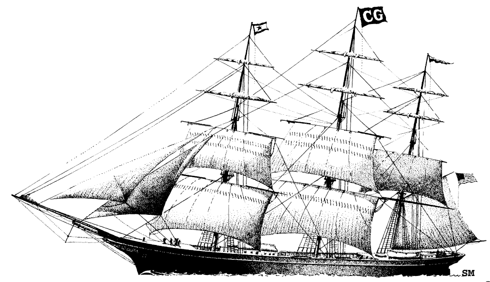
Clipper Notes™
VENDOR-AGNOSTIC EXPLANATIONS AND ADVICE FOR THE INFORMATION TECHNOLOGY BUYER



Navigating Information Technology Horizons℠

Published Since 2007 Report #TCG2008056 Updated October 21, 2008

Disk and Tape Square Off Again — Tape Remains King of the Hill with LTO-4

Analysts: David Reine and Mike Kahn

Management Summary

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

Read this bulletin to understand what this really means.

Progress happens in many ways. At times, progress can be *evolutionary*, as technology advances one-step at a time, as we have seen in the development of the microprocessor, from Intel's 16-bit *8086* in 1978 to the 64-bit multi-core CPUs of 2007. It can also be *revolutionary*, as with the invention of the cell phone that literally turned the business world upside-down with mobile communications. Sometimes, we can have a revolutionary, new technology competing with the evolution of an older technology for the hearts, minds, and wallets of the business community. The newer technology usually has a lot of buzz about it, creating an excitement to be the first to implement, rather than rolling out the latest iteration of the legacy solution. Both sides make claims as to the functional advantages of their solution. Both sides also readily admit that there are additional costs associated with the new technology; replacing a legacy system is expensive. The question that persists, however, is "how much?" The CIO may be anxious to push forward with the newer technology, but the CFO may want to apply traditional measuring sticks to the acquisition. The enterprise needs to determine the total cost of ownership (TCO) of both solutions and the impact on the enterprise bottom line before making a choice.

This very situation exists in today's mid-sized data center where new technologies are competing with tape, *again*, for the favors of an I.T. staff struggling to support a rapidly expanding storage base while protecting the assets of the enterprise. The purpose of this bulletin is to review these new technologies against the value of retaining, or upgrading, your current tape environment. To do this, we will try to establish a framework of comparison – using a backup/recovery and archive scenario – between the latest disk technologies and the latest iteration of tape storage, LTO-4 technology.

We chose backup and archive for two reasons. First, every business does it, or at least should! Second, it is an excellent example of a very large and growing storage requirement, from which many conclusions can be drawn. For example, once you have one or more disk arrays deployed in the solution, adding another will not change the cost/TB by very much, except for the underlying reductions in cost per terabyte that come with time. For a tape library example, if you assume that the enterprise initially acquires a configuration with maximum tape automation and minimal number of tape drives (i.e., more than enough room for later growth), then incremental costs for additional drives and media are largely linear. While this might front load, somewhat, the total costs of ownership for tape, it does show a very favorable payback over disk for long-term storage. Our finding is that for long-term storage over our five-year study period, the cost of disk is about 23 times that of tape, while the cost of energy for disk is about 290 times that of tape. Read on, for the details.

IN THIS ISSUE

The backup scenario looms large in every enterprise IT budget, so it is a good example of *something large and stored for a long time*, but it is only an example, as backup is not the focus of this study, which is comparing the costs of disks and tape for long-term storage. With the constant growth of data, backing up and, especially, recovering a modern data center has become a constant struggle between performance and cost, as the backup window continues to shrink while the volume of data continues to expand. The need to backup and recover storage in a 7x24x365 Internet-driven environment has forced the data center to look for innovative techniques, such as disk-to-disk (D2D) backup solutions in order to keep the enterprise performance objectives viable. **D2D, however, is not a replacement for tape; it should be a complement to tape.** The costs associated with a pure D2D scenario are simply too great for any mid-sized or larger business to consider using this technology to establish a comprehensive policy to save their entire data store for any extended length of time. Because of the urgency associated with short-term backup and recovery, D2D can make sense to retain some number of days worth of backup data on disk. For this study, we have chosen to store the most recent 90 days (13 weeks) worth of backup data on disk to help meet the performance objectives defined in service level agreements. We will then save the *13th weekly full backup* as the *archived quarterly backup* and save it to a disk system and to a tape system – to compare the TCO costs of each long-term storage solution. Noteworthy is that the cost of a 13-week disk cache is not included in the main comparison, as the same baseline costs (for the disk cache) will be common to both the disk and tape solutions.[1]

Failure to recover quickly from a loss of data can cost the enterprise millions of dollars in lost revenue. However, after 90 days, where is the likely urgency? Does the convenience outweigh the massive budget hit? Is there an economic advantage to deploying a tape solution, an LTO-4 solution? Best business practices dictate that the enterprise retain weekly, monthly, quarterly, and end-of-year data for many months, or even years, in order to recover from a problem identified later or to retrieve critical information required by litigation or compliance. This may happen as part of normal audit procedures or due to seasonality or irregularity of a process or customer interaction.

Thus, the enterprise, however, can justify a blended D2D2T environment (disk-to-disk-to-tape), on acquisition costs alone. When you include factors involved in determining the TCO, costs associated with maintenance, training, floor space, and energy consumption, the result is a no-brainer. **Our study shows that tape continues to be the most economical solution for long-term storage requirements for the mid-sized data center**. You even gain the additional ecological advantage of re-using tape media to help keep the environment green.

To be fair, we chose to take a conservative position whenever there were options that could slant this analysis one way or the other. For example, we used the same capacity utilization rate for tape and disk. Previous studies have identified a significantly higher utilization factor for tape; in fact, by as much as 10% to 15%. If your data center has measured a lower rate for either disk or tape, then you can factor that into your analysis. We, arbitrarily, chose a seven-day window for the transition of quarterly backup data from disk to tape.

