

Enterprise Backup and Restore technology and solutions



LESSON VII

HP Data Protector Overview and Concepts Part V

Veselin Petrunov Backup and Restore team / Deep Technical Support

HP Bulgaria Global Delivery Hub Global Operations Center November, 2013





HP Data Protector Overview and Concepts

Part V

Veselin Petrunov / November 2013

Contents

Part V

8. Disk backup

- 1. Overview
- 2. Disk backup benefits
- 3. Data Protector disk-based devices

9. Synthetic backup

- 1. Overview
- 2. Synthetic backup benefits
- 3. How Data Protector synthetic backup works
- 4. Restore and synthetic backup

10. Split mirror concepts

- 1. Overview
- 2. Supported configurations

8. Snapshot concepts

- 1. Overview
- 2. Storage virtualization
- 3. Supported configurations
- 9. Microsoft Volume Shadow Copy Service
- 1. Overview
- 2. Data Protector Volume Shadow Copy integration
- 3. VSS filesystem and disk image backup and restore
- 10. Data Protector Deduplication
- 11. VM environment backup
- 12. Backup scenarios
- 13. Further information



Disk backup

Overview

Industry has requirements for increasingly faster methods of backing up and restoring data. In addition, it has become more and more important that the time required for data backup and restore be reduced to a minimum so as not to interrupt the day-to-day running of company applications.

Many applications and databases frequently make small changes to existing files or produce many new files containing business-critical data throughout the working day. These files need to be backed up immediately to guarantee the data in them will not be lost. This requirement means that a fast medium that can store large amounts of data that works without interruption is necessary for storing data.

Disk-based storage media have become increasingly cheaper in recent years. At the same time, the storage capacity of disks has risen. This has led to the availability of low-cost, high-performance single disks and disk arrays for storing data.

Disk backup (also known as disk-to-disk backup) is becoming ever more important. In the past, tape storage was the favored medium for backup and restore because of its price and effectiveness in meeting disaster recovery requirements. Today, more and more businesses are augmenting their tape storage backup solutions with faster disk-based backup solutions. This ensures faster data backup and recovery.



Disk backup benefits

There are many situations in which it is advantageous to use disk-based devices when performing backups. Disk-based devices are, in fact, specific files in specified directories, to which you can back up data instead of or in addition to backing it up to tape. The following list indicates some situations in which disk-based devices are particularly useful:

- Many applications and databases continuously generate or change a large number of files, which contain business-critical data. Under these circumstances, it is necessary to continuously back up the files concerned, in order to guarantee the capability of restoring them without data loss. In these environments, tape devices typically have to operate in stop/start mode, because they do not receive a constant data stream. This may result in the tape device limiting access to the files concerned. In addition, the lifetime of the backup device may be greatly reduced.
 Alternatively backups can be performed to any disk-based device, overcoming the limitations described. As a short-term backup solution, this is adequate in itself. If a longer term backup solution is required, the data in the disk-based devices can be moved periodically to tape to free up the disk space. This process is known as disk staging.
- In environments that have fast, high-capacity disk drives and slow tape drives, you can shrink the backup window by performing backup to disk-based devices first and moving the data to tape later.



Disk backup benefits

- Using disk-based devices for backup enables you to take advantage of advanced backup strategies such as synthetic backup.
- Disk-based devices are useful for providing fast restore capability for recently backed up data. For example, backup data could be kept in a disk-based device for 24 hours to enable fast, convenient restore.
- Mechanically, a disk-based device is quicker to use than a tape. When using a disk-based device there is no need to mount and unmount a tape. When backing up or restoring a small amount of data, a disk-based device is quicker because it does not need the initialization time that a tape drive requires. With a disk-based device there is no need to load or unload media, which consumes a significant amount of time in a small backup or restore. The advantages of using a disk-based device are even more evident when restoring from an incremental backup.
- The risk of media problems such as faulty tapes and tape mounting failures are reduced to a minimum. The availability of RAID disk configurations provides protection of data in cases where a disk fails.
- Overhead costs are reduced because there is no need for tape handling.
- Overall, disk-based storage space is becoming increasingly cheaper even if compared to tape-based storage.



