

SPECIAL EDITION

VMware & Deduplication

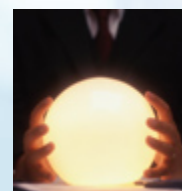
HIGHLIGHTS

- » Virtualizing Storage for VMware
- » Deduplication and VMware DR
- » Deduplicating VMware VMs
- » Resources:
 - Deduplication Guide
 - VMware Best Practices

Future Past and Future Present

Dave Hitz, founder and EVP, NetApp

Ever wished you had a crystal ball to guide you in planning your data center? In this Q&A session, Dave shares his current and past vision of how today's innovations will help shape tomorrow's data center.



[More »](#)

BLOGGING WITH DAVE

Dave Hitz, NetApp Founder and EVP

"Activision implemented remote disk-to-disk copies (SnapMirror®). A few Sundays ago, disaster recovery was 'tested' by two back-to-back power failures. They recovered in less than 4 hours, with no business interruption."

» [Dave's Blog](#)

DRILL DOWN

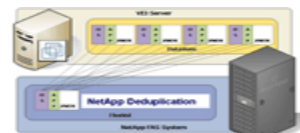
- » **Deduplication Guide**
Read this deployment and implementation guide to get configuration examples, known caveats, and recommendations.
- » **NetApp And VMware VI3 Storage**
Facing VMware® ESX backup and restore challenges on NetApp? Get this 41-page Best Practices report.

TIPS FROM THE TRENCHES

The Perfect Virtual Marriage — Deduplication and VMware

Larry Freeman, deduplication evangelist, NetApp
Bill May, deduplication evangelist, NetApp

Deduplication can reduce your VMware storage footprint by 50% or more. Find out how you can benefit from this unique synergy.



[More »](#)

Virtualizing Storage for VMware Success

Vaughn Stewart, manager
VMware Technical Alliance, NetApp

If you've virtualized your servers but not your storage, you've done only half the job. Learn how FAS deduplication and other NetApp technologies can help you:

- Cut storage costs in half
- Reduce your VMware RTO to minutes
- Increase your overall operating efficiency

[More »](#)

ADMIN RESOURCES

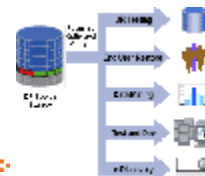
- » **Know Your Space Savings Potential**
Calculate potential storage savings for your VMware environment with the updated deduplication calculator.
- » **Deduplication for VMware Webcast**
Hear how Virginia Credit Union changed its way of provisioning and managing storage with NetApp deduplication for VMware.

ENGINEERING TALK

Combining Deduplication and VMware DR

Jeremy Merrill, product engineer, NetApp

See how applying deduplication to VMware simplifies DR by reducing storage and bandwidth requirements. Plus, get tips on how to leverage your DR data.



[More »](#)

[FEEDBACK](#)

TECH ONTAP ARCHIVE



Dave Hitz

Founder and Executive Vice President, NetApp

As executive vice president, Dave Hitz is responsible for the future strategy and direction of NetApp. He and James Lau founded NetApp in 1992 with the goal of simplifying storage the way Cisco simplified networking. Dave previously worked as a senior engineer at Auspex Corporation and in engineering at MIPS Computer. Before starting his career in the computer industry, Dave worked as a cowboy, getting valuable management experience while herding and branding cattle.

Future Past and Future Present

Podcast transcript: Q&A with Dave Hitz

In a recent NetApp TechTalk podcast, Eric Brown, senior director of Worldwide PR and Internal Communications, asked Dave Hitz about technological changes in the data center. Dave shares with us his view of changes he didn't anticipate and changes he sees headed our way. The full interview transcript is available only from Tech OnTap.

Eric: Dave, let's talk about fundamental shifts in the data center. What is coming up now that you didn't anticipate? What do you see happening in the foreseeable future?

Dave: Five years ago, if you asked me what I saw coming, I would never have said VMware®. It wasn't even on my horizon. Since then, VMware has swept the world in a way that I never would have predicted. Also, I doubt that I would have predicted that our technologies would have proven to be so useful in VMware environments.

In our booth at Oracle Open World, we had a member of VMware's partner engineering team telling customers why he thinks NetApp storage is awesome for VMware. His job is figuring out what VMware can do with partners to make VMware better. Using an Oracle® Database installed on Oracle's Linux®, he showed how fast he could make 25 copies of VMware virtual servers using FlexVol® on NetApp storage. It took only a handful of seconds to get his 25 copies. Cloning is the most compelling feature of FlexVol flexible volumes. For years we've been focusing on FlexVol for test and development. Then our partner comes along and says that he suspects there are more areas we haven't even thought of where FlexVol could be useful. I suspect he's right. Now we're wondering, "How else could you use cloning?"

Another technology that we've had for a while is deduplication—also known as dedupe. Dedupe is just a good idea in general. But when it comes to VMware, dedupe can provide astounding data reduction. By definition, VMware clone copies are redundant, making them good candidates for space reduction through deduplication. There are some zones in which you have a virtual desktop interface where you're just going to create 1,000 copies of something. Each one of the thousand is almost just the same. You can use dedupe in this instance to gain space savings of 1,000:1. You want the savings to be better? Create 5,000 virtual desktops.

**OnDemand
TechTalk Webcast:**

- [Future Past and Future Present](#)
- [The Future of Storage](#)
- [Technologies That Shape Consumer Experience](#)
- [Public Policy and Technology Solutions Address Data Breaches](#)
- [When It Makes Sense to Be an Early Adopter](#)

RELATED INFORMATION

- [The Future of Storage](#)
- [The Future of NetApp: Riding the Technology Curves](#)

NetApp and VMware Webcast Series

The interest in combining VMware and NetApp technology has been intense. To help satisfy the demand for information, NetApp has kicked off a series of Webcasts on various important topics.

Two of the Webcasts have already occurred. You can view them on demand and also read answers to participant questions from each associated chat session:

- [Driving Lower TCO for Virtualized Environments \(chat transcript\)](#)
- [Consolidating Your Windows Environments on VMware and NetApp \(chat transcript\)](#)

Interested in hearing more? Two future Webcasts in the series are scheduled:

- **Go Green: Maximize Power Savings and Minimize Floor Space in Your Data Center**
Date: January 31, 2008
Time: 11:00 a.m.–noon Pacific Time
- **Disaster Recovery and Backup Strategies for Virtualized Environments**
Date: March 20, 2008
Time: 11:00 a.m.–noon Pacific Time

Three or four years ago everybody at Storage Networking World was doing some form of deduplication. There are different ways to do it. We've been doing it for ages just with Snapshot™ copies because they enable us to have lots of copies while keeping most of the blocks the same. Lots of people are using dedupe for backup. We use dedupe ourselves for disk-to-disk backup of our secondary storage systems. But we're also doing something that's unusual. We're using dedupe for primary storage as well. At this time, I think we're unique in the industry in allowing people to use dedupe for primary storage.