With regard to energy, we could have given tape an edge, since it uses little power, by using a cost based on the energy rates in major urban areas such as Boston or New York, around $.18 per KWH. We did not. We chose a figure a third lower, $0.12 per KWH, but still above the national average of $.10 per KWH[2], and we kept the figure constant for the five-year study period. Whenever we chose any variable, we have explained our rationale. We even considered data deduplication in terms of a VTL comparison, reducing the volume of data stored on disk by a factor of 20:1; more on this later.

We began this study to see if **the decreasing costs of disk subsystems and the increasing capacity of disk drives, especially second-tier SATA storage, might have made the TCO for disk more attractive versus tape in the long-term storage of data. *It did not.* We thought that the cost of energy would be a noticeable factor in favor of tape. *And it is.***

The advantages to the enterprise of retaining tape libraries as part of a tiered storage strategy to enable a D2D2T environment are quite clear. To learn more about the details of this analysis and see how you can save significantly on your long-term storage costs by leveraging *LTO-4* tape, please read on.

Assumptions for a Mid-Sized Enterprise's Long-Term Backup Scenario

For the purposes of this study, we have chosen to look at the data center of a mid-sized business.

Backup/Archive Assumptions

[1] You still have to pay for the cost of the disk cache (itemized separately in the middle table of Exhibit 3) but we have excluded these costs because they are the same in the disk and tape solutions and our focus is on the cost of long-term storage.

[2] From the Energy Information Administration rankings for June 2007.

We will compare an *all-disk solution* (D2D) with a *tape library solution* (D2D2T). We assumed an initial capacity of 50TBs of primary storage, with planned growth of 50% annually. We looked at the disk and tape library requirements over a five-year period to establish a finite period for the study. Our mid-sized data center will be doing daily incremental backups and weekly full backups, with these retained on disk for 13 weeks, in order to meet backup and recovery SLAs. After the initial 13 weeks, the last of the 13 full backups (i.e., the "quarterly") will be retained on the disk system or copied to a tape system. **The cost comparison, over five years focuses on storing these "archived quarterly backups" on the two alternatives, disk and tape.**

Disk Assumptions

The disk arrays will store all archived quarterlies for the defined period in a RAID architecture, with one hot spare drive per drawer. This will enable the D2D solution to use standard maintenance warranty terms.

Tape Assumptions

Copying the archived quarterlies to tape will assume the standard LTO compression ratio of 2:1; you might do better or worse, but that will depend on your data content. The data center will recycle all tapes at the conclusion of their retention period. The tape library must be able to retain all quarterly data *nearline*, i.e., within the robotic library, in order to expedite recovery. We ignored the fact and the potential benefits that the data center can write LTO-4 media as a Write Once Read Many (WORM) cartridge and encrypt it at the device level for security.

Our mid-sized data center will utilize a strategy of *n+1* tape drives[3] to ensure device availability and standard 1-year warranty and 24x7 maintenance terms for response. The transition to tape will take place in background, using the robotic library for continuous operation and enough drives to complete the tasks within a seven-day period.

Energy Assumptions

The energy available to run any data center may be (probably is) limited. Many sites have already reached the limit of electricity available to run their data center. The cost of energy is also a factor, as the rate per KWH is not going down. Energy access and energy rates do vary, however, across the country. In an attempt to determine the cost of the various storage options, we will use an average urban rate[4]. If you are in Boston or New York, prepare to pay more, *if you can find it.*[5] If you are in Boise, you will be pleasantly surprised. **We have also opted to hold the cost of energy as a constant over the five-year term in order to simplify the comparison. This decision favors disk, as higher energy rates will accelerate the energy costs for disk to a greater extent than tape**. You should factor in local rates and annual increases, as appropriate for your region and situation. We have also adopted the industry norm and assumed a dollar to cool the environment for every dollar spent to run the data center's infrastructure.

Accounting Assumptions

We chose to have a five-year period over which to study the total costs of ownership of the long-term storage of data (i.e., the archived quarterlies), which included cost of equipment, media, electricity, and data center floor space. This is complicated by the fact that disk systems tend to be kept for three years and tape systems tend to be kept for more than five years. Given these realities in our model (or imperfections, if you prefer), we tried to keep all of the other accounting as simple as possible. The cost for backup software, backup servers, network operation, and human interaction is not part of this analysis.

Fundamental Building Blocks

For disk, we have selected a SATA-2 platform from a leading Tier-2 storage array supplier, using 750GB SATA drives with scalability to over 350TBs of raw data in a single configuration. We chose a Fibre Channel interface to a Storage Area Network, with a RAID 5 format to ensure data access and a hot spare.

Our tape solution will consist of a scalable library with support for over 6,000 LTO cartridges *nearline*, i.e., within the library, with a compressed capacity of up to 10PB and support for more tape drives than this application will ever need. We have elected to use the latest tape architecture, *LTO-4*[6], with a native data transfer rate of 120MB/sec, native capacity of 800GB per cartridge, and the expectation of an average 2:1 compression ratio. It is also worthwhile to mention that LTO-4 comes with support for WORM cartridges and data encryption at the drive level, enabling the data center to both secure their data from unauthorized access and to ensure its credibility against any future litigation, but WORM did not affect our economic model.

[3] n = the number of LTO-4 tape drives required to complete a backup in the defined window.

[4] Also from the Energy Information Administration rankings for June 2007.

[5] Many office buildings can't get additional capacity.

[6] See the issue of Clipper Notes dated July 12, 2007, entitled *LTO-4 Pounces into the Data Center with New Life, Greater Capacity, and Higher Performance*, and available at http://www.clipper.com/research/TCG2007073.pdf.

