SPECTRA

DEEP STORAGE EXPERTS

Enterprise Tape: Performance

July 2014

800-833-1132 • 303-449-6400 • (Fax) 303-939-8844 • 6285 Lookout Road • Boulder, CO 80301 USA SPECTRALOGIC.COM

Contents

Abstract

Data access times are a crucial consideration for tape library purchases in processing-intensive environments. Using the right metrics and understanding them in the context of how the library will be deployed is important in selecting the right system. Multiple variables and use cases should be considered in order to acquire the right library for use today and well into the future.

Introduction

The best way to determine the true speed of a system and the best tape library to choose is to fully research and evaluate several characteristics of the drives and library robotics, and consider how they interact with data processing requirements.

IT professionals managing enterprise storage environments face many challenges, including processing staggering amounts of data that grow larger every year, and completing backups and archiving in short periods of time. Finding a tape library that meets processing speed requirements is therefore very important; however, doing so is not as straightforward as looking at a simple performance metric. Furthermore, each data center processes data differently according to internal processes and requirements, so the best library for one organization might not be the best for another.

The best way to determine the true speed of a system and the best tape library to choose is to fully research and evaluate several characteristics of the drives and library robotics, and consider how they interact with data processing requirements. Failing to do risks choosing a solution that is poorly suited to the storage tasks it needs to perform.

Overview of Attributes Affecting Performance

Enterprise Tape performance is important in order to reduce non-productive backup window times and improve data response times. For tape automation systems, two major components affect performance:

1. Tape Drive:
    a. Speed at which tape drives access data, or Average Time to Access
    b. Rate at which they subsequently read or write data, or Tape Drive Throughput
2. Library Robotics:
    a. Speed of the tape robots in retrieving tape cartridges and providing them to the tape drives.

Additionally, the above items are of varying importance and impact depending on the data and processing requirements in each IT environment.

Tape drive and library vendors provide a number of features and innovations designed to improve performance within both of these arenas. The following sections cover the essentials of the attributes, then provide specifics on how Spectra

libraries with TS1140/TS1150 technology tape drives and other innovations provide value for these attributes.

- Tape drives
- Library robotics
- Library use case for drives and robotics

Tape Drives

When considering the overall workload of an automated tape system, tape drive reads and writes normally constitute the bulk of work since they may run continuously while tape robots operate intermittently. For example, a library robot may spend seconds locating and moving a tape to a drive while that same tape drive may spend hours writing data to the tape. Since tape drives do the majority of the work within the automated system, tape drive performance is the first area tape users should understand.

There are two important drive attributes: average time to access, and tape drive throughput.

Average Time to Access

Multiple operations or factors are included in assessing the average time to data access of a tape drive. These factors can include:

1. The time needed to load a tape into a drive and thread the tape media through the drive tape path so it crosses the tape read / write heads.
2. Time to locate the appropriate data block on the tape.
3. Speed at which the tape is traveling as it moves to the data block.
4. The total length of the tape. The longer the tape, the longer it takes to reach the last data block.

The average time to access data is a critical service-level metric for a drive. Consequently, users should consider not only the individual performance factors in the previous list, but should look at them as a sum of the parts when comparing enterprise and midrange tape drive performance as part of the total solution.

Average Time to Access and TS1140/TS1150 Drives

Figure 1 shows that the TS1140 tape drive provides better average time to data access than the competing midrange drives as well as enterprise tape drives from Oracle (T10000 series).

