



Cost of Disk-Based vs. Tape-Based Archives

Bob Covey, Qualstar Corporation

Summary: Appropriate application of tape and disk storage technologies yields real cost savings and operating efficiencies. Getting it wrong, however, leads to poor performance, risk of data loss and skyrocketing costs. Tape libraries are ideally suited for archive storage applications, delivering substantially lower acquisition and operating costs than disk-based systems. Savings of 75% initially and more than 7x operating cost reductions over 5 years are readily achievable when using tape libraries to store archived data.

The cost of storage has steadily fallen as new developments in magnetic recording have been introduced for both disk drives and digital data tape. With the enticing low prices offered for high capacity Serial Advanced Technology Attachment (SATA) hard disks, it is tempting to just store archive data on rotating media. But, is this the right solution? This white paper looks at some of the obvious and not-so-obvious issues of archiving data on both disk-based and tape-based systems.

Let's first address some of the traditional comparisons between disk and tape, leaving out cost for the time being. These are:

- Capacity
- Performance
- Data Availability
- Security

In the past, data tape cartridges always had more capacity than a hard disk. With the introduction of 2 Terabyte (TB) SATA hard disk drives the capacity gap seems to have reversed. LTO 4 tapes holding 800 Gigabytes (GB) appear to be slipping behind, but LTO 5 tape cartridges with capacities of 1.5 TB are about to enter production. Note that this whitepaper ignores the use of data compression that can increase capacity by 2-300% when storing most, but not all kinds of data on LTO tape.

Traditionally, data could be transferred faster from hard disk than from tape, but the latest LTO tape drives can deliver sustained data rates well over 100 Megabytes (MB) per second. This is much faster than SATA RAID 5 or RAID 6 disk systems, which can manage roughly half that rate (with 7200 RPM drives and a 5-drive RAID set).

So two of our traditional values, capacity and speed, are turned "topsy turvy". What about the others, data availability and security?

For archive systems, the retrieval speed of disk is moot, because archived data is accessed infrequently. The performance edge that disk has delivering small amounts of data is also irrelevant when the overwhelming majority of the time data is being written to the archive, not retrieved from it. Unlike backup systems, fetching a file from the archive is rarely time-sensitive. Tape's speed advantage when writing archive data is substantial.

Finally, data security must be a key consideration. Losing archived content is simply not acceptable. Different RAID schemes provide varying levels of security. The compromise is, as always, security (and speed) versus cost. For example, RAID 10 improves performance and security but is very costly, while RAID 5 is more cost effective but with far lower assurance that the data is truly safe.

Modern tape technology with error correction and Read after Write verification is very stable. Apart from physical damage to a tape, the data can be considered to be non-volatile. Housing media in a tape library virtually eliminates the risk of handling damage. Making a second copy of a tape is simple, inexpensive and can be automated using the archive software. Indeed, two copies can be retained on site and a third copy kept in a separate location for disaster protection, simply for the cost of the media.

The same cannot be said for disk based systems. The SATA disk drives used in the large capacity RAID sets needed for archive systems do fail, and given the number of individual drives required, it is not unlikely that two drives may fail within a very short period as they age or are subjected to the same environmental conditions. This effectively rules out using a single RAID 5 system because the second failure in the same RAID set effectively makes the data irrecoverable. Data security in disk-based systems means replicated RAID sets.

Rebuilding a disk in a RAID set seriously impacts performance and risks data loss. The percentage of system time spent on rebuilding the RAID set can usually be adjusted. Set it to a low percentage to maintain performance and the rebuild time becomes very long. You risk losing another drive during the rebuild which means risking (RAID 6) or losing (RAID 5) your data. Rebuilding 2 TB disks can be a long, slow process often taking days to complete.

Data written to RAID systems must also be backed up. In the purely disk environment presented here, that means that disaster protection must be part of the archive systems design.

