

► Tech OnTap Home

February 2006

HIGHLIGHTS

- **Architecting Storage for Resiliency**
- **Early Review: NearStore® VTL**
- **D2D Interactive Online Event**
- **Tips for Data Migration**
- **Data ONTAP® Simulator**

NetApp's Strategy for Improving Your Backups and Recoveries

Q&A with Manish Goel, VP and GM, NetApp

How you can leverage disk-to-disk backup in your existing infrastructure plus a sneak peek at the NetApp engineering roadmap.



More ►

BLOGGING WITH DAVE

Dave Hitz, NetApp Founder and EVP

"Since value-add drives are so profitable, drive vendors hate it when we figure out how to use commodity drives in high-end storage systems."

► **Dave's Blog**

DRILL DOWN

- **D2D Interactive Online Event**
Presentations, special offers, and online chat transcripts.
- **SnapMirror® Best Practices Guide**
A 66-page planning/deployment guide.
- **iSCSI Multipathing Possibilities**
Evaluate pros and cons of each option.



ADMIN RESOURCES

- **Data ONTAP Simulator**
Test functionality, export NFS/CIFS shares, and even simulate clustering without purchasing new hardware.

TIPS FROM THE TRENCHES

Tips for a Successful Data Migration

Will Titherington, Professional Services Consultant, NetApp Global Services

Field-proven advice on how to choose your tools, develop a test plan, and maintain realistic timelines. Includes a detailed anatomy of a six-month migration project.

More ►

Early Customer Review: NetApp NearStore VTL

VTL600 Beta User

"We saw a 4-5x performance improvement ... I flew to our DR facility for implementation and testing. I didn't need to fly back or call anyone on-site for the rest of the trial."

More ►



ENGINEERING TALK

The Private Lives of Disk Drives NetApp Protects Against 5 Dirty Secrets

Rajesh Sundaram, Storage Resiliency Architect, NetApp

A former WAFL® engineer outlines how NetApp addresses universal disk weaknesses with unique resiliency features such as RAID-DP™, SyncMirror®, unresponsive drive protection, and lost write protection.

More ►



Manish Goel

Vice President and General Manager, NetApp Secondary Storage Business Unit, NetApp

Manish Goel drives NetApp strategy for data protection and retention solutions. Evaluating and tracking technology trends, market offerings, and customer needs, Manish and his team guide the NetApp product roadmap for data backup, archiving, security, and disaster recovery products. Manish has a background in electrical engineering plus more than 20 years of technical and operations experience.

NetApp's Strategy for Improving Your Backups and Recoveries

Q&A with Manish Goel

Q: NetApp recently announced that it is extending the benefits of disk-to-disk backup to multi-vendor storage environments. Could you tell us a little more about what this means?

NetApp was one of the first storage vendors to offer disk-to-disk (D2D) backup solutions and is currently a leader in the D2D backup space. Over 1,000 NetApp customers have adopted our disk-to-disk backup offerings to support NetApp storage environments and/or centralize backup for local storage at remote offices.

However, today only 8% to 9% of the world's data sits on NetApp storage. Given our confidence in NetApp capabilities and solutions and the fact that backup is a high-priority pain point, we think it is a natural next step to extend the power of our solutions to all storage environments.

Launching the NetApp NearStore® VTL solution and expanding our relationship with Symantec has enabled us to round out our portfolio and extend disk-to-disk backup technology to any type of environment. Now, we can not only help customers improve backup and recovery in NetApp and remote office storage environments, but also help them implement effective solutions for protecting data sitting on EMC, Hitachi, HP, IBM, or any other storage platform.

Q: Why should companies, especially those not currently using NetApp storage, look to NetApp to help them improve backup and recovery?

There are three reasons you should seriously consider NetApp disk-to-disk backup solutions:

- **First, NetApp offers the most comprehensive disk-to-disk backup solutions portfolio in the industry.** Most other storage vendors have either been slow to embrace D2D backup or have embraced it with limited point solutions. NetApp provides a range of options using a common hardware platform. This gives you the ability to redeploy existing NetApp equipment with new software as your backup and recovery challenges change over time.
- **In addition, NetApp enables you to enjoy the benefits of disk-to-disk backup without major infrastructure changes.** If you use NetApp storage today, we have excellent solutions to help you back up and protect your data. If you're not using NetApp storage, you can take advantage of technologies like NearStore VTL and Open Systems SnapVault® to complement your existing environment. NetApp also enables you to leverage familiar management tools from industry-leading backup vendors including Symantec, Tivoli, CommVault, Syncsort, and Bakbone.
- **Finally, disk-to-disk backup is a strategic priority and major area of technology development investment for NetApp.** Most of our competitors are very tactical in their approach to D2D. The EMC CLARiiON Disk Library (CDL), for example, is an OEM solution from a third party. It's not owned by EMC, and it doesn't get EMC engineering mindshare. Not only does NetApp have the most innovative data protection and retention solutions in the marketplace; we're also committed to maintaining our leadership position going forward.

RELATED INFORMATION

- [NetApp D2D Online Event](#)
- [Disk-to-Disk Backup Demo](#)
- [NearStore VTL: Early Review](#)
- [Data Protection Strategies for NetApp Storage](#)
- [Architecting Storage for Resiliency](#)
- [Is NetApp Storage Simpler?](#)

Disk-to-Disk Backup: The Entire Family

NetApp's comprehensive family of products can help you address critical backup issues in any data center or remote office environment.