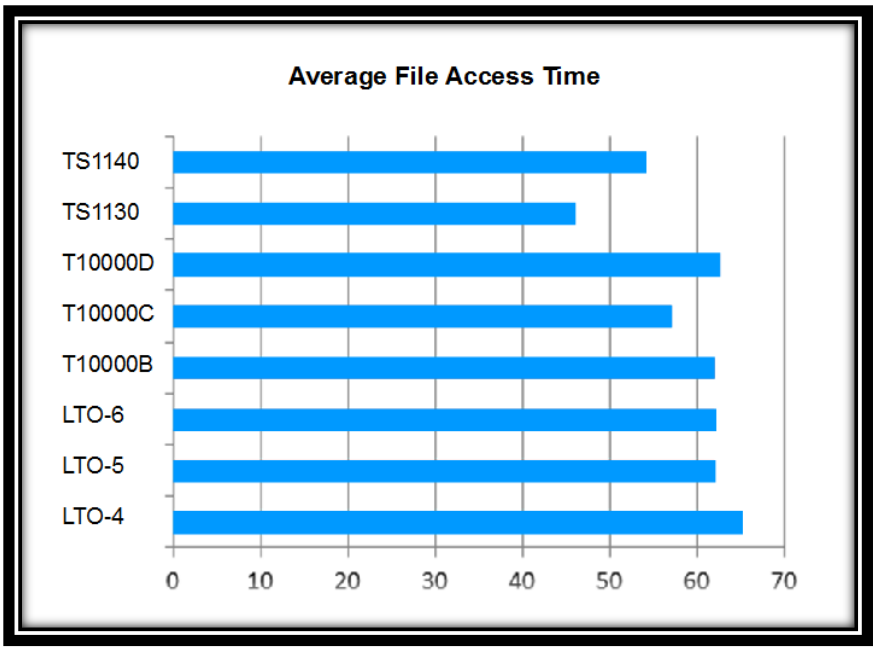

Figure 1: Average file access time for each drive technology[1]

Tape Drive Throughput

Tape drive throughput, or speed, is the rate at which a tape drive either writes data to a tape or reads data back from a tape. This metric is normally provided for both uncompressed and compressed data streams and is expressed in Megabytes per Second (MB/s). The faster the speed, the more quickly a drive can write or read data. This means the drive can either move more data within a given period of time than a slower drive, or move the same amount of data within a shorter period of time than a drive with a lower speed.

Tape Drive Throughput and TS1140/TS1150 Drives

Increasing drive speeds with each new generation of drive is critical given the ever-expanding data stores users are faced with managing. Drive roadmaps must demonstrate a commitment to advancing drive speeds in future drive generations in order to keep pace with growing data storage requirements. Tables 1 and 2 provide historical tape drive speeds and roadmap projections for expected performance increases.

[1] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.

Table 1: Midrange tape drives[2]

	LTO4	LTO5	LTO6	LTO7[1]	LTO8[3]
Native throughput (MB/s)	120	140	160	300	472
Compressed throughput (MB/s)	240	280	400	700	1,180
Release dates	2006	2010	2012	2015[2]	Est. 2018

Figure 2 shows the average file access time per technology.

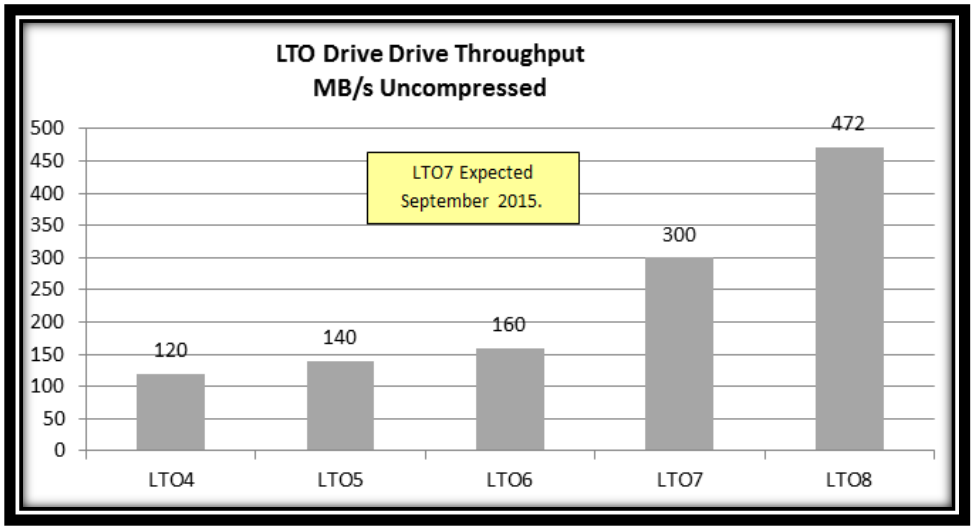

Figure 2: Average file access time for each drive technology[3]

Table 2: Enterprise tape drives[4]

	TS1120	TS1130	TS1140[2]	TS1150[2]	Gen 6[2]
Native throughput (MB/s)	104	160	250	360[2]	540[2]
Compressed throughput (MB/s)	208	320	650	700[2]	1,350[2]
Release dates	2005	2008	2011	2014	Est. 2017

[2] Source: IBM. GA expected September, 2015 (IBM). Compression expected to be 2.5:1 with 8 Gb/s FC interface. 3LTO Consortium Roadmap.
[3] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.
[4] Source: IBM TS1140 Tape Drive White Paper, October 2011. Compression expected to be 2.5:1 with 8 Gb/s FC interface.

