Arkivum
Every bit archived

ArkivumLimited
R21 Langley Park Way
Chippenham
Wiltshire
SN15 1GE
UK

+44 1249 405060

info@arkivum.com
@Arkivum
arkivum.com

How Green is Your Archive?

Whitepaper

Introduction

Businesses looking to reduce the amount of energy and money consumed by archiving often turn to datacentres for their storage needs. However, datacentres themselves are facing a growing strain on the amount of electricity they use. Research by data centre think tank, The Uptime Institute[1], reveals 36% of the 525 datacentre owners and operators surveyed expect their facility to run out of power, cooling and/or space in 2011-2012. Although worrying, this will come as no surprise to many, with Google[2] calculating that server power plus associated Power Usage Effectiveness (PUE) overhead can be more than 20% of the total cost of running servers in a datacentre. PUE is a measure of data centre efficiency, and the long-term cost of data centre power and cooling can be more than the cost of the server equipment itself.

The benefits of LTO

For archive data, LTO data tape can offer a much more cost and energy efficient alternative to powered hard disk servers. This is why we use data tape at the core of Arkivum's data archive service. Data tape has the huge advantage of only needing electricity to run the tape drives when data is being accessed. If data is not being used then tapes simply sit in their slots in a library needing no power at all. Typically, all that is needed to serve the needs of a petabyte of archive data is just two or three tape drives and a few hundred watts of power – no wonder that data tape is often called a 'cold storage' technology.

Disk versus tape

To illustrate this point, several independent reports compare the energy efficiency of data tape with other approaches, such as hard drives in servers - often called 'spinning disk'. One of the most comprehensive is the 2011 Information Storage Industry Consortium's (INSIC) Magnetic Tape Storage Roadmap[3], which includes INSIC's own research into tape compared to spinning disk, as well as comparable analysis done by The Clipper Group[4]. The Clipper Group examined what proportion of the Total Cost of Ownership (TCO) was taken up by the energy needed to run a data archive over a 12 year period. It estimated that the energy costs for running a hard drive solution accounted for 12.6% of the TCO, while for a tape-based solution this was only 0.4%.

Additionally, Clipper highlighted that as the TCO for hard drives is 15 times greater than that for data tape, the money spent on power for the hard drive solution is actually 238 times more than data tape. Indeed, Clipper calculated that the cost of power alone for a hard drive solution over 12 years would be more than the entire cost of using data tape.

There is evidently a wide gap between the energy consumption of spinning disk archives and those on data tape. Those that run hard drive storage in datacentres for archive applications are attempting to bridge this chasm by using an approach where hard drives are 'spun down' when not in use. This is often called MAID (Massive Array of Idle Disks), but making this work with as little energy consumption as data tape is extremely challenging.

Unlike data tape, hard drives need to be in arrays to make them robust against failures. When a data file needs to be accessed it requires not just the ‘spinning-up’ of one disk but many, therefore increasing power consumption. Additionally, hard drives are by their very nature not designed to be left 'spun down' for long periods of time.

The solution to this is to 'spin up' the drives on a regular basis, check the integrity of the data on them, and then send them back to sleep - all of which uses energy. On the other hand, data tape used for archive can be left for a year or more before it needs to be checked. So, while 'spun down' or 'sleepy disk' approaches do have many benefits, when they are used for actively accessed archive content or where a very high level of data safety needs to be assured, they can use more power than one might think.

Options to reduce power consumption from hard disks include reducing spindle speed; stopping drives from spinning; or powering them off completely, which can create energy savings of anything between 15 and 85%[5]. While this is impressive, even in a generous scenario of MAID being 10 times more energy efficient than constantly spinning disk systems, the power consumption is still several times higher than data tape - and the technology can cost significantly more too.

De-duplication in an archive

Another argument presented in favour of hard drives is that de-duplication technology can be used to reduce the amount of data being stored. This can work really well for backup applications, but de-duplication is not recommended inside an archive. For example, block level de-duplication, which offers the highest storage savings, also makes data far more complex to reconstitute back into a useable form. This typically requires proprietary software and systems - not good for the long-term safety of data, especially when seeking to achieve portability and avoiding vendor lock-in. That is not to say that putting duplicate files into an archive should not be avoided - but this comes down to good data management by the user rather than the need for complex technology.

But suppose you did use de-duplication in an archive - How much energy would this save? The Enterprise Strategy Group (ESG) concluded that backup to hard drives requires significantly more power and cooling than a tape-based solution. When using Virtual Tape Library (VTL) technology to make hard drives look like data tape, even when de-duplication is used to lower number of drives

needed, the power requirements are still 38% higher[6]. For archive data the difference would be greater still.

MAID and de-duplication do offer the opportunity to use less energy and less storage space, but they have yet to achieve the energy efficiency of data tape in archive applications. Data tape has added benefits of being a comparatively simple, reliable and low cost technology. Data tape might not have the same very low latency as spinning disk, but that isn't a problem in most archive applications – and when you do get to the data you want after a minute or two waiting for the tape to load, retrieval speeds can be very high. Overall, data tape is a compelling choice for long-term archiving of data, and that includes archive data that gets used on a regular basis.

Summary

Disk of course has its place in archive applications. Fast random access and easy scaling make it ideal for caching or short-term storage, especially for data that is being accessed very frequently. An effective archive solution will typically require a combination of both disk and tape in a tiered solution – and a way to ensure that the right data ends-up on the right tier at the right time. It would be unfair to dismiss disk altogether and indeed at Arkivum we use disk servers in our service – but only for a small proportion of the data we store.

Data tape for archiving can be a great choice from both an economic and green perspective, and this is exactly what Arkivum has done!

About Arkivum

The company was launched in 2011 following a four year research project at the University of Southampton IT Innovation Centre. Partners in the project, which benefitted from UK government funding, were BBC R&D, Ovation Data Services, University of Edinburgh, Xyratex and IT Innovation. The aim was to develop a new approach to planning and managing large- scale, sustainable and integrated digital archive solutions.

Arkivum now has established data centres in the UK and is working with clients from a number of industry sectors on archiving solutions which provide a 100% guarantee that all data can be recovered at any time.

[1] http://uptimeinstitute.com/publications
[2] http://www.morganclaypool.com/doi/pdf/10.2200/s00193ed1v01y200905cac006
[3] http://www.insic.org/news/A&S%20Roadmap.pdf
[4] David Reine and Mike Kahn (2010) In Search of the Long-Term Archiving Solution - Tape Delivers Significant TCO Advantage over Disk. Dec 23, 2010. Report from the Clipper Group. http://www.lto.org/pdf/2010_December_Archive%20TCO.pdf
[5] http://www.nexsan.com/library/automaid.aspx
[6] Mark Peters (2011). A Comparative TCO Study: VTLs and Physical Tape. With a Focus on Deduplication and LTO-5 Technology February, 2011. Enterprise Strategy Group (ESG). http://www.lto.org/pdf/ESG_WP_LTO_TCO.pdf