See presentations from NetApp, Symantec, and Decru experts and executives at the [D2D online event](#).

Disk-to-Disk Backup

D2D backup has emerged to provide significant advantages over the tape-based systems traditionally used for backup and recovery. This began with three recent developments:

1. Less expensive SATA drives began to be used with RAID redundancy, hot-swappable drives, and other high-reliability features, making them a technically and economically feasible backup media.
2. Backup requirements began to exceed the capability of tape—that is, jobs could no longer be completed in available backup windows because of exponential data growth and 24x7 working environments.
3. Data recovery requirements have become increasingly stringent.

Q: The "most innovative data protection and retention solutions in the marketplace" sounds great, but what does that mean?

The biggest pain point for most customers is that as their data repositories have grown, they are over-running their backup windows, and can no longer back up all the data that they need to, in the increasingly 24x7 environments. For data stored on NetApp storage, we help customers eliminate backup windows by frequently transferring only the changed data to their D2D backup device.

The goal of doing backups is to be able to restore user data when needed. In the NetApp approach, by storing backup copies in native file formats, restores are near instantaneous.

Q: Could you provide a few specific examples?

Let's start with remote backups. Most companies have some type of established process for backing up data in their data centers. It may not be perfect, but for the most part it works. Outside the data center, however, the situation is often very different. Remote offices have an increasing amount of business-critical data, but most customers still rely on tape backup and have limited or even nonexistent IT resources at the remote office.

In 2003, NetApp launched an open systems version of SnapVault that enables you to back up data from local storage to a centralized location. What makes this solution unique is that we leverage NetApp Snapshot™ technology to create an incremental copy of only the data that has changed from the last copy and can very quickly send this data copy across extremely limited bandwidth connections. We have customers with oil rigs so remote they have to communicate via low-bandwidth satellite connections. These customers depend on Open Systems SnapVault to send backups to their main data center.

Some backup software vendors are beginning to focus on this space, but they cannot bring the same level of combined network and media efficiency that NetApp delivers.

Do you realize that no other VTL solution on the market can effectively use the compression available on a tape drive and still fully utilize a physical tape?

Another area of innovation involves the newly announced NearStore VTL solution. Do you realize that no other VTL solution on the market can effectively use the compression available on a tape drive and still fully utilize a physical tape?

NearStore VTL uses something we call tape smart sizing to take into account tape drive compression when sizing a virtual tape. The NearStore VTL samples data for compressibility as it is backed up to the VTL and adjusts the size of each virtual tape to deliver optimal utilization of the corresponding physical tape. The result is a two-to-one savings relative to other VTL solutions in the number of physical tapes that must be purchased and managed.

We also offer self-tuning performance to automatically balance backup streams across all the available disks in the VTL, which maximizes throughput and eliminates the need for ongoing manual tuning.

Q: Where does NearStore VTL fit into the NetApp D2D product offering?

We are continuing to build a solution portfolio that meets the requirements of any customer in any environment. We recognize, for example, that many companies want to take advantage of D2D, but do not want to disrupt their operating environment in any way, shape, or form. The best solution if you're in this position is to start with a VTL implementation.

NearStore VTL enables customers to immediately benefit from D2D technology without requiring changes to existing software, process, or infrastructure. In addition, NearStore VTL is part of a broader solutions portfolio and runs on exactly the same hardware platforms as all of our other disk-to-disk backup solutions. This means that if your requirements change over time, you can simply install new software and repurpose the original hardware for other types of NetApp solutions. This type of investment protection simply isn't an option with an OEM VTL solution.

Q: How does integration with NetBackup fit into the NetApp D2D portfolio?

Companies today need to restore critical data instantaneously, not two or three days out.

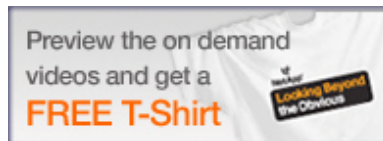
D2D is now an affordable option that helps minimize or eliminate the lengthy backup/restore problems of tape systems.

Learn more about [D2D backup](#).

NetApp Is a Leader in D2D Backup

- NetApp was one of the first storage vendors to offer D2D backup (including the first nearline storage system)
- NetApp offers the most comprehensive D2D offering available
- 1,000+ customers use NetApp storage for disk-to-disk backup
- 100,000TB+ of SATA disks shipped by NetApp as of December 2005

This integration is the result of roughly 18 to 20 months of joint engineering development between NetApp and Symantec. We now offer a joint solution that lets customers get all the benefits of our SnapVault technologies combined with the benefits of Symantec NetBackup management for both NetApp storage and storage from other companies. Customers will see significant media savings achieved through data deduplication and will be able to access backup copies in native file format, so end users can actually restore their own files. And they won't have to deal with any operational disruption in their environment—it will continue to work exactly as it does today.



NetApp also has very strong working relationships with other major backup software vendors such as Syncsort, CommVault, Tivoli, CA, and BakBone.

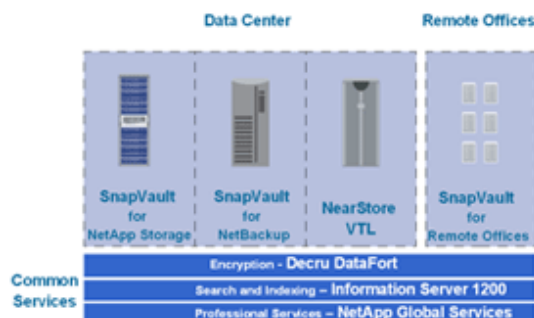
Q: What are some other topics that are top of mind for your organization?

