EXAGRID® SYSTEMS

TOSS THE TAPE!

7 Key Steps for Overcoming the Limitations of Tape Backup.

INTRODUCTION

Although tape for backup has seemed fairly inexpensive and most organizations still use tape to backup data nightly, IT professionals know well that managing tape backup is hardly a straightforward job. With data growth putting increased pressure on tape—and projections of 60% annual increases in data through 2011— the processes of backing up to tape, restoring files from it, and managing and administering tape have become time-consuming, costly, unreliable and complex.

Magnetic tape was first used to record data and programs in 1951, and today, it's not a matter of *if* your tape backup is going to stop keeping up with the job, but *when*. In the same way that DVDs have largely replaced VHS tapes and disk has largely replaced tape as a working medium in most computing applications, it's inevitable that disk will replace tape for backup. Most IT organizations are no longer asking *if* they will replace tape with disk, but *when*?

Disk has always been an easier, more reliable alternative. However, until recently, its high acquisition cost has made it untouchable for many organizations at the capacity levels needed to backup and retain many months worth of data.   Now, as an increasing number of IT directors and managers are discovering, new disk and data deduplication technologies have converged to make disk-based backup available at about the same price of tape backup systems.

This document is a 7-step guide for making the move from tape to disk backup with deduplication. Within these steps, the Guide provides a concise summary of the downsides of tape backup, and then examines the superiority and economy of second-generation disk backup with data deduplication. It also presents key functional requirements for a disk backup system with deduplication as well as a framework for assessing the cost savings of a disk-based backup system with data deduplication vs. tape.

Business Impact of Failed Backups

The downsides of using tape can affect virtually every person who works within an organization or who does business with it. Users and customers are negatively impacted, not to mention the IT department.

- *Users* spend hours every day 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.

  > According to the National Computer Security Association, without adequate backup it takes 21 days and $19,000 to recreate 20MB of lost accounting data, and 42 days and $98,000 to recreate 20MB of lost engineering data.

- *Customers* are also impacted by a failed backup.. It can lead to a business not only losing its reputation, but losing 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 a major U.S. trucking company (and winner of the 2009 Info Security Products Guide Award) suffered a recent crash, they achieved significant savings with a quick recovery that resulted from   using disk with deduplication instead of tape. They avoided two days of system downtime and a loss of $200,000 in business and productivity that would have resulted if they had to restore their database from their previous tape backup system.

- *IT departments* spend a considerable amount of time dealing with tape issues, failed backups and lengthy restores. This detracts from IT being available to help business users with a wide array of other priorities, it impacts how IT resources are spent, and it dissipates the goodwill that the department has created with its users and management.

  > In one highly-publicized incident, a 2007 data backup failure by the Alaska Department of Revenue wiped out data on a $38 billion account and cost the state $200,000 to restore the data.
  >
  > For another organization, a leading provider of independent investment research with operations in 18 countries, their network administrator spent the majority of a work-day and work-week managing tape backups—rotating out up to 40 tapes each day and up to 90 tapes each weekend.

Step 1: Realistically assess the shortcomings of tape and potential consequences.

For years, traditional tape backup has been a low-cost method of protecting valuable data. 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 off-site for disaster recovery. In virtualized server environments, many backup applications do not support tape, and tape does not allow use of advanced options in backup software.

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 backup data at night so system and network performance are not affected. However, many organizations face long nightly backup windows, and in many cases, backups can't always be completed within the allotted timeframe. That means backups are often aborted, leaving incomplete data that may be unusable in the event of an emergency. And even if a backup is performed successfully, the data isn't always available 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
- 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.

Some Sobering Statistics

Data loss occurs much more frequently than most people may realize. According to many industry researchers, the odds are against you if you do not backup your data reliably every night.

