

Continuing the Search for the Right Mix of Long-Term Storage Infrastructure — A TCO Analysis of Disk and Tape Solutions

Analysts: David Reine and Mike Kahn

Overview and Findings

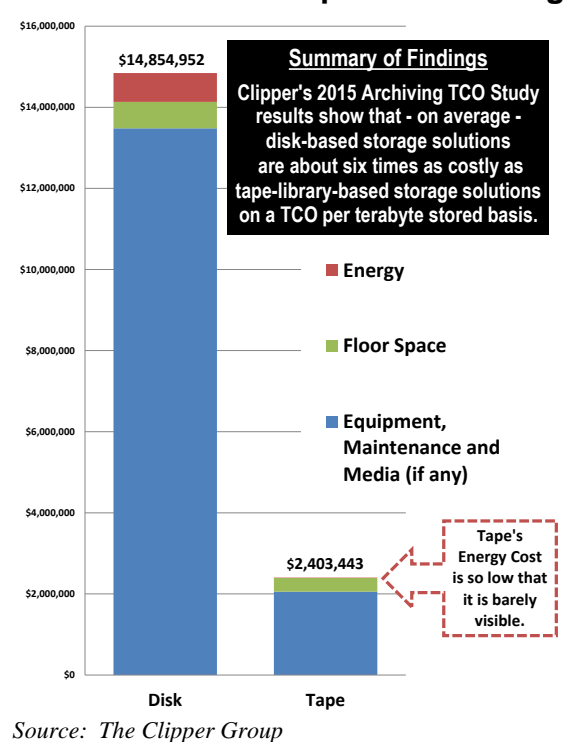
We are surrounded by data and very close to being overwhelmed by it. The desire to retain data for years, if not decades or longer, so that it might be leveraged in the future (perhaps even tomorrow), is not new. It is the scope and volume of data that is new and the scale to which archived data is being used that brings us to ask again what has become the proverbial question in data centers around the world: *Where should I put it?* Today, there are more choices than before, but the decision process to answering that question only has two dimensions, all other things generally being equal: *time* and *money*.

Time primarily is about both *duration* (how long you want to keep it) and *retrieval speed* (how fast do you need to access it). Money is about paying the least amount (on a continuing basis) per unit of storage (think about terabytes¹, because that is something sufficiently tangible to which you probably can relate), as long as the business requirements are being met. The latter is mostly about retrieval speed – the quicker you want it, the more it is going to cost.

Answering the where-to-put-it question is valid only for a given point in time. Beyond a short period (think three-to-six years), we are challenged to predict what storage technologies will be available and economically viable. Fortunately, the need is front and center now and not deferrable until later – because there is much that has to be preserved today. So we need to look at what is available today, and extrapolate that a little into the future while making some important assumptions.

For those readers who just want a quick answer, **we found that for archiving, disk-based storage solutions are, on average, more than six times more costly per terabyte stored than ones based on tape libraries, as shown in Exhibit 1 on this page.** Explanations and insight require more details and discussions, which follow. Please read on.

Exhibit 1 — Comparing the Average TCO for Disk and Tape for Archiving



¹ A terabyte (TB) is a thousand gigabytes. A petabyte (PB) is a trillion bytes, a thousand terabytes, and a million gigabytes.

IN THIS ISSUE

- Overview and Findings..... 1
- The Details of the 2015 TCO Study 3
- Conclusion 17

In 2010² and 2013³, The Clipper Group, a 22-year-old computer industry analyst firm and publisher of this report, addressed the economic challenges of storing archived data. This year, Clipper has self-funded another study (herein called the *2015 Archiving TCO Study*), with some variations from what had been done previously, as will be explained. **This report describes the findings and methods of this third-generation total-cost-of-ownership (TCO) study comparing disk and tape as the major targets for storing many petabytes, or maybe even exabytes, of data for the long term.** It is important that you focus on this study and its results and not try to compare them too closely to the prior two studies, because enough has changed in our TCO model to make that problematic, especially if you choose to project our findings into the future.

What went into the TCO calculation? In addition to the acquisition cost of the required hardware, we included the cost of maintenance, floor space, energy, and media⁴. Everything was calculated based upon vendor list price, except for the tape media which has broader availability from a number of sources. Tape cartridges were priced slightly above where we estimated the street price to be. It should be noted that disk solutions typically come with a three-year maintenance and support warranty, although some vendors included a six-year warranty with an adjusted price. Tape libraries typically carry a one-year warranty, with paid annual maintenance contracts after that.

Need to Ignore the Declarations of Death

In the past, many pundits have claimed that “Tape is Dead”. Well, years later, tape continues to thrive and reach new capacity and performance milestones. Now, some are ready to declare that “Disk is Dead” (referring to traditional rotating disks), as the cost of solid-state storage (such as in SSDs) appears on a downward trend, while the growth in the capacity of rotating disk drives seems to be slowing. However, we suspect that disk also will continue to improve in the years ahead (at least in the near future), though it is not

clear by what technological methods nor exactly at what rate of capacity growth. As a result of neither being dead, this year we continue to focus on both tape and disk as the primary storage infrastructure for today’s long-term archiving solution.

Why Limit This Study to Just Disk and Tape?

In this archiving study, we are looking at the long-term preservation of data stored as files (and, indirectly, as objects) and need to do this at the lowest possible cost that will accomplish our archiving objectives. Remember, that we’re talking about *archived data*. **The urgency to retrieve immediately the vast preponderance of archived data tends to be much less than with transactional data.** This is exactly why we have determined that today disks and tape are required for archiving and SSDs are not⁵. **If archived data needs to be retrieved somewhat quickly (as a valid business requirement), it should reside on disk (or maybe even on some form of solid-state storage). However, if a minute or two for retrieval is acceptable and it can be stored at a much lower cost, then tape is the place for archived data to reside.**

In fact, in our TCO model we presume that only 15% of the growing collection of archived data will be touched in any given year.⁶ **Saving about 84%⁷ of the cost of using disk (by using a tape library solution as the target archiving media) usually makes a lot of sense (especially for the archived data not being retrieved) unless there is a critical need requiring near-instantaneous delivery of the requested data.**⁸

What has Changed, in Terms of the Underlying Storage Technologies?

In general terms, disk and tape technologies are much improved (in terms of capacity per disk or per tape cartridge and costs per terabyte) than in the beginning of 2013, when we did the previous study, although there is a notable exception,

² See the issue of *Clipper Notes* dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at <http://www.clipper.com/research/TCG2010054.pdf>.

³ See the issue of *The Clipper Group Calculator* dated May 13, 2013, entitled *Revisiting the Search for Long-Term Storage – A TCO Analysis of Tape and Disk*, and available at <http://www.clipper.com/research/TCG2013009.pdf>.

⁴ Tape media is priced separately from the library that holds it. You need both.

⁵ On a more practical level, none of the storage solutions offered for this study used SSDs as part of the archiving solution. However, there is no doubt that this is coming.

⁶ 15% is consistent with the reality of many archived data collections, where the dominant IT activity is preserving the data for possible later use and the secondary IT activity is using some small fraction of what has been preserved. Thus, we tend to be more focused on the time it takes to store (write) the rapidly-growing new data than on the time it takes to retrieve a single item of data.

⁷ 84% is derived from the 6.18:1 TCO ratio of disk over tape.

⁸ This comparison is based on the average costs across multiple solutions of the same type.

as will be explained shortly. We looked at the many costs of storing a lot of data on rotating disk infrastructure and in tape libraries.⁹ The degree and nature of the improvements (or lack thereof) will be discussed in detail in the following pages.

Our fundamental expectation has not changed (from our two prior studies) – most larger enterprises almost certainly still will need a mix of both disk and tape infrastructure for their archiving solutions – and certainly this will be true for the next three years.¹⁰

High-Level Summary of Our Findings

The bottom line of the 2015 Archiving TCO Study is the same as in our earlier studies and should not come as a surprise – **disk-based archiving solutions cost a lot more than tape library solutions – more than six times as much as tape (based on the TCO per terabyte stored) – 6.18 times – to be exact.**

Thus, tape library-based archiving still costs significantly less per terabyte to store data than with rotating disks¹¹. As before, disk-based solutions typically still are much faster at retrieving modest amounts of data than when using tape in a tape library.

