the Availability Digest

www.availabilitydigest.com

# Archive Storage – Disk or Tape?

November 2010

In our companion article this month, Does Data Replication Eliminate the Need for Backup, we talked about the perils of depending upon data replication as a backup method. Data replication protects the system, but it does not protect the database. In the event of the loss of both the production and standby databases either due to data corruption or due to a simultaneous loss of both databases, only a database backup can provide the facility for database restoration by going back to a known good point in time.

Backups can be made either to magnetic tape or to disk (typically as magnetic-tape backup images). Disk is faster. Tape is cheaper. Which should be used? An interesting study by The Clipper Group has shed some light on this question. We review its conclusions below.

## Backing Up

There are strong reasons for backing up to disk – so called virtual tape or D2D backup – rather than to magnetic tape (D2T). Disk backups are less operator-intensive and are faster. Consequently, backups can be taken more frequently, leading to less data loss should the database have to be recovered. Furthermore, recovery is faster and more reliable from disk than it is from tape. The use of virtual tape backups can reduce data loss and recovery times to hours as compared to days, as is often the case with magnetic tape.

## Archiving

But the recovery of data in the event of system failure, accidental deletion, or malfeasance is only one purpose for backing up the database. Another important purpose is archiving. Long-term retention of data is often required for regulatory compliance, for litigation, or by company policy. Archived data may need to be accessed for other purposes, such as tracking down security violations.

**The Cost Ratio for a Terabyte Stored Long-Term on SATA Disk versus LTO-4 Tape is about**

**23:1**

**For energy cost, it is about**

**290:1**

**From Clipper Notes, published by The Clipper Group**

At some point, the retention of database backups for recovery purposes loses its value. For instance, a company may decide that the liklihood of having to go back more than three months to retrieve a backup quickly so that the system can be restored or so that erroneously deleted data can be retrieved is minimal. It then faces a choice on how to archive older data – continue to maintain the backups on disk or archive the backups to magnetic tape – the D2D2T strategy.

Certainly, by archiving old backups to tape, more time will be required to access or restore any data that has been archived to tape. On the other hand, it seems that tape archiving should be significantly less expensive than disk archiving. But what are those savings?

## Disk Versus Tape

David Reine and Mike Kahn of The Clipper Group (www.clipper.com) published an in-depth analysis of a typical business operation facing this problem in their report entitled Disk and Tape Square Off Again – Tape Remains King of the Hill with LTO-4.[1] They compared the monetary costs and the energy costs of archiving to disk (D2D) or to LTO-4 tape from a disk-based store that holds backups for ninety days (D2D2T). They concluded that long-term disk archiving cost twenty-three times as much as tape archiving and consumed 290 times the energy as that consumed by tape over a five-year period. If data deduplication is used for disk backup, the cost difference is reduced to a factor of five.

### *The Strawman Company*

The study assumed that a company held its backups on disk for one calendar quarter. The study then compared an all-disk archive solution (D2D) with a tape-library archive solution (D2D2T) for storing backups after the initial ninety-day period. The cost to store the backup tape images on disk for the intial ninety-day period was not considered as it was the same for both solutions.

Ninety-Day Online Backup

At the beginning of the five-year cycle, the database size was 50 terabytes. The company performed a full backup of the entire database once a week with daily incremental backups. 5% of the database changed each day, and those changes were captured in the incremental backups.

The full and incremental backups were kept on disk for thirteen weeks (one quarter). After that, the thirteenth week was archived to disk or tape as a quarterly backup. The incremental and full backups were deleted after they were ninety-days old.

The company's database grew at a rate of 50% per year. The disk storage required for storing ninety days of incremental and weekly backups was 845 terabytes the first year and grew to 4,278 terabytes by the fifth year. The cost of this storage was not included in the comparison as it was the same for both disk archiving and tape archiving.

Archiving Volume

Backups were archived every quarter. At the end of the first year, the archive had 200 terabytes of data stored in it (four quarters at 50 terabytes per quarter). Each year, the quarterly backups for that year were added to the archive. With the database growing at a rate of 50% per year, the archive grew to 2,384 terabytes of data by the end of the fifth year.

General Assumptions

Assumptions applicable to both disk and tape approaches include:

- Power cost was assumed to be $0.12 per KWH. This was higher than the national average of $0.10 per KWH but lower than the $0.18 per KWH found in some areas.

[1] David Reine, Mike Kahn, Disk and Tape Square Off Again – Tape Remains King of the Hill with LTO-4, *Clipper Notes*; February 13, 2008.
http://www.dell.com/downloads/global/corporate/iar/Clipper_Tape_v_Disk_2008.pdf.

