An Oracle White Paper
September 2013

Maximizing Tape Performance with StorageTek T10000 Tape Drives

Introduction

Oracle’s StorageTek T10000C and StorageTek T10000D tape drives provide 252 MB/sec of native throughput, more than 1.5 times the performance delivered by the midrange LTO-6 tape drive. However, if tape storage applications are not designed to take full advantage of tape drive performance gains, the extra speed can be wasted. Host applications might be simply unable to match the throughput of today’s tape drives due to the application itself or the server hardware on which it operates. Or, the application may be designed to write discrete files to tape rather than continuous streams of data.

Both scenarios cause tape drives to start and stop frequently during data transfers, potentially increasing backup and tape migration times. Oracle understands the performance challenges that each of these scenarios creates when data is transferred to tape, so Oracle has included three important features in StorageTek T10000 tape drives to tackle the challenges head on.

Detrimental Effects on Tape Drive Performance

Transferring data to or from an application fast enough to ensure that a tape streams smoothly is critical to maximizing overall throughput. In a *streaming mode* scenario, where a host application is sending a continuous stream of data to the tape drive at a slower rate than tape drive is writing the data to tape, the tape drive must stop and wait frequently for its internal data buffer to fill before emptying the contents to tape again. The tape must be repositioned each time this occurs for the tape drive to begin writing at the location where the previous write operation stopped. This reposition is called a *backhitch* (also known as a *football* or *shoe shining*), and this mechanical operation can result in data transfers being held off, typically for three to five seconds during each backhitch, significantly decreasing overall throughput.

In addition to speed mismatches between the application and the tape drive while in streaming mode, there is also a *file mode* scenario in which the application sends discrete files, each separated by a *file sync* that causes the tape drive to backhitch, causing slower performance. Applications insert file syncs in the data stream as a means of forcing the tape drive to confirm the data that it just received from the application was written successfully to tape. To maximize tape drive performance, file syncs should be used as infrequently as possible, but that is not always the case depending on the application.

Maintaining High Performance Despite Low Data Transfer Rates

This article describes the following three key performance features of Oracle's StorageTek T10000 tape drives:

- 2 GB data buffer
- File Sync Accelerator
- Tape Application Accelerator

Benefits of a Large Buffer

A large data buffer can help a tape drive deliver multiple performance benefits. With its 2 GB data buffer, StorageTek T10000D utilizes a larger buffer than any other tape drive.

Eliminate Performance Drops

Common to all tape drives is an internal data buffer, which has a primary purpose of serving as a holding tank for data as it is transferred to and from the tape media. The benefit that the buffer can bring is that backhitch operations can occur with no negative impact on throughput, as seen by the application.

Typically, the size of the buffer is chosen so it contains at least the maximum amount of data that the drive would transfer during the time required for a backhitch operation. Using the StorageTek T10000D tape drive as an example, with throughput being 252 MB/sec and with roughly estimated backhitch time to be 4 seconds, a buffer size of 1,008 MB (252 x 4) or greater eliminates any drop in throughput due to backhitches.

However, the StorageTek T10000C and T10000D tape drives include a 2 GB buffer, more than double the size required for making backhitches transparent to the application. The extra buffer capacity dramatically improves the effectiveness of other performance enhancing features of StorageTek T10000 tape drives that are covered later in this paper.

Eliminate the Need for Speed Matching

One approach to the speed mismatch problem in the streaming mode scenario is to try to eliminate backhitch occurrences altogether by dynamically matching the speed of the tape drive with the speed of the application. In this approach, an algorithm in the tape drive measures the median data rate of the host application over some interval. Then, if necessary, the algorithm commands the drive to stop transferring data and switch to a new speed (resulting in a backhitch).

From a drive- and media-reliability perspective, there may be a benefit to reducing the number of backhitches that occur in this scenario. This is especially true for LTO tape drives, which are not designed to withstand the demands of enterprise data centers, and this approach explains why LTO drives utilize a number of different speeds in the streaming mode scenario.

The problem with speed matching schemes for streaming mode scenarios is that performance sometimes can be impacted negatively by dramatic changes in the rate at which data is being sent and by changes in the compression ratio of the data. In Figure 1 and Figure 2, the performance of two drive configurations is compared in streaming mode, with the assumption that the host transfer rate varies between two extremes. In both figures, the columns on the left provide the performance over time for a drive operating at a single speed of 250 MB/sec and possessing a 2 GB buffer. The columns on the right provide the performance for a drive operating at speeds ranging from 90 MB/sec to 250 MB/sec and possessing a 1 GB buffer. In this scenario, there are three key assumptions:

- The compression ratio of the host data stream remains constant.
- The backhitch time is 4 seconds.
- The speed calibration time is 5 seconds (applies to Drive B only).

Figure 1 shows the impact on drive throughput performance during the 15 seconds following a *decrease* in the speed of the host data stream from 270 MB/sec to 90 MB/sec. A backhitch occurs in both drive configurations, but the buffer in each drive prevents any drop in throughput. Even when an additional backhitch occurs, as is the case for the multispeed drive, Drive B’s buffer prevents any impact to performance.