There are a lot of places you can get a little bit of benefit from dedupe. There are a few places you can get a jaw-dropping benefit. We recently ran a test with a VMware customer who was running lots and lots of systems that were essentially identical. With virtual desktops, you can have literally thousands of systems that are the same. In the tests we ran, we saw a 95% data reduction because of dedupe. Now, I want to be really clear: I'm not saying dedupe reduces your data by 95% everywhere, but in certain applications you can get very large space savings with dedupe.

Eric: Let's go back to VMware for a second. People used to talk about storage virtualization and server virtualization as two separate concerns. Do you think there's still a difference between them?

Dave: First, I have a big problem with the word "virtualization." "Virtual" has been used in computer science for so many years that whenever you just say the word "virtual," nobody knows what you're talking about. The key notion with VMware is that virtual systems address the fundamental problem of underutilized systems in data centers. It solves the problem of how to drive utilization up.

Virtualization on the storage side is also a lot about utilization. Thin provisioning is all about not wasting space when someone comes to you and says, "I'm going to create a 3TB database. Give me a 3TB LUN." And you look at them and ask, "Are you sure?" And they reply, "Oh, yeah, it's going to be really hot. In fact, it's going to be so painful to grow, give me 4TB." And then you look back a year later and discover they're only using half of their allocated storage. Virtualization is so important in that context.

Today, I think whenever anybody says "virtualization," what they really mean is VMware, Xen, and Viridian. The real question is "What are the issues around virtualization in the data center?" Because, when you virtualize the servers, right behind the servers there's a storage problem waiting to happen. With virtualization you can provision a new server instantly, but without storage virtualization, how long is it going to take you to provision the new storage behind it? If you can provision a new server without adding a CPU, the next question is "Can you provision some new storage without adding disk drives?" What's virtualization VMware? It's a thin CPU.

Eric: Are there any other developments in the data center that have surprised you?

Dave: I didn't really understand where application integration would take us. I could not have predicted that we would have teams of people investigating what we could do to make Exchange better from a storage perspective, what capabilities we could offer to an Oracle Database administrator, and what we could do directly for the database administrator (DBA) using the SnapManager® family of products. The challenge for storage suppliers and customers is that most people don't think about storage as a concern until after they've thought of everything else. Most people think first about how to solve their critical business problems and the applications they need to do that. First they think about Oracle Financials, SAP, and Exchange—then they think about storage. The closer you can be to that application, the more relevant your products will be to them.

Eric: Let's drill down a little bit on application integration. After buying Oracle, SAP, or Microsoft® SQL Server™, they think SAN, or NAS, or direct attached when they think of storage. What should they be thinking about?

Dave: One of the biggest challenges in selling storage systems with advanced features is figuring out how to give the folks who actually use the storage access to those advanced features. Let me create a concrete example. I gave a presentation at Oracle Open World with Neil Armstrong, the CIO from Activision. He runs the whole Oracle suite—the database and the e-business suite. He loved the Hyperion

[Register](#) for either or both of these live events.

The Future of Storage (2007)

Every year we ask Dave to talk about the important trends that are impacting the data center and affecting storage deployment. If you read the transcript at left, you already know what's on his radar screen for the coming year.

Last year, Dave was thinking about a variety of technologies, including:

- Storage virtualization and Data ONTAP® 10G
- iSCSI
- Encryption
- The impact of inexpensive ATA disk
- Storage grids

[Read the transcript](#) or [listen to the podcast](#) to learn more.

Storage Evolution

Like any technology, storage adoption generally moves in waves. New technologies emerge in response to the limitations and challenges of existing technology, while other technologies mature and move into decline.



In June 2006, NetApp founder and EVP Dave Hitz outlined his predictions for the next 20+ years of storage technology evolution.

Don't miss [Examining the Future](#).

acquisition because he was already a Hyperion user. The challenge he faced was how to create copies of his entire production database for the various things that he needed to do. The process for creating a single copy was a week long. In fact, creating copies was such an ordeal for him that he pushed back on anybody who wanted one. It was just too painful for him. He thought he was doing well when after major reengineering he got the whole process down to four days. He was thrilled. After he brought in NetApp with FlexClone®, creating a copy of his entire production database took just four hours.

Here's the challenge. Not all storage systems are created equal. NetApp systems have cloning capabilities that allowed Activision to create a copy of the database in just four hours. Their previous environment simply didn't support that. But that only addresses part of the problem. More to the point: How are DBAs going to access that storage feature? They're going to have to call the server guy who's connecting the LUNs, and then the server guy has to talk to the storage guy, and then the storage guy ... What a mess!

The whole point of application integration is creating an application that speaks the database administrator's or Exchange user's language. To do this you have to customize application integration to focus on things that the DBAs or users care about. In Oracle, one of the big issues is test and development. With Oracle, your production environment may be running just fine, but you still need to be able to create clones for test and development. Neil creates these clones all the time. He has said to me, "I've probably got 12 of them. Some guy wanted to run a class and said, 'Can I have a clone just to do some training?' 'Yeah, no problem.' I never would have given them a clone before." What gives a DBA the ability to say "no problem"? That's where application integration comes in, and that's where you need to optimize your storage solution for the particular task.

With SnapManager for Exchange, what's important is how easy we make it to restore an individual mailbox or the whole mail environment. Each SnapManager product we offer is tuned for a specific application environment. SnapManager creates a bridge from the storage world, which is kind of esoteric if you're not a storage person, up to the application level, where more companies are investing their time from a business perspective.

Eric: Let's go from that big picture to some specific technologies. iSCSI or IP SAN seems to have been really successful over the last couple of years. Where do you see that evolving?

