Special Advertising Section

NETWORK Storage QUARTERLY

The Definitive Guide to Storage

The 4th report in a continuing series

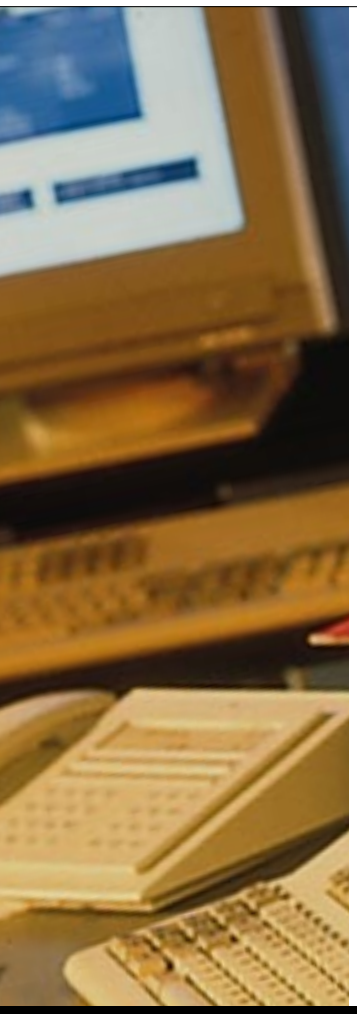
Backing Up A Database Can Result In Terabytes of Headaches

By David Braun

One of the most popular and effective ways to keep a Web site (or an Intranet or Extranet site, for that matter) interesting and useful is to link it with various corporate databases.

But the challenge of backing up and maintaining such constantly changing databases is quite a nightmare, even if the Internet wasn't making matters more complicated. To compound matters even more, databases themselves are growing at an exponential rate.

Storage



This report will also appear on Network Computing Online at http://techweb.cmp.com/nc/marketing/s_toc.html

There are several enterprise-class storage management products that cover hot backup for the major databases, said Mark Nicolett, a Gartner Group research analyst in storage technology, operations and resources. "So backup is not the main problem. Rapid recovery is."

The issue of backing up an application while it is active is essentially solved as long as there is enough capacity, CPU and I/O to handle both the backup load and the application, Nicolett said.

"The problem is recovery time. The bulk movement of data to tape requires hours, usually, if it is a large database. An organization needs to be willing to tolerate recovery times in terms of hours, even as much as a day, depending on the configuration."

But Internet access to database applications by outside customers has resulted in undefined access requirements. There is no schedule, so companies require very high availability for their database.

"What we think will happen over the next few years is that more and more organizations will look at point-in-time replication as being the first source in a recovery," Nicolett said.

Instead of having to do bulk data movement from tape for the whole database and then rolling the logs forward, most companies are going to have to make periodic point-in-time replicas, at least for the portion of the applications that have rapid-recovery requirements. Recovery would then proceed.

"Right now, for the most part, we are talking about physical duplication of the data," Nicolett said.

Raj Sehgal, product manager for Sybase's adaptive server enterprise unit, agreed that the recovery window is as important as the time and throughput of the backup process itself.

"A lot of our customers actually set up their criteria in terms of what's more important to them: the restore window or the backup window," he said. "The whole purpose of backup is to guarantee that you can recover the data in a timely fashion. This is so that you don't actually disrupt your operations should you have a corruption or some hardware failure that's not covered by some other availability method such as hardware redundancy, clustering and things like that."

Backing up large open databases involves hardware, software and sometimes the network itself.

Of these components, software is the most important factor today, said Marc Farley, director of marketing at Crossroads Systems and author of several books on network backup.

"The software drives the intricate process of database backup and determines how easily restores are done," Farley said.

The difficulty begins with the need to keep the integrity of the database intact as transactions are made during the backup process.

"To achieve high performance, the backup software reads large blocks of data that are read sequentially. Backup software has no view of what kind of data is held within those blocks. Some of them may be pointers inside the database, while other blocks may be records," he said. "As backup proceeds sequentially, database updates are written randomly on disk, making it likely that a single update operation could change some disk blocks that have already been read for backup and change

> **"an organization needs to be willing to tolerate recovery times in terms of hours, even as much as a day."**
>
> *—Mark Nicolett, Gartner Group*

Author Marc Farley argues that traditional cold backups are no longer practical.

other blocks that have not been read yet. Without some way to account for this, the backup copy won't have integrity. If you don't clean out the whole garage in the one chance you have, you end up with junk."

