WHITE PAPER
June 2005

Comprehensive comparison of tape and disk archives

ABSTRACT

Measuring total cost of ownership for a magnetic disk or magnetic tape storage solution requires looking comprehensively at all of the components. At first glance, disk storage media costs seem more attractive (prices have dropped as much as 30 percent annually). But when storage media, hardware, maintenance and personnel are factored in, the total cost of tape can be approximately *one-half* the cost of a similarly configured disk solution. Ultimately, a combination of both technologies may prove to be a more cost-effective and appropriate long-term storage solution.

1 Executive summary

Determining a long-term storage solution begins with a close examination of all hard and soft requirements for an organization's application and requires the discipline to look beyond falling prices and claims of increased performance. Companies need to consider storage performance, which can be defined as the efficiency of the storage processes for equipment and personnel, the ease with which data can be recovered or repurposed, and the reliability of the system.

This analysis compares the total cost of ownership (TCO) for disk and tape solutions for 1 petabyte (PB) of usable dual-copy data stores over a three-year period, based on prices and performance measurements collected during the first quarter of 2005. Data storage varies widely to meet the demands of different companies, but one tendency affects organizations consistently: As archive systems continue to grow over time, 500-terabyte (TB) or possibly 1-PB data stores will become more common.

The kinds of companies, industries and applications that require these very large data stores include banks (for check images), research laboratories, biomedical research institutions, large universities, oil and gas exploration companies, intelligence agencies, hospitals (patient data), high-performance computing sites, space exploration and earth observation agencies, retail point-of-sale transactions and genome research facilities.

For disk option pricing, this study uses data from www.pricewatch.com. Tape pricing assumes vendor list cost and reasonable discounts. The TCO elements include servers and software, software maintenance, administration, environmentals (e.g., floor space, power and cooling), hardware maintenance, and hardware and media.

During the first quarter of 2005, the sizes of SATA disks and LTO Ultrium Generation 3 tape drive media were equal at 400 gigabytes (GB) apiece. Yet considering conservative archive capacities to meet the same data storage requirements, traditional LTO subsystems are approximately *one-half* the cost of SATA disk subsystems sold with warranties and customer support. Although it is not included in this study, an SCSI/FC disk architecture can cost 5 to 10 times as much as a tape subsystem.

However, the best long-term storage strategy would combine the strengths of FC disk, SATA disk and FC tape technologies.

2 Business case: Why is data storage important?

All businesses need to keep track of information. Some companies make their money off the data they sell (e.g., a telephone directory publisher). But even when a company's revenue doesn't come from its stored data, records of business practices are equally indispensable. Some companies have regulations that require them to store data that may never be retrieved, but the cost of losing this data and experiencing an audit later can be crippling. Data storage is unavoidable. And the performance of the data storage system carries bottom-line consequences for productivity.

Data storage is unavoidable for businesses, even when the data does not generate revenue.

Data storage system performance carries bottom-line consequences for productivity, so companies carefully examine productivity metrics and cost before choosing a strategy.

As this TCO study reveals, the cost of continuing software and hardware maintenance and of managing enterprise-class archives is substantial, no matter what strategy a company employs. Considering efficiency (equipment and personnel), ease of access and reliability, companies should carefully examine productivity metrics and cost before settling on a storage strategy.

3 TCO comparison methodology and results

3.1 Six primary measures

The comparisons between disk and tape strategies in this study examine 1-PB usable dual-copy data stores over a three-year time period. Dual-copy is used to eliminate any single point of failure from the data store. Judging by historical trends, companies that currently manage 100 TB of data today may well face a 500-TB or perhaps a 1-PB data store within two or three years.

This study considers six components: servers and archive software, software maintenance, administration, environmentals (e.g., floor space, power, and cooling), hardware maintenance, and hardware and media.

- Two components — servers and archive software and software maintenance — are effectively identical because the challenges are similar for both systems. The servers and software costs for disk and tape over a three-year period are $500,000 apiece, and the software maintenance figure is also $500,000 for each system.
- Administration for a tape system is approximately $1.2 million for this time frame, compared with $2.4 million for a disk system. Although their base salaries may be equivalent, each tape archive administrator can handle about 500 TB, but each disk archive administrator can manage roughly 250 TB.[1]
- Environmental cost estimates run $40,000 for a tape system and $350,000 for disk, primarily because the duty cycles for tape drives are much lower than for disk drives, and tape media doesn't consume any power.[2]

Considering these first four components together, disk is $1,510,000 more ($3,750,000 versus $2,240,000).

