WHITE PAPER

Overcoming the Challenges of Tape with a Disk-Based Backup Solution

Abstract

All organizations use tape to back up data nightly. Tape is fairly inexpensive and low-tech, but managing and administering tape, backing up to tape and restoring files from it can be time consuming, unreliable and complex. Disk has always been an easier, more reliable alternative, but until recently its high acquisition cost has made it untouchable for many organizations. Fortunately, new disk and data reduction technologies have recently converged to make disk-based backup available at about the same price of tape backup systems. This white paper examines the pros and cons of both technologies and will present a cost analysis of ExaGrid's disk-based backup system with data de-duplication and compression vs. tape and standard SATA disk.

Tape Backup: Low Cost Yet Cumbersome and Unreliable

Traditional tape backup has been a low-cost method of protecting valuable data for years. However, tape can be extremely cumbersome and time consuming and can require a considerable amount of manual intervention to successfully perform regular backups. Tape backups must be monitored, equipment needs to be maintained and heads cleaned for backups to be run properly, and tapes must be loaded and changed, labeled correctly, and physically transported offsite for disaster recovery.

As an organization's data grows, so does the amount of time it takes to back the data up. In order to avoid impacting user productivity, most organizations choose to back up data at night so system performance is not affected. However, many organizations face long nightly backup windows, and in many cases, backups can't always be completed within the allotted timeframe, so backups are often aborted, leaving incomplete data that may be unusable in the event of an emergency. And even if a backup was performed successfully, the data isn't always there for a number of reasons. Tapes may be blank or become jammed in the tape reader, files are corrupt or missing, tapes may be damaged due to dirty heads, humidity or temperature, and tapes may have been mislabeled, lost or stolen. If a file needs to be restored and data is unavailable, the recoverable data may be days or weeks old.

Gartner Group estimates that approximately 15 percent of all backups fail, and the failure rates on restorations range from 10 to 50 percent. And the age of the data is critical. Gartner believes that data up to one year old has a 10 to 15 percent failure rate. The failure rate of data five or more years old is 40 to 45 percent.

The Business Impact of Using Tape

The impact of using tape goes beyond administrative and management of backups. It also affects users, customers and IT.

- *Users* spend countless hours each year creating data. When data is lost, the user loses productivity while waiting for the data to be restored. If the data needs to be recreated, users can miss critical deadlines and risk lost business.

- If *customers* are impacted by a failed backup, the business could lose not only its reputation, but it could lose revenue as well. If tapes are lost or stolen, then customers and business partners must be notified and the business could face fines or other actions.

- When *IT departments* spend a considerable amount of time dealing with tape issues, failed backups and lengthy restores, the impact hurts not only the business but greatly affects how IT resources are spent and the goodwill that the organization has created with its users and management.

Using Disk to Alleviate Tape Headaches

Fortunately, it's now possible to significantly streamline the backup process with disk. Disk provides numerous benefits over tape, so that backups are painless and successfully completed each night.

Tape	Disk
Labels fall off	No manual labels
Tapes are mislabeled	Correct labeling is automatic
Backups fail	Writing to disk is reliable
Restores fail	Retrieving data from disk is reliable
Slow backups/restores	Fast backups and restores, shorter backup windows
Lost tapes	Disk is in a fixed data center rack
Stolen tapes	Disk is in a fixed data center rack
Damaged tapes	Disk is in a data center with proper AC and humidity
Tape is inexpensive	SATA disk with compression and data de-duplication is inexpensive

The reason more organizations haven't turned to disk is simple: cost. Until now, tape has had a significant cost advantage over disk. Not only was disk expensive to acquire, but it was also costly because most organizations tend to keep weeks to months of backup history onsite. Without effective data de-duplication and compression technology, the amount of disk needed and the sheer cost of it put disk-based backup beyond the reach of most organizations.