- The cost of energy for cooling was assumed to be equal to the cost of energy to run the disk or tape equipment.
- Power cost was assumed to be constant over the five-year period.
- Space cost was $300 per square foot per year.
- 200 terabytes of storage were required the first year for archiving and grew to 2,384 terabytes the fifth year.

***The Disk-Array Solution***

The study's authors assumed the use of a SATA-2 platform from a Tier-2 storage array supplier. The fibre-connected storage area networks (SANs) used 750 gigabye disks in a RAID 5 configuration and were scalable to 245 tereabytes in a single configuration.

After the third year, the data center replaced its three-year old SANs with new technology that doubled the SAN storage capacity and that cut the SAN footprint in half. The result was a requirement for one array the first year, growing to five arrays the fifth year, in order to hold the quarterly archives.

The following SAN parameters were assumed:

- Disk capacity utilization rate was 85%.
- A SAN cost $2.6 million dollars.
- A SAN could store up to 245 gigabytes of data.
- After the third year, new SANs could store up to 490 gigabytes of data.
- SANs would be replaced with the new higher-capacity SANs when they were three-years old.
- SAN acquisition cost, including maintenance, was about $14 million over five years.
- A fully configured SAN required 39 square feet of floor space, including aisles.
- A fully configured SAN required 11 KW of power, or about 96,000 KWH per year.

***The Tape-Array Solution***

If tape archiving was used, the ninety days of online backups were still held on the disk subsystem. The tape archive held the quarterly backups.

An LTO-4 tape robotic library was assumed. It had an ultimate storage capacity of 10 petabyes so that tape cartidges did not have to be removed (at least, over the five-year study period).

The number of tape drives was chosen to ensure that the quarterly backup could be accomplished within a seven-day window. Two tape drives were configured with this in mind. In the early years, only one tape drive was needed, giving a spare tape drive. In later years, both tape drives were in use.

Additional tape parameters that were assumed included:

- An LTO compression ratio of 2:1 was achieved.
- A cartridge held 800 gigabytes of data (today's LTO cartridges can hold 1.5 terabytes).
- A tape drive could achieve an effective transfer rate of 734 gigabytes per hour.
- The tape-system acquisition cost was $242,000.
- Tape-system maintenance cost was $33,000 over the five-year period.
- Tape cartridges cost $12,600 per hundred.
- The tape system required 100 square feet of floor space.
- The tape system during the first year drew 185 watts, or about 1,600 KWH per year.

- The fully configured tape system for the fifth year required about 527 watts, or about 4,600 KWH per year.

***Disk/Tape Comparison***

The resulting five-year total cost of ownership (TCO) for disk archiving and for tape archiving is shown in the following table. These costs applied only to archiving. Backup for ninety days was provided by disk in both cases.

| | Five-Year TCO | |
|---|---|---|
| | **Disk** | **Tape** |
| Acquisition Cost | $13,000,000 | $241,600 |
| Maintenance Cost | 1,140,000 | 33,312 |
| Media Cost | | 207,703 |
| Energy Cost | 380,000 | 1,307 |
| Space Cost | 180,000 | 153,000 |
| Total Cost | $14,700,000 | $636,922 |

The result was that disk archiving cost 23 times as much as tape archiving and used 290 times as much energy. In contrast, the disk-backup subsystem required for both disk and tape archiving cost over $37,000,000, of which the energy cost was $840,000.

***Disk Deduplication***

The authors did not include the capability of disk-to-disk deduplication, in which each quarterly archive includes only the changes from the previous quarter's archive. They did, however, state that if a compression ratio of 20:1 could be achieved, disk archiving would cost about $3,000,000, including the cost of the data-deduplication software, virtual-tape engines, and virtual-tape software. This reduced the disk/tape cost ratio for archiving from 23:1 to 5: 1.

## Summary

There are two needs for backing up a database – fast recovery in the event of a partial or full database loss and long-term archiving. Fast recovery can only be achieved with disk-based backup. Disk backup can reduce restoration time for a full database from days to hours and also can reduce from days to hours the amount of data lost.

However, the cost of fast recovery afforded by disk may not be justified for long-term storage, for which fast recovery is usually not required. The costs of long-term archiving of data can be significantly reduced by using a disk-to-disk-to-tape strategy for backup and archiving. Recent backups of data are stored on disk for fast recovery, but older data is archived to tape for economy.

The Clipper Group study indicated significant savings when using tape instead of disk for archiving. Assuming the last database backup is archived each quarter, tape can be 23 times more economical than disk if quarterly backups are saved in full. Tape archiving can still be five times more economical if data deduplication is used for the disk archive. In addition, tape archiving is green. It uses significantly less energy than disk archiving.

The use of only disk or the use of only tape may not address all of an organization's goals. Disk and tape in a tiered D2D2T solution provide complementary values to achieve a company's goals of recovery, data protection, compliance, energy, and TCO objectives.