But the final cost components — hardware maintenance and hardware/media costs — are significantly lower for a tape subsystem. Maintaining the hardware for a disk subsystem for three years would cost $1,034,222, assuming the one-year warranty that most vendors offer and assuming next-day service. But a tape system for the same time frame would be $380,380 (also assuming one year of warranty and next-day service). The hardware and media requirements for a disk system can be almost twice as much as for tape ($3,977,778 versus $1,683,888; see Figure 1).[3]

Considering maintenance and storage, a disk archive system can cost almost twice as much as a tape archive system for the same amount of data.

Taking all six components into consideration, a disk system can run $8,762,000, more than twice as much as the $4,304,268 required for tape.[4]

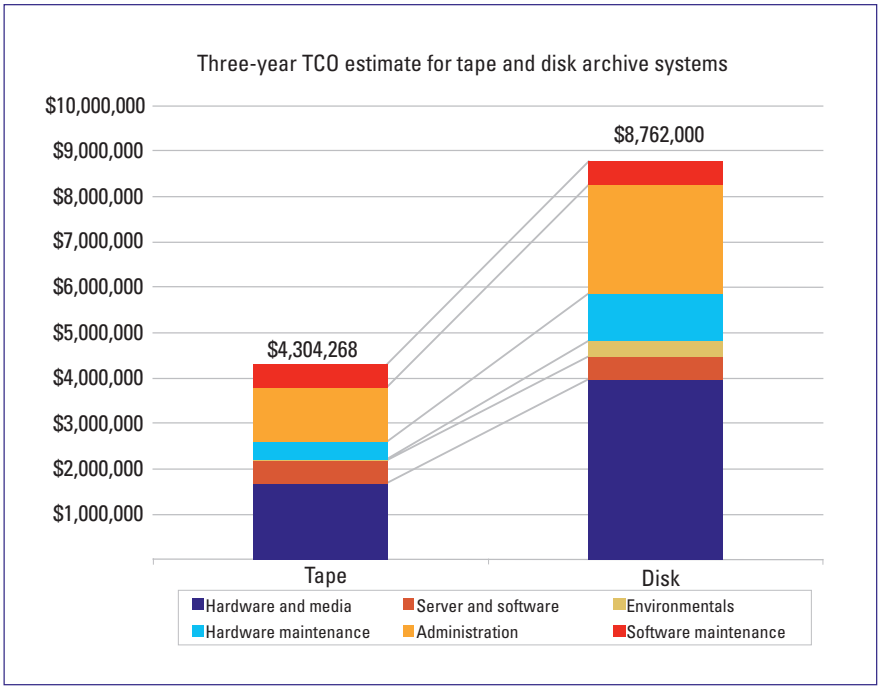
[1] Full set of assumptions behind these are documented in the Appendix.

[2] Environmental cost was estimated using StorageTek's Storage Value Tool.

[3] Warranty cost estimated using StorageTek's Storage Value Tool.

[4] Models and assumptions used to generate these estimates are detailed in the Appendix.

Figure 1: Comparative cost of tape and disk archive



3.2 Apples and oranges

Other components in disk and tape architectures do not correlate directly, so this study assumes approximate equivalents:

- One tape cartridge equals one disk drive.
- The total tape library, robot and slot assembly equals the total disk cabinet and enclosure assembly.
- One dual-disk controller equals two tape drives.

For more details about the comparison components used in this study, see the appendix.

3.3 Serial ATA disk cost parameters (excluding software, environmental and administrative costs)

At the beginning of 2005, the raw cost of a single SATA disk drive purchased from www.pricewatch.com ranged between $0.75/GB and $0.90/GB. Factoring in some vendor overhead and margins, this translates to approximately $360 for each 400-GB SATA disk drive.

Before factoring software, environmental and administrative costs, a disk system can be $3.98/GB. A tape system for the same size data store can be as low as $1.68/GB.

A 1-PB raw dual-copy archive requires purchasing 5,000 individual 400-GB disk drives to accommodate 2 PB of raw storage. At $360 per drive, this comes to $1,800,000.

Considering 15 disk drives per enclosure and 10 enclosures per cabinet, this would require more than 34 disk drive cabinets. At $2,000 each, this comes to $68,000 for the cabinets. With a one-to-one correspondence between cabinets and dual-disk controllers, the cost for 34 dual-disk controllers is $340,000 ($10,000 apiece). Enclosures usually cost $2,500 each, which amounts to a cost of $850,000 to enclose each disk drive.