15%	failure rate estimated for all backups[1]
10-50%	failure rate on restores. Data up to one year old has a 10 to 15% failure rate, and the failure rate of data five or more years old is 40 to 45%. [2]
$12 billion	Annual cost to U.S. businesses of data loss.
34%	of companies backup their data to tape but do **not** test their backups.[3]
77%	of those companies who do test their tape backups found back-up failures [4]
7 of 10	small firms that experience a major data loss go out of business within a year. [5]
30%	of IT professionals escalate backup problems or failures to a VP or C-level executive at least once every 3 months. Half of these IT professionals escalated problems as often as once/week or once/month. [6]
35%	of IT professionals reported that their backup times exceed the available window by times ranging from 1-2 hours up to more than 4 hours.[7]
30%	of IT professionals have to shut down backups without fully protecting their data at least once a month because the backup time is outside the backup window. 11% shut down backups at least once/week or several times/week.[8]

---

[1] Gartner
[2] Gartner
[3] Gartner & Storage Magazine
[4] Boston Computing Network, Data Loss Statistics
[5] DTI/PriceWaterhouse Coopers, 2004
[6,7,8] ExaGrid Systems 2009 survey of 3,000 IT professionals

Step 2: Determine your key drivers and align with budgetary priorities.

Projects to replace tape are invariably tied to a failure of existing systems to execute reliable backups, a need for IT productivity or cost savings, or a new business initiative. Below are six of the most common drivers for replacing tape with disk, and questions you should ask that will help align the right organizational and budget support for replacing tape.

1. Your backup times exceed the available backup window
- Have you done everything practical to bring backup times back in the window, such as reduce the amount of data, change rotations, and/or change configurations?
- What is your projected data growth over the next six to twelve months? If you are barely keeping up now, where will you be in 6 months time?
- Have you and your organization yet concluded that the shortcomings of tape will not be overcome by a new tape library?

2. Your tape library is dying or dead, or the IT maintenance costs have become excessive
- Have you and your organization yet concluded that the shortcomings of tape will not be overcome by a new tape library?

3. You have a project to improve disaster recovery, and need something better than tape for RPO and RTO
- Assuming you are not looking at live intraday DR, is your organization looking to have DR from nightly data, but faster recovery and more reliable than tape?

4. You currently write to straight disk, and the amount of disk is growing and getting expensive
- Are you willing to go to all disk—at least on-site—and eliminate tape on-site?

5. You have a VMware/virtualization backup project, and you want to go all disk and shut off tape
- Are you going to use your existing backup application VMware agents?
- Are you going to a separate VMware backup product such as Vizioncore?
- Are you going to eliminate tape but looking to avoid rip-and-replace of your backup application?

6. You are doing a data center consolidation project
- Are you being mandated to go with disk backup and replace tape as a part of this initiative?

Step 3: Make a side-by-side comparison of tape headaches and disk benefits .

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 data deduplication is inexpensive

One of the main reasons why more organizations haven't yet turned to disk is cost, or at least perceived cost. Until recent years, tape 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 on-site. Without effective data deduplication technology, the amount of disk needed and the sheer cost of it put disk-based backup beyond the reach of most organizations.

Other reasons why organizations have delayed replacing tape with disk include a lack of time to learn about the new technology, fear of trying something new, and concerns about eliminating tape for long-term retention. These concerns have been addressed at thousands of deployments at businesses of all sizes across virtually every industry sector.

Step 4: Understand how data deduplication and compression technologies are putting 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 deduplication:** a technology that stores only the bytes that change from backup to backup, instead of storing full copies. Data deduplication can reduce the amount of disk space needed by a range of 10 to 50:1 or more.

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 deduplication and compression, backup data can be stored in anywhere between 1/10th to 1/50th the storage space, greatly reducing the disk needed for backup and therefore cutting the overall cost. Using these systems, organizations can reduce or eliminate tape and shorten backup windows by as much as 30 to 80 percent with the small footprint of disk.

For example, suppose a 100MB file is stored, and then three letters are changed in the file. Data deduplication technology only stores the changes from backup to backup. So instead of storing the full file copy, it stores only the small fraction of data that actually changes, significantly reducing data volume.