New Data De-Duplication and Compression Technologies Put Disk-Based Backup Within Reach

The cost of disk-based backups has been driven significantly lower due to a combination of the following technologies:

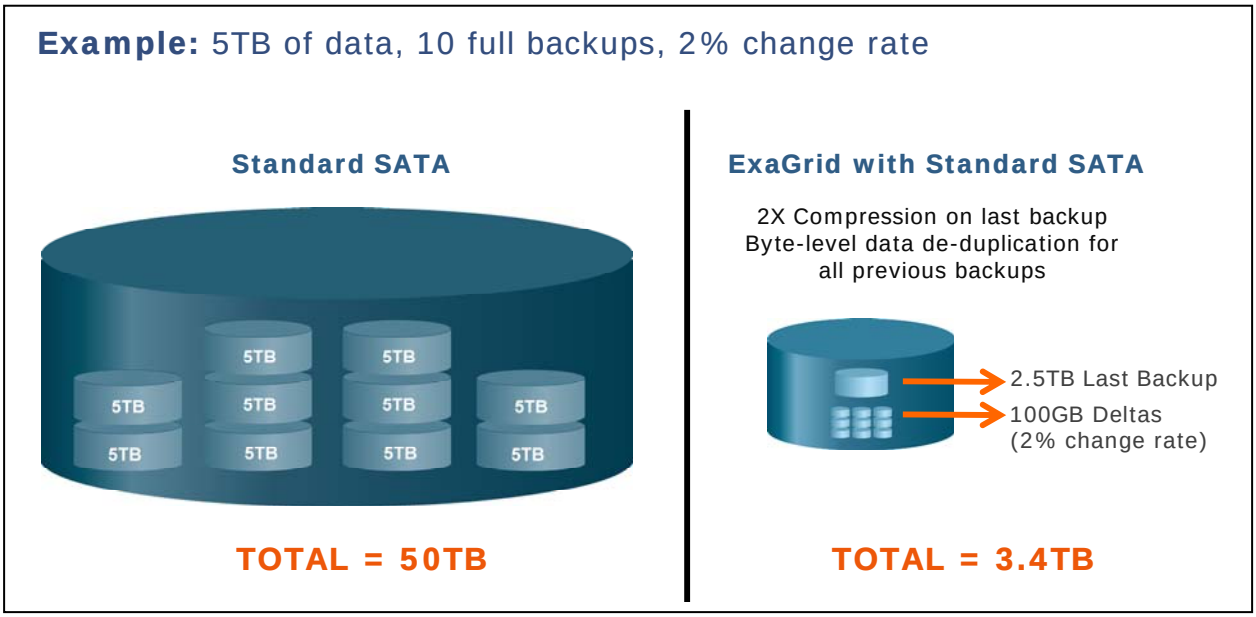
- ***SATA disk:*** highly reliable and available disk that is about 25 percent the cost of standard high speed storage.
- ***Data compression:*** a technology that compresses the data down to a smaller footprint.
- ***Data de-duplication:*** a technology that only stores only the bytes that change from backup to backup instead of storing full file copies. Data de-duplication can reduce the amount of disk space needed by a range of 10 to 50:1 or more.

ExaGrid's disk-based backup systems combine these three technologies to provide fast and reliable disk-based backups and restores at about the price of tape. Utilizing data de-duplication and compression backup data can be stored in anywhere between $1/20^{th}$ to $1/50^{th}$ the storage space, greatly reducing the disk needed and therefore greatly reducing the overall cost. Using ExaGrid systems, organizations can reduce or eliminate tape and shorten backup windows by as much as 30 to 80 percent with a small foot print of disk.

For example, if a 100MB file is stored, and then three letters were changed in the file, only a small fraction of the data actually changed. Data de-duplication technology only stores the

changes from backup to backup, significantly reducing the amount of data that needs to be stored instead of storing the full file copy.

If an organization typically keeps 10 full backups on-site or off-site, without compression and de-duplication, 50TB of disk would be required to backup 5TB of data, at a total cost of approximately $250,000 ($5,000 per TB, with controller and RAID 5).



Using ExaGrid to back up the same 50TB of data would be significantly less expensive. Using standard data compression for the most recent backup and data de-duplication for all previous backups, the most recent 5TB backup is compressed 2 to 1, to 2.5TB. Because only the bytes that have changed are stored for each previous backup, and only two percent of bytes typically change from backup to backup, therefore in this 5TB example only about 100GB of changes at the byte level change per week. The result is a 2.5TB compressed most recent backup plus 9 weeks of byte-level deltas as 100GB each or a total of 3.4TB (2.5TB + 9 x 100GB). What would normally take 50TB would take 3.4TB with compression and data de-duplication. A 5TB ExaGrid system would cost less than 25% the price of using disk alone.