		DRIVE A Buffer: 2 GB Speed: 250 MB/sec					DRIVE B Buffer: 1 GB Speeds: 90 to 250 MB/sec				
time (sec)	host transfer rate (MB/sec)	buffer fill rate (MB/sec)	buffer drain rate (MB/sec)	buffer level (%)	drive data rate as seen by the host (MB/sec)	running average of drive data rate as seen by the host (MB/sec)	buffer fill rate (MB/sec)	buffer drain rate (MB/sec)	buffer level (%)	drive data rate as seen by the host (MB/sec)	running average of drive data rate as seen by the host (MB/sec)
-2	**270**	270	250	70%	250		270	250	70%	250	
-1		270	250	71%	250		270	250	72%	250	
0		270	250	72%	250		270	250	74%	250	
1	**90**	90	250	64%	90	90	90	250	58%	90	90
2		90	250	56%	90	90	90	250	42%	90	90
3		90	250	48%	90	90	90	250	26%	90	90
4		90	250	40%	90	90	90	250	10%	90	90
5		90	250	32%	90	90	90	0	19%	90	90
6		90	250	24%	90	90	90	0	28%	90	90
7		90	250	16%	90	90	90	0	37%	90	90
8		90	250	8%	90	90	90	0	46%	90	90
9		90	250	0%	90	90	90	250	30%	90	90
10		90	0	5%	90	90	90	0	39%	90	90
11		90	0	9%	90	90	90	0	48%	90	90
12		90	0	14%	90	90	90	0	57%	90	90
13		90	0	18%	90	90	90	0	66%	90	90
14		90	250	10%	90	90	90	90	66%	90	90
15		90	250	2%	90	90	90	90	66%	90	90

Back-hitches occur when buffer empties, causing no impact to performance

Speed calibration period causes no impact to performance

Figure 1. No impact on drive throughput performance occurs following a *decrease* in the speed of the host data stream from 270 MB/sec to 90 MB/sec.

However, Figure 2 shows the effect on the two drive configurations when the host transfer rate *increases* from 90 MB/sec to 270 MB/sec. In this scenario, the single-speed drive instantaneously delivers its maximum throughput of 250 MB/sec, but the multispeed drive must continue to transfer data at the slower rate until its speed-matching algorithm determines a new speed at which the tape drive should operate. By the time the algorithm makes its decision and commands the drive to backhitch and change speeds, the 1 GB buffer has filled to capacity. Since there is no room in the buffer for additional data, the drive must hold off the host application while it completes the backhitch, further dropping the throughput, as seen by the host, from 90 MB/sec to 0 MB/sec during the four-second backhitch period. In the 15 seconds following a low-to-high change in host transfer rate, the single-speed drive with 2 GB buffer outperforms the multispeed drive with 1 GB buffer by 85 percent.

		DRIVE A Buffer: 2 GB Speed: 240 MB/sec					DRIVE B Buffer: 1 GB Speeds: 90 to 250 MB/sec				
time (sec)	host transfer rate (MB/sec)	buffer fill rate (MB/sec)	buffer drain rate (MB/sec)	buffer level (%)	drive data rate as seen by the host (MB/sec)	average data rate as seen by the host (MB/sec)	buffer fill rate (MB/s)	buffer drain rate (MB/sec)	buffer level (%)	drive data rate as seen by the host (MB/sec)	average data rate as seen by the host (MB/sec)
-2	90	90	250	10%	90		90	90	10%	90	
-1		90	250	2%	90		90	90	10%	90	
0		90	0	7%	90		90	90	10%	90	
1	270	270	0	20%	250	250	270	90	28%	90	90
2		270	0	34%	250	250	270	90	46%	90	90
3		270	0	47%	250	250	270	90	64%	90	90
4		270	250	48%	250	250	270	90	82%	90	90
5		270	250	49%	250	250	270	90	100%	90	90
6		270	250	50%	250	250	0	0	100%	0	75
7		270	250	51%	250	250	0	0	100%	0	64
8		270	250	52%	250	250	0	0	100%	0	56
9		270	250	53%	250	250	0	0	100%	0	50
10		270	250	54%	250	250	270	250	100%	250	70
11		270	250	55%	250	250	270	250	100%	250	86
12		270	250	56%	250	250	270	250	100%	250	100
13		270	250	57%	250	250	270	250	100%	250	112
14		270	250	58%	250	250	270	250	100%	250	121
15		270	250	59%	250	250	270	250	100%	250	130

Back-hitch occurs when buffer empties, causing no impact to performance.

Speed calibration occurs after incoming data rate increases. Drive continues to run at lower data rate during this period.

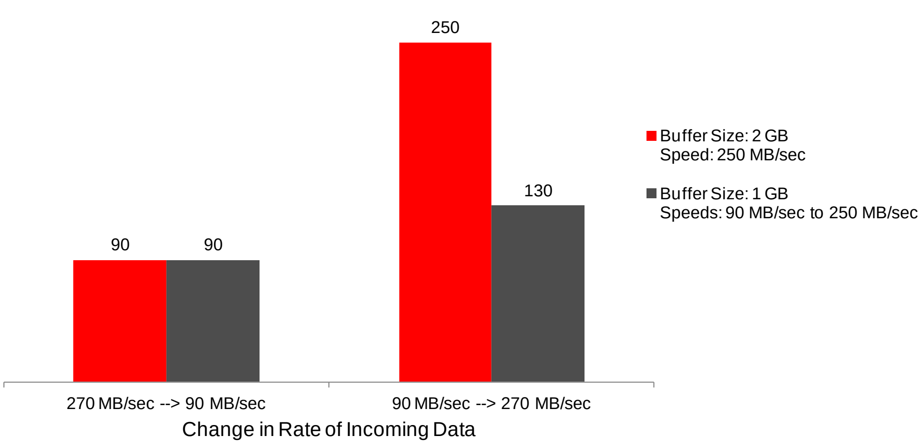
Back-hitch occurs when speed change is required. Since buffer is full, hold-off occurs, dropping performance to 0 MB/sec.

