

STORAGE SWITZERLAND

BREAKING THE TOP FOUR MYTHS OF TAPE vs. DISK BACKUP



George Crump, Senior Analyst

Over the past few years disk as a backup target has become a key enhancement to most backup infrastructures. Disk is believed to be faster, almost as cost effective and more resilient. In reality tape has its own unique value in each of these areas. When the fourth myth that tape must be treated separately is broken and tape is integrated tightly with disk, the combination resolves many of the backup storage challenges facing data centers today.

Myth 1: Tape is slower than disk

One of the most common assumptions is that backing up to disk is faster than backing up to tape. Reality is that when the raw speed of tape is compared with the raw speed of disk, tape is actually much faster. The speed difference actually becomes more significant when the additional housekeeping that most disk backup systems perform is factored into overall throughput. Most disk backup systems offer some data redundancy, but to keep capacity costs down, they also employ some sort of RAID protection strategy. While it is a more capacity efficient technology than mirroring, RAID suffers a noticeable

performance loss in write-heavy conditions. It should come as no great revelation that backup processes are extremely write-heavy.

To make disk more affordable to the backup process, most disk based backup systems leverage some form of deduplication to eliminate redundant data from backup storage. While deduplication has shown to provide as much as a 20:1 capacity efficiency gain, given the high ingest rates of data that are typical with backup jobs, deduplication can cause performance issues as it consumes processor cycles. This means that extra CPU horsepower must be invested in the disk backup device in order to maintain acceptable performance or the deduplication process must be done after the backup completes, something which requires additional disk capacity, increasing the price premium of disk backup.

In short, disk has to include a lot of complicated processes in order to make it feasible as a backup solution, such as RAID for data redundancy, error checking for data integrity and deduplication to narrow the price gap with tape. But these processes can severely eat into disk performance, making its speed disadvantage even worse.

Tape is relatively simple when it comes to writing data and in this case, simple means faster. As stated earlier, based on specifications, tape is faster per drive than disk and has less to do as it writes data so there is no degradation of that advantage. There is no RAID, or in most cases deduplication that takes place. Tape is already affordable, there is no need to add data protection or capacity optimization techniques that consume I/O performance. If redundancy is needed, an extra copy can be made with little concern over cost.

A more accurate description of disk backup's performance advantage is that it's more 'patient' than tape, which must be fed with data at a consistent rate in order to maintain its high performance. When the input data stream is inadequate, the tape drive must slow down, wait for data, and then spin back up. Disk does not have to go through this process. However, when disk is integrated with tape, a small and simple disk area, one not encumbered with sophisticated data protection or capacity optimization techniques, can still deliver the best of both worlds: cost effective and high performing backups.

Myth 2: Disk is almost as affordable as tape

Two factors have led disk to be the first stage in many data protection processes. First, the capacity per drive has continued to increase, bringing disk's cost per GB, now per TB, down significantly. Second, techniques discussed above, like compression and deduplication, have allowed even more data to be stored in the same physical capacity. This combination plus the "patience" factor described above has led to disk backup's emergence. The capacity reduction techniques and the increased density per drive have led some disk based backup vendors to claim cost parity with tape, or at least costs that are "close enough."

In most cases, these vendors are making a couple of convenient assumptions that may not apply to all data centers. First is the assumption that the data will indeed be compressible and redundant enough that a best-case deduplication ratio (~20:1) will be achieved. In reality, not all data can be compressed and/or deduplicated. There are several types of data where this is the case -- rich media files are a good example. Also applications with a high data turnover rate, document scanning systems for example, won't benefit much from deduplication.

A second major cost of disk backup systems is the cost of upgrades. When a disk backup system fills up, either retention times on-disk need to be decreased or more likely, an additional disk backup system needs to be purchased. Since most systems are stand alone units, their internal upgradability is limited, which means the cost of more capacity must include a whole new controller and power supply as well as disks. Even with scale-out storage systems, an additional node has to be purchased when more disk capacity is needed. While these systems more evenly spread out the capacity investment, they are not as price competitive as a tape capacity upgrade, which simply requires buying another tape cartridge.

LTO-5 tape media can deliver 1.5TB native and 3TB compressed capacity per cartridge for less than \$100. There is no amount of deduplication or compression that's going to match \$33 per TB any time soon. Disk of course, has its role and the affordability of the platform is important. An integration of the two would allow disk to be leveraged for its strengths but tape would allow disk capacities to remain small and help avoid very expensive capacity upgrades.

Myth 3: Tape isn't as Resilient

One of the appeals of disk backup systems is their perceived reliability. First, most disk backup systems use some form of RAID to protect from drive failure, and redundant power supplies and dual-ported connectivity are becoming increasingly common. However, the concern with disk is the amount of risk exposure they can cause should one of these components fail. For example, if the system experiences a drive failure, both backup and recovery performance plummets. If during the RAID rebuild a second drive fails, or a third under RAID 6, then 100% of the data is lost. While the chances of dual or triple drive failures may seem unlikely, the ramifications are so great that concern must be given. Also, as drive capacities increase, something that disk based backup systems are quick to adopt because of pressures to narrow the price gap with tape, the time it takes for the rebuild process to complete also increases. The longer the rebuild process, the greater chance for the unlikely to become likely.

While most tape systems have redundant power and connectivity, they do not typically have a RAID style of data protection. Redundancy is most often achieved by making a secondary copy of the tape after the backup process completes. While possibly more time consuming, it's a far more granular protection method; if one tape fails none of the other tapes are impacted. Data can still be read from the other tapes, with no performance impact either. Most importantly, if tape and disk are integrated, the disk system could create two identical tape copies simultaneously at very high speed, which resolves the extra time involved with tape duplication.

Myth 4: Tape and Disk must be separate

The introduction of disk based backup solutions has created yet another silo of storage to be managed in the environment. It was functionally simpler for suppliers to deliver a stand alone platform than it was to try to integrate it with multiple tape libraries. Some vendors did try to come out with integrated tape and disk solutions, but those required that the existing tape solution be replaced. Since the service time of the typical tape library is longer than that of a disk system, most data centers were not prepared to replace their tape library and purchase new disk backup hardware in a single transaction. The result was that most customers purchased stand alone disk backup systems. Those vendors tried to convince users that tape was "dead" because doing so meant they did not have to worry about integration. In reality most users struggled with how to get the two to work together.

This need for easy integration of the two storage types has finally been met by backup virtualization solutions like those available from [Tributary Systems](#). Backup virtualization abstracts the backup storage hardware from the backup software enabling the backup software to write and read data from a single virtual storage device. This allows disk and tape to work in tandem without having to constantly fine tune the environment.

As a result, the best attributes of each platform can be leveraged. In fact, even sub-categories can be leveraged. For example, a small but simple high speed disk cache can be used to store inbound backup data. Then, as time allows, it can be simultaneously directed to a deduplication-capable disk system and a tape library. The cache area can be used for high speed recovery of the most recent copies of data, the data deduplication system can be used for medium term recoveries of data, and the tape system can be used for long term retention of data. All of this can be managed across operating system platforms and backup application types, greatly simplifying the overall backup process.

Summary

Tape has strengths that are often overlooked because of concern over its shortcomings. Continued advancements in the technology plus the capabilities brought forth by integrating tape and disk with a backup virtualization solution lead to a fast, reliable and cost effective solution that all data centers should consider.

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://www.storage-switzerland.com>

Copyright © 2010 Storage Switzerland, Inc. - All rights reserved