Exhibit 1 – Annual Backup Storage Requirements for D2D

Year	Storage Base (TBs)	Daily Back Ups (TBs)	No. of Days in 13 Weeks	Daily Sub-Total (TBs)	Weekly Sub-Total (TBs)	13-Week Total (TBs)	Quarterly Total (TBs)	Historical Quarters Total (TBs)	Rolling Total (TBs)	Total No. of Disk Arrays Required
1	50	2.5	78	195	650	845	150		995	4
2	75	3.8	78	293	975	1268	225	200	1693	7
3	113	5.6	78	439	1463	1901	338	500	2739	12
4	169	8.4	78	658	2194	2852	506	950	4308	14
5	253	12.7	78	987	3291	4278	759	1625	6662	18

Source: Clipper Analysis

The Disk Array Solution

Beginning with a base storage requirement of 50TBs by the end of year 1, we can see in Exhibit 1 (above) that daily incremental backups, estimated at 5% of a full backup, would require 2.5TB/day of disk capacity during the first year. With a requirement to retain 78 days[7] of daily differential backups to support the SLAs for recovery, we start with 195 TBs in the array. Weekly (full) backups will add another 650TBs for the most recent 13 weeks, for a total of 845TBs. With the addition of three quarterly reports by the end of year 1, the total capacity required will rise to 995TBs at the end of year 1. As you will quickly recognize, the real enemy here is the continuing growth of data!

During the second year, assuming 50% annual growth of storage, the enterprise will replace the 2.5TB incremental backup with a 3.8TB version, and the 50TB weekly backup with a 75TB backup, freeing up the 13-week space from year 1, retaining only the four quarterly backups from year 1 (as archived quarterlies). The data center will repeat that process in the third year, resulting in a cumulative requirement of 2.7 PBs of storage. The enterprise will require more than 2.7PBs of *disk capacity* by the end of year 3.

In order to ensure reliability and recoverability, we have assumed a RAID5 format with two RAID5 arrays assigned to each drawer. This will consume two drives as parity drives, with a third drive in reserve as a hot spare for both RAID groups, leaving a net of 12 drives per shelf. We have assumed a very generous utilization rate of 85% for disk consumption. This will result in the capability to store 7.65TBs of data per drawer, or 245TBs in a single 32-drawer disk storage system[8]. In order to support the enterprise backup requirement for the first year, the data center will have to acquire four fully-configured disk systems. They will also have to acquire an additional three systems for the second year and five more systems for the third year. This means that the data center will have to deploy 12 disk systems to satisfy the requirement for the first three years! The floor space alone for a deployment of this size might require the construction of a new data center.

After the third year, the data center can plan on replacing the four arrays acquired in the first year, rather than absorbing new maintenance charges after the expiration of the initial 3-year maintenance warranty. We have assumed that the new arrays will have twice the capacity of the older models, at the same price. Combining this replacement with the growth of enterprise storage will result in the acquisition of six new arrays for the fourth year and seven more for the fifth year, replacing the three systems acquired in the second year, for a total of 18 fully-configured arrays at the end of year 5.

Floor Space Requirements

A single 40U rack occupies 6.5 sq. ft. of floor space. A fully-configured array consists of three racks or 19.5 sq. ft, which we have assumed must be doubled to allow for aisle space. The first-year requirement of 4 arrays would occupy 156 sq.ft., with a total of 702 sq.ft. of data center floor space needed by the end of the fifth year. At an average rate of $300/sq.ft./year, this would add $590K to the 5-year TCO of the data center, expensive but a drop in the bucket when compared to the disk acquisition expenses.

Hardware Costs

Vendors usually do not make price lists publicly available for storage systems in this category. However, we have learned that the acquisition cost for a single, fully-configured disk array (32 drawers) with 750GB SATA drives is in excess of $2.6M, including 3-year warranty, at list price. The first year cost

---

7 During the most recent 13-week period (91 days), the data center will do 78 daily incremental backups and 13 weekly (full) backups, the last of which will be retained as an archived quarterly.

8 In actuality, this figure will be slightly less as a minimum of 5 drives will be required in the first drawer for system use. We have ignored this factor in an attempt to keep the analysis as simple as possible.

Exhibit 2 – Annual Energy Requirements for Each Fully-Configured Storage Array

In Watts/Hour, unless otherwise noted

	Processor Chassis	Standby Power Supply	Per SATA Drawer	No. of SATA Drawers	Total Array Power (KW/H)	Power per Day (KWH)	Annual Power (KWH)	Annual Cost @ $0.12/ KWH
Typical Power	430	34	325	32	11	264	96,360	$11,563
Maximum Power	800	300	425	32	15	360	131,400	$15,768

Note: The energy cost does not include costs for cooling.

Source: Clipper Analysis of vendor specifications and data

of four-fully configured systems, needed to hold the daily, weekly, and quarterly backups described above is **$10.57M**, again at list price[9], $10.71M including energy and space. By the end of the third year, this accumulated total will rise to almost **$30M**, with an additional **$21.78M** added for years 4 and 5. **This is a total of $51.73M for acquisition costs, energy, and space**. (See Exhibit 3 on the next page.)

Energy Requirements

Assuming that the enterprise has enough energy available to install this much spinning media, how much energy is required to run this configuration? For this computation, component figures are available. In fact, the data center can calculate typical power usage for a normal system operation, maximum usage when all drives are active (reading/ writing) simultaneously, and a recharge usage, when the battery backup subsystem is being recharged after usage. Obviously, the data center would have to be configured for the recharge power consumption value; we, however, will use the "typical" value in an attempt to treat the energy consumption of the D2D environment fairly.