Figure 2. Negative impact on multispeed drive performance follows an *increase* in the speed of the host data stream from 90 MB/sec to 270 MB/sec. A 2 GB buffer and single high-speed operation provide 85 percent faster performance after a low-to-high change in host transfer rate.

Figure 3 summarizes the performance results of the two drive configurations for both the high-to-low and low-to-high host transfer rate scenarios. In the case of low-to-high host transfer rate changes, single-speed operation offers 85 percent faster performance than multispeed operation.

Average Performance as Seen by the Host (MB/sec)

(During 15-second Period after Change in Rate of Incoming Data*)



* changes in rate of incoming data may be due to changes in host application throughput and/or changes in compressibility of the data stream.

Figure 3. Single-speed operation with a larger buffer offers faster performance than multispeed operation.

Improving Performance When Writing Files

In the file mode scenario, sync commands such as *tape marks* or file syncs located between each file can have a significant detrimental impact on overall data throughput from the host application to the tape media. Every sync command requires the drive to empty the contents of its buffer to tape, provide confirmation to the host that the file was written successfully to tape, and then backhitch in preparation for receiving the next file in the sequence from the host application.

StorageTek T10000 tape drives offer two advanced features to improve throughput performance when writing files:

- File Sync Accelerator
- Tape Application Accelerator

File Sync Accelerator

File Sync Accelerator consists of a sophisticated algorithm and a collection of tools at the algorithm's disposal for automatically improving performance when writing files. Figure 4 is a block diagram that shows a list of incoming user data metrics that are continuously monitored by the algorithm so that it can intelligently choose which tool or combination of tools to utilize for optimizing file writing performance.[1]

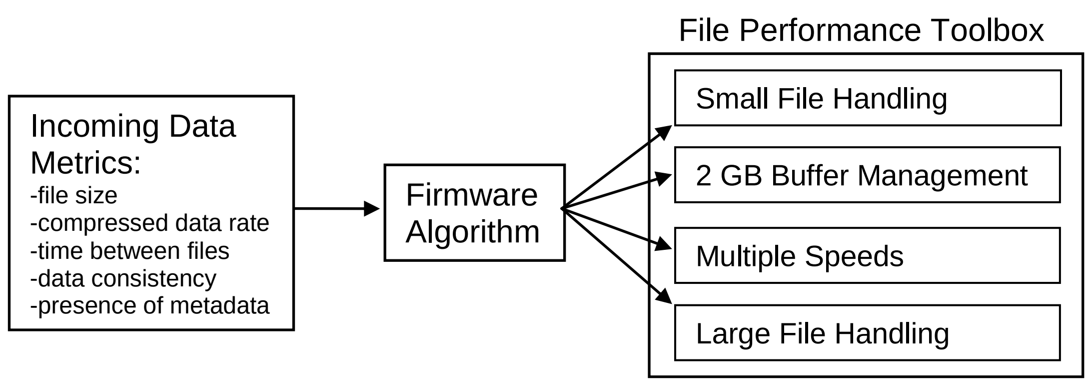


Figure 4. File Sync Accelerator monitors incoming data metrics before intelligently choosing which tool or combination of tools to utilize for optimizing file writing performance.

[1] Buffer management, multiple tape speeds, and large file handling capabilities are featured in the StorageTek T10000D File Sync Accelerator, but are not featured in the T10000C File Sync Accelerator.

Small File Handling

Small file handling is another tool available to File Sync Accelerator. It automatically optimizes file mode performance. Figure 5 illustrates how this two-step process works for writing many small files, less than 50 MB in size, to tape. The first step entails writing incoming data to two unused wraps as a temporary location. Instead of causing the tape drive to empty its buffer to tape and backhitch in anticipation of the next file, File Sync Accelerator detects the high frequency of sync commands and signals the drive to keep tape moving by writing the data "sparsely" to the two unused wraps. Sparsely written data is data that contains gaps of pad data between each file. The length of the pad data gaps varies proportionally with the wait time before the arrival of the next file. Data continues to be sparsely written to tape without stopping until one of two conditions is encountered:

- The tape drive's buffer is approximately half full.
- Half the length of tape has been traversed on the first of the two wraps designated for sparse writing.

When either of these two conditions is encountered, the tape stops briefly and begins writing in the same fashion in the opposite direction on the second unused wrap. Nominally, the buffer is approximately full when the tape drive reaches the location on the tape where data was being written prior to File Sync Accelerator enabling small file handling. Following a quick reposition, the second step is executed.

As shown in Figure 5, the second step entails continuously writing the entire contents of the 2 GB buffer (without progressive write gaps) to the wrap where the data originally was intended to reside.

