



TCO Analysis Elements and Arguments

Nicholas Pearce-Tomenius
& Jon Morgan

Table of Contents

Introduction	3
The View from Object Matrix	3
Analysis Parameters	4
Analysis period.	4
Data Protection Levels	4
Initial storage requirements.	4
Annual growth requirements	5
Solution hardware and software costs	5
Cost of Disaster Recovery	5
Base Media Costs	6
Annual Change in Cost of Media	6
Percentage of Usable Capacity	7
Media Rollout	7
Servicing/Support Costs	7
Workflow changes	8
Exit Costs	8
Infrastructure Costs	9
Power	9
Cooling	9
Network	10
Floor Space	10
Failover Facilities	10
Data Management and Manpower Costs	11
Cost of Employees	11
Maintenance costs	11
Obsolescence and Migration	12
Duplication, Replication & Failover	13
Restoring Content (or Not)	14
Full Workflow Support	14
Exit Costs	14
Opportunity/Legal Benefits	15
Horror Stories	16
Conclusion	17
References	18

Introduction

Is there such a thing as an unbiased total cost of ownership (TCO) paper? The writer of a paper on TCO is highly likely to have either been commissioned to do so after receiving a sizeable donation from a large disk/tape vendor or, as myself, works for a company for which a favourable TCO analysis is the aim.

This document, whilst written from a biased source, tries to identify and ensure that the TCO analysis contains ALL the elements that are relevant today. The information used has been gained from both tape and disk advocates. The document includes references from articles/white papers that can be found on the web as well as those gathered from storage and media communities, our blog and Twitter groups.

Traditionally, TCO calculations take into account only the infrastructure costs of owning a large digital information archive. The factors typically taken into account include: purchase cost, floor space, electricity consumption and tape costs. This is the method favoured by tape vendors because on that basis alone tape comes out favourable against rivals such as disk. But is that a true picture of TCO? What about media lifespan, standby units, time spent by employees maintaining the system? There are many factors that can also be taken into account, that arguably create a truer TCO picture.

A TCO analysis of digital information archiving solutions should take into account:

- Analysis Parameters (e.g., the period of analysis)
- Infrastructure Costs (e.g., the hardware costs, expansion costs)
- Management Costs (e.g., time)
- Opportunity/Legal Costs

NB: It is also important to note that not all disk solutions are equal. This report is focused on the advantages of clustered storage of the self-healing, self-managing sort. A good example of which is “The MatrixStore” from Object Matrix.

The View from Object Matrix

The long and the short of any comparison of TCO is that the inputs and resultant comparison will be vastly different for between organisations. Major factors that affect the the TCO that can give different results particularly include the amount of manpower required to run one solution over another. Other “macro”/overriding factors may be:

- If quick, multi-user or global access to assets is key disk is the only technology that can scale to achieve that
- If functionality is a key player (e.g., regulation compliance, regular verification of content authenticity, security that data wont “walk”) is required then this TCO model may need major adjustment and may require new inputs. In general, disk can do everything that tape can do, and more, except (normally) being able to be sent in the post.
- If you do not have the in-house technical resources to manage storage technology, or, feel uncomfortable employing technologies that require specialist skills which may become difficult to find in the future, then that may be a reason beyond TCO to choose one technology over another.

So, read this paper with your own needs in mind. The conclusion may not be relevant to your particular organisation or requirements but most of the TCO elements within it will be important to consider and may have a dramatic effect on the way you calculate that your own TCO.

Analysis Parameters

The input data for the TCO model starts with what we call analysis parameters. The primary parameters are the period of time over which the TCO model will be calculated and the amount of data that will be stored. But how long should the period be? How “safely” does data need to be kept?

Analysis period.

For long-term preservation of digital content the analysis period should be for approximately 6-12 years. Such a period allows for the costs of media format change, content migration to take place and for clearly identified industry trends to take place. The period covers from when ingest of content onto the platform commences through to the eventual potential moving of that data to a new platform.

Data Protection Levels

Object Matrix applies four maxims here:

- If you don't have at least two copies you do not have it
- If you cannot find it you do not have it
- Your data should not potentially become reliant on your vendor's existence/technical support
- Your data should not become reliant on “reader's” that can become difficult to replace (think Digibeta & BetaMax tapes/recorders)

Other data protection costs to consider are:

- How (if necessary) to keep authenticated data copies in geographically separate locations
- Is data kept in standard formats (When is a file not a file? When it has been chopped up into multiple parts and unusual algorithms have been applied to it), and if not, what is the cost to recover that data in exceptional circumstances
- How is data re-protected should a copy of the data become corrupted or lost, and how much manual intervention does that involve

As an analysis parameter it is necessary to calculate how well data needs to be protected. A calculation of risk and data value vs cost can be applied in that decision. Object Matrix recommends having at least two RAID5'd and verified copies of data on disk, or 3 copies of data on tape.

