The Clipper Group Calculator™

VENDOR-AGNOSTIC TCO STUDIES FOR THE INFORMATION TECHNOLOGY BUYER

Navigating Information Technology Horizons℠

New Publication in 2013 Report #TCG2013009LU May 13, 2013

Revisiting the Search for Long-Term Storage — A TCO Analysis of Tape and Disk

Analysts: David Reine and Mike Kahn

Management Summary

The Clipper Group, a 20-year old computer industry analyst firm, independently funded **a total-cost-of-ownership (TCO) study to compare the costs of backend storage on high-capacity SATA/SAS disks and LTO tape for holding archived data.** This is the fourth iteration of this TCO study and is in many ways an update to the one done in 2010. **The study's goal was to determine the relative economic relationship of the cost of storing archived data on disk versus tape in a large enterprise over a forward-looking nine-year period**, which was long enough to reach long-term conclusions. **We found that the average cost for a disk-based backend storage solution costs about 26 times as much as the average cost of a tape-based solution using an automated tape library (ATL).** This is not a declaration that tape is better than disk, as each serves a valuable purpose in archiving and you probably need a mixture of the two. **The practical question is *"How much of your enterprise's archived data belongs on tape?"*** <u>There is no right answer.</u> As you can see from the summary box above, **this is a multi-million dollar question, with a very real potential for significant savings.**

Major Study Findings at a Glance

For long-term archiving of many petabytes of digital data growing from 1PB to more than 28 PBs over 9 years:

(1) The TCO (including equipment, media, maintenance, energy and floor space) of the average disk-based solution costs **26** times the TCO of the average tape-based solution.
- Tape costs about $1.5M while disk costs about $38.5M

(2) The cost of energy alone for the average disk-based solution exceeds the entire TCO for the average tape-based solution.
- Energy for tape costs about $47K while disk costs about $4.9M, about **105** times

(3) Disk required about **4** times the floor space of tape.

Thus, the more data preserved on tape, the lower the overall TCO.
- With 50% on tape, the TCO is reduced by **48%**
- With 90% on tape, the TCO is reduced by **87%**

You probably need both disk and tape, in an appropriate mix for your business. Read this report to find out why.

Source: The Clipper Group

The data was presumed to be large binary files (like medical images or video) and already stored in a compressed form, i.e., it was presumed that the data was unable to be compressed further. The initial amount of data was 1 petabyte (PB), which grew at a rate of 45% per year. At the end of the 9-year study period, this would total more than 28 PBs. Data was presumed to be so valuable or covered by legal requirements for retention that none was deleted during the study period. Many disk and tape vendors provided confidential pricing and configuration data that was combined with publicly-available information. All comparisons of equipment and maintenance were done using list prices. Maintenance costs were normalized to adjust for varying warranty coverages. Equipment upgrade and replacement decisions focused on meeting the requirements at the lowest reasonable cost. Many presumptions were made and conclusions drawn. They are described and discussed in detail. Please read on.

IN THIS ISSUE

Introduction to the TCO Study

For many, the family budget keeps getting squeezed every month, while family expenses remain constant or even rise. You have to pay your monthly mortgage, or rent, your utilities for power and water, and, of course, the weekly trip to the supermarket for the family's food. You feel that you have to cut back somewhere, *but where?* The monthly bill for heating your home keeps going up as the temperature goes down. Great, with summer approaching, the temperature will go up, but when it does, you may need to turn on the A/C to keep cool, possibly consuming even more energy. One solution might be to turn down the thermostat in the heating season to burn less oil or gas, and turn it up in the summer to conserve electricity. Unfortunately, that may result in everyone being too cold, or too hot, and once done, that new level of energy consumption becomes the "new normal", while income and expenses continue to work against you.

Another way to reduce expenses might be to reduce the amount of fuel that you are pouring into the family car. If the price of gas isn't going to come down significantly in the foreseeable future, then you may need to use less of it. Unfortunately, this may mean a reduction in leisure travel as there is not much you can do about commuting expenses, unless you work from home or switch to public transportation, which also can be expensive. One means to consume less gas would be to acquire a hybrid automobile. Unfortunately, this may mean a healthy hit to your budget for acquisition (even for a previously-owned vehicle), even if it does reduce your future operational expenses. **Bottom line – you need to look at the total cost of ownership to determine your next move.**

Concern for the budget is not unique to the home environment, as similar issues plague the CFO and CIO of every enterprise. Unless your enterprise's management has bestowed an unlimited budget for the data center, the IT staff is faced, on a daily basis, with trying to do more with less. In order to stay competitive, that staff must satisfy the compute performance and storage capacity requirements of a wide variety of users – enterprise management, internal workers, partners, and, especially, customers. Every time that the data center is required to deploy a new application, additional server capacity must be found. Unfortunately, there may be no additional power available. At this point, the IT staff has learned to replace existing server infrastructure with new servers designed to consolidate and virtualize mission- and business-critical applications more efficiently. Regardless, this does not resolve any of the issues associated with storage capacity and its prominent place in the data center budget.

In the past few years, almost every data center has experienced a very high growth of data being collected and parsed for any value that might contribute to enterprise profitability. The storage volume required to save, preserve, and secure it usually exceeds the existing capacities of most enterprises. In fact, the majority of this data may not be accessed frequently, rarely, or perhaps never. This problem is very real but is different than keeping data for a required retention period, say emails or office documents for a seven-year statute-of-limitation period (in the U.S.). In many cases, such data is retained out of fear, because it has to be kept to meet statutory or other requirements. Thus, retention is seen as a liability (and retained as a reaction to potentially severe penalties) that must be managed to limit exposure, as opposed to retention of an asset so valued that it must be preserved for a long time. There are many digital assets that might need to be kept for a long time.

We started this study thinking about data that has a very long useful life. In discussions with enterprises and storage vendors, retention periods of both 50 and 100 years were 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.

A good example is medical records, especially high-resolution medical images. Depending on the jurisdiction, medical records may need to be retained for as long as the life of the patient plus several decades, long enough to be considered *forever*.

This report is not about whether disk costs more than tape, or not; it is about having the right mix of disk and tape to access and preserve valuable digital data for a long time, taking advantage of the strengths of both, i.e., the low-cost, high-capacity of tape and the rapid response of disk.

The time period requirement for preservation usually ranges between a decade and a lifetime,

Publisher's Note – *The Clipper Group Calculator* is our new publication focusing on TCO studies, which we deemed worthy of its own series. This is the first issue.

and possibly forever. The center of our focus is on the compounding quantities of commercial and institutional digital data, which are deemed so important that someone in the enterprise says: ***We need to keep this forever!*** This usually means for longer than can be imagined in normal, 21st-Century human terms, where we have trouble thinking about the short term (measured in weeks and months), much less several years or decades into the future.

From a practical point of view, anything being kept for more than a couple of years, needs to be considered in the same way as items being kept longer (maybe even forever), because the required procedures for handling five years of retention, or twenty, may be very similar or identical. Certainly, the goals are very similar, if not identical. You want to be able to get to the data that you want in short order, measured in seconds to at most several minutes.[1]

In the data center, we make decisions in the present, with a hope that our decisions are sufficiently good and durable for a longer period, but still in the near future, say a few years. *Why only a few years?* Because the pace of technology innovation and related economic changes requires us to reconsider the solutions deployed every few (think three-to-five) years. This is especially true for data that might be growing at 50% to 100% per year. Continuing into the future with what now may be a less-effective and/or less-efficient solution makes little sense, with that much data being added each year to the archive. **Thus, you need to figure out what is the best and sufficiently-durable solution based upon what is available today.** Only then can you decide whether to change what you have been doing, either from that point going forward, or also encompassing what was previously preserved. ***The question is how to do that, reasonably and in the most cost-effective manner, thinking not just about what to do now and for the next few years, but preparing for the ensuing decades of continuing data growth, technology change, and increasing long-term preservation requirements.***

This report describes the results of a major Total Cost of Ownership (TCO) study funded independently by The Clipper Group, Inc., the publisher of this report, using cost and operating data provided confidentially by many storage vendors. It focuses on the underlying media on which data is kept "for now and for the foreseeable future" (i.e., forever), which we see as being at least a decade. We have designed it to be a generic study, one that is not applicable to just a single problem area or industry, but to any business or organization needing to preserve petabytes[2] of data, especially digital data like images, videos, seismic recordings, astronomical data streams, etc. This is the fourth generation of this study and builds heavily on the modeling done in our last report in 2010.[3] In that study, we made many presumptions, most of which remain, although some have been refined or extended. In this 2013 report, we added several additional disk and tape vendors, some new parameters, and did some additional analyses. While the two reports have some comparable conclusions, it is important that the reader understand that there have been changes and improvements that limit direct comparison of some of the microdata in the models generally incomparable, at least without a long explanation.