The common practice for years has been to shut down the database and run a cold backup covering the entire database. If the user doesn't have the time to backup the whole database, copies are made of tables and a schedule is created to copy them at different times. For many customers, cold backups are not practical anymore due to high availability requirements.

The challenge therefore is to run hot backups when the database is open for updates, Farley said. "The standard approach is to try to take a snapshot of the database at a particular point in time, between updates. This means that database updates that occur during backup must not impact the integrity of the data."

This is typically done using a form of caching or disk redirection of the updates that occur during backup. After the backup is done, all the suspended updates are applied by the database.

"A problem with this method is that if this whole system goes down in the middle of the backup process, you need to be able to ensure the integrity of the suspended updates that may be pending," Farley said.

Another way is to copy-on-write, which is done by temporarily suspending writing disk updates until after the old data from the block is written to cache somewhere. When the backup software comes to a block on the disk where an update has occurred, it instead reads that information from the cache.

Copy-on-write capability is enabled for databases through a set of APIs provided by the database vendors and used by backup software companies.

"The database companies themselves have to provide a facility that enables backup software companies and end-users to work within the intricate architecture of each database system," Farley said. "That is the only way to achieve the tight integration that's needed."

Oracle developed such a facility for its databases several years ago and continues to work on it. Informix and Sybase have been slower off the mark, but now understand that customers will not buy their products if they do not address backup and recovery adequately, Farley said.

"As a result, both have provided better support in the latest versions of their products," he said.

Back-up TIPS

Crossroads Systems

Marc Farley, Director of Marketing

"I don't tell my clients to remove bottlenecks because, when you try to find a specific bottleneck, you discover another one right around the corner. Instead, I say to buy everything as big and as fast as you can right from the start. Brute force always wins in backup."

Oracle's Merrill Holt, director of enterprise OLTP for Oracle's server group, said that he couldn't address what either Informix or Sybase are doing, but stressed that Oracle certainly is providing such integration.

"We provide the interface boundary not at the operating system utility point, but with an API that integrates more tightly with the various storage sub-system vendors," Holt said. "As we've moved from Enterprise Backup Utility (EBU), it's primarily been a certification effort to make sure that the media-management product vendors that have been integrated with EBU continue to work with Oracle8."

Backup interfaces such as Oracle EBU, Oracle8's Recovery Manager, Informix's On-Bar, IBM's DB2/6000 backup utility and Sybase Backup Server are all involved in preparing data streams of database transactions to the storage management

Storage

product for backup, said Denise Reier, vice president of marketing at SCH Technologies.

"From there, the backup product must efficiently move the data to tape as quickly and efficiently as possible. When done correctly, this provides efficient backups of critical data and allows for 24x7 availability."

Oracle8's Recovery Manager maintains detailed information on when backups are performed, exactly which parts of the database are backed up, and where the files are stored, Reier said. "Should a recovery be necessary, Oracle8 analyzes the state of the database and determines the operations necessary to repair the database. It also minimizes the space needed for backup creation by supporting incremental backups (even when the database is open) and thus reducing the amount of time backups take as well."

Sybase's Sehgal disagreed with Farley that his company has been slow in providing interfaces for software companies and end-users to work within the architecture of its database.

"We have focused on this online backup issue probably earlier than most database vendors have," Sehgal said. "Perhaps it is something we certainly have not marketed as significantly as Oracle has with their backups in recent years."

For a database product without hot backup capability, a method called extraction might work, Farley said.

Extraction depends almost completely on the database engine to deliver a backup copy. Although this ensures integrity through the internal workings of the database, it also creates an enormous load on the database itself. Instead of copying blocks of data from disk, a query is submitted to the database to get a picture of what the database looks like.

"You end up getting the database's logical representation of what's inside it, as opposed to its physical representation and it takes a long time to do," Farley said. "It's very time consuming, creates a huge load on the database engine and, frankly, it's really not practical for today's environments."

Users will extract what they need to another disk volume or another machine and then make a backup copy of that. It's fairly simple to extract specified tables so the entire database doesn't have to be done. The database administrator may, for example, decide that one or two tables out of 25 or 50 are the most important. They can run an extraction on just those one or two to make copies of them and back them up and get the others later, in different cycles.

