

 SearchStorage.com All-in-One Guide

Advanced Storage

Chapter 2: Backup/Data Protection

It takes more than blank tapes or cheap disks to handle enterprise data backups — it's all about keeping the business running while meeting the compliance standards that are appropriate for your industry. If you're already familiar with the basics of backup and data protection, this guide — which focuses on the advanced features and latest developments in tape, disk, and remote and strategic storage concepts — is a must read.



Advanced Storage Chapter 2: Backup/Data Protection

Table of Contents

[Overview](#)

[Tape backup: Best practices for long-term tape archives](#)

[Tape backup: Unreadable magnetic tapes — How to deal with tape errors](#)

[Tape backup: Integrate a virtual tape library with real tape](#)

[Tape backup: Commvault tapes stop writing before full](#)

[Tape backup Expert: Channel extending a tape environment](#)

[Tape backup Expert: Monitoring tape media duty cycle](#)

[Disk backup: Best practices — Optimizing your backups](#)

[Disk backup Expert: Hardware mirror migration](#)

[Disk backup Expert: Mirroring HP EVA4000 data](#)

[Disk backup: How to troubleshoot your D2D2T system](#)

[Disk backup Expert: Narrow down your replication or CDP options](#)

[Disk backup: Virtual tape evolves to survive](#)

[Disk backup: Case in point — Replacing tape backup with Avamar](#)

[Remote backup: Take full advantage of the remote replication process](#)

[Remote backup Expert: Oracle replication for failover](#)

[Remote backup Expert: Hot-hot replication with EMC's SRDF?](#)

[Remote backup: Remote replication gets out of the array](#)

[Business Continuity: Cost-effective legacy data protection](#)

[Business Continuity: Risk management — Know your storage risks](#)

[Business Continuity: Restoring data — Increase your efficiency](#)

[Business Continuity: Better DR and BC planning](#)

Overview

By Stephen J. Bigelow, Features Writer

Aug 1, 2006 | SearchStorage.com

It takes more than blank [tapes](#) or cheap [disks](#) to handle enterprise data backups — it's all about keeping the business running while meeting the compliance standards that are appropriate for your industry. Faster tape drives can speed a conventional backup, but disks are filling a wide range of important data backup roles; both locally and across a (wide area network) [WAN](#). Deciding on the best backup strategy is also more complicated since an organization must maintain security and ensure proper adherence to government regulations. If you're already familiar with the basics of [backup and data protection](#), this guide — which focuses on the advanced features and latest developments in tape, disk, and remote and strategic storage concepts — is a must read.

Tape backup

On the surface, it's easy to select tapes — just buy cartridges that fit the tape drives in your enterprise. But today, the choice is more complicated than that. Tapes are expensive. High-end LTO-3 tapes can cost over US\$60 for a 400 GB-to-800 GB cartridge, while a Quantum digital linear tape (DLT) 0.8-to-1.6 terabyte (TB) tape can start at US\$120 each. Limited availability might also impact backup plans. For example, back [in late 2003, Imation Corp. became the sole distributor for VXA and Mammoth tapes](#) from Exabyte Corp. A limited manufacturing and distribution base can restrict stock and inflate tape costs — two crucial considerations when selecting a tape technology.

Although [tape backups](#) are slower than disk backups, tape drives are getting faster. For example, today's SuperDLT drives can support 60 megabytes per second (MBps), uncompressed, while late-model linear tape open (LTO) drives can handle 80 MBps, uncompressed. [Sun Microsystems Inc. introduced the T10000 tape drive](#) in late 2005, touting a data throughput of 120 MBps with uncompressed capacities up to 500 GB on a single cartridge. The biggest difficulty with “fast” tape drives is that host backup servers generally cannot accommodate those data rates, often resulting in wasted efficiency — fast tape requires careful design of the communication path between the drive and backup server, and between the server and storage.

Today, groups within an organization are sharing fewer tape libraries, so there's a push to pay only for the library capacity being used — yet maintain management control over the tapes in that portion of the library. Consequently, tape libraries are incorporating partitioning and chargeback features, as well as greater scalability for consolidation. The Scalar i500 from Advanced Digital Information Corp. (ADIC) provides partitioning and chargeback, and scales from one to 18 LTO drives, and 36 to 404 tape slots in a single frame. Partitioning can also be found in libraries like IBM's TS3310, the TLS series from Qualstar Corp. and the CSM200 from Sony Electronics Inc. It's interesting to note that library features like multiple tape media support, high availability and large numbers of tape drives are not as popular as once thought.

Backup software is moving beyond the role of scheduling and reporting. Users are turning to backup software to reduce the sheer data backup size. Compression had been used to fit more data onto a given tape, but

data deduplication, also called intelligent compression or commonality factoring, is starting to appear. Deduplication works by saving a single iteration of a file or block — providing only pointers to duplicated data. That is, instead of saving 10 copies of a 10 MB sales presentation, only one copy is actually saved to tape.

Disk backup

Low cost and relatively high performance have made hard disks the preferred target for many data backup tasks. Regardless of the actual storage platform, there are several important trends worth considering. First, data deduplication is appearing on data backup platforms like [virtual tape](#) libraries (VTL) and [content addressed storage](#) (CAS) archives. Deduplication reduces the number of disks needed for storage or fits more data into available space — lowering the disk investment.

As the amount of data relegated to secondary disk expands, the retention period for secondary disk is also increasing. Data backups might typically reside on a secondary disk platform for several weeks or perhaps a month before being offloaded to tape for long-term off-site storage. This is changing as larger arrays of inexpensive disks enter service. CAS systems are already managing long-term data retention on disk, and VTL storage arrays might soon hold a year's worth of data.

Users must also consider the effects of power, cooling and reliability in large data backup disk installations. Arrays with hundreds of disks can consume thousands of watts of power which is difficult to cool properly, and cumulative disk vibration can cause premature disk failures — especially among SATA drives. Array manufacturers like Copan Systems are developing MAID systems where 80% of the disks are idle. The idle disks are powered on and tested periodically. Data is migrated between disks to ensure that all disks are used for the same time — reducing power and improving mean time between failures ([MTBF](#)).

Remote backup

The ongoing challenge with remote data backups is the cost of bandwidth. A company must budget for connectivity that supports an appropriate backup volume within an acceptable backup window. Too much bandwidth wastes money; too little bandwidth wastes time. Deduplication and selective backups reduce data and lower bandwidth needs.

The choice between synchronous and asynchronous replication can have a significant impact on data backups. Synchronous replication offers the lowest recovery point objective (RPO) and recovery time objective (RTO), but the latency of long geographic distances can render this impractical. Asynchronous replication is a little easier, can work across longer distances and is tolerant of WAN outages. But asynchronous RPOs can range into hours because remote writes can lag significantly behind local writes.

WAN reliability is another consideration that is often overlooked. A failed WAN link can disable the data backup process, potentially leaving critical data at risk. Organizations should investigate an alternative that can protect data during a WAN interruption. For example, users may implement a backup to local disk and then pass the disk backup to an off-site VTL or other disk system. If the WAN fails, there is already a local backup, and the remote backup can be retried or completed once the WAN is available.

Other backup concepts

Traditionally, data backups were implemented to suit the individual needs of the organization, ensuring that important data could be recovered in an emergency. Mirroring, replication, snapshots and [continuous data protection](#) (CDP) technologies are still commonly employed for exactly that purpose. Many disk storage platforms routinely include applications to support these features. For example, the Clariion CX3 Model 80 from EMC Corp. includes SnapView software for local replication and MirrorView software for remote replication.

Today's data backups are increasingly influenced by [compliance and corporate governance](#) concerns that require data to be integral, accessible and retainable for a prescribed length of time. Backup administrators must understand what data should be backed up; how the data should be backed up and protected; and how the data is accessed in the face of legal discovery or disaster. Backup planning for compliance should involve business units across the enterprise — not just IT. CAS systems are often employed to meet compliance obligations since CAS platforms offer data deduplication, security and data management/search tools that are suited to long-term data retention and retrieval.

