In Search of the Perfect Archive Solution

The computing world has been using some form of tape backup for more than 30 years (longer in the mainframe world, but that's a different creature). From 12" reel to reel to quarter inch cartridges, 4mm DAT, DLT, AIT, VXA and then today's LTO devices (among many others), tape has long been used for backup and archival to store and retrieve data. Most admins refer to tape-based storage as offline storage since the data can't be accessed directly from the tape.

For many years, specialized software developers have worked very closely with tape drive manufacturers and the ANSI T10 committee dedicated to the creation of an interface standard to provide reliable mechanisms for copying and retrieving data from the myriad tape devices available. Since accessing tape is not a science that a majority of programmers understand, the number of development teams that actively create tape-aware storage applications is a relatively small group (compared to database developers, for example). These organizations and teams have spent many years - in most cases - getting to a point where the software that they create can operate properly with any new tape technology that a hardware vendor introduces.

However, looking back 18 or so years as the IT world expanded and data storage needs exploded (remember when a 310MB ESDI hard drive was considered large?), system administrators moved away from tape storage to large scale disk arrays because disk had become relatively inexpensive and disk access was (and "was" is the key word here) exponentially faster than tape access. Add to this the fact that disk manufacturers had very deep marketing pockets to sway an audience and you begin to see a recognizable reduction in the adoption of new tape technologies. Up until LTO-3, the fastest tape device was only topping out at speeds of 25MB/sec with most accessible devices hovering around 10MB/sec. Additionally, per-tape capacities of 25GB to 200GB were an issue in environments where data capacities were beginning to exceed 1TB. All of these issues were leading to tape losing its pertinence for backup and archival in the computing world.

To combat this perception, over the past few years a new effort has been promoted by the LTO Program team, LTO.org, which is designed to allow anyone to access tape as if it were a disk. This effort, named "Long Term File System" by it's originating design team at IBM and now known as "Linear Tape File System" or simply "LTFS" as other vendors have gotten involved, was created in an attempt to make tape pertinent once again. Unfortunately, the original design team didn't consider that the new performance and capacity provided by LTO-5 would already do this. Sustainable Read and Write performance of 140MB/sec and native uncompressed per-tape capacity of 1.5TB made the new LTO-5 mechanisms very pertinent when used with the right software. With a native, uncompressed per-tape capacity of 2.5TB and a rated read/write speed of 160MB/sec, LTO-6 was providing even more in the way of pertinence for modern, high- performance,

Contact BRU Sales

Email: brusales@tolisgroup.com
T: 480.505.0488
F: 480.505.0493
W: www.tolisgroup.com

Technical Support

W: http://support.tolisgroup.com
http://knowledgebase.tolisgroup.com

TOLIS Group™

high-capacity environments.

Because the LTFS specification promised a self-defining format with full cross-platform capability that worked “from the users’ desktop”, the marketing machines at a number of software and system integration houses went into high gear since they could now offer tape support for their platforms, even though they had no idea how to write to tape. In taking this leap, many quickly learned that even with LTFS in the mix, there was far more to putting data onto tape than simply copying it there.

While the enterprise environment was slow on the uptake of LTFS, mainly because of both their existing investment in software and staff workflow training, a relatively new kid on the data generation block was very quick to adopt it. The market space known as Media and Entertainment (M&E) had long depended on other technologies - video tape, CD/ DVD, and film - to store and archive their content. The movement of the industry to digital production operations - digital cameras, digital recording system, and digital editing, led them to search for other storage solutions.

Disk was the first consideration for many of these organizations because the non-computer users - camera operators, editors, producers, directors, and interns who normally only worked with specific applications on computer systems - could easily understand dragging files from one folder to another as a way of creating backup copies.

Unfortunately, the copying of data from one disk to another as a backup is sort of like “borrowing from Peter to pay Paul”. The number one reason for backing up data is due to the relatively high failure rate of disks. By copying data to another disk, the production teams now had to worry about two sets of data becoming lost because of disk failure. While tape was the obvious solution to anyone familiar with older, existing data retention practices, the difficulty in learning the very IT-centric workflow of most existing backup solutions was something for which the majority of M&E users had neither time nor training or hiring budget.

