

DEDUPE—What to Use and Where to Use It

Data deduplication can reduce the amount of storage space from a factor of 10 to a factor of 100. So why aren't more channel pros talking about it? By James E. Gaskin

THE COST OF DISK STORAGE for backup devices keeps dropping, but not nearly as fast as business data is growing. Data deduplication and compression can reduce the amount of storage space needed by a factor of 10, up to a factor of 100 (or more), keeping backup storage costs down. But even more important in today's distributed environment is reducing the amount of data transmitted to backup destinations, whether to another office or to a cloud backup service.

As with most technologies, choosing the best dedupe option is a trade-off based on specific needs. Anurag Agrawal, CEO at SMB

research firm Techaisle LLC, says, "People talk about the cloud, about virtualization, but not about dedupe. You need to spend some time to get to know it."

Examine your basic data safety procedures, including disaster recovery, archival copies, and data compliance mandates, and data deduplication justifications will appear. Common reasons to jump into dedupe are reducing storage costs, shrinking backup windows, regulatory demands for ever-longer data retention, and the always popular stretching the backup budget by reducing disk needs and rack space, as well as power and cooling costs.

Trade-off decisions include where to perform the actual data deduplication, whether to use dedicated hardware or just software, and integrating dedupe with virtual environments. Add another critical requirement: Data deduplication means combining multiple backup locations to a single location. The dedupe storage device must be absolutely reliable, because the *only* backup copy of much of your data will be in that location.

Where to Dedupe?

Different vendors offer dedupe tools for different points on the network. Data can be deduplicated and compressed at the client, the server, an appliance on the network, or by the endpoint storage device. "Dedupe really helps the bandwidth requirements of moved data between devices, from corporate to mobile devices, and the like," says Agrawal.

In one presentation, the Storage Networking Industry Association (SNIA) outlines eight different dedupe points in a typical remote office linked to a data center. Data to

be transmitted over the network should be compressed before transmission, either by the client, an appliance, or a gateway. Data stored in the cloud should be compressed and encrypted before transmission.

The most common areas for deduplication on the remote side are at the client with a software agent, a software appliance (particularly with virtualized servers), or a hardware appliance (inline deduplication). At the data location, the most common locations are at the storage system with specialized software to compress data after storage (inline), at the gateway of a storage system, or at a hardware appliance on the network.

Tools for Dedupe

Leading the dedupe market, with a 60 to 65 percent share, is Hopkinton, Mass.-based EMC Corp. "EMC's Data Domain and Avamar lead the market," says Deni Connor, founding analyst with SSG-NOW. "But dedupe is becoming a checklist item for users who want to save on storage space. Appliances such as the Dell DR4100 or Quantum's NDX-8d are catching on in the SMB." EMC purchased Avamar Technologies, a dedupe specialist, in 2006.

"It's up to the vendor and channel partner to convince their customer whether to go with a hardware or software solution," says Agrawal. "Hardware solutions are kind of winning at this stage because those vendors are [well] known."

(A deduplication solutions chart appears on page 30.)

JAMES E. GASKIN is a freelance writer and former reseller based in Mesquite, Texas.

SOURCE: IDC, WORLDWIDE PURPOSE-BUILT BACKUP APPLIANCE 2012-2016 FORECAST AND 2011 VENDOR SHARES