Moving beyond traditional comparisons of disk versus tape, let's look at some of the other tangible components that add up to the total cost of ownership of an archive:

- Initial purchase cost
- Electricity costs to run the archive, including air conditioning
- Cost of replicated media and disaster recovery
- Transportability costs
- Post warranty replacement costs
- Upgrade costs

Figure 1 shows a typical small archive containing 35 Terabytes (TB) of content. The LTO 4 tape-based archive uses archive software running on a dedicated Windows server containing 2 TB of disk as a temporary cache. Forty-four LTO 4 data tapes are located inside the tape library, and a second replicated set of data is stored on tapes in a fireproof vault. This system is compared to a pure disk-based solution. The disk solution uses replicated RAID sets to ensure data security of the archived content. It still doesn't protect from physical disaster at the site unless the second replicated RAID system is located remotely. That would require a dedicated broadband connection and additional systems that are not covered in this study. Most storage management professionals would argue that the disk-based archive is not secure, unless it is backed up off site.

Only brand name systems were researched for this white paper. In all cases, the servers and the disk subsystems are IBM, Dell or HP models. As you can see there is a significant initial

cost savings when purchasing the tape based system. The tape-based system costs \$34,770 compared to \$76,900 for the disk-based system. Both systems include the first 3 years of on-site support.

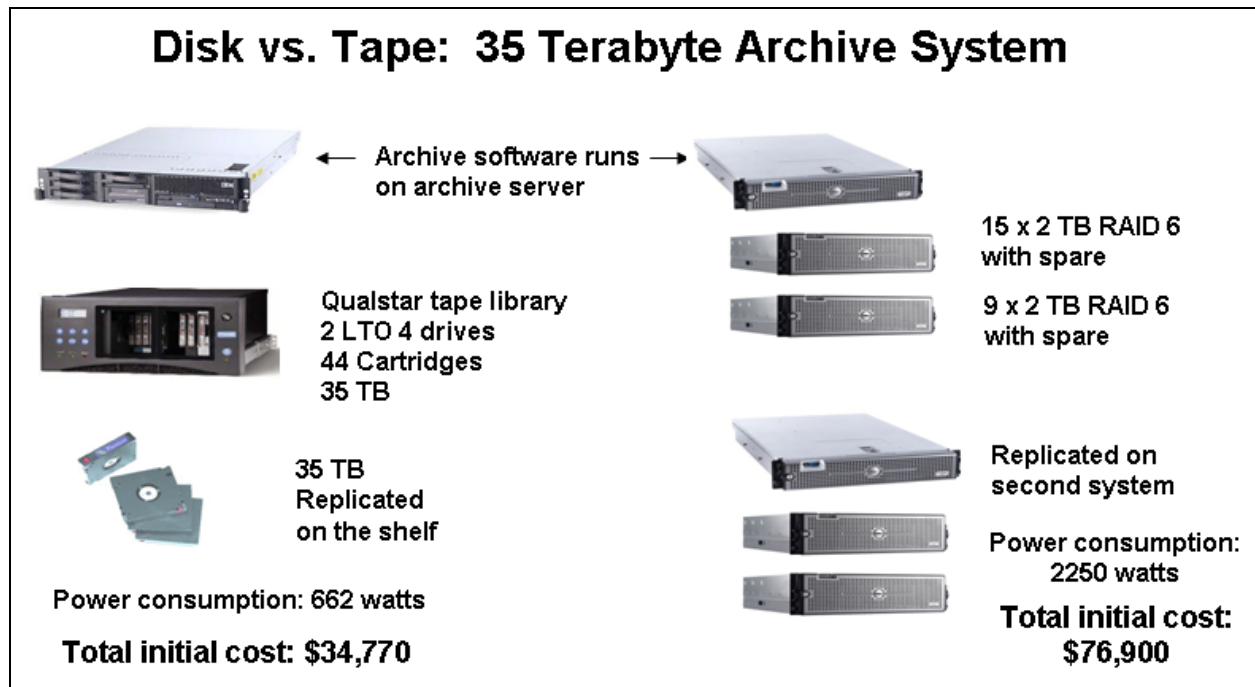


Figure 1

Significant operating costs are associated with the electricity needed to run the system and air conditioning to cool the location where the system resides. The tape-based archive consumes a maximum of 662 watts which is mostly from the archive server and the disk drives in the server cache. The tape library's consumption is minimal. In contrast, the disk-based system requires over 2 KW of power to operate. What does this mean in terms of cost? Using California commercial electricity rates, over five years the cost to power and cool the tape-based archive is \$6,085, while the disk-based system costs \$23,910, almost four times more.