Dave: First of all, I have to say NetApp takes Ethernet storage of all types very seriously. We demonstrated Fibre Channel over Ethernet (FCoE) at the recent Storage Networking World expo. FCoE is not yet very mature, whereas iSCSI is shipping, and a lot of people are using it in surprisingly high-end production environments. I wonder whether FCoE is really that much better than iSCSI or whether it is a ploy by people who aren't ready for iSCSI to slow down the market. It may turn out that FCoE is great. But if you're thinking that block-based storage over Ethernet might be something you're interested in, don't wait. If you buy systems from NetApp, you can use iSCSI today, and you can switch to FCoE someday—it's not an issue. But don't be fooled into waiting.

Eric: Well, tech is definitely about instant gratification, so if you don't have to wait, why wait? Let's switch from protocols to disk drive technologies. Serial attached SCSI or SAS drives seem to be coming on big in the market. What's your take on that?

Dave: Here's my big picture view of the disk drive industry. They sell two kinds of disks: the expensive ones with high margins and the cheap ones with low margins. So be thinking filet mignon and McDonald's.

Whenever a new disk drive technology comes out, you always have to figure out which style it is. When you look at the expensive disk drives, there are reasons why they're expensive. They have more revolutions per second and typically have faster seek times. They're better. Fibre Channel drives are also more reliable. The cheap stuff spins more slowly, takes longer to get your data off, and the reliability is not as high. So you really have to understand where in the market the disk drives will be used.

I'm concerned by what the disk drive industry is saying about serial attached SCSI (SAS): that it would be the best of both worlds—cheaper but still the best. I don't really believe it. My instinct is that as SAS gets more mature it will be the replacement for Fibre Channel. The interconnect has some very nice characteristics. What I love especially is that the interconnect allows you to plug in either SAS or SATA. You can build a shelf that supports either one and that will allow people to choose more easily between the cheap one and the expensive one. Look! The same plate holds both filet mignon and a McDonald's hamburger.

Eric: Now let's talk about the future of the disk overall. Is the hard disk moving toward oblivion? What does flash mean for the disk drive business?

Dave: Is the hard disk moving to oblivion because of flash? First of all, absolutely not. Capacity improvements in the disk drive industry actually slowed down for a little while, but they're back up now with perpendicular recording. Currently, disk drives are improving about as fast as flashes, and they're improving from a capacity point of view. So, people who need a lot of storage are not going to be using flash any time soon.

What is flash good for? Think about the iPod. Some early versions had disk drives spinning in there. The latest generation just has flash. The same thing's going to happen with the laptop. In fact, I'm running my laptop today—with all my personal data—on flash. Even in the high-end enterprise storage environment, flash will be good for speeding up performance. If you look at flash, it's about 10 times more expensive than disk. For random access, it's a hundred times faster. So it makes a lot of sense to incorporate flash into your architecture. In the same way that you can use RAM to make your cache bigger, you can use flash to make your cache way, way bigger. I think that's the direction where flash is going.

Eric: Can you talk a little bit more about backup technologies? What future do you see for virtual tape libraries?

Dave: Let me tell you a story about virtual tape libraries. When we bought Alacritis to acquire our virtual tape libraries, they were surprised. They knew that we already had many of the capabilities in disk-to-disk backup that we acquired with Virtual Tape Library, and they wondered why we would buy them. I think it is important to understand that tape is deeply embedded in the backup process for large enterprises. They like the whole disk-to-disk idea. Tape is just not fast enough for most applications—certainly not for recovery—so they start looking for intermediate technologies. And VTL is perfect for that. If you think about it, though, most of them are not replacing tape with VTL. They're saying, "You know, I used to go to tape every night, but now maybe I'll go once a week or possibly even once a month." They are also using VTL to buffer the data on the way to tape, because it actually slows down tape a lot when it starts and stops. They can get so much more throughput out of their existing tape infrastructure if they buffer it through a VTL.

Eric: What are your thoughts on continuous data protection, or CDP?

Dave: The big question is what exactly is meant by "continuous"? Some people say that the thing they hate about backups is that if they lose their data, they'd have to go back all the way to the night before. To them, continuous backup might mean once an hour. NetApp has been doing Snapshot copies once an hour for 15 years now. That's not new at all. But some people think that it's not continuous unless the Snapshot interval is all the way down to the exact millisecond. The question I have is "How small an interval do you need"? I'm skeptical of that last millisecond. In a lot of environments, when you're looking at CDP as a backup technology, you're talking about a lot of downtime anyway. You're going to be replaying log files, and once you get into replaying log files, the question is whether or not you can get the data out of the log file anyway. My instinct is that an interval of an hour probably is not fast enough. An interval as short as a minute is getting a lot closer, but an interval of seconds is in the zone. I'm not saying that an interval of milliseconds is never useful, but it's probably going to be required only by a very select subset of your environment, which really puts a limit on the cost of implementing it.

Eric: Let's talk about storage and security. How big an issue is security, and what advancements can we expect?

Dave: I think data security still is a big issue—especially for financial services and healthcare. It makes sense because those are the most personal topics you can think of. A lot of companies have Sarbanes–Oxley and security issues regarding confidentiality of their financials, and it's a relatively small part of their overall business. I believe that in terms of who's going to use security features, the big pressures for every company will come from headquarters, M&A, and financial departments. In the healthcare and financial industries, security will be much more broadly deployed until it's just built in as a standard part of every system. When that happens, a lot of people will say, "Oh, well, since it's there, I will turn it on." That's how I see that playing out.

Eric: Do you see security moving into the fabric?

Dave: Tom Georgens, our EVP of Product Operations, has an interesting theory. He believes that most data center environments only have a few zones of investment. They have the applications that run on the servers, they have the servers, they have the network that connects everything, and they have the storage. Tom's theory is that people really don't want another zone of infrastructure to manage. As a result, they may deploy something like a storage virtualization device or a Decru DataFort® data encryption device temporarily. Long term, their strategic question is whether they really want to manage another piece of infrastructure permanently or whether they could attach it to one of the zones they already have. When you look at the data encryption technique for storage, you see there are two natural places to put encryption—into the fabric or the storage system. And my guess is that it will go into both places over time.

Eric: How do you see the storage services industry changing?

Dave: I love storage services. A lot of our channel partners do storage services as part of their business. When you look at our business model, we focus on products. However, if we need to sell services to help our customers install products, use the products, and get the best performance out of the products, we'll sell those services. Because we focus on products, if partners can do the services business just as well as us or, in some cases, better than us, we would like to have our partners delivering those services. That said, we're committed to delivering a consistent, high level of services around the world.

Eric: One last question. How would you sum up what you've learned from tracking developments in the data center over the last year?