During any File Sync Accelerator cycle, the incoming data stream is re-evaluated. If the tape drive determines that data still can be written to tape more efficiently with the small file handling tool, the tape drive continues operating in this way. Otherwise, the tape drive automatically switches back to the normal mode of operation.

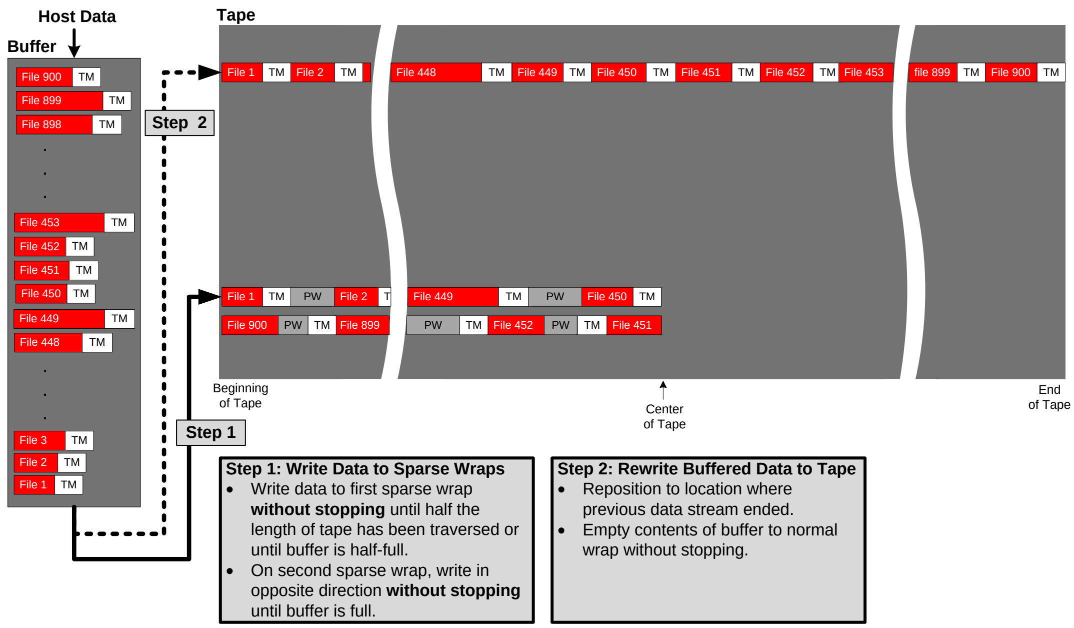


Figure 5. Small files (less than 50 MB) are handled using a two-step process that maximizes streaming performance by eliminating almost all of the backhitches that occur due to sync commands (e.g., tape marks) between files.

Figure 6 provides a performance comparison of StorageTek T10000D, StorageTek T10000C, IBM TS1140, LTO-6, and LTO-5 when writing 25 GB of 13 MB files to tape. A 2 GB buffer enables small file handling in the StorageTek T10000C and T10000D tape drives to deliver 27 percent faster performance than IBM's Virtual Backhitch feature, which is offered in the TS1140 tape drive. Compared to LTO-5 and LTO-6, the small file handling capability of StorageTek T10000C and T10000D tape drives delivers eight times better performance. Furthermore, small file handling in StorageTek T10000C and T10000D tape drives causes the tape drive to backhitch only 13 times, which is 0.6 percent of the 1,923 backhitches that occur when using LTO-5 or LTO-6 to write these files.

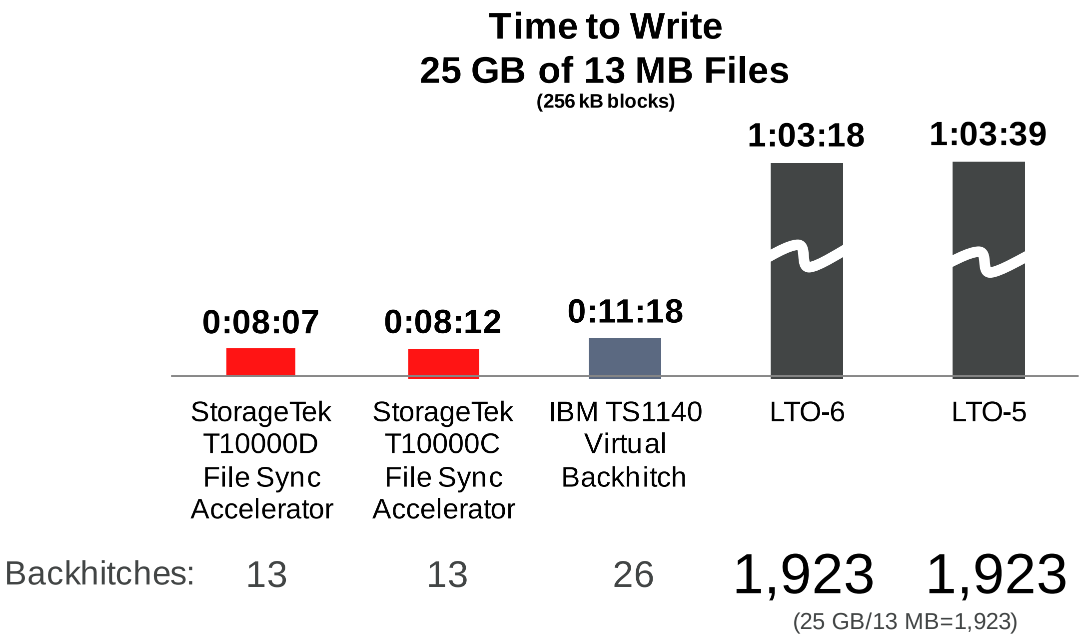


Figure 6. When small file handling is enabled by File Sync Accelerator, data is transferred eight times faster than with LTO and with 0.6 percent of the backhitches.