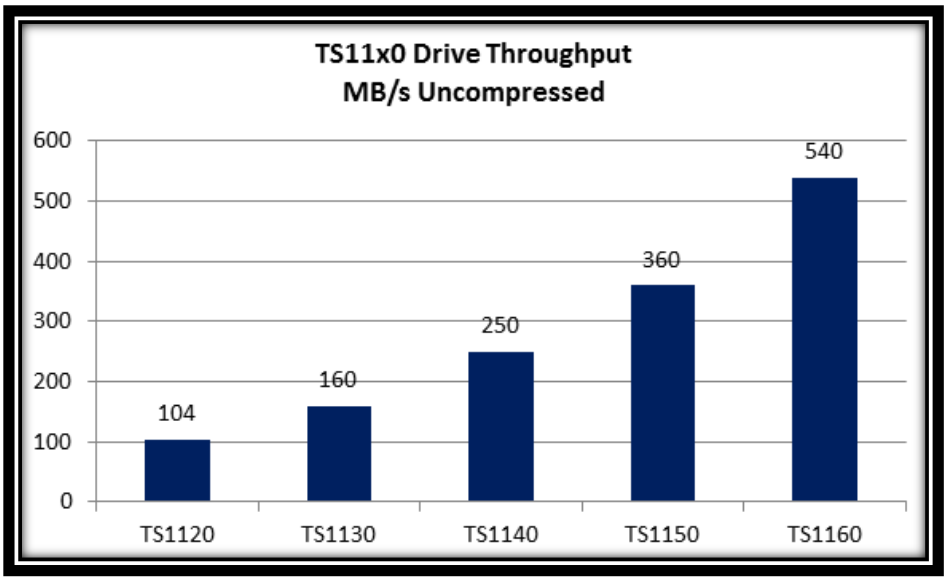

Figure 3: Average file access time for each drive technology[5]

Based on this information, it is easy to see that enterprise tape drives like the IBM TS11x0 family have provided superior performance relative to midrange drives over time. Furthermore, they are expected to continue increasing their performance advantage.

Additional Drive Speed Attributes

Tape Reformatting

Tape drives generally provide a level of backward compatibility with older generations of media that allows a new generation of tape drive, such as LTO-6, to read back two generations (LTO-4 and LTO-5) while writing back to one generation (LTO-5) of tape media. This functionality helps protect users' investments in older media by extending the useful life of that media. This functionality also helps minimize the media migration process required to move data from one generation of media to the next.

Tape Reformatting and TS1140/TS1150

The TS11x40 Technology tape drive offers the ability to reformat a prior generation of media for re-use.

While this read / write compatibility function is standard, it may also mean that drives using older media will be operating at slower speeds. However, there is an enterprise tape drive that offers something more. The TS1140 Technology tape drive offers, and subsequent generations will offer, the ability to reformat a prior generation of media for re-use. In the case of the TS1140 specifically, it can reformat TS1130 media to hold 60% more data while reading and writing at a data rate that is 26% higher it would normally provide when used in TS1130 format within a TS1130 tape drive.[6] Customers with TS1130 media in their inventory, which they can re-use

[5] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.
[6] Karp, J. IBM TS1140 Tape Drive – Technology, Features, and Function. October, 2011.

from the scratch pool, receive 60% more storage space and 26% greater performance for no additional cost when deploying TS1140. This is a very compelling performance factor.

The reformatting story becomes even better with the release of TS1150 which will be able to reformat the TS1140 media to hold up to 7 TB of uncompressed data. This increase represents a 75% gain in data storage![7]