Dave: The innovation we're seeing today isn't coming from technology itself. Today's innovation is coming from new uses of old technology and the business processes that exploit existing technology.

TECH ONTAP ARCHIVE



Larry Freeman and Bill May
Deduplication Evangelists, NetApp

Larry Freeman (left) and Bill May have worked extensively with NetApp deduplication customers since they initiated the deduplication Early Access Program (EAP) a little over a year ago. Since then, they have seen nearly 2,000 NetApp deduplication systems go out the door and observed countless examples of customers using deduplication in a variety of environments.

With over 30 years of IT and data storage experience, Larry specializes in SAN technology, storage management, and the protection and retention of critical data. Bill has extensive experience with networked storage and data protection technologies and is a frequent and popular presenter at seminars, conferences, and briefings.

The Perfect (Virtual) Marriage: Deduplication and VMware

By Larry Freeman and Bill May

VMware® has become one of the more popular use cases we've seen for NetApp deduplication. Shortly after the release of deduplication with Data ONTAP® 7.2.2, customers began reporting great success deduplicating VMware virtual machines (VMs), both in traditional V13 environments and also emerging VDI environments. Naturally, we wanted to take a closer look and discover why they were so excited. The answer came quickly: *They were consistently seeing space savings of 50% or more with virtually no performance impact. Some were obtaining storage savings as high as 90%. Here's how..*

NetApp Deduplication

The unique advantage of NetApp deduplication is that it can take any NetApp flexible volume (FlexVol® volume)—regardless of how the data was written into the volume—and easily identify and eliminate duplicate blocks within that volume. If two or more blocks are the same, we eliminate the duplicate blocks and change the data pointers so that all the duplicates are redirected to a single data block.

It doesn't matter what the blocks are or what application they belong to; if the blocks are the same, the duplicates are eliminated. This is in sharp contrast to most other deduplication products out there, which are predominantly limited to use with a single application—typically backup.

Another NetApp advantage is that you can deduplicate existing data volumes. You don't have to have deduplication running from the start. You can take a volume that's been in use for a long time and recover significant disk space through deduplication.

How NetApp Deduplication Works

When deduplication is enabled on a volume, it creates a list of the digital fingerprints that represent all blocks in use. These fingerprints are already part of the Data ONTAP metadata, so it is not necessary to create a new one for each block. By comparing these fingerprints, it is relatively easy (that is, system overhead is low) to determine which blocks are duplicates. (Possible duplicates are compared to ensure they are indeed the same.) Then, it's just a matter of bookkeeping to change the reference pointers and eliminate the duplicates.

The fingerprints are only used to identify duplicate blocks; they are not used to look up or access data. Thus, data access remains fast and is not subject to data corruption due to the deduplication process.

RELATED INFORMATION

- In this issue:
 - [Virtual Storage and VMware](#)
 - [Deduplication and VMware DR](#)
 - [Deduplication Comes of Age](#)
- [Using NetApp Snapshot with VMware](#)

Top 10 NetApp Deduplication Questions

Want to learn more about NetApp deduplication? A recent Tech OnTap article answered some of the most often asked user questions, including:

- How do I add deduplication?
- Why do I need a NearStore license?
- Can the system do other work while deduplication is running?
- Can I estimate my space savings ahead of time?
- Can previously written data be deduplicated?

[Read the article](#) to find out the answers to these questions and 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:

By the way, this is the same basic process we've been using with our Snapshot™ technology for over a decade: using one "physical" data block to represent many "logical" data blocks. The deduplication process is simply run on a volume periodically whenever you need to reclaim storage space. Because of its low overhead, NetApp deduplication can be used with a wide range of workloads. [A more detailed explanation of how deduplication works appeared previously in Tech OnTap.](#)

VMware Environments

VMware is a terrific technology that reduces the number of servers needed in the data center by consolidating several physical servers into one "virtual" server. VMware accomplishes this by allowing users to first create a master template for each application environment, then to "clone" these templates into many VM images. Once the clones are created, they are installed concurrently as "guests" on a single server. By virtualizing your server environment, you can utilize your servers much more efficiently.

VMware users typically run six to 10 VM guest operating systems per physical server, although we have heard from some customers that they are running up to 70 VMs on a single server.

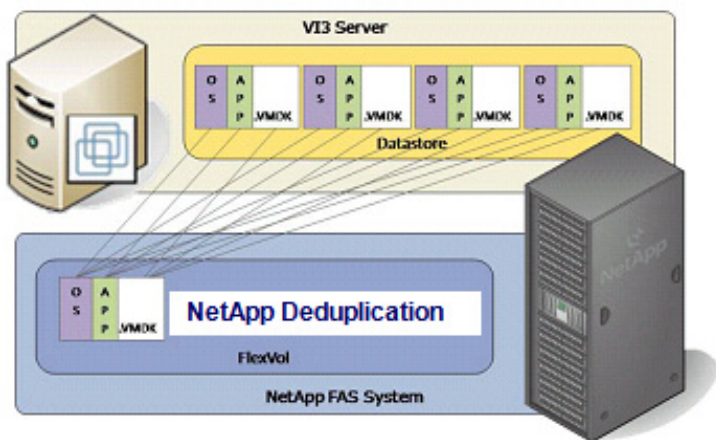


Figure 1) Deduplicating VMware VMs.

The Perfect Marriage

While VMware provides a valuable cost benefit by consolidating your servers, it is not quite so efficient at consolidating the storage used by VMware clones. That's where deduplication comes into the picture.

Each cloned VM image requires the same amount of physical storage space as the template from which it was created, but it is largely redundant. This makes them good candidates for space reduction through deduplication, but—because VMware is a primary storage application—users are reluctant to impose any additional load on these servers, which might degrade end user read/write response times.

NetApp deduplication solves this problem. Because it provides deduplication with minimal system performance intrusion, users can substantially reduce the amount of storage capacity required to house VMware clone copies without affecting business workflow.

How is this possible? NetApp deduplication is an intrinsic part of Data ONTAP and its WAFL® file system. Unlike other forms of deduplication, NetApp deduplication utilizes many characteristics inherent within the storage operating system. There are no need to create complicated hashing algorithms, no lookup tables to search for and reconstitute data, and no rewriting of data during the actual deduplication process.

All that's required is a small digital fingerprint for each 4KB WAFL block (they already exist in the system), a quick comparison of these fingerprints, and a simple block-redirect process to rereference the original data block. Duplicate data blocks are then released back to the system.

