



Tech OnTap

Highlights

June 2008

- [VMware Backup Challenges](#)
- [Deduplication Safety](#)
- [The Right Backup Solution](#)
- [Dedupe Q&A Expert Answers](#)
- [Tech OnTap Archive](#)

Special Edition: Backup and Recovery

Choose the Right Backup Technology

David Chapa, Director, Backup and Recovery Solutions



Difficult recoveries, long backup windows, reliability issues. Solving these problems is no longer a hard choice between deploying the best technology and staying within budget.

Find out about the four classes of backup solutions and how to choose the best option to meet your needs.

➤ [More](#)

Tips from the Trenches

Assessing Deduplication Technologies: They Are Not ALL the Same

Larry Freeman, Sr. Marketing Manager



Deduplication is hot, but can you really count on it at restore time?

Find out the two essential criteria for assessing the safety of the technology before you commit.

➤ [More](#)

Blogging with Dave



Dave Hitz
NetApp
Founder
and EVP

Which Blog Is Dave Reading When He Is Not Blogging?

Cloud Computing by Jay.
What it is, and why it is like
an alien abduction.

➤ [Dave's Blog](#)

Engineering Talk

Tackle the Top 3 VMware Backup Challenges

Darrin Chapman, Product Engineer



Virtual infrastructure backup and restore can be tricky. What works well with physical servers may not once you've virtualized.

Take our quick quiz to reveal any hidden botches in your backup plan and find out the best ways to tackle virtualization backups.

➤ [More](#)

Tech OnTap Spotlight

➤ 2007 NetApp Innovation Awards

With over 185 submissions, 2007 is our most successful year yet. Congratulations to our innovative winners and finalists!

➤ Tough Questions/Real Answers

What kind of real world questions are people asking about dedupe? View a "no holds barred" transcript from a recent live chat.



- [netapp.com](#)
- [Tech OnTap Archive](#)

Tackle the Top Three VMware Backup Challenges

By [Darrin Chapman](#)

The rapid rate of VMware® deployment in modern data centers has created new challenges. Backup and recovery methods that worked well for physical servers may not be up to the job of data protection once you virtualize.

I want to discuss the three main VMware backup challenges:

- Reducing the risk associated with consolidating tens of physical servers onto one ESX server
- Improving backup performance
- Eliminating the complexity of backing up tens or hundreds of virtual machines (VMs)

I'll begin by discussing these problems in a general way, followed by details about specific NetApp® solutions that may help you enhance your VMware backup.

Are You Fully Protected?

When you consolidate servers, you also consolidate risk. If you take 10 or more physical servers and consolidate them onto a single VMware ESX server, you eliminate a lot of physical complexity, but you put a lot of eggs in one basket. Protection levels that seemed adequate for standalone servers may be exposing your operations to unnecessary risks.

Here are a few quick quiz questions to ask yourself:

- **Is my infrastructure as robust as it could be?** Outside of the server itself, such things as RAID-protected storage, replicas, redundant HBAs, networks, SANs, and so on can improve resiliency.
- **Are my backups consistent at both the application and VM level?** For performance, VMs and many common applications cache data in memory. This information has to be flushed to disk, writes to the file system must be temporarily suspended, and the application or VM has to be quiesced to create a consistent backup. More than one early adopter of VMware was surprised to discover that the backups they were creating were not consistent. A consistent backup allows operations to be restarted when a backup is restored.
- **Am I backing up my data frequently enough?** With more applications tied to a single server, you may need to increase your backup frequency, providing multiple recovery points.
- **Is the backup workload on my ESX server affecting business operations?** You can't afford to have backups running outside of your designated backup window and slowing down your business.

A careful review of these questions may reveal some hidden gotchas in your backup plan.

Improving Performance

An ESX server hosting multiple virtual machines may not have the same aggregate I/O or CPU capability as the physical machines it replaces. Bandwidth-intensive backup and restore operations run in VMs can result in server bottlenecks that extend backup windows and decrease restore performance to unacceptable levels.

There are two approaches that you can take to improve backup performance:

- Decrease the workload associated with backup by moving less data
- Offload your backup workload from servers to storage

Eliminating Complexity

Some IT shops run tens of VMs per physical server and may manage hundreds of VMs in total. That's a lot of backup agents to install and a lot to track and manage, and it can be difficult to ensure that everything is properly protected. Some things that may help are:

- Reduce complexity by working at the ESX server level rather than at the individual VM level
- Allow your storage to process backups, retention policies, and schedules
- Choose appropriate management tools

NetApp Backup Solutions for VMware

Traditional backup methods can put a significant load on your ESX servers, resulting in extended backup windows and reduced performance for critical applications when backups are running. NetApp offers a complete suite of solutions designed to reduce risk, improve performance, and eliminate the complexity of backup and restore in VMware (and other) environments while reducing or

Explore

Focus on Backup and Recovery

This special issue of Tech OnTap drills down on critical backup and recovery topics:

- [Choosing the Right Backup Technology](#)
- [Top Three VMware Backup Challenges](#)
- [How Safe Is Deduplication?](#)

