

Tape- vs. disk-based backup, part III

Can you elaborate on some of the new, cutting-edge backup systems that you mentioned in your first column on disk-enabled backup systems?

There are a number of really snazzy backup technologies on the market today, and a number of other options will soon be available. These products rely on disk as secondary storage media and offer some very powerful advantages over conventional tape backup—even tape environments enhanced with disk staging or virtual tape.

However, as I mentioned in the July 2003 column (see p. 24), disk-enabled backup (or disk-to-disk [D2D] backup) isn't necessarily faster than conventional backup. In fact, a well-designed tape system, which takes advantage of multiple streaming tape drives, can still significantly out-perform a single disk array.

That said, there are still some clear advantages to using disk, rather than tape, for primary backup. Because the new class of disk-based products takes full advantage of disk's random-access capabilities, it can provide more flexible, efficient, and faster backup than traditional tape backup systems, which write data in a sequential manner.

Writing data in a sequential fashion requires users to back up redundant sets of data (e.g., alternate weekly full backups with daily incremental or differential backups), which means that files that may never change are backed up over and over again.

Why? Because the less efficient the backup, the easier the restore process. In tape

backup, you always trade efficiency for ease of restore. However, with inefficiency come more problems. In today's 24x7 world, it is simply not practical to run tape backups more than once or maybe twice a day.

However, if your data is changing quickly enough, nightly backups might provide an acceptable level of protection: If your system fails at 5:00 PM, can you afford to roll back to the previous night's backup?

Disk's random-access encoding, in comparison, allows for extremely efficient backups without compromise. Instead of moving large sets of data in batches (the way conventional backup software works), only the segments (or blocks) of data that have changed are moved. The backup software reassembles the encoded data into complete files during the restore process.

With this level efficiency, you can afford to back up every hour, half hour, or even continuously. A variety of new D2D products allow you to back up continuously. These products are commonly referred to as *continuous backup* or *real-time recovery* software. Some of these products claim to be more continuous or granular than oth-

If you have a question you would like to ask one of our experts, please e-mail Heidi Biggar at heidib@pennwell.com.

ers; each has unique features. However, for the purpose of this discussion, you can think of them as hybrids of more-traditional data replication, snapshots (or checkpoints), and backup/restore products.



JACOB FARMER
Cambridge
Computer

Replication—Replication software has been around for quite awhile. It monitors the volumes or file system of a disk array and then replicates changes to a second storage system. Replication systems move data in real-time as the data changes. And they are quite efficient. The one drawback: Data corruption is replicated, too.

Snapshot (or checkpoint)—The industry seems to love the word *snapshot*; unfortunately, the word is commonly misused. I prefer to use the word *checkpoint* to describe the process of preserving a point-in-time copy of the data that can be rolled back in time prior to the point of corruption or data loss. Many applications and file systems have transaction tracking, which offers a form of checkpoint. You simply back out from the logs and retrieve lost data from a specific point in time.

Backup/restore interface—If you combine replication and snapshots, you have a great foundation for a data-protection scheme. Replication gets the data off your host storage and the checkpoint allows you to roll back to point-in-time versions of data. All you need is an interface that facilitates the process. Navigating through snapshots without a good interface can be really annoying. If you take 100 snapshots

over the course of the day, but don't know exactly when a file (i.e., a spreadsheet) was overwritten, you might have to look at 100 versions of the file before you find the one you want.

While the new breed of continuous backup products has some pretty neat capabilities, none of them fully replace the need for conventional backup software. Conventional backup products offer comprehensive tools for managing backups for all major platforms and applications, and, equally important, they have withstood the test of time.

Continuous backup products still have a place, however. They might be able to help you better back up Oracle, Exchange, or remote offices, or they might offer you a bare-metal restore function or help you quickly create a test environment. And because many of these products interface well with conventional backup systems, they can be used in conjunction with existing tape backup processes to provide you the best of both worlds: super-granular backup-and-restore capability with the safety net of doing backups the old-fashioned way.

THE BOTTOM LINE

The bottom line is that continuous backup is relatively new but can be extremely powerful. My advice is to check out these products for specific pain points or areas where you are having problems meeting specific backup-and-restore objectives. Use these systems in tandem with conventional backup, or at the very least, keep your personal resume on a floppy! □

Jacob Farmer is the CTO of Cambridge Computer. He can be reached at jacobf@cambridgecomputer.com.