Buffer Management

File Sync Accelerator monitors the rate of incoming data and other factors to determine whether the 2 GB buffer management tool should be utilized to improve file mode performance. When utilized, the buffer management tool adjusts the data threshold (aka high-water mark) at which the tape drive begins to write data out of the buffer, emptying its contents to tape. For optimum performance, the goal is to minimize the amount of data residing in the buffer in anticipation of the next sync command so that less time (on the order of milliseconds) is required for the tape drive to write the buffer contents to tape before accepting the next file from the host. Without the buffer management tool, performance suffers in proportion to the data level in the buffer when a sync command is received. In this case, the tape drive will hold off the host for a longer period of time (multiple seconds) while writing the buffer contents to tape. When utilizing the buffer management tool, File Sync Accelerator typically will operate the tape drive at its maximum tape speed. However, as input conditions change, File Sync Accelerator may begin utilizing other available tape speeds to optimize performance.

Multiple Tape Speeds

File Sync Accelerator has multiple tape speeds at its disposal to optimize file mode performance. The primary goal is to choose a speed that enables the drive to deliver the highest possible performance in consideration of various factors including incoming file sizes, host rate, and other factors.

Large File Handling

Files that are larger than 50 MB are written using File Sync Accelerator's large file handling capability, which trades off a very small amount of tape capacity for faster performance. This capability works similarly to step 1 shown in the small file handling diagram in Figure 5. The difference is that the tape drive streams the data to its intended wrap without stopping, instead of using the buffer and sparse wraps to hold the data temporarily before streaming the data to tape. A small capacity tradeoff occurs when pad data is written to tape during the host time delay between files, consuming no more than 2 percent of total cartridge capacity by default and no more than 10 percent if the user chooses to select this larger tradeoff option using the Virtual Operator Panel (VOP) application when configuring the tape drive. Multiple speeds are utilized when large files are handled to minimize the capacity tradeoff. For example, a slower tape speed may be used if the delays between files are relatively long, and faster tape speeds may be used if the delays are shorter.

Figure 6 provides a performance comparison of StorageTek T10000D versus LTO-6 when writing 1 GB files. Ideally, the lines on this graph would indicate a linear 1-for-1 match of the tape drive data rate to the host data rate. In this example, the StorageTek T10000D uses *buffer management* to achieve linear performance at lower host data rates (less than 160 MB/sec) and utilizes *multiple speeds* with *large file handling* to achieve near-linear performance at higher host data rates (greater than 160 MB/sec) up to its maximum native throughput of 252 MB/sec. Note that LTO-6 does not utilize buffer management or multiple tape speeds when writing files and therefore cannot provide a linear 1-for-1 performance match. A maximum performance of 75 MB/sec is delivered by LTO-6 when writing 1 GB files, 53 percent lower than its advertised native throughput of 160 MB/sec.

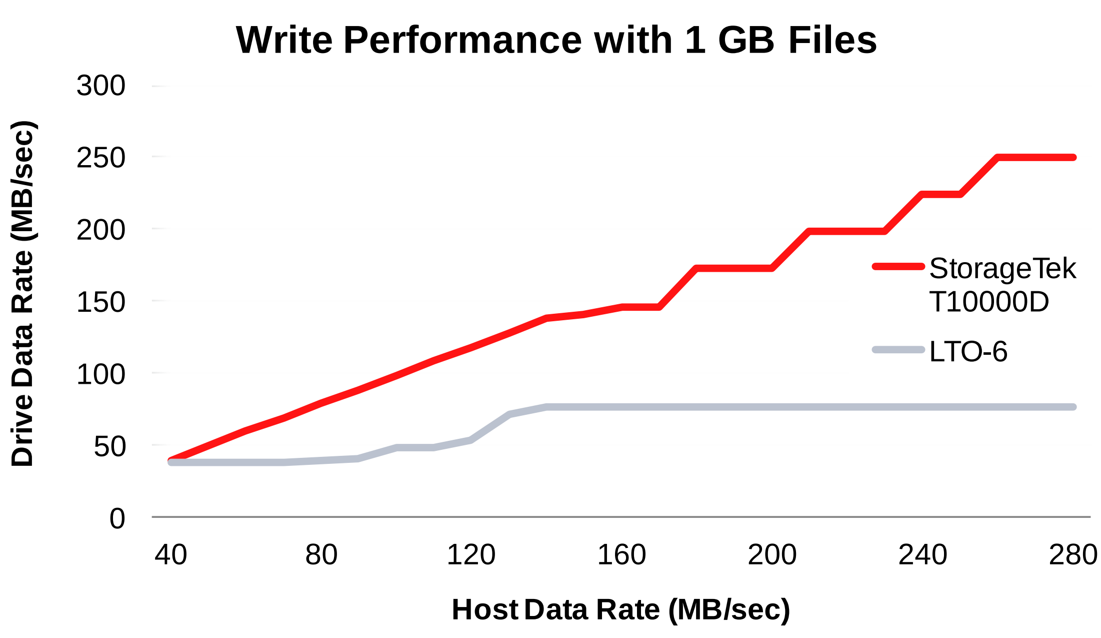


Figure 6. StorageTek T10000D tape drive delivers near-linear performance matching when writing files by utilizing buffer management and multiple tape speeds. LTO-6 does not offer these capabilities.