If the database should fall apart, it

Database Backup Techniques: Pros and Cons

ADVANTAGE	NAME	WEAKNESS
Allows for a hot backup with incoming data cached for later synchronization with database.	Copy-on-write	Not yet widely commercially available.
It backs up only those parts of the database deemed critical.	Extraction	Time-consuming, creates huge database load.
Transactions continue to accumulate on the primary disk while the mirror is taken offline for backup.	Mirroring	Tricky synchronization of primary disk with mirror when mirror comes back online.
Allows for backing up only the changed portion of the data	Incremental database backups	Useful only where relatively small amounts of data change as a percentage of the overall database
In some cases with the very largest databases, cost and time may make backing up inefficient, if the data exists elsewhere.	No backup at all	The risk of losing data due to device failure or corruption.

Storage

should be possible to rebuild it from the different components. "Rebuilding it will be quite a strenuous operation, but that's one of the reasons database administrators make so much money," Farley said.

A third way to backup open databases is to use mirroring technology. Data is written to two different physical disks and then the mirror can be taken off line and copied.

Several hardware vendors provide mirroring functions, said Gartner Group's Nicolett.

"EMC's TimeFinder is an example of a point-in-time mirror that can be accessed by an application so you can basically get this replica and then access it to back it up to tape," Nicolett said. "This is not in lieu of tape backup, but it is hopefully the way the database would be recovered."

"Of course, what happens is that when that mirror goes off line, you need some way to synchronize the primary with the mirror when the mirror comes back online, and this is a different trick," Farley said. With mirrored disks, all the transactions continue to accumulate on the primary disk. "It's not a database synchronization anymore, it's a disk synchronization," he said.

TimeFinder mirrors to multiple volumes so it has at least three RAID cabinets mirrored. It always has one mirrored pair in case one of the RAID systems crashes, and then it has an additional mirror to take off-line to do the backup. Bringing it back online requires resynchronization of a mirrored pair with another RAID subsystem. This synchronization does not work on a logical level to process database updates. Instead, it must synchronize block updates.

Back-up TIPS

NSI Software

Mike Kidd, VP of Marketing

"The time required for a tape restore is not the tape's raw data rate and it's not the typical backup/restore I/O speed. In reality, the big factors are getting the original or replacement hardware configured and working, and then going through the human steps that need to be done."

NSI's Mike Kidd asks how much data can you afford to lose?

"I don't believe mirroring solves the backup issue," said Sybase's Sehgal. "All that mirroring does is solve the availability issue. If you have system A and you mirror to system B, you still need to address the same set of issues in backing up system B as you would have system A. All you've done is provide higher availability."

"To me, the major issue is still backup throughput and how quickly you can restore," he said. "So there's a backup window and a restore window and how your system is affected while you are doing your backup."

With the very largest databases, people will sometimes actually go without backing up because the data exists in other places, said Oracle's Holt.

In systems with a lot of redundancy at the hardware level, the risk of losing data due to device failure is low. In rare cases of database corruption, it is possible to go in and patch things such that the user does not lose the data completely, Holt added. The question is whether "it's actually worth the risk to not do a backup because the cost and time of doing backups is quite large."

For extremely large databases, the cost of backup media can also be a factor.

Holt said, however, that backup capabilities are advancing to the point that even the largest systems will find it worthwhile to do backups. "Tape devices are getting much higher speeds, so I think we will see that capacity improving so people can do backups."

Incremental backups are also a growing trend, Holt said. "For a very large warehouse, you can do a full backup at one point and then just do a series of incrementals, especially in

Storage

cases that have just relatively small amounts of data change as a percentage of the overall database. The incremental strategy can work well in terms of reducing the amount of time as well as data that has to go with that backup."

One advantage of boosting storage-layer capability, Holt said, would be to reduce the demands on the CPU and memory associated with backup.

"For example, if the recovery manager can issue a command to the storage system to backup a file, as opposed to reading the information off the disk and handing it off to the storage subsystem to put on tape, and that can be done at that lower layer of the system, you potentially avoid the CPU cost and memory associated with backup," he said.

Storage sub-systems are already getting more intelligent in terms of being able to do things like maintain copies of changed blocks down at the block level, Holt said.

"If I say 'Start at this point and maintain a snapshot of this particular file,' the storage subsystem itself will maintain the image of the disk blocks as of that point in time and maintain a set of changed blocks subsequent to that," he said. "I could then interact with the storage subsystem and say, 'Give me all of the changed blocks for this particular file' and it could do that extremely rapidly and put it out on tape."