While not directly comparable, in our 2013 study this was about 26:1. There are many reasons for this narrowing, but the largest factor is that *Ultrium* tape-based library solutions today are using the same *LTO-6* drives and cartridges as

when we did the 2013 study two years ago.¹²

Even with this narrowing of tape's TCO advantage, there continues to be a significant economic justification for tape (measured in millions of dollars), as long as the business use cases can tolerate a retrieval time that usually will be measured in many seconds to a couple of minutes rather than fractions of a second to maybe a few seconds that disks would deliver.¹³

In summary, the question is not “tape or disk”, but when to use tape and when to use disk, on the assumption that it is very likely that you will use both for archiving, driven largely by your enterprise's time and money parameters. What you care about is the 6:1 ratio of the average TCO of disk divided by the average TCO of tape. Be careful not to jump to conclusions, because the timing of this study (with respect to the announcements of next generation of disk and tape storage devices) was an important factor in the results.

Please read on to get these details and much more.

The Details of the 2015 Archiving TCO Study

In addition to the generational incongruities mentioned briefly, this study has somewhat different parameters and presumptions than the prior two studies. This was deliberate. In our earlier studies, block storage was the common denominator for both disk and tape. We didn't address what was being stored inside the block, except for the basic presumption that it was irreducible (uncompressible) bit streams (which most likely already had been compressed and thus could not be compressed further). By making this presumption, we avoided the messiness of files and the hierarchies that they bring. **This time, we chose to focus on files,** which had a significant impact on the solutions submitted by some of the

⁹ While solid-state storage is a reality, for archiving we deemed it's not to be a low-cost per terabyte solution today and thus not appropriate for large-scale archiving. This likely will change, especially when looking out more than three years. However, today's archiving solutions must be based on what is available and practical today and affordable.

¹⁰ We consider the next three years (think 2015 to 2017) to be defined in terms of the storage technology that is available today and at today's costs. While the technology inside archiving solutions beyond the next three years will include disk and tape and solid-state storage (like SSDs), it gets murkier the further one projects into the future, especially regarding TCO per terabyte stored. We only can use our professional judgment to estimate what will be available (in terms of capacities and write speeds) and what it will cost to purchase and operate. In making these future-looking estimates, we assume that the next generations of disk and tape drives/cartridges will be viable (that they will be worthwhile, workable, and cost effective). No doubt, rotating disks will be more challenged to meet the growth in capacity that we forecast for the next two generations than will be true for tape, which already has demonstrated research prototypes of very high tape capacities.

¹¹ See the issue of *Clipper Notes* dated July 5, 2014, entitled *Is Tape the Best Low-Cost Technology for the Preservation of Data,* and available at <http://www.clipper.com/research/TCG2014015.pdf>.

¹² The next generation of Ultrium tape – *LTO-7* – is expected to be announced late this year and we expect that it will be generally available early next year.

¹³ You have to be careful when thinking about retrieval times, because they most certainly are dependent on the volume of data being retrieved. If the volume is small, then disk will be much speedier than tape. However, if the volume is large, say 10GBs or more (like a high-definition video file), the elapsed time to receive the last bit will be a much closer race, because the time it takes to mount a tape cartridge and get to the data will be outweighed significantly by the time it takes to retrieve the entire volume. Thus, the required time for retrieval is very use dependent (on whether you need to access all of the data before beginning to use it, or not. For example, if you are viewing an MRI, having only some of the image may be of little value.

vendors, especially the tape library solution vendors, as will be described.

What follows is a discussion of the many interrelated parameters that we considered (and some that we chose to ignore), including the details of each, the presumptions made, and our rationale for how we handled each within our TCO model.

The Realities of an Imperfect World

Timing is everything. For example, when new car models are introduced, typically in late summer and early autumn, they are marked with a “new model year”. If you buy near the beginning of the model year, you tend to pay more than you would 10 months later, when the next model year is approaching. Such uniformity of cycle has advantages for the buyer who can time whether to buy now or buy later, depending on his/her priorities, because s/he will find most new car dealers selling at the same place in that annual cycle.

Unfortunately, there is no model year when it comes to buying storage at enterprise scale. Changes in storage “models” (storage infrastructure) almost always are driven by the arrival of a new generation of underlying physical storage technology.

For disk and tape, the primary indicator of improvement is storage capacity per “physical unit” – where you get more capacity in the same physical space. For disk, the physical container is a 3.5-inch or 2.5-inch disk module.¹⁴ For tape, it is a cartridge, although there is less uniformity in physical dimensions for tape cartridges (and, as a result, they each usually will have different dimensional and other requirements for the library “slots” in which they are housed). **The problem for disk and tape drives and tape cartridges is that the arrival of new models isn’t synchronized, although there are times when the cycles of one coincide with that of the other.**

In a perfect world (from a TCO modeler’s perspective), announcement of the next generation would line up nicely and most of the vendors would be using similar capacities as their storage building blocks. In the spring of 2015, we weren’t so lucky. As you will learn in greater detail shortly, the disk solutions that we have included in our 2015 Archiving TCO Study have

¹⁴ Today, the largest 3.5” disk drive module will hold more data per cubic inch or centimeter than the largest 2.5” disk drive module. All of the submitted disk solutions in this study used 3.5” disks.

been chosen from a wide-range of disk drive capacities, including 4TB, 6TB, and 8TB 3.5-inch drives. We asked disk vendors to submit data and pricing on their “best *files-based* solutions” and got much more variety than we had expected, including multiple entries from the same vendor (which offered different capacity disk at differing costs per terabyte of raw capacity).

This required us to select the “best offering” from each disk-based solutions vendor, i.e., the one with the lowest TCO per terabyte. In addition, we decided to eliminate some disk and tape solutions that we concluded were way out of line (much higher in cost) from the rest of the solutions received. These outliers, for whatever reason¹⁵, cost many times other solutions and we deemed that no reasonable buyer would pay so much more for a similar solution.

As a result of these decisions, we only included configurations using 6 TB and 8 TB disk drives in our final calculations. We excluded the 4 TB disk drive entries as “outliers” because they were significantly more expensive than the others, including some entries from the same vendor. If we had included configurations based upon 4 TB disks, the ratio of 6.18 would have been considerably higher (because the average TCO for disks would go up).

Something similar is true for tape cartridges (and the drives that can read and write them). In this study, we received some automated tape library solutions based on *Linear Tape Open (LTO) drives and cartridges*¹⁶ and others based upon *enterprise tape drives and cartridges*¹⁷. We

¹⁵ One likely reason is that the list price was extremely high but almost always would be accompanied by a very hefty discount (think greater than 65%). Another may have been the 4 TB capacity disk drives that were being used, which required more cabinetry, electronics, floor space and energy than larger capacity disk drives.

¹⁶ *What is LTO?* LTO is a high-capacity tape solution developed and enhanced by a consortium headed by IBM, HP, and Quantum. It is a powerful, scalable, and adaptable tape format that addresses the growing demands of data protection in the open systems community.

¹⁷ *What is an enterprise tape?* While the characteristics of enterprise tape are similar to LTO tape (i.e., they all use a cartridge to store the data and a tape drive to write and read it), the differences are in the specifications, details, and origins. Today, for those tape library vendors that offer both LTO and enterprise tape solutions, the same tape libraries are used for LTO and enterprise tape solutions. Different vendors have different accommodations to handle the physical differences, but the fact that the libraries can handle both speaks to the commonality among them.

Enterprise tape has its roots in the mainframe world of decades past, where superior quality and performance were (and still are) strong requirements. While still used for mainframes and other computers, enterprise tapes applicability has broadened

included both in the 2015 Archiving TCO Study, because we believed that the significant capacity advancement in enterprise drives would bring down the average TCO per terabyte significantly. Including both is a major difference from our previous studies, which only focused on LTO.

Both LTO and Enterprise tape do WORM¹⁸, inline compression, application-managed encryption, and partitioning for LTFS¹⁹. LTFS is another new requirement for tape libraries in the 2015 Archiving TCO Study.

Tape products (tape drives and cartridges) tend to be on 2.5-to-4 year generational cycles, but there have been surprises, when the “leap ahead” happens sooner than expected²⁰, due to the highly competitive game of leapfrog, especially between the two enterprise tape manufacturers: IBM and Oracle.²¹ **This is why something almost always is incongruous (out of line) within and between the tape and disk worlds, except when all of the next generations magically align nicely.**

While the enterprise tape providers made leaps forward in the last couple of years²², today

to include use with open systems of all types. Enterprise tape has higher capacities, higher performance levels, and higher reliability than LTO tape. While once being labeled as both *proprietary* (because there are two enterprise tape drive manufacturers, with incompatible formats – IBM and Oracle) and *expensive*, both of these conclusions tend to be erroneous today, at least when used at large scale with open systems servers and applications.

