

# STORAGE SWITZERLAND

## UNSTRUCTURED DATA TAPE ARCHIVING EFFICIENCY



Colm Keegan, Senior Analyst

### Dealing With the Unstructured Data Deluge in Higher Education

Colleges and Universities have many of the same issues with efficiently protecting critical data archives as private corporations but arguably, their challenges are even greater. While IT budgets have typically remained flat for private enterprises, some industry sources estimate that state funding for higher education has been reduced by as much as 28% in the past five years. This dearth of state funding combined with the ongoing explosion in the growth of data, is forcing IT planners in academia to find solutions which can both cost effectively manage long-term data archives and provide users with fast access to information.

#### **Big Data Conundrum**

It has been well documented that the growth of unstructured data now accounts for the majority of all new data creation; this is certainly also true for higher education. Sources of unstructured data include machine created data, user files, PDF images, power point documents and audio/video files.

For higher education, the data deluge is especially complex as there is a broad range of data intensive applications to support. From digital learning courses and research projects to video surveillance systems and high

performance computing applications that consume big data; unstructured data repositories throughout academia are vast and continuing to increase. The vast majority of this data is infrequently referenced again after it is initially created but end users still need to have the ability to quickly access this information on demand.

As data grows on primary storage systems, additional pressure is placed on other resources downstream. Backup jobs, for example, start to become elongated and bandwidth connections begin to become saturated as backup data volumes increase; potentially disrupting other sources of data traffic in the environment. With higher education IT budgets becoming increasingly constrained, adding capacity or upgrading existing primary storage and backup systems to accommodate this growth is difficult and at times not an option.

Compounding this problem is the fact that the various academic departments within college and university environments are all vying for the same IT budgetary dollars to meet their needs. Consequently, infrastructure planners in academia need to find very creative ways to stretch their funds to address a broad range of end user storage and data protection requirements that can address all parties concerned.

## **Cloud Relief?**

Some Universities may be considering public cloud solutions which offer “cheap and deep” storage capacity as a way to efficiently archive their inactive information. However, there are significant challenges, first is that applications need to be able to support cloud protocols to interface with their storage technology—introducing cost, complexity and delays with implementing a solution.

Secondly, while some cloud storage providers offer attractive “teaser” rates for archiving data in their cloud, these costs will quickly begin to increase as data grows; making this option less attractive as a long-term storage archive solution. Essentially with the cloud, the institution has to pay for storage all over again every month. While there may be a large upfront expenditure with purchasing physical storage outright, over time it is still less expensive compared to the cloud model. Retrieving this information could be painfully slow as it would more than likely be traversing network links with limited bandwidth.

## **Disk-on-the-cheap?**

Another option for storing archive data is low cost disk like SATA. There are multiple suppliers that offer multi-PB, direct attached or SAN attachable storage enclosures that can be very cost effective on a cost per GB basis. Moreover, these enclosures can be quickly racked, powered up and provisioned out to satisfy growing application data stores. The challenge with this approach, however, is that rotational disk media requires a significant amount of power and cooling. The year-over-year costs of powering and cooling potentially thousands of disk drives would be prohibitive and consequently, not an ideal approach given the budgetary challenges in academia today.

## **Cache Powered Tape**

What is needed is a solution that can leverage the economies of scale of a low cost storage medium, like tape, to efficiently archive Petabytes (PBs) worth of information. To satisfy application service level agreements (SLAs) however, the solution should also be capable of providing rapid access to data retrieved from the tape archive through an efficiently sized front-end solid state storage cache.

In this manner, as data went from an inactive to an active state, it would get retrieved from the tape archive and immediately promoted up into the cache so that users would have rapid access to this information. The first access would be at tape speed but subsequent accesses of the same file would be at the speed of SSD.

## **NAS Converged Tape**

An ideal architecture to implement this solution would consist of a Networked Attached Storage (NAS) appliance combined with a fully integrated intelligent back-end LTFS (Linear Tape File System) tape library. The NAS and LTFS combination makes for an ideal archiving platform because unstructured data is typically stored as files and both technologies natively support file system storage access. NAS is also an easy storage technology to deploy since it can plug into existing ethernet network switches and present a CIFS or NFS network share to allow connecting servers to access unstructured files.

The open nature of an LTFS tape storage repository is also beneficial since data is stored in its native format. Prior to the introduction of LTFS, tape archives could only be referenced by the proprietary backup or archiving application that had originally written the data to tape. With LTFS, any application loaded with the LTFS driver is capable of accessing the information in the tape archive. This is a critically important capability since data archives need to be accessible far into the future. Keeping the archive in an open format helps make this possible.

From a user perspective, the process for accessing data would be completely transparent. If the file is on primary storage, it will be served up from that repository. If it is on tape, there will be an initial delay while the data is restored to the NAS disk cache, however, it will not require a separate process to retrieve the information. And as stated above, subsequent access to the restored data will be at cache speeds. These solutions could still play a role in cloud storage solutions providing a hybrid approach where data is not totally dependent on the cloud.

### **tNAS Fueled ROI**

This type of approach, also referred to as a tNAS (for tape NAS), would enable IT planners to implement a tiered storage strategy to reduce primary storage costs. By migrating or copying inactive data from primary storage systems to the tNAS via simple file systems, infrastructure managers could free up primary storage space and potentially defer the need for additional disk storage purchases. The resulting savings produced from freeing up primary storage assets alone, could make for a quick return on investment (ROI) of the tNAS.

In addition to deferring costly storage expenditures, there could also be savings on disk power and cooling and a reduction in the amount of data center floor space required to house additional disk storage systems. In fact, when these operational overhead reductions are taken into account over a multi-year timeframe, the collective savings could be quite substantial; particularly for those environments supporting PB's worth of information.

Lastly, with LTO-6 tape technology's ability to store 2.5TB's of data natively on a single cartridge, academic institutions could build a highly cost effective tape archive to satisfy their multi-PB unstructured data archive needs far into the future. What's more, this archive would obviate the need to run separate backup processes for unstructured data once it migrated to the archive; helping to reduce backup traffic induced network latency and delivering additional savings in backup software licensing costs. All of this combines to a significant reduction in total cost of ownership while improving the standard of protection.

### **Conclusion**

Storage archiving technologies, like the StrongBox offering from [Crossroads Systems](#), are making it financially feasible for budget strapped higher educational institutions to cost effectively manage their growing unstructured data storage repositories. By combining the efficiencies of a highly dense tape storage archive with a front-end NAS disk cache, academic IT organizations can implement a tiered storage infrastructure which seamlessly plugs-in to their existing primary storage environment to reduce ongoing storage capital and operational expenditures.

### **About Storage Switzerland**

**Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://storageswiss.com/>**

*Copyright © 2014 Storage Switzerland, Inc. - All rights reserved*