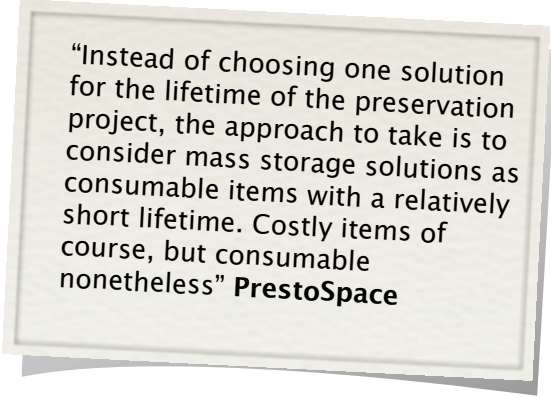
Initial storage requirements.

Another analysis parameter is how much data storage needs to be purchased up front. However, this may not be as simple a parameter to calculate since the amount of storage required is a product of how long data ingest will take (does it need to be digitized from another form of media?), what is the minimum size of storage that needs to be purchased in the solution type, and how do the internal company budgets work.

So, for instance, with MatrixStore it is assumed that the price of storage per TB will drop up to 50% year-in-year-out due to improvements in underlying disk capacity, speed and energy utilisation. Thus, if the ingest of data will take many months, then it may not make sense to purchase all the data storage required upfront but be wiser to purchase each additional unit of storage as and when it is required.

When comparing two solutions, if content can be data compressed on one type of storage it almost certainly can be compressed on another, and the same compression ratios should be used. Data de-duplication is sometimes possible on disk (not yet tape) based solutions but is only generally relevant if the same data is being stored multiple times. Compression and/or de-duplication always messes with the data format and which makes data harder to recover in disaster scenarios.

PrestoSpace is a European Funded project that has a mission about: “Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe.” Their take on disk based storage is:



“Instead of choosing one solution for the lifetime of the preservation project, the approach to take is to consider mass storage solutions as consumable items with a relatively short lifetime. Costly items of course, but consumable nonetheless” **PrestoSpace**

Buying a 1200 slot tape library for £120k+ is a bit of a risk if the technology does not fit the workflow as desired. Also, buying a smaller library does not allow for seamless growth. Clustered storage solutions like ‘The MatrixStore’ are designed to start small and grow, thus reducing initial outlay risk.

Annual growth requirements

Following the initial storage requirements, most projects can foresee what the approximate annual growth of the archive will be required. See “Estimating the future price of media” below.

A risk consideration is “what if we exceed the estimates?” Will the whole solution need to be replaced earlier than foreseen in that circumstance? It may be that the upper limit in scale of the selected solution comes into play there.

Solution hardware and software costs

These analysis inputs need cover all the hardware and software costs required for the solution to function (other than the infrastructure costs covered later).

This starts from the software tools required to work with that particular storage solution (backup tool plugins, HSM tools, specific monitor tools etc) through to the licenses and nuts and bolts of the solution itself. In the case of a tape based solution, the cost should also include any disk based cache that lies in front of the tape part.

What prices should be input? The list prices? If possible accurate quote prices should be used (on both sides of the equation) but sometimes when making the first analysis the only semi-public information we have to go on are the list prices. Finding list pricing is another game in itself. Luckily some vendors don’t hide their pricing and there are a few useful blogs out there (storagemojo.com for instance) to shed light on how much products cost from vendors who seem to have something to hide.

Other questions to consider and potentially to factor into the analysis are, does the license cover the asset or the media? Is the license transferable when media fails or gets upgraded? What upgrade licenses will be required for software changes? Will changing the backup software require new licenses or plugins?

Cost of Disaster Recovery

The cost of data being offline can be crippling to companies. Indeed, there are many examples of companies who have folded due to relying on a single tape or a single RAID5 or RAID6 system which has been permanently or temporarily lost. For good levels of business continuity data must be kept in two geographic locations and readable from both. If high data availability is required then there must be two geographically separated tape libraries and two clustered disk bases solutions.

The analysis parameters here are to enter the up-front infrastructure, software, communication, contractual and hardware costs. Other considerations are the ability to maintain two copies when things go south. Can the platform self-heal across geographic locations or is it a manual process? What are the

