

STORAGE SWITZERLAND

HOW TAPE CAN FIX THE UNSTRUCTURED BACKUP PROBLEM



George Crump, Lead Analyst

The fastest growing data set in just about every organization today, and typically the largest, is unstructured data. This is data that's outside of a database; essentially file data, stored on file servers or Network Attached Storage (NAS) systems. The unprecedented growth of unstructured data combined with its increased importance has created an overwhelming data protection challenge that is crippling most backup processes.

What is Unstructured Data

Unstructured data is most commonly thought of as files from office productivity applications that users create and collaborate on. This type of data is rapidly growing both in size and volume as users begin to embed rich content like video and audio clips into their documents. But unstructured data is now far more than just office productivity documents, this data now includes stand alone rich media, architectural or engineering drawings as well as scanned images of paper documents. For many companies this data is the company's product and for others it's an essential part of the company data set.

Why is Unstructured Data Growing

The growth of unstructured data is being caused by several factors which are brewing to create a disastrous storm for data protection professionals. First, there is an increase in the sheer volume of data. Data rarely starts out or ends up on paper anymore. Most data is created, modified and then stored for safe keeping - all digitally.

Additionally, almost every task, process or device now creates some kind of data. For example, smartphones and tablets are used to create documents. Notes are taken electronically and pictures are often embedded into those notes but stored separately. Machines all have sensors on them that create data about the environment that they are working in or the work they're performing. All of this information is typically stored as file data and needs to be protected.

But it is more than just the size and amount of the data that's causing unstructured data management problems, it's also the length of time that this data needs to be stored and remain accessible. The need to retain this information is not just compliance driven, although that is certainly a contributor. Unstructured data has a growing level of importance to the organization as a whole. It may be mined to help with decision support or even be 'monetized' - used for future business opportunity.

Unstructured data needs to be protected and retained to the same standards that structured data sets like Oracle and other databases are. This means that frequent backups have to be successfully completed and the information moved off-site. It also means that information needs to be stored in an easy to access format so it can be available when analysis needs to be performed.

How is Unstructured Data Breaking Backup

It is now easier to protect a database environment (structured data) than it is a file system full of unstructured data. The database, while important, typically consists of only a few files that are individually large. Databases also have built in backup procedures as well as customized applications that allow for controlled, online, consistent backups.

A file system is just the opposite. First, it is generally full of millions if not billions of small files. The files often vary significantly in terms of size and importance. The file system itself typically has no built in procedures that allow for controlled backups of these files.

This means that a file system has to be manually scanned each backup session for data that has changed since the last backup. Called a "file system walk" this process can take longer than the actual transfer of data to the backup device.

Large file systems full of unstructured data also create problems for the devices designed to optimize the backup process, such as purpose-built disk backup appliances (PBBA). PBBAs are often designed to leverage deduplication, but many large unstructured data sets are full of data that doesn't deduplicate well (or at all). First, the data is often unique and net new. The result is that since these systems can't benefit from deduplication, the capacity of these devices needs to grow at a significantly faster rate (as much as 5X) than the primary storage systems they are designed to protect.

Even if a file is only changed slightly, a subsequent copy of that file will appear to the PBBA as a net new file. As a result storing the backups of unstructured data in disk pools is often inefficient. Finally, many of these file formats are pre-compressed (Microsoft Office, Video formats, Audio formats) so even compression won't help.

How to Fix Unstructured Data Backup with Tape

It may come as a surprise to some but one of the best solutions for fixing the data protection problem is to leverage tape technology. It's inexpensive per GB, fast and can require little to no operational cost to store for a long period of time. The ideal solution would be to merge tape into a NAS based platform so that protection becomes an integrated part of unstructured data storage. This is a technique that [Crossroads Systems](#) has brought to market with its StrongBox solution.

These solutions essentially create a NAS with a tape integrated backend. A Tape Integrated NAS (tNAS) solution looks to other applications and servers like a CIFS or NFS file share. Initially, data can be copied from the main file servers or NAS systems using built-in, intelligent data copy solutions like rsync or robocopy to the tNAS. These applications can be set to run as frequently as the data owners want. Once on the tNAS data can be managed by policy.

For example, Crossroads StrongBox is an integrated and seamless disk/tape solution. Its software has the intelligence to manage disk as a cache to tape. Essentially data is automatically copied to tape in a near continuous fashion. This removes the need for a backup window as well as the aforementioned problems with walking the file system. Since the backup process is now integrated with the file system itself it's instantly aware of which data has been protected and which has not. A process can also be run to create a second copy of each file on a second tape so that it can easily be moved off site for vaulting.

Despite this movement of data from disk to tape, the user sees no impact. All data appears as if it were still on disk. They access the files as they always have, meaning no changes in workflows or applications. If the requested data is still actually on disk it can be accessed directly from that mount point, no need to copy it back to the primary NAS. If the data is on tape only, it is automatically restored. While the user has to wait a few minutes for that restore to complete, they don't need to go to a special interface to trigger a separate restore process.

As confidence builds in the tape integrated NAS the solution can be used to protect databases and other forms of data too. Databases could use the tNAS area to dump transaction logs to during the day as well as for making database copies at the end of the day. Finally, the use case can be expanded again to have the tape integrated NAS store an increasing amount of unstructured data that is on primary storage. Essentially, it becomes the ultimate, automatically protected, tier-2 NAS.

Conclusion

Tape integrated NAS solutions like those from Crossroads Systems solve several unstructured data management problems. First, data no longer needs to be protected as part of a separate process. Data can be copied to the tape integrated NAS with standard operating system utilities and the tNAS takes care of the rest. Second, it can eliminate the capacity growth of the secondary disk backup appliance further reducing costs. Finally, it can become a destination for database backups as well as grow into a primary storage area for tier-2 file data.

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