A single 32-drawer configuration, running in a "typical" environment will consume about 11,000 Watts per hour. (See Exhibit 2, above.) Using a conservative rate of $0.12/KWH, a single system operating in "typical" mode will consume $11,563 worth of energy in a single year. The data center, however, requires an additional, equal amount of energy, in order to cool the environment from the heat generated. With four systems required for the first year, the enterprise should plan to spend at the rate of $92,500/year for energy alone. By the end of the third year, this figure will be approaching $260K, holding the KWH rate constant, and over $420K for year 5. If the installation is in one of the larger urban areas, that figure would be at least 50% higher, over $600,000 for electricity the fifth year.

We have also shown, in Exhibit 2, an entry for "maximum" power. If the disk array is extremely active with all drives processing concurrently, the cost of power and cooling will increase proportionally, with a single system consuming 15,000 Watts per hour, at an annual cost of $31,536 per system. Costs to the enterprise for the first year alone would be over **$125K**, over **$375K** for the third year, over **$.55M** for the fifth year, at $0.12/KWH.

Disk Summary

While energy and space costs are significant, they only represent 4% of the total cost. They are dwarfed by the cost of the hardware. Please note that we have not included the cost of administration or backup software, which all contribute further to the TCO equation.

The Tape Solution

Despite the fact that the data center *can* remove cartridges from the tape library when the immediate need for recovery is low (i.e., after six or nine months), we have configured an LTO-4 library to support all recoveries with nearline response for all five years. This significantly adds to the costs of tape in our economic model. In order to simplify acquisition and installation costs, we will initially deploy an open systems library for a mid-sized enterprise with the capability to expand to over 6,000 cartridge slots, but with only the number of slots required for the study. Again, in order to satisfy the near-term backup SLAs, all incremental and weekly backups will be written to a secondary disk array in a D2D2T architecture. That is, the daily incremental backups are not written to tape (in our base model), as the SLAs for recovery will be satisfied within the framework of a D2D environment. Therefore, just as in the disk solution, the most recent 13 weekly backups will be retained on disk for immediate recovery purposes, with only archived quarterlies stored to tape. **This will require an 845TB secondary disk deployment to handle the initial backup period. As previously noted, this disk deployment is common to both the disk and tape solutions and is not being counted in the TCO for either scenario.**

9 We did not attempt to negotiate a discount rate as your purchasing department would. We have done all of our calculations at list price.

Furthermore, a user could choose to keep more or less than 13 weeks to meet individual SLAs.[10]

With a native capacity of 800GB and throughput of 120 MB/sec, an LTO-4 drive has a compressed capability to write at 240 MB/sec, or 864 GB/ hour. We have adjusted the theoretical maximum performance downward by 15% to account, conservatively, for the possible effect of network or application bottleneck scenarios that could occur. *Your performance may vary*. Considering these factors, we end up with a practical throughput of 734GB per hour for a single LTO-4 drive. As stated above, the goal is to complete the archive process, in less than seven days of continuous, unattended operation each quarter[11]. Each LTO-4 drive can back up almost 6TBs in a single shift, about 18TBs in a 24-hour period. This equates to over 52 TBs within a 72-hour period, which we will round down to 50 TBs. Backing up the 50TB first-year storage requirement will require, therefore, only one drive. We will deploy a second drive to conform to the *n+1* configuration requirement, with the understanding that the second drive can be used to help satisfy the increasing weekly demands during the first three years.

We will not need to add another LTO-4 drive for the second or third year, as the initial drive can back up the 75TB requirement for the second year in less than 4.5 days. The library can also complete the third year backup (113TBs) within one week with one drive, or a half-week when utilizing both. In years four and five, we will plan to utilize both drives with the full understanding that if one drive does fail, we will still have more than enough time to replace the failed drive under the standard maintenance contract and complete the backup within the required window.

Exhibit 3 – D2D Solution Cost (in Millions of Dollars)

(Values have been rounded.)

D2D Solution (Sum of Two Tables Below)				
Year	Acquisition Cost ($M)	Energy Cost ($M)	Space Cost ($M)	Total Cost ($M)
1	$10.57	$0.09	$0.05	$10.71
2	$7.41	$0.16	$0.08	$7.65
3	$11.11	$0.26	$0.13	$11.50
4	$8.33	$0.28	$0.14	$8.76
5	$12.50	$0.42	$0.18	$13.11
5-Year Total	$49.92	$1.22	$0.59	$51.73

D2D - Backup Cache Only (13-Weeks)				
Year	Acquisition Cost ($M)	Energy Cost ($M)	Space Cost ($M)	Total Cost ($M)
1	$8.97	$0.08	$0.04	$9.09
2	$5.55	$0.12	$0.06	$5.73
3	$7.71	$0.18	$0.09	$7.98
4	$5.52	$0.19	$0.10	$5.80
5	$8.03	$0.27	$0.12	$8.42
5-Year Total	$35.78	$0.84	$0.41	$37.02

D2D - Backup Archive Only (Quarterlies Held Long Term)				
Year	Acquisition Cost ($M)	Energy Cost ($M)	Space Cost ($M)	Total Cost ($M)
1	$1.59	$0.01	$0.01	$1.61
2	$1.86	$0.04	$0.02	$1.92
3	$3.40	$0.08	$0.04	$3.52
4	$2.82	$0.10	$0.05	$2.96
5	$4.47	$0.15	$0.07	$4.69
5-Year Total	$14.14	$0.38	$0.18	$14.70

Source: Clipper Analysis of vendor specifications and data

Hardware Costs

Given the procurement plan and specifications just described, what will this cost? **There are six components involved in each years' calculation (tape library, frames, drives, cartridges, floor space, and energy), plus the cost of maintenance on the first three, after the end of the standard warranty period.** These are all shown in Exhibit 4, on the Page 8.[12]

[10] If one assumes that no disk cache is required and that writing directly to (and restoring from) tape would satisfy specific SLAs, then the very low costs shown in the blue box in Exhibit 4 (on Page 8) would apply.