We consider both enterprise tape libraries and LTO tape libraries to be equally viable solutions, assuming that they can satisfy your requirements. In particular, if you have one or the other (and not both) then you may prefer to stick with the format that you currently are using. We see and treat the tape solutions similarly and report average costs across all of them. However, because there are more LTO solutions than enterprise solutions in our sample (not a surprise because all of the tape library vendors have LTO solutions but only a few also offer enterprise tape solutions), the cost averages are weighted more by LTO solutions than enterprise solutions.

¹⁸ WORM=Write Once Read Many. This is a protection scheme to guarantee that data stored has not been altered.

¹⁹ *LTFS (Linear Tape File System)* is a self-describing file system for file data stored on tape media (a.k.a. “files on tape”) and the implementation of specific software that supports the special format to provide a file system interface to the data stored on them. These tapes are formatted into two tracks: track 0 for the metadata describing the files, and track 1 for the files themselves. This metadata identifies files stored in a typical hierarchical directory structure and LTFS presents that file structure to users and applications accessing the data.

²⁰ By pundits and prognosticators (like Clipper) and by the data center buyers of storage technology.

²¹ The LTO library vendors tend to be synchronized with respect to LTO drive and cartridge availability, since they all are working off of the same product specifications.

²² See *The Clipper Group Captain's Log* entitled *Enterprise Tape for Archival Storage? – Why This Just Might Make Sense*

we find tape library vendors using the same generation of Ultrium tape (LTO-6) as when we did our 2013 study. The next generation, *LTO-7*, with a significant expected increase in capacity²³, is due to be announced at the end of this year and we expect will begin to ship early in 2016. This, of course, means that LTO-6 technology – especially the media – is priced much lower per cartridge than when it was introduced in 2012²⁴. That has been factored into this study, but our adjustments aren't enough to offset the higher TCO per terabyte stored caused by the lesser capacity of LTO-6, because more slots and frames (and thus requiring more square feet of floor space) are required to hold the many more cartridges.

You might be thinking, “Why didn't we delay this study until next year, after the LTO-7 generational clock ticks?” We did consider that, at length, but decided that enough had changed since our 2013 study that we couldn't wait, just as you as a storage technology buyer usually cannot wait to buy more storage “until the next model year”, when things likely are to be “better” and cheaper per terabyte than they are today. As a result, you will see some interesting results and we will attempt to point out and interpret many of the time-dependent incongruities.

Spending No More than Required

The enterprise IT budgetary balancing act requires that the data center figure out how to store archived data without exceeding budgetary limits. **The compromise tends to be in how fast data gets stored and retrieved. The goal is to accomplish this *within the longest acceptable time frame*. You do not want to pay more to go faster than you need to, because there only is increased economic cost (without any additional economic benefit). The goal is the right balance of processes – some high(er)-performing and some not.** However, if instantaneous retrieval is *not* a serious business requirement for which you are willing to pay more, then a lower cost solution should be a preferable option, unless you have money to burn.

dated March 31, 2013, and available online at <http://www.clipper.com/research/TCG2013005.pdf>.

²³ According to the LTO Road Map; see <http://www.lto.org/technology/what-is-lto-technology/>.

²⁴ See **The Clipper Group Navigator** entitled *Magnetic Tape Turns 60 – The IT Industry Receives Another Gift*, dated July 12, 2012, and available at <http://www.clipper.com/research/TCG2012015.pdf>.

A New Focus on Files instead of Blocks

Files and objects represent the fastest growing segment of data, driven by many trends, including the mega-growth of media collections (videos, photographs, medical images, etc.) and the desire to do analytics on all kinds of data (especially on semi-structured and unstructured data). Equally important is the continuing trend to want to save just about everything for future use, which often is driven by government or industry regulations or standards but, increasingly, is driven by competitive demands for information. While archiving largely used to be about “save and forget” or, maybe a little less sarcastically, as “save and rarely retrieve”, *active archiving* (using archived data in regular operations) is becoming very important.

We chose to focus the 2015 Archiving TCO Study on files for several reasons.

1. Files continue to be the prevalent type of data stored.
2. Files on tape generally are available from all tape library vendors (using LTFS storage and retrieval methods.)
3. Files on disk are something that almost everyone understands and can relate to.
4. Doing a generally-equivalent TCO analysis of files on disk and tape was possible now (and would have been much more difficult in 2013 and earlier).

We chose not to include objects at this time because the solutions are less uniform across the vendors, which would have made the comparisons more difficult, and we felt strongly that when all things were considered and done, the TCO of storing a petabyte of objects would be very similar to the TCO of storing a petabyte of files.²⁵

Can't Clearly See Too Far into the Future

Underlying disk and tape technologies will continue to change as the new generations are announced. That is certain. **What is not as certain is the pace of capacity improvements and the increments in storage density that will occur.** Additionally, it isn't clear whether or when the TCO for solid-state storage (whether Flash-memory based or other) will be less per

²⁵ Objects clearly are the up-and-coming darling of the information industry and will become very important, but managing objects (as opposed to files) requires different software and, thus, further TCO modeling complexities. Since our goal was to focus on and compare the underlying storage infrastructure (and not the TCO of broader, a multi-purpose archiving system), focusing on files gives us the meaningful results that we believe can be extrapolated to other situations.

terabyte than is available on rotating disks – especially for the long-term storage of data where cost per terabyte stored usually is more important than a very fast speed of retrieval.

Thus we based our 2015 Archiving TCO Study on the storage infrastructure (technology) that is available now and what it costs now to procure and operate, and extend these parameters into later years by extrapolation and expert assumptions (described herein). We did make different rate-of-capacity-growth assumptions for disk and tape; these are our best estimates of what we think will happen. Of course, all storage infrastructure is procured in the present based on the presently-available technologies. **Therefore, this study should be considered valid for 2015 and will need to be updated as new technologies and alternatives are added to the storage vendors' arsenal, and especially when LTO-7 is available.**

Defining the Relevant Time Frame

As in the previous study, we presumed three cycles (periods) of three years each; thus resulting in a nine-year study period. This is as far into the future that we feel comfortable in estimating the future capacities. Nine years also is a sufficiently long time for the rapid growth of data to be felt and for the “weight of forever” to be felt.

For this study, we presumed that Cycle 1 began on the first day of 2015 and thus the three cycles are:

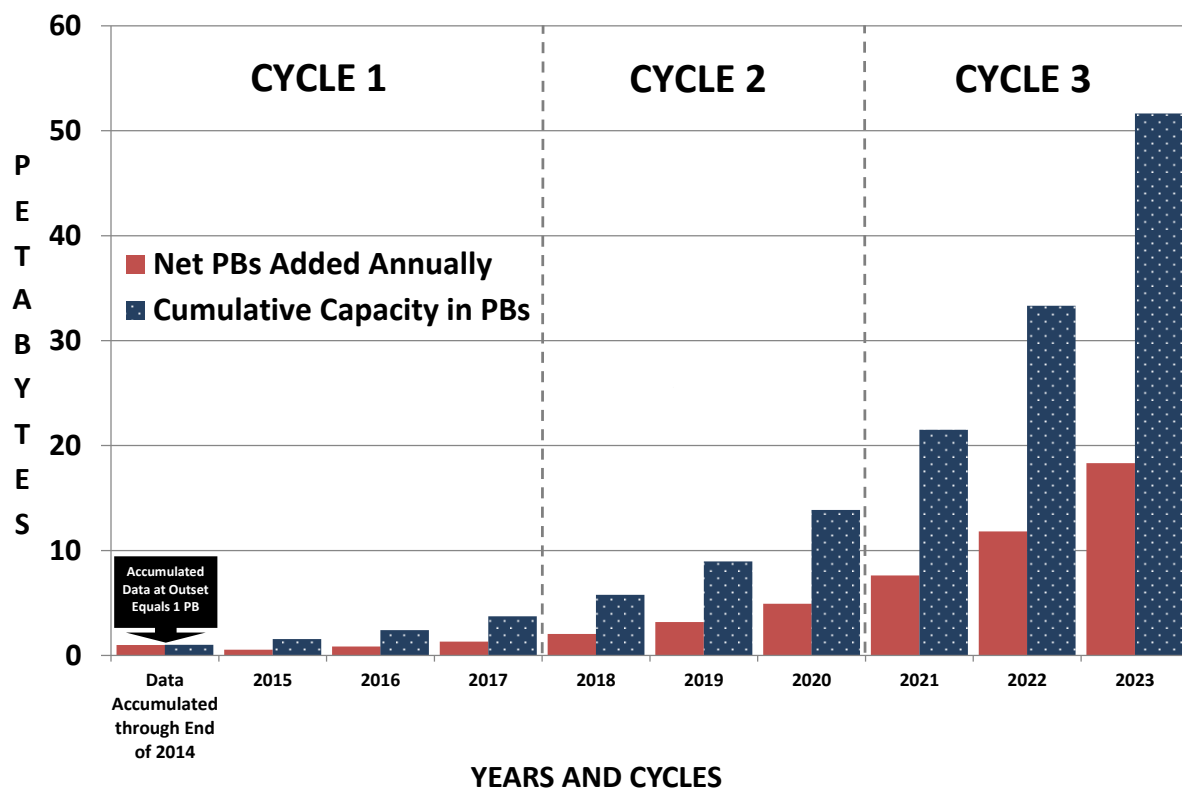
- **Cycle 1 – 2015 through 2017**
- **Cycle 2 – 2018 through 2020**
- **Cycle 3 – 2021 through 2023**

The years represented by each cycle and the presumption that the years are whole calendar years are arbitrary but quite useful for convenience of communication. When you see reference in this paper to a specific year, it should be mapped back to the cycles listed above.