Figure 2 compares 105 TB archive systems. Here the cost savings are even more dramatic: acquisition costs of \$78,650 for a tape-based archive versus \$193,700 for the disk-based system. Power consumption is an even greater factor with these larger archives. After adding in the cooling expense, the annual cost to operate the tape-based archive is \$1,500, compared to \$10,360 for the disk-based system. Over five years this amounts to \$7,500 and \$51,810, respectively, making the disk-based system seven times more expensive to operate.

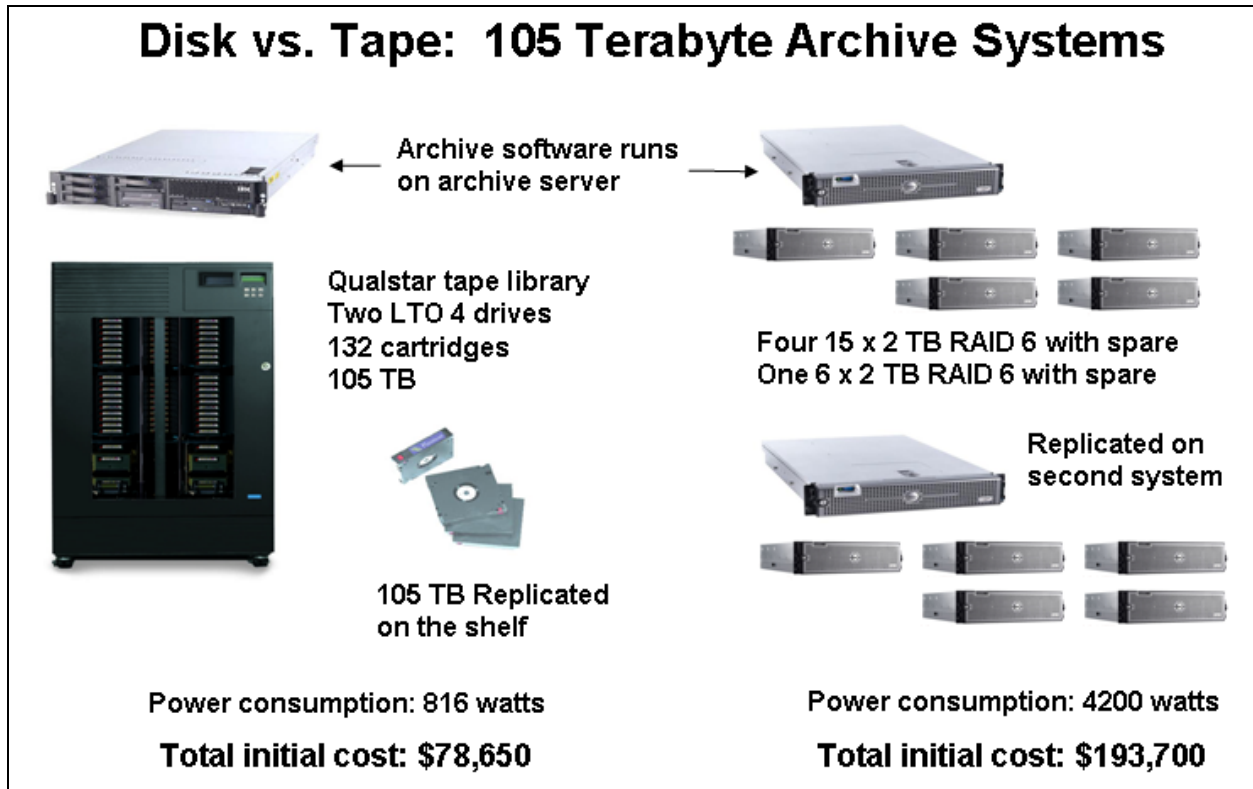


Figure 2

Disk-only systems are extremely expensive to provide multiple copies of the archive's contents. While two copies do reside on the mirrored RAID's to protect against local hardware failures, there is no provision for disaster protection of the data unless one of the RAID systems is remotely located. The cost of adding a third copy becomes really prohibitive.

Making multiple copies of the archived data with a tape-based system is inexpensive, just the cost of the media. A third set of tapes for offsite disaster protection does not add much to the overall cost of the system while providing significant failsafe protection for the data.