Data Protector disk-based devices

Data Protector has the following disk-based devices:

- Standalone file device
- File jukebox device
- File library device

Recommended disk-backup device

Hewlett-Packard recommends using the file library device as the preferred disk-based backup device. The file library device is the most flexible and intelligent of the set of disk-based backup devices. It can be reconfigured at any time during use and is capable of performing more sophisticated disk space handling than any other disk-based backup devices. Furthermore, it enables the use of advanced backup strategies such as synthetic backup.

Data format

The data format of the disk-based devices is based on the tape data format. Data Protector converts the data to be backed up into tape format before it writes the data to the disk-based device. With file libraries used for **virtual full backup**, distributed file media format must be used. Select this format in the device's properties.



Synthetic backup

Overview

With the volume of data increasing and backup windows shrinking, performing a full backup often presents a problem in terms of time and storage space. On the other hand, having many incremental backups can be problematic because each incremental increases the time needed to perform a restore.

As backup to disk is gaining popularity due to the high performance and capacity as well as increasingly lower price of disks, new opportunities have arisen. The industry's requirements are to minimize the backup window, minimize the load on production servers and the network, and enable a quick restore. These requirements are met by synthetic backup. Synthetic backup is an advanced backup solution that produces a **synthetic full backup**, an equivalent to a conventional full backup in terms of data, without putting stress on the production servers or the network. A synthetic full backup is created from a previous full backup and any number of incremental backups.

Performing synthetic backup eliminates the need to run regular full backups. Instead, incremental backups are run, and subsequently merged with the full backup into a new, synthetic full backup.

This can be repeated indefinitely, with no need to run a full backup again.

In terms of restore speed, a synthetic full backup is equivalent to a conventional full backup. The restore chain consists of only one element, so a restore is as quick and simple as possible.



Synthetic backup benefits

Synthetic backup brings the following benefits:

- It eliminates the need for full backups. After the initial full backup, only incrementals are performed, which significantly reduces the time needed for the backup.
- Consolidation of backed up objects is performed on the device server, putting no stress on either the production servers or the network.
- A type of synthetic backup, called virtual full backup, is even more efficient. Virtual full backup consolidates data using pointers, which eliminates unnecessary duplication of data.
- A restore from a synthetic full backup is as fast as from a conventional full backup, as there is no need to retrieve data from incremental backups. This eliminates the reading of each incremental backup in the restore chain, and if tape devices are used, also loading and unloading of several media and seeking for object versions.



How Data Protector synthetic backup works



Virtual full backup





Synthetic backup and media space consumption

If synthetic backups are performed frequently, and the sources are kept, this typically means significant space consumption on the backup media. However, if virtual full backup is performed, the backup media space consumption is minimized.

With virtual full backup, the space consumption largely depends on the size of the backed up files. If the files are significantly larger than the block size used, virtual full backup achieves maximum savings of the space compared to normal synthetic backup. On the other hand, if the files are smaller than the block size, the savings are rather small.



Restore and synthetic backup



Full and incremental backups





Restore and synthetic backup



Regular synthetic backup

Synthetic backup and object copy



How data protection periods affect restore from synthetic backup

Data protection of a conventional full backup and all incremental backups that precede synthetic full backup does not compromise a successful restore.

By default, the last synthetic full backup in the backup chain is used for restore, irrespective of whether the preceding backups are still valid or their protection has already expired and the objects are removed from the IDB.

For additional safety, set data protection to permanent so that data on the media is not overwritten unintentionally.



Split mirror concepts

Running Environment



3. Un-freeze the disks (Take DB/Appl out of backup mode/online)4. Perform backup from the mirror



5. Resynchronize the mirror





Local mirror - dual host (full performance, zero downtime backup)



Benefits

- True on-line backup for Oracle and SAP
- No performance impact on Application/Database during backup
- Maximizing uptime of business critical applications.
- Fast Recovery of online backup
 - small amount of archive logs generated due to short backup mode time
- Automated full integrated solution
 - integrated with Oracle RMAN
 - integrated with SAP brbackup



Snapshot concepts

Overview

The rapidly expanding requirement for high availability storage configurations has led to the introduction of new zero downtime backup (ZDB) technologies. The advances in storage virtualization technology have provided the opportunity for an alternative to conventional split mirror technology.