How long is data retained?

In discussions with enterprises and storage vendors, retention periods of both 50 and 100 years have been mentioned. Since most of us won't be around in 50 or 100 years, we began to talk about *forever*, which we defined as longer than we can imagine or beyond which we can plan reasonably. In reality, it's very hard to differentiate between what will need to be kept forever and what will need to be kept for five or ten years. From today's perspective, it all looks very much like “forever” when talking about what you need to do in the present, which is to store it and

Exhibit 2 — The Growth of Archived Data Throughout the 9-Year Study Period



Source: The Clipper Group

make it accessible for retrieval.²⁶

More Data to Store

In our prior two studies in this series, we started with a petabyte of already-accumulated data and incremented it at a rate of 45% per year. We presumed then and now that no data is deleted from the archive once it has been added. While this presumption was made for modeling convenience, it does reflect what is going on in many enterprises.²⁷

For this analysis, we looked at different annual growth factors for the nine years covering three cycles, including 45% (which we used in the prior two studies), 55%, and 65%. Starting with a petabyte of data on day one of the first year, the

amount of storage grew over nine years to 28 PBs with a 45% growth factor, to about 52 PBs with a 55% factor, and to over 90 PBs with a 65% growth factor. **We chose to use 55% as the reasonable growth rate for this study.**

Of course, doing so significantly increased the amount of *usable* file storage capacity that will be needed during the study period (more on this to follow); basically, it almost doubled by the end of year 9, when compared to the previously used 45% rate. While this isn't an enormous amount of data (which today might be measured in exabytes for many enterprises), we are confident that it is sufficiently large enough for us to come up with some near-steady-state TCOs per petabyte stored (i.e., the cost of storing the next petabyte should be nearly the same as the prior petabyte). (See Exhibit 2, above, for a chart showing the growth of archived data in our study model.)

Interestingly enough, the number of tape drives required to support the reading and writing, as required by the study, has remained fairly low with a maximum of 12 drives being required to satisfy the 55% growth factor for the smaller-capacity LTO solutions. The enterprise tape

²⁶ Of course, much, if not most, of the data being archived will be held longer than nine years, but calculating farther than nine years into the future would not have changed the results (the key disk-to-tape TCO ratio), because our model is extrapolated largely from what is available today.

²⁷ While this “no delete” scenario is valid for many (but not all) enterprise situations, our goal was to rapidly increase what was being stored over the nine-year study period. To arbitrarily decide to delete a certain percentage of the stored data (say due to aging) would not have changed significantly (or at all) the key TCO ratio between disk- and tape-library-based archiving solutions.

**Exhibit 3 — Estimated Average Usable Capacities
After Low-Level Formatting for Protection Scheme (if required),
High-Level Formatting for Files (if not already done) and
Adjusting for Maximum Fullness (of 85%)**

LTO-6 Tape	71.4%
Enterprise Tape.....	80.8%
Rotating Disks.....	65.0%

Source: Data provided by tape library vendors. Assumptions and calculations by The Clipper Group.

solutions required fewer tape drives, because they write faster than LTO drives, which was the major determinant in sizing the number of tape drives needed for each cycle.²⁸

How Much Data and When?

Unless your enterprise's management provides an unlimited budget²⁹ to the data center, the data center staff is faced, on a continuing basis, with trying to do more with less. This 2015 Archiving TCO Study is about the compromise and balance between time and money for the storage of a lot of data which is constantly growing at a very high rate and needs to be retained and retrievable for a very long time, probably for decades, or perhaps forever, as the data center may not know when that data will be needed.

We made the following presumptions

- **There will be “a lot of data”**, by which we mean it will be measured in many petabytes (and starting with one petabyte at the beginning of 2015 and growing to about 52PBs by the end of the nine-year study).
- **The data will be in files**, which likely will be large and irreducible³⁰.
- **The data will be growing at a constant rate of 55% per year**. This refers to the growth in

²⁸ These drive quantities are the total drives in the solution at the end of the nine-year study period and include older drives still housed within the library but not being maintained. These older drives will be used to supplement the newest generation of drive when it comes to reading files on older cartridges.

²⁹ If you have an unconstrained budget, this paper still may interest you, as even the rich don't want to overspend without an adequate return.

³⁰ By “irreducible”, we mean that it can't be compressed, either because it already is in a compressed format or because of the nature of what is in the file. Compression ratios are very dependent on the contents of a file or object. We chose to assume that compression is not going to offer any savings. Doing so simplifies our model, but does ignore that LTO and enterprise tape drives are capable of compressing data as it arrives at the drive and decompressing it in real time as it is retrieved, at no extra cost. This might give tape an additional 2-3 times advantage over disk (which typically isn't compressed without extra costs). We did not consider deduplication.

volume of new files and not because the previously-stored files are growing individually.

These three presumptions create a set of requirements that we think are “enterprise scaled” and which can be extrapolated linearly to larger collections of data. If you have much smaller capacity requirements, your TCO per terabyte no doubt will be higher than what we present herein.

Need to Adjust Each Vendor's Product Specification to Deal with Truly Usable Capacity and Not Raw Capacity

What we care about is usable capacity for storing files and not the total raw capacity of the devices being used in the underlying infrastructure. Media (disks and tape cartridges) needs to be low-level formatted, in general, then needs to be adjusted for protection mechanisms that are deemed to be necessary, high-level formatted to accommodate files, and then adjusted to a point of likely average maximum capacity. (See Exhibit 3, above.) The methods for doing this are different for disk and tape.³¹

Challenges Presented by Including Both LTO and Enterprise Tape Solutions

As mentioned earlier, in prior studies we included only LTO-based tape solutions and used the LTO Program's official roadmap to represent the expected storage capacities of future LTO generations. This made the TCO analysis simpler because all tape library vendors' solutions were using the same class of drives (as they all were designed to the same specification). In addition, at the time of the 2013 study (before the latest round of leapfrog), enterprise tape did not seem to

³¹ Disk-based solutions have more overhead than tape primarily because of the protection mechanisms employed to prevent data loss or outages due to the failure of a disk drive. Disk drives have a much higher failure rate than cartridges. These mechanisms include RAID-6 or replication to multiple disks. Additionally, we assumed that all of the solutions would fill the disk drives and tape cartridges to a maximum of 85% of usable capacity. (We call this the *fullness factor*.) This is a more generous assumption for disk, as disk-based solutions tend to be less fully filled than tape cartridges).

be priced competitively against LTO on a cost per terabyte basis, so it wasn't included.

Much has changed since then, as next generations of both formats of enterprise tape and cartridges (*IBM TS1150* and *Oracle T10000D*) were announced and became available since the publication of the 2013 Archiving TCO Study.³² As mentioned earlier, LTO technology's "next leap" forward hasn't happened yet, so we are using LTO-6 drives and cartridges in Cycle 1.³³

In simple terms, current enterprise tape appears now to cost less per terabyte than LTO-6³⁴. In addition to not yet making the expected big leap to LTO-7, the higher densities of the enterprise cartridges means that many fewer cartridges will be needed (in comparison to LTO-6), with an equivalent reduction in the number of slots needed in the tape library, which means fewer frames and less data center floor space. The enterprise drives also read and write faster than LTO-6, meaning that fewer drives may need to be procured (to satisfy our model's writing capacity presumptions).

Of course, not all of the tape library vendors offer enterprise tape solutions. Those that do handle the variety of physically-different cartridge dimensions in different ways. We paid close attention to this, as the number of slots in base and expansion frames was different for each vendor.

