

## To collocate or not to collocate

**I need to arrange for off-site data storage for a small data center with three servers (e-mail, database, and file server) and roughly 500GB. I cannot cost-justify a pair of replicated storage area network (SAN) storage arrays, nor can I swing the monthly costs of disaster-recovery hosting services. Are there any lower-cost options?**

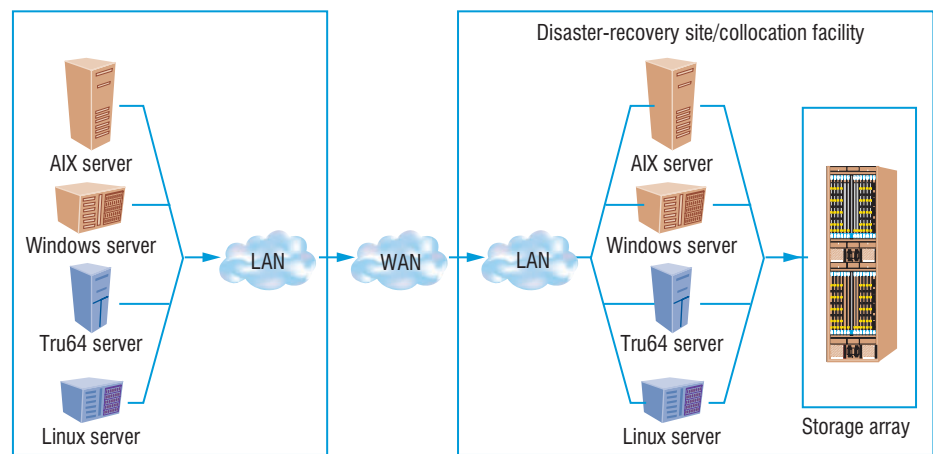
With the collocation option, your backup will be limited to the bandwidth of a T1 (1.5Mbps), so you will need software or hardware that is capable of asynchronous replication. While asynchro-



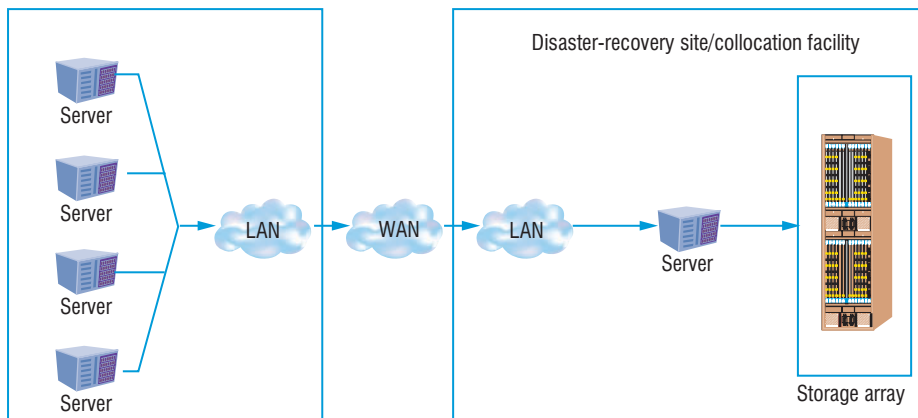
JACOB FARMER  
Cambridge  
Computer Services

There are low-cost alternatives for replicating data off-site. The obvious one is to back up to tape each day and then send your tapes off-site. But I will assume that you want something better than that—and that you are willing to pay an additional premium for that level of service. In my neck of the woods (Boston), off-site tape storage goes for about \$1,250 per month for daily pick-up. For ±\$1,500 per month, I can rent a half of a rack at a collocation facility with a dedicated T1 back to my office. I can build a mini-disaster-recovery site at the collocation facility.

### Host-based replication: One to one



### Host-based replication: Many to one



nous replication does add a degree of risk to the backup process (since there is a delay between when data is written to disk in your data center and when it is safely stored at the hot site, you will lose some data in the event of a catastrophic failure), it is less risky than relying on periodic tape backups (you will lose far less data doing asynchronous replication than you would from reverting to the last full backup on tape). And you have the added benefit of disk's faster restore capabilities.

Asynchronous replication can be ei-

ther host- or storage-based. In the host-based approach, software is installed on each server that you want to back up. The software monitors the data for changes. All changes are sent over the WAN to a second server, which then writes the changes to disk (see figure). Most host-based products allow either one-to-one or many-to-one replication.

One-to-one replication allows you to have a replica of each server—hardware and software—at the collocation facility, while many to one allows you to replicate data from many servers to one server. The data from each source server can be written to separate directories or disk volumes on the target server. Many-to-one systems typically require the servers to have the same operation systems. Since you have only three servers, the one-to-one approach may be right for you.

Replication software is usually priced per server, typically running from \$2,500 to \$7,500, though high-end servers may cost more. Additional expenses include the cost of buying a server (or servers) for the collocation facility and the cost of the storage. For storage, I recommend investigating today's new breed of ATA storage arrays. You can get a rock-solid half-terabyte of storage for about \$5K.

As for storage-based replication, you've got several options. In these types of configurations, the storage system monitors changes made to disk and then forwards them asynchronously to a similar storage system on the other side (see figure). This capability is associated with high-end enterprise arrays, but it can also be performed by storage virtualization systems. Storage virtualization software mimics the functionality of high-end storage arrays using off-the-shelf computers and storage devices.

If you are not interested in replacing all of your storage with a virtualization system of some sort, you might consider deploying a simple virtualization system and using host-based mirroring to copy data into it. That is, one side of the mirror is your existing storage and the other side is a partition inside the virtual storage array. Once data is in the virtualization system, it can be replicated to the collocation facility. While this type of configuration could easily cost you upward of \$100K (for all the

necessary hardware, software, and services for virtualization), the result is a very powerful, flexible system that might have other cost justifications.

If neither of these options (host-based or storage-based) appeals to you, you will soon have a third alternative: in-band data replication. In this scenario, an in-band data replication appliance monitors traffic to your existing storage devices and replicates it asynchronously over the WAN. Look for products like these over the coming months.

Whichever solution you choose, don't forget about security. After all, you are copying all of your company data to a remote location. If you only take half of a rack, someone else has the other half—and they could gain access to your data. What can you do to safeguard your data? Find out next month. □

**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](mailto:jacobf@cambridgecomputer.com).

*If you have a question you would like to ask one of our experts, please e-mail Heidi Biggar at [heidib@pennwell.com](mailto:heidib@pennwell.com).*