From a hardware perspective, consideration must be given to the potential limiting effects of the server's CPU, memory, disk and tape devices, and their associated I/O paths.

"The ideal software performance level can only be achieved by selecting a backup product that provides tailorable software components for maximum throughput in the backup process," said SCH's Reier. "The desired approach is to maximize disk-read and tape-write data rates, through multitasking, parallelism and data stream consolidation."

When selecting a backup product, consider how it is currently implemented at customer sites including the amount of data to be backed up, the customer's availability requirements, and the impact on CPU and network performance.

Vendors that traditionally developed great products for the smaller organization ended up with a product that was simple to install, yet lacked the sophistication of enterprise-class products, Reier said.

"Consequently, the architecture of these products do not scale up, and are unable to address the performance, availability, data integrity and platform coverage required," Reier said.

Two-tier architecture, for example, communicates file-level detail information to and from the backup and server components. This substantially increases network traffic and prohibits the product from addressing large

Legato's Ed Cooper suggests looking closely at what your datebase vendor is doing.

Back-up TIPS

Legato Systems

Ed Cooper, Director of Strategic Communications

"Look at what comes bundled with the database or application. That, generally, is a good tip as to what type of solutions the supplier actively endorses."

Consideration must be given to the potential limiting effects of the server's CPU, memory, disk and tape devices, and their associated I/O paths.

Storage

data and high availability requirements, Reier said.

SCH Technologies' HyperTape Enterprise Backup is able to minimize network traffic by building on a three-tier architecture. The product provides:

- Control Node (the component that maintains backup objects and initiates unattended backups),
- A Service Node (the server to be backed up) and the
- Backup Node (where the data is being written.)

HyperTape segregates the control and administration functions from the backup server, allowing central control over many decentralized backup nodes.

Synchronization and the technologies available to achieve it are only one of the challenges facing database administrators wanting to backup open systems. To Farley, the next big hurdle is scalability.

"Quite frankly, the thing that breaks all database backup systems is that they don't scale," he said. "If the databases get bigger and the availability requirements go to zero down time, scalability is really important."

Then there is the matter of bottlenecks. "I don't tell my clients to remove bottlenecks," Farley said, "because when you try to find a specific bottleneck, you discover another one right around the corner. Instead, I say, 'Buy everything as big and as fast as you can right from the start.' Brute force always wins in backup."

The reason for this is that performance is never fast enough for database backup. No matter what process is used, whether it is copy-on-write or block synchronization, it needs it to run for as short a time as possible to reduce the processing overhead and risk of data loss.

SCH's Denise Reier says it's critical to maximize disk-read and tape-write data rates.

But the approach must be balanced. It's no good having a big, fast tape drive if the I/O capabilities can't drive it.

Planning for backup really should begin at the outset, when systems are being designed and installed, Farley said.

Most people, according to Farley, don't understand that backup is an extremely intense application. When people purchase database system platforms, they need to size them to enable them to be backed up quickly and not just for their expected transaction operating performance. When backup is added in, it may well require extra horsepower and memory.

"Backup is hard on a system just because of the amount of information that has to move through the I/O system," Farley said. "If there's a lot of small pieces being backed up, that means there is a whole lot of record-keeping for the backup system to do, which means that the backup system's internal database can wind up being a fairly intense application itself. It's a big load on a system and it's no wonder that problems exist because you are really loading up a system."

Farley's advice is to imagine the entire network—including all servers and PCs, machines with databases and applications systems, operating systems and file systems, fabric and storage equipment—as having the sole purpose to run backup operations. In this systems approach to backup, he said, it should be possible to see where problems could occur.

A quandary for many database administrators is that, even if they do think it through

Back-up TIPS

SCH Technology

Denise Reier, VP of Marketing

"The desired approach is to maximize disk-read and tape-write data rates, through multitasking, parallelism and data stream consolidation."

and identify their needs for backup, the technology needed may not be available.

"In the last five years, big strides have been made in the database system technology to support backup and the backup software companies have been developing good products. Tape technology is also getting faster with more capacity and more affordable," he said. The real problem today, according to Farley, is that system I/O capabilities have not developed at the same rate as all the other technology. SCSI has made strides, but has not made the big leaps that other technologies have.

Farley said the solution to the issues of I/O speed lies with Fibre Channel.

