

White Paper: Breaking the Top Four Myths of Tape vs. Disk Backup

Disk as a backup target has become a key enhancement in many data centers. Disk is believed to be faster than tape, almost as cost-effective as tape, and more resilient than tape. In reality, tape has its own unique value in each of these areas. And when the fourth myth of tape—that it must be treated separately—is debunked, and tape is integrated tightly with disk, the combination resolves many of the backup storage challenges facing data centers today. Tributary's Virtual Tape Library solution, Storage Director®, blends disk with tape, and adds sophisticated software to automate and optimize your IBM i backup. Storage Director works seamlessly with your existing backup software and procedures.



both reduces cost and improves performance, providing superior return on investment on both sides of the ROI equation. On the cost savings side, Storage Director reduces demands on the host system, space utilization, manpower charges and power consumption. On the benefit side, Storage Director provides much faster recall times, more reliable and realistic disaster recovery, a smoother migration path to new storage technologies, and unified data protection for heterogeneous servers, including Linux partitions.

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. **The reality is that when the raw speed of tape is compared with the raw speed of disk, tape is actually much faster.** When the additional housekeeping that most disk backup systems perform is factored into overall throughput, disk is even slower. This is because most disk backup systems offer some data redundancy but, to keep capacity costs down, also employ some sort of RAID protection strategy. While RAID 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. In addition, again 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 been 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, which requires additional disk capacity, which increases the price premium of disk backup.

In short, to be feasible as a backup solution, disk has to include a lot of complicated processes like 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 here 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. No RAID or, in most cases, deduplication takes place. Tape is already affordable, so 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. When the input data stream is inadequate, the tape drive must slow down, wait for data, and then spin back up, while disk does not have to go through this process. **However, when tape is integrated with a small and simple disk area—that is, disk not encumbered with sophisticated data protection or capacity optimization techniques—tape delivers the best of both worlds: cost-effective and high-performing backups.**

Myth 2: Disk Almost as Affordable as Tape

Several factors have led disk to be the first stage in many data-protection processes. The capacity per drive has continued to increase, bringing disk cost per GB down significantly, and techniques like compression and deduplication allow even more data to be stored in the same physical capacity. This combination plus the "patience" factor described above has led to the emergence of disk backup. These 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." But this ignores some significant cost factors.

For one, it ignores the fact that not all data is compressible and redundant enough to achieve the best-case deduplication ratio (~20:1). Among data that can't are rich media files or data used by applications with a high data-turnover rate, such as document scanning systems.

For another, it also ignores the major cost of upgrading disk backup systems. When a disk backup system fills up, the solution is to decrease the data retention times (not always feasible due to compliance and other requirements) or, more likely, to purchase an additional disk backup system. Since most systems are standalone units, there is a limit to internal upgrades, 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.

Tape cartridges are easy to add and deliver high capacity for backup.

For example, LTO-5 tape media can deliver 1.5TB native and 3TB compressed capacity per cartridge for less than \$100. No amount of deduplication or compression will match that \$33 per TB any time soon! Of course, disk has its role, and the affordability of the platform is important. Integrating the two platforms—disk and tape—leverages the strengths of each and helps avoid very expensive capacity upgrades.

Myth 3: Tape Isn't As Resilient as Disk

One of the appeals of disk backup systems is their perceived reliability. 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. The concern with disks, though, 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 percent 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.

Moreover, 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 more chances 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, this is a far more granular protection method. **If one tape fails, none of the other tapes are affected. 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 eliminates 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 data center environment. It was functionally simpler for disk backup vendors to deliver a standalone platform than 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 libraries and purchase new disk backup hardware in a single transaction. The result was that most customers purchased standalone disk backup systems. By convincing users that tape is "dead," vendors do not have to worry about integration. In reality, most users struggle with how to get disk and tape to work together.

This need for easy integration of the two storage types—disk and tape—has finally been met by backup virtualization. Also known as virtual tape library (or VTL), this approach allows disk and tape to work in tandem without having to constantly fine-tune the environment.

Data is backed up at high speed into the disk-based cache in the VTL system, from where it is automatically backed up to legacy or new storage devices, including disk and tape from IBM and other vendors.

This approach leverages the best attributes of each platform and optimizes the backup environment. For example, a small but simple high-speed disk cache can be used to store inbound backup data, thereby reducing the backup window and freeing the server for its production tasks. From the VTL disk-based cache, as time allows, data can be simultaneously directed to a deduplication-capable disk system and a tape library.

Backup data is stored on the media most appropriate for its recall needs. 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. 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.

Conclusion

Tape has strengths that are often overlooked because of concern over its shortcomings. Continued advancements in tape 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. ♦