Security is also gaining importance, and backup administrators must protect sensitive or personal information against loss. Backup software like Symantec Corp.'s NetBackup offers encryption as an option, allowing tape or disk data to be encrypted during data backup, or decrypted for recovery. Still, there is some debate about just "where" encryption should take place. Software encryption is effective, but it reduces performance and locks an organization into the backup software product. By comparison, encryption can also be performed at the tape drive itself (like Sun's T10000), through a dedicated appliance, such as the CryptoStor family from NeoScale Systems Inc., or the DataFort family from Decru Inc. In the disk-to-disk (D2D) storage realm, VTL products are embracing encryption, and FalconStor Software Inc., includes a Secure Tape Transport Service module with its VTLs.

Tape backup: Best practices for long-term tape archives

Greg Schulz

August 22, 2006

What you will learn from this tip: Best practices for magnetic [tape](#) for data retention and how to store tapes for long duration.

The advantages of using tape to store long-term archival and compliance data include low cost per gigabyte and relative ease of transportability, assuming the tapes do not get lost in transit. Several options exist to secure stored data, including drive or [tape library](#) level encryption.

For the near term, tape is still being used in some environments to support [backup](#), archiving and retention of data for compliance. Some environments have shifted from tape to disk while others have embraced a mixed tape and disk approach to support long-term data retention. The decision to use tape instead of other media, including optical, disk or remote network-based services comes down to cost, preferences and service requirements. When looking at the cost of tape vs. other media, make an apples-to-apples comparison by considering the cost of handling of the media, applicable hardware and software costs and ongoing maintenance. With high energy prices, also include power and cooling costs into your analysis. Requirements include what service-level objectives you need to support for data availability and survivability.

Periodically audit your backups and archives by randomly selecting a tape and verifying that you can read the tape on a tape drive in a different location than where it was written. Also, you should verify that you can restore selected data to an alternative location. This audit confirms that what you think has been backed up is in fact on the tape. If your data is important enough to backup, then it should be important enough to make multiple copies that can be stored in different locations either online, near line or offline.

Protect your tapes and the data they store by implementing appropriate levels of physical and logical security. Store your tapes in a comfortable place, free of dust and other containments. As an added safeguard, store your tapes away from any sources of magnetic interference including monitors, motors and microwave ovens. Also, keep your tape drive heads clean, as recommended by the drive manufacturer — using the recommended cleaning tools.

Manufactures of tape drives and media including HP, IBM, Imation, Quantum, Sony and Sun/STK among others have specific recommendations for the care and handling of media for you to review. In general, if you are not comfortable with the habitat in which the tapes are stored, chances are the tapes are not comfortable either. If you need to move tapes from a climate controlled environment, allow time for the tape to acclimate and adjust to the new surroundings before using the tape.

Determine what the lifecycle of the media that you are using is and at what point in time you will retire a cartridge or migrate to a different storage medium. Unless you plan on being in the media conversion business or maintaining an active tape technology museum, avoid islands of technology to preserve backwards compatibility, and instead look into migrating data to alternate mediums for preservation where practical.

Once you have migrated or retired old tape media, make sure to dispose of it properly (by yourself or with a qualified disposal firm). Simply throwing your old tapes in the dumpster is no longer an option.

Tape remains an option for various environments depending upon your preferences, budget and requirements. Look to the future as to what technology you will eventually replace tape with. Also, consider when and how you will go about migrating data from one medium to another. In the meantime, take care of your tapes so that your data will be safe and secure when you need to access it.

Tape backup: Unreadable magnetic tapes – How to deal with tape errors

Greg Schulz

June 27, 2006

What you will learn from this tip: Learn what you can do when you are faced with errors when trying to read your magnetic [tapes](#).

There can be many reasons why a magnetic tape cannot be read; some of the issues can be temporary and others are more permanent. Some tape errors can be attributed to operational or pilot error or mishandling, while other errors can be the result of damaged media or tape drives.

Below are some reasons you may encounter tape errors:

- Heat, smoke, fire and water damage
- Damaged or broken tapes and cartridges
- Accidental deletion or overwritten data and reformatted tapes
- Tape drive and other mechanical wear and tear
- Dirty tape drives
- Lack of proper tape storage and handling
- Exceeding manufacturer suggested lifetime of the media

When you encounter a tape media error, verify that the tape media is in fact bad, and that something else is not preventing you from reading your tapes.

The following questions can help you to verify and diagnose the problem:

- What errors are being logged by your tape drive when you try to read the tape?
- Does the error occur if you try reading the tape on a different tape drive?
- Are other tapes readable on the same tape drive using the same utilities?
- Is there any obvious physical or visible damage to the tape media or cartridge?
- Are you able to read the tape from a different server?
- Is the tape [encrypted](#) or [compressed](#), thus preventing the media from being read?
- Has the tape media been properly acclimated if recently moved or transported?
- Is there visible physical damage to the tape drive device?

A first step should be to check with your technology provider, or the vendor who sold you (and provides the servicing for) your tape drive. Some vendors have Web sites that list frequently asked questions and other troubleshooting tips that may help your diagnose your tape errors.

There are several firms providing services for recovering data from damaged or corrupted tapes. While this list isn't complete, a few companies that provide these services include: Imation, Quantum, Exabyte, Sun/STK, HP and IBM, among others.

In 2003, for example, after the tragic challenger space shuttle disaster, technicians at [Imation](#) along with NASA and other experts, were able to recover critical data from the severely damaged flight data recorder tapes of the stricken space shuttle to support the accident investigation.

There are specialty firms that deal with recovery of media from disk and/or tape as well as optical media. A Google search of the phrase "[tape media recovery](#)" will yield a list of different service providers that may meet your specific needs. Inquire if the vendor provides physical as well as logical (data) recovery of damaged media or if an external third party is involved in the process. Look for a recovery service provider that has experience and credibility in recovering damaged media matching your needs and requirements. For example, be familiar with which tape formats, type of tape drives, [operating system](#) and applications or backup utilities and data formats are they familiar with. Since critical and sensitive private data will reside on your tape media, look into and ask a service provider about their security and privacy policies — including who will have access to the media and any recovered data.

To help minimize the chance of damaging tapes, exercise caution and follow manufacturer suggested best practices for tape handling and media storage. In addition, periodically and randomly audit and test tapes along with your data protection and archiving to tape media processes. As part of an audit, verify that the correct data is stored on the media and that it can be restored. Another safeguard is to test the restoration using a different tape drive to an alternate location to verify that the tape is good. Also, test and verify that your backup or archiving application parameter settings and configurations are correct as you intended. For critical data, make an alternate copy of the data on the same or different type of media and store in a climate controlled environment to minimize lost data.

Tape backup: Integrate a virtual tape library with real tape

Rick Cook

April 24, 2006

What you will learn from this tip: Storage expert Rick Cook explains how hanging a tape drive or library off of a VTL and pruning your data are two ways to integrate your VTL with tape.

Sooner or later you're almost certainly going to need [tapes](#). For archival storage, tape's combination of capacity, cost and storage life is hard to beat. That doesn't mean, however, that you're going to be backing up directly to tape. Increasingly, backups are going to disks, either in a disk array in a D2D configuration or to an array organized as a virtual [tape library](#) (VTL).

So, how do you integrate a VTL with tape? The answer is, "it depends." First, it depends on the architecture of your system. Beyond that, it depends on the specific VTL and backup software you are using. Another factor is how much data pruning you do between backup and archival storage — and how sophisticated it is. In general, the considerations involved in integrating a VTL with tape are conceptual and architectural rather than technical. VTL interfaces are highly standardized and are designed to appear as tape libraries or as disk arrays when exporting data.

The good news in all of this is that while you may still need tape, you don't need nearly as much of it. Not only is the volume of data written to tape much smaller with a VTL in the loop, but the architecture is much less complex (and expensive) and the time requirements are much more flexible. When writing from a VTL to archival tape, you are writing from one data source (the VTL) instead of multiple servers and with an extremely relaxed time frame. This means you also avoid the problems associated with multiplexing data streams from multiple servers into a single tape library.