NetApp deduplication is performed as a low-priority background process. This process can be run automatically any time the VMware data grows beyond a predetermined

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

Learn more. Read [Five Ways to Use NetApp Snapshot Copies in VMware Environments](#).

How Data Deduplication Fits into the NetApp Master Plan

The following is an excerpt from a recent post on NetApp founder Dave's blog:

Buying less storage is the small picture. The big picture is that we want to help customers create a disk-based copy for all of their primary storage. ...

Interesting things start to happen when you create a disk-based copy of everything. Instead of doing searches on primary storage, which could hurt performance, why not search the secondary copy? If the people running decision support systems want their own copy of a critical database, why not clone the secondary instead of paying for a whole new copy? Why not create lots of cloned copies for the test and development team preparing to upgrade to the next version of Oracle or SAP?

When you create a copy of everything, and add functionality like Snapshot copies and clones, what you end up with is a smart copy infrastructure that can completely change the way you think about data management.

This won't happen overnight. We understand that. But anything that helps people reduce the cost of creating copies helps us achieve our vision more quickly. In the short run, data deduplication helps customers save space and save money, but what's more important is that by reducing the cost of copies, it helps us achieve our master plan.

Read the [full Dave's blog post on deduplication](#).

Deduplication Deployment and Implementation Guide

Want to learn even more about NetApp deduplication? This comprehensive guide gives you everything you need to understand, plan, and deploy deduplication across your NetApp environment, including:

threshold, or it can be scheduled to run only during convenient off-peak hours.

Sounds Good, How Do I Start?

To get started, you first have to add NearStore® and deduplication licenses to your system. Then you can enable deduplication on your desired volume(s) with a simple CLI command. This will trigger the process of gathering fingerprints in each enabled volume. If you have existing data in the volume, NetApp deduplication can optionally scan that data too. Once deduplication has been enabled, it's simply a matter of deciding how often you want to reduce your volume space requirements by running the deduplication process. Most customers run deduplication nightly, since their daily data change rate is normally low enough that the deduplication process will run quickly.

Summary

Customers have been pleasantly surprised to find that NetApp deduplication fits easily within their VMware environment. Commonly heard assessments of the technology are "it couldn't be simpler" and "it just works."

- A complete technology overview
- Supported configurations
- Setup and operation details
- Best practices for deployment
- Using deduplication with cluster and DR configurations
- Troubleshooting

[Download the guide.](#) (PDF)

TECH ONTAP ARCHIVE



Vaughn Stewart

VMware Technical Alliance Manager, NetApp

Vaughn Stewart has been with NetApp since 2000, and during the past seven years he's worked with hundreds of customers on their VMware® implementations. Vaughn has coauthored four VMware on NetApp white papers and worked with VMware on multiple joint solutions. Vaughn currently holds industry certifications on solutions offered by Cisco, IBM, Microsoft, NetApp, Sun Microsystems, and VMware. Prior to NetApp Vaughn worked as a systems administrator and an MSSP consultant architecting and supporting a wide range of solutions.

Virtualizing Storage for VMware Success

By Vaughn Stewart

Server virtualization has made day-to-day life a whole lot simpler for systems administrators; however, the benefits have not been shared by storage administrators and backup operators, who face new challenges as the result of data encapsulation. Storage infrastructure and operational practices that were designed for physical server architectures are not well suited to meet the demands of a virtual infrastructure.

Backup and disaster recovery practices have to change radically. New light needs to be shed on the importance of data protection and recovery time objectives (RTOs). Any consolidation effort also consolidates risk. With VMware the point of risk has shifted away from servers to shared storage. A RAID failure in a virtual infrastructure is catastrophic, as it potentially impacts dozens of virtual servers and applications. Any restore operation that requires data to be copied from another form of media can take days.

NetApp storage virtualization technologies solve these challenges and alleviate the burden on storage and backup teams. In this article, I'm going to describe how NetApp can:

- Cut storage costs by half or more
- Reduce VMware RTO to minutes
- Increase efficiency to decrease soft costs

If you've virtualized your servers but not your storage, you've only done half the job. Virtualizing your storage infrastructure is critical to addressing the unique storage needs of VMware.

Cut Storage Costs in Half

The combination of rapidly growing data storage coupled with the need for greater data availability has significantly increased storage costs. NetApp directly reduces the cost of VMware storage with three distinct technologies:

- RAID-DP™ improves disk availability without doubling cost.
- FAS data deduplication eliminates redundant data in production data sets.
- Flexible volumes allow dynamic storage management without disruption.

RAID-DP

As individual disks have increased in capacity, the odds of a double-disk failure occurring have become a mathematical certainty. With a traditional array you have to choose between reducing storage costs (RAID 5) or increasing data protection (RAID 10). NetApp RAID-DP solves this issue with a dual-parity implementation that provides the protection and performance of RAID 10 at a RAID 5 price point. Released in 2003, RAID-DP is implemented on more than 95% of all NetApp storage volumes

RELATED INFORMATION

- In this issue:
 - [Deduplication and VMware DR](#)
 - [The Perfect \(Virtual\) Marriage](#)
 - [Deduplication Comes of Age](#)
- [Using NetApp Snapshot with VMware](#)
- [Love VMware but Hate Backups?](#)
- [Achieving 99.99% Availability with Virtualization](#)

Love VMware but Hate Backups?

While no one disputes the obvious benefits of VMware and server virtualization, the radical change in architecture has significant implications for backup, HA, disaster recovery, and so on.

In an earlier article, this month's author and NetApp VMware expert Vaughn Stewart tackled some of the thorniest questions posed by end users. Examples include:

- The fact that a single service disruption could affect my entire environment is keeping me awake at night. Do I have to live with this level of insecurity?
- My backup processes haven't changed, but suddenly I'm spending almost all my time managing backups. How do I stop the madness?

Read [Love VMware, Hate Backups?](#)

in production, so you can feel comfortable the technology is mature.