In addition, most tape libraries scale differently and reach maximum capacities at different slot capacities. For very large archives³⁵, this might require that some solutions add one or

more additional libraries. We needed to pay attention to this, as there always are extra costs associated with an additional library. However, all tape library solutions included in the study required only a single library to hold the 52PBs of data.³⁶

For Cycles 2 and 3, we went with LTO-7 and LTO-8, using the capacity and performance data presented in the published LTO Program road map. For enterprise drives in Cycles 2 and 3, we estimated what we thought were conservative capacities (20 TBs and 35TBs per cartridge, respectively), based on published research data from the enterprise tape drive vendors and the leading manufacturers of tape media, and applied the rates of improvements in write speeds in the LTO road map to the existing enterprise tape write speeds.

Cost of Media

For tape library solutions, the cost of media is a significant budgetary factor. In simple terms, tape media is extra cost and not part of the infrastructure price of a fully-working tape library. While many data centers buy their cartridges from their library vendor, the market for cartridges tends to be more commodity-like. Whereas everything else in the 2015 Archiving TCO Study is based on list pricing³⁷, this turned out to be a big problem for tape media. Some vendors have exceptionally high list prices for tape cartridges while others offer something similar at a fraction of the cost per terabyte. On the assumption that a very large number of cartridges will be procured during the 9-year study period (from more than 2400 (for enterprise cartridges) to more than 8400 (for LTO cartridges), depending on tape type), we know that this represents a "big buy" procurement and thus would receive an accompanying significant discount.

Based on available data, we have priced the cartridges a little above what we think is street price. We treated each vendor's cartridges differently, when there is a difference (i.e., all LTO cartridges are priced about the same, regardless of tape library vendor), but different prices are used for the different enterprise tape cartridges.³⁸ It

³² The IBM TS1150 cartridges hold 10 TBs of native block data and the Oracle T10000D holds 8.5 TBs. (Native means without compression and formatting for files.) When formatted for LTFS and after an adjustment is made for fullness (average percentage of the capacity actually used, which was set at 85% for both disk-based solutions and tape library solutions), each cartridge holds less data than its native (unformatted and uncompressed) physical capacity when filled to the maximum (as shown in Exhibit 3, above). All of these factors have been taken into consideration.

³³ LTO-6 cartridges natively hold 2.5 TBs of data and less after formatting for LTFS and adjusted to an average 85% fullness (as shown in Exhibit 3, on the previous page).

³⁴ As indicated earlier, we chose not to break out the TCO details on LTO versus enterprise tape. In soliciting the cooperation of storage vendors for this study, we had promised not to reveal any vendor's cost or configuration data and decided that there were too few enterprise tape vendors in our sample to make this possible. (While there are only two enterprise tape drive formats from two tape drive manufacturers, some other library vendors procure enterprise drives from one of the manufacturers and thus there are more than two tape library vendors selling enterprise-drive-based solutions.)

³⁵ With capacity requirements well beyond the 52PBs that we modeled for this study.

³⁶ While mid-sized tape libraries possibly could be used to deliver a 52PB solution, we deemed that most enterprise data centers would not go this route. All of the tape libraries included in this study are the largest of their type offered by a vendor, i.e., they all are "enterprise class".

³⁷ Typically, enterprise data centers will be offered a sizable discount by vendors of tape libraries.

³⁸ For LTO, we used a price of \$50 per cartridge and for enterprise tape we used a price of \$270 per cartridge. The

almost always pays to shop around for tape media.

We asked each tape vendor for their recommended best quantity to order tape cartridges (i.e., the best economic lot size) and were surprised that they were uniformly low, from less than 10 to about 30. In our TCO model, we used the quantity provided by the vendor. One advantage of this was that there were only a few unused tape cartridges at the end of each cycle, which we chose to ignore because their cost was very small.

Challenges in Estimating Future Disk Drive Capacities

Rotating disks usually require new technologies and/or designs to deliver increased capacities. While disks once were doubling capacities every 12-to-24 months, the pace has slowed significantly in recent years, as older methods and materials approached critical limits. New technologies and methods have begun to arrive, but it still is unclear how often the devices will be able to double their capacity. Accordingly, we have used 12 and 20 TBs for the estimated capacities for Cycles 2 and 3, respectively, which are sizable jumps from today's prevalent 6 TB drives (and some 8 TB drives) that were used in Cycle 1.

The Need for a "File Controller"

Both disks and tape require a file controller to manage what is placed on which device or cartridge. We solved the problem on the disk side by asking vendors to provide us with their best (lowest cost per PB) file-storage solution (which might be generalized as network-attached storage (NAS)). We did not specify a storage architecture, and have both scale-up and scale-out disk-based solutions in our sample.

Files on tape present a different problem. While LTFS allows a single tape cartridge to be mounted as a file system, this would only account for one cartridge's contents. When you have many thousands of cartridges containing files (as we do in our TCO model), you need something (typically a piece of file storage management software running on a server) to keep track of which files are on which tape cartridges. We call these "LTFS servers", and each LTFS tape library solution needs one (or several). The TCOs that will be presented for tape include the extra costs of LTFS servers, software and maintenance. If an

sizable differences are due mostly to the capacities provided but also represent where the products are in their life cycle, with LTO-6 being discounted heavily since it is much later in its cycle. In general, enterprise cartridges cost more per terabyte than LTO cartridges.

LTFS server could failover (most could), we doubled the cost to include two of them, as one would represent a single point of failure.

We did not require a single namespace solution (i.e., one that expands across many NAS storage solutions and/or many tape libraries). Some have this capability, while others don't. This may, or may not, be important to you, depending on how you can subdivide your archive storage needs.

Protection

Disk and tape employ different mechanisms for error correction and device failure. We needed the archived data to be available and accurate. All devices are subject to failure, but some more than others. For example, failure rates are much higher for a single disk drive than for a single tape cartridge, because disk drives always are spinning at high speeds³⁹. This requires a lot of sensitive components in comparison to tape cartridges (which are much simpler electro-mechanical devices). In addition, disk drives tend to operate at much higher temperatures than tape drives; this adds a lot of stress to a disk drive's internal components.

This especially is true for many of the high-capacity, lower-performing disk drives typically used in NAS-based archiving solutions. So for disk, we specified a two-points-of-failure-solution, such as RAID-6 or something at least as equally good.⁴⁰ If a disk drive fails unexpectedly (or is about to fail and has issued warnings), that drive will be replaced and rebuilt. In the meantime, data continues to be available. Related to this, we specified next day service for disk maintenance, because having spare drives around to be replaced by someone from the data center staff always seems to be less costly than signing up for two or four-hour response times.

For tape, what is most likely to fail is a tape drive, because they are electro-mechanical devices that often are operating continuously.⁴¹ To account for this possibility, each library procured has an additional "spare" tape drive of the

³⁹ This is even true for "low RPM" drives.

⁴⁰ With two points of failure, if one disk is lost, no data is lost and the failed disk can be replaced and repopulated while the data still is in use and being stored. It takes a second disk failure within the same RAID group to create a major fault. That is why archives need to be backed up, preferably to a remote location, as a single data center is subject to a catastrophic happening (such as a tornado, hurricane, flood, earthquake, or fire).

⁴¹ This does not mean that they are regularly prone to failure, as they are not. We mention this just to explain how we prepared for a failure.

latest generation. However, because there may be single points of failure within a tape library (say in the cartridge picking or transport mechanisms), we have specified same-day response times on tape library maintenance, with its higher cost than next day service.

Longevity of the Solutions

The useful life of tape library components (frames that hold the cartridges and tape drives) tends to be very different than the useful life for disk-based NAS solutions. This significantly influences how long the storage infrastructure assets will be retained.

Tape Library Solutions

Tape library storage frames (where the cartridges are stored) tend to last for a very long time (measured in decades, typically) and physical cartridge formats tend not to change, allowing many generations of cartridges to occupy the same slots. The frames that hold the tape drives⁴² also last for a long time (again, measured in decades, typically⁴³).

However, tape drives tend to have a shorter *useful economic life*, not because of operational (physical) durability (because they can be maintained for decades), but by a policy-driven need to limit how many older generations of drives are retained and maintained (to go along with retention of older generations of tape cartridges). By and large, tape drives can read back two or three generations and write back one generation. So the mix of drives in the library will depend on how long you keep older cartridges, how many there are, and how often they are read.