This study focuses on <u>disk *and* tape</u>, <u>not disk *or* tape</u>, as the principal storage media for archiving[4], or the long-term preservation of data, taking into consideration the functional requirements of the enterprise and the best attributes of each. We hope to report on other archival storage possibilities, including those in the Cloud, later this year.

Repeating what was said at the outset: few enterprise data centers will archive solely on disk or tape; the answer almost always is a mix of both. *How should you decide what goes where?* There are two general criteria: speed and costs. Most archiving solutions[5] use disk[6] for speed and

---

[1] If your data set is very large (many TBs), this time frame is for when the data will begin to flow to its target. There is no magical potion for moving large volumes of data instantly from rotating media.

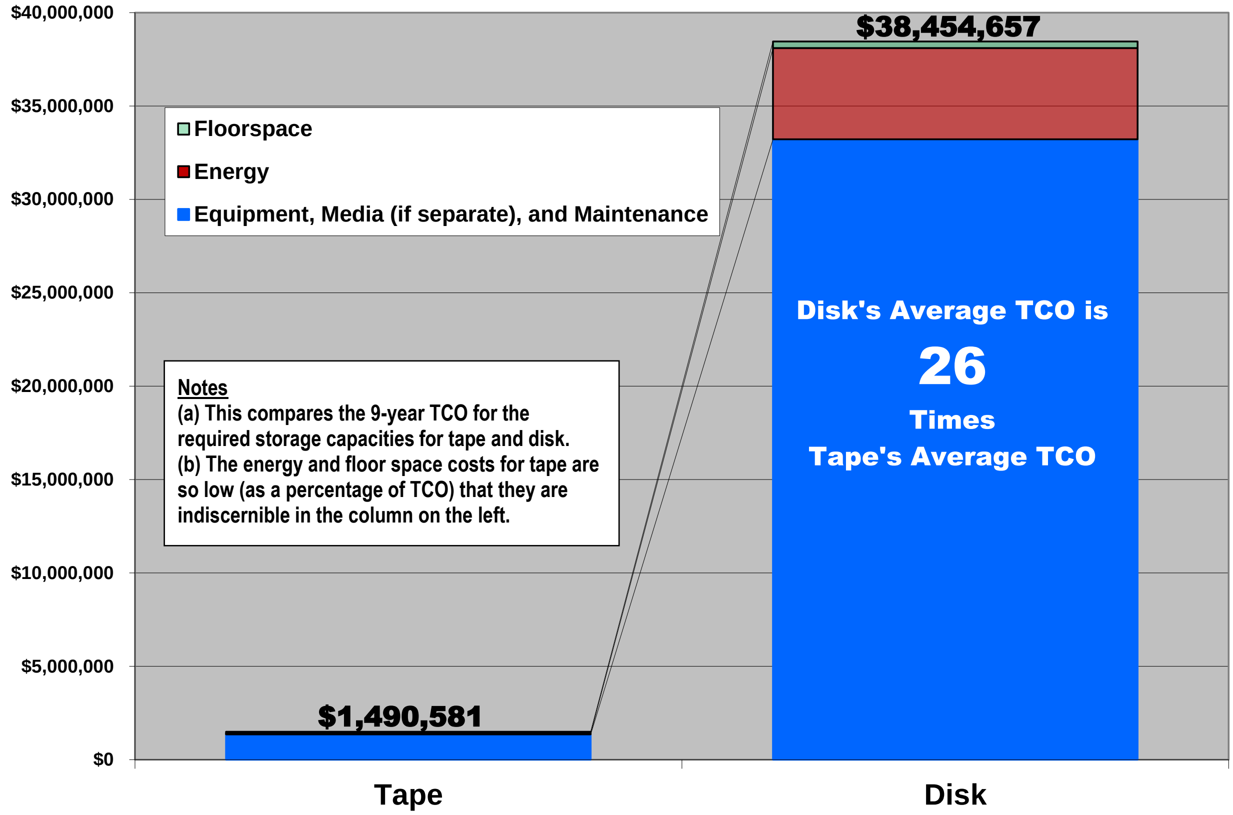
[2] A *petabyte* is 1000 terabytes and a *terabyte* is 1000 gigabytes.

[3] 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.

[4] It is important to understand the difference between *backup* and *archiving*. *Backup* is about saving a copy of stored data (either blocks or files) to a safe place, so that it can be *recovered* (either locally or at a distance), if the need arises, whether due to system problem, software error, virus, human mistake, etc. The focus is on recovering reasonably quickly. To meet this recovery time requirement, today most backups are stored on disks or on virtual tape appliances, so that they can be recovered quickly. ***Archiving*** **is different and is where data is stored for long-term preservation and, usually, only occasional accessed. Depending on the nature of the data and required time to get to the saved data, disk, tape and/or other storage devices may be used.**

[5] An *archiving solution* is an application that is hosted on one or more servers (and may be delivered as an "appliance"). It manages the data to be stored, indexed, protected, and retrieved. The archiving solution typically manages the storage

Exhibit 1 — Comparing 9-Year TCO for Tape to Disk for Archived Data



Source: The Clipper Group

tape for cost savings. If you have data that has a critical component to its "instant availability" then only disks (hard drives or SSDs) are likely to satisfy your needs. A good example of this might be a database of known criminals. Accessing the needed information quickly might have life-or-death consequences to a police officer inquiring about the vehicle just pulled over. However, if your request for data can wait a little longer (think "several seconds" to "a couple of minutes"), then tape probably is your answer.[7]

Knowing what each will cost (say per terabyte (TB) or petabyte (PB)) might help you make a good decision. Consider this imaginary example. If premium gasoline (also called "high octane") costs ten times what regular gasoline costs, might that influence your next vehicle-procurement decision?[8] You most likely would assess very carefully whether you really needed a vehicle that would cost ten times as much to operate. If your family had two or more vehicles, might you consider whether one or more should be "highly efficient", to bring down the average cost per mile of operating your "personal fleet"?

That is akin to what is at the heart of this paper.

- *What is the difference in the cost of storing a lot of data to be held for a long time on disk versus tape?*
- *How might your understanding of the differences in the underlying costs affect your storage procurement decisions? In other words, would a good understanding of how much more disk costs than tape affect the mix of your storage types for archiving?*

If you want a very long, one-sentence summary answer, here it is.

resources that sit behind it. This study is about those storage resources and their TCO.

[6] Disk includes rotating hard drives, which may be accelerated by SSDs in some way. For this study, we consider SSDs to be used only as an application accelerator and not as an archival target, because SSDs cost too much to use for historical data. Thus, this study only considers rotating hard drives as the alternative to tape.

[7] Typically, when cartridges are stored in an ATL, it takes only tens of seconds (or less) to automatically pick (retrieve) a tape cartridge from its slot and then move it to an available drive and mount it. No human is involved. Getting to the right place on the tape additionally might take several-to-many tens of seconds, with the average access time being around 50 seconds.

[8] Right now, in the U.S., high-octane gasoline sells for about a 10% premium.

Exhibit 2 — Clipper's TCO Study Results Summary

TCO by Expense Category for Archived Data Growing from 1PB to More than 28 PBS over 9 Years (3 cycles)	Average 9-Year TCO for TAPE	Average 9-Year TCO for DISK	9-Year TCO RATIO Disk Average to Tape Average
Equipment, Media (if separate), and Maintenance	$1,348,907	$33,221,012	25
Energy	$46,569	$4,874,845	105
Floorspace	$95,106	$358,800	4
Total 9-Year TCO	$1,490,581	$38,454,657	26

Source: The Clipper Group

Using a nine-year TCO model for the long-term data retention of many petabytes of digital data, based upon vendor-supplied list pricing for an archiving application of large binary files with a 45% annual growth rate, the average disk solution (including acquisition costs, maintenance, energy, and floor space) was found to cost more than 26 times what the average tape solution would cost.

(See Exhibit 1, at the top of the previous page.) **As part of this total, the average disk solution was found to consume 105 times more energy than the average tape solution, costing four times more than the entire nine-year TCO for tape.**[9] (See Exhibit 2, above.) If you want more of this story, then read on to learn how we came to these conclusions, including the important presumptions and details that framed our TCO analyses.

Our 2013 Study and the Results

This study focuses on one overarching question: ***When should you use disk and when should you use tape for the long-term storage of digital data, taking into consideration the back-end costs associated with storing it and the features that bestow an advantage to each?*** The bottom-line TCO-driven answer already was disclosed. **However, if the time required to access any of your vast, long-term archived data needs to be a few seconds or less[10], archiving to disk may be a better solution, but it will be much more costly.**[11]

Just like the shock of having to pay ten times as much for premium gasoline (over regular) probably would force you to reconsider your new vehicle requirements, **we think that the results of this study should force you to look at what you store where and why.** The fact that the average disk solution costs 26 times more than the average tape solution should drive you faster than the imaginary premium gasoline example that was fabricated to get your attention, which it likely did with only a ten-times multiplier.