One popular way to connect a VTL and tape is to hang the tape drive or library off of the VTL as [direct-attached storage](#) (DAS). This is cheap, simple and minimizes the load on the network. However, like all DAS, it is inflexible. If you think you may have to support additional VTLs you may want to use an alternate method, such as communicating with the tape unit over your SAN.

Data pruning is a consideration because you seldom need to archive everything you back up. Backups are designed to let you restore your files or system to a known state, almost always less than three months old and usually 30 days old or less. Archival data is data you need to store for the long term — usually for years at a time and sometimes permanently. Typically, archival data will amount to much less than half of the data you back up.

So, how do you prune the data going from VTL to archival tape? The easiest way is simply not to save certain file extensions to tape. Unfortunately, this is seldom satisfactory, especially in the world of [SOX](#) and [HIPAA](#) requirements. That means you're going to need some 'storage smarts' in order to enforce business rules on what goes to tape, if you want to prune effectively. Some VTL products, such as those from

FalconStor Software and Adaptec Inc., can handle policy-based replication of selected data to tape. Other approaches include having a separate server to handle your archival tape library or using the features of your backup software to monitor and prune the data feeding to tape.

Finally, keep in mind that although VTLs have been around for a while, it is still not a highly standardized technology. Consider VTL integration with tape carefully when choosing a VTL product and understand that not every vendor will offer all the features you want. Sometimes you can't get everything you want in one package.

Tape backup: Commvault tapes stop writing before full

Pierre Dorion

April 1, 2006

Why do our CommVault tapes stop writing before they are full? We get a message that there is no media. When I check the tapes, there is plenty of valid data space and the tapes are not marked full.

There are a few situations where this condition can be encountered. However, without more details about the actual error and hardware configuration, it is difficult to pinpoint exactly the source of the problem. Below are a few suggestions:

Tape errors: Sometime the tape drive can encounter a media error that is interpreted by the device as an EOT (end of tape) signal. This leads the device to mark the tape as full and stop writing. This error condition does not typically occur with all media in a library making it easy to diagnose but hard to predict. If this is your case, this could be the starting point for troubleshooting.

Configuration:

- Check the "Appendable" and "Use Appendable Media for (n) Days option" from the Media tab of the library properties; the number of days might be set lower than it takes you to fill a tape.
- A job option which has the Start New Media option enabled, will not use appendable media.
- Synthetic full backups will not use appendable media.

Tape backup Expert: Channel extending a tape environment

Christopher Poelker

July 2, 2005

We are looking to consolidate Mainframe CPUs from site A into site B. We are also looking to leave the existing tape environment (VTS/ATL, STK silos, floor drives) at site A and channel extend them from site B (approx. 250 miles). This would be for full time 24x7 tape operation, not just peer-to-peer replication or for offsite backups.

Do you have any recommendations, dos and don'ts, pitfalls, risks? Is there anyway to obtain information on others' experiences with channel extension used in this manner?

There are a number of companies that provide channel extension hardware and appliances. I will assume from your question that the current tape subsystems are [SAN](#) fabric attached, and you are currently using [FICON](#) connections from the mainframes to a switched fabric.

Please be advised that moving tape backup streams over extended channels would require substantial bandwidth between sites A and B. I'm not sure if you currently have dedicated [dark fiber](#) between sites, but that may be the minimum that would be required for multiple concurrent backup streams. 250 miles is also very far for extended channels. 100KM to 200KM is usually the top-end distance for either Fibre Channel ([FC](#)) or FICON extension before latency or "droop" occurs (ESCON extension is MUCH less tolerant to distance).

200KM is about 124 miles as the crow flies, so even using dark fiber connections and great extension gear will only get you halfway there for channel extension. Using an [IP](#) bridge to your fabric with a leased IP connection would be a better solution for that type of distance. Buffering, packet shaping, compression and spoofing of the connection can reduce a lot of the latency issues. Common Mainframe PPRC, XRC and GDPS configurations can utilize connections such as this. An even better approach may be to use [SONET](#) as the connection medium, since some of the FICON to SONET vendors support high-bandwidth trunking, and distances up to 40,000KM.

Your bandwidth requirements will depend on how much data you need to move across the links. I use a simple formula of 10 Mbit of IP bandwidth required for every 1 megabyte (MB) of data per second transmitted. As for do's, don'ts, pitfalls, and risks, there are too many to discuss using this as a venue. You should sit down with your storage, tape and switch vendors and have an in-depth planning session to go over all the details.

The tape vendors you mentioned in your question have a lot of experience in this area, and may be a very useful resource for you, as would your FICON switch vendor. You did not mention the storage vendor you are using, but I would also include them on the planning session.

Tape backup Expert: Monitoring tape media duty cycle

Ashley D'Costa

December 5, 2005

Recently, we have gone to do a couple of restores and found that the tapes (SDLT 1) were bad. I have been trying to find information on tape life. I know that there are many variables, such as how well do the tapes stream, etc. I found one article saying that the tapes should be good for a 1,000,000 passes of the head. How you would calculate that? The only information I have is that I know we filled the library about two years ago. We have replaced some tapes that were frozen and others that seemed to always get errors, but we have a number of tapes that have been in since we received the ESL9595 library. I ran a report that shows me the number of mounts. How do I know when to remove tapes that may be getting unreliable?

This is a tricky question to answer. I've always found it hard to get a tape media manufacturer to specify realistically how reliable their media is in terms other than laboratory results (thus, the 1,000,000 passes number). It's the safest way for them to express the reliability of their product, although very difficult to get statistics on and, therefore, impractical to monitor. I will try to address the question by discussing how the issue is currently being handling in most IT shops and how to avoid errors in the first place.

Tape media duty cycle

With tape media, you typically don't know that a tape is bad until it fails — at which point it's too late. As a result, the most typical way that this issue is addressed is to pull tape media out of circulation after a certain amount of time and assume it is unreliable even if it has not had any errors. This would be considered its useful duty cycle. I've found that linear tape technology ([DLT](#)-based/[LTO](#)-based) that is used daily typically gets pulled out of circulation after about a year. Helical scan technology is more prone to wear, and therefore, is pulled from circulation sooner — generally after about three to six months (with the most extreme cases after only 10-12 passes).

Most [backup](#)/recovery products have the ability to enforce a tape duty cycle by attributing an expiry period to the tape media (this is different from the expiry period of the data on the tape that defines the data's retention). Because you state you have "frozen" tapes, I'm assuming you have Veritas NetBackup as your backup/recovery product since "frozen" is a NetBackup term. I know NetBackup does have an expiry date you can assign to your tapes, after which the tape becomes read-only until all data on it expires and then the tape is no longer used.

Archival life

A tape's reliability is also dictated by how long it has been sitting around — its archival or shelf life. This doesn't come up as often with respect to tape reliability, mainly because of the long shelf life that most modern tapes can have. For example, depending on temperature and humidity, SDLT 1 has an advertised archival life of 15-30 years. Of course, this assumes no tape handling at all.

Tape handling

The best way to prolong your media's duty cycle is avoid having errors altogether. Errors are most times the result of damage to the physical medium within the tape cartridge rather than a result of a defect to the medium. This damage to the medium occurs most commonly because of incorrect tape handling procedures.

Tape media is typically handled because of offsite vaulting requirements. As a result, most tapes will be handled at least several times through the course of their duty cycle.

Modern tape technology (e.g., LTO-based, Magstar-based, DLT-based media) pack a significant amount of tracks on the medium — far more than in the past. The margin for error is orders of magnitude smaller than 10 years ago. In the past, there were far fewer and much larger tracks. The margin for error was significantly larger, and therefore, tapes in the past could be handled more robustly.

Due to this legacy thinking, there is still a tendency to man-handle modern tape media the way it was handled in the past. For example, a common problem today is edge damage. If a tape is dropped, it is highly likely that the edges of the media could get crimped. Since, in most linear tape-based products, this is where the servo track information is kept (a track that allows the tape drive head to stay aligned with the tape) it is possible that media errors could result because the head can no longer “stay on track” so to speak.

