Technical white paper

Protecting VMware Virtual Machines with HP StoreOnce D2D systems and the VMware vStorage APIs for Data Protection

Table of contents

Executive summary 2
Introduction 2
Technology overview 2
HP StoreOnce D2D systems key features and benefits 2
HP StoreOnce D2D systems in small to large data centers 3
Overview—why D2D for VMware virtual machine backup? 3
VTL vs. NAS—advantages of each for VMware VM backup 3
HP StoreOnce D2D systems for VMware VM backup 4
VMware virtual infrastructure backup components 4
VMware virtual machine backup structure 5
Capacity planning 5
Capacity planning usage models 6
Sequential virtual machine backups 8
Simultaneous virtual machine backups 9
Changed Block Tracking backups 10
Data protection application backup block size 11
Virtual machine disaster recovery with D2D remote replication 11
Recovery scenarios 12
Recommendations 13

Conclusion 13
For more information 13
Executive summary

VMware VM are increasingly prevalent in today's business atmosphere. VMware customers need efficient, high-performing, and reliable backup systems that are easily integrated into the vSphere environment. With increasing virtual machine count, keeping backup infrastructure costs under control can be challenging. HP StoreOnce Disk-to-Disk (D2D) systems provide a disk-based data protection platform while addressing data growth by applying HP StoreOnce deduplication software for efficient, longer-term data retention.

This document describes the benefits of using HP StoreOnce D2D systems combined with a data protection application and the vStorage APIs for Data Protection (VADP) to back up VMware VM. This document also recommends backup and recovery implementations for VMware environments. Key recommendations for VMware VM backup:

- To improve backup storage utilization: Back up VM images sequentially for quickly rising deduplication ratios that are maintained over time.
- To reduce backup windows: Back up multiple VM images simultaneously to improve backup throughput performance. Interleaving backup data will cause a small decline in deduplication ratios; however, the benefits of data deduplication will still be realized.
- To decrease network load: Configure VMware Changed Block Tracking (CBT) to reduce the amount data in incremental backups. The D2D data deduplication will further reduce the amount of data stored for the backup.
- To increase backup speed and deduplication ratio: Use a larger backup block size for faster backup throughput performance and better D2D data deduplication ratios.
- For the trade-off between backup impact and ease of recovery: Configure a weekly full and daily incremental backup schedule to reduce the amount of end-to-end data and decrease the time required to run daily backups. The D2D data deduplication will further reduce the amount of data stored for incremental backups.
- For efficient and cost-effective movement of backup data offsite: Use the D2D remote replication feature to seamlessly replicate all VM backup images to an appliance in a remote facility for simpler recovery in the event of a disaster.

Introduction

HP StoreOnce D2D systems are a disk-based backup system that delivers leading price/performance, and deduplicates VMware VM image backups. The D2D can be used to automate and consolidate the backup of many virtual machine images onto a single, rack-mountable device while improving reliability by reducing errors caused by media handling.

For business environments with remote offices or a disaster recovery site, the D2D can be used to replicate data to a central data center or remote facility.

HP StoreOnce D2D systems are ideal for virtual machine backup images. Proper configuration of VMware VM backups with a data protection application and the VMware VADP to the D2D provides the best backup throughput performance and data deduplication ratios. D2D systems integrate seamlessly into current IT environments and offer the flexibility of both VTL and NAS targets (common internet file system (CIFS) only on legacy systems; CIFS and network file system (NFS) on current systems) for VM backups.

Technology overview

HP StoreOnce D2D systems key features and benefits

- **HP StoreOnce deduplication to store more data on disk**
  
  HP StoreOnce deduplication reduces the disk space required to store backup data sets without impacting backup performance. Retaining more backup data on disk for longer enables greater data accessibility for rapid restore of lost or corrupt files and reduces downtime.

  Deduplication ratios are strongly influenced by two factors—data change rate and retention periods of the data on the deduplication appliance. Low data change rates and data retained for longer periods of time yield higher deduplication ratios.
• **Deduplication Enabled replication**

HP StoreOnce deduplication is the technology enabler for HP D2D Deduplication Enabled replication, which allows fully automated replication over low-bandwidth links to a disaster recovery site, giving ROBO and small data centers a cost-effective DR solution for the first time.

• **Rapid restore of data for dependable, worry-free data protection**

HP StoreOnce D2D systems offer immediate access to backups for rapid restores. StoreOnce deduplication allows more data to be stored closer to the data center for longer periods of time, which offers immediate access for rapid restores.

• **Automate, simplify, and improve the backup process**

HP StoreOnce D2D systems automate your backup processes, allowing you to reduce the time spent managing your data protection. Implementing hands-free, unattended daily backup is especially valuable for environments with limited IT resources such as remote or branch offices.