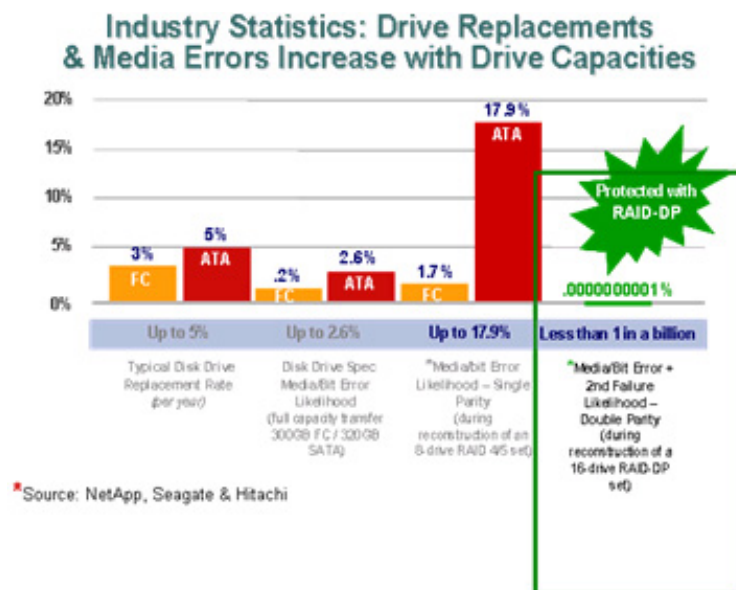


Figure 1) Drive replacement and media error data.

Data Deduplication

VMware provides a method of provisioning a new server from an existing server template. A template is a server image that includes operating system binaries, patches, and standard application sets. Provisioning a new virtual server by cloning a template creates an identical server that is operational within minutes.

NetApp FAS deduplication enhances VMware provisioning technology by transparently deduplicating the identical parts of each virtual server to reduce production storage footprints by 50% or more with some obtaining storage savings as high as 90%. Since virtual servers deployed from templates are identical by design, this is an ideal usage of FAS data deduplication. You can free up significant amounts of storage and delay additional storage purchases. As we'll see later, replicated virtual servers at a DR site also inherit this reduced storage footprint automatically.

Improved Storage Utilization

Unlike traditional RAID arrays, NetApp creates resource pools of storage termed aggregates, which separate the physical realm of disks and RAID from the realm of storage virtualization and storage management. Flexible volumes (FlexVol® volumes) can serve up both NAS and SAN data and allow you to carve up your storage resources more efficiently, analogous to the way that VMware lets you divvy up your server resources.

FlexVol volumes spread each storage volume across a large number of spindles to ensure performance from every volume regardless of size. Using the NetApp FlexShare™ capability you can set I/O policies in a manner similar to VMware SRM in order to shape I/O limits and ensure quality of service to a particular virtual server.



Figure 2) Space utilization in traditional RAID versus NetApp. NetApp spreads all disk volumes across a large number of spindles to achieve higher utilization without sacrificing performance.

Leveraging Virtualization to Increase Availability

Availability is crucial to the healthcare field. Medical centers are increasingly dependent on digital technologies, driving a need for critical applications and data to be accessible at all times

As a leading healthcare provider, Sentara is committed to leverage the latest technology for the best possible customer care. To meet its availability needs, Sentara has undertaken a number of recent IT initiatives:

- Installed NetApp V-Series systems to virtualize existing storage from other vendors.
- Implemented NetApp SyncMirror® to maintain two fully consistent copies of critical data at all times.
- Consolidated servers using VMware VI3 and NetApp storage.

[Read the full article](#) to find out the details.

87% Space Savings in 60 Days

The County of Sacramento has a backup environment that includes approximately 2.1TB of data.



The agency saves daily incremental Snapshot copies for two weeks and keeps weekly full backups for 60 days. Every day, approximately 5% of files and 0.5% of blocks change.

By deploying a NetApp storage solution that includes NetApp deduplication technology, the agency saw results that included:

- ~ 87% space savings at 60 days
- Backup 16x faster
- Restores 6x faster: five minutes vs. 30 minutes

[Hear Keith Scott, County of Sacramento IT analyst, describe the deployment](#) in a recent Webcast.

Reduce Your VMware RTO to Minutes

Backing up data from a large virtual infrastructure is hard. There are many overlapping factors beginning with the disproportionately large amount of data that is addressed by each physical ESX server, coupled with limited CPU cycles and I/O bandwidth to run backup jobs. The result is that traditional backups are impractical.

The mere thought of the job scheduling, time requirements, I/O bandwidth, and media required to periodically move full, unduplicated copies of data to backup storage is intimidating. It's like emptying a swimming pool with a straw. It can be done, but it is long, hard work.

NetApp provides a set of technologies that offload backup operations from the VMware server. The NetApp approach combines our patented Snapshot™ technology with SnapMirror®.

By creating Snapshot copies at the shared storage level, backups are done in seconds, and you avoid having to manage backup schedules and jobs for every individual VM. The result is immediate, consistent backups that are available online on the production storage system. Creating Snapshot copies has virtually no performance impact, and storage utilization is very efficient because space is only consumed as blocks are changed.

Snapshot copies provide two immediate benefits:

- A VM can be restored immediately if it fails for any reason.
- You can create more frequent restore points and retain a greater number. This becomes a necessity as you virtualize more and more of your business—critical applications.

NetApp SnapMirror allows you to take your entire production data set, including saved Snapshot copies, and replicate the whole thing to another site. If you've deduplicated your primary VMware environment, you've already significantly reduced your DR storage (and bandwidth) requirements, making this approach cost effective.

With the same amount of storage required to run a production copy of VMware on a traditional array, NetApp lets you support a production and a DR copy of the data. In case of a disaster, you can bring up anything from a single VM to an entire data center in minutes, running directly off the replicated data sets. The whole process is hours or days faster than tape, and recovery is much faster than other backup solutions that don't allow you to fail over your environment directly to the backup. You can learn more about VMware disaster recovery in [a companion article in this issue.](#)

Increase Operating Efficiency

Combining deduplication with NetApp Snapshot and SnapMirror gives you local backup in seconds, local or remote recovery in a matter of minutes, and significant storage savings (50% or higher) in both your primary and DR locations. In addition, you no longer have to micromanage backups. NetApp software manages everything at the storage level.

In a traditional storage environment, every time you add another VM you increase the storage capacity you need and increase the time needed to do backup, and someone has to make all the configuration changes that are necessary and monitor the results. With NetApp deduplication, your storage requirement for a new VM is minimal, and your new VM will inherit the same data protection policies that have already been applied to other VMs in the volume, so there's no work to do. Storage scales easily with your VMware environment, reducing management overhead and increasing efficiency. Because all NetApp storage supports multiple protocols (NFS, iSCSI, and FC SAN), you can easily deploy whichever storage protocols make sense for your environment.