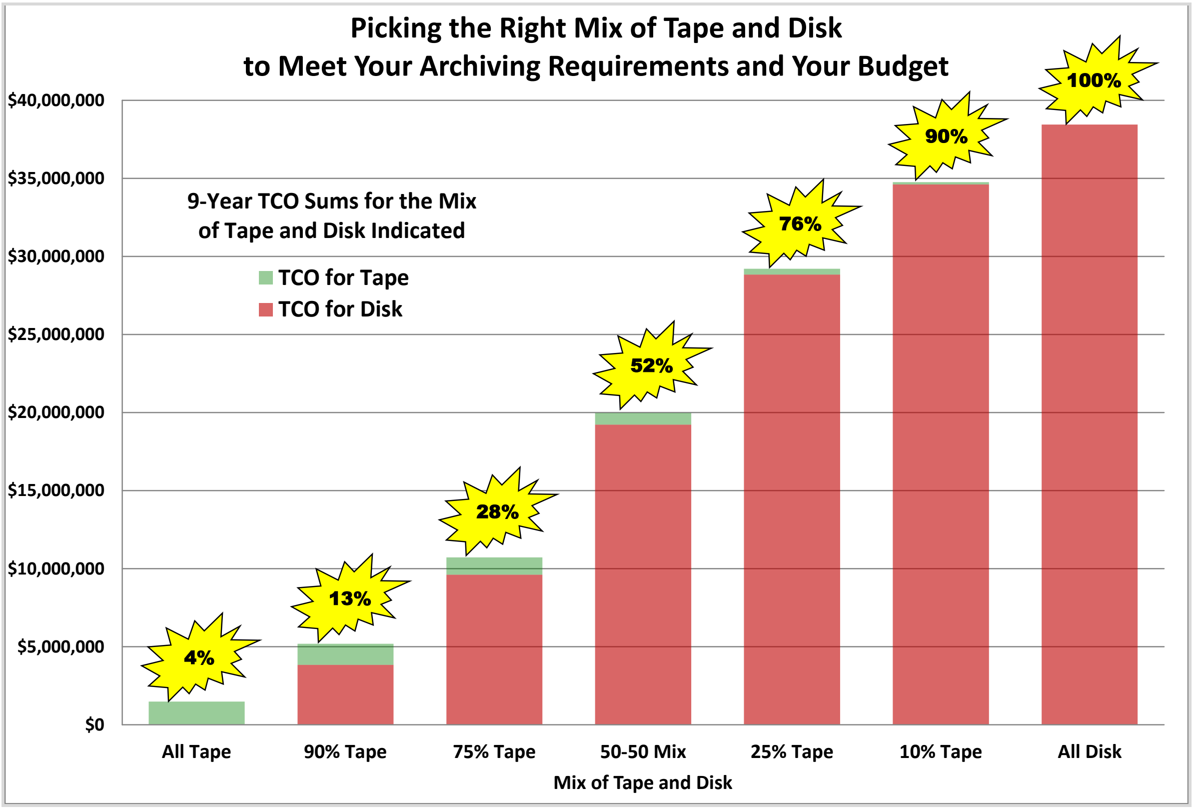
In fact, the most practical and economic solution almost always is a mix of both disk and tape, say with 90% tape and 10% disk. As a rule of thumb, for most uses, the vast majority of archived data should reside on tape, assuming that it is lightly accessed, which we define as no more than 15% of what has been archived is accessed in any year. Exhibit 3, at the top of the next page, shows why. **Simply put, the more tape that you have in your mix (as a percentage of your archive's backend capacity), the less expensive will be your total cost for storing the data.** As Exhibit 3 shows, if you could put 75% of your archived data on tape (with the rest on disk), your TCO is about 28% of the cost of putting it all on disk. With 90% on tape, it costs about 13% of the all-disk solution. *Is this compelling?* We think it is, but it is not an absolute. You need to make a well-reasoned decision based on your requirements. Some of your data

[9] Do note that in Exhibit 1, the energy and floor space costs for tape are visible only as a thin line (at the top of the column bar on the left) because these costs represent a very small percentage of tape's TCO,

[10] The concept of time will be discussed later, as will the caching presumptions.

[11] Nonetheless, do remember that even though there is access latency with tape, once the location of the file on the tape has been reached, the transfer-retrieval time can be very fast. For example, with LTO-6 tape the transfer rate is up to 160MB/second (uncompressed). This fast tape transfer speed can outperform many disk system transfer rates. For large data sets (big files), the real metric may not be how fast you get to the first bytes of data but how long it takes to get all of the data to where it is needed.

Exhibit 3 — The Mix of Disk and Tape Greatly Affects TCO



Note: The percentage shown above each column is the percentage of the TCO for an all-disk solution that can be realized by using tape for the percentage of data shown on the x-axis.

Source: The Clipper Group

belongs on disk. The question is *"How much?"* Only you can answer that.

Regardless, for large quantities of data, tape always is much less expensive than disk and always uses much less energy and floor space, when measured on a per-petabyte basis. Tape should be used whenever its somewhat slower retrieval times are acceptable.

That's the summary, at a high level. However, you may want the details and there are a lot of them – study methodology, data, equipment specifications, and economic presumptions (along with more than a few caveats) that will take many pages to explain. If you want to understand how we did what we did and look at some of our model's components, presumptions, and findings, please read the entire report.

Defining the Storage Challenge

Imagine an archiving process, where there are three rooms that are separated by walls, through which only networking cables pass. The rooms might be separated by many miles or only a few feet. A user sits at his/her desk in what we abstractly call "Room 1" and retrieves data (think about a file, for simplicity) that has been stored by an archiving appliance (in "Room 2") in a prior action. That individual has no idea how or where the data was archived and only can judge the success of the archiving solution by three very important criteria.

(1) Whether it stored what it had been given,

(2) Whether the user gets back exactly what was stored previously, and

(3) Whether this happens in a timely manner.

The archiving appliance manages all that is necessary to store the data and accomplish these three criteria. Vendor and brand of the archiving appliance are not important to this study, but certain generalizations can be made about its characteristics.

The archiving appliance typically has its own operating storage, an *archival cache*, traditionally stored on rotating hard disks. Depending on the requirements, it may hold data for days, weeks, months, or even years, in order to meet the users' service level requirements. Eventually, based on some criteria, the data usually needs to be archived to the least expensive storage available, for possible later use or for legal retention requirements or other reasons. This data is sufficiently valuable that it must be preserved for a long time, beyond the study period.

To preserve it, the appliance moves it to the archival storage (low-cost disk or tape, or a blend) in "Room 3" (near-by or farther away). Remember, the users of the appliance can only judge the adequacy of the archival storage in terms of (1) whether it stores what it is given, (2) whether it is able to retrieve it without loss, and (3) whether this happens in a timely manner.

Let's presume that the archiving appliance is dependable, of high quality, and, thus, always stores and retrieves (and caches) properly, regardless of where it is storing the data to be preserved (i.e., on disk or tape). Thus, the only remaining critical question, or success factor, is whether this happens in a timely manner.

Some business needs demand that data be retrieved instantly, such as police records, especially when a pursuit is involved. Other applications are price sensitive, due to limited budgets and/or huge volumes of archived data, and its users may be willing to wait (to varying degrees) for the data to be made available, if this all can be done at a significantly lower cost. These applications might be seasonal or occasional and its users willing to wait a little longer (say, several minutes) for the data, especially if its use was not anticipated. **In the end, for many business uses and users, there is a balancing act between time-to-availability of the needed data and the TCO of the archived storage that is holding it.**

Typically, the more data that you have, the more likely you are to be looking for ways to reduce the unit cost per stored PB (for archival and retrieval). This study focuses on those enterprises that archive very large amounts of data for a long time and are looking for ways to keep the costs down.

Study Presumptions

In a nutshell, what did we do, both in 2010 and again now, in 2013? We compared lower-cost, high-capacity disk solutions against high-end *automated tape libraries* using *Ultrium LTO* tape, the most common format used for open systems data. We started with the current generation of disk and tape technologies and made future generation presumptions based on published road maps, historical extrapolations, and our best judgment.

What were the most important presumptions? First, we presumed that an archiving appliance controls the indexing and placement of data on the long-term storage media, whether disk or tape. Since the same appliance is used in both cases, we did not include its cost or the cost of the disk that serves as its operating cache as part of the TCO for disk or tape storage that "sits behind" the archiving appliance, which we also call backend storage.