D2D systems can back up multiple servers via a standard Ethernet or Fibre Channel network simultaneously to a disk-based solution at peak speeds of up to 2.4 TB per hour instead of sequentially to a tape drive or autoloader, meaning that you can substantially reduce your backup window.

D2D systems can be intuitively managed and configured by using the built-in Web browser’s administrative interface. And for large deployments of replicating StoreOnce appliances, the StoreOnce Replication Manager can monitor multiple backup systems throughout the data center. D2D systems are self-managing backup appliances that require little, if any, routine maintenance. Unlike other disk-based storage devices, D2D systems do not require virus protection or logical unit number (LUN) provisioning.

### HP StoreOnce D2D systems in small to large data centers

#### Overview—why D2D for VMware virtual machine backup?

As customers of small to large virtual machine implementations prepare to upgrade or migrate data protection environments, an opportunity exists to improve functionality, performance, and total cost of ownership with HP StoreOnce D2D systems. Customers with VMware VM may wish to migrate or consolidate disparate small systems into a scalable D2D system.

Migration of VM backup from traditional physical tape devices to the D2D can be seamless because customers may use existing processes and policies on the new platform. Customers with complex backup schedules will find the D2D is a platform that provides an opportunity to migrate to a disk-to-disk backup environment maintaining existing processes used with physical tape using VTL.

Customers who use network-attached storage (NAS) file devices on general-purpose shares as backup targets may migrate to a D2D to mitigate contention and performance issues found in VM I/O intensive backups, business, and other applications sharing the same storage array.

#### VTL vs. NAS—advantages of each for VMware VM backup

HP StoreOnce D2D systems may be configured with VTL and NAS targets for VM image backup. This section describes VTL and NAS configuration differences and HP recommended uses.

D2D systems are designed as target storage for backup applications that need rapid restores and reduced backup windows. As such, a D2D configured as VTL or NAS uniquely offers:

- Backup to disk, including VMware VADP based backups
- Easier setup and maintenance
- Data deduplication, allowing more backup data to be retained on disk for longer periods
- Data replication for cost-effective offsite storage

VTL and NAS configurations benefit from the same deduplication and replication features; however, note the comparisons and recommendations in table 1.
**Table 1: Comparison of NAS targets and virtual tape devices**

<table>
<thead>
<tr>
<th>VTL targets</th>
<th>NAS targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommended for customers looking to leverage their current investment in tape hardware and software; easing migration using the same VMware VADP backup policies</td>
<td>Recommended for customers in non-tape environments, used to the simpler method of backup and recovery; HP recommends not using as general-purpose NAS share</td>
</tr>
<tr>
<td>Appears to the backup application as physically connected tape devices, with backup and recovery jobs managed in the same manner</td>
<td>Seen by the backup server operating system, application, and users as a NFS or CIFS disk share</td>
</tr>
<tr>
<td>Acts as a staging post to tape because backups are in tape format; supports tape offload using a backup application copy</td>
<td>Enables use of backup application functionality that is only available with file shares, such as Disk-to-Disk-to-Tape implementations</td>
</tr>
<tr>
<td>Requires a backup application that supports backup to tape and media management</td>
<td>May be used with backup applications that do not support backup to tape</td>
</tr>
<tr>
<td>Integrates into the current backup routine and requires no special backup jobs</td>
<td>Requires a new backup job, but is simple to set up</td>
</tr>
<tr>
<td>May require purchase of a “tape backup” license for the backup application</td>
<td>Licensing is unique per backup application</td>
</tr>
</tbody>
</table>

**HP StoreOnce D2D systems for VMware VM backup**

An important part of virtual machine administration is maintaining a consistent set of virtual machine image backups available for recovery. When data is lost due to user error, system failure, or site catastrophe, there is a need for virtual machine image and possibly application data recovery. HP StoreOnce D2D systems integrated with a well-planned data protection strategy that includes regular virtual machine image backups will maintain a consistent set of virtual machine images for recovery purposes.

**VMware virtual infrastructure backup components**

**VMware ESX server**

VMware ESX server is virtual infrastructure software for portioning, consolidating, and managing computing resources.

**VMware Virtual Machine**

VMware VM run on ESX servers and emulate various servers based on different operating systems.

**vCenter server**

The vCenter server runs a Windows® service or can be installed as a VM appliance. The service acts as a central administrator for VMware ESX servers that are connected on a network.

**vStorage APIs for Data Protection (VADP)**

The VADP enable data protection software to protect system, application, and user data in your virtual machines in a simple and scalable way.

**Backup server**

For complete integration with VADP and HP D2D backup systems, a physical backup server is best. The backup server uses data protection software working with VADP to backup VM images.