You'll be hearing a lot from NetApp in the next few months about security, search, and services.

There is a big requirement for encryption, whether it's done as the backup data is being created or as it's offloaded onto a tape. Our acquisition of Decru enables us to combine Decru® encryption technology with our VTL and other disk-to-disk backup solutions.

One key benefit of the new NetApp Information Server 1200 technology is the ability to search and index content within files of unstructured data across an enterprise. These capabilities are helpful, of course, for assessing and classifying data for rapid retrieval of backed up and archived data and for data migration projects.

NetApp Disk-to-Disk Backup and Recovery Solutions



Finally, NetApp Global Services can help you implement and tune our backup solutions to meet your recovery point objectives (RPOs) and recovery time objectives (RTOs). This is a set of offerings that we have developed and refined over the years. We've codified a lot of the best practices in implementation and installation so that they can be systematically applied to any new infrastructure deployment. We can consistently apply the same best practices over and over again, as opposed to having customers have to discover them on their own with each new implementation.

Q: You mentioned earlier that disk-to-disk backup is a major area of investment for NetApp. How is this impacting NetApp's product development strategy?

On the media side, our focus is on driving down costs. Disk-to-disk solutions are already more economical than tape backup, even though the media costs are higher. This is because the amount of data you need to store using disk-to-disk backup and NetApp Snapshot technology is significantly lower relative to tape environments. Each time you create a backup copy you're only storing the changed blocks, as opposed to having to make another copy of the entire data set every single time.

Still, tape is a very cheap medium. For disks to truly become the dominant media, disk pricing will have to be at par with or better than tape-based systems. This means NetApp Engineering is continuing to focus on things such as compression technologies, eliminating duplication, and media savings.

On the backup capabilities side, the goal is to be able to do backups in such a way that backup windows are eliminated, allowing customers to operate their online environments in a 24x7 mode. Intrusive backups are a very 20th-century infrastructure paradigm that will not scale into the 21st century.

NetApp technology development is focused on getting around the requirements of having backup windows and being able to provide seamless restores for our all environments. You can expect to see increased application integration and increased integration in the environments where customers want to use disk-to-disk backup.



Will Titherington

Professional Services Consultant, NetApp Global Services

Will Titherington is a NetApp Professional Services consultant and "resident expert" on data migration. In the past two years, Will has worked with over 50 customers to successfully migrate over 160TB of data. Prior to joining NetApp, Will spent five years as an administrator for a NetApp customer and was responsible for migrating 17TB of Oracle® data ... every week!

Tips for a Successful Migration Project

by Will Titherington

A successful migration project is not just about moving data from one location to another, but about making sure that everything goes as expected. This can be challenging for even the most sophisticated IT pro. In fact, an independent study found that more than 80% of data migration projects exceed their timelines, go over budget, or fail to meet their goals.

Through long experience and thousands of migration projects, NetApp Global Services (NGS) has established a variety of standard methodologies to ensure projects involving data migration are completed on budget and on time. (For a complete description, see the [NGS Data Migration Methodology](#).)

Rather than rehash information that is available elsewhere, this article uses the example of a large-scale migration project in which the customer brought in NGS to help migrate 57TB of engineering data as part of a massive storage technology refresh. To complicate matters, the customer allowed planned downtime only one weekend a year, and the data migration was just one element of a move to a new data center. (See sidebar for details.)

The result? The six-month project was completed on time and on budget and has been deemed a complete success by the customer.

There were many factors that contributed to the success of this effort, including the customer's decision to involve NGS in the initial planning phases of the project. In this article, I'll share three tips to help ensure that your next migration project runs smoothly:

- Make sure you choose the right tools for the job
- Test, test, test
- Be prepared to reevaluate timelines as new information becomes available

Tip 1: Choose the Right Tool for the Job

Automated tools can make a huge difference in the amount of time, resources, and pain associated with a migration, but nothing takes the place of careful, upfront planning and a thorough discovery process. That said, you should choose your tools carefully and always consider capabilities versus costs.

There are two main migration phases in which tools can be useful:

- **Discovery.** This process can be manual (using command-line options to gather information to complete a checklist) or automated (via a discovery tool). A standard NGS best practice is to use automated discovery tools whenever possible because they help minimize the chance of error.
- **Migration.** Information collected during discovery helps guide the selection of tools to physically move data from the source (existing location) to the target (new location).

Unfortunately, no single tool (despite any marketing claims to the contrary) does everything well. Although you'll find that some commercial tools have both discovery and migration capabilities, in our experience, it is extremely unlikely that any one tool will meet all your needs for either task (let alone both!). To complete a complex migration, you will almost

RELATED INFORMATION

- [NGS Data Migration Methodology](#)
- [Simplifying Exchange Migration](#)
- [Novell Netware Migration to Windows Server 2003](#)
- [Architecting Storage for Resiliency](#)
- [Is NetApp Storage Simpler?](#)
- [NOW Customer Site](#)
- [\(password required\)](#)

Anatomy of a Complex Data Migration

Company: A large engineering software and services provider

Data set: 57TB of engineering data (NFS with a small amount of CIFS)

Migration goal: Consolidate storage from 25 storage systems down to 10 and refresh storage technology

Special considerations/problems:

1. Only one weekend/year of planned downtime (Friday 6:00 p.m. to Sunday 6:00 p.m.).
2. In addition to the migration, the company was completing a data center move and firmware upgrade (effectively rebooting every server in the company) that same weekend.
3. Corporate networks were at 97% utilization, so the IT team had to install separate network infrastructure to carry out the migration.
4. Company had 811 separate mountpoints that had to be managed.