Or consider the case of an organization that keeps 12 full backups on-site or off-site. Without data deduplication, 60TB of disk would be required to backup 5TB of data, at a total cost which would make disk backup impractical. Using disk backup with deduplication to back up the same 60TB of data is far less expensive. With standard data compression for the most recent backup, the most recent 5TB backup is compressed by a factor of 2:1, to 2.5TB. Then, data deduplication is used to reduce the amount data stored from all previous backups. Only the bytes that have changed are stored for each previous backup, and since just two percent of bytes typically change from backup to backup, the result is that only the 100GB that change at the byte-level per week (out of the original 5TB) need be stored. Taking it one step further, the most recent backup (compressed to 2.5TB) plus 11 weeks of byte-level deltas at 100GB gives you a total of 3.6TB (2.5TB + 11 x 100GB). Instead of taking 60TB, the use of compression and deduplication brings the total down to 3.6TB.

Example: 5TB of data, 12 full backups, 2% change rate
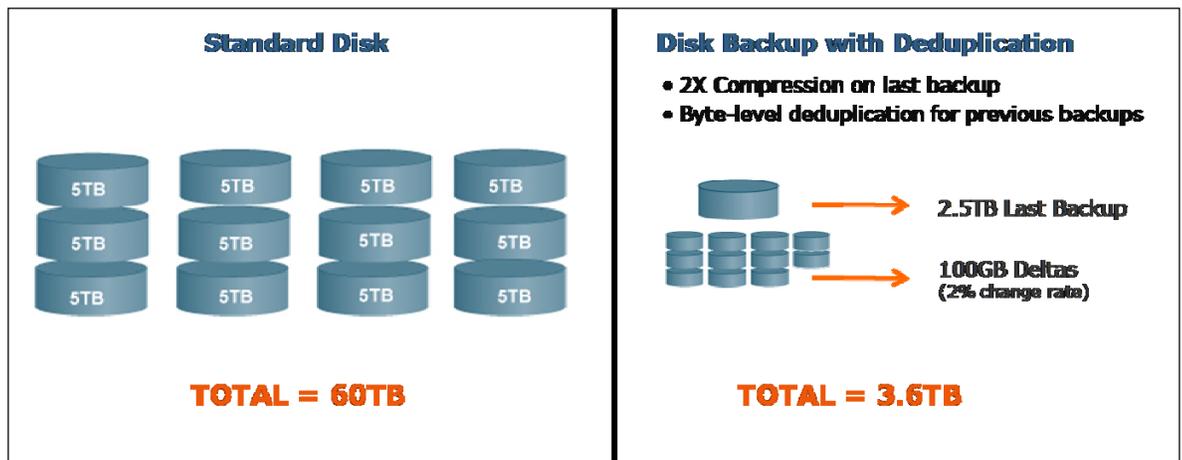
Step 5: Define functional requirements to meet business goals.

Once it has been determined that tape is going to be replaced by disk backup with data deduplication, the following essential capabilities should be considered:

1. Storage efficiency ranging from 10:1 to 50:1 via byte-level data deduplication
2. Quick installation with a turnkey plug-and-play appliance
3. No change to your existing backup application via NAS interface
4. Shortest backup window, using post-process deduplication
5. Fastest restore and tape copy by storing a full recent backup copy
6. No performance degradation as data grows via GRID scalability
7. No fork lift upgrades as data grows via GRID modularity
8. Fast disaster recovery via WAN-efficient replication and rapid restore
9. Deduplication and replication status via job-level reporting
10. Lowest price and IT operational costs through purpose-built architecture

Step 6: Eliminate tape entirely—and cost-effectively replace it with disk.