Figure 1 illustrates a backup and recovery topology with a backup application that supports D2D backup systems and VMware VADP.
VMware virtual machine backup structure

To successfully recover a VMware virtual machine from all types of failure, the following may be required:

- ESX host installation media
- ESX host configuration information—disk storage, networking, licensing, etc.
- Virtual machine image backups

It is not common practice to back up ESX hosts as they are easily rebuilt from the installation media.

Capacity planning

The amount of required backup storage for VMware VM backups depends on the following:

- Size and number of virtual machines
- Site backup retention policy
- Type (full, incremental, differential) and frequency of backups
- Rate of change of the VM data
- The deduplication ratio achieved by the D2D

HP StoreOnce D2D systems do not deduplicate across VTL or NAS shares. Each VTL and NAS share is an independent deduplication domain. A unique D2D VTL or NAS share should be created specifically for virtual machine backup images. Multiple VTL or NAS shares may work best for larger environments with disparate VM operating systems.

Note: The rate of change of a VM refers to the amount of data that would be contained in a differential backup as a percentage of a full backup. A 100 GB full backup with a subsequent 5 GB differential backup before the next full backup would be a 5 percent rate of change.
A VMware administrator may desire to have 2 week’s worth of backups stored on the D2D for quick recovery access. Data deduplication provides more backup space without increasing the physical capacity of the backup device; however, a dynamic VM environment with changing data affects the backup data deduplication ratio.

Figure 2 shows the data rate of change on deduplication ratios when backing up VMware virtual machine images. In performing these tests, HP used standard Windows and Linux virtual machines and a customer representative dataset with realistic structure and content. For the chart below, data was updated between each backup until the desired rate of change was reached.

**Figure 2: VMware VM rate of change on D2D deduplication ratios**

[Bar chart showing D2D deduplication ratio per VMware VM rate of change]

### Capacity planning usage models

A VMware environment with 100 VMs of 75 GB each with a 14-day backup data retention requirement can have several D2D backup usage models. Usage models change based on variables such as the following:

- **Backup schedule type**
  - Daily VADP full image backups deduplicate well but use more server and D2D compute and bandwidth resources during a backup.
  - Weekly full with daily incremental backups do not deduplicate as well but use less compute and bandwidth resources.
  - End-to-end data compaction for weekly full with daily incremental backup schedules is comparable to daily full backups and may be slightly better.

- **VM daily rate of change**—Lower change rates result in better deduplication ratios and require less D2D backup storage.

- **Backup block size**—The backup block size set by the data protection application can affect D2D deduplication ratios. A larger block size will usually result in higher deduplication ratios.

- **Sequential or simultaneous VM backups**—Multiple VM backups running simultaneously typically have better backup throughput but affect D2D deduplication ratios.

- **VM operating system type**—Windows VMs tend to deduplicate at a higher rate than Linux VMs.
Figure 3 compares six usage models for VMware backups with the following common characteristics:

- Number of VMs: 100
- VM size: 75 GB per VM
- Backup schedule: daily
- Retention period: 14 days

Each usage model shows the overall size of the VM backup data without deduplication (14 daily backups of 100 VMs each with 75 GB = approximately 100 TB) versus the size of the data on D2D after deduplication.

**Figure 3:** Capacity planning usage model storage requirement with and without HP D2D data deduplication

Note: The deduplication ratios are a factor of the data type, daily rate of change, sequential vs. parallel VM backup, VM operating system type, backup block size, and backup retention period as observed and measured during testing. Deduplication ratios will vary according to the environment.
Sequential virtual machine backups

With VMware VADP, it is possible to back up one VM at a time or multiple VMs simultaneously (see VMware documentation for limitations). If VM backup throughput performance is not the highest priority, we recommend backing up a single VM at a time for better deduplication performance.

Figure 4 illustrates how D2D deduplication ratios rise quickly and are maintained over time for sequential VM backups when compared with simultaneous VM backups.

**Figure 4:** VMware VM backup deduplication ratio trends over time
Simultaneous virtual machine backups

Backup applications supporting VMware VADP can back up multiple VMs simultaneously. If backup throughput performance is important, sending multiple VM images to the D2D in parallel will result in better performance.

Figure 5 illustrates how D2D backup throughput benefits from multiple VM backups simultaneously.

Figure 5: Simultaneous VM backup

Leveraging VADP, backup applications can send multiple VM images simultaneously during a backup.

When using D2D VTL devices, the backup application should be set up to use a separate virtual tape device for each parallel VM image to enable better throughput performance.

When using a D2D CIFS target, the backup application should be set up to use a disk writer for each parallel VM image to enable better throughput performance.

For example, four simultaneous VM backups will perform better if writing to four D2D VTL tape devices or a single D2D CIFS share with four disk writers.

Note: A backup application that sends simultaneous VADP VM backups to the same D2D VTL or NAS share will result in interleaved backup data. Interleaving the data could cause a decrease in D2D data deduplication ratios.
Changed Block Tracking backups

With VMware Changed Block Tracking (CBT) enabled for VM image backups, an increase in backup performance can be expected; however, the D2D deduplication ratio will decline.