[11] Obviously, this leaves the tape infrastructure idle most of the time, for recoveries, when needed, and for other uses.

[12] For a more detailed look at the components acquired each year and the acquisition costs, please see Appendix A.

Exhibit 4 has two tables, one shaded in blue and the other shaded in yellow. The second (yellow) one represents the costs of storing only the quarterly backup copies, for the purpose of long-term retention (archiving). Not included in these costs is the cost of the 13-week disk cache on which we are holding the 13 weekly and 78 daily incremental backups, for rapid recovery to meet RPOs. The middle table in Exhibit 3 (without any shading) represents the cost of this 13-week disk cache. Because the same 13-week disk cache is the same in the all-disk scenario and the mixed short-term disk and long-term tape scenario, it is OK to look at only the differences in the costs of the long-term storage (the fundamental scenario of our study). This is represented by the two yellow-shaded tables in Exhibits 3 and 4. **For long-term storage over our five-year study period, the cost of disk is about 23 times that of tape.**

The first (blue) table of Exhibit 4 (on next page) represents another scenario – the cost of storing everything on tape (consisting of the quarterlies, the weeklies, and the dailies). This would only make sense if the nature of what you were backing up could be done directly to tape within the available time windows and if you could tolerate the slower speed of recovery for going directly from tape to disk. For some enterprise situations, this is very plausible. One would need to compare the costs of this all tape scenario against the cost of an all disk scenario. The blue-shaded boxes of Exhibits 3 and 4 allow for this comparison. **The all-disk solution costs about 29 times that of the all-tape solution.**

There are a couple of conclusions that can be drawn from this data.

(1) In reality, most enterprises probably don't need more than four weeks' of ready-to-restore-quickly backup on disk, allowing another nine weeks to be on tape. So your answer may be somewhere in the middle.

(2) While the cost of disk is declining, so is the cost of tape, even though all of the tape costs were based on current LTO-4 technologies. LTO-5 is expected midway in our study period and probably should be what is purchased in the later years (both additional drives and cartridges). However, since we didn't want to price what is not yet announced, we stuck with the present LTO-4 pricing.

(3) Because we front-loaded the costs of the largest tape library that would be needed at the end of the study period (and included the maximum floor space that would be occupied for the entire period), it is safe to say that the cost of storing a TB on tape gets cheaper as the volume of data stored increases. (While the above ratios are 23:1 and 29:1, a heavier-use scenario for tape would lower and accelerate the unit costs of storing a TB per year and might result in a 40:1 or higher ratio. This might happen if you started with 100TBs of data or were growing at a 100% rate, for example.)

Media Costs

LTO-4 cartridges are available in lots of 100 for $12,600 at list price. The cost for 125 cartridges needed for the first year will be about $16,000, which will be expensed.

Floor Space Requirements

Each frame occupies 10.2 sq. ft. of floor space. When you double that to allow for aisle space, at $300/sq.ft., the cost for the first year would be about $6K. However, because you should plan for the four additional frames to be used in the following years (for additional slots), you need to reserve sufficient adjacent floor space, at a total cost per year of $30K.

As with disk, we could consider an upgrade to *LTO-5* in year 4, with a compressed cartridge capacity of 3.2TB and a projected cost of $20 - $25K. This would reduce the number of new cartridges required for year 4 from 1610 to 805 and for year 5 from 2540 to 1270. Quite clearly, this would ease the burden on the number of new frames required (at $30,000 each) and floor space, although it would not have much impact on cartridge cost, as the industry has maintained a rather constant cost per TB for the media. As before, we chose not to give any undue advantage to tape in this analysis. However, you should keep this capability in mind as LTO cartridge capacity advances from LTO-4 (1.6TB) to LTO-5 (3.2TB) and, eventually to *LTO-6* (6.4TB).

Energy Costs

Now, let us consider the energy component of the tape TCO equation. **The active tape library, without drives, consumes 185 Watts[13] of energy per hour. Let me repeat that: only 185 Watts of electricity!** Each additional active frame adds 110 Watts to the total. Each active LTO-4 drive, power supply, and associated infrastructure consumes about 50 Watts while engaged in reading/writing, 12 Watts when idle. Therefore, **the initial deployment of the library, with two drives, will consume about 1150 KWH during the first year, significantly less than the energy consumed to make your morning toast and coffee for a year**, and nowhere near Al Gore's hit list! At $0.12 per KWH, this will add $250 to the enterprise's annual energy bill, including the energy to cool the heat generated. In year 5, the data center will only add $304 to its energy bill. The enterprise CFO can even afford to paint the library green! (See Exhibit 4, on the next page.)

---

[13] We assume that an idle library will consume less than half of the active energy; therefore, we have assumed a value at 50% for idle energy use.