Organizations use disk backup with deduplication to cost-effectively eliminate tape on-site, and optionally to also eliminate tape entirely for both on-site backup and long-term off-site retention. Disk backup with deduplication is extremely economical for a second site because the data deduplication technology only moves changes, so minimal WAN bandwidth is required. In a typical scenario where only two percent of the data has changed, disk backup with deduplication delivers as much as a 50:1 data efficiency over the WAN. Once the first backup is sent to the remote site, only the bytes that change are sent. And the software on most systems is intelligent enough to merge those bytes that change into the off-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. A combined on-site and off-site system would completely eliminate tape, provide significantly faster backups and restores, and reduce IT intervention, management and maintenance. It also ensures superior security, since the system and its data sits in a data center environment.

Despite the clear operational advantages of shorter backup and recovery times, greater reliability, reduced risk of failure, increased data retention, and enhanced IT productivity, improving backup via disk-based backup with deduplication requires an investment whose payback can be difficult to quantify. Fortunately for IT managers, significant performance and productivity gains are achievable at a cost equivalent to that currently being spent, with paybacks reported in as fast as several days to 6 months.

Examining the budget for operating and maintaining tape libraries reveals an opportunity to reinvest for a better outcome in areas including:

- Cost of tape media and storage: Monthly costs for tapes and media replacement, tape storage, retrieval, transporting tape to a second location, tape administration and tape library maintenance.

- Labor savings: Productivity costs include monitoring during tape back-ups, maintaining equipment such as cleaning heads, loading and changing tapes, labeling and physically transporting to off-site for disaster recovery. This indirect savings would allow the time otherwise spent on backup to be utilized on other projects, and also reduce the amount of evening and weekend time required for IT backup management and troubleshooting.

- Hardware cost avoidance: Costs avoided or redirected to disk backup by eliminating the need for IT to buy additional tape hardware to complete backups within the available backup window. Also includes cost avoidance to perform tape backup in remote disaster recovery sites.

> A large manufacturer of recreational vehicles and manufactured housing reduced backup times by two-thirds, achieved their goal of a 24-hour recovery period from disaster (estimated previously at 3-5 days with tape). They also achieved the following savings:
>
> - Cost of tape media and storage: In addition to operator salaries, the company has been able to reduce tape costs by more than $5,000 per month and tape storage fees by $4,000 per month.
>
> - Labor savings: Reduced number of full-time data-center operators to perform and manage tape backups from seven to three and redeployed staff to other key initiatives because the backup process is now more efficient.
>
> - Hardware cost avoidance: Redirected $40,000 from new tape library purchase to investment in disk backup with deduplication.

Step 7: Entrust your backup to vendors and resellers with a track record of success.

Customers looking to replace tape with disk will typically turn to vendors or resellers with whom they already have a trusted relationship. But since many larger IT vendors do not offer a best-of-breed disk backup with deduplication solution, it's important not to exclude smaller vendors with superior technology who may have leapfrogged the established players. Also, make sure you can verify the track record of successful deployments for the specific disk backup product you are considering. While some tape-drive providers are introducing new disk-based solutions, they may not have a track record of successful deployments with the new technology, leaving your investment and data protection at risk. The combination of high quality SATA drives along with data deduplication technology in a plug-and-play appliance 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 look for customer success stories with other similar users, and also perform a cost/benefit analysis to evaluate which backup technology best meets their budget and functional needs.

About ExaGrid

ExaGrid offers the only disk-based backup appliance with data deduplication purpose-built for backup that leverages a unique architecture optimized for performance, scalability and price.

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 deduplication technology for all previous backups. Customers can deploy ExaGrid at a primary site and at a second site to supplement or eliminate off-site tapes with a live data repository or for disaster recovery. The combination of post-process deduplication, most recent backup cache, and GRID scalability enables IT departments to achieve the shortest backup window and the fastest, most reliable restores, tape copy, and disaster recovery without performance degradation or forklift upgrades as data grows.

With offices and distribution worldwide, ExaGrid has more than 1,600 systems installed at more than 460 customers, and more than 130 published customer success stories.

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

Cost Effective Disk-based Backup with Data Deduplication

http://www.exagrid.com

Phone: 1-800-868-6985