Discovery: Performed manually using spreadsheets/checklists to gather necessary data.

Migration: Performed using NetApp NDMPcopy. Scripting capability made it easy to automate process for the large number of mountpoints.

Pre-migration Timeline

- **N—six months:** Begin data collection.
- **N—four months:** Provision and test new storage, validate tools and methods.

certainly use a variety of "tools," including checklists, spreadsheets, operating system commands, tools provided as part of your business applications, and storage system utilities, in addition to—or instead of—commercial tools.

You should also be aware that new commercial migration tools are emerging on a regular basis, and old ones are being enhanced. When NGS first started planning the migration project mentioned above, for example, we had to perform all discovery manually because none of the tools available at the time met our requirements.

If we were starting this project today, there are several automated tools which do a particularly good job finding and classifying data. I might also use ISI Snapshot to inventory the hosts and hardware, which would have saved two to three days of effort spent on manual data gathering.

Because the migration required organizing data below the qtree level, we relied on NetApp NDMPcopy as the primary migration tool. Typically we use SnapMirror® for migration between NetApp systems because it allows very frequent mirroring, has a minimal impact on the infrastructure, and can be throttled to overcome bandwidth or CPU limitations. However, SnapMirror isn't an option below the qtree level. An added advantage of NDMPcopy is that it allows scripting, which saved time managing the customer's 811 mountpoints.

If you are moving more diverse data types, be prepared to use different tools to handle the migration of the different types (for example, NFS/UNIX®, CIFS/Windows®, databases and business applications, and block-oriented SAN data).

In addition, your choice of tools will be driven by your goals and requirements. For example, if you have a situation in which you have minimal or no allowable downtime to cut over to the new storage, NeoPath File Director can be helpful. For migrations from UNIX to NetApp, NGS consultants often use simple, host-based tools such as rsync or rdist. If part of a CIFS deployment includes implementing a global namespace, you may be able to use VFM™ for the implementation and then seamlessly do the migration under the covers.

Path lengths may also be a consideration in migration tool selection. Many Windows tools (for example, Explorer, CACLS) have a path length limitation of 256 characters. Since paths on NetApp systems can easily exceed that, if you want to use a Windows tool, you might look at the Windows 2003 version of Robocopy (part of the [Windows Resource Kit](#)) which does not suffer from the 256 character path limitation.

Depending on your situation, this is probably either more than you want to know or not nearly enough. Because of the plethora of tools available and trade-offs associated with each, if you are planning a complex migration that involves NetApp storage, I recommend having NGS help you identify the best tools for your situation.

Tip 2: Test, Test, Test

Even if you've carefully chosen your toolset, nothing takes the place of testing to increase confidence and ensure a smooth migration. Often when something goes wrong during a migration, it could have been identified with upfront testing.

The standard NGS testing methodology includes:

- **Verifying the infrastructure.** You don't want to hunt down and fix network problems, such as duplex and jumbo frame mismatches or IP conflicts, after the migration begins. Test your infrastructure before you do any other testing. Make sure you are able to achieve the read and write performance you would expect.
- **Testing the migration tool(s) with real data.** Based on discovery, NGS typically identifies a representative sample of the data and tests the migration tool(s) on it to ensure that everything works as expected. In addition, the test will provide a good estimate of how long the migration will take. Test data selection criteria include data type (NFS, CIFS, block, and so on), complexity of permissions and ACLs, depth of directory structure, and so on.
- **Executing a complete dry run up to the cutover point.** A dry run is the best way to identify all potential problems and find out exactly how long a data migration will take. This is highly recommended but optional; a dry run is not always feasible due to a lack of timing or resources.
- **Validating the completed migration.** Before the migration project can be considered a success, someone must validate the post-migration environment and

- **N—three months:** Dry run. Complete data transfer took 13 days. (This time frame allowed for test to be rerun if needed.)
- **N—13 days:** Data transfer initiated.

Cutover Weekend

- The data transfer initiated 13 days ago was completed within 15 minutes of predicted time.
- **Saturday 7:00 a.m.:** Shut down storage systems and hosts for move to new data center.
- **Saturday, 2:00 p.m.:** Hardware reaches new data center, and reinstallation commences. QA begins.
- **Saturday, 10:00 p.m.:** Validation testing commences.
- **Sunday, 2:00 a.m.:** Validation testing complete.
- **Sunday, 2:00 p.m.:** All work requiring downtime completed with four hours to spare.

Results:

- Not a minute of unplanned downtime
- No impact on end users
- Only one issue was identified (this was later found to be an accidental file deletion not associated with the migration)

Performance Testing Tip

In some situations it is helpful to do a test migration and see if performance fits expectations.

For example, if you use Robocopy (part of the [Windows Resource Kit](#)) with the default settings, open files can cause a lot of delay. You can reduce migration time by changing settings (/R:2 [retries to two] and /W:3 [wait three seconds between retries]).