Exhibit 4 - Tape Library Cost

(in Dollars, unless Otherwise Stated)

Tape - Backup Everything (13 Weeks of Weeklies and Daily Increments plus Archive All Quarterlies) to Tape (i.e., No Disk Cache)												
Year	No. of (Extra) Frames Acquired	No. of Drives Acquired	Library Acquisition Cost ($)	Library Main-tenance ($)	Frame Acquisition Cost ($)	Frame Main-tenance ($)	Cartridge Cost ($)	Drive Acquisition Cost ($)	Drive Main-tenance ($)	Space Cost ($)	Energy Cost ($)	Total Cost ($)
1	1	2	$76,000	$0	$30,000	$0	$82,278	$45,600	$0	$68,850	$599	$303,327
2	2	0	$0	$3,216	$60,000	$840	$72,702	$0	$4,272	$68,850	$599	$210,479
3	2	1	$0	$3,216	$60,000	$2,520	$124,614	$22,800	$4,272	$68,850	$704	$286,976
4	3	0	$0	$3,216	$90,000	$4,200	$202,860	$0	$6,408	$68,850	$704	$376,238
5	6	1	$0	$3,216	$180,000	$6,720	$320,040	$22,800	$6,408	$68,850	$809	$608,843
5-Yr. Total	14	4	$76,000	$12,864	$420,000	$14,280	$802,494	$91,200	$21,360	$344,250	$3,416	$1,785,864

Tape - Backup Archive Only (Quarterlies Held Long Term)												
Year	No. of (Extra) Frames Acquired	Number of Drives Acquired	Library Acquisition Cost ($)	Library Main-tenance ($)	Frame Acquisition Cost ($)	Frame Main-tenance ($)	Cartridge Cost ($)	Drive Acquisition Cost ($)	Drive Main-tenance ($)	Space Cost ($)	Energy Cost ($)	Total Cost ($)
1	0	2	$76,000	$0	$0	$0	$15,750	$45,600	$0	$30,600	$233	$168,183
2	1	0	$0	$3,216	$30,000	$0	$23,625	$0	$4,272	$30,600	$251	$91,964
3	0	0	$0	$3,216	$0	$840	$35,438	$0	$4,272	$30,600	$251	$74,616
4	1	0	$0	$3,216	$30,000	$840	$53,156	$0	$4,272	$30,600	$268	$122,353
5	2	0	$0	$3,216	$60,000	$1,680	$79,734	$0	$4,272	$30,600	$304	$179,806
5-Yr. Total	4	2	$76,000	$12,864	$120,000	$3,360	$207,703	$45,600	$17,088	$153,000	$1,307	$636,922

Source: Clipper Analysis of vendor specifications and data

Tape Summary

Keep in mind that, as discussed above, in any D2D2T environment, the data center needs to incorporate a disk cache – the middle "D" – into the architecture. This storage will provide instant access to short-term recovery data and needs to be large enough to support three months worth of incremental backups and thirteen weeks of weekly backups, 845TBs for the first year. This equates to four of the same disk arrays described earlier, at a 5-year cost of $37M, as shown in the middle table of Exhibit 3.

However, as we are examining the cost for the "long-term" storage of data, we will remove the cost of this disk cache from both the disk and tape solutions to come up with a comparison for long-term storage.

Cost Comparison Summary

Assuming away the cost of storing 13 weeks of backup data in a ready-to-restore disk cache, over our 5-year study period, the cost of disk is about 23 times that of tape, for the kinds of long-term storage that we have described. (Again, the $37.2M for the 13-week disk backup cache is not included in either solution's costs.)

VTL with Data Deduplication for Tier 2 Disk

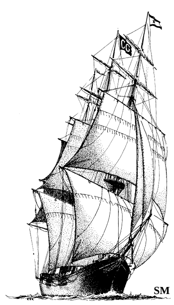
The data center could consider pushing the backup stream to a Virtual Tape Library (VTL) and use data deduplication to reduce the size of the VTL's disk cache. Data deduplication vendors make various claims regarding the reduction of data. We have seen claims from 10:1 to 100:1. We chose a factor of 20:1 in our analysis, although higher than most would see; it seemed to be the most reasonable given the number of unknowns surrounding data content. If you assume a compression ratio of 20:1 for a data deduplication VTL solution, the resulting 20X straight-line reduction could reduce the cost of the disk archive solution to around $3M, including the approximate cost of the data deduplication software, VTL engines, and VTL software, yielding about 5 times the cost of the tape component in the D2D2T solution.

Conclusion

This study began as a marginal-cost analysis of disk and tape in a scenario for long-term storage of archived quarterly backups. We quickly realized that we had to consider the value that each architecture brings to the table. With D2D, the data center acquires the ability to meet every SLA with regard to their backup window and recovery time. The enterprise gets rapid access to all of their data for whatever period they measure. That comes with costs: the acquisition cost to deploy the D2D solution, the energy costs to power and cool the systems and floor space costs.

With a D2D2T solution, the data center retains rapid access to the most recent data, 13 weeks in this analysis, and nearline access to five years of historical data. The cost of the tape library is nominal when compared to the cost of the disk arrays necessary to support the environment. The cost of energy is negligible, compared to acquisition costs or the energy costs of an all disk solution. In addition, with LTO-4 technology the data center acquires a solution with ancillary values of automatic compression, WORM capability, and encryption option at the tape drive level enhancing the security and reliability of their data.

As we look at the two architectures, there is no question as to the complementary value of disk and tape in a tiered D2D2T solution to help the enterprise address its performance, compliance, data protection, energy, and TCO objectives. All of one, or all of another, may not address all of the goals.[14] Furthermore, having data on tape, detached from the system, reduces the risk of accidental or intentional corruption. Moreover, tape is portable, enabling low cost data protection strategies. *Tape is not dead. Far from it!* Tape, especially with the reliability and speed of LTO-4 technology continues to provide the fiscal responsibility and functional value that enterprises require in the twenty-first century. The data center needs to continue to consider tape as an integral component of any long-term storage architecture and you need to focus on how much of your stored data really needs to be held in the disk cache.



Please see the Addendum on the next page and the Appendix that follows.

[14] Unfortunately, the real world has more recovery requirements than can be addressed in any single analysis. We have assumed a standard of 13 weeks of online backups, which certainly made our analysis simpler. Your requirements definitely will vary, depending on the size of your data store and the number of weeks of data that your enterprise needs to retain in a disk cache for "rapid" recovery. This model can be applied to any environment, however, once a given set (or collection) of RTOs and RPOs are established.