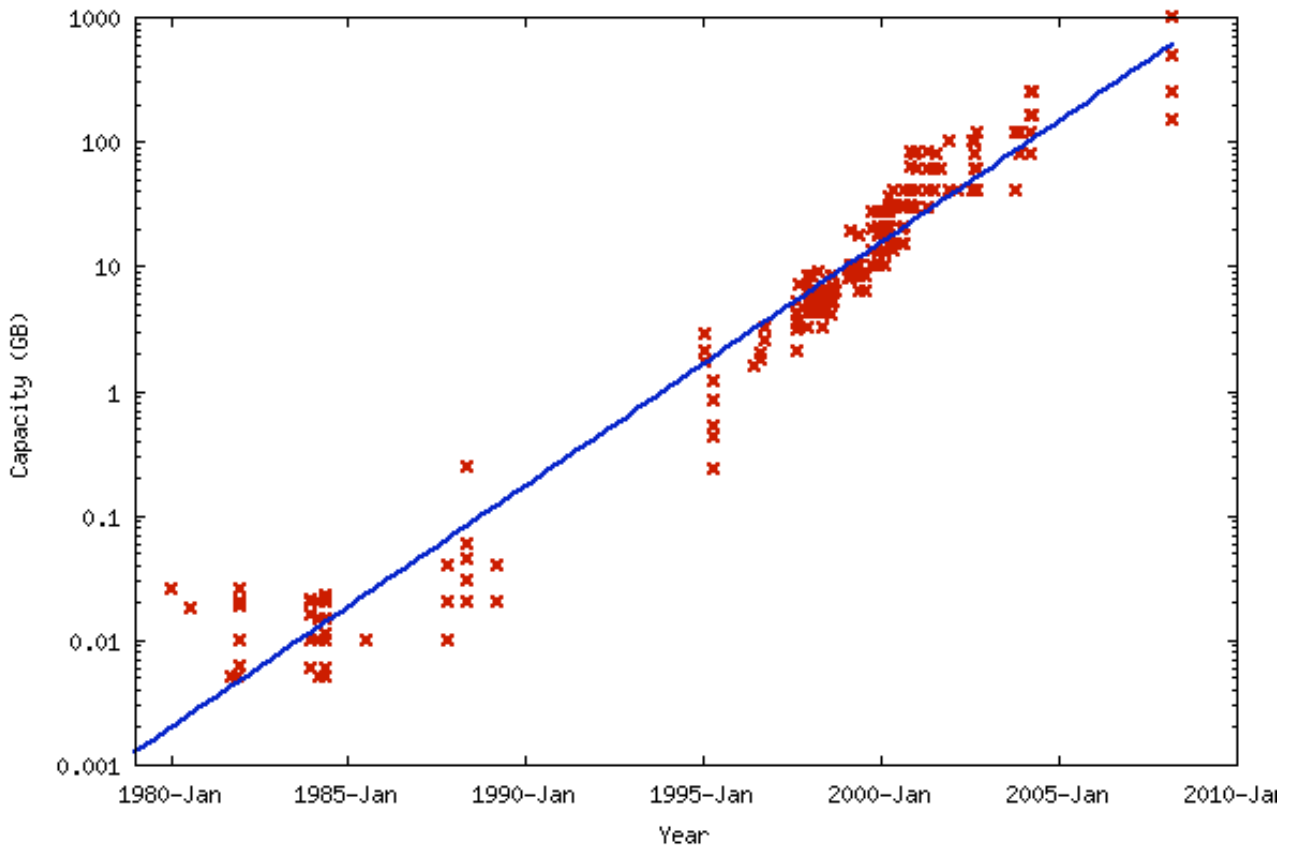
daily “synchronisation” costs for transmitting or carrying data between sites? How up-to-date do the sites need to be with each other?

Base Media Costs

Analysis should include the cost of each tape/disk required, e.g., LTO4 tape at time of writing comes in at about £38 for a 0.8TB tape.