Once LTO.org recognized this need, they began a very active marketing campaign aimed at the M&E customers. Imagine a tape drive that could be accessed like a disk right from your desktop. Some recognized this as sounding too good to be true, but a number of industry managers thought that any technology that would allow them to better assuage their underwriters with regards to the new completely digital productions and keep the working teams happy because they wouldn’t need to learn any new technologies was worth digging into.

As more and more production teams were swept up in the high profile marketing campaign, a growing number of them began recognizing and complaining about the results, Problems such as difficulties in finding the right drivers for their systems, the fact that tapes they thought that they created yesterday were not accessible today, and that tapes they sent to other organizations couldn’t be read. Regardless of these headaches (and lost data), the LTO.org marketing machine continued its promotion efforts and the users continued trying to adopt LTFS.

Four of the main LTFS marketing points, as it were, are - the format was self defining, the format was cross platform for interchange compatibility, the format didn’t require specialized software, and the open source status meant no vendor lock in. Unfortunately, aside from being mostly self defining, the remaining points soon started to fall apart because of the non-curated open source and multi-platform nature of the project.

The first problem a user faces is to determine what drivers and other software components are needed to get LTFS functionality on your system. Depending on which vendor’s tape drive you are using, which version of your operating system you are using, and - in the case of Linux - which kernel and desktop manager release you are using, you have a multitude of options that you have to wade through. This defeats both the “no specialized software” being required (you have to find and install

the drivers and contingent component software) and the “no vendor lock in” claims (you can’t use IBM’s drivers with an HP drive, or Tandberg Data drivers with an IBM drive). Additionally, since LTFS is open source, no single party takes responsibility for assisting a user in getting things properly configured.

If you are lucky enough to properly configure your system, the second issue is that you then need to learn arcane command line syntax to create, mount, check, and eject LTFS tapes. It has also been discovered that an improper decision made during the formatting process could translate to lost data depending on how the tape was later mounted or ejected. This leads to more frustration since these command line operations are often even further outside of the user’s experience on Windows and OS X systems where things are very graphically-oriented. Many users don’t even know how to launch a terminal, let alone perform low-level commands. Again, since LTFS is open source, if problems arise, the only option for users is to turn to online forums and hope for assistance since no vendor takes responsibility if things start failing.

Since the LTFS public release in 2010, there are 21± different recognized LTFS implementations - from simple GUI’s that assist in formatting, mounting, and ejecting LTFS tapes, to dedicated appliances, to full asset management suites that create “wrappers” around the LTFS standard in an attempt to make up for many of the recognized weaknesses of the LTFS format. Features such as providing a checksum mechanism for the files copied to the LTFS volumes (LTFS does not provide an built in verification mechanism), splitting up large data sets to fit on single tapes (LTFS does not provide for data to span multiple tapes), and to provide offline cataloging functions (your system will only know about a tape that is mounted) are the most recognized limitations covered by these wrappers. The added side effect of using these wrappers is that you must use the wrappers to get the features that they add to LTFS. And, you have to pay for these wrappers - in some instances far more than normal backup applications for the same platform. And, you face even more vendor lock in.

As for cross platform compatibility, because of the open source nature and lack of a single controlling entity (SNIA is trying t step up to this requirement), many of the splintered variations result in incompatibility between versions and platforms. Why would you spend the time and effort getting your data segmented into tape-sized chunks, perform manual checksumming of those files and then not be certain if the data that you’ve copied to the tape can be read by anyone else?

Summarizing, the M&E industry is looking for an easy to use solution that is stable, reliable, and truly cross platform. Unfortunately, at this point in time, LTFS is proving to be anything but those things without a lot of help. There are a number of existing ISV backup solutions available and even the long standing “tar” formats from the Unix realm provide better cross platform compatibility than the existing 3rd generation of the LTFS specification.