Source: Andrew Bond, NGS Professional Services Consultant in the UK

confirm that all expectations have been met. At a minimum, network access, file permissions, directory structure, and database/applications need to be tested. Only the owner of the data can truly validate that the migrated data is correct and usable, so ideally, all owners should participate in the validation.

In the project mentioned above, data collection began a full six months prior to the scheduled cutover date. The new storage systems were then installed, and testing began. We almost immediately identified network bandwidth as a showstopper issue and had to install a separate network to accommodate migration traffic without affecting day-to-day operations. Three months before cutover, we did a complete dry run, which told us the live data migration would take 13 days to complete. Doing the dry run so far in advance gave us enough time to do the test again if things didn't go as expected. After the migration was completed, validation was performed by a dedicated customer team.

Even if you're not facing a fixed time window, insufficient testing can be costly. NGS was recently brought in to troubleshoot an issue with a large aerospace company that inadvertently idled 7,000 users for 36 hours when the chosen migration tool encountered a previously unknown bug. This problem would have been identified with proper testing.

Tip 3: Reevaluate Timelines as New Information Becomes Available

Customers who have experience with small migrations are often surprised that multiterabyte migrations designed to minimally impact ongoing operations take many days or even weeks to complete. The planning process alone can easily span months. Even if every other aspect of your migration goes perfectly, the migration won't be considered successful if the timeline isn't accurate or people don't know and buy into it. The results of pre-migration testing should be constantly reviewed and used to finalize your project timeline.

In the project referenced above, timelines were so tight that we only had a single weekend to complete the data migration and cutover to the new storage. By doing exhaustive upfront testing and monitoring data changes after our tests completed, we identified a four-hour start time window in which we had to commence the actual data movement. If we had started sooner, the data movement would have completed too soon (before work activity ceased). If we started later, the data movement would not have completed in time to do all the things we needed to accomplish during the 48-hour downtime window. (Storage systems were being moved to a new data center, so they had to be packaged, moved, and reinstalled.)

Conclusion

Complex data migrations are rarely executed without a certain amount of stress, but it is possible to accomplish them on time and with minimal business disruption. Careful attention to tool selection can eliminate a lot of potential problems. Testing validates your tools and methods, helps you identify potential problems before they affect the migration, and allows you to better understand your migration timeline.

If you need help, [NetApp Global Services](#) has years of experience with data migrations at businesses of all sizes. We are structured to provide help where and when you need it. We can do the whole migration, help with the upfront planning to ensure you make the right planning decisions, or just complement your in-house expertise.

I'd like to be contacted by NetApp Global Services about a migration project!

Contact Me!



Several NetApp customers participated as beta sites for the newly announced NetApp **NearStore® VTL**, an appliance-based virtual tape library solution. One contact at a multi-billion dollar technology company volunteered to share his feedback with Tech OnTap, but was vetoed by his company's marketing team. He talked to us anyway. We've agreed to keep his identity secret, but can tell you that for this review we interviewed a senior systems analyst who leads the evaluation, implementation, and testing of new products for his company's data center.

Early Customer Review:

NetApp NearStore VTL

Beta system: NearStore VTL600

For the initial VTL test, the IT team chose the company's disaster recovery (DR) facility. The reasons were simple: the relatively low volume of tape backups allowed testing with a minimally configured appliance and it enabled the testing to take place in a production environment since tape backups represent the third and even fourth copy of critical information.

The VTL system was installed between a NetApp NearStore R150 and a Spectra Logic 20K tape library. The physical tape library remained connected to the VTL to allow testing of export functionality and restores from tape if needed.

Prior to putting the system into production, the beta user successfully tested:

- Backup and restore procedures (between the R150 and the NearStore VTL)
- Tape exporting (backing up from the R150 to VTL to physical tape)
- Tape importing (importing data from tape to VTL)

The VTL solution fully replaced the tape library for the duration (about four weeks) of the trial. During this time, our contact ran continuous backups from NearStore to VTL. Production runs involving several terabytes of data were backed up flawlessly and there was not a single VTL backup failure.

So what did this beta user think of the NearStore VTL solution?



Installation/Implementation	Installation was remarkably simple. All I did to connect the NearStore VTL to the physical tape library and existing backup infrastructure was switch out a few cables. The VTL recognized all of the physical connections to the robot and installed everything automatically. It also integrated extremely well with VERITAS® NetBackup. Everything was self-configured, and no zoning was needed on the fabric.
Backup Performance	We used the activity monitor in NetBackup to determine throughput speeds and kept a spreadsheet to track performance pre- and post-installation. Within minutes of plugging in the VTL—and without any kind of customization or performance testing—we saw a 4x to 5x throughput improvement.
System Reliability	The NearStore VTL system ran fully unattended without a single failure for the duration of the trial. This immediately freed up a couple hours each day. [Note: The IT team's current tape libraries have a 10 to 20% tape media failure.]