If you have a high rate of tape errors, I recommend to reviewing your current tape handling procedures and then reviewing the best practices for handling the media from the tape manufacturer's Web site. DLT and SDLT media are manufactured by Quantum. Click on the following link for Quantum's very informative [Care and Handling Guide](#) (pdf).

Disk backup: Best practices — Optimizing your backups

Pierre Dorion

August 15, 2006

What you will learn from this tip: Most organizations back up data one way or another nowadays — but just how many do it well? This tip will show you some of the areas of backup you should consider optimizing.

There is definitely a line to be drawn between data backups and good data backups. If your organization does not yet back up data and you just stumbled on to this storage site by mistake, the moral of the story is simple: Start backing up now! For the rest of us, there are a number of items that are often overlooked when performing data backups. Many administrators focus on backup success or failure and performance but often forget about the most important aspect of backups: restorability.

Optimizing backups is not always about storage device performance and tweaking software configuration settings. Here are some other aspects to consider:

- **Know what you are backing up:** Are you backing up everything you should be? Do you back up more than you need? There might be critical workstations or PCs that are missed daily while users assume they are covered. Conversely, you might be consistently backing up static or redundant data (i.e., some OS files, archived database tables or exports). A series of discussions with the business side may provide some valuable insight.
- **Disk backup:** D2D backups or data replication are probably some of the best ways to optimize backups (and restores) from a performance perspective. Unfortunately, bandwidth for remote backups and long-term storage capacity costs remain major hurdles for most organizations. This capability should be deployed based on data and application criticality.
- **Prioritization:** Data must be categorized based on application restore priority. The application criticality dictates the RTO and drives the priority ranking. This same RTO and restore priority should be used when choosing the backup methodology (i.e., D2D backups or [replication](#)). This warrants further discussion with the business.
- **Full and incremental backups:** Many backup products are now “incremental always” capable. Some of these products also offer the ability to create “synthetic full backups” by concatenating incremental backups into a current state image. This feature can dramatically reduce the amount of storage as well as time spent backing up unchanged files.
- **Number of tape drives:** When backing up to tape, the number of tape drives should be equivalent to the number of desired (or required) concurrent data streams. However, network bandwidth must also be considered when adding tape devices.
- **Network and system performance:** Ensure that the network bandwidth and disk subsystem will be capable of handling an unusually large amount of data, such as simultaneous full restore streams. While full backups are typically scattered throughout a week, simultaneous full restores may not offer the same flexibility. LAN-free backups (SAN) can also offer an alternative backup and restore path.
- **Data interleaving:** Avoid using interleaving to optimize backup performance. It may be tempting as a means to increase the number of simultaneous backup and restore streams, but it carries a cost in terms of performance.
- **Scheduling:** Backup and administrative task schedules should be reviewed for opportunities to spread the workload over a 24-hour cycle when possible.
- **Monitoring:** Proactive monitoring is by far the best way to ensure and maintain backup and restore optimization. Backup products all offer various levels of monitoring capabilities that can be enhanced with third-party products.

Disk backup Expert: Hardware mirror migration

Ashley D'Costa

June 2, 2005

I have a server running Win2k Server, it is a P4 2.0Ghz (EpoX). It has two drives ATA133 wired and the board doesn't have [SATA](#) sockets. I heard that software [mirroring](#) isn't that bright, can you please tell me how to migrate to a hardware mirror without losing any data and maybe not having to reinstall the [OS](#)? We recently suffered a crash where an imaging program was used to create a clone on a second drive. It wiped out the servers system drive and also the other PC's C drive that I used to create the clone with. It's a mystery. Please help.

First off, I would strongly suggest that you discuss this problem with the manufacturer of your OS imaging software (i.e., place a technical support call). In no way should any kind of imaging software wipe out any data residing on the source. I would suggest that unless the software is a hack, it is possible that you cloned the OS in reverse by mistake (i.e., cloned the blank data from the destination PC to the source PC which would result in two blank systems as was the end result in your case).

Once you've determined the problem and resolved it, I would say that you are on the right track. You've chosen a good easy option which is to use OS imaging software that can do a recovery (such as Symantec Live State); although, in your case, the software might be of the more dubious nature if it's not working as advertised. The more tedious method is backing up and restoring the OS using [backup](#)/recovery software such as NT backup that is built into Windows or whatever third-party backup software you may have. The only final suggestion I would have is perhaps to consider server virtualization software such as VMware or Microsoft Virtual Server. These products allow multiple Windows systems to run simultaneously as "virtual machines" (VMs) on one physical host server and OS. The VM's drives are simply large files that the VM boots its OS from. As a result, this makes your OSs highly portable since all you have to do to move an entire server is to simply shut down the VM its running in, copy the files associated with the VM to another set of drives on the same server (or another physical server running VMware/Virtual Server) and then boot up the VM again to have the entire system back, but on completely different storage (or completely different server).

Some companies have only one VM running on their physical servers just so that they have the option to move them around while they are still running using tools such as VMotion from VMware. Also using special tools, you can create virtual machines from existing physical servers. VMware has a tool called P2V that creates virtual machines out of physical servers, thus allowing you to migrate to a server environment without having to reinstall and start from scratch. VMware also allows you to import images created by Symantec Live State. I've successfully imported Live State images into VMware and have had a running virtual machine on a different server exactly the way it was on the original server in less time it takes to install the OS, to install all the applications and to recover the data associated with the applications.

To sum up, I believe that you are on the right track with using an OS imaging tool to move your OS, but I would first reconsider the software you are using or at least follow up with the manufacturer to solve "The Mystery of the Wiped out PCs" before moving forward.

Disk backup Expert: Mirroring HP EVA4000 data

Pierre Dorion

June 02, 2005

I have a setup of two HP EVA4000 [arrays](#). These are connected to two nodes participating in a cluster via [Fibre Channel \(FC\) switch](#). I want to mirror the entire data of one EVA4000 into the other EVA4000. I do not want to implement Continuous Access Synchronous Replication.

Is there a reliable method to keep a mirror copy of the EVA4000 data? I want to implement host-level [mirroring](#) or software mirroring that can do this. Can you let me know how I can do this?

There are quite a few reasonably priced software products available to achieve that particular level of data protection. Neverfail, Datalink, Data Domain, XOSoft, TimeData and FalconStor (just to name a few) provide various features and levels of functionality. However, simply naming "the best product" is not that easy. In your particular case, you indicate that you want to avoid using HP's Continuous Access Synchronous Replication. The situation seems to indicate you are seeking a point-in-time or snapshot type of [replication](#) rather than [synchronous](#) mirroring; the assumption is that you probably want to mitigate the risk of suffering data loss or corruption on two copies simultaneously.

In any case, you must answer a few questions before selecting the type of product that will best meet your business requirements:

- **What type of data are you trying to protect?** Certain products are database-aware and will take into consideration logs and configuration files during creation of the copy or snapshot. These products interface with the database application to ensure data integrity and provide the ability to "roll-forward" using transaction logs.
- **How "new" [cluster-aware](#) must the software be?** Some of the products listed above will range from fully automated [failover](#) capability to failover with manual intervention (i.e., Datalink, XOSoft or NeverFail).
- **What level of granularity do you require with respect to the recovery point objective (RPO)?** Some software products can "journal" the changes allowing you to "play-back" changes to a specific point-in-time. (i.e., TimeData) or allow multiple point-in-time copies (FalconStor).

By first answering these questions, you will be able to narrow down your search to the products that meet your requirements. The rest of the selection process can then focus on added functionality and cost.

Disk backup: How to troubleshoot your D2D2T system

Rick Cook

April 4, 2006

What you will learn from this tip: Checking your tape drive, inspecting your logs and adjusting your buffers can help you to determine what may be wrong with your D2D2T system.

The most common problems with [disk-to-disk-to-tape](#) (D2D2T) [backup](#) systems are related to performance. It's not that the system doesn't work; it's that it doesn't work well. Usually this means one of two things: Either the parameters, such as buffers, aren't set properly or your tape's performance is less than optimal.

