Available online at www.prace-ri.eu

Partnership for Advanced Computing in Europe

Media and Technology Appraisal for Long Term Preservation

Florent Marceteau, Olivier Rouchon, Johan Raber, George Tsouloupas

(CINES, SNIC, CaSToRC)

Abstract

Reliability, performance, costs and return on investment are key factors in the long term preservation of digital data. They differ from one technology to another. The different media and technologies used for storage and transfer will be compared, with a particular focus on disks and tapes.

1. Introduction

Long-term preservation technologies are constantly changing. For the long-term preservation of data, this can be a disadvantage that needs to be managed. But it may also lead to advantages when it is possible to make the most of it. It is necessary to anticipate and adapt to the various technological advances to enable good data preservation. The aim then is to enable the data to be preserved physically, but also read and understood. Luckily, by using some of these advances, it is possible to improve some aspects of long-term preservation.

In order to keep data for a long time, reliable media must be used as a basis. Loss of data is not an option. Unfortunately, no support is 100% reliable and thus the maximum must be done to limit and anticipate equipment failures. This can be done by using technologies to anticipate failures, and by setting up a suitable environment.

Installing and maintaining such equipment represents an investment in both human and economic terms. Indeed, the costs of such equipment are not negligible. For this reason, it is useful to choose equipment that best meets the needs of the users. This then enables expenditure to be targeted better.

2. Reliability

2.1. Environment

Whether it is for hard drives or tapes, the environment plays a key role in long-term equipment reliability.

Temperature, humidity, dust and electromagnetism represent different factors able to affect disks and tapes. For this reason, it is necessary to properly plan the rooms where the storage equipment will be located. Disks produce heat and will have to be located in a sufficiently air-conditioned room. Transport of these media may also present risks, in particular if the previously mentioned factors are not taken into account. Although some technologies may limit their impact, shocks are particularly risky, especially for disks. In unsuitable conditions, error percentages of up to 10-15% after one year and 40-45% in 5 years are possible [1]. Hence the importance of properly taking accounts of these various factors. Temperature is a more important constraint for disks due to the constant heat given off by the equipment.

A study [2] shows that the temperature at the surface of the disk affects the risks of failure. With a manufacturer specified temperature between 30°C and 40°C at the disk surface, it is preferable not to allow the temperature to increase. At temperatures of over 45°C, the risk of disk failure doubles after three years. Too low temperatures are not desirable either; the risks also increase for temperatures that are too low.

2.2. Age

For long-term preservation, the equipment must be able to last for this "long term". Equipment ageing must therefore be assessed to be able to anticipate failures and thus replace the equipment before the risk of losing data becomes too high.

Tapes are relatively reliable over time. Ageing is only a concern if the media or tape drives are improperly maintained. In good conditions, very high reliability can be obtained, with 99.945% of tapes aged 2 to 12 years are fully readable (with 32.7% of tapes aged 12 years) [3].

Hard drives however, are far more sensitive to the passage of time. As they are constantly powered, even when not in use, some components are always active. For this reason, disk wear is inevitable.

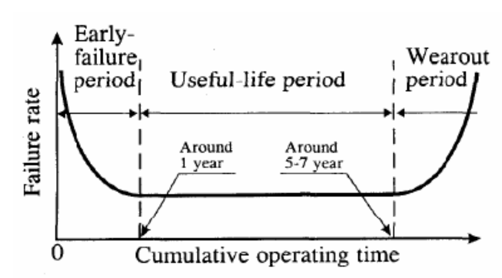

Figure 1: Lifecycle failure pattern for hard drives [4]

This is not a constant phenomenon. On the contrary, wear occurs in the form of a bathtub curve. Thus, several steps can be observed in disk wear, with an error percentage curve forming a bathtub shape. The failure rate is high at the start and end of the equipment life cycle. Several studies describe this phenomenon.

Considering this model [4], the first year is characterised by early errors. The failure rate remains constant in disks from 2 to 5 years of use. After this period, wear on the disks causes increasingly frequent failures.