Five Tips to Increase NetApp Resiliency

A recent Tech OnTap article examined five little-known things that you can do to improve the resiliency of your NetApp storage systems, including:

- Multipath HA
- Proper sparing
- SyncMirror®
- Nondisruptive upgrade
- Automated tools

➤ [More](#)

NetApp Snapshot Copies and VMware

NetApp Snapshot technology is ideally suited for use with VMware. A recent Tech OnTap article explored five uses of Snapshot and other NetApp technologies derived from it in VMware environments:

- Near-instantaneous VM backup
- Fast and flexible VM recovery
- Accelerated data management through cloning
- Disaster recovery
- Application backup and management

➤ [More](#)

Combining Deduplication and VMware DR

DR for VMware Virtual Infrastructure 3 (VI3) requires that all your VMs (virtual machines) be regularly replicated to a remote site, consuming significant storage and network bandwidth. By using NetApp deduplication on your primary VMware storage systems, you can substantially reduce the amount of data in your primary storage environment. This reduction results in a cascading benefit to your downstream infrastructure, reducing the bandwidth necessary for replication and the storage required at the DR site.

➤ [More](#)

eliminating server load. These solutions can dramatically shrink your backup windows and accelerate restores.

This discussion focuses on five solutions:

- SnapManager® for Virtual Infrastructure
- Other SnapManager Suite solutions for application backup
- SnapVault®
- Open Systems SnapVault
- Protection Manager

If you're already using NetApp storage systems as primary storage for VMware (either by using VMFS or over NFS), SnapManager for Virtual Infrastructure and SnapVault solutions are good options for local and disk-to-disk backups, respectively. If not, you can use Open Systems SnapVault in combination with NetApp secondary storage to achieve the same benefits for non NetApp primary storage. NetApp Protection Manager can simplify backup management in these environments, especially if you have large numbers of replicas to maintain.

SnapManager for Virtual Infrastructure

This data management tool allows you to create fast, consistent Snapshot™ copies of VMs running on VMFS or using NFS datastores. (Raw Device Mapping [RDM] is not currently supported.) It eliminates backup processing from your servers and moves the processing requirement to underlying storage.

Using the SnapManager for Virtual Infrastructure GUI, you can create a backup and retention schedule to regularly protect your VMs. Backups can be performed at either the VM or the data store level. Once a backup is created, it can be mounted by ESX to verify the backup before it is needed.

When a backup is scheduled, SnapManager for Virtual Infrastructure communicates with VirtualCenter and NetApp storage as necessary to coordinate the backup, as illustrated in Figure 1. Because backups occur very quickly and put almost no load on ESX servers, you can create many more backups per day for greater data protection.

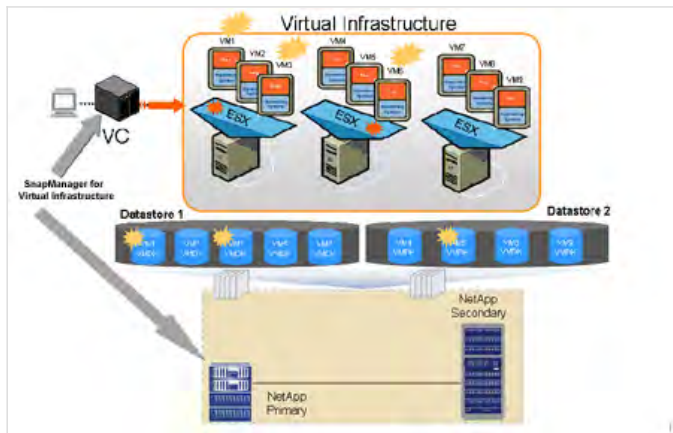


Figure 1) SnapManager for Virtual Infrastructure Snapshot creation. SnapManager for Virtual Infrastructure signals VirtualCenter (VC) to place selected VMs (VM1, VM3, and VM6) in hot backup mode. Next, it triggers NetApp primary storage to create appropriate Snapshot copies. Finally, it signals VC again to take the VMs out of hot backup mode.

Once a Snapshot exists on primary storage, it can be retained locally, backed up to tape, or replicated to secondary storage by using NetApp SnapMirror® for disaster recovery. These features are controlled directly from SnapManager for Virtual Infrastructure. You can also manually coordinate your Snapshot schedule with SnapVault to back up your captured Snapshot copies to a local or remote secondary storage system, as described later.

SnapManager for Virtual Infrastructure restore gives you the ability to restore specific VMs or entire datastores, including the ability to restore VMs that have been deleted from VirtualCenter. The VM to be restored must be in the "powered-off" state. VMs are powered off before a restore is done.

SnapManager for Virtual Infrastructure is fully VMotion™ aware. It communicates with VirtualCenter, so it knows where a VM resides after VMotion has been run. A VM cannot be backed up until the VMotion activity is complete.