RELATED INFORMATION

- [NearStore VTL Overview](#)
- [NearStore VTL Tech Specs](#)
- [eWeek: NearStore VTL Review](#)
- [Evaluating VTL Solutions](#)
- [D2D Event: VTL Deep Dive](#)
- [NetApp D2D Backup Strategy](#)

Beta Customer Backup Environment

This beta customer relies on tape backups at a main production site and remote disaster recovery facility. At both sites the customer uses:

- NetApp NearStore nearline storage
- NDMP to dump data from the NearStore system to tape
- VERITAS NetBackup 5.0 MP5 to manage the process
- Spectra Logic tape libraries
- Sony AIT-3 tape drives and media

Main production site

- NearStore R200
- SpectraLogic 64K tape library
- 15 tape drives
- Over 150 tapes written each week
- Policies require the retention of patented engineering data for 20 years

Remote DR facility

- NearStore R150
- SpectraLogic 20K tape library
- 4 tape drives
- 30-day retention policy
- Capacity is self-contained; tapes are rewritten as they expire

Seven Questions to Ask a Potential VTL Vendor

1. Can performance and capacity scale to keep pace with my data growth?
2. What is the penalty for using compression?

Management	I flew to our DR facility for implementation and testing. I didn't have to fly back or call anyone on-site for the rest of the trial. There were no surprises, and I didn't have to learn anything new. It's like having a physical robot with the flexibility of configuring it remotely.
Features and Flexibility	I like the fact I can write everything to VTL without worrying about different formats and then pick what I want to export to tape. Otherwise I would have to do separate VTL backups and then do a backup to tape. Also, I don't need a VTL in place to do the restore, which is important if we need to restore at other sites. Finally, the ability to reuse shelves in other NetApp systems is definitely helpful. We have a variety of other NetApp systems, and the ability to swap hardware makes this purchase easier to justify. [Note: Because the NearStore VTL writes tapes in the native backup application format, a VTL system is not needed for restore]
Overall Evaluation	Instead of spending money increasing tape and tape drives, a VTL provides more flexibility and room to grow. It has been very straightforward and easy, not something you need to have additional training to understand. For someone looking to enhance backup and recovery processes, I definitely think VTL might be an option. The NetApp NearStore VTL solution is the first we've tested, and so far we're very happy with it. We have a history with NetApp products and professional services, so we have reason to be confident in their ability to deliver the best solution. We've done four major projects with NetApp. They have been very successful and NetApp has always been very prompt in resolving any issues we encountered.

3. Can the solution deliver maximum performance without manual configuration and tuning?
4. Are physical tapes written in the identical format as the backup application?
5. Is the VTL fully compatible with the leading backup software and tape devices?
6. Can the VTL create tapes directly for optimum performance?
7. Can the VTL fully utilize the speed and media savings provided by tape drive hardware compression?

Read the full [Guide to Evaluating Virtual Tape Systems](#).

Based on the success of this trial, the team has ordered additional shelves and shipped the NearStore VTL system to its production site for a phase II evaluation. The team is "very keen" to increase capacity and conduct additional performance testing to see the full benefits of NearStore self-tuning technology. What's next?

"We would like to remove the physical robot completely in our DR facility. For our production site we won't ever be able to remove the physical robot because we'll always need tape for long-term retention. We can add VTL to the environment, though, and it would be ideal to write anything [with a retention policy] under one year to VTL instead of tape. What would make this really cool is to add indexing and search capabilities. Using VTL for archival with an ILM solution to do the indexing and search and archiving would be the way to go."

Comment on this article.



Rajesh Sundaram

Storage Resiliency Architect, NetApp

During his 8½ years in NetApp Engineering, Rajesh Sundaram has worked on many of the most significant resiliency projects in the company. In addition to being an early member of the team that worked on the WAFL® file system, Rajesh helped lead the rearchitecture of the RAID subsystem and the development of SyncMirror®. Rajesh is currently focused on designing unique new resiliency technologies and improved on-site drive diagnostics.

The Private Lives of Disk Drives

How NetApp Protects Against Five Dirty Secrets

By Rajesh Sundaram

NetApp builds resiliency into its storage systems at every level to ensure that critical data is always protected. If you've been involved with NetApp for a while, you've probably heard a lot about technologies such as SnapMirror®, SnapVault®, and Snapshot™ that protect you from events ranging from sitewide disasters to user and application errors. NetApp also offers a unique degree of resiliency against problems that occur within disk drives themselves, but you've probably heard much less about these technologies.

You may be surprised by some of the "secret" problems that still lurk inside disk drives despite their remarkable dependability. Below are five of the most troublesome disk problems and the resiliency technologies that NetApp Engineering has developed to protect against them.

Secret 1: Drives fail suddenly!



NetApp solution: unique RAID-DP™ technology that provides extra protection against failure.

Okay, this isn't really a secret. Despite their dependability, we all know that disk drives still occasionally fail. When you consider the relatively short production lifecycles of disk drives (most models are only manufactured for a year or two) combined with the huge production volumes for popular enterprise disks (tens of millions of units annually), it's obvious that some problems are going to occur. Occasionally, a component change, manufacturing facility change, generational drive transition, or some other perturbation will result in the production of a less reliable lot. NetApp uses stringent drive screening criteria that meet or exceed industry norms, but some failure modes are extremely time-dependent. This not only causes drives to fail after a significant amount of time has passed but also increases the likelihood of two drives failing near the same time.

One common failure mode is for a disk to suddenly cease functioning. One moment it's working fine, and the next it's gone with no warning. Everyone knows that the way to protect against this sort of failure is RAID, but what if two drives fail in the same RAID group or an uncorrectable media error occurs during RAID reconstruction? Given the rapid market adoption of large-capacity SATA drives, many customers don't realize the odds of "double failure" are increasingly stacked against them.