Here is a checklist you can use that can help you determine what may be malfunctioning in your system.

The Easy Stuff

- **Check your tape drives**

When the performance of a D2D2T product deteriorates, either suddenly or slowly, the first thing to check is the tape drives. At times, the drive needs cleaning, and because of this, it is forced to rewrite sectors. Cleaning the heads will often fix the problem.

- **Check your tapes**

A related problem is worn or poor-quality tapes. Try using new tapes from a major manufacturer and see if performance improves.

Tuning

D2D2T systems are inherently complex. That means they have a lot of opportunities for tuning — and a number of places where improper settings can compromise performance.

- **Start with your logs**

You should leave logging enabled on all parts of the backup system. The performance cost is minor and the information in the logs is invaluable.

When you suspect a problem with your D2D2T system, inspect the device logs to see what is actually happening — meaning, check the logs for the network connecting the backup disk array to the main disks and the network connecting those disks to the tapes, as well as the logs for the devices themselves.

Ideally, you should save a set of base logs containing 'normal' data that you can compare the current logs against. You can also find typical throughput parameters in the system's documentation. Use this information to narrow your search for the problem.

- **Pay attention to throughput**

In theory, every section of the D2D2T system should be working as hard as it can.

The most critical throughput parameter is feeding information to the tapes. Tape devices are notorious for reacting poorly when data arrives too slowly. This can result in shoeshining and increased wear on the tape mechanism as well as poor performance. Of course, trying to feed the tapes data faster than they can absorb it will also cause problems for tape devices. Make sure that your system is feeding information to the tapes at the recommended speed.

- **Add data streams**

Many D2D2T systems can handle multiple data streams between the primary storage, the backup storage and the tape devices. Additional streams can considerably increase performance by optimizing data flow. However, trying to handle too many data streams can reduce performance as well. Make sure you have the correct number of data streams to maximize throughput.

- **Adjust your buffers**

D2D2T systems have a number of buffers that can be used to match the throughput of the various parts of the system. The default buffer sizes are set by the vendor or system integrator to match the expected loads. However, if your system's loads are unexpectedly large, or especially bursty, the buffer sizes may be inadequate. Increasing buffer sizes can even out the flow of data and make the system perform better.

A word of warning, though: Increasing the buffer sizes can mask problems as well as solve them. Buffers are only there to handle temporary mismatches in data flows. Always try to understand the conditions that cause you to need to increase the buffer sizes and correct the underlying conditions if possible.

Disk backup Expert: Narrow down your replication or CDP options

Pierre Dorion
January 9, 2006

I'm considering making disk backups of a share disk in my network, aprox 150 GB, but I don't know if there a software to do this fast enough. I would like some software that will allow me to have a replica of my disk and update files like the synchronization on PDAs.

The list of software products that can help you accomplish this task keeps growing. There are host-based [replication](#) products, hardware-based replication products (at the disk array level) and there are now a growing number of [continuous data protection](#) (CDP) products available. Given the relatively small amount of data you have to replicate, a software-based product is probably better suited unless you already own two storage [arrays](#) that are compatible for hardware-based replication.

Below is a VERY PARTIAL list of products that can help you with this (alphabetically):

- Data Domain
- Doubletake (NSI)
- Falconstor CDP
- Neverfail
- Replistor (Legato)
- LiveState (Symantec)
- TimeData (TimeSpring)
- XOSoft

The above are not necessarily listed as the only recommended products and there are other equally viable alternatives.

To narrow down your options, you need to consider the following criteria:

- Distance between data source and destination (this will determine whether you can do synchronous or asynchronous replication. This will influence performance to various degrees.
- Features and functionality (i.e. how many point-in-time copies, block- or file-level replication, [restore](#) granularity, etc.).
- Ease of management -- How user-friendly is the interface and what kind of training will be required (if any)?
- Cost (capacity-based versus host-based licensing should be compared).
- Scalability — This will depend on your anticipated future growth.

The first step in selecting the right product for your organization is to clearly establish your business requirements (number of copies, retention, recovery time objectives, etc). You can then do a little research on the Web using keywords such as "data replication" or "CDP" to gather a listing of products available. Once you have narrowed down your choice of products, many vendors offer trial software or can set up demos for you.

Disk backup: Virtual tape evolves to survive

Alex Barrett, Trends Editor

July 19, 2005

What you will learn from this tip: Information about virtual tape libraries (VTLs) and the changes currently occurring in the VTL world.

Disk backup has arrived, and it's not going anywhere anytime soon. But VTLs have a long and winding road ahead of them.

In a nutshell, a VTL works by "faking out" the backup software and presenting a disk array as a tape device, says Frank Sloodman, president and CEO at Data Domain, Palo Alto, Calif., which makes a disk-

based backup system. This is useful if your backup software doesn't natively support disk targets, or if you're making inefficient use of your tape devices and need to improve your backup performance.

The list of VTL-based disk backup products is long, and includes ADIC's Pathlight VX, EMC's Clariion Disk Library, Quantum's DX-Series, StorageTek's Virtual Storage Manager (VSM) and VSM Open, and the latest entrant, Hewlett-Packard's StorageWorks 6000 Virtual Library System. There's also a bevy of smaller firms peddling VTL products, including Diligent Technologies, FalconStor (which also resells its software through third parties), Neartek and Sepaton. Backup software vendors such as Arkeia and Atempo include VTL features in their software, and others are rumored to be following suit.

But for some customers, VTL just doesn't cut it. Take Scott Roemmele, for example, SAN engineer team leader at Quicken Loans, an online mortgage lender in Livonia, Mich. Last year, his company began looking for a way to speed up mailbox restores for its bankers — a quest that led Quicken Loans to look at different disk-based backup products. "We've found that most mailbox restores happen within 14 days of deletion, so if I can keep up to 21 days of backups online, I don't have to go to tape," says Roemmele.

The VTL products he looked at, however, were priced too high to justify multiweek retention times, and the enhanced tape-media utilization VTLs provide didn't bring enough benefit either. "A VTL was just too costly for what it brought to the table," says Roemmele. "It seemed kind of frivolous." Instead, Roemmele installed a DD200 from Data Domain which, thanks to specialized capacity optimization technology, can store approximately 10 terabytes (TB) of data — or 14 to 18 days of Exchange backups — on only 1 TB of physical hard disk drives.

However, VTLs aren't only about enabling longer retention periods — they're also about achieving better backup performance than you can get with generic disk arrays. "You can't just put a disk on the network and call it a backup," says Shane Jackson, director, enterprise product marketing and strategic alliances for Quantum's storage systems business unit. "That can cause more problems than it solves." Backup data, he explains, is written sequentially; if written to a random-access device like disk, it can result in poor performance and disk fragmentation. A good VTL package compensates for those discrepancies and can deliver optimal performance with the fewest possible tape drives. Quantum's DX30 and DX100, for example, are designed to back up 1 TB and 2 TB of data per hour, respectively, reports Jackson.

Furthermore, their tight integration with standard-issue backup software such as Veritas NetBackup, make VTLs a great "risk mitigation play," says Ray Anderson, director of IT at Egenera Inc., a blade server vendor in Marlboro, MA. Egenera is a Sepaton customer, and internal tests have concluded that the presence of the VTL appliance will dramatically improve restore times, although Egenera has yet to try and do a real restore.

VTL vendors have also started to heed the call for greater capacities. Sepaton, for example, has increased capacity with its latest VTL appliance, the S2100es, which goes beyond the previous limit of 200 TB all the way to 1 petabyte (PB). Furthermore, the appliance supports software-based compression, which compresses the data by roughly 2:1.

Quantum has added optional hardware-based compression to its DX30 and DX100 VTLs, which according to

Jackson, enables customers to increase their backup retention periods to 30 days or more.