For example, if the density of cartridges (the amount of data that each can hold) is doubling every generation (not out of the realm of possibilities) then, after three generations, the newest cartridge will hold eight times as much as the $n-3$ generation cartridge. Thus, if one were to copy the contents from the $n-3$ cartridges to a new cartridge (a manageable task for the LTFS server to handle automatically as a background task), one new cartridge could replace eight of the older cartridges and seven of eight slots holding the old cartridges would be reclaimed. The slot-related savings are economically significant, as the reclaimed slots would reduce the need to procure more frames to hold the new incoming data plus the data center floor space associated with those new frames. Add to this the savings by not

having to maintain as many older drives (probably under maintenance contracts) and by reusing the old drive space to house fewer, newer drives, and you begin to understand why tape replacement policies are economically important.⁴⁴

Accordingly, we presume that tape cartridges have a practical life of three generations and the files on 9-year-old media will be copied to new media when the $n+3$ generation of drives is available.⁴⁵ Because the 2015 Archiving TCO Study period is nine years and because we presume that a new tape generation is acquired every three years, the conversion of old tape cartridges (e.g., LTO-6) to new ones (e.g., LTO-10) falls outside the boundaries of this study. Additionally, we presume that maintenance is not paid for older drives retained in the library (assuming that there is space to hold them without buying an additional frame). The use of these drives is not a determinant in our computation of the number of new drives required for each cycle; however, these older drives are readily available for reading older cartridges.⁴⁶

One might question whether buying the next generation of tape drives (every three years or so) is a good policy and a practical solution. Given that our requirements resulted in a very high

⁴⁴ We've done the TCO calculations for this every-fourth-generation copy-and-replace strategy numerous times, and it always comes out as a cost-savings approach, even if more of the latest generation of drives is required to do the copying. Some data centers now choose to replace older cartridges after two generations. A lot depends on how busy your tape drives are and whether there are any hours of each week where demand slacks off. In this study's TCO model, in order to calculate how many drives will be needed in a cycle, we limit writing to the latest generation of drives to 10 hours a day for 50 weeks of the year, leaving the rest of the hours for retrieving the files. Do recall that there is an extra latest-generation tape drive in each library (which is not considered in the computation of how many drives will be needed). You have to look carefully at how long it will take to copy from the old cartridges to the latest and whether additional drives (and, possibly, the frames to hold them) will be required.

⁴⁵ For example, files stored on LTO-6 cartridges would be copied to LTO-9 cartridges, when the latter is available. Thus, there would be no reason of compatibility to have anything but LTO-9 drives, because they can read back two generations, i.e., both LTO-7 and LTO-8. Because data is not being deleted, we are only concerned about reading older cartridges, since new data will be written only to the latest generation of cartridges. However, in our model, we keep older tape drives installed (but not maintained) as long as we have space to house them. This gives a lot more reading capacity to meet unexpected access needs.

⁴⁶ We didn't include the extra maintenance costs to maintain consistency with our earlier studies. However, if maintenance costs for older tape drives had been included, there wouldn't have been a significant increase in the TCO, as the bulk of the tape library solution's cost for satisfying our archiving scenario is for the frames (and slots) and for cartridges.

⁴² Many drive-holding units also store cartridges.

⁴³ Manufacturers typically claim a 30-year life for the media.

slots-to-drives ratio, we were much more focused on reducing the number of slots and cartridges than on the cost of acquiring new drives. Thus, we deemed that it was beneficial to always be writing new archive files to the latest generation of drives and cartridges. Our policy solution was to buy one more new drive (in each cycle) than we determined was necessary to write the new data in 10 hours per day; this is the “spare” mentioned previously.⁴⁷ In the second and third cycles, we retained the older generation(s) of drives as long as there was room for them (i.e., without buying another expansion frame to hold those that wouldn't fit). These older drives were held as long as they worked (i.e., without paying for a maintenance contract).

Disk-Based Solutions

Disks are different, primarily because they are rapidly-spinning, always-on electro-mechanical devices. In the past, it seemed reasonable to assume that an entire disk solution would be replaced every three years, because the costs of out-of-warranty fourth and later years of maintenance were prohibitively expensive, especially in light of the new capacities (within generally the same energy and floor-space envelopes) being offered by the next generation. However, in this study we were more open-minded and asked what it would cost for maintenance of disk systems in years four through six. To our surprise, for some solutions it made sense to keep a disk solution for six years.

Thus, there are some disk solutions in our sample that are replaced every three years and others replaced every six years. Do note, that the impact of this on the TCO is less than you might casually think, because most of the data growth comes in the second half of the nine-years of the study (i.e., the bulk of the data growth is back-loaded, due to the effects of compounding), as shown in Exhibit 2 on page 7.

Which Products to Include

Because this is about the cost of different tiers of storage, we relied primarily on the tape library vendors to also provide their best (most cost-effective) NAS solution, as all of them are in the business of providing many tiers of storage, including the two that are the focus of this study. *Might you pay less for a NAS solution from*

non-library vendors? Quite possibly, but likely not by enough to upset the conclusions in this report (i.e., low-cost disk solutions will not become less expensive than the most efficient tape library solutions and, most likely, any of the tape library solutions included in this study).

A Lot of Variety

The storage solutions, both disk and tape, included a lot of variety, from the types of hard disks used to the density of tape storage frames.⁴⁸ Whenever there was a solution component that would offer a lower TCO, we chose that component. There is an exception. All tape library vendors that offer enterprise tape solutions also offer LTO solutions. We kept both in the average, even when one cost less than the other, as each is a viable offering capable of satisfying the requirements. Additionally, some vendors offered multiple library solutions, i.e., more than one tape library platform. As long as the platform met the requirements, we considered it a viable alternative and included that in the average TCO.

However, there is a wide range of TCO for both disk-based and tape library solutions. For example, if you compare the least-efficient disk-based solution (with the highest TCO of those included in the study) to the most-efficient tape library solution (with the lowest TCO), the TCO ratio rises to about 9.5:1. Conversely, if you compare the most-efficient disk-based solution (included the study) to the least-efficient tape library solution (included in the study), the TCO ratio would narrow to about 3.5:1.

The bottom line here is that there is a lot of diversity from which to choose and you need to use your own data center policies, perspectives, and priorities to find the ones that are right for you.

Floor Space is a Wildcard

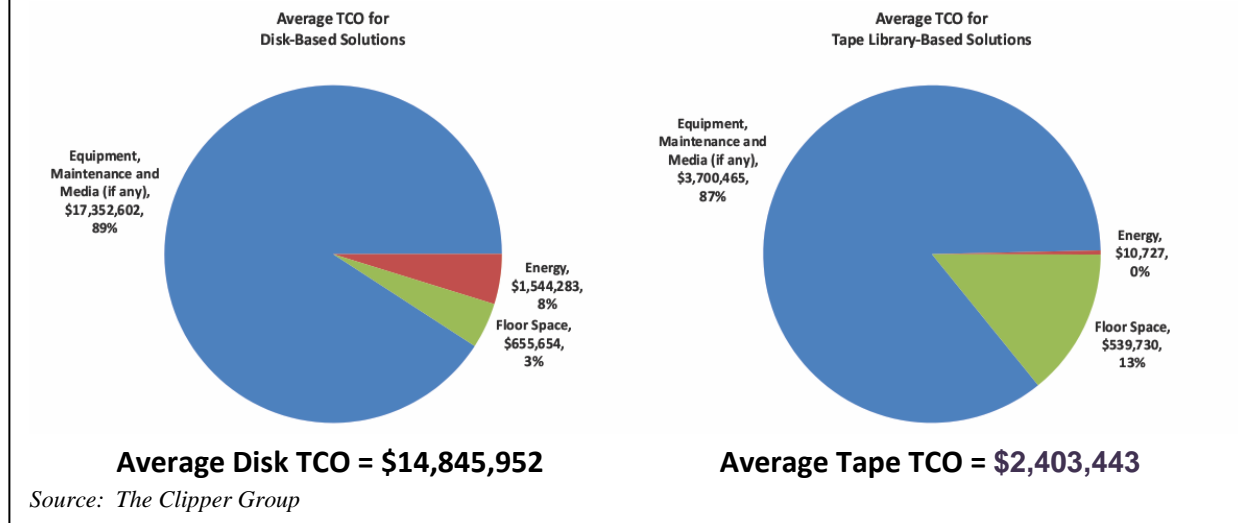
For all storage solutions, we included and accounted for sufficient aisle space in the front and back to enable the opening and closing of all doors and drawers of each of the configured storage solutions and to allow for the movement of data center staff. Typically, this was about 36", but it did vary somewhat. In general, we calculated the square footage required by each configuration including added space for the aisle in front of the solution.⁴⁹

⁴⁷ It was assumed that the latest generation of drives would be available for accessing stored files during the remaining 14 hours a day. Also available for reading was the latest-generation spare drive plus all of the older drives. We assumed a 50 week year, to allow for down time.

⁴⁸ Think about number of cartridges held per square foot of floor space.

