

INFOSTOR[®]

newsanalysis&trends

READER-EXPERT I/O

Capacity at a crossroads

I need to build a digital archive for scanned images and other content, and I am looking for the most cost-effective solution. I expect to have about 5TB today with upward of 20TB in the next three to five years. Do I need a jukebox, or can I get away with disk?

This is an interesting question because your capacity needs put you at a crossroads. You could go the IDE hard-disk-drive route or you could go with a jukebox solution that uses DVD, magneto-optical, or tape to simulate a cohesive disk volume. While there truly is no "right" answer, there are some things to bear in mind.

True, disk storage keeps getting cheaper and cheaper, but cost isn't everything (see below). If you buy IDE storage in the form of RAID arrays (as opposed to NAS servers), you can connect multiple units to a single computer to build an enormous storage system, *but* with scale comes a variety of headaches.

First, there is the issue of management. You need a file-system technology that can reliably manage multi-terabyte volumes. Second, you need a backup technology that is capable of protecting all this data without being wasteful of tape media. One way of dealing with the backup issue may be to replicate data to an off-site facility, but this drives storage costs up. Not only do you need to budget for the additional off-site storage off-site, but there is also the cost of the WAN connection.

So, while some really cool products may soon incorporate all this functionality into a turnkey package, until then you're stuck with cobbling together a complete solution.

Perhaps the biggest disadvantage of a juke-

box storage system is the simple fact that, unlike disk storage, the data isn't online, but near-line. This means there could significant latency between the time a file was requested and the actual response, depending on where the actual file resides (e.g., on disk or DVD) and how efficient your hierarchical storage management (HSM) application is.

So, if you decide to go with a jukebox system, it is imperative that you understand the usage patterns of your data so that most frequently accessed or important files are waiting on disk before requests are made. This technique, called "pre-fetching," is commonly used by medical imaging systems to ensure medical records are readily accessible.

An understanding of your data usage patterns can also tell you how much disk capacity you need to front-end the jukebox and how many drives you need to service requests. It also helps determine the type of media that you should use, which ultimately affects the storage bottom line.

If data usage patterns are hard to predict or you anticipate many users on the system, it may be difficult, perhaps even impossible, to minimize latency. In these situations, requests are queued up on a first-come, first-served basis, with latencies ranging from

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

minutes to hours.

As for backup, many jukebox solutions have the ability to copy and/or mirror their media. Thus, the cost to back up the jukebox system may amount to nothing more than the cost of the additional media. Some jukebox systems can mirror themselves to another jukebox over a LAN or WAN connection, greatly reducing the human involvement in administering the system.



JACOB FARMER
Cambridge
Computer Services

PAY ME NOW, PAY ME LATER

Though disk storage is getting cheaper and will get even cheaper over time, it is still variable. Today, 1TB of disk storage (IDE) goes for about \$10,000. A year from now, it'll likely cost you about \$5,000. By the end of the decade, some say 10TB may cost you less than \$1,000.

So, if the growth of your storage system is likely to mirror the decline in storage prices, then disk may be the right solution for you. However, if your storage needs grow faster than you anticipated, you might find yourself digging deeper into your pockets.

Jukebox systems, in contrast, hit you with an up-front cost, but give predictability in return. Because you buy your jukebox, drives, and management software up-front, the only variable cost is that of media, which, if anything, will drop. This makes it possible to safely predict the cost-per-TB of purchasing a jukebox. □

Jacob Farmer is the CTO of Cambridge Computer Services, a storage technology integrator and training provider based in Boston, MA. His team is currently writing a book on SAN and NAS technologies to be published in the spring/summer of 2002. He can be reached at jacobf@cambridgecomputer.com.