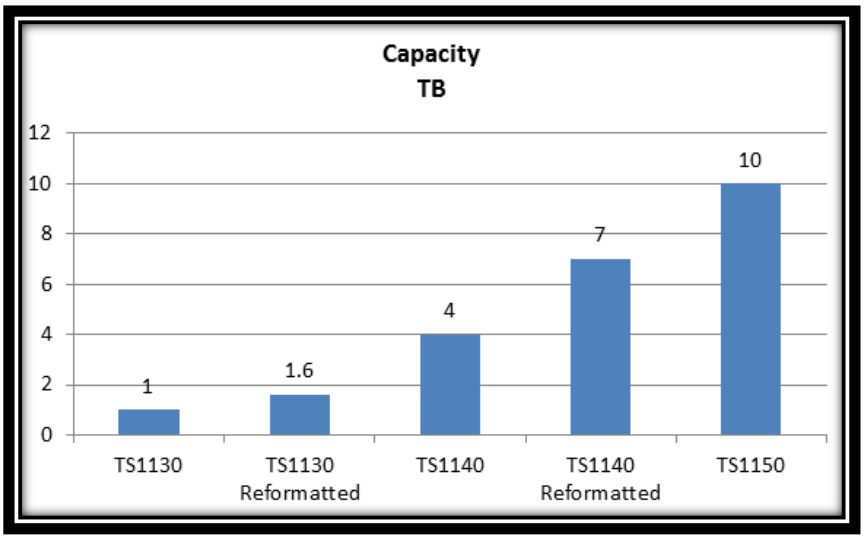

Figure 4: Average file access time for each drive technology[8]

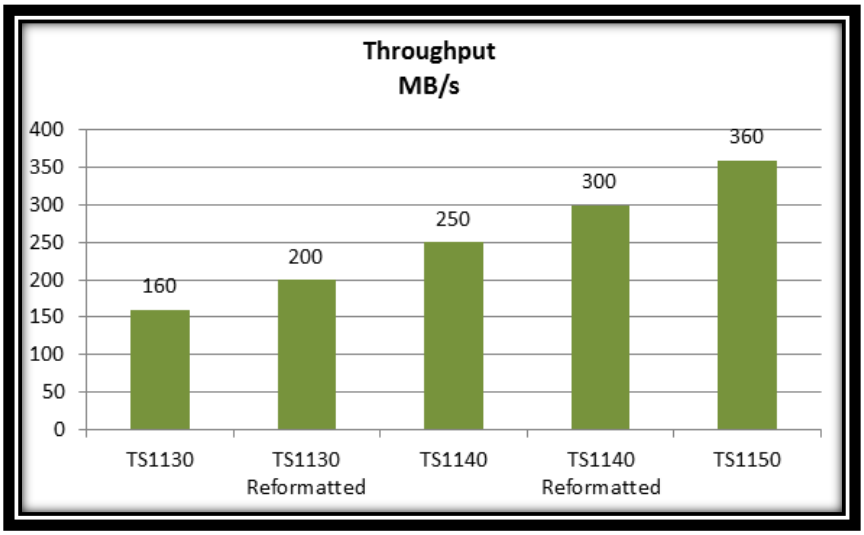

Figure 5: Average file access time for each drive technology[9]

[7] IBM Tape Drive Update to Spectra Logic, July 16, 2014.
[8] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.
[9] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.

Speed to Locate Data Blocks

When a tape drive receives a command to locate a block of data on a tape, the drive doesn't know if the data block is located at the beginning of the tape, the end of the tape, or somewhere in the middle. Working through a lower speed read process to reach the correct data block increases the latency for the data retrieval time.

Speed to Locate Data Blocks and the TS1140/TS1150 High Resolution Tape Directory

Enterprise tape drives like TS1140 and TS1150 have incorporated a feature called the High Resolution Tape Directory. This directory function allows a tape drive to forego the lower speed search iterations that reduce tape drive performance and instead make a single, high-speed pass directly to the required data block. The directory maintains logical block content information for each index of the length of the tape. This index of locations is what enables the drive to move directly and quickly to the desired location, thereby reducing latency in response time by eliminating the search iteration process.[10]

To ensure that this directory is protected, the drive stores multiple redundant copies on the tape in use while putting a low-resolution version in the tape cartridge memory. In addition, the drive will rebuild the directory upon each read operation in case previous copies have become unreadable. As a result, the High Resolution Tape Directory feature provides not only improved performance, but delivers a level of redundancy for that feature as well.

Speed Tape Moves Across Tape Heads

While the High Resolution Tape Directory allows the drive to go directly to a data block location on the tape, the speed with which the tape is moving across the tape heads during this operation also affects the drive's performance. Obviously, if you can drive your car directly to your destination, doing so at 60 miles per hour instead of 30 improves your arrival time. The same can be said of the search speed at which an enterprise drive will move along the tape.[11]

Speed Tape Moves Across Tape Heads and TS1140/TS1150

In the case of TS1140, the drive will move tape across the tape heads at the rate of 12.4 meters per second or just over 40 feet per second (about 27 miles per hour) which is quite fast for a tape drive. In comparison, the LTO-6 tape drive delivers a tape rate of approximately 10 meters per second, in which case the enterprise TS1140 drive is moving over 20% faster than its comparable generation midrange drive.

[10] Karp, J. IBM TS1140 Tape Drive – Technology, Features, and Functions. October 2011.
[11] Ibid.