Second, we made important presumptions about the nature of the data being archived. We presumed that it was blob-like digital data that could not be compressed or deduplicated. One example is scientific data streams, such as soundings from geophysical explorations, astrophysical data from radio telescopes, and detailed medical images. This kind of data is either too expensive to reacquire or cannot be reacquired at all, so it is presumed to be very valuable and, thus, may need to be kept for a long time, probably measured in decades.

Digital data also is the most challenging for archival purposes because it usually does not benefit much from deduplication or compression technologies and, therefore, represents a straight comparison between the benefits and costs of disk and tape as the media choices for long-term data retention.[12]

Third, we presumed that a three-year procurement cycle exists for both disk and tape and have based our 9-year model on 3 three-year cycles, as shown in Exhibit 4, at the top of the next page.

We believe that this is sufficient to draw a long-term conclusion but, more importantly, it clearly calls for tape to be the media of choice for long-term archiving for the next three years, the only period for which current decision-making is required, as long as its seconds-to-minutes access time is acceptable.

If we limit the term of this study to just the next three years (Cycle 1 only), the TCO ratio of disk to tape is 15:1, less than for the 9-year period because some ATL acquisition costs have been front-loaded at the beginning of the first cycle and because the 3-year old disk arrays do not get

[12] If your data is mostly office documents, heavily duplicated and/or capable of being compressed significantly, adjustments will have to be made to our model, which focuses on irreducible digital data. Do recognize that the adjustments go both ways, since it usually costs extra to deduplicate and/or compress disk files and objects, while LTO tape has compression built-in at no extra charge. (In addition, you will have to retain and maintain the necessary versions of the data deduplication and decompression software for future retrieval of your archived data.)

Exhibit 4 — Defining the Three 3-Year Cycles and Storage Capacities

Capacity Assumptions	Uncompressed Capacity in TBs		Rate of Capacity Growth	
Study Cycle	Tape	SATA/SAS Disk	Tape	Disk
Cycle 1 (LTO-6)	2.5	3	n/a	n/a
Cycle 2 (LTO-7)	6.4	6	256%	200%
Cycle 3 (LTO-8)	12.8	9	200%	150%

Source: The Clipper Group

replaced until the beginning of Cycles 2 and 3. While a 15-times higher cost for disk over tape is a compelling ratio for your decision-making, we think that it is very narrow-sighted to look only at the first three years, because it is too short to include the expected and significant later (long-term) costs, as will be discussed shortly.

If you need to act in 2013, then you must choose between today's tape and disk or a combination thereof. If something new appears, or if tape or disk fails to keep pace as we presume, then you can make a reasonable change at the beginning of your next procurement cycle.

Fortunately, the archiving appliance will make these transitions largely painless and transparent, since it handles the transfers from one generation of technology to the next, even when transferring to a different storage technology.

Fifth, we ignored the cost of a remotely maintained copy for disaster recovery. This presumption is explained in detail later in the report in the section entitled "Data Protection Presumptions" on Page 14.

Lastly, we tried to be fair in all of our presumptions, bending over backward to reduce disks' presumed higher costs, as described in Appendix A. This report, however, was not designed to simply compare acquisition costs for backend storage behind archiving appliances; it has been designed to compare the total cost of ownership of the different media choices, including continuing maintenance of hardware and middleware and upgrade/replacement costs, energy, and data center floor space.

In the end, such favorable bias toward disk really didn't make any difference to our conclusion that tape is significantly less expensive than disk for long-term archiving on a TCO per PB basis, since the TCO differential was so great.

We consider our presumptions to be very conservative. However, you need to compare your situation and IT procurement practices to our business case and adjust accordingly.

Our TCO Model

For our 2010 study, we used a 12-year period to examine storage costs, which we divided into four three-year cycles. We chose three-year cycles, because that matches (closely enough) the generational advancement of tape and disk products. While 12 years isn't forever, it clearly was more than long enough to see the trend.

Back in 2010, from the LTO Program roadmap[13], we had three future generational data points for Ultrium tape, i.e., *LTO-6*, *LTO-7*, and *LTO-8*, in addition to the then-current generation of *LTO-5*. Today, the current generation of LTO tape is LTO-6, which was announced last year.[14]

As we started the current (2013) study last autumn, the LTO Program roadmap *only* had two (remaining) future generations identified – LTO-7 and LTO-8. Even though the LTO vendors missed their LTO-6 capacity target of 3.2TB (only delivering 2.5TB uncompressed capacity), the goals for LTO-7 and LTO-8 continue to remain the same. We have chosen to accept that each of the next two generations of LTO tape will be delivered as described in the current LTO roadmap.[15] Rather than speculate on the future specification for LTO-9, for this 2013 study we decided to work with the three publicly-disclosed generations – LTO-6 through LTO-8, figuring that nine years still was long enough to see the trends and draw reasonable conclusions. We believe that this roadmap, both for the past decade and the next decade, provides the proof that tape is alive and well. [16]

[13] The LTO Program is a consortium of tape drive manufacturers that sets the specifications for Ultrium (LTO) generations. See http://www.lto.org/technology/roadmap.html for the roadmap (accessed April 21, 2013 5:02PM).

[14] 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.

[15] As you will see in a conclusion stated later, even if you stayed on LTO-6 for the next nine years, tape would still cost less than disk.

[16] The future roadmap for increasing tape densities is a little easier to accept than for disk, given the linear nature of tape and the fixed heads employed in tape drives. There still is a lot

We have used a price of $90 per cartridge through the study period (the same as in 2010), obtained from a reputable media source; anecdotal evidence suggests it is high. To satisfy our model, 5749 tapes will be needed, as follows.

- LTO-6: 1435 in Cycle 1
- LTO-7: 1709 in Cycle 2
- LTO-8: 2605 in Cycle 3

The cost of the cartridges is included in the Total Cost of Equipment and Media (TCEM), which also includes maintenance costs beyond included warranties. In order to ensure the best pricing for media, we have established an acquisition pattern to obtain cartridges in lots of 100[17], with the acquisitions taking place annually, potentially resulting in some of the tapes in the last batch of 100 cartridges not being used by the end of a given cycle[18].

At the same time, the advancement of disk recording densities and capacities seemed to be slowing down, due both to technological challenges and the coming of age of solid-state disks (SSDs), which are changing the face of high-performance storage and consuming more attention from the disk vendors. The capacities for future generations of high-capacity 3.5-inch rotating disk drives, beyond 4TB, are debatable, at best. Our decision only to look ahead two generations for tape also seemed to make sense for disk. So that is what we did. **Based upon the generation jumping that we have seen over the last decade, we feel comfortable presuming that each new generation (of disk and tape) will come at 3-year intervals.**

The uncompressed capacity for each generation of disk and tape used in our model is shown in Exhibit 4, at the top of page 8.[19] Uncompressed capacity makes sense for our model because most digital data (like video and medical images) already is compressed in its native format. The growth rates are calculated for uncompressed data.[20]

Data Presumptions

In both the 2010 study and this one, we presumed that the data center started with an existing collection (archive) of digital data with a capacity of 1PB. We presumed that an additional 45% of new data was added in each year, including the first.[21] This progression would grow to 28.3PBs by the end of the ninth year. This was more than big enough to represent what smaller and much larger collections would require. We presumed that no data was deleted during the nine-year period.

Noteworthy is the weighting that occurs by compounding (at 45% per year). What this means is that much more data is stored in Cycle 3 than in Cycle 1, as clearly shown in Exhibit 5, at the top of the next page.[22] The volume of data (and when it occurs) significantly affects the TCO, because we presumed that data will be stored more cheaply in later years, whether on disk or tape. It is fortunate that the cost of storing data is going down (as discussed in next section), as we'd all be in trouble if we had to store all of that future data at today's costs.

Cost Presumptions

As shown in Exhibit 2 (on Page 5), over nine years these are million-dollar expenditures. The amounts in this table show the costs for tape and disk solutions to accumulate and hold the 28.3PBs. Looking at the TCO totals, you can see how we determined the disk-to-tape TCO ratio of 26:1. It is interesting to note that the factor of 26 is significantly higher than the factor of 15 in our 2010 study. Here's why.

- We did this year's study for a 9-year period rather than the 12-year period used in 2010.
- There is significantly less data to contend with, 28PB (at the end of nine years) as

of room on tape to increase the density of what is being written without pushing the laws of physics. This is a very technical discussion that is not undertaken in this report. Both disk and tape are given the "benefit of the doubt", as you will see.

17 In our 2010 study, we used a lot size of 500, but the cost per unit turns out to be about the same for 100.

18 In the real world, of course, these tapes would be consumed. Regardless, the financial effect of not consuming the entire last lot of 100 is not significant.

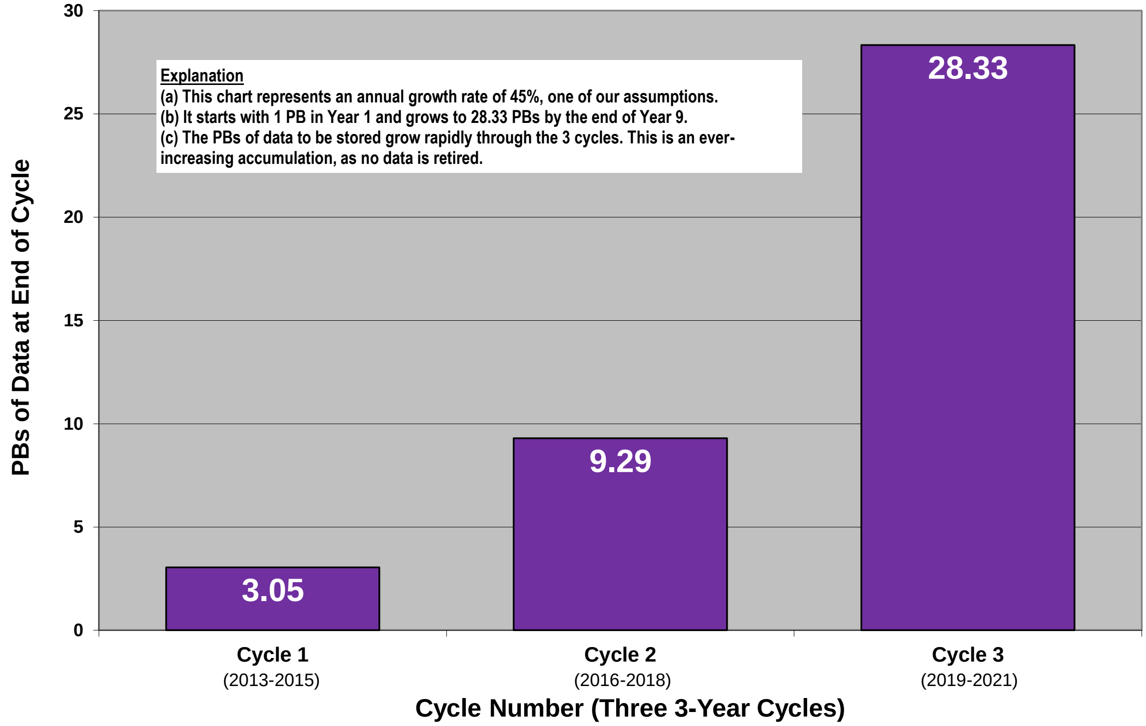
19 Some might say that we are favoring tape over disk by giving LTO the benefit of the doubt but not disk. As you will see later, we did test for the sensitivity of our model to have 12TB drives in the third cycle and the economic savings were modest, at best, as will be explained.

20 It should be noted that LTO tape drives have compression built into the drives (at no extra charge), while most disk systems charge more for compression, if available, whether done by software or hardware. Regardless, we presumed compression was not done.

21 *Why did we choose 45% and not a significantly higher growth rate?* There are two reasons. First, not everyone is doubling their data each year, as frequently reported. Second, it doesn't make any difference to this cost study, as the quantities of data are the same for both tape and disk and the model consumes enough storage for the incremental TCO per petabyte to flatten out and become linear (for both disk and tape). If we had twice as much data, the ratio between disk and tape wouldn't change significantly.

22 If your growth rate is higher, this effect will be exaggerated.

Exhibit 5 — Presumed Data Growth Over 9 Years
(as Required at the End of Each Three-Year Cycle in PBs)



Source: The Clipper Group

opposed to the 86PB (at the end of 12 years, which was the basis for the 2010 study). This large difference is due to the compounded growth rate over the extra three years.

- Because the 9-year period did not require copying the first generation of tape cartridges to fourth generation media, as we did in the 12-year study, the conclusions of our two studies are not always comparable. The copying isn't just an expense, as there is a significant consolidation of library slots after copying old data to denser media, which, in the long run, saves money.

It is important to reiterate what is not in the scenario described above. The cost of the hardware (servers and disk cache) for the archiving solution (appliance) is not included, because these costs would be the same regardless of whether the archival storage behind it was disk or tape, or a mix of both.

The costs of procuring the backend storage, including maintenance, energy and floor space was included. Purchase and maintenance prices were "List Prices". Your actual costs should be less, possibly much less, as there is a lot of competition. Most of this is between vendors of like media (tape versus tape, etc.). Vendors that sell both seem to be hesitant to price one against the other, but that may be exactly what you need to force them to do. In the end, those that sell both would like to sell you both and there may be certain technological, support, and cost advantages to you from buying both from the same vendor.

Also not included is any administrative (data center) labor, which would be fraught with presumptions that would be debated ad infinitum. While there may be differences in the labor for disk versus tape, we think that this is not significant, because the archiving solution does most of the "heavy lifting" of moving data from one generation of disk to the next and storing and cataloging the location of data on tape cartridges in the ATL, where all of the cartridges are stored[23].

All pricing comparisons were calculated based upon vendor-supplied and public configuration data and list prices[24], provided in confidence to Clipper by the vendors. We had more than twice the number of vendors in this study over

[23] For our scenario, there are no cartridges are in "cold storage sitting on a bookshelf", somewhere outside the ATL.

[24] While you can get a valid sense of street pricing on consumer electronics by surfing the Internet, this is not true for high-end IT procurements. Thus, we have chosen to use list pricing as the best available common reference point. Additionally, if we chose to use the same discount across the board, the ratios would have been the same.

2010, providing us with some very rich data points. The exact number will not be disclosed, as that might help one vendor to reverse engineer the confidential data from another vendor. Our goal was to come up with comparable averages for disk and tape from a sufficiently large number of vendors. That has been done. The data was gathered at the end of 2012 and the first quarter of 2013.

We presumed that procurement was done at the beginning of each three-year cycle for the entire cycle and that cash was paid up front. No interest was included in the study. Maintenance costs were normalized to account for varying warranty periods and coverage. These are not separated out, but included in the cost of equipment, which also included the cost of media (cartridges) for tape solutions. No attempt was made to categorize costs as capital or operating.

If you look carefully at the 2010 study and this one, you might notice that some of the average TCO component costs actually went up on a per terabyte basis, all while the densities continued to improve at little or no additional cost per disk or cartridge. While we won't dwell on this here, this anomaly can be traced to the increased number of vendors in this study, some of whom have less-efficient or more-focused solutions, in one way or another[25]. The larger sample is more representative but, as a result, includes a greater range of distortions. **In the end, you need to look at the bottom line offered by each vendor for each class of storage**, something that we have done but are not at liberty to disclose.

If there is one metric that is both on target and comprehensive, that is "TCO per PB". However, one has to be very careful with this metric, especially when the time dimension may span many years. Data stored in the first year of our study period stays resident for nine years and data written in the ninth year is resident for only one year. So when you say "TCO/PB", you must be very specific about how you deal with time and how you count PBs. The data summarized in Exhibit 2 (on Page 5) is the total cost (over nine years) for storing the accumulated mass of data that ends with 28.3 PBs at the end of the ninth year. It is not an annually-weighted average, so don't try to analyze this on a per-annum cost per TB basis.[26]

Disk Storage Presumptions

All of our disk vendors provided us with configurations based upon 3TB SAS or SATA disks assembled into fully-configured arrays[27], which can be considered Tier-2 storage. At the time that data was gathered, none offered 4TB disk drives.[28] We presumed that SAS/SATA drives continue to increase in capacity every three years. Based on current trends, and loud disagreement about potential technological limitations, we have chosen disk capacities of 6TB for Cycle 2 and 9TB for Cycle 3. This is our best judgment, but we also tested for variations, as you will see.

Trying to find a good mix of price points and recognizing that some archiving solutions are ambivalent about the type of backend storage, and also because some presume NAS[29], we included both SAN[30] and NAS disk array solutions in our study. One might expect that NAS would cost more (per PB) than SANs, because of the "NAS heads" (servers that act as file managers) involved, but that was not the case. The inclusion of NAS storage in our 2013 mix of solutions brought down the average costs for disks. **In the end, you want the lowest costs for your backend disk storage, so do compare SAN and NAS alternatives, if workable with your archiving solution.**

As we did in 2010, we presumed that the disk arrays would be replaced every three years, because the economics of buying new always seemed to trump keeping less-dense arrays around with seemingly very high out-of-warranty maintenance costs that tend to be based on the price paid three or more years ago. To someone outside of the computer industry, this may seem ridiculous, but this is the norm, driven by the pace of technology improvement, the fact that disks are

[25] One vendor's pricing was way out of line (vastly higher per PB) from what all of their competitors in the category were offering. Rather than letting this vendor's data distort the averages, we chose not to include their offering in the study, as no rational buyer would have chosen to overpay that much more for little or no apparent incremental benefit.

[26] Additionally, don't try to compare the 12-year TCO from the 2010 study to the 9-year TCO in the 2013. Not only are the total periods not equivalent, but the data in the 2010 study reached more than 86PBs at the end of the twelfth year. Thus, more than two-thirds of the total data volume was created in Cycle 4, with a significant downward pull on the average TCO per PB because of the Generation 4 capacity per disk drive or tape cartridge presumptions.

[27] Do not let the word "array" connote any particular scale-up or scale-out architecture. We use it to represent a delivered storage product.

[28] Later, there is a discussion of what 4TB drives might have done to the TCO.

[29] NAS=Networked Attached Storage, where files are the objects being stored.

[30] SAN=Storage Area Network, where the objects are blocks of disks called LUNS.

mechanical devices that are more likely to fail over time, and, of course, with the reality that the vendors desire to replace their products with newer, better offerings, and have set their out-of-warranty maintenance pricing accordingly. We presumed that the costs per drive stayed the same, even as capacities advanced.

We presumed that SAS/SATA disks were formatted for RAID-6, which has extra two drives for reducing the likelihood of the RAID group failing.[31] If a spare (blank) disk drive was allowed, we allocated one in each drawer or other physical storage collection; this reduced the capacity somewhat but is what most likely would be done because the rebuild time (after failure) on a 3TB drive is noticeably long and having a spare already installed would allow rebuilding to begin immediately. Data is accessible during background rebuilding.

We presumed that disk would be formatted and could be filled to 85% of formatted capacity. While this may seem very high, for storing data behind an archival system, this is believable. Remember that the data is being written once and never deleted; thus, much of the data is written contiguously or can be arranged in this manner.

Tape Storage Presumptions

Our tape comparison point began with Ultrium Generation-6 (LTO-6) tape cartridges and drives, housed in a fully-configured, enterprise-class automated tape library. We presumed that at the beginning of each cycle, the needed number of the then current generation tape drives would be procured and that the older drives would be retained in the library as spares (but not maintained), if there was adequate space (without extra cost). We did not use any of the older drives in our data writing calculations, looking at them as a free bonus for retrieving data from older cartridges, but to be used only when all of the latest generation drives are busy. We presumed that there was no salvage value to these older drives, after they no longer are installed in the library.

We also used an 85% maximum capacity factor, although the tape solution vendors might argue that this is too low for this application. In addition, we presumed that the necessary frames would be acquired at the beginning of each cycle, with sufficient capacity to hold the accumulated number of cartridges plus the new ones to be written within the three-year period. If available and if appropriate, high-density tape library frames were used as the ATL was scaled out.

We presumed that cartridges would be retained only for three generations, because LTO drives can read back two generations and write back one generation. At the beginning of the fourth generation, all of the first-generation cartridges would have to be copied to the latest-generation (and the old ones scrapped or cleaned for resale), thus allowing much more data to be stored in fewer library slots. This is economically favorable to acquiring more frames (and keeping and maintaining very old drives). However, because this was only a three-cycle study, no copy-and-replace components are included in the TCO calculations.

In estimating how many tape drives would be required for each study cycle, we presumed that one more (a spare) would be acquired than was necessary to complete all of the writing of the new data during a 10-hour daily window. This would leave all drives available for retrieval for the other 14 hours a day (ignoring the time for occasional maintenance). The spare was not included in the writing-volume/time-to-write calculations and any older drives that remained also were ignored.

Be aware that your tape drive reading and writing activities may require more drives than needed to satisfy our data model's requirements. The number of tape drives required to satisfy our model was modest, primarily because we presumed that *n+1* drives of the current generation could handle read/retrieval requests. If you need to mount and use many tapes simultaneously while you are writing into the archive, you will need more tape drives than were included in our model.

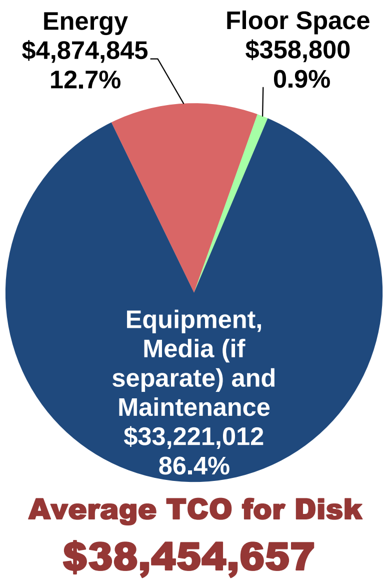
Energy Presumptions

Since 2010, we all have seen the yo-yoing of energy prices, especially gasoline and heating oil. However, when we went to research the per kilowatt hour (KWH) cost for electricity in a commercial setting, we were surprised. In 2010, we presumed $.15 per KWH, rounded up from 2010 U.S. Energy Administration data for New England.[32] When we revisited the same source in

[31] For a RAID-6 group to fail, two of its active drives must fail individually. If one fails, the RAID group continues to perform operationally, possibly with some degraded performance, will the bad drive is replaced and rebuilt from information on the remaining drives. With RAID-5, a single drive failure can disable the RAID group. Thus, RAID-6 is like wearing a belt and suspenders (at the same time).

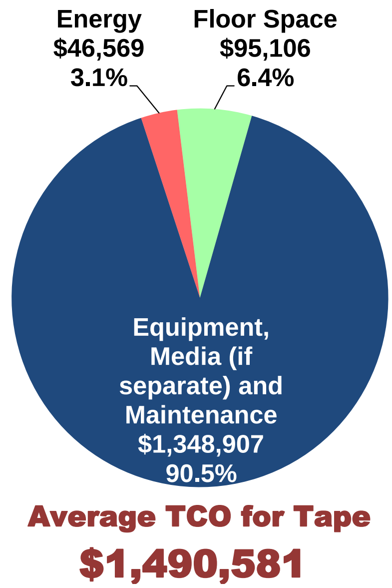
[32] We chose New England because it had higher costs than many other regions and was more representative of larger cities. If you happen to be in one of those lower-cost regions,

Exhibit 6 — Distribution of TCO for Disk



Source: The Clipper Group

Exhibit 7 — Distribution of TCO for Tape



Source: The Clipper Group

2013, we found that the average cost fell by a penny per KWH to $.14, again rounded up from 2012 data.[33] In addition, we found that the rate of change was essentially flat. In spite of the pain that we feel when we fill up our vehicles with fuel, for 2013 we decided not to build in an inflationary escalator. (In 2010, we presumed that energy costs per KWH would go up 5% per year.) This change does not affect our conclusions.

In fact, the rate charged for electricity is not the relevant metric; the kilowatt-hours-consumed is what is important. Most likely, the per-KWH rate will be whatever it is in your location. There's not much you can do about that in the short term, especially if you do not own the building. We need to be concerned about how many KWH are consumed by each solution. We presumed that disk and tape drives were spinning all of the time, which makes sense for the disk arrays but a lot less sense for the tape drives, which we expect all to be in use only occasionally. Configured KWH requirements were doubled to cover the cost of necessary cooling and air handling.

As you can see in Exhibits 6 and 7 above, energy takes a bigger bite out of disk's TCO pie than tape's TCO pie.[34] Don't be alarmed by this. The percentage of energy is going up (compared to 2010) because the costs of tape library components, maintenance and cartridges have gone down, significantly. On a KWH per PB basis, tape (and disk) are getting more efficient with each generation. Be aware that there can be significant energy consumption differences among similarly-appearing tape solutions and disk solutions.[35] This requires close scrutiny when you are doing your own TCO analysis.

Don't try to compare the energy slice percentages to each other (in Exhibits 6 and 7) without recognizing that the cost of energy for disks is 105 times greater than the cost of energy for tape, as shown in Exhibit 2 on Page 5. Regardless of the size of the slice of the TCO pie for tape's energy costs, **tape is exceptionally green, when compared to rotating disks.**

Floor Space Presumptions

Careful attention was paid to the floor space required for tape libraries and drive racks. This can be more complicated for tape, since the layout for some of them is non-linear. Adequate aisle space was built into our calculations. As we did in 2010, we used $200 per square foot for a raised-floor, properly powered and cooled data center. As with energy, the rate that you pay for

your energy costs will be proportionally less. See the rates by year by region at http://www.eia.gov/electricity/data.cfm.

33 *Ibid.*

34 It is easier to see tape's costs for energy and floor space in Exhibit 6 than it was in Exhibit 1, because Exhibit 6 shows the percentage of tape's TCO for each category, while Exhibit 1 shows the amounts spent in dollars.

35 This refers to comparing one ATL to another ATL or one disk solution to another disk solution.

floor space is mostly out of your control, unless the amount of floor space forces you to acquire additional floor space, and then you likely will pay more. As with energy, the percentage of the pie (again, see Exhibits 6 and 7) that goes to floor space isn't the important metric; what is important is the amount of storage that can be delivered per square foot of floor space. As was shown in Exhibit 2 (on Page 5), **you need four times the floor space for disk over tape.**

Data Protection Presumptions

Of course, there is more to consider. 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 the archiving appliance is responsible for keeping the remote site up-to-date.
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.

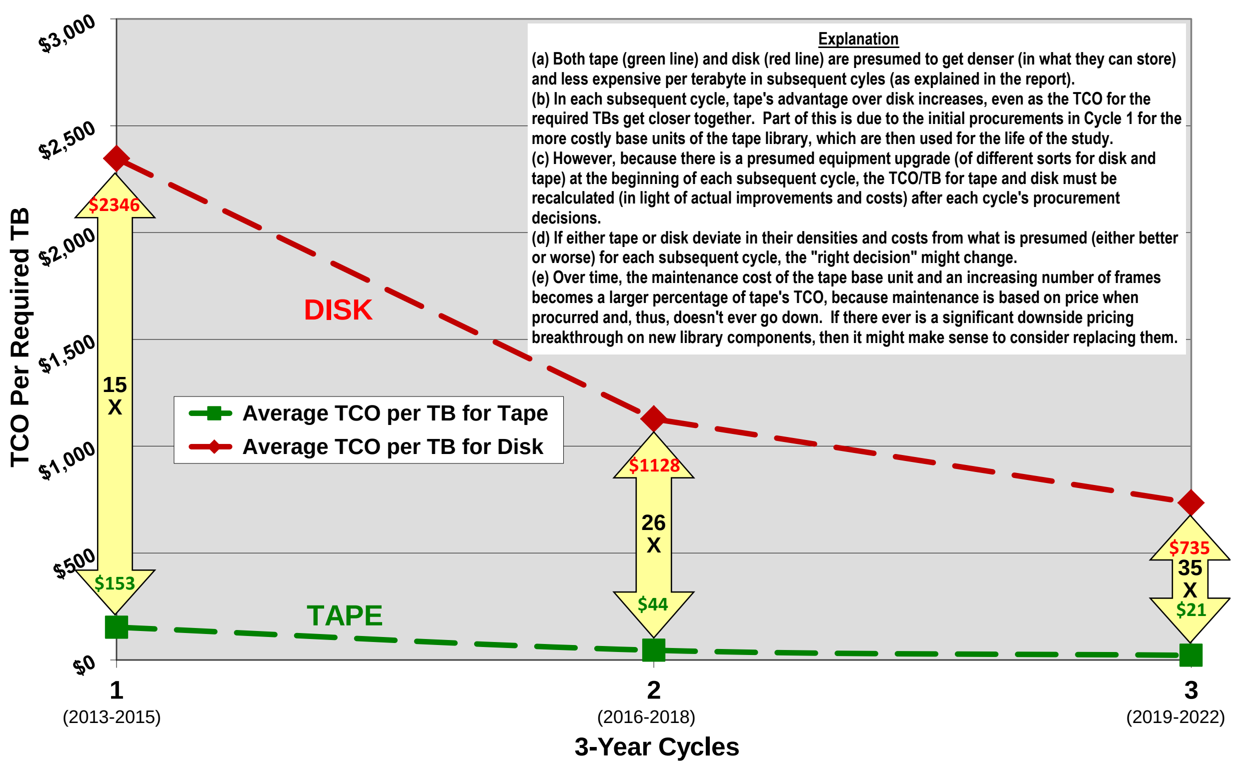
Economic Presumptions

We have chosen to ignore the time-value of money in this study. Here's why.

1. Data growth over the 9-year period is significant. Most of the data to be archived arrives in the second half of the study period.
2. If you were buying most of the equipment and incurring most of the expenses in year 1 (as one might when buying a railcar), it might have made sense to consider the time-value of money. However, by breaking the procurement into three "natural" cycles, we compare costs within each cycle and the time value of money is not very significant. This is especially true when interest rates are very low, as they are now. Unfortunately, we cannot predict when this will change and to what degree.
3. Our economic analyses are focused on Room #3 where the archival (backend) storage resides. Our cost analyses do not include the contents of Room #2, since that is the same regardless of what storage sits in Room #3. Thus, our decision to duplicate what is in Room #2 at a remote location (including whatever extra software and networking costs this will bring) does not affect our Room #3 analyses.
4. With the costs of Room #2 presumed away (for the purposes of our comparison of tape and disk), the three backup and recovery sub-scenarios (1a, 1b, and 1c, from the first column on this page) are all economic derivatives of the cost analyses that we did for Room #3 at the primary site.
   (1a) If the tape solution is the same at both sites, then the cost of the archival storage at the remote site can be presumed to be the same as the costs at the primary site.
   (1b) If the disk solution is the same at both sites, then the cost of the archival storage at the remote side can be presumed to be the same as the costs at the primary site.
   (1c) If disk is used at the primary site for archival storage and tape is used at the remote site, then the two-site solution is a combination of the cost for disk at one site and tape at the other.

These are calculations that you can do yourself from the cost data that we have calculated for the backend storage in Room #3. The simplicity of

Exhibit 8 — Average TCO/Three-Year Cycle (for PBs Required in Each Cycle) for Long-Term Archived Data (Using Clipper's TCO Study Model)



Source: The Clipper Group

doubling the cost of all-tape or all-disk archival solutions is easy to comprehend. If the cost of all-tape is less than all-disk, then the mixed mode remote scenario's cost will be between the all-tape and all-disk, similar to what was shown in Exhibit 2 (on Page 5). However, if all-disk is less expensive, then there is no cost reason to use tape at the remote location, although, there may be data protection reasons to consider tape (since tape is offline and therefore not susceptible to the data corruptions that can occur with online data).

Testing for Other Presumptions

Exhibit 8, above, shows that the per-TB costs decline in later cycles. This should be intuitive, at least to folks in the data center. What may be counter-intuitive is that the ratio between disk and tape actually is getting larger in later periods. Here's why. If you look in the right two columns of Exhibit 4 on Page 8, you will notice that the rate of growth for the capacity of disks is slowing down, when compared to LTO's roadmap. *Is this set in stone?* No. We could be surprised, but we think our presumptions are as good as anyone else's. *Is this a big deal?* That's a good question, one that we did explore. We did some sensitivity testing.

- ***What if 4TB drives had been available for disk arrays in Cycle 1, at the same cost as the current 3TB drives?*** This would only affect Cycle 1. We found that the 9-year TCO for disk would go down 3.8%, a pretty small amount. The ratio between disk and tape would drop to 25 from 26. <u>This would not affect the conclusions that we have reached.</u>
- ***What if 12TB drives were available for disk arrays in Cycle 3 (instead of 9TB drives)?*** We found that the 9-year TCO for disk would go down 12% and the ratio would drop to 23 from 26. While this is driven by the fewer disk drives and arrays that will be needed, do not lose sight of the fact that there is a lot more data to be stored in the third cycle. At 23 times the TCO of tape, <u>we still think that our conclusions are valid.</u>
- ***What if both of the two factors just discussed (resulting in 4TB drives in Cycle 1, 6TB drives in Cycle 2, and 12TB drives in Cycle 3) happened in the same modeling run?*** While we consider this an overly-generous disk-favoring set of presumptions, we tested for it anyway. Disk came in at 22 times that of tape; thus, <u>we still think that our conclusions are valid.</u>

While all of these variations are interesting exercises to consider, they do not change the conclusion that **disk always is more than twenty (or more) times as costly as tape.** Here's the reason why this generalization is true, as shown in Exhibit 8 – ***there is no way for disk to catch up with tape (without a technology breakthrough).*** **In fact, tape would be less costly than disk even if you stayed with LTO-6 infrastructure for nine years** (i.e., acquiring more LTO-6 drives and media in later cycles instead of buying LTO-7 and LTO-8 technology when it became available), because the TCO per PB for disk never gets as low as it is for tape in Cycle 1.

Also noteworthy in Exhibit 8, the TCO for each cycle goes down significantly for both disk and tape as we progress through the study period. This is to be expected, as we presumed that disk and tape densities continue to improve and that drive and cartridge costs stay the same or go up a little.[36] However, please note that the TCO ratio of disk to tape for each cycle continues to increase, from 15 times to 26 times to 35 times, mostly because there is so much more data to be archived in each ensuing period and the cost per PB continues to go down in each later period.

Conclusion

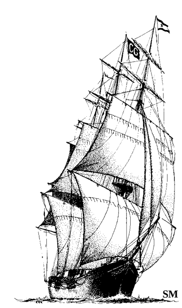
In the course of this study, we have reviewed the required configurations from a variety of vendors for both disk arrays and tape libraries and taken the average of the systems studied for the comparison. **As a result of this comparison, we show that tape is the better value in terms of total cost of ownership for the long-term preservation of irreducible images and binary data, if you can wait for several seconds to several minute for retrieval.** Here's a summary.

- In terms of TCO under our scenario, the cost to implement a disk solution for long-term archiving is over \$38M, about 26 times the cost to deploy a tape solution of almost \$1.5M.
- A very large expense (proportionally) in deploying a tape archiving solution might be, for some, an operational expense – the cost of media. The cost of tape cartridges, about \$378K, represents about 25% of the TCO for tape, and has been included in the total cost for the equipment and maintenance.
- As we saw in 2010, the cost of energy continues to be a significant factor in the operational costs of maintaining a disk subsystem for long-term storage. In fact, the cost to run and cool the data center for a long-term archival storage on disk is about \$4.9M. This represents a ratio of 105:1 in comparison to the energy required for a tape archive, about \$47K.

For decades, because of its cost, most enterprise data centers have used tape as the low-cost media of choice. Now, with any number of storage solutions dotting the landscape, the data center staff is faced with a potentially budget-busting question: *Does your data center need to spend tens of millions of dollars for sub-second response time for archive retrieval requests or is a response time of up to several minutes sufficient to satisfy the service level agreements that exist with your user community at 1/26th of the cost?*

Restated in terms of this TCO study: *Is near-instantaneous response worth over \$38M over 9 years for <u>all</u> of your archival data?* If all of your archive retrieval processes demand one-second response time (or less) for millions of queries during which expensive people are just sitting and waiting to get the data (or worse, as when people are at risk), then the answer probably is "yes, spend the money."

If some of your data does not require instantaneous response or your policies generally will retain what requires speedier retrieval in the archiving appliance's cache, then a tape solution will save a lot of money. If the ratio of TCO for disk over tape was only 10:1, then the answer should be clear; but at 20:1 or more, it should be compelling. ***Think about it … long and hard!***



[36] After the introduction of a new generation of tape drives, the cartridges for that generation tend to cost a lot more than they will six to nine months later when more tape vendors have compatible cartridges. Our average cartridge price is for the more competitive scenario. In reality, most writing onto the new format takes place after the initial period after availability.

Appendix — Explanations of Model Variables, Presumptions, and Bias

#1 – USABLE CAPACITY – *Model Bias Favors Disk…Potentially Significantly*

- **Presumption – Both tape and disk are filled to 85% of their usable capacity.**
- **Explanation** – For disk, this is a generous presumption, since most disk solutions tend to become "filled" at a much lower percentage. For tape, being written in serial order (as it comes in, until there is no more room), 85% tends to be an austere limitation. If disk is 10% too high and tape is 10% too low, then there is a significant (20%) bias that favors disk.

#2 – COSTS OF TRANSITION – *Model Bias Favors Disk…Somewhat*

- **Presumption** – **Disk and tape both have negligible costs for transition between generations.**
- **Explanation** – Because the archiving appliance manages data migration from one generation of solution to the next, neither will be as traumatic as manually migrating a disk array, LUN by LUN, and dealing with the associated addressing issues. However, we presumed that each generation of disk is replaced by the next every three years, while older tape cartridges are consolidated and rewritten every nine years, which is beyond the three cycles of this study. Physically managing the migration of even thousands of tape cartridges is a much less demanding physical chore than installing new disk solutions and, after the data has been migrated and cleansed, removing the old disk solution. This requires staff time, extra space, extra energy, etc., none of which are included in the disk costs.

#3 – ENERGY – *Model Bias Favors Disk…Somewhat*

- **Presumption – Disk and tape consume energy at their maximum operating level.**
- **Explanation** – For disks, this makes sense, since the media is spinning all of the time even when not being accessed. None of the involved vendors' disk solutions use any form of spin-down (quiescing). However, for tape, this is an extreme presumption, as it implies that all tape drives are moving, reading, or writing, all of the time. We have calculated the time needed to write the volumes of new data to tape, but presumed that this must be done in a 10-hour window and without using the spare drive, leaving the other 14 hours and the spare drive for retrieval. This presumption means that the energy required for tape may be much too high, but the tape solutions use so little energy that the bias against tape is not significant.

#4 – COMPRESSION – *Model Bias Favors Disk…Potentially Significantly*

- **Presumption – Data is uncompressible.**
- **Explanation** – We know that even with blobs, some "lossless" compression may be possible. A compression capability is built into LTO tape (at no extra charge) but not into any of the included disk solutions. Thus, the tape capacity required by the study's model might be higher than necessary, which provides a significant bias toward disk.

#5 – MAINTENANCE – *Model Bias Favors Disk…Somewhat*

- **Presumption – All hardware is maintained 24 by 7 with a 4-hour response time.**
- **Explanation** – By assuming that the disks are RAID-6 configured with a spare for approximately every 12-to-16 drives, we feel that the data is protected (in its archived location) and operations can continue uninterrupted while rebuilding takes place and the bad drive(s) are replaced. The same is true for tape, as there is always a spare tape drive of the latest generation. However, the spare tape drive actually can be used operationally (for writing and for reading) while the spare disk drives cannot. Additionally, we have presumed that tape drives purchased in a prior generation will be kept installed (but not maintained) as long as there is room for them within the tape library. Thus, there always is at least one extra tape drive and potentially several additional older tape drives, generally increasing the throughput potential, for use when needed. This presumption is a bias against tape, which favors disk.

#6 – FLOORSPACE – *Model Bias Favors Disk…Somewhat*

- **Presumption** – **All hardware requires only the specified floor space, including space needed to provide access to all functional panels and components.** We presumed that floor space in the data center is valuable and available. We make decisions in each cycle that inherently minimize the floor space used.
- **Explanation** – We have considered all specified installation requirements. Racks full of spinning drives generate significant heat (unlike tape libraries and drives) and additional space may be needed to balance the placement of the racks of disks. For example, there may be an airflow plan for hot aisles and cold aisles, potentially increasing the square footage required for disks. Since we did not consider this, this presumption probably favors disk.

#7 – EXCESS CAPACITIES – *Model Bias Favors Disk…Somewhat*

- **Presumption** – **Disk and tape solutions are bought in optimal capacities. Additionally, tape cartridges are procured in an economic order quantity of 100.**
- **Explanation** – At the end of each 3-year cycle, we presumed that disk solutions are replaced (with no salvage value) and nothing is carried over to the next cycle. However, for tape we make similar replacement presumptions for the tape drives, but we retain the old drives for use as spares. In addition, because cartridges are bought in lots of 100, there almost always are some virgin cartridges left over, which we have presumed will not be used. In reality, both the older tape drives and the extra cartridges likely will be used, but we do not estimate this potential cost-lowering contribution to TCO that, in the end, increases the TCO for tape and thus favors disk.

#8 – CONFIGURATION VARIATIONS DO NOT SIGNIFICANTLY AFFECT ACQUISITION COSTS – *Model Has No Bias*

- **Presumption** – **We have endeavored to make the configurations and pricing equivalent for all vendors in the same class of product.**
- **Explanation** – There are some small configuration variations among the like offerings. We have adjusted for these variations as best as we could. We are confident that they are generally equivalent. Given that discounts should be expected for multi-million dollar procurements, the differences can be considered as insignificant to your procurement decisions. We do not think that these variations affect the averaged TCO significantly.

#9 – PEOPLE COSTS – *Model Bias Favors Disk…Somewhat*

- **Presumption** – **No IT personnel costs are included in the model.**
- **Explanation** –Since the archiving appliance sits in front of the backend archival storage (disk or tape), most of the "work" is done by the archiving appliance. There are generational upgrades for both disk and tape, which for disk are forklift upgrades (wholesale swapping of old arrays for new ones). Thus, there may be a little more IT staff work being done for disks, which results in a small bias to disk.

#10 – PACE OF TECHNOLOGY CONTINUES AT VARYING RATES, BUT OUR MODEL IS FROZEN IN TIME – *Model Has No Bias When Published, but Might Favor Tape Over Time*

- **Presumption** – **Tape and disk technology were represented as having three-year refresh cycles, driven by the generations of new drives. The TCO was determined early in 1Q2013.**
- **Explanation** –While it is true that new and or improved disk arrays arrive on the market somewhat continuously (and often independent of the advancement of hard disk generations), tape's improvement cycles tend to be more tied into the tape drive generations. New or improved ATLs do get announced but there is no discernible pattern for this. The bottom line is that disk technology may get "better" (due to the inclusion of SSDs and/or auto-tiering, for example) or less expensive (because of bundling of software with the hardware or because hardware components become less costly). Both tape and disk solutions might go down on a cost per PB basis, for competitive reasons, as well. You need to assess what is available and what it will cost when you are making your decisions, as change is to be expected; maybe a little more frequently for disk than for tape.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twenty-first 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 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 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 20 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.