These are wise moves, says Tony Asaro, senior analyst at the Enterprise Strategy Group, Milford, Mass., but in the long term, they probably won't be enough. "The VTL guys will have to add more and more capabilities if they're going to make it more than just [tape] emulation," he says. Besides capacity optimization, Asaro sees a need for additional replication capabilities and features such as synthetic full backups. "Unless the VTL guys start doing this...it's going to be a stop-gap technology."

However, it will be a very long time before VTL becomes extinct, if ever, says Brad O'Neill, senior analyst and consultant at Taneja Group, Hopkinton, Mass. "The complete replacement of tape-centric technologies will take the better part of a decade," he says, during which time "we'll have a co-existence of VTL disk technologies alongside pure, disk-as-disk-based data protection solutions."

Disk backup: Case in point — Replacing tape backup with Avamar

Jon William Toigo

April 18, 2005

Ed Holmes grew up using tape for data retention and protection. Over a period of many years, his company built what was, in his view, a robust tape environment. But their last round of infrastructure upgrades, which were intended to serve the company's tape needs for the next four years, ran out of steam after only 18 months.

That's when Holmes started looking for an alternative. And that's when he discovered Avamar Technologies in Irvine, Calif.

Ed's situation is popping up in more and more companies of every size and in every industry segment. Like many others, Holmes came to manage storage as a sideline to his server administration role. He built a tape backup solution that steadily consumed more resources over time.

"Growing our tape environment translated to steadily increasing costs in terms of hardware platforms and upgrades, supplies, management time and service time. Media management was creating a manpower issue, and new regulations were adding more complexity than before," Holmes said. "At last count, we had about nine terabytes (TB) of data going to tape every month: 30 days of rotation in the library, and more offsite. I started to see it as money sitting on a shelf somewhere."

Moreover, he said, analyst reports pointing to vulnerabilities of tape media were becoming a concern for him, "A Gartner analyst told me that between 20% and 40% of tapes on average failed on restore. I had seen a small number of bad tapes in my career, but never on that scale." It concerned him that he might need to devise a program to test and recycle tapes more closely, he added.

Holmes' doubts about tape reliability also increased as tape densities improved, "The industry has been

squeezing more and more data into the same space on their media. At one time, if you lost a tape, you lost, at most, a few megabytes. If you loose a tape today, you are looking at about a half-terabyte of data." All things considered, Holmes decided to look at the alternatives. He says that the idea of copying data directly from disk to disk (D2D) — particularly to large, inexpensive Serial-ATA (SATA) disk — piqued his interest. He tested several products under workload to see whether the technology had matured beyond the "renegade, cutting edge" phase.

He quickly dismissed D2D products that made the second disk tier a surrogate tape target, "At first, it seemed like a good idea to do tape emulation on disk, but there was no convenient way to move it offsite [so the data was still exposed to disasters]. Plus, it was a waste of network capacity to run backups from a server or desktop system."

It was then that he was approached by Avamar, and the more he learned about the company's Axion solution, the more impressed he became.

"It sounded like a more powerful computing solution. Basically, you would let the storage solution roll up your data and put it away for you," he said. "Axion technology parses data on existing storage, and migrates it, in accordance with user defined policies, to a content-addressed storage repository. In the process, data is compressed to a fraction of its size on the primary disk."

Holmes says that he invited Avamar to present their solution several times before he licensed the product, "They needed to show me better density than what I had with tape in order to make their solution attractive," he said "They had to prove that their numbers were legit."

In the middle of 2004, he initiated a live test of the product in his shop, running Axion in parallel with his existing tape solution. At that point, Axion was "not an off-the-shelf, plug-and-play" solution, he said.

"The first snap up [copy of all data in the test storage platforms] took a long time. Afterwards, it was very fast. Still, we waited for changes and upgrades to be made, and we took an additional 3 to 4 months to make sure that we were choosing the right approach," he said.

By the beginning of 2005, his confidence in the solution was unshakeable. "We recently did a test data restore comparison: it took 15 minutes to recover a data set from tape, while Avamar's Axion did it in three-and-a-half minutes," he said "Bottom line: we were ready for a full-fledged roll-out. And what's more, I didn't just want to use Axion for the protection and management of 17 TB of data in this data center, I wanted to use it throughout the company for all of our data: the whole data enchilada." Holmes is currently rolling out the Axion solution within his Milpitas, Calif., data center, owned by Adaptec, where he serves as a network systems administrator. Holmes was concerned that his Axion solution might be passed over by readers who would view his situation as unique. However, he argues that, when it comes to its IT systems, Adaptec is as conservative and budget conscious as any company.

Given the cost savings compared to tape that are expected to accrue to the Avamar solution, he said, "We would have had to have adopted it even on a futures basis."

Avamar's Axion is appearing in more and more competitive bids for data protection solutions. The product is offered in three different versions, including a software-only option that consumers can implement on whatever storage hardware they choose. Worth a look.

Remote backup: Take full advantage of the remote replication process

Pierre Dorion

April 18, 2006

What you will learn from this tip: With the increasing availability of cost-efficient products, many view remote data replication as the DR strategy of choice to address tighter RTO and offsite media handling issues. This tip outlines the other considerations you should keep in mind that can help you take full advantage of this technology.

We have seen a significant increase in the number of low-cost data [replication](#) products available to [small and midsized](#) (SMB) IT departments in the past few years. Given the technology's enhanced performance from a [backup](#), and, even more, a restore perspective, many see it as an efficient and cost-effective answer to tighter recovery time objectives (RTO). Decreasing bandwidth costs have also contributed to the technology's gain in popularity.

The ability to quickly access an online (or nearline) copy of the data at an alternate location, the elimination of lost or damaged tape media typically associated with handling and the reduction in manual interventions are all factors that make remote data replication an appealing [disaster recovery](#) (DR) strategy.

However, this rapid recovery shifts our attention to other considerations that may not have been as high on the priority list with traditional tape backups.

The list below can help you to take full advantage of this technology:

- **Data categorization:** An organization would likely not enable remote replication for all data for cost-saving reasons. This introduces the need to identify business-critical data and assign priorities. In essence, this is the development of service levels or tiers.
- **Data access:** With traditional tape backups, data access issues could be addressed while data was being restored to a replacement system. The ability to almost instantaneously access a remote copy of the data means that applications and users must be redirected to that data just as quickly. This implies some planning around network path redirection.

- **Data synchronization:** You should identify dependencies between various data sets during categorization. Replication may allow a recovery point that no longer matches that of other data backed up to tape, thus introducing data synchronization issues. It may become necessary to promote certain lower criticality systems to a higher priority level based on interdependencies.
- **Documented procedures:** Remote data replication is definitely the DR strategy of choice but does not constitute a DR plan (DRP). No matter how seamless and automated the replication, configuration and procedures must be documented and integrated with the DRP and, ultimately, the business continuity plan. Maintain documentation on configuration and procedures documentation on a regular basis — or when significant storage changes are made — so it remains current and relevant.
- **Monitoring, notification and testing:** As with any other backup procedures, you should implement monitoring and replication failure/success notification. Validate the replicated data on a regular basis to avoid unpleasant surprises. Test the system's ability to acquire replicated data sets regularly to ensure that any issues or undocumented changes are captured and remedied.

You should also consider the data protection requirements at the remote location. In the event of a disaster causing the loss of the production copy of the data, the remote replica is promoted to production status until the primary site is restored or repaired, which can take quite some time. This replica is now the only production copy of the data and should therefore benefit from some form of data protection to reduce the risk of data loss. A corporate decision must be made with respect to the service levels for data availability while the organization operates in recovery mode.

Like anything else, a recovery strategy is only as good as its weakest element. Without some forward thinking and comprehensive planning, remote replication is like driving to an unfamiliar destination without directions; you might get there...eventually.

Remote backup Expert: Oracle replication for failover

Pierre Dorion
August 19, 2005

We are planning a Oracle data replication (using Veritas VVR or Sun SNDR) between a primary and secondary site. The application is Web-enabled and the users access the application/database through the Internet. In case of failure of the primary site, how will the users get automatically routed to the secondary site so that services can continue without any downtime?