Addendum to the Original (February 13, 2008) Bulletin

Since our publication of February 12, 2008, issue of Clipper Notes comparing the long-term benefits of preserving backup and archive data on tape rather than on disk, there have been several new developments. The most significant of these deal with increases to the capacity and performance of the various media types used in the study. (This Addendum contains the only changes to the original paper.)

General Use of 1TB SATA Drives

When the study began, 750GB SATA drives were the largest low-cost disk devices available and qualified for deployment in a Tier-2 storage environment. Since that time, most disk arrays have been certified for the newest 1TB devices. The installation of 1TB devices in our scenario would certainly reduce the quantity of disks required to store the same amount of data, and, therefore, the amount of energy needed to power the arrays and cool the environment. On the other hand, however, the deployment of 1TB disks would almost certainly necessitate the use of RAID-6 instead of RAID-5, to protect that data further from loss during the long recovery of the failed drive, doubling the number of parity disk drives required. If a 1TB device fails, the rebuild time to restore a hot spare device to operational status could take 12 to 24 hours, or longer. The data center cannot afford to take the risk of having a second failure, thus losing the data in the RAID group. Quite clearly, the IT staff would achieve some capacity gain by transitioning to 1TB devices, however, that density advantage might be overshadowed by the additional overhead from the extra parity drive, and the extra processing required to calculate the two parity blocks. It is possible that the data center staff believes that their platform of choice is so reliable that additional RAID protection is not required but, generally, we disagree.

Considering 1.5 TB Drives as Mid-Life Kicker

It should also be noted that while we used 750GB drives to initiate the study, we did incorporate 1.5TB devices in the model in the year-4 upgrade, taking advantage of anticipated improvements in disk architecture. Many customers will take advantage of this opportunity to avoid incurring significant maintenance costs because of expiring warranties. Furthermore, we constrained the overhead by continuing to use RAID-5 rather than incorporating RAID-6 (which our rubrics dictated), in order to bend-over-backwards to be fair to disk-based solutions.)

Considering the Next Generation of Tape Drives

At the same time, it can reasonably be expected that LTO-5, and possibly even LTO-6, will be available during the 5-year period analyzed. Even so, we did not replace the LTO-4 drives with a mid-life-kicker in the tape environment because the penalty incurred from going on maintenance was expected to be insignificant, when compared to the advantages gained from the additional performance and capacity of these newer technologies. Because LTO-5 will be able to read LTO-4 cartridges, the data center would not have to replace existing media (a big advantage) while gaining the advantages of a 2X capacity improvement for newly-created cartridges (after upgrading to the next-gen tape drives. With this, the results would have been skewed heavily in favor of tape.

Cost of Tape Drives is Negligible

It should also be noted that the cost of tape drives is a minor factor in the TCO calculation. The cost of media is far more significant. The data center can scale performance as required simply by adding more drives, either LTO-4 or LTO-5. Migrating to LTO-5, when it becomes available, might cut significantly the future media expenditures, plus halving the frames, slots, and floor space required.

Data Deduplication

Although we did not include data deduplication in the long-term study, we did incorporate it in our analysis of a VTL solution, assuming a 20:1 data compression. Since that time, it has come to our attention that this ratio might be overly generous, as most industry experts and many customers have been typically realizing between a 10:1 and a 15:1 data deduplication ratio across all stored data. (It really depends on what is being backed up. While there are many data types that might be deduplicated significantly (say, 40:1), there are also many that cannot be deduplicated at all.) Combined with the additional cost of the data deduplication software, this might further skew the relative cost per GB in favor of tape.

On the other hand, we may have been too restrictive in our assessment of the compression that can be achieved using standard tape compression algorithms. Some in the tape industry claim that you can achieve better than 2:1 compression using LTO-4. While this may be true for some data sets, other sets will not be able to achieve even that much (for the same reasons involving the variability of what is being deduplicated and how often, as discussed above). Therefore, we will stand by our use of the 2:1 standard for compression because changing it arbitrarily to 1.75 or 2.25 does not alter the costs significantly, in comparison to the costs for disk.

Used Consistent Assumptions

We used the same assumptions for both tape and disk in calculating the cost per TB. We used readily-available list pricing for commodity disk and tape offerings. We did not attempt to assume any discounts, leaving that to enterprise purchasing departments who would not take our word on pricing, anyway.

APPENDIX A – EXPLANATION OF ACQUISITION COSTS IN COST MODEL

We have chosen to expense all equipment in the year of acquisition, over the chosen 5-year study period. Typically, disk arrays are replaced after three years.

We have assumed that arrays bought in year 4 and 5 will have twice the capacity but the same cost and footprint as the arrays at beginning. We chose to ignore the planned introduction of LTO-5 drives and cartridges in the later years. If we had done so, the cost per TB stored would have gone down, as would have the number of additional frames needed and the floor space to house them. Speaking of Floor space, we made a further assumption that is biased against tape. We have chosen to allocate the same space needed to accommodate all of the frames that would be needed in year 5 for all five of the years, so that adjacent equipment would not have to be relocated for the tape solution to expand.

Additionally, as mentioned in the body of the paper, we have also chosen to have all tape cartridges nearline (as opposed to having, say, all tapes older than two years being kept on a shelf nearby, but not in a library slot, and possibly not on raised data center floor, with its higher costs per square foot).

Lastly, we have chosen not to account for the time value of money. If we had done so, it would not have materially changed the results.

The result of these assumptions is clearly an intentional bias against tape. So when you see the ration of 23:1 on the first page, recognize that this could have been concocted to be biased more in favor of tape, resulting in a much higher ratio. We put tape to a challenge and it clearly was beneficial, in spite of all of the just-described biases. The bottom line is that you need to do this kind of analysis with your data and assumptions or work with a consultant, like Clipper, to get your own calculations.