This study assumes a RAID 5 9+1 design, which defines the equipment needs as well as a RAID overhead of 10 percent. In addition, disk utilization for a SATA system has been factored at 85 percent because the archive data is not dynamic. Taken together to create a parallel factor for cost, the combined RAID overhead and utilization percentage for the actual data has been multiplied by 1.31 to equate with 1 PB usable storage, which raises the price by the same factor to $3,977,778, or $3.98/GB.

3.4 Tape cost parameters (excluding software, environmental and administrative costs)

LTO Gen 3 400-GB native media is roughly $200 per cartridge. Considering a 1-PB raw dual-copy archive system with no data compression, this comes to 5,000 total cartridges ($1 million).

Assuming one tape drive per 75 cartridges (or two tape drives per dual-disk controller) with an average cost of $17,000 per LTO Gen 3 drive and a total of 67 drives, the drive cost reaches $1,139,000.

One 3/4-configured StreamLine™ SL8500 modular library system is enough to accommodate the required tape drives and media for this sampling (for a cost of $324,000).

Factoring all the storage system costs together with an estimated street price of 33 percent off of list price, the subtotal is $1,650,210.[5]

With a utilization factor of 98 percent for a tape storage system, this study includes a factor of 1.02 to equate the disk space used with the true equipment cost. This comes to $1,683,888, or $1.68/GB for an equivalent 1-PB usable dual-copy archive (less than half the cost for a disk solution).

3.5 Other cost parameters not included

This study did not factor compression for tape archives, nor did it include installation costs and the price of professional services over time.

Disk can be the perfect target for either the primary or secondary copy of the archive.

Tape, on the other hand, is less expensive, and the capabilities have increased as the prices have dropped.

But to accommodate unpredictable data needs, a combination of strategies might be the best solution.

4 Recommendations

The factors that drive data storage decisions stem from regulatory requirements, business practices, historical patterns in an organization and data use forecasts. Even within a single company, some data may require backups that will probably never be retrieved, whereas other data needs to be constantly accessible. And, with the passage of time, data that may have been marked for purging may become essential for another, unanticipated purpose. For example, an oil company's 1970s-era analysis of alternative fuel sources may have been shelved, only to require careful attention in the early 21st century.

The terms *online, offline* and *nearline* cover the range of data storage needs. Information that requires constant accessibility with high duty cycles benefits from an online strategy, which is best executed with a disk storage system. Long-term data that allows longer retrieval times can benefit from an offline strategy that might entail manual tape backups and remote storage. And nearline storage combines automated processes that fall somewhere between manual and fully automatic strategies.

[5] *The modeled disk subsystem is a 'white box' configuration, modeled at street price. A street price estimate is is generated for the tape configuration to make a realistic comparison.*

Disk, in addition to being easily accessible, can be the perfect target for either the primary copy or the secondary copy of the archive. And even though the cost is significantly higher than for tape, a disk configuration with Fibre Channel connections can provide increased performance.

Fibre Channel disk can support heavy input/output-per-second and high megabyte-per-second workloads. Consequently, FC disk systems are best suited for "primary storage." Conversely, SATA disk systems are best suited for applications that require high megabyte-per-second but lower input/output-per-second workloads. This is normally called "secondary storage."

Tape, on the other hand, is less expensive, and the capabilities have increased as the prices have dropped. Also, this study did not factor data compression for a tape archive. Although compression ratios vary according to file type, compression lowers the number of drives and cartridges required, thereby reducing the TCO of a tape system even more. If additional cartridges are required over time, administrators save money because they can add low-cost cartridges to an archive workload while maintaining the same number of drives (as long as they fit within an existing library structure).

To accommodate unpredictable data needs, a combination of strategies can help. As this cost analysis shows, tape storage provides the lowest price per gigabyte today. But accessing tape data frequently is not the best use of library and tape drive equipment. Therefore, combining a disk cache and tape can provide a bridge between long-term and short-term needs. If data in the archive needs to become more fluid and more consistently accessible, a combination storage strategy allows administrators to move that data from primary FC disk to optional secondary SATA disk and then to SCSI/FC tape (while continuing to protect changing data with back-end tape).

Ultimately, an organization's analysis of its data drives the offline, nearline and online designations. With effective data classification, companies can determine the primary, secondary and perhaps tertiary paths for each data type, and the resulting combination can accommodate the tasks of moving data between the extremes.

5 Glossary

ATA Advanced Technology Attachment or AT Attachment. A protocol that defines I/O specifications. (See ***SATA***.)