SnapManager Suite

SnapManager for Virtual Infrastructure quiesces underlying VMs so that they are consistent when backed up. Unfortunately, it cannot ensure that applications running in the VM are also quiescent and consistent. This requires more application specificity. For consistent application backups, you can install NetApp SnapDrive® along with one of the other SnapManager solutions (providing tailored protection for applications including Microsoft® Exchange, SQL Server™, and Oracle®) to create consistent application Snapshot copies. [A recent Tech OnTap article](#) provides more information about these products as applied in physical server environments. Like SnapManager for Virtual

infrastructure, the load that these SnapManager products put on the server is minimal compared to traditional backup.

NetApp SnapVault

NetApp SnapVault enables you to vault Snapshot copies to secondary storage, either in the same data center or at a remote site, for longer-term retention. Unfortunately, there is currently no explicit coordination between SnapVault and SnapManager for Virtual Infrastructure or other SnapManager products, so you have to write a script to coordinate the Snapshot schedule that you create with these products with your SnapVault schedule.

Because all the work occurs on the storage system, SnapVault operation has no impact on your ESX servers. Each SnapVault backup is a read-only version of a file system at a particular point in time. These file systems can be shared or mounted and used for various purposes such as cloning, testing, recovery of VMDKs, and so on.

The first step in creating a SnapVault relationship is to perform a baseline transfer that makes an exact replica of the data store (containing VMs) to be protected. Subsequent SnapVault backups transfer only data blocks that have changed since the last backup, so it is highly efficient in terms of both network bandwidth and storage space. Most backup methods back up entire files, even when only a single block has changed.

SnapVault backups avoid the high level of duplication that occurs with traditional backups that create copies of the same files day after day, even when they have changed very little. Only blocks that change are replicated and stored by SnapVault. This makes SnapVault backups highly efficient in terms of space consumption and changes the economics of disk-to-disk backup versus tape. (See the [companion article](#) in this issue for more details.)

Once your backups are vaulted on secondary storage, you have several options:

- You can recover entire data stores (qtree restore). Use of Protection Manager (optional) simplifies this function.
- You can recover individual VMs (with NFS datastores). Use of Protection Manager (optional) simplifies this function.
- You can recover single files from within a virtual machine by using FlexClone®.
- You can clone the replicated VMs on your secondary storage with NetApp FlexClone and use the clones for testing and development.
- You can replicate your backups off site for regional disaster recovery.

For more information about using SnapVault with VMware, see [TR 3610](#).

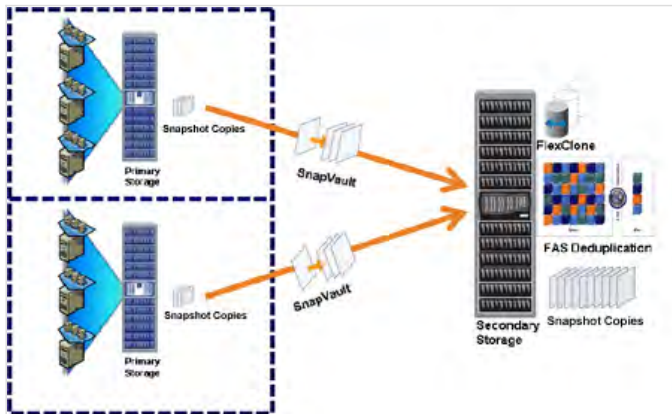


Figure 2) Using SnapVault to provide a centralized backup repository for VMware. Once backups are stored on secondary storage, you can apply additional NetApp technologies such as NetApp deduplication for further data reduction or NetApp FlexClone to create copies for testing and development and so on .

Open Systems SnapVault

Historically, Open Systems SnapVault has been a successful solution for backing up platforms such as Windows®, UNIX®, and Linux®. Its use of block-level incremental transfers—like those described earlier for SnapVault—makes it especially suitable for remote offices where slow network connections typically impede centralized backups, while reduced IT staff makes local backup and data management problematic.

Open Systems SnapVault moves data management tasks from your remote sites to a centrally managed location. Like SnapVault, it uses network bandwidth and secondary disk storage in an extremely efficient manner, making disk-to-disk backup more economical.

With the release of Open Systems SnapVault version 2.6, you now have two options for backing up your virtual infrastructure with Open Systems SnapVault, making it a good option if your VMs are not stored on NetApp primary storage:

- You can install the Open Systems SnapVault agent in each individual VM and back them up to secondary storage just as you would a physical server. The data retains the same format as the original virtual server.

- With Open Systems SnapVault 2.6, the agent software can be installed in the Service Console of the ESX server.

The latter approach makes it possible to back up the individual files that make up each virtual machine. In other words, you can now back up . vmx, . vmdk, . nvram, and . log files, making bare metal and entire VM recovery possible.

Because you only have to install and manage Open Systems SnapVault on each ESX server rather than on each individual VM, the complexity of your backup environment is significantly reduced.

Like SnapMirror for Virtual Infrastructure, Open Systems SnapVault is VMotion aware. You can learn more about using Open Systems SnapVault with VMware in the [OSSV Best Practices Guide for Virtual Infrastructure](#).

Protection Manager

The final NetApp tool I want to talk about is Protection Manager. Protection Manager simplifies the configuration, management, and monitoring of Snapshot copies, disk-based backup, and replication across a NetApp storage infrastructure; and it can also incorporate Windows, UNIX, Linux, and VMware server backup through Open Systems SnapVault. It is the preferred management interface for both SnapVault and Open Systems SnapVault.