Cost-Effectively Eliminate Tape

Organizations can also use ExaGrid to cost-effectively eliminate tape entirely. ExaGrid is extremely cost-effective as a second site because its data de-duplication technology only moves changes, so minimal WAN bandwidth is required. In a typical scenario where only two percent of the data has changed, ExaGrid delivers a 50 to 1 data efficiency over the WAN. Once the first backup is sent to the second site, only the bytes that change are sent. And the ExaGrid software is intelligent enough to merge those bytes that change into the second site backup so that both backup copies are entirely up to date with the most recent backup. The systems can also cross-protect, so two sites can act as disaster recovery sites for each other.

Tape vs. SATA vs. ExaGrid

Media	Tape	SATA Disk	ExaGrid with Compression & Data De-duplication
Includes	Tape Library Maintenance Onsite Tapes Offsite Tapes	94TB SATA disk RAID 5 Maintenance	ExaGrid Onsite ExaGrid Offsite Maintenance
3-Year Cost	**$35,000**	**$805,000**	**$42,300**

3-Year Cost of Tape:		Standard SATA Disk Cost:	
Tape Categories	**Tape Costs**	**Number of Copies**	**Disk Required**
New Tape Library	$12,000	2 weeks of nightly backups	2TB
Maintenance – 3 years	$6,000	10 weeks of Friday full backups	10TB
Onsite tapes	$2,000	Offsite monthly for 7 years	84TB
Offsite tapes	$6,000	15% RAID5 overhead	14TB
Pickup/storage – 3 years	$9,000	30% room for 1 year growth	30TB
		Total TB	**140TB**
		Cost of SATA Disk	$700K ($5,000 per TB)
		3 Year Maintenance	$105K
Total for Tape	**$35,000**	**Total for SATA Disk**	**$805K**

In the examples above, a two-site ExaGrid system capable of backing up 1TB of primary data, with ten weeks of retention onsite and 84 copies (seven years of monthly backups) of data off site would cost only $42,300 for ExaGrid, while a new tape system would cost $35,000, and SATA disk would cost $805,000.

A two-site ExaGrid system would completely eliminate tape, provide remarkably faster backups and restores, less IT intervention, management and maintenance, and provide better security because the ExaGrid system sits in a data center environment.

Conclusion

The combination of ExaGrid's high quality SATA drives, along with data compression and data de-duplication enables organizations to move from cumbersome, unreliable tape backup systems to next-generation disk-based backup. As older tape backup technologies need to be replaced, organizations should perform a cost/benefit analysis to evaluate which backup technology best meets their needs.

Intelligent Data Protection

ExaGrid's turnkey disk-based backup system combines high quality SATA drives with byte-level delta data de-duplication, delivering a disk-based solution that is more cost effective than standard SATA drives. ExaGrid's byte-level delta de-duplication technology stores only the changes from backup to backup instead of storing full file copies, reducing the amount of disk space needed by 10 to 50:1, or more, resulting in a solution that is 25 to 30% the cost of standard SATA drives.

ExaGrid is easy to install and use and works seamlessly with popular backup applications, so organizations can retain their investment in existing applications. ExaGrid can be used at a primary site while maintaining tape for offsite or can be deployed as a two site solution eliminate offsite tapes with a live data repository or for disaster recovery. When a second site is used, the cost savings are even greater because ExaGrid's byte-level delta de- technology moves only changes, requiring minimal WAN bandwidth.

For more information about ExaGrid, please visit our website or call us at 1-800-868-6985.

About ExaGrid

ExaGrid is the leader in cost-effective disk-based backup solutions. A scalable system that works with existing backup applications, ExaGrid is ideal for companies looking to quickly eliminate the hassles of tape backup while reducing their existing backup windows. ExaGrid's innovative approach minimizes the amount of data to be stored by providing standard data compression for the most recent backups along with byte-level data de-duplication technology for all previous backups. Customers can deploy ExaGrid at a primary site and at a second site to supplement or eliminate offsite tapes with a live data repository or for disaster recovery.

ExaGrid Systems, Inc | 2000 West Park Drive | Westborough, MA 01581 | 1-800-868-6985 | www.exagrid.com

EXAGRID SYSTEMS

Cost Effective Disk-based Backup with Data De-duplication

Compliments of: EXAGRID SYSTEMS

ESS ENTERPRISE Storage Solutions
3835R E. Thousand Oaks Blvd. #315
Westlake Village, CA 91362
Tel 877.230.2837 ~ Fax 805.435.2500 ~ www.ESS-Direct.com