Optimizing LTO Backup Performance

Consider software and hardware optimization for optimal backup performance

Dell Product Group | Dell PowerVault Tape Engineering
November 2015

A Dell White Paper

Revisions

Date	Description
November 2015	Initial version

Acknowledgements

Table of contents

Introduction

Since its inception, Linear Tape Open (LTO) technology has grown to become the dominant tape hardware backup solution. Each successive generation of LTO drives vastly increases the throughput and capacity of the previous generation. In order to achieve the greatest performance, customers must consider many software and hardware optimizations.

There are four main areas to consider when optimizing throughput on the LTO tape drive:

- Host"system"configuration
- LTO"tape"hardware"configuration
- Data"protection"software"configuration
- Data"set

This white paper describes these four factors and how they affect total performance. This paper also covers how to validate the tape drive performance using the IBM Tape Diagnostic Tool.

Host system configuration

Host system random access memory (RAM)

While the host system RAM may not have a direct effect on the data transfer performance of the tape drive, it is extremely important to note that larger system RAM can provide benefits in a data protection solution. Most data protection software applications run a SQL server instance to maintain information regarding the data protection environment, tape drive configuration, and various other components. As your data protection environment grows, increased RAM significantly aids in processing this information.

Additionally increasing the host system RAM is beneficial when using multiple agent and media servers in a data protection solution. These resources utilize the additional memory to process read and write requests on the master media server.

RAID configuration

Storing the data to be protected on a RAID disk configuration will provide a significant performance increase against a non-RAID setup as well as add additional redundancy. A RAID setup increases performance because it stripes data across multiple disks and allows reading and writing of data faster than a single hard drive can typically support. Adding more physical disks to the RAID array will also increase performance as it allows the host bus adapter (HBA) to take advantage of the additional blocks of data striped on the extra physical disks.

Hard drive spindle speed

Dell offers a variety of hard drive and solid-state drive (SSD) choices in PowerEdge™ servers as well as across Dell's storage portfolio. For enterprise applications, the hard drive spindle speeds range

from 7200 rpm to 15k rpm. There are many benefits to storing the data to be protected on the fastest rpm hard drive or SSD available. The faster the rpm, the faster the data can be read from the hard drive platters. This translates into an overall disk performance increase. The faster the data are read, the faster the tape drive can process it resulting in a higher throughput to the tape drive.

HBA firmware

As with the tape drive, it is always important to ensure the HBA has the most up-to-date firmware available. For a list of the latest HBA firmware, please refer to the Drivers & downloads section for your HBA at Dell.com/support.

It is important to consider the type of HBA to which your tape device is connected. Using an unsupported HBA can result in poor performance or result in the host system failing to detect the tape device. Dell qualifies its tape devices with a multitude of HBAs to ensure they are properly supported and provide optimal performance. For a complete list of compatible HBAs, please refer to the tape compatibility matrix at Dell.com/support.

Host system tape drive configurations

When connecting the tape drive or tape automation library to your host system, several configurations that could inhibit throughput.

- Connecting multiple tape drives to a single controller: Due to the amount of processing necessary to manage both devices, the controller's CPU may inhibit the tape drive from operating at full performance.
- Connecting tape drives and hard drives to a single controller: This is an unsupported configuration. Always attach tape drives and hard drives on separate controllers.
- Connecting multiple drives to a single host system: While there is no maximum specified number of tape drives per host system, it is very important to recognize the effect this will have on each tape drive's performance. For example, if you attach four tape drives and all four drives are streaming data, your disk subsystem will be constantly stressed to pump the necessary amount of data.

LTO tape hardware configuration

Firmware updates

Dell releases firmware updates and enhancements on a regular basis. These often improve the performance of the tape drive or automation device. Furthermore the latest firmware versions contain the most recent fixes and will make your LTO tape solution more robust. To get the latest firmware updates for your tape device, please visit Dell.com/support.

Cleaning

Routine maintenance is important to ensure continued optimal drive performance. A tape drive head can become dirty from particles from the tape media as well as dust in the operating environment. A dirty tape drive head increases error rates that cause the data protection software to perform retries during the read or write operations. These retries can significantly decrease tape drive performance.

Cleaning requests can occur at different intervals. The first is defined by the amount of tape drive usage while the second cleaning request can occur any time a threshold error rate is reached during a read or write operation. Dell PowerVault™ LTO drives and automation products notify users when cleaning is required. For more information on how to identify when your tape device requires cleaning, please refer to the appropriate product user guide at Dell.com/support.

Media

Tape media is a necessary component of a data protection solution and is often overlooked as an impediment to performance. Similar to the tape drive read/write head, tape media wears as it is used. If a tape cartridge is constantly used and not part of a media rotation, the potential to induce read or write errors during a backup operation is greatly increased. When the data protection software encounters read or write errors, it will perform retries on the media which decrease the backup performance.

Proper media handling and care are essential for optimal performance. Ensuring the media is stored securely and in the proper thermal and humidity environment is recommended. Dropping the cartridge against a hard surface, even from a short distance, can cause edge damage to the media inside. This can affect performance and if the damage is severe enough, cause data loss. The Dell PowerVault LTO media handbook provides detail on proper media handling procedures for all LTO media. The handbook is located at Dell.com/support under Manuals on each tape product page.