"The most difficult thing in backup is scalability, so people should plan on implementing the fastest technology that scales. That's why I'm in the Fibre Channel business, because Fibre Channel is very, very fast and will deliver scalability."

"By comparison, neither SCSI nor faster networking technology will scale for big storage requirements. That's why I urge people to plan to get something that scales and to start getting experience working with it," Farley said. "Even though Fibre Channel products are a little harder to find in the market, and the technology is still developing, you can start working with it."

Apart from scalability, Farley said, there are other advantages to Fibre Channel. "If you talk about 7 x 24 high availability environments, there's a very strong argument for Fibre Channel's superiority, too. The key element for HA with Fibre Channel is the fact that storage is physically independent of the server, so if a server crashes, that data is still available on the network and another machine can access it. In the SCSI world we have today, if a machine crashes, you lose access to its data. To get access to it again, you have to pull all the cables apart, maybe you have to pull out its SCSI cards, bring in a new machine, reinstall and configure it, attach all the cables and cross your fingers that it all works when the power comes up --it's really ugly. The fact that you can save all of this physical work by simply logically connecting via Fibre Channel significantly increases the availability."

Another common problem with backup is operators inserting the wrong tapes into tape drives, which may keep the backup from finishing. For database backup, where several tapes drives are being used in parallel, this becomes a bigger problem. Farley recommends using tape automation products because tape libraries and tape autoloaders can reduce the likelihood of having the wrong media available.

Media management also becomes an issue with larger organizations with lots of data, Reier said, especially those distributed across many servers and locations. Centralizing control and administration of media management is as important as centralized control of media policy.

"Many organizations have made substantial investments in robotic libraries requiring widely dispersed functionality such as sharing of robots, mixed media support, abstraction of physical hardware changes," she said.

The problems of backing up networks are already difficult enough without adding the complexity of database systems to the mix, Farley said.

"Originally, network database products were not really designed to be open 24 hours a day," he said.

But times change and today hot backup is a big problem for a lot of customers, Farley said. "Unfortunately, the database and backup

Storage

software technology to run hot backups is just catching up with the problem. Because of the enormous complexity involved with database integrity and rollback mechanisms that backup software must take into account, there is no one product that works equally well across the various database systems. Different databases require different products and expertise. You have to specialize to get the best results for the database that you've got."

When it comes to working with the three big Unix database companies—Oracle, Informix and Sybase—Farley said, a common goal of backup software companies is to provide consistent performance and management across all three. However, he added, compromises made for the sake of a consistent management interface might not be able to provide optimal protection equally well for each database system due to the significant differences in the backup and restore facilities provided by each database .

Farley sees Legato Systems as the most active Unix backup company developing specialized database backup products. Legato has a targeted strategy for database backup through their BusinesSuite product line and they have established both development and business relationships with the leading database companies.

SQL-Backtrack from BMC Software in Houston provides database backup plug-in modules that work with Oracle and Sybase as well as a wide number of Unix network backup products, such as IBM's ADSM, HP's Omniback II, Legato Networker, Veritas' Netbackup, Software Moguls' SM-Arch and Spectra Logic's Alexandria, Farley said.

SQL-BackTrack is available for both Oracle and Sybase and has been the leading product for backing up Sybase databases running on Unix. This has been partly due to the lack of backup facilities and APIs within the Sybase database engine, which has historically made it more difficult for software developers.

"BMC Software has done a good job serving the Sybase niche, although Sybase has developed more capability in their latest version, Sybase 11, which has already led to the release of alternative products for Sybase from other Unix backup companies," Farley said.

Sybase's Sehgal said it is true that several of the company's customers use SQL-BackTrack.

"But this is not because of our lack of backup capability or APIs, which we've had for a couple of years now," he said. "It's primarily because SQL-BackTrack does have certain management capabilities and they do other things like compression and so on. We've actually partnered well with BMC."

With regard to Oracle and Informix, these companies both have developed the APIs and internal technology necessary for third party hot backup, Farley added. "As a result, backup software companies continue to develop their own products, rather than depending completely on BMC."

Veritas, Farley said, uses a mix of its own internally developed modules and SQL-Backtrack to cover a broad set of Unix databases with its Netbackup product. Since the acquisition of OpenVision a year

"Compromises made for the sake of a consistent management interface might not be able to provide optimal protection."

ago, Veritas has been able to leverage its leading position in file system technology to build momentum for Netbackup.