Protection Manager provides policy-driven data management that can simplify your backup environment by eliminating many of the repetitive manual processes that are required in typical backup environments. ([A recent article provides more detail.](#))

Conclusion

To achieve optimum data protection from VMware backups, you need to eliminate risk, accelerate performance, and reduce complexity. The first part of this article offers some general guidelines for achieving these goals. I've also described a set of unique NetApp solutions that can have significant benefits in VMware environments.

By implementing SnapManager for Virtual Infrastructure and/or SnapVault, you can do VM-consistent backups faster and more frequently while moving the backup workload off of your VMware servers to significantly reduce business risk. You don't have to install software in each VM you want to back up, and you can do backups at the data store level to further reduce complexity. You also have the option of installing software from the NetApp SnapManager Suite of products to back up applications running in VMs.

Open Systems SnapVault provides similar benefits for non NetApp primary storage environments. Although the backup workload is not completely removed from your VMware servers, Open Systems SnapVault substantially reduces I/O requirements to better match server capabilities, and installation at the ESX server level substantially reduces the complexity of configuring backups that may need to include tens or hundreds of VMs.

Finally, Protection Manager works with both SnapVault and Open Systems SnapVault (plus NetApp SnapMirror software for disaster recovery) to significantly reduce the complexity of managing environments with a large number of backup and/or replication relationships.



Darrin Chapman

Data Protection Subject Matter Expert and Technical Marketing Manager
NetApp

Darrin Chapman is the person you turn to for just about any question involving disaster recovery or backup and recovery at NetApp. He's been involved with almost every NetApp best practices guide about data protection since 2002, and in his spare time he designs training courses for customers and NetApp technical staff.

Originally schooled as an electrical engineer, Darrin's background includes several years in systems architecture for AT&T, Nortel, and EMC .



- [netapp.com](#)
- [Tech OnTap Archive](#)

How Safe Is Deduplication?

By Larry Freeman

Unless you've had your head in the sand recently, you're probably well aware that deduplication is hot. It seems like every storage vendor you've ever heard of (and many more you haven't) is touting deduplication technology as a way to reduce the cost of disk-to-disk backup.

The question you're probably asking yourself is whether deduplication of your data is safe. When the time comes to restore data from that deduplicated backup, will you actually be able to get your data back?

Assessing the relative safety of a deduplication technology really comes down to two basic components:

- The algorithms that are used to identify and eliminate duplication
- The reliability of the underlying hardware and software

In this article, I'm going to take a look at deduplication technology in light of these two criteria. I'll also explain the choices that NetApp has made to enhance the reliability of our own deduplication technology. Because we support deduplication for both primary and secondary storage—whereas most other vendors provide deduplication only for backups—data safety is of the utmost importance for us.

Identifying Duplicate Data

Most existing deduplication products operate at the block level—new blocks are compared against previously stored blocks to determine whether an identical block has been previously stored. If it has been previously stored, the "new" block is discarded in favor of a pointer to the stored block.

So how do you determine whether two blocks are identical? The most common method is that for each block, you compute a "fingerprint," which is a hash of the data contained in the block. If two blocks have the same fingerprint, they are usually assumed to be identical.

However, there is a small but nonzero chance that two nonidentical blocks will yield the same fingerprint or hash value. This is termed a "hash collision" and can result in a unique data block being accidentally deleted.

As you might expect, reducing the probability of a hash collision requires a more complicated algorithm, which typically consumes more CPU resources to compute the hash and yields a larger output value. So there's an obvious trade-off between reliability and speed. Also, longer hashes consume more space for fingerprint storage.

When you are evaluating deduplication technologies, you need to find out how a vendor identifies duplicates and ask about the risk of hash collisions with the chosen algorithm. Many vendors will argue that the chance of a hash collision is lower than the probability of a disk failure or a disk drive error or tape error that corrupts a data block. I don't know if that's a truly comforting thought or not, but I believe that most of us would prefer to minimize as many risks as possible.

Because NetApp supports deduplication for both primary and backup storage, we take a more aggressive approach to preventing hash collisions. We use a fingerprint algorithm like most everyone else, but we use it only to identify potential duplicates. When that happens, we do a byte-by-byte comparison of the two blocks to check that they are identical before discarding any blocks. NetApp technical director Blake Lewis provided a more [detailed explanation of how NetApp identifies duplicate blocks in a previous article](#).

Reliability of Underlying Hardware and Software

Deduplication is only as reliable as the underlying hardware and software. In fact, although it may not be immediately obvious, with deduplication, reliability becomes even more crucial.

For example, suppose that you run a fairly standard backup schedule with nightly incrementals and weekly full backups. Now suppose that you create a file at the beginning of the month and then make no changes to it. With traditional backups, you'll have four copies of the file at the end of the month, one for each week's full backup. If you need to restore the file at that point, the probability is high that you'll be able to restore at least one of the four copies, even if your backup medium is not reliable.