Duty cycle a measure of the frequency and intensity of storage use.

EIDE Enhanced Integrated Drive Electronics, developed in 1994 to increase the maximum data throughput to 16.7 megabits per second, increase the size limit from 528 megabytes to 8.4 gigabytes, and support two controllers to increase the maximum number of disks from two to four. (See ***IDE***.)

Enclosure a chassis that holds disk drives with power, cables and interfaces to the controllers.

FC Fibre Channel, a set of standards for rapidly transferring data between devices.

IDE Integrated Drive Electronics, an interface for connecting internal storage devices.

I/O input/output.

LTO Linear Tape Open, an open tape storage technology format.

LUN logical unit number, the designation for one of eight specific SCSI connections (typically a disk drive).

RAID redundant array of independent disks. Designs can be simple or complex, but the idea is (through "redundancy") to ensure data integrity if one of the many drives goes out. To the user, it may appear as one drive (LUN) or as broken up into multiple LUNs.

SAN storage area network, a network of storage devices connected to a network and each other.

SATA Serial ATA. This enhancement of the AT Attachment interface supports serial connections, thereby increasing the speed over ATA. (See **ATA**.)

SCSI Small Computer System Interconnect, an ANSI standard for using I/O adapters to connect storage devices to hosts.

6 Appendix

These tables show each cost component for this comprehensive disk and tape archive comparison, assuming a dual-copy factor for both systems.

Disk Archive Worksheet

Storage system

	Primary	Dual-copy redundancy	Cost
Targeted raw capacity (GB)	1,000,000		
Drive raw capacity (GB)	400		$ 360
Subtotal disk count	2,500		
Disks per enclosure	15		
Dual controller			$ 10,000
Controllers per cabinet	1		
Total enclosures	166.67		
Total enclosures rounded	167	334	$ 835,000
Number of drives	2,500	5,000	$ 1,800,000
Enclosures per cabinet	10	20	
Disks per cabinet	150	300	
Total number of cabinets	17	34	$ 68,000
Total number of controllers	17	34	$ 340,000
Storage equipment cost subtotal			*$ 3,043,000*
RAID 5 9+1 overhead	10.00%		
Available capacity (GB)	900,000		
Usage efficiency	85%		
Usable capacity (GB)	765,000		
Adjustment to 1 PB usable capacity	*1.31*		*$ 3,977,778*
Total cost for usable capacity ($/GB)	*$ 3.98*		

Software, environmental and administration

Archive server and software		$ 500,000
Software maintenance		$ 500,000
Environmentals (e.g., power and space)		$ 350,000
Maintenance cost (two years, next-day service)		$1,034,222
Administration		$2,400,000
Overhead subtotal		*$4,784,222*
	Total	*$8,762,000*

Tape Archive Worksheet

Storage system

	Primary	**Dual-copy redundancy**	**Cost**
Targeted raw capacity (GB)	1,000,000		
Cartridge raw capacity (GB)	400		$ 200
LTO Gen 3 drives			$ 17,000
Cartridges per drive ratio	75		
Compressed capacity (GB)	400		
Required cartridges (base)	2,500		
Required cartridges (total)		5,000	$1,000,000
Required drives	67		$1,139,000
SL8500 library system	1		$ 324,000
Subtotal			*$2,463,000*
Discount to street price	33%		
		Storage equipment cost subtotal	*$1,650,210*

Available capacity (GB)	1,000,000	
Usage efficiency	98%	
Usable capacity (GB)	980,000	
Adjustment to 1 PB usable capacity	*1.02*	*$1,683,888*
Total cost for usable capacity ($/GB)	*$1.68*	

Software, environmental and administration

Archive server and software		$ 500,000
Software maintenance		$ 500,000
Environmentals (e.g., power and space)		$ 40,000
Maintenance cost (two years, next-day service)		$ 380,380
Administration		$1,200,000
Overhead subtotal		*$2,620,380*
	Total	*$4,304,268*

ABOUT STORAGETEK®

Storage Technology Corporation (NYSE: STK) is a $2 billion global company that enables businesses, through its information lifecycle management strategy, to align the cost of storage with the value of information. The company's innovative storage solutions manage the complexity and growth of information, lower costs, improve efficiency and protect investments. For more information, visit www.storagetek.com, or call 1.800.275.4785 or 01.303.673.2800.

WORLD HEADQUARTERS

Storage Technology Corporation
One StorageTek Drive
Louisville, Colorado 80028 USA
1.800.877.9220 or 01.303.673.5151

TT 0010 A 06/05