Someday, the vendor said, Veritas may be able to integrate its backup and file system technology and provide advantages for database backup, but only if customers start using the Veritas File System for their database data.

Although lesser known than either Legato or Veritas, Intelliguard Software has excellent technology for backup of Oracle and Sybase databases, Farley said.

"What Intelliguard lacks in breadth and exposure, they make up for by being able to focus on the fewer products they do have."

Similarly, Spectra Logic is not as well known, but provides high performance products through add-on modules they call personalities to its Alexandria network backup product. They offer support for Oracle, Informix and Sybase. It also has a CATIA backup personality for high-end Unix CAD systems.

EMC offers an integrated network backup product called EDM, for EMC Data Manager, Farley said. EDM consists of dedicated Unix systems with attached tape libraries. It also offers hot backup capabilities for Oracle, Sybase and Informix.

"Cheyenne Software has, by far, the most comprehensive product suite for database backup in the NT market, including several important specialized backup agents for the major database products running on NT," Farley said. "Their Arcserve product for NetWare has been the market leader for several years and their NT products also seem to be doing well."

It appears that Cheyenne has shifted its focus from NetWare to NT and has become an NT centric technology company, Farley said. This puts Cheyenne in the position of developing its products using NT as its primary platform, an approach that may give Cheyenne an advantage if NT databases become more popular, because its major competitors for database backup today are primarily Unix companies.

What this means for customers is that new database technology and backup support is more likely to be developed first by Cheyenne on NT, while its competitors focus on their Unix products.

However, many of the backup companies with Unix backgrounds are also moving quickly to establish NT products.

Although Cheyenne also has Unix products, they are not generally considered to be the best of breed in that market. They have a lot of catching up to do to compete with the entrenched Unix backup companies for Unix databases, Farley said.

For databases running on Netware, the best product available is St. Bernard's Open File Manager, Farley said.

NetWare is not often used as a general purpose database platform, Farley added, but there are many vertical industry products with database systems running on Novell. St. Bernard OFM has been providing hot backup capabilities for these customers for the last three years.

There is also a version available for NT Server. It does not work to protect open files on NT workstations, although it can be administered from an NT workstation, he said.

Cheyenne also sells a product it calls the "Backup Agent for Open Files" that does

Storage

the same thing as St. Bernard's products, but has not been sold in the market as long, Farley said.

Oracle has several different technologies that give high availability at a different cost point and different response times in terms of recovery operations, said Holt, Oracle's director of enterprise OLTP.

"For example, we can exploit the cluster technologies in several different ways. One of our premier products is Oracle Parallel Server. This lets us take a cluster configuration and run a single database that has multiple nodes accessing that data concurrently in both read and write operations and the parallel server does the parallel cache management across those loosely connected nodes. It allows us to tolerate single node failures. If I have a node that fails, that database in effect continues to run and users can switch their work from the node that failed to one of the surviving nodes."

Incorporating even more redundancy and the ability to tolerate catastrophic failures, Oracle offers a standby database where the primary system is matched with a fully redundant system. This means there are two complete systems and a primary system. The logs generated by the primary system are archived and applied to the standby system. If there is a failure, there can be complete recovery on the standby system and the system can come up very quickly with the users switching over to the second system.

The disadvantage of the stand-by system is that the company has limited use of the secondary system. It can't run transactions against it while the primary system is running.

"the disadvantage of the stand-by system is that the company has limited use of the secondary system."

A system that gets around this problem is symmetric replication on two master sites. "I can be running transactions against both sites and, if either site fails, I can fail over to the surviving side," Holt said.

Other availability strategies operate below the level of the database, Holt said. "We're seeing companies such as the storage vendors providing geographical separation by duplicating disk writes down at the disk I/O level so it's transparent to us."

"One of the things we're doing with Parallel Server is making the whole cluster work like a single computer. We're also working on a variety of techniques to allow that system to scale in a write-intensive or a combination of read and write," Holt said. "So you can do things like having one of the nodes doing decisions to support operations while another is doing, OLTP-type operations. Since it's a common database, we can move data easily between tables and things like that. The ability to support both read/write and write/write on multiple nodes and that also scales well is something we're working on. That's really unique technology in the industry. The write aspects are extremely difficult to engineer."

A mirroring product called Double-Take is produced by NSI Software. Double-Take can be used to mirror a database to a second volume, which can then be used for backup purposes with minimal impact on the database system. ❖