There are additional benefits to being able to create multiple copies on tape. It is easy and inexpensive to transfer data between facilities, particularly facilities located in countries that have limited broadband infrastructures. Indeed, several facilities already use LTO tape as a cost effective means to transfer material between remote locations. LTO 4 tape costs less than \$.06 per Gigabyte, making it especially attractive for distributing large datasets.

Now let's take a look at the cost of post-warranty support. The systems used in the comparisons have the first three years of on-site support built into the initial purchase price, but what happens when we look at years four and five of the 5-year life cycle costs?

The costs to provide post-warranty support for the 35 TB systems are \$4,375 per year for the tape-based system and \$7,500 for the disk-based system, for a total cost of \$8,750 and \$15,000 respectively. That's more than enough to pay for a third set of media.

The post-warranty support cost comparison for the 105 TB systems are even more striking. The tape-based system costs \$5,115 per year while the disk-based system cost is \$19,500. This amounts to \$10,230 for the tape-based archive and a whopping \$39,000 for the disk-based

system for years four and five. The reason that the disk-based systems are so expensive to maintain is that service and support costs are calculated on a per-chassis basis and there are twelve chassis in the disk-based archive.

Disk drives are powered and rotating 24 hours a day. Since there are 138 disks in the arrays for the 105 TB archive, two disk drives will likely fail each year during the warranty period. The probability of multiple disk drive failures causing complete loss of data on one of the RAID systems increases significantly during the post warranty period. Using mirrored systems with replicated data mitigates this risk, but at a significant cost.

What happens if a tape drive fails? Well, the data is secure unless a tape is damaged in the failed drive, a very remote occurrence, but replicated media solves that problem. Since the archive system uses multiple tape drives in the library, it can continue to operate while a replacement tape drive is changed out.

Upgrading the system is another factor to be considered when selecting an archive system. With data doubling every 18 months, the cost of keeping up with that growth is a critical factor. Expanding disk-based systems adds a lot more hardware because the mirror must be maintained, incurring proportionally higher operating costs. Tape-based systems can be upgraded much less expensively. LTO tape drives are read/write compatible back one generation and read compatible two generations back. Each new generation typically doubles capacity, so a 35 TB archive becomes a 70 TB system by simply replacing the tape drives and media. Upgrade the system a second time and the capacity quadruples while still reading the original media.

Let's look at an example of doubling the capacity of both 105 TB systems. The tape-based system will require two new tape drives and 264 pieces of media. Using \$98 as the cost for a new LTO 5 1.5 TB cartridge and \$12,000 per tape drive, the cost to expand the tape-based archive is: $(2 \times \$12,000) + (132 \times \$98) = \$36,936$. And, we still have our existing data stored on tape.

With disk-based systems, doubling the size of the archive means purchasing another complete set of RAID hardware at a cost of \$193,700. Remember, this also doubles the operating costs, too. Similarly, the 35 TB disk system will cost \$76,900 plus operating costs to upgrade, while spending just \$16,560 to double the comparable tape-based archive.

In summary, using tape for archive data storage is substantially less expensive to purchase, operate, support and upgrade compared to a similar sized disk-based system. Disks are clearly not the best medium for archiving data.

Tables 1 and 2 summarize the costs discussed in this white paper.

Description	Tape-based Archive	Disk-based Archive
Initial purchase cost	\$34,770	\$76,900
Operating power costs per year	\$811	\$2,391
Cooling power costs per year	\$406	\$2,391
5 year operating costs	\$7,095	\$23,910
Service and support costs 24 x 7 years 4 & 5	\$8,750	\$15,000
Total cost of archive	\$50,545	\$115,810
Cost to double the capacity of the archive	\$16,560	\$76,900

Table 1. Typical 35 TB archive system life cycle costs

Description	Tape-based Archive	Disk-based Archive
Initial purchase cost	\$78,650	\$193,700
Operating power costs per year	\$1000	\$5,181
Cooling power costs per year	\$500	\$5,181
5 year operating costs	\$7,500	\$51,810
Service and support costs 24 x 7 years 4 & 5	\$10,230	\$39,000
Total cost of archive	\$96,380	\$284,510
Cost to double the capacity of the archive	\$36,936	\$193,700

Table 2. Typical 105 TB archive system life cycle costs