But when you bring deduplication into the picture, at the end of the month you have only one physical copy of the file—the copy created by the first full backup—plus three sets of pointers to the same file blocks. Based on this simple example, it should be clear that you want to be sure your deduplicated backups have been reliably stored on resilient hardware with good RAID protection. Over the course of a year you could have hundreds of backups that actually reference most of the same data blocks.

Explore

Focus on Backup and Recovery

This special issue of Tech OnTap drills down on critical backup and recovery topics:

- [Choosing the Right Backup Technology](#)
- [Top Three VMware Backup Challenges](#)
- [How Safe Is Deduplication?](#)

Tough Questions, Real Answers

What kind of real-world questions are people asking about deduplication? View a "no holds barred" transcript from a recent live chat.

➤ [More](#)

Manage Data Growth with Deduplication

A recent TechTalk Webcast zeroed in on some of the biggest problems in backup and recovery today, including:

- How to cut costs with deduplication
- Low bandwidth WAN backup
- Making the move from tape to disk-based backup

Watch the Webcast to learn more about these important topics.

➤ [More](#)

Questions About Dedupe?

Want to learn more about NetApp deduplication? Larry Freeman hosts a forum on the topic on the NetApp communities site. You can read past postings, share your experiences, ask questions, and get feedback from other users.

➤ [More](#)

ESG Report on NetApp Deduplication

In a recent independent study, ESG evaluated the effectiveness of NetApp deduplication in terms of both capacity savings and performance impact.

Environments evaluated included:

- Virtual server storage
- Backup
- Archiving

Read the full report to find out the results.

➤ [More](#)

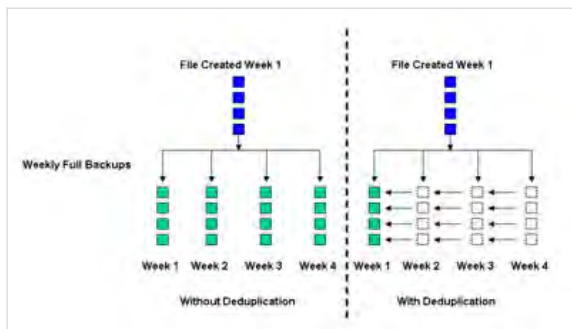


Figure 1) Four weeks of full backups of the same unchanged file. Without deduplication you have four discrete copies of the same file. With deduplication you have a single copy. This yields substantial space savings, but at the same time it makes reliable storage of that single copy more critical.

There are a wide variety of deduplication products out there. Some are software only and may use a variety of underlying hardware; some include both hardware and software (possibly obtained from a variety of sources through licensing or OEM arrangements). Before making a decision, you should assess how mature the software is, how robust the underlying hardware is, and how well the two integrate together.

NetApp Reliability

With NetApp® storage, deduplication is an integral part of the Data ONTAP® operating environment that runs across our entire product line. Data ONTAP has been under continuous co-development with NetApp hardware platforms for more than 15 years. Unique features of the NetApp WAFL® technology actually simplify the implementation of deduplication and make it possible to deduplicate any stored data, not just backup data.

Proven reliability features in NetApp hardware and software result in data availability of more than 99.999% as measured across the NetApp installed base. [A recent analyst report](#) describes the NetApp methodology and many of the features that contribute to NetApp's reliability.

One example of our attention to detail involves the well-known fact that disk drive bit errors can develop over time—or even during the manufacture of disk drives. Every drive has built-in error correction that detects and usually corrects such bit errors. If a string of errors is too great to be handled by ECC, the drive reports back that the sector is unreadable, at which point RAID algorithms fix the error from the information stored on other sectors. NetApp, however, also uses a checksum scheme for further protection—we use an additional portion of the drive as overhead to store checksums that move with the data through the system to check that what was written is returned perfectly during data restoration. In essence, we provide a third level of protection.

To protect the reliability of data committed to disk, NetApp also developed RAID-DP™, a high-performance, dual-parity RAID 6 implementation that protects against double disk failures without sacrificing write performance. You can [read more about RAID-DP and other NetApp enhancements to protect against misbehaving disk drives in a previous Tech OnTap article](#).

Conclusion

To protect your backup data, deduplication technology must use appropriate algorithms to avoid discarding unique data blocks and also provide the fundamental hardware and software reliability necessary to safely store deduplicated data for later recovery.

Because NetApp deduplication technology is used for primary data stores as well as for backup data, we take extra care to protect data reliability. NetApp deduplication uses a combination of fingerprints plus byte-by-byte block comparisons so that unique data blocks are never erroneously deleted due to hash collisions. Deduplicated data is stored on NetApp storage systems using hardware and operating software that have been proven reliable and resilient through years of field deployment, so you can be confident that when it comes time to recover data, you'll get back the data you backed up.



Larry Freeman
Sr. Marketing Manager for Storage Efficiency
NetApp

During his 25-plus year career in data storage, Larry has held various positions with companies including Data General, Telex Computer Products, NEC Information Systems, and Spectra Logic.

A frequent speaker and author, Larry's current role at NetApp is evangelizing data storage efficiency technologies, including deduplication. Larry is the founder and co-chair of the SNIA Data Deduplication and Space Reduction Special Interest Group and active in the SNIA Green Storage Initiative.