Data protection software configuration

Dell offers NetVault® Backup as our data protection software solution certified with Dell PowerVault LTO tape hardware. Other data protection software vendors support Dell PowerVault LTO tape hardware. Please refer to the software vendor website for supported products.

The performance differences between data protection software products can vary depending on configuration and settings. Selecting the right software solution should be based on the features offered to fit your data protection goals.

Software updates

Each data protection software vendor typically releases hot fixes and service packs several times a year that contain enhancements and fixes. These enhancements may provide improvements to the overall performance of the data protection solution. To ensure your software is at the latest version, please refer to the software vendor website for information on downloading the latest patches, service packs, or hot fixes.

Reported backup job performance

All data protection software report throughput to the user during the course of backup and restore operations. It is important to understand these calculations are not always a good indication of true hardware performance. These throughput calculations could be off by as much as 10 to 20% and are largely dependent on when the software starts its internal timer. The timer start can range from the submission of the backup job to the start of the first SCSI write command depending on the software product.

Block size

The block size in the data protection software determines how much data is streamed to the device during each write operation and is dependent on the block size the tape drive can support. Performance is largely affected by whether the data set being backed up is optimized for different block sizes. Most data protection software allows the user to modify the block size and the default is usually 64KB. Increasing the block size can result in a performance increase in certain scenarios depending on a variety of factors. Different models of HBAs may perform differently at varying block sizes depending on how each HBA is optimized. The data set can also impact the performance depending on the block size chosen.

There is no simple rule to follow when adjusting the block size. Perform several backup tests using various block sizes to determine which block size maximizes performance with your data set and HBA.

Buffer size

The tape drive must write data at a constant speed but the speed at which the data is received from the host can vary. The buffer size determines the size of the tape drive data buffer that is used to temporarily store the data in preparation for writing to tape. Some data protection software may allow modifying the buffer size. Please refer to your software vendor website to determine if buffer size modification is supported and for instructions to change the buffer size.

Increasing the buffer size is recommended to improve performance when backing up data sets consisting of many small files. For mixed data sets of varying sized files, the performance increase is generally less than 5%. It is important to note that increasing the buffer size often requires more CPU and memory utilization on the host system. As with block size, you should experiment with buffer size to determine the ideal size based on your typical data sets.

Data set

The largest contributor to the overall tape drive performance is the type of data being written to tape. There are several factors that can affect performance including the compressibility of the data, directory structure, and file size.

Compression ratio

The LTO specification for each drive generation defines a maximum native and compressed capacity for the tape media. LTO-5 and earlier generations are specified at a 2:1 compression ratio. LTO-6 and later generations are specified at a 2.5:1 compression ratio.

Unfortunately most data in data protection environments is not 2:1 or 2.5:1 compressible and thus will never take advantage of the data rate or capacity compression of the LTO drive. The compression ratio of every file is different. A simple way to gauge the compression of a particular file is to use a compression utility like Winzip to get a general idea of the compression ratio.

Compression is enabled by default on Dell PowerVault tape drives. As some data sets grow in size when run through a compression engine, the drive checks the post-compression data size and writes the smaller of the compressed or uncompressed data to the media. This allows for the smallest possible data size on the media regardless of the compressibility of the data.

Directory structure

The directory structure of the data is also a contributing factor to performance. Thousands of files buried several directories deep requires the data protection software to perform additional processing on all the file pointers to each file in each level. Even though this is a tiny amount of additional data processing per file, thousands of files structured in this manner can affect the throughput of the tape device.

File size

File size typically has the greatest effect on performance of all of the variables mentioned in this white paper. Small files can drive additional hard drive overhead when reading the files for the backup job. There is additional data protection software overhead incurred when the data protection software agent must open, read, and then close each file during the backup job. With data sets utilizing larger files, lower overhead can deliver a more consistent rate of performance throughout the entire backup set.

Drive performance validation

The IBM Tape Diagnostic Tool (ITDT) can be used to validate the performance of the tape drive. The procedure below sends test data directly from memory to the tape drive. This procedure removes any performance impact from the data protection software or disk configuration.

If the tape drive is performing as expected, the data rate during this test will be within 10% of the maximum data rate specified for the drive generation. Please refer to the product documentation located at Dell.com/support for the maximum specified data rate for the tape drive in your configuration.

Please note this test may take up to two hours to run depending on drive generation.

Download ITDT from Dell.com/support. ITDT is available at the Drivers & downloads tab on each product page.

From the ITDT CLI, follow these steps to validate the drive performance:

1. Select [S] to scan the bus.
2. Enter the drive number to select the drive to test.
3. Select [W] to enter the full write test.
4. Enter the number for transfer size desired. Values less than 64 should not be selected as performance may suffer.
5. Select [O]k.
6. Select [C] OR [I] for compressible or incompressible data.

Conclusion

There are many factors that affect performance of your data protection solution as this white paper demonstrates. There are many important decisions to make when designing the right tape solution to fit your needs. There is no single "golden rule" for maximizing performance for your tape solution, but by utilizing the appropriate best practices detailed in this paper, you can greatly increase your overall tape throughput. It is vitally important to remember that any performance increase will vary depending on the data set and overall hardware configuration.