Tape Application Accelerator

Tape Application Accelerator is an innovative feature exclusively available in the StorageTek T10000 tape drives. It offers another means of increasing write throughput to tape despite an application's propensity for writing files rather than streaming data to the tape drive. This feature works similarly to File Sync Accelerator by enabling the tape drive to write data that contains a high frequency of sync commands without stopping the tape.

However, unlike File Sync Accelerator, Tape Application Accelerator does not cause the tape drive to write the data sparsely to a temporary location on tape before writing the data to its final location. Instead, data is written directly to its final location without any gaps of progressive write data.

Prior to enabling a Tape Application Accelerator, users must determine how their particular application utilizes tape marks and file syncs.

(Tape mark and file sync are defined differently in FICON and Fibre Channel environments. The glossary at the end of this document provides additional detail on this difference in nomenclature.)

When enabled in StorageTek T10000 tape drives, Tape Application Accelerator causes the tape drive to respond to tape mark and file sync commands differently than when disabled:

- A tape mark received by the tape drive is treated as a *buffered tape mark*.
- A file sync received by the tape drive is treated as a *no op* command.

Since buffered tape marks and no op commands do not cause the tape drive to empty the contents of its buffer to tape and backhitch, the data is written to tape in significantly less time.

Figure 7 illustrates how data is transferred to tape when Tape Application Accelerator is enabled. Data is written to tape without stopping, and the buffer is emptied to tape only if it is completely filled or after the application provides indication of the end of file (EOF) or end of volume (EOV). All tape marks or file syncs received prior to either of these indicators are treated as buffered tape marks or no op commands, respectively.

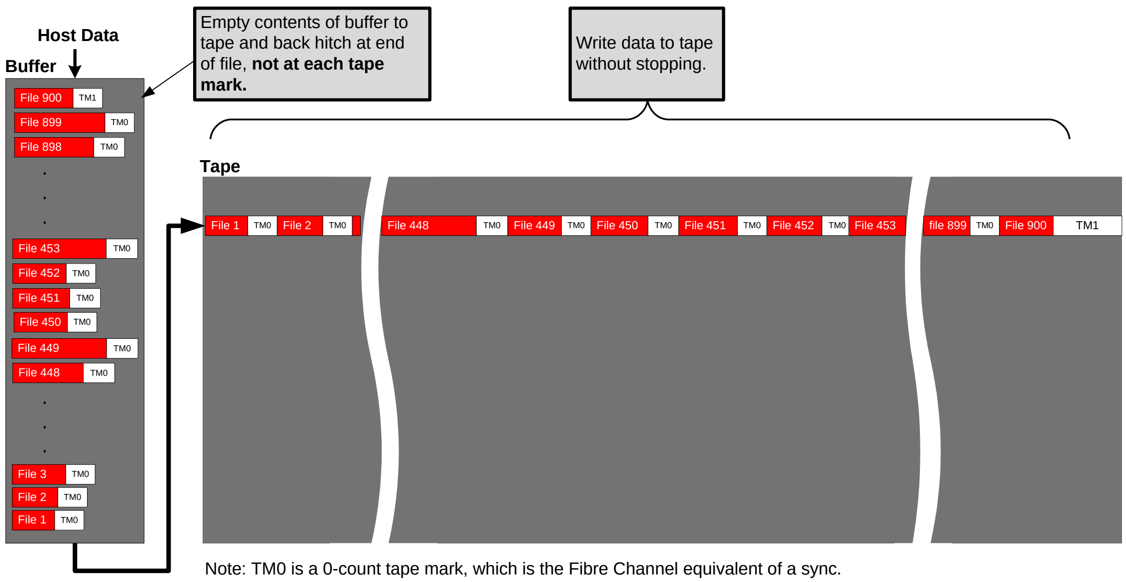


Figure 7. Tape Application Accelerator enables streaming of files (of any size) to tape without stopping.

In Figure 8, the small file performance of StorageTek T10000D tape drives with Tape Application Accelerator enabled is compared to performance with the small file handling capability of File Sync Accerator, IBM's Virtual Backhitch feature in the TS1140 tape drive, and LTO-6. Tape Application Accelerator enables StorageTek T10000D tape drive performance that is 6 times faster than IBM TS1140, 35 times faster than LTO, and requires zero backhitches.

