

Does Tape Still Make Sense?

An investigation of tape's place in a modern backup system



May 2009

Does Tape Still Make Sense?

Contents

Abstract	3
Disk makes tape better	3
Throwing the baby out with the bathwater?.....	3
What do you lose when you lose tape?.....	3
Recovery After a Disaster: Restoring Remote Data	3
Economical Storage of Uncompressible Data	4
Elimination of Single Points of Failure (SPF)	5
Use of Legacy Backups	6
An Easy Compliance Solution.....	6
Familiar and Affordable Data Protection	6
Conclusion.....	7
About Spectra Logic	7

BlueScale, Spectra, and the Spectra Logic are registered trademarks of Spectra Logic Corporation. All rights reserved worldwide. All other trademarks and registered trademarks are the property of their respective owners. All library features and specifications listed in this white paper are subject to change at any time without notice. Copyright © 2009 by Spectra Logic Corporation. All rights reserved.

Abstract

While disk-only backup systems provide some advantages over tape, they also lose some advantages that are exclusive to tape. This whitepaper investigates the attributes of tape technology in a backup system and explores the ramifications of eliminating tape. By addressing the issue of what you lose when you remove tape this paper strives to answer the question of whether tape still has a place in a current data protection system.

Disk makes tape better

Disk eliminates many of tape's shortcomings in a data protection system. Adding disk uncouples the serial nature of tape from the backup process and can reduce the size of the tape infrastructure required. Disk may¹ enable faster backups and does enable faster restores. Disk is generally less prone to read/write errors than tape and it also enables deduplication. But does this mean that backup systems are better with *all* disk and *no* tape?

Throwing the baby out with the bathwater?

There certainly is no shortage of writing in the storage press about the advantages of disk in a data protection environment. The 'pain points' of traditional tape backup systems (mainly around tape handling, media failures and overall reliability) have been well covered. But in the interest of embracing new technology, it is easy to overlook some of the benefits that tape has historically provided. Eliminating tape from the backup system also eliminates some important functionality that disk alone cannot provide.

The following lists the attributes of tape that are not available through disk, in a data protection system:

- **Lower Cost** – tape and its automation have lower acquisition costs, use less energy, and store more data in less space through greater density
- **Scalability** –tape supports capacity expansion by the simple addition of more tape cartridges
- **Physical copies** – tape backup enables physical separation of data set copies for off-site storage
- **Portability** – tape allows fast and easy transportation of data set copies
- **Incumbency** – tape is the existing technology for backup in most environments

What do you lose when you lose tape?

Recovery after a Disaster: Restoring Remote Data

Disk-based disaster recovery (DR) solutions that use change-based replication and deduplication offer attractive advantages. They reduce the amount of data actually 'walking the wire' to a DR site, and the bandwidth (and time) required to complete the transfer. According to a survey by

¹ Historically, tape drives have not been the bottleneck in a typical backup system. But disk's ability to support multiple backup streams and widely varying data rates can improve overall backup speed.

Does Tape Still Make Sense?

Forrester, over half the respondents said bandwidth costs represent between 10 and 30 percent of the total replication or remote backup solution system costs.²

But what about the **recovery** part of the disaster recovery system? How do you get your data back to the primary site after a disaster? One answer is to have the applications in the primary data center failover to the DR site. This means buying servers, software and network infrastructure to support a second data center and implementing a failover process to bring these data online. This also means bandwidth, as the users of these applications will probably not be operating out of the DR site. This is a hardware-intensive method of restoring operations that is not feasible for many organizations.

Organizations that don't have this failover infrastructure must use the disk-based DR system itself to get the data back to their primary site. When faced with recovering anything more than a few files, these DR systems fall short. The 'skinny pipes' that were so appealing when designing the system (with 20-to-1 deduplication for example), are now inadequate to actually do a recovery of any magnitude, since they have to transfer the entire data set rather than just data changes. Typically, organizations don't have the bandwidth required to permit a timely transmission of the entire recovered data set.

Without tape, the only option to restore data quickly may be to physically move the arrays that hold the data in the DR site back to the primary data center location. In a worst case scenario, this could be your only copy of data, since the primary copy was lost in the DR event. Moving that disk array or DR appliance can be a significant (and risky) undertaking.

For the price of an autoloader, a small backup server and \$100 per TB (the cost of LTO-4 tape) you can have a copy of your data in the DR site that's a lot easier and safer to ship than a disk array. As one analyst put it, "Never underestimate the bandwidth of a Station Wagon full of tapes."³

Economical Storage of Uncompressible Data

Data reduction technologies, like deduplication and change-based replication, are critical to the economics of disk-based backup. Given the cost per TB of tape compared with the cost per TB of disk, (disk cost estimates range from about 3.5x to 23x more expensive⁴), disk backup would not be feasible if it had to actually record 100% of the data that ran through the system. For anything more than caching, data reduction is essential.

But some data doesn't deduplicate or compress well. Examples are: Images, CAD or graphics files, and files that were previously compressed or encrypted. Operational factors also affect data reduction. Storage Network Industry Association (SNIA) lists a dozen factors that impact

² Disaster recovery and data replication study conducted by Forrester Consulting commissioned by F5 Networks, January 2007

³ "The Advantages of Tape for Data Backup" Rick Cook, *TechTarget*, July 14, 2008

⁴ "The Cost of Tape vs. Disk Data Backup", Russ Fellows, *SearchDataBackup*, April 2008

"Disk and Tape Square Off Again", David Reine and Mike Kahn, The Clipper Group, Feb 2008

Does Tape Still Make Sense?

how well data deduplication works on a given data set – including how often the data changes, how long it's kept and how large the scope of deduplication is.⁵

In an environment with a large percentage of uncompressible data, these data sets can ruin the economics of a disk-heavy backup system. Tape's economy doesn't rely on data reduction technology and is unaffected by uncompressible data.