This curve remains theoretical and indicative. Another study [2] also shows these results:

Thus there is an infant mortality failure over the first 3 months, but this is far lower than normal wear on the hard drive which appears after several years. It can also be noted that the error rate in years 2 to 5 is not influenced by age. This confirms the previous diagram with a failure rate relatively constant over the first years.

These two models present the infant mortality phenomenon which exists over the first period of use of hard drives. However, in actual fact, this risk is not dominant compared to hard drive failures requiring 6 to 8% of them to be replaced every year.

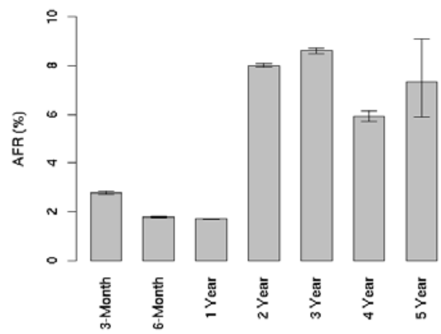

Figure 2: Annualized failure rates broken down by age groups [2]

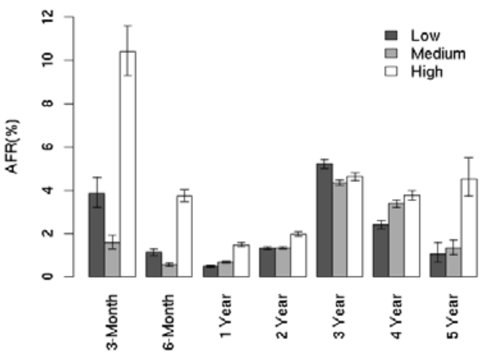

Figure 3: Utilization [2]

More or less intensive use of hard drives has an impact on reliability. This is shown in a study [2] presenting the change in several populations of hard drives classified according to their mean read/write time.

The mean usage time of hard drives does not have as obvious an influence as we might expect. Intensive use of hard drives has an impact on long-term reliability, with almost five time greater risk of failure after five years of use. The infant mortality rate is also globally higher if the hard drives are used more.

However, the behaviour observed in the intermediary years is not as obvious.

There is an unexpected result in the 3rd year with higher error rates for the least used drives. This could be explained by the low infant mortality rate.

2.3. The risks

As presented previously, there are several risk factors.

Tapes may become unreadable, partially or over the whole tape. The data must therefore be recovered from another backup copy. But this should be kept in perspective. According to a study by the Enterprise Strategy Group [3], of the 24000 tapes they verified after a minimum 2 years of use, 13 were entirely unable to be read. In suitable hardware conservation conditions, reliability of 99.945% can be obtained. On the 13 tapes in question, certain blocks were no longer readable, thus reducing the risk of loss of data.

Within the CINES, we have also noted that these failures may occur from the first use.

The hard drives may present data corruption problems. Sometimes, this corruption may be silent, as when some data are incorrectly copied or recopied in the wrong place. Cyclical redundancy errors may also occur. There are many possibilities for data corruption. The RAID system can be used to overcome most of these errors by offering the possibility of recovering data if necessary.

2.4. SMART technology (Self-Monitoring, Analysis and Reporting Technology)

Hard drives are complex equipment, with numerous components that can cause failure. For this reason, for preventive reasons, it is preferable to check several indicators to anticipate the risks better. SMART technology (Self-Monitoring, Analysis and Reporting Technology) allows for self-monitoring of hard drives, in particular by means of error scan parameters, and memory reallocation counts. According to a Google study, although these parameters are relevant to determine a certain error probability, they are not accurate enough to be used as a prevention tool alone.

Nevertheless, they enable advanced analysis of hard drive wear.

When drives scan the surface of disks, errors may occur. SMART technology keeps a visible trace of these errors – good indicators of the reliability of a hard drive. Hence, the failure risk is ten times greater on drives containing errors. It is also useful to specify that the more errors there are, the higher the risk. One study [2] also shows that the risks of failure within sixty days following the first error are thirty nine times higher than for others.

Sometimes the drive detects sectors with read, write or check errors. In this case, the data are transferred to another location. These reallocations are counted in the SMART data and can be used to know how many times the problem has occurred. Thus, the more the problem occurs, the more the drive had to reallocate and the greater the risk in continuing to use it. Reallocations are therefore another variable that can be used to assess drive ageing. The failure risk after an initial error is also very high, with an almost fourteen times greater risk of failure in the next sixty days.