- [netapp.com](#)
- [Tech OnTap Archive](#)

Choosing the Right Backup Technology

By David Chapra

When people think of NetApp, they usually think “primary storage.” While that’s true, NetApp also has a wealth of experience in designing and implementing backup and recovery solutions worldwide, from the largest data centers to small remote offices with limited IT staff. Over 60% of Fortune 100 companies use NetApp for long-term data protection; we have thousands of backup and recovery installations. NetApp has vast experience in helping IT teams choose and implement solutions for difficult data protection problems.

This article examines:

- Today’s backup and recovery challenges
- Different backup technologies to address these problems
- When to choose specific solutions

Key to this discussion is balancing cost constraints against the never-ending need for better performance while leveraging new and emerging technologies. We’ll start by discussing the sources of backup and recovery problems, talk about general classes of solutions, describe selection criteria, and then provide some information on specific NetApp® backup offerings, including NetApp virtual tape library (VTL) solutions and NetApp SnapVault®.

	Pros	Cons
Traditional tape	Familiar operation and technology	Growing complexity, poor reliability
Transitional disk	Improved reliability and recoverability	Difficult to provision, expensive
Virtual tape	Simple to deploy and scale, disk turned for high backup performance	Relies on existing tape processes
Replication-based backup	More frequent backups, long distance replication, storage efficient	Different methodology for backup

Figure 1) Pros and cons of various backup and recovery alternatives.

Traditional Backup and Recovery Challenges

Most organizations, regardless of size, face a set of common challenges:

- Inability to get nightly backups done within available backup windows
- Difficulty recovering data quickly enough
- Reliability issues that cause restores to fail

At branch and remote offices, these issues are compounded by additional challenges:

- Backups often fail to complete
- No reliable off-site disaster recovery
- Staff too busy to handle the complexity of tape storage or backup applications
- Lack of standardization and limited visibility into operations

Tape technology is a significant contributor to long backup windows, long recovery times, and reliability issues that can plague traditional backup environments:

- It can take up to 3 minutes to simply load a tape and position the tape heads to begin reading or writing. And that’s assuming the right tapes are in the library.
- If the backup or restore spans multiple tapes, similar delays are incurred each time a tape is changed.
- If you can’t stream data fast enough to or from the tape drive, it can slow down considerably, leading to “shoe-shining” as the tape drive repeatedly repositions. Most modern tape drives have stepping algorithms to limit this, but there’s still a performance impact that can lengthen your backup window or slow down recovery.

When it comes to recovery, tape reliability often becomes a significant issue that can add delays and uncertainty. Tape is a passive medium. Once you write to a tape, it just sits in a slot in your library until it’s needed again, with no active monitoring to make sure that it’s okay. If there’s an issue with a tape, you don’t detect it until you use it again, and if the next use is a restore, it can significantly extend the time needed to do the restore—if it can be completed at all.

Explore

Focus on Backup and Recovery

This special issue of Tech OnTap drills down on critical backup and recovery topics:

- [Choosing the Right Backup Technology](#)
- [Top Three VMware Backup Challenges](#)
- [How Safe Is Deduplication?](#)

A Quick Primer on NetApp Data Protection Software

NetApp offers two alternatives for data protection software: SnapMirror® and SnapVault software.

SnapMirror is a replication technology intended for disaster recovery. The mirror is an exact replica of data on the primary storage and allows rapid failover to the most recent system state.

SnapVault, in contrast, is intended for backup and recovery. SnapVault generates backup images to allow recovery back to one of many point-in-time images and enables a much longer retention period than Snapshot copies alone.

Get the details. Read the reports:

- [SnapMirror Best Practices Guide](#) (PDF)
- [SnapVault Best Practices Guide](#) (PDF)
- [Data Protection for NetApp Storage Systems](#) (PDF)

Technical Overview of NetApp Protection Manager

Protection Manager is an intuitive, policy-based management application for NetApp backup and replication technologies including SnapMirror, SnapVault, and Open Systems SnapVault. This tool enables administrators to define and apply policies to their data, eliminating ambiguity and the potential for error inherent in manual management.

A simple dashboard shows comprehensive data protection information at a glance, including unprotected data, longest lag times, and so on. The application automatically correlates the data sets and underlying physical storage resources, so you don’t need to think about storage infrastructure.



A 25-page report explains the details.

[More](#)

Is Disk an Economical Answer?

It probably won't come as a surprise to you that adding disk backup to your overall data protection strategy may solve a lot of these problems. Disk solutions increase performance and avoid the load/unload/seek delays that are inherent with tape as well as the throughput penalty for "shoe-shining". [Learn more about assessing the needs of your specific backup environment](#).

For recovery, disk solutions largely avoid the reliability problems that can occur with passive tape media. Disks are continuously monitored through the use of technologies such as regular media scans. If a problem does occur, you are much more likely to find it with disk—before you need to restore that critical file or server; and RAID protection allows you to recover from media errors even when they occur during restore.