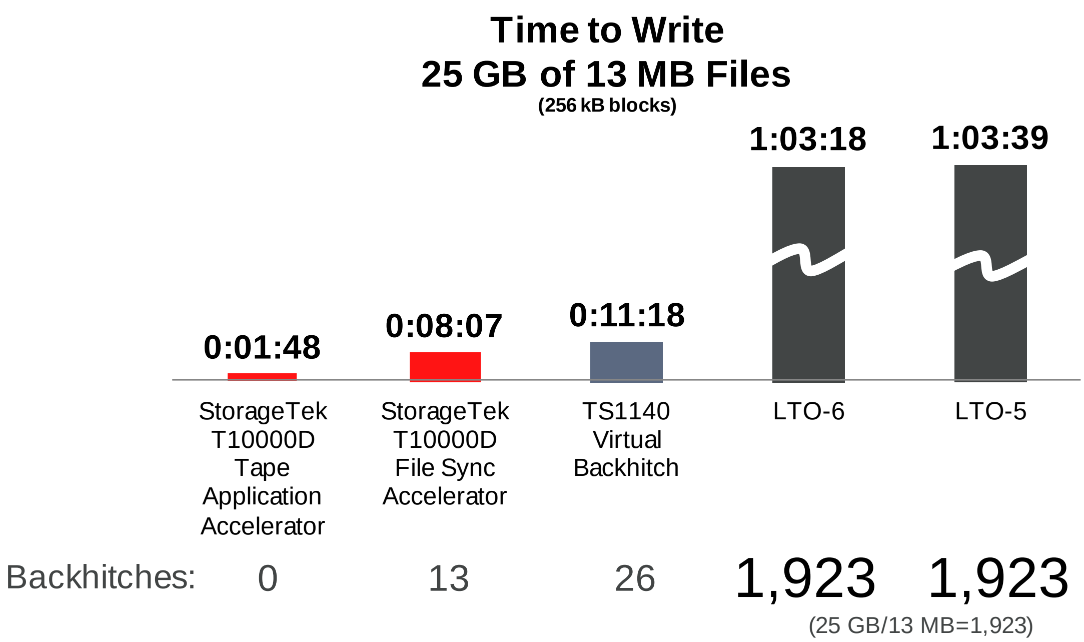


Figure 8. Tape Application Accelerator provides the fastest performance by streaming files to tape without stopping.

It is important to note that when Tape Application Accelerator is enabled, multiple files with tape marks or file syncs may reside in the 2 GB buffer of the StorageTek T10000 tape drive prior to being written to tape. Consequently, there is potential for data loss in the event of an individual tape drive power failure or reset. The best practice for minimizing this risk while maximiming performance depends on whether the StorageTek T10000 tape drive is operating in an MVS environment (FICON) or an open systems environment (Fibre Channel or FCoE).

Best Practice in MVS Environments

When Tape Application Accelerator is enabled in StorageTek T10000 FICON tape drives, file syncs are always converted to no op commands and tape marks are always treated as buffered tape marks. Therefore, when utilizing Tape Application Accelerator in MVS environments, *Oracle highly recommends duplexing output-type jobs to two tape drives*. Please see Appendix A for additional information. By default, this feature is disabled and can be enabled through the Virtual Operator Panel (VOP) application.

Best Practices in Open Systems Environments

In open systems environments, there is a choice between two best practices. Selection depends on whether the user's storage application automatically restarts a job following a power failure or reset event. Most, if not all, open systems storage applications are designed to restart a job immediately following a failure event. When using StorageTek T10000C or T10000D tape drives with applications that restart jobs after failure events, Oracle recommends Best Practice #1, as described below. When using StorageTek T10000C or T10000D tape drives with applications that *do not* restart jobs after failure events, Oracle recommends Best Practice #2, as described below.

Best Practice #1

When operating the StorageTek T10000C or T10000D tape drive with open systems applications that are designed to restart a job following a failure event, *Oracle highly recommends configuring Tape Application Accelerator to convert file syncs to no op commands WITHOUT treating tape marks as buffered tape marks*. In this mode, the tape drive will deliver faster performance by not stopping after writing the contents of the buffer to tape each time it receives a file sync command from the application. However, when the tape drive receives a tape mark command, it *will* stop after writing the contents of the buffer to tape. Risk of losing buffered data becomes negligible in this mode by taking advantage of the fact that any failed job will be restarted by the application and the tape mark command at the end of each job will cause the tape drive to immediately empty the contents of the buffer to tape before waiting for the next job.

Note:
The converse mode is also available for the unlikely possibility that an application inserts frequent tape mark commands within a single job before completing the job with a file sync command. In this mode, the StorageTek T10000C or T10000D tape drive will deliver faster performance by not stopping after writing the contents to the buffer each time it receives a tape mark command. However, when the tape drive receives a file sync command, it *will* stop after writing the contents of the buffer to tape. Risk of losing buffered data becomes negligible in this mode because any failed job will be restarted by the application, and the file sync command at the end of each job will cause the contents of the buffer to be written to tape before waiting for the next job.

Figure 9 provides a comparison of the time required to perform a 50 GB backup of a NetApp appliance via NDMP using a StorageTek T10000D tape drive versus LTO-6. Tape Application Accelerator—when configured according to Best Practice #1—enables up to 45 times faster backup performance than LTO-6.

