OpenVMS Technical Journal V19

Backup management with D2D for HP OpenVMS

Table of contents

Overview

HP StoreOnce D2D Backup Systems is a disk-based storage appliance that emulates tape devices, which can be used to back up host network servers or PCs. This technical journal describes the difference between the traditional tape backups and modern D2D backups and also covers a case study of the issues faced by the customer while performing the HP OpenVMS backup on D2D device. It also covers few OpenVMS backup performance considerations for D2D devices.

Introduction

Backups are critical for any operating system and HP OpenVMS is no exception. One of the constant challenges faced by the system administrators for the backup is the exponential growth of the data. They often treat storage capacity as a limitless resource and make little or no effort to conserve storage space. If capacity starts to run low, they will typically bring more capacity online rather than asking users to conserve storage capacity. Data storage demands are being driven at a high rate across the business making it essential for system administrators to manage existing resources more efficiently. Simply investing more on storage hardware to solve the problem is no longer an option as business must strictly control IT expenditure to remain competitive.

The D2D devices have been designed to manage the storage space more efficiently. D2D device supports data deduplication where duplicate copies of data are replaced in the backup storage by a reference to a single instance of the data resulting in preserving space for future usage.

The following sections describe D2D device and the difference between traditional tape backup and D2D backup. If you are familiar with these topics already, then skip to the case study section.

What is a D2D device?

As the volume of stored business critical data continues to increase, a larger storage capability is required. Hence it is required to migrate the backup management to new disk-based data protection technology to address the challenges of creating a fast, user-friendly, reliable, secure, and scalable disk backup strategy. These new products and technologies make the backup management easier than the traditional tape backups.

The D2D device is a disk-based storage device designed for businesses to eliminate the need to manage several standalone tape drives and wants to consolidate, automate, and simplify the whole backup process. The D2D device emulates tape drives using low cost, high-capacity SATA-based disk drives. The key benefit of this approach is that the virtualized tape devices appear to the backup applications as though they were real physical devices directly attached to the server. This allows the D2D device to integrate seamlessly into an existing backup strategy while providing a high level of backup consolidation and automation, all in a single device.

Traditional tape backup vs. D2D backup

The tape-based backup solutions have been the default medium for offsite storage. Tapes are used in corporate data centers for over last 50 years, for serving as the primary data backup-and-restore media. However increasing data storage needs, shrinking backup windows, the need to recover critical applications quickly, and declining disk costs have combined to push tape out of the spotlight in favor of industry-standard SATA (Serial ATA) disks.

For backups, disk-based solutions like D2D have been accepted as the optimal alternative media. Backing up directly to disk will speed up the backup process and also significantly reduce restore times for individual files. For disaster recovery planning, it is also widely accepted that data should be stored offline in some form of removable media. For most businesses, their archiving strategy consists of moving data from disk to tape at some point. This backup strategy is commonly known as disk to disk to tape or Disk2Disk2Tape.

Advantages of D2D technology over tape technology

1. Faster restores:
   The random access nature of disk drives enables the instant merging of incremental backups. Compared with tapes, multiple incremental backups are often spread across multiple tapes and likely to be far from the start of the tape.
2. Reduced backup windows:
   Disk-based backups reduce the backup windows in two ways.
   - Random access and higher reliability of the disk media.
   - Multiple clients can be backed up simultaneously with disk-based backup which results in higher throughput.
3. Accelerated backup and archiving:
   Another benefit of disk-based backups is performance. Disk-based solutions leverage RAID performance and reliability to quickly backup and restore data.
4. Greater reliability:
   Disk-based solutions are more reliable as they make use of the RAID to protect against disk failures. Disk drives are less sensitive to environmental conditions, so they can be reliably used for a longer period of time.
5. Easy to use:
   Disk-based solutions are easier to manage overall. The interfaces and concepts of file system, disk, and network attached storage (NAS) are familiar to all, meaning less training and fewer mistakes.

Case study

Here is the case study of the OpenVMS customer who replaced their traditional tape backups with D2D backup.

Scenario:

The customer had automated their backup process. This automated process includes two operations, first, initializing the free and expired tapes using INITIALIZE command and second, performing the scheduled backups on the initialized tapes.

Requirements:

The typical size of data processed by the customer is around 10–12 TB daily. Hence they have created the virtual cartridges of capacity around 1.6 TB. This is one of the requirements of the customer to move from their traditional tape backup to D2D device.

Implementation and problem faced:

The customer had migrated from traditional tape devices to D2D device and automated their backup process. Automation had two parts:

- Initializing free and expired virtual cartridges
- Performing backup operation

Initialization operation had been given less priority by the customer when compared with backup operation as backups were critical. As a result, the free and expired virtual cartridges were getting accumulated in the initialization queue. Since the sizes of the virtual cartridges were around 1.6 TB, initializing these cartridges using INITIALIZE command takes a long time, resulting in increased total backup window affecting the backup process. The customer wanted to reduce their total backup window and required an alternate solution in OpenVMS to initialize the virtual cartridges including the spanned virtual cartridges.

OpenVMS solution:

The customer was suggested to use the OpenVMS backup utility for initializing the tape as an alternate solution. Instead of using the INITIALIZE command and then performing a backup operation, BACKUP itself can be used to initialize the tape and perform the backup operation. Backup qualifiers /REWIND and /LABEL can be used to initialize the tape. The /REWIND qualifier rewinds and initializes the volume. The /LABEL qualifier allows you to specify the volume label. During the tape span over, BACKUP initializes the second and subsequent tapes.