Naturally, this is where performance versus cost enters the picture. It's true that a TB of tape is still less expensive than a TB of disk. However, there are ways to mitigate disk costs. Many solutions include compression technologies, and new deduplication technologies can significantly change the cost dynamics of disk versus tape backup.

Compression offers data reduction similar to that found on tape drives, reducing overall capacity requirements. Deduplication uses more advanced algorithms to further increase data reduction. Typical backups have an inherently high level of duplication. Every time you do a full backup, you back up mostly the same files as previous backups, most of which have changed little if at all. Even incremental backups may include many files that have been changed only slightly since the last time they were backed up. Deduplication can yield up to a 20:1 (95%) space savings, significantly reducing the amount of disk you need versus tape, as illustrated in Figure 2. (A companion article in this issue discusses [the reliability of deduplication in backup environments](#) .

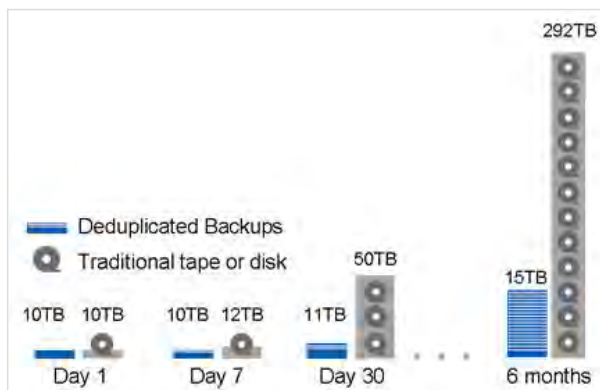


Figure 2) Deduplicated backups versus traditional tape or disk backups for a typical environment.

Of course, tape may still have a valuable role to play in your data protection strategy; the scope is just narrower than it once was. When you consider that most restores occur within the first week after backup, you can get a lot of leverage from a small amount of disk storage while continuing to use tape for less critical data and offsite storage.

Disk-Based Solution Options

When it comes to disk-based backup hardware, you have two options:

- Virtual Tape Libraries (VTL), which use disk arrays, but emulate the characteristics of tape libraries
- Traditional disk arrays

VTL Advantages

VTLs offer the following advantages; they:

- Fit into your existing backup infrastructure without requiring major changes, making them simple to deploy
- Eliminate the reliability issues associated with physical tapes and libraries
- Are optimized for maximum backup performance
- Offer a simple path to physical tape when it's necessary to send data offsite

Traditional Disk Advantages

Traditional disk solutions (SAN solutions using LUNs, or NAS) also offer increased reliability and performance, but without the specializations of VTL. Potential advantages of traditional disk include:

- Standard disk interfaces that make it easy to manage data using familiar methods and tools
- Ability to be used for both general disk storage and backup functions (depending on performance demands)
- Enables use of replication-based technologies for disaster recovery and backup/recovery.

Replication-Based Backup and Recovery

The use of replication technologies is also possible with VTL, but it is currently much less frequent due to the need to keep the backup application catalog up to date with the exact position of backup

images and all replicated copies. Expect this to be a key area of advancement over the next few years as VTL vendors (and storage vendors in general) increase integration with backup applications to provide additional value-added services (SnapVault for NetBackup™ is one example).

For disk array products, the most common use of replication technology is for disaster recovery. Replication can occur at the array or operating system level but the end result is a synchronous or near-synchronous copy of the data set allowing recovery from system or site-level failures. Although this is not truly backup and recovery (which provides multiple point-in-time images for recovery) this type of replication opens the door for unique opportunities.

The integration of Snapshot™ technology with replication makes it possible to provide high-speed, low-overhead backup and recovery. Implementations vary, but the goal is to replicate changed data from a source system, either at the array or server level, to a target system and then generate a backup image via a Snapshot copy. This provides three key advantages:

- Increased backup and restore performance because less data is moved
- Reduced capacity utilization because only changed blocks, not whole files, are stored as part of each backup
- Efficient WAN backup because only a limited amount of data must be transferred

The end result is a high-performance, low-bandwidth, space-efficient backup and recovery strategy. Depending on the vendor, tape integration may also be available.

NetApp Solutions

NetApp offers a full range of disk-based backup solutions, including two hardware options: VTL and FAS storage systems; and our replication-based backup solution, NetApp SnapVault.

NetApp NearStore VTL

The NearStore VTL product line includes three models. Our flagship model—the NearStore® VTL1400—offers raw capacity up to 672TB (1,100TB compressed) and sustained write throughput up to 2.5GB/sec. All models feature hardware compression, self-tuning for optimal performance, smart sizing for efficient use of physical tapes, and shadow tapes for restores even when physical tapes are stored offsite. Entry-level systems can be upgraded in place to higher models.

NetApp FAS Storage

NetApp FAS storage systems support both high-performance Fibre Channel and economical SATA disks in a single array, allowing you to mix both primary and secondary storage in a single system, or to create SATA-only systems exclusively to support backup. With FAS storage systems, you can immediately leverage all the proven capabilities of the NetApp Data ONTAP® operating environment for protecting and managing your backups, including NetApp RAID-DP®, Snapshot, deduplication, and FlexShare™ for quality of service.