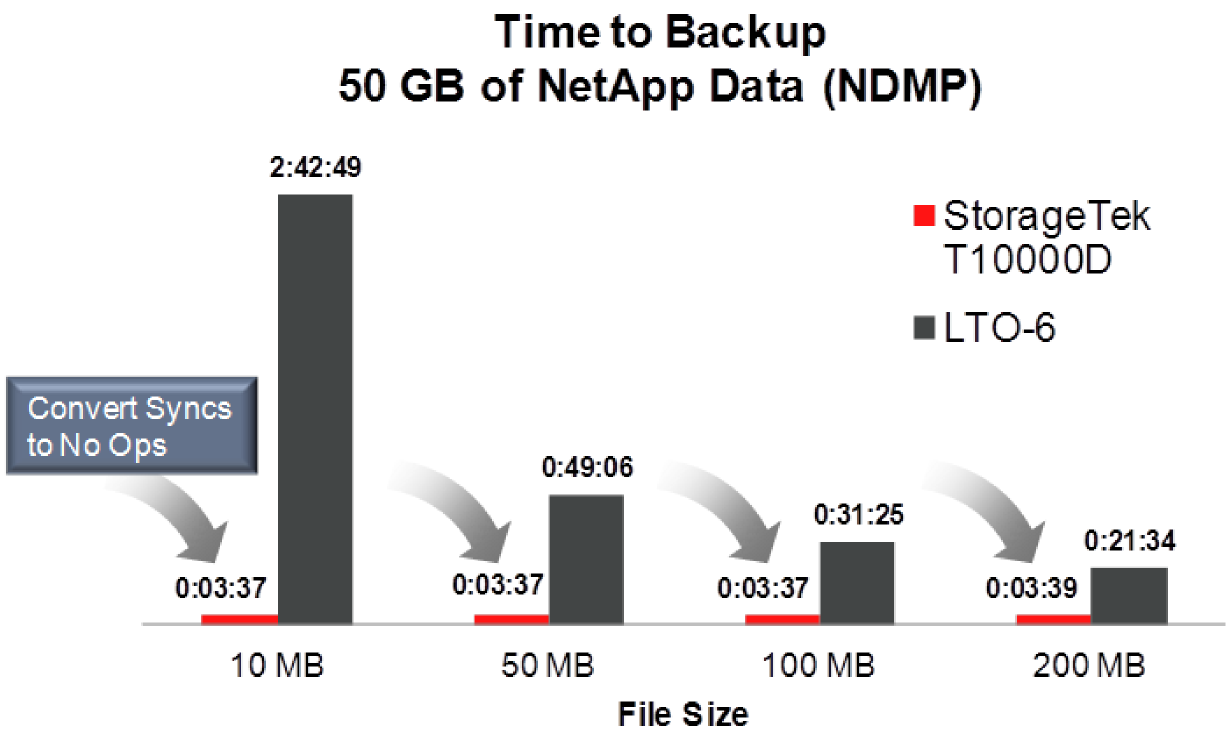


Figure 9. Tape Application Accelerator can reduce a 50 GB NetApp backup job from nearly three hours with LTO-6 to three and a half minutes with a StorageTek T10000D tape drive.

Best Practice #2

When operating the StorageTek T10000C or T10000D tape drive with open systems applications that are not designed to restart a job following a failure event, *Oracle highly recommends duplexing output-type jobs to two tape drives.* In this scenario, the fastest possible write performance is achieved by enabling Tape Application Accelerator from the VOP pull-down menu. In this mode the tape drive will convert file sync commands to no op commands and will treat tape mark commands as buffered tape mark commands.

Conclusion

The StorageTek T10000C and StorageTek T10000D tape drives from Oracle are designed to deliver the fastest performance in the industry for any tape storage environment. Whether a user has slower server technology or an application that writes discrete files rather than long data streams, StorageTek T10000 tape drives are engineered to store data faster than any other tape device. Together, the features described in this paper are capable of shortening backup windows, reducing tape migration times, and enabling new methodologies for data protection and archive storage.

Appendix A—Glossary

Backhitch
A repositioning of the tape relative to the recording head. A backhitch includes the time required for the tape drive to decelerate the tape to a stop, accelerate in the opposite direction, and decelerate to a stop again. Also commonly referred to as *football* or *shoe shining*.

Buffered Tape Mark
A command that separates record data from standard labels or other record data. This command only adds the tape mark to the data buffer without buffered data synchronization to the tape media.

File Sync
A sync operation or command used to force data to tape. For FICON: The command protocol is a tape mark command without a count field. For Fibre Channel: The command protocol is not an actual sync command. It is a tape mark command that includes a count field. A count equal to zero indicates a file sync.

Progressive Write
Progressive write technology serves as a read reliability feature by remapping data to a location further down the length of tape rather than trying to rewrite it in the same location. When small file or large file handling is underway, the gaps of "fill" data between files consist of progressive write data.

Sparse Wrap
An unused portion of the tape that serves as an intermediate location to store data while small file handling is underway. Data is "sparsely" written with gaps of progressive write data between each file.

Tape Mark
A command that separates record data from standard labels or other record data. This command implies buffered data synchronization to tape media. For FICON: The command protocol is a tape mark command without a count field. For Fibre Channel: The command protocol includes a count field. A count greater than zero indicates "count" number of physical tape marks must be written to tape.

Oracle is committed to developing practices and products that help protect the environment

Maximizing Tape Performance with
StorageTek T10000 Tape Drives
September 2013

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com