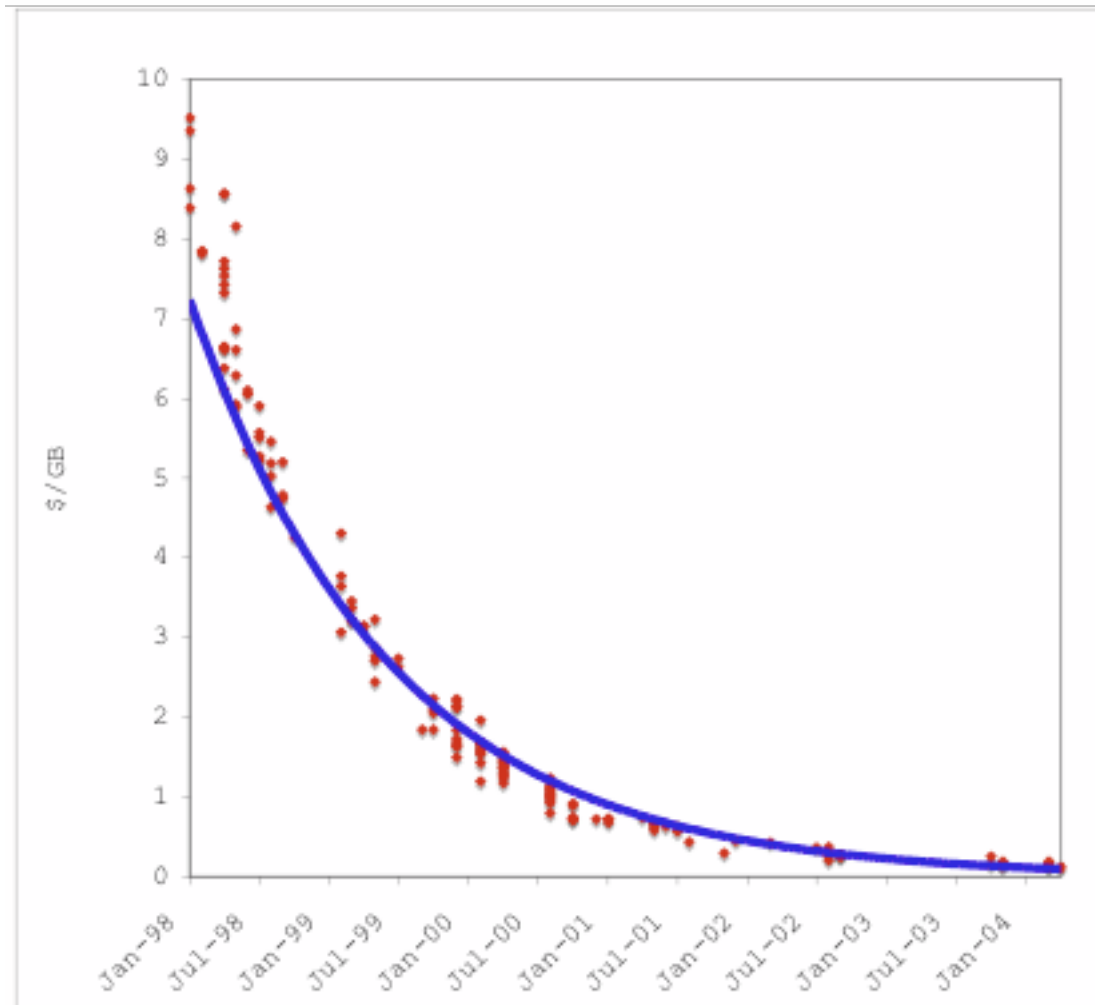
Annual Change in Cost of Media

Based on historical data and published plans (e.g., it is known that there will be a general move from LTO4 to LTO5 within the next 6 years) it is surprisingly reliable to predict the change in price and capacity of disk and tape technologies. Historically HDD capacity has increased linearly and demonstrates a 100 fold increase in capacity every 10 years:



can be seen in the diagram, we’ve risen from a 1GB HDD to a 100GB disk in 10 years, and from a 10GB HDD to a 1TB HDD in 10 years.

Meanwhile, the fall in price per GB for a HDD has been considerable. In 1980 you would be looking at a staggering \$233,000 for a GB of storage space, where as today a GB costs less than 20 cents. The following chart has been created by placing sample data of HDD prices from 1998 to 2004:



So, whilst the growth in disk size is linear the fall in disk \$/GB exhibits exponential decay.

Tape technology has also improved in the same period but nowhere near at the same scale. Since its launch in 2000 the LTO standard has gone from offering 100GB (LT01) raw capacity per tape to 800GB (LTO4) today. Thus, when plugging in the costs of media, what one should never do is to assume that the cost of the media will be the same in 5 years time as it is today. Not unless, that it, the same disks or tapes are being purchased! It can be safely assumed that with a solution, e.g., such as MatrixStore, the HDDs used to expand the size of the archive in 2 years time will be at least twice the capacity of the HDDs purchased today.

Percentage of Usable Capacity

It is all very well purchasing 1200 800GB tapes to give you 960TB of storage capacity if, in reality, only 60% per tape of that storage capacity is regularly utilised. Storage capacity utilised will depend upon the usage scenario and the ability of the software/administrators storing data. An average tape utilisation is probably somewhere nearer to 80-85%. Clustered solutions such as The MatrixStore can use all but 5% of the capacity available in the cluster.

Media Rollout

Covered somewhat in the Maintenance Costs below, if a media being stored to has a life expectancy, then the hard and soft costs of replacing that media need to be factored in.

Servicing/Support Costs

Usually an annual fee based on a percentage of the purchase cost. One question that may make a

difference there is, “can the solution be remotely monitored and managed?” Such remote work might also include software servicing and data verification tasks.

Workflow changes

No platform is an island. Software and applications are required to drive the hardware but also to provide the intelligence as to where data should be protected and at what level. As mentioned above, tape requires some sort of HSM solution to move data in and out which requires its own software licenses and possibly its own hardware. Double that cost if a standby or failover unit is required. Solutions like ‘The MatrixStore’ have software and failover built in to their prices.

In terms of fitting into an organisations workflow to provide seamless integration with automated processes one also has to look at the cost of the software required to provide that functionality if it is indeed possible. Clustered solutions with an API can be customised to fit tightly into workflow/process and providing data and metadata on demand.

Exit Costs

The cost of moving from a platform or format is also a consideration. If you have 10 years of content on LTOx then you are likely to have in excess of 20,000 tapes. Not only is there the management costs of migrating that to another platform, but also safe data deletion and physical disposal costs.

With an online disk based solution data migration and safe deletion can be automated, with tape those processes may be available but only if the solution is fully robotic.

Infrastructure Costs

Tape vendors tend to point at power consumption as the “hidden TCO” Achilles' heel of disk which as can be demonstrated is misleading to say the least, but that is a consideration that vendors of disk based solutions must take seriously. Other infrastructure costs include network bandwidth costs, floor space and cooling.

Power

Let's start with the big one. The claim sometimes made is that disks use power and that tapes do not. True in some ways false in others. Take a robotic tape solution that can handle 2000 tapes. It is immediately plain to see that the tape solution uses a lot of electricity, especially when it needs to constantly wind through tapes to reach the searched for piece of data. But how does that compare to disk? And what about non-robotic tape solutions? They require extra staffing resources, correct temperature tape storage, and possibly the manual movement of tapes between sites. Meanwhile disk based solutions are consuming less electricity per TB stored and per HDD used all the time and can store more data in a smaller footprint than can tape. New ‘green’ drives include firmware to power down disks to differing levels dependent on access, and the move to solid state technology is likely to bring more energy efficiencies in the future also. Furthermore, a good disk based archive will allow old disks and/or nodes to be rolled out of the storage solution and new (higher capacity, lower energy) disks/nodes to be rolled in as and when it is convenient.