SnapVault

The NetApp SnapVault family is a replication-based backup and recovery solution designed to enhance and simplify data protection, especially at remote offices. SnapVault leverages NetApp Snapshot technology to back up and restore systems at the block level. SnapVault identifies and copies only the changed blocks on a system (not changed files) to secondary storage. This increases performance by limiting the amount of data transferred during backup and restore operations, as well as limiting the capacity needed to store backups, allowing you to perform backups more frequently if desired. On average, customers report a 95% reduction in both backup and restore time. This methodology also gives customers the option of backing up directly from a remote site to a core data center, because only a limited amount of data is transmitted.

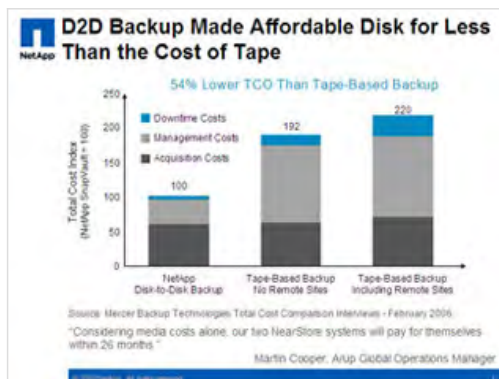


Figure 3) SnapVault disk-based backup versus tape.

For consistent backup and restore of important databases and applications such as Oracle®, Microsoft® Exchange, SQL Server™, and others, SnapVault can be used with the NetApp SnapManager® suite of products.

For non-NetApp storage, Open Systems SnapVault allows you to incorporate standalone servers and diverse storage systems while offering many of the unique advantages of SnapVault. A companion article in this issue discusses the use of [Open Systems SnapVault specifically for protecting VMware® environments](#).

In case you missed previous SnapVault coverage in Tech OnTap, the following recent articles highlight case studies of customer deployments:

- An [Experts OnTap feature](#) details how one IT team evaluated its backup options, including a detailed Q&A with NetApp experts about SnapVault.
- A [customer-written case study](#) describes Agilent Technologies' experience in moving to disk-based backup with SnapVault.

Making the Right Choice

Choosing among the available backup technologies can seem daunting, but there are actually a few simple guidelines that will help guide you to the right solution. At the highest level, it's probably easiest to consider choices in terms of your target environment—data center or remote office.

Data centers typically have a dedicated IT staff with knowledge of backup and recovery. Some level of tape has undoubtedly been in use, and administrators have designed best practices around its usage. In addition, there is probably a mandate for offsite tape storage to meet compliance guidelines. Most data centers are driven by demands from the business units, which require:

- Integration into best practices and an easy way to migrate data to tape
- Scalable performance to meet restore times and backup windows
- Guaranteed reliability for completing backups and restores

To meet these needs, a VTL is the best option. Tuned for backup and providing a traditional tape interface, VTL offers performance, scalability, and integration into existing best practices. Whether through backup application manipulation or value-added technology, VTLs also enable a simple path to tape for long-term offsite retention.

You may also want to evaluate traditional disk solutions, depending on what value-added services they provide. Offerings such as deduplication and backup-application-specific data reduction can be extremely compelling for cost-sensitive environments, but be sure to understand their performance capabilities. If the VTL tuning and interface are not as important, this may be the appropriate path. But again, be sure to understand all the performance, cost, and scaling capabilities of any solution.

At the remote office, the situation is slightly different. There is often limited or no staffing for data protection, and the key goal is to eliminate as much complexity as possible so that backups simply *happen*. First and foremost, this means removing or significantly limiting tape to reduce configuration and management overhead, as well as the complexity and cost of media management. You should look at disk solutions not only for simplicity but also for cost-saving features such as deduplication.

Additionally, offices may want to look at augmenting disk backup with replication-based data protection to limit reliance on backup applications. Although backup applications offer deep and rich feature sets, they require a high level of knowledge and deliver more functionality than is really needed—especially if actions like tape offsite and cataloging can be consolidated to the data center. By stripping out the backup application from the remote office and pushing long-term retention tasks to the data center, you can significantly reduce data protection complexity and the chances of data loss as well as violation of regulations. In addition, replication-based data protection can further reduce capacity requirements and speed up backups through the use of block-level incrementals, where only changed blocks are moved and stored. For the most efficient solution, you should consider replication-based data protection that backs up data from remote to core.

If you've been hesitant to consider disk-based backup because you thought it was too expensive, it may be time to look again. The latest deduplication technologies can significantly reduce the cost of disk versus tape, and the reduction in complexity can reduce management costs as well. NetApp offers a full range of hardware and software tailored to help you solve virtually any backup problem. If you never thought of NetApp as a backup vendor before—think again!



David Chapa
Director, Backup and Recovery Solutions
NetApp

David has over 20 years of industry experience focusing specifically on data availability, data disaster recovery, and business resumption practices. He is co-author of *Implementing Backup and Recovery: The Readiness Guide for the Enterprise* and is recognized as an authority on backup and recovery, disaster recovery, and business resumption practices.