Within the Data Protector ZDB solution, different disk array technologies are combined with the latest developments in the snapshot technology, to create snapshots of application or database data stored on a disk array. These snapshots can subsequently be kept on a disk array as point-in-time copies of the original data for instant recovery purposes or can be used to produce ZDB-to-tape sessions on a backup system.

Storage virtualization

The term "storage virtualization" is used to describe the technology that separates the logical representation of storage from the actual physical storage components. This means the creation of logical volumes out of a pool of physical disks residing in a disk array. A logical volume is limited by the boundaries of the pool, but may span over any number of physical disks within the disk array. Logical volumes can be presented to one or multiple host systems. You cannot have control over the exact allocation of logical volumes on physical disks, but you can influence it with a choice of protection characteristics.



Snapshot backup





Multiple disk arrays - dual host





Multiple application hosts - single backup host





Disk array(s) - single host





Microsoft Volume Shadow Copy Service

Overview

A traditional backup process is based on the direct communication between the backup application (application, which initiates and performs backup) and an application to be backed up. This backup method requires from the backup application an individual interface for each application it backs up.

The number of applications on the market is constantly increasing. The necessity of handling application specific features can cause difficulties in backup, restore, and storage activities. An effective solution to this problem is introducing a coordinator among the actors of the backup and restore process.

VSS

Volume Shadow Copy Service (VSS) is a software service introduced by Microsoft on Windows operating systems. This service collaborates with the backup application, applications to be backed up, shadow copy providers, and the operating system kernel to implement the management of volume shadow copies and shadow copy sets. The idea of the Volume Shadow Copy Service is to provide a unified communication interface that can coordinate backup and restore of any application regardless of their specific features. With this approach, a backup application does not need to handle each application to be backed up specifically. However, this approach is applicable to a backup application only in case it conforms to the VSS specification.



Microsoft Volume Shadow Copy Service

What is a shadow copy?

A **shadow copy** refers to a volume that represents a duplicate of the original volume at a particular moment in time. The volume shadow copy technology provides a copy of the original volume at a certain point in time. The data is then backed up from the shadow copy, not from the original volume. The original volume continues to change as the backup process continues, but the shadow copy of the volume remains constant.

Shadow copy is basically a snapshot backup, which allows applications and users to continue writing to data volumes, even if they are in the middle of a backup process, while the backup is getting data from a shadow copy of the original volume. A shadow copy set is a collection of shadow copies created in the same point in time.

What is a writer?

A **writer** refers to any process that initiates change of data on the original volume. Writers are typically applications (for example, MSDE Writer for MS SQL Server) or system services (for example, System Writer and Registry Writer) that write persistent information on a volume. Writers participate in the shadow copy synchronization process by assuring data consistency.



Microsoft Volume Shadow Copy Service

What is a shadow copy provider?

A **shadow copy provider** to some entity that performs the work involved in creating and representing the volume shadow copies. Shadow copy providers own the shadow copy data and expose the shadow copies. Shadow copy providers can be software (including a system provider, MS Software Shadow Copy Provider) or hardware (local disks, disk arrays).

The example of the hardware provider is disk array, which has its hardware mechanism of providing point-in-time state of a disk. A software provider operates on physical disks and uses software mechanism for providing point-in-time state on a disk. The system provider, MS Software Shadow Copy Provider, is a software mechanism, which has been a part of Windows operating systems starting with Windows Server 2003.

The VSS mechanism guarantees that all hardware providers will be offered for creating shadow copy before all software providers. If none of them is able to create a shadow copy, VSS will use the MS Software Shadow Copy Provider for the shadow copy creation, which is always available.

Data Protector and VSS

The Volume Shadow Copy Service enables coordination among the backup application, writers, and shadow copy providers during the backup and restore process.



Actors of the traditional backup model





Actors of the VSS backup model





VSS benefits

VSS benefits

The advantages of using Volume Shadow Copy Service are as follows

- A unified backup interface for all writers.
- A unified backup interface for all shadow copy providers.
- Writers provide data integrity at application level. Intervention from the backup application is unnecessary.