NetApp protects you from these types of failures with RAID-DP (RAID Double Parity). RAID-DP adds a second parity stripe to drastically increase data availability without sacrificing performance or capacity utilization. Aggregates and volumes using RAID-DP can withstand up to two failed disks in a RAID group—or the increasingly common event of a single disk failure followed by an uncorrectable bit read error from a second disk during reconstruct. (Disk capacities continue to increase, while the media error rate stays about the same. A disk reconstruction must read many more bits of data now than in the past, significantly increasing the risk of a bit error.)



RELATED INFORMATION

- [NetApp Implementation of RAID Double Parity for Data Protection](#)
- [Ready for RAID6 \(InfoStor\)](#)
- [NetApp D2D Backup Strategy](#)
- [Is NetApp Storage Simpler?](#)
- [NOW Customer Site](#)
- [\(password required\)](#)
- [Data Protection Best Practices](#)
- [\(password required\)](#)

RAID-DP vs RAID4

RAID-DP significantly increases data protection with zero to minimal impact to capacity utilization or performance versus RAID4. And, since RAID-DP is an integral part of Data ONTAP, there are no hidden costs associated with it.

RAID-DP offers:

- Protection against up to two disk failures in the same RAID group
- Protection against single disk failure + uncorrectable bit error during the reconstruction time frame
- No significant read, write, or CPU consumption differences
- Larger allowable RAID groups, which mean that capacity utilization stays about the same (one in eight disks dedicated to parity)

Learn more about the [NetApp Implementation of RAID Double Parity](#).

RAID-DP offers the data reliability of mirroring (RAID1) at the price of RAID4. Before the release of RAID-DP, storage administrators typically limited the size of RAID groups to protect against these types of failure. With RAID-DP, NetApp customers can feel confident using larger RAID groups and aggregates.

Secret 2: Drives slowly degrade.



NetApp solution: *unresponsive drive protection takes the drive offline and regenerates data from parity.*

Another common disk failure mode is for a drive to slowly degrade away, resulting in a steady performance decline over time. This can happen for any number of reasons. If you've seen this problem, you know that you can often read all the data stored on the drive before it fails completely.

You may not be aware of the impact such a drive can have on storage performance. A drive with multiple media errors—or a drive with a servo problem—may take several minutes retrying a read until it succeeds. In server environments, the resulting long I/O response times can lead to unwanted connection terminations and noticeable delays on clients.

NetApp engineers specifically designed the Data ONTAP® operating system to anticipate and circumvent potential performance issues. In the event an unresponsive or semiresponsive disk emerges within the system, Data ONTAP ceases all I/O operations to the affected disk, marks it as offline, and serves reads from parity while queuing writes until the disk recovers. If the disk fails to recover, it is marked as failed and reconstructed to a spare.

The innovative disk offline feature (which is available only from NetApp) ensures high performance consistency that is critical to applications that demand consistent quality of service in from the storage subsystem.

Secret 3: A bad drive can lock up an entire FC loop.



NetApp solution: *dual pathing, ESH2, and local SyncMirror prevent data lockout.*

Sometimes firmware bugs or disk failures can result in a single disk locking up an entire Fibre Channel loop, blocking access to up to 84 drives. In these scenarios, the remaining drives are in perfect working condition but temporarily inaccessible until the communication path is unblocked. An errant drive may generate a LIP (loop initialization primitive) storm; the drive continuously issues LIP requests that interrupt ongoing data transmissions.

NetApp offers multiple levels of protection for this problem. Every NetApp drive uses dual pathing in which two independent loops are connected to each drive. If one loop is down, the other provides continued access. If a rare drive failure blocks both loops, dual redundant shelf I/O modules containing second-generation electronically switched hubs (ESH2) detect and bypass disk drives that can disrupt FC operations. In fact, the ESH2 module with firmware revision 15 (FW15) and higher is designed to specifically protect against LIP storms. By electrically isolating these drives from the loop via intelligent point-to-point switching, the ESH2 provides a safety net in addition to dual pathing.

For maximum data availability, customers can deploy NetApp SyncMirror to achieve a level of resiliency that no other storage vendor offers. SyncMirror is local RAID mirroring between two separate volumes on the same storage system. While it also provides improved read performance (similar to RAID1+0) and is an instrumental part of the NetApp MetroCluster disaster recovery solution, SyncMirror stands on its own for customers demanding the ultimate level of local storage resiliency. By ensuring two mirrors are stored on separate failure domains, SyncMirror protects your data against a wide range of rare and unpredictable failures, including dual cable breaks, power strip failures, dual loop failures, disk shelf backplane failures, HBA failures, and even up to five concurrent disk failures on mirrored RAID groups if also using RAID-DP.

For unparalleled local storage resiliency, NetApp recommends SyncMirror for business and mission-critical applications requiring the highest level of data availability.

Secret 4: Firmware bugs can cause silent data corruption.



NetApp solution: checksums and RAID scrubs ensure that correct data is always returned.

It's a well-known fact in the storage world that firmware bugs (and sometimes hardware and data path problems) can cause silent data corruption; the data that ends up on disk is not the data that was sent down the pipe. To protect against this, when Data ONTAP writes data to disk, it creates a checksum for each 4kB block that is stored as part of the block's metadata. When data is later read from disk, the checksum is recalculated and compared to the stored checksum. If they are different, the requested data is recreated from parity. In addition, the data from parity is rewritten to the original 4kB block, then read back to verify its accuracy.