The comparison in energy consumption should be made taking all those factors in mind, and for disk certainly should never be based on the archive solution's energy draw today multiplied by the capacity of storage required of tomorrow.

Cooling

“Temperature plays little role in errors - just as Google found with disk drives - so heroic cooling isn't necessary. Good news for data center air economizer architectures.” **StorageMojo.com**

A large part of power consumption is the cost for cooling and whilst several of these comments are more generic than purely dealing with subjects around the archive, the knowledge is too valuable to ignore in our drive to reduce energy consumption:

Intel recently conducted a 10-month test to evaluate the impact of using only outside air (also known as [air-side economization](#)) to cool a high-density data center in New Mexico, where the temperature ranged from 64 degrees to as high as 92 degrees. Intel said it found “no consistent increase” in failure rates due to the greater variation in temperature and humidity. “This suggests that existing assumptions about the need to closely regulate these factors bear further scrutiny,” Intel concluded.

Yahoo last month announced plans to build its greenest data center in Lockport, N.Y. Yahoo co-founder and Chief Yahoo [David Filo explains in his blog that the data center will be designed to use 100% outside air to cool the servers](#). And the resource used to power the data center and the servers will be the renewable hydroelectric power from the Niagara Falls.

Filo writes that the data center design will have an annualized average PUE (power usage effectiveness) of 1.1 or better. The company's other facilities in Washington are powered by wind and hydroelectric sources. Outside air (or what the industry terms “free cooling”) is used for cooling most of the year, helping Yahoo to reduce energy consumption by 40% to 50%.

Search rival Google uses free cooling almost year round in its data center in Belgium, where the average summer temperature is between 66°F and 71°F (19°C and 22°C), writes Data Center Knowledge. That temperature range falls way below the 80°F (27°C) [temperature that Google maintains for its data centers](#), adds Data Center Knowledge.

Object Matrix's conclusion is that many data centres might benefit economically and environmentally by reconsidering the cooling mechanism that they are currently using, but however for non-economised data centres, we tend to work out the cooling energy required at a 1:1 ratio (1 BTU).

Network

Some questions to answer when looking at how a solution may impact on your existing network infrastructure are:

- Does the platform require extra cabling and switching, be it fibre or ethernet?
- Will traffic to from the solution cause the network to slow down and thus force the organisation to upgrade or add to their existing infrastructure?
- How does content get moved off site, is there a cost element to that and will you require further tools or applications to do so?

Floor Space

"I did some calculations using the biggest, fully populated disk system I could find, powered it on all the time and calculated how much power that would cost. It came out to something well under \$10,000 a year. That translates into 1/5-1/10 of the cost of an FTE, and I'd argue that maintaining that disk system is a heck of a lot less expensive in labor dollars than maintaining the tape system. Regarding the footprint, disk systems beat tape systems any day on a GB-per-square-foot comparison." **Curtis Preston**

In general, floor space required by disk based solutions is similar or can be less than that used by their robotic tape equivalents since a HDD can hold 2.0TB in a 3.5" format, but LTO4 tapes hold 0.8TB. However, actual solutions greatly vary.

Failover Facilities

The cost of this element lies in the power, cooling, networking and physical buildings/floorspace for the remote storage platform. Other hardware and software costs are discussed in the analysis parameters, and management costs are discussed in the next section of that name.

Data Management and Manpower Costs

The total employment cost of just one person is likely to exceed the entire cost of a digital archive over the lifespan of the technology by up to 500% (depending on the size of the archive), so any TCO model that ignores the manpower required to run the archive is ignoring the biggest single factor involved.

Somewhat conveniently tape and optical vendors omit the management cost since it whilst the difference in manpower required to maintain a self-managing disk based archive vs a robotic tape solution (which at the very least requires thousands of tapes to be managed, and at worst requires physical movement of tapes) is always going to end up being beneficial to the disk based solution. Furthermore, disk-based solutions lend themselves to multi-user workflows that don't require going through individual "tape managers", and allow all archive users to get on with their jobs quicker than having to wait in a tape queue, or time to first byte.

With the advent of self-managing disk based archives such MatrixStore from Object Matrix almost all of the manual tasks required to maintain a tape/optical architecture are rendered redundant. The paradigm of automated self-management (be it robotic tape or disk) also eliminates human error, one of the most frequent causes of data loss in an organisation.

When comparing a disk-based data protection solution to a tape-based solutions that require a manual effort, the value of improved restore times and increased reliability can be substantial. Enterprises continue to pay hidden costs for lost productivity due to long restore times or worse, the cost of reconstructing data for restores abandoned or never initiated due to the limitations of tape backup.

Cost of Employees

Take into account:

- 1.How many staff do you need to manage the platform. Studies put the numbers of staff at 2 per 500TB of tape and 1 per 1 Petabyte of clustered storage.
- 2.Staff required to do daily write and/or read operations, per platform type.
- 3.What is the cost of certification and how often does the vendor certification need to be renewed?
Are those certifications required to maintain warranties?
- 4.What are the other costs of training and retaining staff?
- 5.What are the costs of losing and replacing staff?
- 6.The total employment cost (including overheads) of those staff.