This depends primarily on how the Internet users currently access the application. Without knowledge of your environment and supporting infrastructure, it can be assumed that if users access the application via a URL and DNS, it should be practically seamless to the users provided the IP address of the secondary (backup) system is same as the failed system. In this context, the procedure is not much different from when an application is migrated to a new server as part of an upgrade.

Some vendors offer [replication](#) and synchronization products that can replicate Oracle data remotely and take care of client connectivity updates in the event of a [failover](#). XOssoft comes to mind but there are other products available. That said, many of these products include a high availability component that may need to be considered as part of your current plans.

There are also failover-capable network appliances that can help with automatically rerouting network traffic to an alternate location. However, this implies that the network redirection device is not affected by the outage.

Remote backup Expert: Hot-hot replication with EMC's SRDF?

Evan Marcus

June 10, 2005

Is hot-hot replication possible with EMC's SRDF? Is it a fair statement that the only way to be hot-hot is to have non-SRDF storage on both the production and [disaster recovery](#) sites?

It's not really a shortcoming of SRDF that prevents it from doing hot-hot [replication](#). Hot-hot replication is also called bi-directional replication, which means that you can update the data on one side of a replicated link, and the changes will be replicated on the other side.

It is nearly impossible to support bi-directional replication over any distance. If you have a copy of a file on both sides of the replication, and both are written to at the same time, which side's writes should be saved, and which side's should be overwritten?

Bi-directional replication can be performed in some replicated databases, where individual transactions can be controlled, but I would be very careful about implementing that feature. If I needed it, I would test it very carefully before using it in production.

Some organizations are content to have a dataset at site A that is replicated to site B, and then a second, independent dataset at site B that is replicated to site A. That arrangement is fine, since each dataset will only be written to by one side at a time.

Remote backup: Remote replication gets out of the array

Alex Barrett

December 2, 2004

If you're a Symmetrix user and have stringent remote replication requirements, Symmetrix Remote Data Facility (SRDF) is pretty much the only game in town. But SRDF has its limitations: It's extremely expensive and only works within Symmetrix environments.

Last month, EMC chipped away at those complaints with the introduction of EMC Open Replicator for Symmetrix, which copies data between a Symm and IBM Shark, HDS Lightning and HP EVA. In addition, EMC added SRDF/Star, a three-site replication suite that allows a Symm to replicate synchronously and asynchronously at the same time to two separate sites.

Today, the replication market stands at about \$1 billion, and is split about 70/30 between array-based replication such as SRDF and IBM's PPRC, and host-based replication such as Veritas Volume Replicator, says Arun Taneja, founder of Taneja Group. SRDF has "the lion's share" of the array-based replication market, he says. For EMC, "it's the goose that laid the golden egg."

But over the next couple of years, the majority of replication functionality will be network-based, Taneja predicts, and anyone who's still doing array-based replication will "need to have their head examined."

Two startups assailing array-based replication are Kashya and Topio. According to Taneja, they are good choices for environments with heterogeneous storage and "where replication is very, very important."

Architecturally, the two products are different. Kashya implements its software on an out-of-band, SAN-connected appliance, the KBX5000, which handles tasks such as serializing I/Os and performing data reduction. Topio, on the other hand, eschews the appliance model because it can become a bottleneck, says Betty Woychowski, Topio's director of product management. Instead, Topio Data Protection Suite (TDPS) 2.0 consists of the Topio server, located at the remote site, and Topio agents, which rely on a master clock to time-stamp the data and send it directly over the WAN to the Topio server. The remote Topio server then parses the I/Os back into a consistent stream.

Despite their differences, Kashya and Topio share some features: Both use asynchronous replication over an IP network and, according to Taneja, "get an 'A' for thinking about issues surrounding data consistency," adding that they represent "the next generation of data replication."

Business Continuity: Cost-effective legacy data protection

Rick Cook

August 15, 2006

What you will learn from this tip: Not all of the data you need to back up and save was produced today — or last week, or even last year. For business, legal, regulatory and compliance reasons, it is often important to protect data that is a decade old or more. Learn what you can do to cut data protection costs on older data.

Economics rears its ugly head in a big way when dealing with legacy data. Most of the time all of that old data represents a cost sink rather than a profit center. You may need to keep it on hand in case the regulators or lawyers come calling, but you aren't going to be able to generate any more value from it. At the same time, the data may have been recorded on systems that are three or more generations back and you'd like to keep it in a form which is readable by your current systems.

While you can and may just store those old reels of 9-track [tape](#) written in VAX VMS format and hope that you never, ever need that data (or if you do need it you can somehow, somewhere find a system which will let you read) it is usually better to store that data on media and in a format that you can read. Assuming, of course, you can do it cheaply enough.

Fortunately, there are a number of things you can do with that legacy data to reduce the cost of keeping it around in a readable format.

- **Analyze it**

The first step is to figure out what you've got, how often you're going to need it and what form you're likely to need it in. "Legacy data" covers an enormous range of material, with an equally enormous range of value and accessibility requirements.

Some organizations, such as oil companies and scientific institutions, have large amounts, often multiple [terabytes](#) (TB), of data that may have been collected years, or even decades, ago — and which is still frequently used.

Most enterprises will have data that must be preserved for regulatory or legal reasons and which will probably never be looked at again. However, some of that material, like [archives](#) of email messages, will need to be searched through quickly for specific message threads if it ever is needed. You've got to know what to do with it.

Data formats are another important consideration. You not only need to have the data on media you can read, you need to have it in a format your current systems can handle. It doesn't do any good to carefully transfer those old files onto new media if the files are formatted for an application you discarded years ago. You may have to convert the data, as well as translate the media.

- **Prune it**

The real question is how much of this data do you want to protect? Typically a lot of 'legacy' data, perhaps 80% of it or more, isn't needed. It makes sense to do some serious housekeeping before you do any conversion.

Many of the decisions on what to keep and what to discard can't be made by IT alone. They require input from the people who generated the data in the first place, as well as other departments such as legal and accounting.

- **Select the right technology**

After pruning, the data you're left with may have to be kept around forever. This introduces some considerations in storage. Cost per gigabyte isn't the only consideration in choosing a technology for storing old data. All existing media have a certain lifespan and preserving data permanently means rewriting it to media before that lifespan expires. True, that will typically be 10 years or more down the road, but you need to consider the cost of transferring the data when the time comes. It may make sense to choose a longer-lived medium, such as optical disk, even if it has a higher cost, to cut down on the expense of later storage transfers.

It also pays to think ahead, especially in the area of formats. For example, converting text-type data to XML will make it a lot more accessible and easier to manipulate in the future — factors that may pay off. Similarly, it's obvious that you're probably going to want to convert EBDIC data to ASCII, but you might want to consider taking that a step further and putting text data, EBDIC or ASCII into Unicode format.

- **Consider outsourcing**

Even with careful pruning, legacy data can amount to several TB of information. In the case of large or complex data migration projects, it may be more cost-effective to outsource the conversion and associated services.

There are a number of companies which specialize in transferring data, including [Disc Interchange Service Company](#) (DISC), which has a number of brief articles on various aspects of file conversion available on its web site, and [Appian Analytics](#).

- **Store it appropriately**

Storing media under the proper conditions will significantly prolong its life. For most media, especially tape, the most important factors are temperature and humidity.

The other issue is making sure the media containing your legacy data is properly indexed and cataloged. Make sure all the media are properly labeled and you have a catalog showing where each tape or disk is stored. Then make sure it is actually kept in that place.

Business Continuity: Risk management — Know your storage risks

Pierre Dorion
June 20, 2006

What you will learn from this tip: Risk management is a complex discipline and covers a broad area ranging from business and operational risk to the more focused IT risk. This tip narrows down the focus to specifically identifying data storage risk.

Advanced Storage Chapter 2: Backup/Data Protection

At the highest level, an enterprise risk management program would consider elements such as market demand, competition and the state of the economy to be business risks. Operational risks are also considered and business resilience, or the ability to resume business in the event of a disaster, is normally included. This is where [business continuity](#) and IT [disaster recovery plans](#) (DRP) come into play.