Figures 6 and 7 illustrate how D2D deduplication ratios decrease and backup throughput increases when using CBT. The figures show statistics for a weekly full and daily incremental backup schedule.

**Figure 6:** VM image backup deduplication ratio with and without CBT

![VM image backup deduplication ratio with and without CBT](image)

**Figure 7:** VM image backup throughput with and without CBT

![VM image backup throughput with and without CBT](image)
Data protection application backup block size

When doing VADP-enabled VM image backups to the D2D, if a larger backup block size is configured in the data protection application for writes to the backup device, there is likely to be an increase in the deduplication ratio and faster backup throughput performance.

Figure 8 illustrates how D2D deduplication ratios increase when using a larger block size.

**Figure 8: VM image backup deduplication ratios for various block size settings**

![Graph showing VMare backup block size deduplication ratio with 5% rate of change](image)

Virtual machine disaster recovery with D2D remote replication

Most companies recognize the importance of a robust data protection strategy. Enterprise-level customers are likely to invest in site disaster recovery. In addition, many companies, large and small, are protecting VMs in remote offices where untrained IT staffs are expected to manage a daily backup process—generally involving the changing of physical tapes, which is a process prone to human error.

HP Low-bandwidth replication (LBR), available on its D2D backup systems, offers the solution to both of these problems by allowing local VMware backup data to be replicated between sites in a reliable, automated manner. This provides the following flexibility in VM recovery:

- VMs can be recovered at the D2D source site (original VM location).
- VMs can be recovered at the D2D target site (remote location).
- If a total disaster of the VM source site occurs, the target site D2D can be shipped to the source site, or the backup data can be replicated back to the source site for complete VM recovery.
Recovery scenarios

Figure 9 illustrates disaster recovery scenarios that may occur and the recovery path available when pairing two D2D units for remote replication.

**Figure 9:** Recovery scenarios

---

**Scenario 1**  
**Disk failure**

1. Recover failed disk storage
2. Recover VM image

**Scenario 2**  
**Site failure**

1. Recover source site hardware
2. Ship target D2D to source and integrate with backup application
3. Recover VM image

**Scenario 3**  
**Site failure**

1. Integrate target D2D with backup application
2. Recover VM image

Source site  
Target site
Recommendations

- A data protection application that supports the VMware vStorage APIs for Data Protection and D2D backup systems is required.
- Sequential vs. simultaneous VM backup
  - Back up VM images one at a time to enable the best possible deduplication ratio.
  - If backup throughput performance is the highest priority, configure VADP and the data protection application to send multiple VM images simultaneously to the D2D; however, the data will be interleaved on the D2D, resulting in a decline in the deduplication ratio.
- CBT
  - Enabling CBT for VM image backups should result in an increase in backup throughput performance; however, the deduplication ratio is likely to decline.
- Backup block size
  - The backup block size used by the data protection application can affect data deduplication ratios on the D2D.
  - Use a larger block size for improved write performance and deduplication ratios.
- VM full backups vs. weekly full with daily incremental backups
  - Daily full backups deduplicate at a much higher rate than weekly full with daily incremental backups but require more server and D2D processing resources and bandwidth.
  - End-to-end data compaction and D2D data storage required is roughly the same for both types of backup schedules over an extended time period.
  - Daily full backups deduplicate better, but weekly full with daily incremental backups send much less data to the D2D for deduplication processing.
  - If daily full backups are not required, HP recommends a backup schedule that includes incremental backups to reduce the resource load required for VM image backup.
- Disaster recovery
  - D2D remote replication offers a low-bandwidth replication solution to and from remote sites, which is ideal for VMware disaster recovery.
  - VMware VM image backups to a D2D pair configured with remote replication provides recovery for local disk failures or complete site failures by keeping VM image copies at the local and remote sites.

Conclusion

VMware customers demand an efficient, reliable data growth management backup system environment while keeping costs under control. HP provides a variety of reliable data protection storage solutions that address such requirements. HP StoreOnce D2D systems is one such solution. D2D systems offer high performance and reliability while addressing data growth through HP StoreOnce data deduplication technology for small to large customers. D2D systems integrate easily with leading data protection products to protect important data for mission-critical applications. Combining HP StoreOnce D2D systems with leading virtualization products provides a comprehensive data protection solution for VMware VM.

For more information

HP StoreOnce D2D systems
hp.com/go/d2d
hp.com/go/storeonce

HP StoreOnce D2D systems user guide

HP StoreOnce D2D systems Linux and UNIX® configuration guide
Address your requirements for an efficient, reliable data growth management backup system in a VMware environment while you keep costs under control by combining HP StoreOnce D2D systems with leading data protection products. Visit hp.com/go/storeonce.