Maintenance costs

Most organisations have a pretty good feel for how many man hours will be taken up to maintain their archive using existing tape technologies along with how the people could be better used in other areas of the company. Manual disk or tape archives, and even some automated tape solutions will require:

- Tapes to be labelled, loaded and fully written with large blocks
- Bad tapes to be removed and returned to supplier
- Data to be written (can be cripplingly slow in underspec'd solutions)
- Written offline or online media to be verified regularly
- Tapes be unloaded after write operations
- Removable media to be prepared for secure transportation and generally kept secure at all times
- Tapes to be found and retrieved from storage location(s)
- Tapes to be migrated to new formats (sooner or later, see Obsolescence and Migration)
- Capacity upgrades (see Manpower for Capacity Upgrades)

An article in ComputerWorld magazine by Jim Damoulakis (CTO of GlassHouse Technologies Inc.) highlights the following less well known annual processes an organisation should go through when trusting their data to magnetic tape. Many organisations wont need to do all of these actions but may be avoiding doing some that they should really be doing:

Quarterly

- Review off-site tape storage authorization.
- Review and apply any new software version releases as appropriate.
- Perform high-level capacity planning review of backup storage resources.
- Review performance statistics for each backup server.

Annually

- Inventory and audit off-site tape storage.
- Audit backup software licenses for software license compliance.
- Review and update backup/recovery policies for application environments as appropriate.
- Perform detailed capacity-planning and resource-utilization analysis.
- Apply resource utilization metrics to budget request and forecasting.

Daily

- Review successful and unsuccessful backups.
- Resolve failures before the next backup window.
- Check tape drive status.
- Check media status.
- Confirm catalogue backups.
- Review and act on requests: restores, change management, etc.
- Send vault tapes off-site.

Weekly

- Check scratch-tape inventory.
- Check library slots available.
- Check the catalogue tape inventory.
- Review the off-site catalogue pool.
- Review backup server log files.
- Publish weekly metrics reports.

Monthly

- Check backup catalogue disk space usage.
- Review and apply any new software patches as required.
- Perform random restore testing.
- Publish trending metrics reports.

“Rather than wait until a tape is played back to assess its condition, a mass storage solution and systematically check and validate every single storage device within it according to routines that can take into account usage, media age, batch performance and environmental factors. This level of active media quality checking is beyond the capabilities of all but the most extravagantly funded manually managed discrete data stores, and offers tremendous advantages to archives wanting to ensure the quality of their assets over time.” **PrestoSpace**

A good clustered disk based platform can ensure the longevity, authenticity and availability of assets automatically and at a much lower cost as described here by Michael Way of Hitachi Data Systems:

“In fact TCO numbers from IDC point to the fact that TCO continues to outstrip capital acquisition costs significantly whereas the capital acquisition costs are remaining largely flat. What is going on with the various storage vendors out there when they add all of the Reliability, Availability and Serviceability (RAS) features to storage and innovation on top is that you are effectively paying them for reducing your staff. “

Manpower for Capacity Upgrade

If you need to add 200TB in a single upgrade then the maximum chunk of capacity that the platform provides is key. The management effort of unwrapping at least 200+ tapes and potential unloading tapes from the library to make room for the new capacity versus adding a few nodes to a clustered disk solution and flicking a switch is one aspect, along with the management effort for adding extra slots, or replacing an LTO4 unit with an LTO5 one.

Obsolescence and Migration

When considering the TCO of a storage solution one can clearly see how long a lifespan the disks have (before they start holding miniscule amounts of data compared to newer larger disks) and how long tape formats last for before they become superseded by newer, faster, larger capacity formats and platforms,

rendering the tape collection practically obsolete until it can be transferred to a newer format.

When taking into account a period of 5 or more years this becomes a crucial element to the management costs. PrestoSpace works with national archives from around Europe to lay down the technical challenges in content digitisation and long term preservation. The project has plenty of experience with tape management and states:

- “Technical obsolescence happens rapidly in such an archive and it is often the case that a specific mass storage system (e.g. a particular tape robot and tapes) will be come obsolete within 5 years.”
- “A robot that is kept running for a very long time may eventually become very expensive to keep operational, as replacement parts and maintenance skills become more rare with encroaching obsolescence”

Some disk based solutions will face similar challenges of obsolescence, however clustered storage solutions should be capable of “migration in-place” whereby content residing on older, legacy units of storage, automatically migrate to newer units as they are plugged in or when those older units fail.

This section of the TCO calculations should also look at the disruption that might be caused during the upgrade of firmware and/or software throughout the workflow. I.e., from the client software used to store or view data through to the client access machines and/or data caches (in the case of a tape data cache) through to the archive platform itself.

“Long-term storage and proprietary products don’t mix. Along with upgradeability-in-place, this should be high on customer checklists.” **Robin Harris** www.storagemojo.com

One footnote, there is currently some confusion about the future of LTO:

“According to published plans, the all-conquering LTO tape format has no future after LTO 6, which is expected in 2012. This is ridiculous and there must be a secret roadmap for LTOs 7 and 8” **The Register**.

Though one suspects that will sort itself out sooner rather than later.

Duplication, Replication & Failover

“If you don’t have two copies you do not have it” **Nick Pearce, Object Matrix**

The risk element of keeping just one copy of data is so high that the data can be lost at any moment, and usually the discovery of that loss comes just when you needed to get the data back. As discussed in the “Data Protection Levels” and “Failover” in the Analysis Parameters section of this document and “Failover Facilities” in the Infrastructure section, keeping multiple copies of data is both a must and incurs both capital expenditure and infrastructure costs. In addition to those costs, there are a number of employee time management costs to consider.