Data Protector supports the Microsoft Volume Shadow Copy Service at two levels:

- Within the Microsoft Volume Shadow Copy Service integration, Data Protector provides a shadow copy backup and restore of VSS-aware writers, including ZDB and instant recovery functionality.
- Within the Disk Agent functionality, Data Protector provides VSS filesystem backup.

The Data Protector VSS integration supports a consistent shadow copy backup only for VSS-aware writers. Consistency in this case is provided by the writer. Whenever applications are not VSS-aware, a shadow copy is created. The consistency of the shadow copy data is not guaranteed at application level, however, it is improved in comparison to a non-VSS filesystem backup.



Benefits of using VSS

The table below outlines the differences between using Data Protector VSS integration backup, VSS filesystem backup; and non-VSS filesystem backup:

	Data Protector VSS integration backup	VSS filesystem backup	Non-VSS filesystem backup
Open files	No open files.	No open files.	If files are open, backup may fail.
Locked files	No locked files.	No locked files.	If files are locked, backup skips them.
Data integrity	Provided by the writer.	Crash-consistent state (in the event of a power failure, for example).	None (inherent).



Data Protector Volume Shadow Copy integration

The Data Protector integration with Microsoft Volume Shadow Copy service provides full support for VSS-aware writers. This includes automatic detection of VSS-aware writers, and backup and restore functionality.

VSS backup - In case of VSS-aware writers' backup, the consistency of data is provided at writer level and does not depend on the backup application. Data Protector follows the requirements provided by the writers when selecting what to back up. During the backup of VSS-aware writers, Data Protector does not communicate with each writer individually, but through the VSS interface. It uses the VSS integration agent to connect the Volume Shadow Copy Service, which coordinates the backup process. VSS provides Data Protector with the writer-related metadata necessary for performing a consistent backup and restore. Data Protector examines this data and identifies the volumes to be backed up. Data Protector then requests VSS to create a shadow copy of the specified volumes.

NOTE: A **Writer Metadata Document** (WMD) is metadata provided by each writer. Writers identify themselves by the metadata and instruct the backup application what to back up and how to restore the data. Data Protector therefore follows the requirements provided by the writer when selecting the volumes to be backed up and the restore method.

Volume Shadow Copy Service synchronizes the writers and providers. After a backup shadow copy is created, VSS communicates this information to Data Protector. Data Protector performs a backup from the shadow copy volume to the media and then notifies VSS that the shadow copy can be released.



Data Protector Volume Shadow Copy integration

VSS restore

VSS integration restore refers to the restore of data which was backed up using the Volume Shadow Copy Service and a writer. During the restore procedure, Volume Shadow Copy Service coordinates communication between Data Protector and the writers.

When restoring VSS-aware writers, Data Protector first restores all the relevant metadata to identify the backup components and to determine the restore method. It then connects to the Volume Shadow Copy Service and declares that the restore is about to begin. VSS coordinates the writers' activities during the restore. After Data Protector has successfully restored the data, VSS informs the writers that the restore has been completed and the writers can access the restored data and start their internal processing.



VSS filesystem and disk image backup and restore

Some applications are not aware of the Volume Shadow Copy Service. Such applications cannot guarantee consistency of data during the creation of a shadow copy. The VSS mechanism cannot coordinate the activities of these applications in order to perform a consistent backup.

However, you can still benefit from the VSS functionality. The cooperation between the backup application and a shadow copy provider can be still used to assure a higher level of data consistency. Microsoft calls this state of data consistency "crash-consistent state". This means that the VSS mechanism commits all pending I/O operations and holds incoming writing requests during the preparation of a shadow copy volume. In this way, all files on the filesystem are closed and unlocked when the shadow copy is being created.

Microsoft Volume Shadow Copy functionality allows the creation of a volume shadow copy without the participation of the applications being backed up. In this case, the shadow copy volume is created and then backed up by Data Protector. This approach can be used with applications that are not aware of the VSS mechanism.

IMPORTANT: When applications that are not aware of the VSS mechanism are being backed up, data consistency from the applications' point of view cannot be guaranteed. Data consistency is the same as in the event of a power failure. Data Protector cannot guarantee any data consistency when applications are not actively participating in the creation of a shadow copy.