To ensure the accuracy of archive data that may remain on disk for long periods without being read, NetApp offers the configurable RAID scrub feature. A scrub can be configured to run when the system is idle and reads every 4kB block on disk, triggering the checksum mechanism to identify and correct hidden corruption or media errors that may occur over time. This proactive diagnostic software promotes self-healing and general drive maintenance.

To NetApp, rule number 1 is to protect our customer data at all costs. Protection against firmware-induced silent data corruption is an example of NetApp's continuing focus on developing innovative storage resiliency features to ensure the highest level of data integrity.

Secret 5: Committed writes can get dropped!



NetApp solution: lost write protection—the only solution in the industry to protect against this threat.

Brace yourself, because we saved the most insidious disk problem for last. With extreme rarity, a disk malfunction occurs in which a write operation fails but the disk is unable to detect the write failure and signals a successful write status. This event is called a "lost write," and it causes silent data corruption if no detection and correction mechanism is in place. You might think that checksums and RAID will protect you against this type of failure, but that isn't the case. Checksums are written in the block metadata—coresident with the block—during the same I/O. In this failure mode, neither the block nor the checksum gets written, so what you see on disk is the previous data that was written to that block location with a *valid* checksum.

Only NetApp, with its innovative WAFL (Write Anywhere File Layout) storage virtualization technology closely integrated with RAID, identifies this failure. WAFL never rewrites a block to the same location. If a block is changed, it is written to a new location, and the old block is freed. The identity of a block changes each time it is written. WAFL stores the identity of each block in the block's metadata and cross checks the identity on each read to ensure that the block being read belongs to the file and has the correct offset. If not, the data is recreated using RAID. The check doesn't have any performance impact.

NetApp always uses WAFL at the lowest level of disk organization, so even block-oriented, SAN installations have this protection.

Conclusion

Since every storage vendor uses more or less the same disk drives, no one is immune to these problems. Not all vendors, however, can offer equal protection against them. Innovative NetApp technologies, including RAID-DP, unresponsive drive protection, SyncMirror, ESH2, RAID scrubs, and lost write protection, offer a level of security against disk malfunctions that other vendors don't.

Are *you* protected?



Paul Hargreaves

Consulting Systems Engineer and Simulator Junkie, NetApp

A six-year NetApp veteran, Paul is a NetApp expert on Windows® environments and has helped many of the largest companies in the United Kingdom develop their storage strategies. In his spare time, Paul helped write the scripts and documentation for the Data ONTAP® simulator and regularly engages Engineering on updates to the "kernel" of the simulator. When asked why a Windows specialist is involved with a Linux® tool, Paul cites his previous history with Dragon, Spectrum, and Commodore operating systems before commenting, "I've used so many operating systems I've lost count. It doesn't matter what the OS is—this is a really valuable tool."

February's Tool of the Month: Data ONTAP Simulator for Linux

Every month Tech OnTap showcases a free tool that just might make your life a little easier. [Recommend a tool](#) and get a free NetApp cycling jersey.

Author: Network Appliance Engineering

What it is: A tool that gives you the experience of administering and using a NetApp storage system with all the features of Data ONTAP at your disposal.

How it works: The simulator can be loaded onto a Red Hat or SuSE Linux box and looks and feels exactly like Data ONTAP. It has the same code base (with additional wrappers to simulate the hardware) and is included in Engineering's nightly build process. The simulator is available for Data ONTAP 6.4.5 through 7.1RC2.

Why it's cool: Almost anything you can do with Data ONTAP can be done with the simulator. Without purchasing new hardware or impacting your production environment, you can test functionality, export NFS shares, set up fake tape drives, and even simulate two heads on the same box for clustering.

I'm working with NetApp customers using this for...

- **Data ONTAP feature testing.** The simulator includes fully functional license keys for all NetApp software. You can create WORM-protected files using SnapLock® and a simulated set of populated disks, and after the test you just delete the files instead of throwing away drives. You can test NetApp SAN functionality using iSCSI before implementing in a production FC or iSCSI SAN environment. You can also experiment with features such as FlexVol™ and FlexClone™ before deploying Data ONTAP 7.0 on production systems.
- **Application integration.** Application developers use the simulator to experiment with and develop applications that use Manage ONTAP™ APIs. The Manage ONTAP SDK (software development kit) contains documentation and C/C++, Java™, and Perl libraries.
- **Bug fix testing.** The simulator can be used to confirm that a new release fixes a previous issue without having to physically touch a production machine. Assuming the bug fix is proven, you can test it on real hardware and know that time isn't being wasted with upgrades and downgrades.
- **Education.** Every admin on your team can have a personal testing environment.

Caveats: This is *not* a production version of Data ONTAP and should not be used in your production environment. There are inefficiencies (for example, a 1GB disk file will be much larger than 1GB) and performance running on another OS without a disk system behind it will obviously be considerably less than with Data ONTAP. Finally, the simulator can't emulate environments where specific hardware is required (for example, Fibre Channel).

LINKS

- [NetApp ToolChest](#)
(password required)
- [Data ONTAP Best Practices](#)
(password required)
- [Introduction to Data ONTAP 7G](#)
- [NetApp Technical Report Library](#)

Sample Simulator Screens:

