needed last-minute adjustments to its data backup software to optimize its deduplication efforts.

"Historically, the backups were maintained separately at each site. With this solution we needed to bring all the backup servers to the same version of Symantec Backup Exec to ease the management of the backups," says Jamie Rosewitz, IT manager for Pro-Dex.

"The initial backups to the devices were simple and straightforward," she says. "Most of the issues we ran into were due to the Backup Exec software not being installed and used the same way at the three sites."

The inline capability, coupled with virtual private network connections in a hub-and-spoke network with branch offices, allows the bank to do its daily backup in a flash, says Derek Neidermayer, network support supervisor at the bank.

With nonessential data winnowed away by the new system, the bank's storage capacity jumped to 3 terabytes, and data backup took substantially less time to complete, he says.

Ditch the Tape

For Cozen O'Connor, a Philadelphia law firm with offices throughout the United States and in London and Toronto, data storage tape had become a crippling issue. The firm was producing 41 data backup tapes per day and sending them offsite for storage. "That meant a lot of man hours and a single point of failure" per tape, says Thomas Markward, network administrator.

The firm wanted a way to centralize data backup, storage and disaster recovery, so last December it decided to implement a deduplication system in its Philadelphia office to replicate backup data from all of its offices on disk, and then transfer it to Chicago instead of transferring tapes among offices, Markward says. The company met their April 21 deadline to go tapeless, he says, with the Chicago office receiving 4TB of data every night for backup.

Although the technology has substantially increased the efficiency of its data

backups, the firm had to iron out some immediate glitches, adds Markward. Initially, it couldn't transfer and back up data to the new system without significantly impacting its ongoing operations. Transferring to the deduplication system meant a period of overlap between the new and the old systems. Transferring data to both couldn't be done at the same time without significant delays. A special backup system had to be set up one weekend to more efficiently coordinate the operation, he says.

At Grey Healthcare Group in New York, "our expenditure on tape went from \$60,000 to \$80,000 a year to about \$15,000 per year," says Chris Watkis, information technology director. Watkis says the company installed a FalconStor data deduplication appliance about two years ago because it found that backing up the growing volume of data in its customer records, marketing information and high-definition video records was wearing down its traditional tape drive and using up to 15 tapes per day. The FalconStor appliance has smoothed the process and added more functionality to how data can be manipulated and stored.

Pro-Dex, based in Irvine, Calif., develops and manufactures technology-based solutions that incorporate embedded motion control, miniature rotary drive

systems and fractional horsepower DC motors for the medical, dental, semiconductor and aerospace markets.

The company was looking to smooth operations among its offices across the country. It installed three Quantum DXi3500 data deduplication and replication appliances in data centers at its offices in Irvine, Beaverton, Ore., and Carson City, Nev., and networked them for data backup capabilities.

Historically, says IT Manager Jamie Rosewitz, Pro-Dex maintained backups separately at the various sites. The deduplication technology let the company coordinate backup software on servers at all locations, increase efficiency and eliminate redundancy.

"First, I looked for a solution that would provide the same level of backups that

I currently had using tape and wanted to automatically offsite the data at the same time," Rosewitz explains. "Second, we wanted to be sure the circuits between the sites could hold the additional data and not hurt production during high usage hours. Third, it needed to be a good value solution."

"We phased out tape backup and greatly reduced the amount of data that needed to be replicated through the deduplication technology," Rosewitz concludes. **[BT]**

Deduplication is similar to data compression, but it looks for redundancy of very large sequences of bytes across very large comparison windows.