See the tables below and on the next page to see what costs we allocated to each of the years for each of the components of our model.

(I) ALL DISK (ARCHIVING ALL BACKUP DATA ON DISK)

DISK ARRAYS	Acquisition Cost	**Year 1 Costs**	**Year 2 Costs**	**Year 3 Costs**	**Year 4 Costs**	**Year 5 Costs**
Year 1	$10,567,810	$10,567,810				
Year 2	$7,408,088		$7,408,088			
Year 3	$11,112,132			$11,112,132		
Year 4	$8,334,099				$8,334,099	
Year 5	$12,501,149					$12,501,149
Total	**$49,923,279**	**$10,567,810**	**$7,408,088**	**$11,112,132**	**$8,334,099**	**$12,501,149**
					5-Year Study Period Acquisition Costs	***$49,923,279***

(II) TAPE – ARCHIVING ONLY QUARTERLY BACKUPS

TAPE LIBRARY	Acquisition Cost	**Year 1 Costs**	**Year 2 Costs**	**Year 3 Costs**	**Year 4 Costs**	**Year 5 Costs**
Year 1	$76,000	$76,000				
Year 2	$0		$0			
Year 3	$0			$0		
Year 4	$0				$0	
Year 5	$0					$0
Total	**$76,000**	**$76,000**	**$0**	**$0**	**$0**	**$0**
					5-Year Study Period Acquisition Costs	***$76,000***

(This table continues on the next page.)

Source: Clipper Analysis

(II) TAPE – ARCHIVING ONLY QUARTERLY BACKUPS (continued)

TAPE FRAMES	Acquisition Cost	Year 1 Costs	Year 2 Costs	Year 3 Costs	Year 4 Costs	Year 5 Costs
Year 1	$0	$0				
Year 2	$30,000		$30,000			
Year 3	$0			$0		
Year 4	$30,000				$30,000	
Year 5	$60,000					$60,000
Total	**$120,000**	**$0**	**$30,000**	**$0**	**$30,000**	**$60,000**
			5-Year Study Period Acquisition Costs			***$120,000***
TAPE DRIVES	**Acquisition Cost**	**Year 1 Costs**	**Year 2 Costs**	**Year 3 Costs**	**Year 4 Costs**	**Year 5 Costs**
Year 1	$45,600	$45,600				
Year 2	$0		$0			
Year 3	$0			$0		
Year 4	$0				$0	
Year 5	$0					$0
Total	**$45,600**	**$45,600**	**$0**	**$0**	**$0**	**$0**
			5-Year Study Period Acquisition Costs			***$45,600***

(III) TAPE – ALL BACKUP DATA

TAPE LIBRARY	Acquisition Cost	Year 1 Costs	Year 2 Costs	Year 3 Costs	Year 4 Costs	Year 5 Costs
Year 1	$76,000	$76,000				
Year 2	$0		$0			
Year 3	$0			$0		
Year 4	$0				$0	
Year 5	$0					$0
Total	**$76,000**	**$76,000**	**$0**	**$0**	**$0**	**$0**
			5-Year Study Period Acquisition Costs			***$76,000***
TAPE FRAMES	**Acquisition Cost**	**Year 1 Costs**	**Year 2 Costs**	**Year 3 Costs**	**Year 4 Costs**	**Year 5 Costs**
Year 1	$30,000	$30,000				
Year 2	$60,000		$60,000			
Year 3	$60,000			$60,000		
Year 4	$90,000				$90,000	
Year 5	$180,000					$180,000
Total	**$420,000**	**$30,000**	**$60,000**	**$60,000**	**$90,000**	**$180,000**
			5-Year Study Period Acquisition Costs			***$420,000***
TAPE DRIVES	**Acquisition Cost**	**Year 1 Costs**	**Year 2 Costs**	**Year 3 Costs**	**Year 4 Costs**	**Year 5 Costs**
Year 1	$45,600	$45,600				
Year 2	$0		$0			
Year 3	$22,800			$22,800		
Year 4	$0				$0	
Year 5	$22,800					$22,800
Total	**$91,200**	**$45,600**	**$0**	**$22,800**	**$0**	**$22,800**
			5-Year Study Period Acquisition Costs			***$91,200***

Source: Clipper Analysis

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Authors

David Reine **is Director, Enterprise Systems for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. He joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

Mike Kahn **is Managing Director and a founder of The Clipper Group.** Mr. Kahn is a veteran of the computer industry, having spent more than three decades working on information technology, spending the last decade and a half at Clipper. For the vendor community, Mr. Kahn specializes on strategic marketing issues, especially for new and costly technologies and services, competitive analysis, and sales support. For the end-user community, he focuses on mission-critical information management decisions. Prior positions held by Mr. Kahn include: at International Data Corporation - Director of the Competitive Resource Center, Director of Consulting for the Software Research Group, and Director of the Systems Integration Program; President of Power Factor Corporation, a Boston-based electronics firm; at Honeywell Bull - Director of International Marketing and Support; at Honeywell Information Systems - Director of Marketing and Director of Strategy, Technology and Research; with Arthur D. Little, Inc. - a consultant specializing in database management systems and information resource management; and, for Intel Corporation, Mr. Kahn served in a variety of field and home office marketing management positions. Earlier, he founded and managed PRISM Associates of Ann Arbor, Michigan, a systems consulting firm specializing in data management products and applications. Mr. Kahn also managed a relational DBMS development group at The University of Michigan where he earned B.S.E. and M.S.E. degrees in industrial engineering.

- ***Reach Mike Kahn via e-mail at MikeKahn@clipper.com or via phone at (781) 235-0085 Ext. 121. (Please dial "121" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.