Every good DRP should always be based on a recovery strategy that was defined prior to developing the plan itself (hence the term planning). The ideal recovery strategy is not pulled out of a hat, but rather is based on the understanding of the threats to which our IT environment is exposed, its vulnerabilities, the probability of occurrence and the impact to the organization. This essentially summarizes the IT risk assessment process.

Without digging too deep in the specifics of qualifying or quantifying risk, let's examine some of those risks. It should be noted that the following list is by no means exhaustive or complete but is merely a starting point. Risk can vary widely based on geography, climate, level of preparedness, corporate culture and more.

Backup storage and the risks involved

Single copy backups	Exposure to data loss in the event lost or damaged tapes
Daily backups but weekly offsite	Exposure to a much as one week of data loss if the main facility housing the production data is destroyed
The offsite vault is the trunk of your car	Hopefully, this exposure requires little explanation
Backups exceeding available window	Can impose backup schedules that leave the data exposed. For example, full backups are only run on weekends because they take more than 24 hours and are only sent offsite on Monday.
Unencrypted data on offsite-bound media	Can cause a security issue in some cases (industry specific)
Poor or inexistent change management	Poorly planned changes (configuration changes, software upgrades, etc.) are at the root of many failed backups and creating an exposure to data loss.

Disk storage and the risks involved

Replication or synchronization utility errors	If the production copy of a database becomes corrupted or unusable, is it possible to overwrite the replicated copy with the bad copy by mistake in your environment? Is there a mechanism in place to prevent that from happening?
Hardware failure (or SPOF)	Often seen as stating the obvious but single points of failure must be identified from the host all the way to the allocated storage.
Insufficient storage masking, mapping, etc.	Many storage experts agree that storage area network (SAN) storage access should be controlled at the HBA , Fibre Channel (FC) switch and disk array level to avoid device contention between hosts
Poorly documented custom configuration	Exposure to knowledgeable staff being unavailable following a major outage or disaster
Lacking segregation of duty	Too many IT personnel with unrestricted access to storage configuration interfaces or utilities can lead to inadvertent changes or poorly communicated actions
Poor or inexistent change management	Change management is probably one of the most common vulnerabilities but is all too often overlooked because IT personnel typically don't see themselves as a threat agent. However, poorly planned changes are frequently identified as the cause for storage failure or data loss.

Obviously, IT environments are subject to many more internal or external threats that can indirectly affect storage and an attempt at listing them all would exceed the scope of this tip. Some examples include power conditioning, environmental controls, physical security and data integrity. There are a number of publications available on storage best practices and this site offers a lot valuable advice on the subject. Hopefully, this tip will have helped get the thought process started.

Business Continuity: Restoring data — Increase your efficiency

Pierre Dorion

May 9, 2006

What you will learn from this tip: This tip highlights an aspect of backup that most are not as experienced and comfortable with — restoring.

Attempting to make a distinction between [backup](#) and [restore](#) may seem like pure semantics at first; after all — don't we always refer to backup and restore as one discipline?

Chances are, your data backup administrators have become experts at making sure backups run successfully each night. Over the years, they have likely fine-tuned numerous performance parameters and adjusted schedules to better utilize available hardware resources and capture data at the best time. This experience was acquired through years of daily backups. However, the same cannot be said about data restores and this is probably why we seldom use the term "restore administrators."

Ironically, we spend years learning how to master backups and then one day we are faced with having restore an entire environment in 24 hours! Shouldn't we learn to master the restore process instead? The problem is that we don't restore often enough, and when we do, it is typically a partial restore.

As we design our backup strategy, we need to start planning for restores. Here are a few items that should be addressed in order to increase the efficiency of the restore process:

- **Complete restore test:** The ability to successfully restore critical applications from backups should be tested regularly. Avoid shortcuts; every step of the process should be executed and reviewed for possible issues. Restored data must also be validated to ensure everything is as it should be.
- **Documentation:** As the above restore test is performed, each step should be documented. This will ensure that a single employee does not possess all the knowledge. The documented procedures can then become part of a master plan.
- **Restore dependencies:** Some components of an environment need to be recovered before others can be restored. It is surprising to see how often authentication and networking is overlooked. This must be understood and documented.
- **Restore priorities:** It is essential that the restore priority of applications be clearly established ahead of time by the various business functional areas. The IT team can only restore so many applications at once and unless properly informed, IT will set priority based on their understanding of criticality.

- Know what to restore: If your recovery strategy includes a number of standby systems with a loaded operating system and applications, is it necessary to restore any of those files? This requires some backup planning to avoid having to sort out what to restore after the fact.
- Backup data retention: How backups are expired must be closely monitored and managed. For example, the relationship between full database backups and associated logs must be clearly understood. Don't assume that automated deletion always works flawlessly and give it a regular check.
- Have a plan: Last but not at all least is to have a plan. The critical applications recovery procedures must include the data restore and validation procedures. These procedures are then included as annexes to the disaster recovery plan and ultimately, become part of the overall business continuity plan.

Business Continuity: Better DR and BC planning

Greg Schulz

January 12, 2006

What you will learn from this tip: A list of items to consider and address as part of managing and maintaining a business continuance and disaster recovery plan.

Business continuance ([BC](#)) and disaster recovery ([DR](#)) continue to be a main focus from a storage perspective for a number of reasons:

- Information privacy, data protection and data retention are in the spotlight
- Any business is at risk, not just large enterprises
- The risks and threats to security, including cyber attacks
- Regulatory and [compliance](#) requirements ([HIPAA](#), [SOX](#), etc.) are increasing
- The amount of data and its value continues to grow.
- Data is being retained for longer periods of time and in more locations.
- More critical data exists outside of traditional [mainframe](#) environments.
- Time for recovery, [restoration](#) and restart continues to shrink if not already non-existent.

Part of managing, implementing and maintaining a BC and DR plan is understanding what level of data protection, availability, accessibility and service levels, including recovery time objective (RTO) and recovery point objective (RPO), are needed. These [service level](#) items need to be understood, as well as applicable threats and risks for:

- Your entire business (e.g., all applications).
- Specific applications and business functions.
- Different types of data for a given application or location.

Understand what the applicable threats are to your system and categorize your applications, data and storage to enable the appropriate level of protection to counter those threats. Different applications and data have different threats and protection needs, thus requiring tiered data protection. Align the proper RTO and RPO to the specific business function, application, data and storage. Understand your data availability, accessibility and retention requirements, as well as their interdependencies upon other applications and technologies.

BC and DR storage related tips include:

- Leverage fault isolation to contain faults and prevent them from spreading into a rolling disaster resulting from a chain of events. This includes eliminating single points of failure and combining various data protection and availability techniques to achieve resiliency.
- Update your DR and BC plans, documentation and associated procedures on a regular basis as part of change control management.
- Communicating the aspects of the plan to the key people involved — updates on their roles and responsibilities, who to contact when, how the plan will be invoked, as well as how to protect the plan and documents. These items need to be readily accessible to whomever needs them.
- Your documentation should include inventories of the assets (technology) you have, where they are located, how they are configured and who the applicable suppliers are. In addition, maintain information about software licenses and applicable software keys for technology that you may need to recover, restore and restart your environment.
- Leverage multiple techniques and technologies for a 'belt and suspenders' approach to BC and DR. For example, clustering combined with remote replication along with some form of backup technique. This could include leveraging disk-based backup to tiered storage combined with point-in-time copy and snapshots integrated with applications and database systems.
- Look at the overall RTO and RPO from a business, application, server and storage perspective as the true RTO and RPO are the same for all components, not just the time required to recover your storage.
- Distance is a friend and a foe of storage with respect to BC and DR. From a positive standpoint distance enables survivability and continued access to data. The downside is the cost penalty in terms of expense, performance and complexity.
- Data consistency and integrity is important so make sure that your data is copied intact and that it is still consistent and in the proper sequence.
- Test and audit your BC and DR plans, procedures, policies and implementations.
- Include your key partners, suppliers and customers as part of your plan.