⁴⁹ Rear access, if needed, was provided by the aisle behind the configured solution but was not counted in the total square feet per solution, to avoid double counting. Almost all of the

Exhibit 4 —Comparing the Distribution of Average Costs over 9 Years for Disk and Tape Used for Archiving



We have seen data center floor space costs ranging from 25 cents per square foot per month to over \$200 per month. If you have just built a new data center, your cost per square foot might be very high. If you signed a lease a while ago (before the current glut), your cost per square foot still might be high. If you go shopping around today, your cost per square foot might be much lower. If you currently have all of your archived data on older disks (4 TB and smaller), you likely will be able to free up some space with higher density drives (6 TB and 8 TB) and with a tape library, and thus you may now have extra space available.

We used a rate of \$20 per month per square foot, which is somewhat arbitrary. **What is more important is the ratio of floor space costs between disk-based solutions and tape library solutions, which is about 1.92** (i.e., the average disk-based solution uses about 92% more floor space than the average tape-library solution).

However, even if the cost of floor space was zero, disk solutions would still cost more than a tape library solution (even when comparing 8 TB disk solutions against LTO-6 solutions, which required significantly more floor space than enterprise tape solutions).⁵⁰ Of all of the variables in our study, this is the most notable wildcard.

configurations only required access only from the front, although some tape libraries needed side access to the units on each end of a string. We did not account for this left/right access space for the impact of doing so was minimal.

⁵⁰ If you are constrained for data center space, you should look closely at enterprise tape solutions.

Energy Consumption

Energy consumption is a major cost factor for disk-based solutions, but not tape library solutions, as can be seen in Exhibit 4, above.⁵¹ We presume that disks are spinning all of the time and do the same for the latest generation of tape drives (including the spare).⁵² We did not include the energy for the operation of the older drives that had been retained, but this would be minimal, because even if the energy cost for tape was doubled, it still wouldn't be a noticeable percentage of the TCO.

We computed the kilowatts per hour (KWH) consumed when operating all of the hardware (but not at startup, which can be very high for disks) and doubled that to account for cooling. We used the average cost per KWH for the Mid-Atlantic, Northeast and Pacific Contiguous of 14.1¢ (\$0.141).⁵³ This is the same statistic and

⁵¹ The energy cost for tape-based solutions is so low that it doesn't round to 1% and thus is shown in Exhibit 4 (above) as zero percent.

⁵² In comparison to one-to-several-thousands of always spinning disk drives, only a small number of tape drives (a dozen or less) are required to meet our hypothetical requirements, which are more focused on writing data than on retrieving it. In addition, a disk drive tends to consume more energy than a tape drive (because the configured disk drives tend to be rotating at 7200 RPM) and each of the few tape drives requires a small amount of electricity to operate (in comparison to a plethora of disk drives). That is why the energy requirements for a tape library are minimal, even when including the required LTFS servers, which we did.

⁵³ Taken from U.S. Energy Administration's **Electric Power Monthly - Table 5.6.A: Average Retail Price of Electricity to Ultimate Customers by End-Use Sector by State, March 2015 and 2014 (Cents per Kilowatt-hour)**. See

source that was used in the 2013 Archiving Study. In reality, the cost per KWH is only marginally relevant, except for determining energy's share of the TCO. What is most relevant is the comparison of the number of megawatt hours (MWHs)⁵⁴ consumed over the nine-year study period by disk and tape.

- **Average Total MWHs for Disk..... 5053**
- **Average Total MWHs for Tape..... 67**

Thus, the average cost of energy for disk solutions is about 76 times the average consumption of tape solutions.

The Similarity of the Pie Charts

It is worth noting the TCO expense category distributions displayed in the pie charts in Exhibit 4 on the previous page. The Equipment, Maintenance and Media pie slices are very similarly shaped for both disk-based solutions and tape library solutions. The distribution percentages differ significantly for energy and floor space, but the sum of the two percentages is about the same. Of course, these pie charts represent significantly differing TCO expenditures, with disk-based solutions costing more than six times that of tape library solutions.

The Pace of Technology

As discussed previously, new technologies are introduced at their own pace, without consideration for other vendor schedules. For unique (proprietary) technologies, this often results in a game of leapfrog (rather than near-simultaneous happenings).⁵⁵ As analysts, we sit around wondering:

- *When will LTO-7 be available? What about LTO-8? Will the published LTO Program road map be met or surpassed? How far and how fast will the enterprise tapes grow in capacity and write speeds?*
- *When will 10 TB or 12 TB disks attain general availability, and which will be the "standard" capacity for the next generation of high-*

http://www.eia.gov/electricity/monthly/epm_table_grapher.cf?m?t=epmt_5_6_a.

⁵⁴ A megawatt is a thousand kilowatts; this is a rate of consumption. Thus, you need to know how many hours it is running to determine its total consumption, which is why electricity consumption is expressed in KWH or MWHs. As a point of reference, a typical server might consume 2 kilowatts per hour when it is running.

⁵⁵ For LTO, there is no game of leapfrog, as all vendors are using drives and cartridges designed to the same generational specification. However, product announcements and general availability dates will vary, but not by much, as each tape library vendor wants to have its solutions ready soon after the next generation is announced by the LTO Program.

capacity disks used for archiving? How about the generation after that?

We looked at what storage solutions were available early in 2015 and used this definition of "reality" as the basis for meeting the needs of the first three years. However, for later cycles, we estimated the capacities of storage devices and their performance (if that was a sizing factor). By and large, we have been "somewhat conservative", which is a continuing goal. In the end, we made our best judgments. Our presumptions are listed below. However, you might want to make different assumptions.

So what did we use for our present day reality and how did we extrapolate that into the future?

Cycle 1

- **For LTO tape**, we started with LTO-6 (with 2.5 TBs of native capacity), which became generally available early in 2013.
- **For enterprise tape**, we used IBM TS1150 cartridges and drives (with 10 TBs of native capacity) and Oracle T10000D cartridges and drives (with 8.5 TBs of native capacity).
- **For disk**, we used whatever hard disks were in the solutions proposed to us by the participating storage vendors. As a result, both 6 TB and 8 TB drives-based solutions are in our study's mix of the "best solutions". These capacities are the raw capacities of these disks.

You can begin to see some of the cyclical dilemmas that we faced by doing this study now (and not, say, early in 2016). LTO-6 is getting long in the tooth and a significant capacity upgrade to LTO-7 is expected to be available very early next year. (See Exhibit 5, at the top of the next page, for a hypothetical discussion regarding LTO-7.)

Both enterprise tape formats made a big leap in the last two years, so they might be described as being in mid-cycle. 4 TB disk drives have been around for many years and also seem long in the tooth. On the other hand, 8 TB disk drives are young in their life cycle and not very many vendors are shipping 8 TB disk solutions (but we think they soon will). Of course, it is the price per TB of capacity that is most important to us, followed by the amounts of energy required and the floor space consumed. The price per TB is always higher at the beginning of a generational cycle and tapers off as more competitors appear.

Cycles 2 and 3

For Cycles 2 and 3, we made the following presumptions for capacities.

Exhibit 5 — The Hypothetical Impact of Using LTO-7 in Cycle 1

Hypothetically speaking, if LTO-7 had been available on 1/1/2015, Cycle 1 would require 2.56 times fewer slots because of the expected increased native capacity of 6.4 TBs in each cartridge. With fewer slots, fewer frames would be required (certainly over time), resulting in less floor space being consumed. While the cost of these new LTO-7 tape drives may have been higher than for LTO-6 tape drives, the overall TCO of the tape library in cycle 1 would have decreased significantly. More importantly, this may have made LTO-8 available for some portion of Cycle 2 and LTO-9 available for some portion of Cycle 3, significantly reducing the TCO of the second half of the 9-year period, when the largest increase in data growth would occur.

What would the net result be? This could extend tape's TCO advantage from about 6:1 to 8:1, or possibly much more, depending upon mostly on the pricing for the LTO-7 cartridges, which will be higher than we used in our model because cartridges always cost more when announced. Unfortunately, our TCO model with three three-year cycles was fixed to begin on January 1, 2015. Your requirements are different, no doubt. **In addition, your procurements can be more flexible, even if it means adding new drives earlier than originally planned (say, moving up to LTO-7 in 2016).** If your data is growing annually at 50% or higher, the ROI of accelerating technology would surely be in your favor, as drives are relatively inexpensive (as a portion of the TCO).

Source: The Clipper Group

- **For LTO**, we presumed LTO-7 (with 6.4 TBs of native capacity) for Cycle 2 and LTO-8 (12.8 TBs) for Cycle 3. These were taken from the LTO Program road map (cited earlier).
- **For both varieties of enterprise tape**, we presumed the same capacities for Cycles 2 and 3: 20 TBs and 35 TBs, respectively. Based on previous demonstrations and statements, we think that these are achievable and little is gained by trying to handle these two separately. No doubt, the game of leapfrog will continue between the two competing enterprise tape drive vendors.
- **For disks**, we presumed 12 TBs per drive for Cycle 2 (basically a doubling of the very prevalent 6TB drives) and 20 TBs for Cycle 3 (a two-thirds increase over the 12 TB drives). While the latter may be a stretch goal requiring further technological breakthroughs, we feel that disks will have to be close to 20 TBs by 2021 in order to be viable.