Refer Appendix A for more details on backup's tape label processing.

Additional performance considerations for D2D devices

OpenVMS backup generates two types of specially formatted data blocks. The first type contains pure data, which includes user data to be backed up and CRC for error DETECTION.

The second type of data block contains redundancy group data used for error CORRECTION.

Larger the data block size, greater the speed:

The size of each data block is determined by the /BLOCK_SIZE qualifier. The use of large blocks can greatly speed up backup operations and allow more data to be put onto a single reel of tape. However, when using low-density tape devices, care must be taken. Whereas the virtual tapes created in the D2D device are high-density tape devices, increasing the block size speeds up data written to tape and increases disk performance.

The default block size for tape is 8192 bytes and the maximum value that can be specified is 65535 bytes. Use /BLOCK_SIZE=65535 to improve the performance of the backup operation with D2D devices.

Impact of CRC and ECC on performance:

By default, OpenVMS backup does both error DETECTION and error CORRECTION. This is because OpenVMS backup must work on a variety of tape and disk devices and yet still maintain data integrity.

For error DETECTION, backup calculates a CRC value, or “checksum” on each data block. When restoring the saveset, a CRC for the data block is recalculated and compared with the CRC stored within the data block. If the two values do not match, BACKUP knows that the data block is bad. The /CRC qualifier controls the use of this feature. /NOCRC turns off the feature.

If the tape device that does hardware level error detection (check with the manufacturer, most D2D device does), use the /NOCRC qualifier. This will eliminate the overhead of both backup and the tape device calculating CRC values.

For error CORRECTION, backup uses redundancy group data. This data is used to reconstruct bad data blocks being restored. A setting of ten (by default) means that if any one data block in a group of ten data blocks restored is bad, the one data block can be reconstructed using the redundancy group data from the surrounding nine data blocks. The default setting of ten causes BACKUP to devote approximately over 17 percent of each data block to redundancy group data. The /GROUP=nn qualifier controls the use of this feature. /GROUP=0 turns off the feature.

If the tape device that does hardware level ECC (error correction code) (check with the manufacturer, most D2D device does), use the /GROUP=0 qualifier. This will reduce the overhead of having both backup and the tape device calculating and storing ECC data to the tape.

Note: /CRC and /GROUP qualifiers ensure the integrity of backup. HP recommends using these qualifiers if you want to achieve maximum data integrity.

Conclusion

To summarize, the traditional tape-based backup management can be migrated to the D2D backup system to create a fast, user-friendly, reliable, secure, and scalable backup strategy.

Appendix A

Below are few scenarios of tape label processing by backup utility.

1. If the tape label is not specified on the command line, backup uses the first six characters of the saveset name to create a label for the first volume (unless you use the /EXACT_ORDER qualifier, in which case backup preserves the volume label on the tape). For subsequent volumes, backup uses the first four characters from the label of the first volume plus the number of the volume in the sequence. For example, suppose you are saving files that require three tapes and the saveset name is BACKUP. If you do not specify a label, the first tape is labeled BACKUP, the second BACK02, and the third BACK03.
2. If a single label on the command line is specified using the /LABEL qualifier and it matches the label of the first volume, backup labels subsequent volumes with the first four characters of the label from the first volume plus the number of the volume in the sequence. For example, suppose you are saving files that require three tapes and the first tape is labeled TAPE. The second tape gets the label TAPE02, and the third tape gets the label TAPE03.
3. If multiple labels on the command line are specified using the /LABEL qualifier (without the /EXACT_ORDER qualifier), backup uses the labels you specify. If the operation requires more labels than you specified, backup uses the first four characters of the last volume label and the volume number of the tape.
4. The /EXACT_ORDER qualifier can be used in conjunction with the /LABEL qualifier to specify the order in which you want backup to use the labels. Backup continues the operation as long as the label of the tape in the drive matches the corresponding label on the command line. If you do not specify enough labels on the command line to complete the operation, backup prompts you to enter a label for the tape in the drive.

To safeguard against initializing or writing the wrong tape, backup compares the label that is specified on the command line to the label of the tape in the drive. Hence use the /LABEL and /EXACT_ORDER qualifiers to select the exact tape.

For more information

HP OpenVMS System Manager's Manual, Volume 1: Essentials
http://h71000.www7.hp.com/doc/82final/aa-pv5mj-tk/aa-pv5mj-tk.html

HP OpenVMS System Management Utilities Reference Manual
http://h71000.www7.hp.com/doc/84final/6048/6048pro.html

HP StoreOnce D2D Backup Systems—Overview
http://h18004.www1.hp.com/products/quickspecs/13218_div/13218_div.HTML

Support matrix for HP StorageWorks Tape Libraries with ABS and SLS
http://h71000.www7.hp.com/openvms/storage/smstape_matrix.html

Traditional tape-based backup management can be migrated to the D2D backup system to create secure and scalable backup strategy. For more information, visit
http://h18004.www1.hp.com/products/quickspecs/13218_div/13218_div.HTML

Share your feedback or queries on the VMS Technical Journal here.

Sign up for updates
hp.com/go/getupdated

Rate this document

4AA4-4793ENW, January 2013