

This article was originally published in NetApp's Tech OnTap Newsletter. To receive the newsletter monthly and enjoy other great benefits, sign up today at www.netapp.com/go/techontap.

THE PERFECT (VIRTUAL) MARRIAGE: DEDUPLICATION AND VMWARE

BY BILL MAY AND LARRY FREEMAN

LARRY FREEMAN AND BILL MAY

DEDUPLICATION EVANGELISTS, NETAPP

Larry Freeman 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.

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 V3 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.

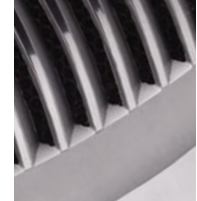
Another NetApp advantage is that you can deduplicate existing data volumes. You don't have to have deduplication running from the start. You can take a volume that's been in use for a long time and recover significant disk space through deduplication.

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

RELATED INFORMATION

- Virtual Storage and VMware (www.netapp.com/go/techontap/mat/ dedup_VM_storage.html)
- Deduplication and VMware DR (www.netapp.com/go/techontap/mat/ dedup_VM_DR.html)
- Deduplication Comes of Age (www.netapp.com/news/techontap/dedupe.html)
- Using NetApp Snapshot with VMware (www.netapp.com/go/techontap/mat/ vmware_snapshots.html)



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.

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 (www.netapp.com/go/techontap/matl/dedup_VM_DR.html).

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.

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" (www.netapp.com/go/techontap/matl/vmware_snapshots.html).

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 (www.netapp.com/news/techontap/top10_a-sis.html) to find out the answers to these questions and more.

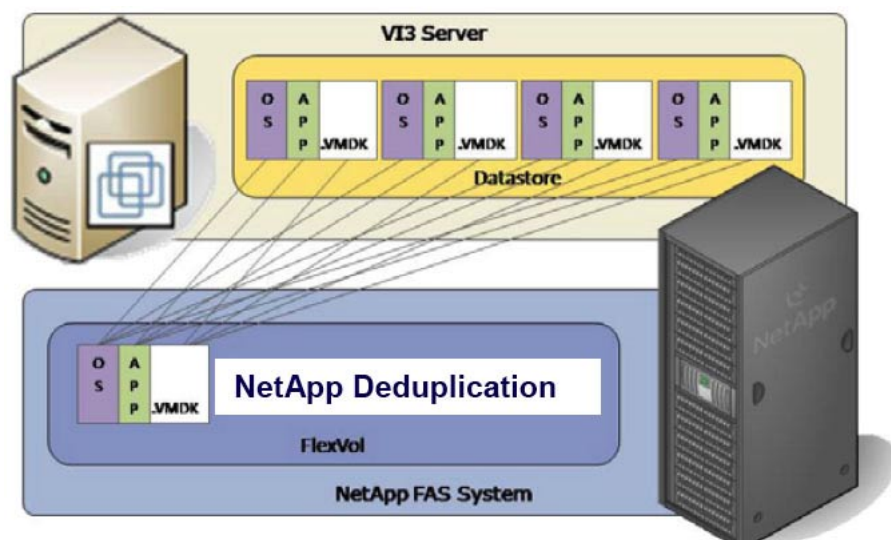
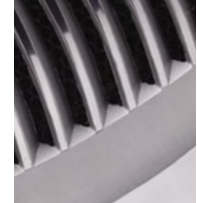


Figure 1) Deduplicating VMware VMs.



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.

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 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.

HOW DATA DEDUPLICATION FITS INTO THE NETAPP MASTER PLAN

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

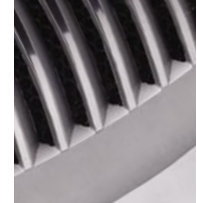
Buying less storage is the small picture. The big picture is that we want to help customers create a disk-based copy for all of their primary storage...

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 (http://blogs.netapp.com/dave/2007/05/how_data_dedupl.html).



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."

Tech OnTap

Insights for Simplifying Data Management

SUBSCRIBE NOW

Tech OnTap provides monthly IT insights plus exclusive access to real-world best practices, tips and tools, behind the scenes engineering interviews, demos, peer reviews, and much, much more.

Visit www.netapp.com/go/techontap to subscribe today.

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:

- 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 (www.netapp.com/library/tr/3505.pdf).