Elimination of Single Points of Failure (SPF)

When multiple copies of data are written to a disk-only backup system, including deduplication, the copies are *physically* located in the same array. The array or backup appliance *is* the single point of failure. If something happens to the controller, all these virtual copies are, at worst, lost altogether, and at best inaccessible until the system is repaired.

Primary disk arrays use many complex mechanisms to reduce the risk associated with this architecture. These include redundant components - disk drives, I/O interfaces, power supplies, batteries, fans and control processors – as well as sophisticated array management software to predict failures. These products are also typically protected by proactive maintenance and support contracts to further increase uptime.⁶

However, all these reliability measures are expensive and not usually provided for secondary storage devices used for disk backup. Disk backup reliability is limited typically to redundant fans and power supplies and email notification of system status, as well as dual-parity RAID 6. In addition, arrays used for long term storage of backups typically run regular integrity checks to identify and correct local errors on individual disk drives.⁷ However, when a system-wide failure does occur, restores are unavailable until the system is repaired. A system-wide failure can come from multiple failures of redundant components, any failure of single components (array controller, deduplication processors, I/O components, etc) or an external event that affects the entire system. The only rapid restore option is having another copy of the data available either on a completely separate disk array or on tape.

Tape's physical cartridge format means additional copies can be written to a separate physical medium, even if the copies are in the same library or on the same shelf. Because each tape is removed from the network and is not stored on disk, a system-wide failure won't affect all backup copies. Also, tape's not a virtual medium. On tape, each copy is complete and doesn't need to be 'reconstituted' or reassembled from other data. This protection against single points of failure can be had for essentially the price of the tape and a tape drive.

⁵ "Data Deduplication – Methods for Achieving Data Efficiency" SNIA Tutorial, Fall 2008

⁶ HP Storage Works XP12000 Disk Array Data Sheet

⁷ Data Domain DD600 Series Appliance Data Sheet

Does Tape Still Make Sense?

Use of Legacy Backups

No universal law of document retention has been mandated.⁸ However, many organizations operate in industries that have some regulations regarding information access and retention, and all responsible organizations keep a certain amount of historical data and data that's critical to the operation. Consequently, most organizations have a significant quantity of backup data stored on tape. Typically, organizations have also saved data on tapes from decommissioned servers and as golden copies. Switching to a disk-only backup system requires migrating these data off old tapes and onto a disk array.

For the cost of an autoloader or a tape drive, the requirement to migrate these backups to disk can be eliminated. These legacy backups can be left on tape and accessed as needed. And, for the cost of additional tapes, organizations can afford to continue keeping backup copies for long-term storage.

An Easy Compliance Solution

Regulatory compliance has forced companies to keep data sets for longer than their own internal requirements may dictate, to support possible future legal and e-discovery concerns. The Federal Rules for Civil Procedure (FRCP) now require that companies *"must not only produce electronic documents in the event of litigation, but produce them as part of the pretrial process. If documents can't be produced, a judge can slam a company with sizeable penalties"*.⁹ The order to produce these documents is called a discovery motion. What this means is companies now must decide which data **might** be needed in response to a future discovery motion.

One answer is to keep less data, reducing the amount that must be saved and searched for in case of a discovery motion. Most organizations decide instead to keep *more* data and keep it for longer periods because they don't know when they'll need to produce it. In a disk-only backup system the cost to store all this data can be prohibitive, even when using less expensive, secondary disk systems. The efforts to manually organize and deduplicate files—that is, not automatic deduplication, but eliminating duplicate copies of files—is also not feasible due to massive effort required to do so and its associated cost.

Tape's economy makes this an easy decision. The "keep it all" strategy for compliance can be implemented for the cost of tape cartridges.

Familiar and Affordable Data Protection

Tape and tape backup applications have been around for decades and they're in place at most organizations. Most IT personnel have handled tape in the context of managing data backup at some point in their careers. Further, for some IT organizations, switching to a disk-only backup system represents a major change. For many organizations, tape is the 'appropriate level of technology' for handling aspects of on-site and off-site data protection.

⁸ "Elements of a Good Document Retention Policy" *Lexis Nexis Whitepaper*, LexisNexis.com/discovery

⁹ FRCP Still Posing Challenges, *Byte and Switch*, Nov 29, 2007

Conclusion

Tape provides many essential operational benefits to a data protection system, perhaps some that have been taken for granted. Tape allows data that's been replicated to a remote site to be restored easily, without a complex and expensive infrastructure. It provides complete, additional, *physical* copies of data, inexpensively. These copies support regulatory compliance, data versioning and provide simple redundancy. Tape is also a superior solution when data sets don't deduplicate well or when you just need an affordable backup system that's well understood by your staff.

For most companies, eliminating tape does NOT make sense. And for these reasons, many actually see tape's role expanding in their data protection infrastructure.¹⁰

About Spectra Logic

Spectra designs and manufactures disk and tape products for data protection, backup and archive. Headquartered in Boulder, Colorado, the company has been in business for 30 years.

¹⁰ Mesabi Group published in Infostor, "Sense and Sensibility About Tape and Disk", December 2008

Does Tape Still Make Sense?



www.SpectraLogic.com

Spectra Logic Corporation
1700 N 55th Street
Boulder Colorado 80303 USA
800.833.1132
303.449.6400

Spectra Logic Europe Limited
Magdalen Centre
Robert Robinson Avenue
Oxford Science Park
Oxford UK OX4 4GA
+44 (0) 870.112.2150

For more information, please visit www.SpectraLogic.com