Backup vs. archiving: It pays to know the difference

By Andy Richards, Plasmon PLC
July 13, 2005 12:00 PM ET

The difference between backup and archival storage

There is often confusion between a data archive and a backup. A classic backup application takes periodic images of active data in order to provide a method of recovering records that have been deleted or destroyed. Most backups are retained only for a few days or weeks as later backup images supersede previous versions.

Essentially, a backup is designed as a short-term insurance policy to facilitate disaster recovery, while an archive is designed to provide ongoing rapid access to decades of business information. Archived records can be placed outside the traditional backup cycle for a long period of time, while backup operations protect active data that’s changing on a frequent basis.

Backup and disaster recovery requirements

- High media capacity
- High-performance read/write streaming
- Low storage cost per GB

Performance is an important factor for backup, but since most backup operations involve large data sets, the ability to quickly stream information to and from the backup media is a first priority. Fast random access to small data sets during restore operations is typically less important. As an insurance policy, it is also necessary to minimize backup expense by reducing the cost of each stored record. The media of choice for backup and disaster recovery applications has traditionally been magnetic tape since it satisfies the performance and cost criteria of most organizations.
Archive requirements

- Data authenticity
- Extended media longevity
- High-performance random read access
- Low total cost of ownership

Archival storage requirements are quite different from those of backup operations. Media longevity and data authenticity feature much more prominently in archive environments. The storage media used within an archive should have a stable, long life to avoid frequent data migration over decades of storage. In order to comply with corporate and government regulations on data authenticity, it is crucial that information be protected from modification.

Unlike backups, the performance bottleneck for an archive is not read/write streaming, but in providing fast access to potentially millions of records requested by thousands of users. For data archives, fast random access is typically the most critical performance consideration.

Archival storage technologies

Although there are traditional roles for magnetic disk and optical technologies, both of these have been evolving in an effort to expand into new markets such as health care and finance. Today, there are magnetic disk products targeted at both the backup and archive market, while optical storage is gaining broader industry acceptance as a medium for long-term archival storage.

One thing is clear: No single technology can fully satisfy all data storage requirements at all stages of the data life cycle. The matrix below is designed to provide a general comparison of different storage technologies when applied specifically to an archive environment. Eight key archival storage attributes are contrasted with the characteristics of disk-based RAID and three optical formats: DVD, MO (Magneto Optical) and UDO (Ultra Density Optical).

<table>
<thead>
<tr>
<th>Archival Storage Attributes</th>
<th>RAID</th>
<th>DVD</th>
<th>MO</th>
<th>UDO</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Write-Once Media</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Media Longevity</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Removable Media</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
RAID and optical formats could both be used for long-term archival storage, but there are trade-offs. For example, RAID offers unmatched performance and high capacity, but it is fundamentally a rewritable media with a life of several years. It cannot be taken off-line for secure vaulting, and the long-term operating cost for RAID storage is high. By contrast, optical is removable, has high capacity and offers fast read/write streaming capabilities.

DVD is a true write-once media with a long life at an affordable cost. But unlike disk, it is a consumer-grade product with low capacity (9.4GB) and modest performance. Older-generation MO meets most of the archive requirements, but its current capacity of 9.1GB means that overall system costs are too high for many environments. Thirty GB UDO addresses all of the archive attributes and is designed for long-term archival storage. Many regulations specify the use of optical storage for data archives because it provides true write-once recording with fast random access plus media longevity.

**Archival storage: Total cost of ownership**

For many organizations, financial considerations are equally as important as technical attributes. Companies will not invest in new products unless they meet their technical requirements and meet their budgets. While active data and backup expenses are analyzed using short-term acquisition and media costs, it is more appropriate to consider the long-term total cost of ownership for an archive environment. Analysis over time provides a much more representative view of the total archive cost over many years.