Some Factors Were Ignored

Some factors were intentionally omitted from the study, such as software management and personnel costs, as they were deemed constants in that they applied to both technologies, equally. **In this study (and the earlier two), we focused on the direct costs of the storage infrastructure.**

Additionally, we presumed that there was no time value for money. With interest rates currently being so low, this is close to being true. Making this presumption (as we also did in the

previous two studies) greatly simplified our TCO analyses without significantly affecting the key ratios that we determined.⁵⁶

Data Protection Presumptions

Of course, there is more that you need to consider, even though we chose not to. *What needs to be done to mitigate the potential for a site-wide disaster or forced shutdown?* There are three major scenarios from which to choose, plus a plethora of variations.

1. **A completely duplicated disaster recovery site that sits far enough away to provide services (in short order) after the primary site fails.** In this case, the archiving appliance in the primary location would be linked to, and synchronized with, another archiving appliance at the remote location. There are three archival storage possibilities.
 - a) Tape is the archiving storage medium in the primary location and tape also is used at the remote site.
 - b) Disk is the archiving storage medium in the primary location and disk also is used at the remote site.
 - c) Disk is the archiving storage medium in the primary location but tape is used at the remote site.

The key characteristic of all of these alternatives is that you need a replication solution to keep the remote site up-to-date. If you use both disk and tape solutions for your archived

⁵⁶ Had we included the time value of money, the burden would have fallen far heavier onto disk-based solutions, because they cost about six times the tape library solutions.

data, then you have many more combinations to consider.

- 2. A peer-to-peer back up strategy is in place for the operating storage and the archival storage.** In this case, each storage solution is responsible for replicating itself to the remote location, although this also might be achieved by writing to the local and remote devices concurrently (i.e., this is a networking solution that also delivers replication).
- 3. A traditional backup/recovery solution is deployed on the operating storage and the archival storage,** with the backup data stored remotely.

In this long-term archiving study, we focused on the scenarios listed in the first alternative above (a completely-duplicated disaster recovery site). We do this because we are most interested in comparing the alternatives economically and Scenario #1 allows us to do that simply.

However, you might want to use different mixes (percentages) of disk and tape at, say, your two data centers (depending, of course, on your requirements). For example, if your solution mix at your primary site was 20% of the data on disk and 80% on tape, you might choose to put more or all of it on tape at your secondary site. This would be mostly an economic decision weighted by the probability of needing to recover from a disaster and the time allowed to do that.

Because of the many possible remote site configuration possibilities and policies, we did not include any costs for the remote site in the results presented in this report. However, you will need to do so. You should be able to use what is presented herein to calculate the additional costs for your required remote site (based on the mix of disk and tape that you will determine).

More Details on Our Procurement Methodology

A Customized Building-Block Approach

Our primary methodology was to identify, configure, and price the necessary storage infrastructure components required to satisfy the many requirements that have been described. Clearly, different building blocks were required for tape libraries and disk-based solutions. When possible, we tried to use the same approach to scaling a system across similar products (i.e., tape libraries and disk-based solutions). While the same rules of the procurement applied to both, the ways that they scaled up inherently are different.

In general, tape libraries scale by adding more cartridge-storing frames to hold the cartridges

required in each cycle. As noted earlier, all of the tape drives that our TCO model required fit into the base cabinets (the first building block) of each tape library. Had more tape drives been required, we would have needed to add frames that can hold the drives. Thus, frames and cartridges are added as needed, once certain minimum configurations (if any) have been satisfied.

Disk-based solutions tend to scale in one of two ways.

- Some scale out by adding components until a maximum capacity is reached. While these may look like scale up solutions (especially those delivered in one or more racks), these days they tend to be built with scale-out modules.
- Others scale out by adding additional (usually identical) modules to a cluster. Many have no real limit on the number of modules that can be aggregated.

We custom-configured each solution properly to reach the capacity required for the cycle. If there was a limit to the capacity, we procured in that maximum capacity, buying as many of these maximum configurations as needed. Other wise, we built up solutions by adding additional modules. In all cases (for both tape and disk solutions), we adjusted for the excess capacities procured, as described below.

We presumed that the costs for storage infrastructure components and maintenance for Cycle 2 and Cycle 3 will be similar to the costs for Cycle 1, but with proportionally greater capacity, using the capacities for tape drives/cartridges and for disks that were described earlier.

Procurements at the Beginning of Each Cycle

All purchases for a cycle are presumed to be made on January 1 of the first year of each cycle. This simplifies the modeling and doesn't affect the TCO, as we have presumed that the cost of money is zero. Thus, it doesn't matter when in the cycle equipment is procured, except for incremental software, maintenance, energy and floor space costs. Because we treated disk and tape solutions the same (in terms of procurement date), the impact of this presumption is very small.

Limited Effects of Over-Procurement

Using the methods just described, we were able to grow our disk-based and tape library solutions to meet the capacity requirements of our TCO model, by one means or another. While this level of aggregation almost always will result in

over-procurement within a cycle, we have mitigated that in the following ways.

- For tape solutions, what we over-procured was slots.⁵⁷ We paid for use of all slots when the containing frame was added. However, we didn't buy the cartridges for the slots until they were needed. This mitigates most of the excess cost, as all procured-but-unused slots would be consumed in the following cycle.
- For disk solutions, what we over-procured is usable capacity. Since only a fraction of the last-procured unit(s) will be used in each cycle, we removed the same share of the costs attributable to that unused capacity (on the assumption that some other application will use the excess storage). Thus, the unused space is fully mitigated for disk solutions.

The Concept of “Petabyte Years”

Our starting point of 1PB of data will grow to about 52PBs by the end of the ninth year, but that is only part of the story. When you bring time into the equation, over the nine years there are about 143 “PB Years” of data being stored and paid for.

What is a “PB Year”? This is the number of PBs of data being held each year. In simple terms, data added to the archive in year 1 is actually being held for 9 years, while data added in year 9 is only being held for 1 year.⁵⁸ Thus, there are more PB Years of data than actual data, because PB Years account for duration held as well as capacity. We have chosen not to focus on PB Years, but you may want to consider this as a valid parameter for analysis. Please note that the growth of data in our model is back-loaded into the later years, which have reduced costs as higher density solutions are being used in those years.

Other Issues

Some disk solutions are delivered as a set of server-like components. We asked the vendor to supply these in racks and to include the cost of the racks. Some did; others didn't. For those that didn't, we selected a rack from a reputable

supplier and added its retail cost to TCO for disks. In terms of the TCO, this was a negligible amount.

Conclusion

Were we surprised by the results of this study? At first glance, most definitely, especially by the significant narrowing of the TCO ratio between disk-based solutions and tape library solutions from our 2013 Archiving TCO Study. Upon deeper inspection, however, it all makes sense, as we have explained throughout this paper.

Nonetheless, tape library solutions still have a significant economic advantage over disk-based solutions on a cost per terabyte stored basis – if you can tolerate retrieval speeds measured in seconds to a few minutes instead of a few seconds or less. This is the important conclusion for you to take away from reading about our study. As we said at the outset, we expect that most enterprises will use both disk and tape for archiving. Now you know why.



⁵⁷ For some tape libraries that have a minimum configuration, we procured that minimum even though it resulted in a significant excess of unused slots in Cycle 1. This had no real long-term economic impact, because we assumed that the time value of money was zero, so it didn't make a difference when looking at the total costs for the entire nine-year period.

⁵⁸ If you were renting this space, this is how you would pay for it.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twenty-third year, is an independent publishing and consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 40 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 a Senior Contributing Analyst for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, 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 co-founder of The Clipper Group. Mr. Kahn is a veteran of the computer industry, having spent more than four decades working on information technology, spending the last 26 years as an analyst including 22 years 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; at Power Factor Corporation, a Boston-based electronics start-up – President; at Honeywell Bull – Director of International Marketing and Support; at Honeywell Information Systems – Director of Marketing and Director of Strategy, Technology and Research; at Arthur D. Little, Inc. – a consultant specializing in database management systems and information resource management; and at 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 Mike.Kahn@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

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, The Clipper Group Calculator, and "*clipper.com*" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "*Navigating Information Technology Horizons*", and "*teraproductivity*" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

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.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

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.