The above-mentioned reallocations mainly concern read and write phases. During disk checking and cleaning phases, there may be sectors with allocation problems. The procedure is the same as previously and the data are reallocated. These offline reallocation errors are a sign of wear on the drive. The risk of failure is then twenty one times higher within the next sixty days.

In general, these data concerning drives do not enable all drive failures to be anticipated. However, in order to preserve data over a long period of time, the risks must be anticipated as far as possible. SMART data can be used to take account of the various factors signalling drive wear and may provide early indications of failure. If preservation on hard drives is chosen, and a highly reliable system is required, it is preferable to use this information.

3. The costs

The costs differ between use of hard drives or tapes for storage.

With CINES installations, given amortisement over 5 years for hard drives and 10 years for tapes, and with current equipment costs, for quantities of data of less than 70TB, hard drives remain less costly for the moment.

However, over 70TB, the cost of the infrastructure required to use tapes is offset by the difference in price between tapes and hard drives.

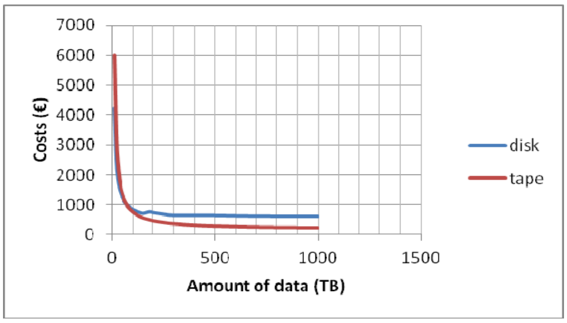

Figure 4: Costs of tapes and hard drives according to storage capacity up to a PB

If we take a closer look at the threshold, close to 70 TB, we obtain the diagram below:

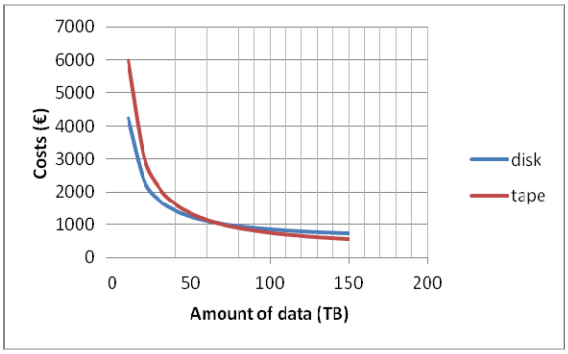

Figure 5: Costs of tapes and hard drives according to storage capacity

Tapes require suitable infrastructures which, although costly for a reduced quantity of data, will be offset by the large quantity of data to be preserved.

These data correspond to the equipment used at the CINES, but the equipment in question is not only used for sustainable data archiving. Other needs such as data storage have been taken into account, enabling faster access to the data.

This diagram provides an overview of the costs of storage, but does not take account of parameters such as air conditioning, equipment contracts and labour.

Moreover, the equipment used will depend largely on needs. For tapes, it is possible to use only one reader, which reduces costs, but increases data access times.

At CINES, the DMF license is used for both disks and tapes, with an initial cost representing 40€/TB.

SNIC NSC has got a comprehensive and rather long record of the costs associated with magnetic tape storage, the currently cheapest back-end storage solution for large amounts of data. The record dates from 2006 and covers three generations of tape storage technology, from Ultrium 3 through Ultrium 5.

The tape storage costs can be compared to NSC's disk based storage. Current cost figures on disk storage at NSC are based on cost of purchase, administration and other running costs such as electricity, cooling and software licenses. Disk cost data comes from a breakdown of the costs for NSC's latest installation of large Lustre file systems and dCache storage pools consisting of 12 2TB disk Dell PE R510, in total amounting to ~2 PB worth of storage. For comparison purposes, they put the disk storage data point in the most favorable light, since they were the most cost effective disk based solutions to date in terms of price per TB, as determined through the vendor bidding process. In essence it boils down to disk being in the same order as the cost of tape based storage. The costs are detailed in table 1 below:

Table 1. Cost comparison of disk and tape storage.Unit is €/TB/Yr.