The management cost will come down to the level of protection desired at local and remote locations and how that level or protection is achieved. Some options there are:

- Dual Copy Single Location
 - One Robotic tape/optical machine (Tape/Optical)
 - One Cluster, Double Nodes (Clustered Disk)
 - Non-robotic tape machines / USB drives and lots of manpower
- Dual Copy Dual Location
 - Two Robotic tape machines and automated data transfer between the two locations
 - One Robotic tape machine and secure manual tape transfer (no failover facility)
 - Two Clusters and automated data transfer (one or more copies of data at each location)

If full failover (in disaster recovery) scenarios is required then management time to achieve that failover needs to be considered.

Cost to consider in this section include the management cost of:

- Ensuring the multiple copies of the content are authentic and immutable
- Replacing corrupt or missing copies in each location
- Quickly working from remote backup location
- Access to content in remote location

Restoring Content (or Not)

“Archiving is not about storage, it’s about access”

Michael Howard, Chairman/CEO OuterBay Technologies

Backup restore can be measured in RTO (Recovery Time Objectives); the amount of time elapsed between disaster and restoration of business functions, but where the archive is being used as a living library of information, those objectives might better be defined in terms of efficiencies and scalability (in terms of numbers of users) gained of one solution over another.

When the solution must meet certain access criteria it could be that disk is clearly the only viable solution. You couldn’t see online search engines such as Google having to access a tape library for instance, they simply have to have the response times that disk provides across their huge libraries of information. Secondly, time to make the restore of data may impact on the ability of individuals to be able to complete their jobs, and if so there may be an element of cost that should be attributed to that. As with all the parameters in this document, restore requirements will be very much down to the individual organisation, but what will be common is that if data is being stored, then presumably the idea is to be able to get it back again at some point in time and if guaranteed access to legacy content is important then the authenticity of the data stored and the costs of data retrieval become very relevant.

“Boeing ... reportedly figures a cost of \$1 million for every 15 emails that must be retrieved in response to any sort of inquiry. Boeing averages two discovery requests per day from its legal staff, say people familiar with the company.” **Byte and Switch**

Full Workflow Support

No storage technology platform should be an island. Take into account the efficiencies to be gained when it comes to organisations moving content seamlessly through the various tiers of storage they have in place versus solutions that require clunky or manual workflows. Having legacy data available at any stage of a given workflow can be of huge benefit. Costs here may include consultancy, integration effort, licenses and staff training.

Exit Costs

The cost of moving from a platform or format is also a consideration. If you have 10 years of content on LTOx then you are likely to have in excess of 20,000 tapes. Migrating that to another platform may take some time. With a disk based solution that entire process can be automated.

There is also the cost of secure deletion of content. Clustered solutions can have their disks scrubbed by an automated process, what’s more if the solution uses non-proprietary hardware the tin can be re-purposed once the data has been “shredded”. The disposal of the equivalent amount of tape is a very manual process.

Opportunity/Legal Benefits

“On average, for every 12TB of installed and usable disk capacity, there is the potential to find/avoid/recover \$1M of net OPEX savings” **HDS Blog**.

Increased productivity, assets re-use, regulation compliance, fine avoidance, results in an annual value of that data to the company of assets being available online and readily available. Potential new revenue streams can be gained by making the data available via the internet, on demand to customers, or any number of other usage scenarios which benefit from instant random access to files. These are some of the opportunity benefits that can be compared between disk and tape based storage, depending on the usage scenarios.

In terms of find the data required, the old adage of “If you can’t find it you don’t have it” comes into its own. TCO models that attach a “lost opportunity” or “lost revenue” cost to the model, can justify even the selection of otherwise seemingly very expensive solutions. A more conservative approach for many might be to translate prolonged restore times into an “employee expense” which is based solely upon the lost productivity of the affected employees.

The following table is a snapshot of a chart provided by ItCentrix. ItCentrix performed a TCO analysis of HCAP from Hitachi Data systems. Interestingly it takes a view that having data on-line would provide a 10% rise in user productivity. According to GISTICS research the average creative person looks for a media file 83 times a week and fails to find it 35% of the time which would suggest that there are greater productivity gains to be had:

	Traditional Digital Archiving Solution	Hitachi Content Archive Platform
Productivity from Increased Availability (200 Users, 10% Usage)	\$29,250	\$30,225
Improvement in Productivity from Improved Access to data (500 Users, 10% Usage)	\$243,750	\$975,000
Expected avoidance of Claims from Adverse Outcomes	\$182,000	\$188,500
Expected avoidance of Claims from Adverse Outcomes	\$143,000	\$156,000
Total Benefit/Year	\$598,000	\$1,349,725

What this chart does not take into account is the knock on affect of losing data in terms of lost opportunities. Acxiom, a data management company in the USA suffered data loss through theft. After the theft became public, prospective customers took more time than usual in deciding whether to do business with Acxiom (Information Week).

Horror Stories

This section is dedicated to real life stories of data loss by going to tape. It is abashed scaremongering and perpetuating the general feeling of fear when trusting your data to tape and optical devices! Don't hide behind the sofa...