Full Virtualization Is the Key

When it comes to building a virtual infrastructure, virtualizing your servers is only half of the job. NetApp storage systems virtualize your storage in the same way VMware virtualizes your servers, with resource pools providing performance and capacity, deduplication to share disk space between multiple VMs, and NetApp FlexShare to give priority to storage used by certain VMs to ensure quality of service (QOS). Backup and off-site replication based on Snapshot alleviate the burden of traditional

backup while ensuring the quickest possible recovery time.

NetApp lets you increase the utilization of your hardware so that you get more from each dollar you spend, while providing unsurpassed data protection and data management for your entire VMware environment.

TECH ONTAP ARCHIVE



Jeremy Merrill

Product and Partner Engineer, NetApp

Since joining NetApp over two years ago, Jeremy has focused exclusively on issues of data protection in both remote offices and data centers. Recently, he has worked on data protection and disaster recovery as it pertains to VMware. Jeremy has spent over seven years in the storage industry, including working extensively with backup, disaster recovery, and storage management in data center environments.

Combining Deduplication and VMware DR Cascading Savings Improves Cost Effectiveness

By Jeremy Merrill

The shift from physical servers to consolidated, virtualized server infrastructures has had undeniable IT benefits. However, the rapid move to VMware has caused traditional approaches to disaster recovery (DR) to become outmoded and added a layer of complexity to DR implementation.

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 necessary at the DR site.

The cost savings created by using deduplication can make DR feasible in situations in which it may have otherwise been cost prohibitive. For instance, one customer reported that after deduplicating his VMware Virtual Desktop Infrastructure (VDI) environment, the storage and bandwidth needed to provide DR for his desktops was relatively minor and made adding DR feasible for his VDI environment as well as his VI3 environment.

In this article I'm going to explore what it takes to implement deduplication with VMware DR. I'll also talk about leveraging the replicated data in your DR environment for DR testing and other purposes.

Implementing Deduplication in Primary VMware Environment

Because each virtual machine in a VMware environment requires dedicated storage for its operating system, there is a large amount of duplication. You may have numerous VMs that are all installed with more or less the same operating system and applications.

If you have 100 VMs running the same OS and each virtual machine requires 10GB to 20GB of storage, that's 1TB to 2TB of storage dedicated to almost identical copies of the same data. Applying NetApp deduplication can eliminate much of this redundancy.

In general terms, if you have X virtual machines assigned to a storage volume, after deduplication you will need approximately 1/X the amount of operating system storage you would require in a non-deduplicated environment. Obviously, the actual results

**OnDemand
TechTalk Webcast:**

Deduplication for
VMware Environments:
A Shift in Storage
Provisioning

RELATED INFORMATION

- In this issue:
 - [Virtual Storage and VMware](#)
 - [The Perfect \(Virtual\) Marriage](#)
 - [Deduplication Comes of Age](#)
- [Using NetApp Snapshot with VMware](#)

Engineering Perspective: Deduplication Comes of Age

At its heart, NetApp deduplication relies on the time-honored computer science technique of reference counting.

When deduplication is enabled on a volume, it computes a database of fingerprints for all of the in-use blocks in the volume (a process known as gathering). Once this initial setup is finished, the volume is ready for deduplication.

To avoid slowing down ordinary file operations, the search for duplicates is done as a separate batch process. As the file system gets updated during normal use, WAFL® creates a log describing the changes to its data blocks. This log grows until one of the following occurs:

- The administrator issues a start command
- The next time specified in the schedule occurs
- The changes to the log exceed a predetermined threshold

[Read the full article.](#)

you achieve will depend on how many VMs you have in a volume and how similar they are.

In practice, customers typically see space savings of 50% or more in ESX V13 environments, with some obtaining storage savings as high as 90%. This is for deduplication of the entire VMware storage environment including application data—not just operating systems. In VDI environments customers typically see space savings up to 90%.

Another advantage of NetApp deduplication is that not only can it run on primary storage, it can run on any existing NetApp volume. Even if your VMware infrastructure is well established, you can run deduplication and free up significant storage space. All that is required is the deduplication license (which is free of charge) plus a NearStore® license on the target storage system.

Configuring for Disaster Recovery

While the reduction in storage use in your primary storage environment is a significant benefit by itself, the true leverage you get from deduplication becomes apparent when you implement disaster recovery with NetApp SnapMirror®. Because deduplication significantly reduces the amount of data that must be replicated, it reduces both the space needed in your DR location and the network bandwidth needed between sites. After deduplication, you may be able to configure DR over a slower link than would otherwise be possible, and it will be easier and faster to get your DR environment up and running.

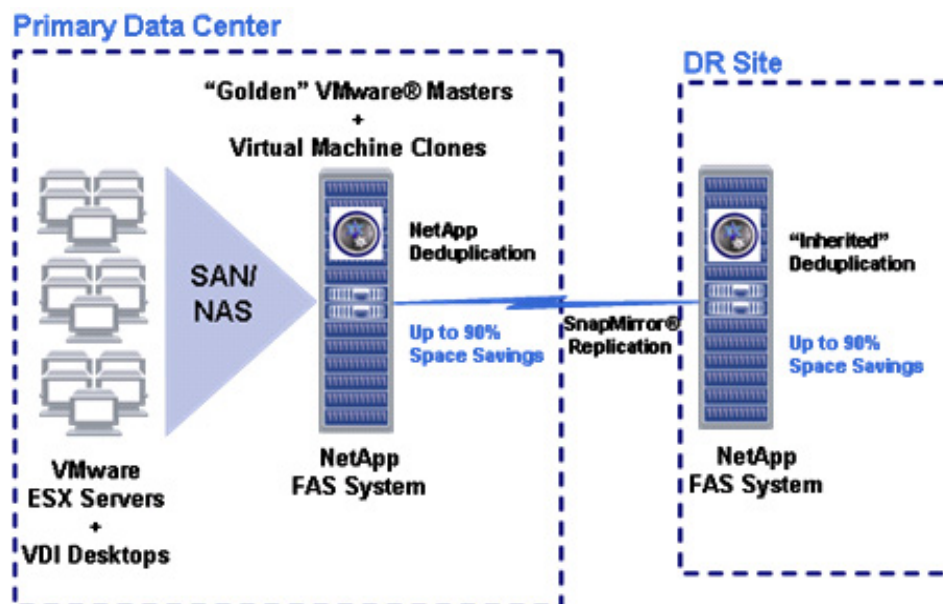


Figure 1) Applying deduplication in a VMware environment with replication for DR.