VSS filesystem and disk image backup and restore

The consistency of data in a VSS filesystem and disk image backup is improved in comparison to a non-VSS backup. VSS allows you to create shadow copy backups of volumes and exact point-in-time copies of files, including all open files. For example, databases that are held open exclusively and files that are open due to operator or system activity are backed up during a VSS filesystem or disk image backup. In this way, files that have changed during the backup procedure are copied correctly.

The advantages of VSS filesystem and disk image backup are as follows:

- A computer can be backed up while applications and services are running. Therefore, applications can continue to write data to the volume during a backup.
- Files that are open are no longer skipped during the backup process because they appear closed on the shadow copy volume at the time of the creation of the shadow copy.
- Backups can be performed at any time without locking out users.
- There is little or no impact on the performance of the application system during the backup process.



About StoreOnce software deduplication

Data Protector's StoreOnce software duplication offers a software-based, block-level deduplication solution. When using StoreOnce software deduplication, note the following:

- Deduplication backs up to disk-based devices only. It cannot be used with removable media such as tape drives or libraries.
- Because Data Protector uses a software-only approach to deduplication (that is, when using StoreOnce software deduplication), no specific hardware is required other than standard hard disks to store the backed up data.
- In the deduplication process, duplicate data is deleted, leaving only one copy of the data to be stored, along with reference links to the unique copy. Deduplication is able to reduce the required storage capacity since only the unique data is stored.
- StoreOnce software deduplication uses hash-based chunking technology to split the data stream into sizeable chunks of data.
- Specifying a Backup To Disk device with the StoreOnce Software deduplication interface in the backup specification tells Data Protector to do a deduplication-type backup.



When to use deduplication

Typically, you would use a B2D device with data deduplication support when backing up an e-mail filesystem which may contain 100 instances of the same 1 MB graphic file attachment. If the system is backed up using a conventional backup technique, all 100 instances of the attachment are backed up. This requires approximately 100 MB of storage space. However, if the backup is done through a B2D device deduplication support, only one instance of the attachment is actually stored. All other instances are referenced to the unique stored copy. In this example, the deduplication ratio is approximately 100 to 1. Although this example is referred to as file-level deduplication, it serves to demonstrate the benefits of B2D devices and deduplication.

Other points to consider when deciding to use deduplication technology:

- Some data is not a good deduplication candidate! Data that is automatically created by a computer does not deduplicate well, for example, database files. Photos, video, audio, imaging, seismic data are all examples of data that do not deduplicate very well.
- Do not compress data before deduplicating it. It will impact on the deduplication and is unnecessary as compression is done following deduplication.
- Do not encrypt data before deduplicating it. This produces a deduplication ratio 1:1, basically, no deduplication.


Advantages of B2D devices and deduplication

Generally, data deduplication increases the speed of the backup service as a whole and reduces overall storage costs. Data deduplication significantly reduces the amount of required disk storage space. Because data deduplication is a disk-based system, restore service levels are significantly higher and tape (or other media) handling errors are reduced. Additional benefits of deduplication include:

- Data deduplication is more appropriate with large volumes of data.
- Data Protector uses well-proven deduplication algorithms to guarantee data integrity (StoreOnce software deduplication uses deduplication technology developed by HP Labs for HP StoreOnce Backup Systems. These systems use hardware-based deduplication.
- Disk-to-disk (D2D) storage with deduplication is rapidly becoming the preferred method for backup and recovery in both local and remote applications.
- The total cost of recovery for duplication-enabled D2D systems is significantly lower than with tape-based systems. Data deduplication backups can provide considerable capacity and cost savings compared to conventional disk backup technologies.



Deduplication performance

There are many factors that can affect deduplication performance. These include hardware and network speed, how the storage disk is set up, the size of the store, the deduplication ratio of the data, and how many concurrent backups are running. Using multiple streams can significantly improve backup performance. The number of parallel streams reading and writing data to a store is limited by the target device.

How deduplication integrates with Data Protector

Data Protector supports various deduplication setups:

- Source-side deduplication data is deduplicated at the source (the backed up system).
- Server-side deduplication data is deduplicated on the Media Agent system (the gateway).
- Target-side deduplication data is deduplicated on the target device (StoreOnce Backup system or StoreOnce Software system).