Human Error

In February, Bank of America Corp. lost a tape with credit card information on 1.2 million customers. In April, Ameritrade Holding Corp. told 200,000 current and past customers that a tape containing confidential account information had been lost or destroyed in transit. Time Warner Inc. reported in May that 40 tapes containing personal data on 600,000 current and former employees had been lost en route to a storage facility. In June, Citigroup Inc. said that a box of tapes holding personal information on 3.9 million customers had disappeared on the way to a credit bureau.

Tape/Optical is Error Prone

Data on tape is notoriously error prone. Recent press reports have spoken about lost tapes (Bank of America, CitiGroup, Time Warner, Iron Mountain) and it is fairly obvious that where human intervention is required, accidents can happen, e.g., tape and optical platters can easily be dropped.

Billion Dollar Fines

E-mail played a central role in a \$1.58 billion judgement against Morgan Stanley and in favour of Ronald Perelman, the billionaire investor who said he was defrauded by the Wall Street company over the sale of a business and focused on the firm's inability to produce documents.

Backup Failure

An article in InfoStor Magazine (August 2005) states that 85.2% of the people in their survey have backup problems to tape, largely due to equipment or tape media failure.

<http://www.eweek.com/article2/0,1895,2054052,00.asp>

Q&A: The founder and digital librarian of Internet Archive reveals storage challenges and solutions.

Q. With tape, were you horribly input-output bound?

A. You can just stop at "horrible." There's almost nothing nice to say about tape.

Conclusion

Choosing a archive architecture to protect and serve your organisations valued digital information is not an easy task. There are many factors that will change the conclusion according to your needs. However, the exercise is well worth carrying out with care and attention since selecting the best solution could not only save your organisation many multiples of the actual cost of the solution, but could even result in revenue earning streams for your organisation's digital assets.

Whilst tape and optical solutions have vastly improved over the years, especially of the robotic genre, it is clear that disk based solutions are not only immediately relevant for archiving but also improving in their capabilities at a faster rate. If current cost and usage ability trend lines are followed, then it is unlikely in the not too distant future that there will be any call for tape based solutions at all. Tape is not dead: it is highly relevant in many data storage scenarios, but once whereas it was the king of the data storage mountain, it is now heading towards being a bit part player. Indeed it was reported that back in 2006, there was more data backed up on disk than on tape (source: PeriConcepts).

In promoting disk Object Matrix would ask:

- Would you like to see an increase in productivity by putting information at the fingertips of your employees?
- Would you like your staff to be working on core competencies more relevant to your business than shuffling tapes around?
- Are you legally required to keep certain types of information for a given period?
- Would you like to increase your chances of avoiding possible legal ramifications of not finding an item of information that you are legally or morally required to protect?
- Could you make use or generate revenue from having all your digital assets available on-line?
- Is business continuity important to you?

But perhaps, most importantly, we'd ask you to perform an in-depth TCO analysis that goes beyond superficial "upfront" costs, and that delves into a realistic, predictable but holistic impact of the data storage choice for your organisation.

References

MatrixStore from Object Matrix Ltd (www.object-matrix.com)

Avamar Axion TCO white paper

Plasmon - Archival Storage TCO Analysis Summary

ITCENTRIX white paper - Hitachi's New Generation Active Archive Platform

Tape Loss - <http://www.computerworld.com/securitytopics/security/story/0,10801,103541,00.html?SKC=storage-103541>

InfoStor Magazine

Network World

Billion Dollar Fines (ComputerWorld.com <http://www.computerworld.com/hardwaretopics/storage/story/0,10801,108687,00.html>)

Lost Opportunities (Information Week <http://www.informationweek.com/security/showArticle.jhtml;jsessionid=E5VMLUFPT0540QSNLQSKHSCJUNN2JVN?articleID=183700367&pgno=2&queryText=>)

Tape vs Disk costs (SearchStorage.com http://searchstorage.techtarget.com/expert/KnowledgebaseAnswer/0,289625,sid5_gci1179948,00.html)

Tape vs Disk quote (itbusinessedge.com <http://www.itbusinessedge.com/item/?ci=5289>)

More Disk than Tape (Periconcepts.com <http://www.periconcepts.com/Reports/2006%20Tapes%20vs%20Disks.html>)

Tape Process (ComputerWorld.com <http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=storage&articleId=9001451&taxonomyId=19>)

More tape process <http://esj.com/Articles/2009/05/12/Tape-Zombie.aspx?Page=4>

PrestorSpace. <http://digitalpreservation.ssl.co.uk/storage/T2/T2-1-2-4.html>

http://www.auststor.com.au/download/resources/archival_storage_tco_datasheet.pdf

http://blogs.hds.com/david/2006/04/storage_economi.html

http://www.channelregister.co.uk/2009/09/17/lto7_and_8/

http://www.byteandswitch.com/document.asp?doc_id=65643

<http://storagemojo.com/2009/10/10/nightmare-on-dimm-street/>

<http://www.datacenterknowledge.com/archives/2008/10/14/google-raise-your-data-center-temperature/>

<http://www.ravica.com/blog/data-center/yahoo-powers-data-center-using-niagara-falls-google-raises-data-center-temperatures/>