To configure DR, you first deduplicate all the volumes in which your data stores reside in your primary VMware storage environment. Then you create SnapMirror relationships between your primary volumes and target volumes at your DR site.

Unlike many other replication solutions, SnapMirror does not require that the target configuration be identical to the source. You can use a different NetApp storage system and less-expensive disk (e.g., SATA instead of Fibre Channel disk) in your DR site if you wish.

When SnapMirror first runs it will sync each source volume with its target. This process is typically the most bandwidth-intensive part of SnapMirror implementation, but since the source volumes are all deduplicated, the amount of data that has to be transferred is much less than would otherwise be the case. This approach is ideal for customers who have slow links and wouldn't otherwise have the bandwidth to do the initial sync but who can manage the incremental updates that occur thereafter.

Note that because deduplication acts at the volume level, you must use Volume SnapMirror to get the maximum benefit. Volume SnapMirror acts on the entire volume, so your mirror always retains the same level of deduplication as the original, saving

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

Learn more. Read [Five Ways to Use NetApp Snapshot Copies in VMware Environments](#).

Choosing a Replication Option

NetApp offers several tools for data replication: SnapMirror, MetroCluster, and ReplicatorX™.

If you have only NetApp storage systems, SnapMirror or MetroCluster is the best replication solution in most circumstances. Functionality ranges from asynchronous to fully synchronous and goes beyond disaster recovery with smart copies for DR testing, test/dev, data migration, data mining, offloaded backup, and so on.

To find out more about SnapMirror, read [NetApp SnapMirror Best Practices Guide](#). (pdf)

Choose ReplicatorX if you have third-party primary storage and you want a tiered storage solution (different, less-expensive storage on the back end) or you have a variety of primary storage systems and need a solution that can cover all of them.

Also, you should choose ReplicatorX if you have federated applications (SAP, DB2 IEE, etc.) and data needs to be recovered at the exact same time across multiple servers. The ReplicatorX solution has patented global clock synchronization technology that guarantees write order consistency across multiple servers at the target location.

To learn more read:

- [Streamlined DR and More for VMware ESX](#)
- [Data Replication and Recovery over Any Distance with ReplicatorX](#) (pdf)

space, decreasing bandwidth utilization, and speeding up the mirror update process.

Once you complete the initial sync, you configure SnapMirror to run on a set schedule to keep your DR site up to date. In each iteration, SnapMirror transfers only the blocks that have changed, so it uses network bandwidth very efficiently.

You will need to run deduplication periodically on the primary site. Depending on your particular needs this can be done:

- On a schedule that you specify
- Automatically whenever there is 20% new data in a volume
- Manually as needed (for instance, after installing a big patch)

With SnapMirror, whatever happens on the primary volumes is automatically reflected on your secondary volumes, so there is no need to run deduplication at your DR site. Since your secondary volumes are mirrors they "inherit" the deduplicated state from the primary volumes.

Leveraging Your DR Environment

Once you've got all your data at your DR site and it's being updated regularly via SnapMirror, it's still not the end of the story. NetApp also makes it possible to leverage the data stored at your DR site for DR testing, development, or a variety of other purposes.

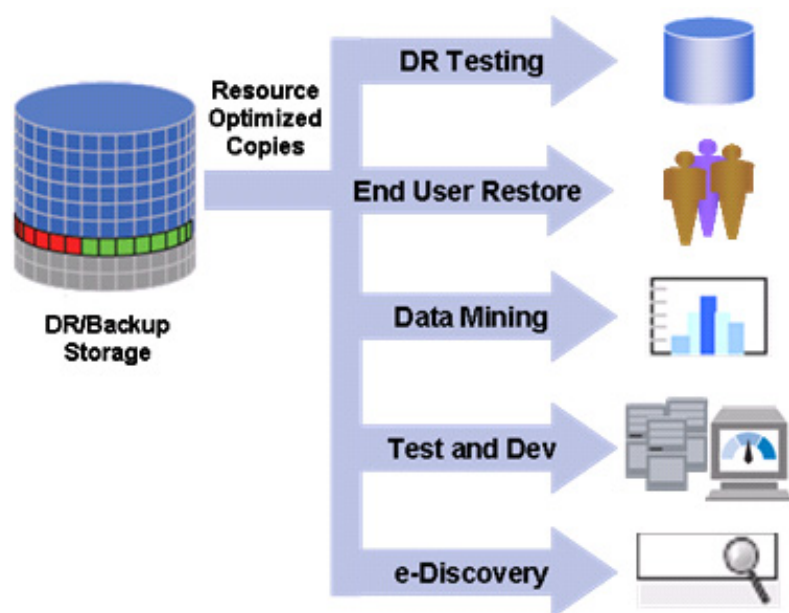


Figure 2) Leveraging FlexClone® at the DR site allows replicated data to be used for a variety of purposes.

In a typical DR testing scenario, all the data for the test must be copied to another set of disks before testing commences. That means you need 2X the storage space right off the bat, and you've got time-consuming copies to make before you can start testing.

With NetApp FlexClone technology, you can make space-efficient, writable clones of any or all of your DR volumes; additional space is only consumed as you make changes to the cloned volume. These FlexClone volumes allow you to capture a static view of your DR data at a fixed point in time without disrupting ongoing SnapMirror updates and without requiring massive amounts of additional storage.

Using FlexClone you can reduce the time it takes for DR testing from 24 hours or more down to a few hours, because the process is fast, reliable, efficient, and far less resource intensive. It's also possible to use FlexClone in a similar fashion for application development work, data mining, patch testing, and so on.

A DR site represents a substantial investment in resources. Using FlexClone, you can leverage those resources for other tasks without negatively impacting DR readiness. By simplifying DR testing, FlexClone makes it easier to meet company-mandated DR

testing requirements to ensure DR readiness.

Conclusion

Applying NetApp deduplication to your primary VMware storage yields substantial benefits across both your primary and DR infrastructure. In a typical scenario, you can reduce your primary storage needs by 40% to 60%. This reduction's trickle-down effect reduces the storage needed at the DR site and the bandwidth needed for DR by a corresponding amount, making DR faster and more cost effective. With NetApp FlexClone, you can leverage the data in your DR site for DR testing, application test/dev, or other activities to maximize resource utilization.