Storage medium	Purchase cost	Licenses	Floor space cost	Electricity and cooling	Yearly cost (total)
Disk	*55*	-	*2.7*	*15*	*72.7*
Tape	41	7	0.44	1.5	50

There is a possibly large degree of variability in the calculation of purchase costs as spread over the life time of the storage medium. We write off the storage medium over a five year period for tape and four years for disk. Since tape media end-of-life is highly depending on the number of overwrites, other usage patterns like for instance “Write Once Read Many” type of usage may extend the life time of the media, thereby potentially changing the yearly cost significantly, say an order of magnitude in favor of tape media. The write off period is included in purchase cost above. The TB count above is based on effectively usable space, for tape we count on a 90 % tape fill grade while for the Dell servers ~20 TB ends up as usable file system space, serviceable up to 100% fill. Personnel costs have been considered equal for disk and tape in table 1.

At CaSToRC, there is no tape yet, but the costs are detailed in table 2 below:

Table 2. Cost comparison of disk and tape storage.Unit is €/TB/Yr.

Storage medium	Purchase cost	Licenses	Floor space cost	Electricity and cooling	Yearly cost (total)
Disk	*53*	*26*	*9.6*	*25*	*114*

In general, it can be noted that for storage, the price per TB is lower for tape than for hard drive, whether for the equipment purchase or energy costs.

Other factors can be taken into account when evaluating costs. The space available to install the equipment and the maintenance required. The price of equipment changes and depends on needs. For example, it is possible to have tape readers that are inexpensive but that have especially slow reaction times. Inversely, if we need a fast tape reader, the price will be much higher.

According to a study [5], with large quantities of data to be preserved over the long term, the least expensive support is undoubtedly tape. Over a period of twelve years, with a large quantity of data of the order of a PB or more, the price of tape readers is largely offset by the quantity of readable tapes. Thus, although the initial price of tape drives is nearly fifteen times higher than hard drives, they allow for considerable savings when compared to hard drives. Hard drives require much more energy to operate well than tapes. Considering the price of energy alone, the cost difference is of the order of 238 times more expensive for hard drives than for tapes.

4. Conclusion

The tape would therefore seem especially indicated for long-term preservation of data due to its reliability and low cost. However, one of the aspects not dealt with concerns data access time. With reaction times of the order of few seconds for tapes, and almost instantaneous for hard drives, if data needs to be used frequently by users, it is better to place them on a hard drive.

In general, for sustainable data preservation, it is preferable to make several copies of the data on several different media. For this reason, where possible, the data should be kept both on tape and on hard drive. Technologies for anticipating the risks of various supports will evolve over the coming years. It is therefore essential to continue monitoring the various technologies to improve different aspects of sustainable data conservation.

5. References

1. Nexsan Technologies Incorporated, "Making Cents of Tape vs. Disk", 2010
2. E. Pinheiro, W. D. Weber, L. A. Barroso, “Failure Trends in a Large Disk Drive Population”, Google Inc, 2007
3. M. Peters, "NERSC: Proving Tape as Cost-Effective and Reliable Primary Data Storage", Enterprise Strategy Group, 2010
4. B. Schroeder, G. A. Gibson, “Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you? “, Carnegie Mellon University – Computer Science Department, 2007
5. D. Reine, M. Kahn, "In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantage over Disk", The Clipper Group, 2007
6. E. Pinheiro, W. D. Weber, L. A. Barroso, “Failure Trends in a Large Disk Drive Population”, Google Inc, 2007
7. L. N. Bairavasundaram, G. R. Goodson, B. Schroeder, A. C. Arpaci-Dusseau, R. H. Arpaci-Dusseau, “An Analysis of Data Corruption in the Storage Stack”, University of Wisconsin-Madison, 2008

6. Acknowledgements

This work was financially supported by the PRACE project funded in part by the EUs 7th Framework Programme (FP7/2007-2013) under grant agreement no. RI-211528 and FP7-261557.