How deduplication integrates with Data Protector





Source-side deduplication

With source-side deduplication (1), a Media Agent is installed together with the Disk Agent on the client that is backed up and thus the client becomes a gateway (a source-side gateway). The deduplication is performed by the Media Agent on the client itself so only deduplicated data is sent to the target device, thereby reducing the overall network traffic. The number of concurrent streams is limited by load balancing settings. Once a Media Agent finishes the backup of local objects, a new Media Agent is started on the next client system.

Note that the backed up system must support deduplication (64-bit Windows systems or 64-bit Linux systems only, for details, see the support matrices).



Server-side deduplication

With server-side deduplication (2), deduplication is performed on a separate Media Agent client (a gateway) by the Media Agent. This reduces the load on the backed up system and on the target device, but does not reduce the amount of network traffic between the Disk Agent and Media Agent.

Note that the Media Agent client must support deduplication (64-bit Windows systems or 64-bit Linux systems only, for details, see the support matrices). Server-side deduplication enables you to deduplicate data from clients on which deduplication is not supported locally.



Target-side deduplication

The deduplication process takes place on the target device (3). It receives data to be backed from Media Agents installed on clients (gateways).

Target-side deduplication using the StoreOnce Software system

The StoreOnce Software deduplication system then writes the deduplicated data to the StoreOnce library (this is the physical store and is sometimes referred to as the deduplication store).

The StoreOnce software deduplication system allows connections from several Media Agents, locally or remotely. It also provides synchronization mechanisms to enable multiple Media Agents to work with the StoreOnce library at the same time. The Media Agent reads or writes data in terms of object versions to or from the StoreOnce library. Each object version is represented as an item in the StoreOnce library. To optimize deduplication performance, Disk Agent concurrency is not supported (this means, one Disk Agent talks to one Media Agent – there is no multiplexing of streams).

Target-side deduplication using the StoreOnce backup system device

The deduplication process looks from the Data Protector perspective very similar to target-side deduplication using the StoreOnce software system. However, there is no separate StoreOnce software deduplication system and the deduplication takes place on the StoreOnce Backup system device itself.



Data Protector backup solutions for VMware

Feature		Filesystem backup	VMware (Legacy) integration	Virtual Environment integration
Online backup			\checkmark	\checkmark
Crash-consistent backup		\checkmark	\checkmark	\checkmark
Application-consistent backup				\checkmark
Granularity		File	VM disk	VM disk
	Full	\checkmark	\checkmark	\checkmark
Backup types	Incremental	\checkmark	\checkmark	\checkmark
	Differential	\checkmark	\checkmark	\checkmark
Is Changed Block Tracking supported?				\checkmark



Data Protector backup solutions for VMware

Backup types

	Full	Backs up the complete virtual machine (disk), including the virtual machine memory file (if specified).		
Incr Backs up the changes made to a		Backs up the changes made to a virtual machine (disk) since the last Full, Incremental, or Differential backup.		
	Differential	Backs up the changes made to a virtual machine (disk) since the last Full backup.		

Snapshot management

Changed Block Tracking	VM Snapshot handling mode	Number of snapshots remaining after backup
	Disabled	0
disabled	Single	1
	Mixed	up to 2
enabled	Disabled	0
	Single	0
	Mixed	0



Data Protector backup solutions for VMware



Data Protector backup solutions for VMware

Full backup (disabled mode)



1. All DP snapshots (if they exist) are removed machine base or to the latest non-Data Protector

Data Protector backup solutions for VMware

Full backup (disabled mode)



1. All DP snapshots (if they exist) are removed DP snapshots are committed to the virtual machine base or to the latest non-Data Protector

4. DP snapshot1 is removed (the active state

Data Protector backup solutions for Microsoft Hyper-V

Feature		Filesystem backup	Volume Shadow Copy Service integration	Virtual Environment integration
Online backup			\checkmark	\checkmark
Crash-consistent backup		\checkmark	\checkmark	\checkmark
Application-consistent backup			\checkmark	\checkmark
Granularity		File	Virtual machine	Virtual machine
	Full	\checkmark	\checkmark	\checkmark
Backup types	Incremental	\checkmark		
	Differential	\checkmark		
Is virtual machine migration within a Microsoft Hyper-V cluster supported?				\checkmark



Hyper-V Image method (VSS transportable)