Speed Range / Speed Matching

Tape drives are typically specified as having a maximum rate of sustained throughput. In the case of TS1150 (due out in the fourth quarter of 2015) that rate is 360 MB/s. However, the sustained rate is not the only rate at which quality tape drives normally run. Most drives do not have only two speeds e.g., 0 MB/s and 360 MB/s. Instead, a superior tape drive will provide a range of speeds enabling it to read and write data slowly or quickly and match the various rates at which a server might be able to provision or accept data.

The inability of a tape drive to match data rates with the attached server across a variety of speed ranges means the tape drive runs too fast or too slow relative to the server data rate. This mismatching of speeds would result in the tape drive experiencing a number of stop / start cycles as it waited for data to buffer up and be ready to write or read. This stop / start motion, called back-hitching or "shoe shining", reduces drive performance while increasing the wear and tear on tape drive components and tape media.

To account for this condition, tape drive manufacturers often build into drives the capability of matching various data rates. The greater the range of speeds a drive can provide for automatically matching data rates, the better the overall drive performance and the more reliably it functions.

Speed Range / Speed Matching and TS1140/TS1150

The TS1150 tape drive provides up to 14 speeds, up from 12 on TS1140, with which to match data rates in both read and write modes. It can match speeds as low as 66 MB/s at the low end of the range while also matching speeds up to 360 MB/s sustained and even 365 MB/s (burst rate) at the high end.

Bit Error Rate

The reliability of the tape drive when writing data to tape far exceeds the speed or capacity that the drive provides.

Great speed and great capacity also require great responsibility. In this case, it means being responsible for reliably writing data to the tape irrespective of the data rates and capacities in question. For TS1140 and TS1150 the uncorrected Bit Error Rate (BER) or rate at which a bit is written to tape without correction is 10^{-20} or roughly 1 bit out of 111,000 PB worth of bits. Even more impressively, the undetected BER, or rate at which bits are written in error without the drive knowing about it, is roughly 10^{-34} for the TS1150. This is fourteen orders of magnitude greater than the uncorrected BER. The reliability of the tape drive when writing data to tape far exceeds the speed or capacity that the drive provides.

Once tape drive performance has been evaluated for an enterprise tape solution, it's important that users consider the performance of the tape library. The library will be conducting many transactions (tape moves) that need to be included when calculating the total data access time for the system.

Library Robotics

After completing the evaluation of drive technology performance, consider the performance metrics of the library itself in assessing the total solution performance.

Total Performance = Tape drive performance + library performance

Total performance is typically viewed as a function of data response times. Users are normally looking for the best overall performance that yields the lowest total response time in completing tape reads or writes. When assessing library performance as the second half of the equation above, consider the following factors:

- Robotic Exchange Rates
- Library Size
- Number of Robots
- Robot Speed
- Library Use Case

Robotic Exchange Rates

Library performance takes many forms. The one most users are concerned with involves media handling; specifically tape exchanges per hour. A complete robotic tape exchange normally consists of the following movements:

1. Library receives a move command and passes it to the tape robotics
2. Robot moves to the tape slot location in the library
3. Robot picks a tape from its slot location within a library
4. Robot moves from that location to a tape drive
5. Robot puts the tape into the drive
6. Robot picks the tape from the drive after the job completes
7. Robot moves the tape back to its home slot location
8. Robot puts the tape back into its slot

The sum of this series of events is considered to be one exchange cycle. The number of cycles completed within a given period of time, usually an hour, constitutes a performance metric called exchanges per hour (EPH) or cycles per hour (CPH). This metric is what many users look at when considering library performance. The higher the EPH a library is capable of sustaining, the more work that can generally be done.

However, EPH figures without some context are of little value. The EPH can vary greatly based upon the following factors:

1. Number of frames within a library or its overall length
2. The number of robots within the library
3. The speed with which the robots move
4. The use case involved e.g., transactional environment, backup, archive?

Library Size

For example, enterprise libraries within only 1 frame in the configuration will be able to generate a far higher EPH figure than will a library with 9 frames in its configuration assuming robots are moving at the same speed. This is due to the fact that a 9-frame library will require the robot to travel longer distances to retrieve tapes in some cases than a single-frame library will. The added distance increases move time resulting in fewer cycles that can be completed. This fact alone means compact libraries requiring shorter move distances can have a performance advantage.

Library Size and Spectra Logic

Spectra Logic libraries are all designed with space efficiency and high density in mind. In fact, Spectra enterprise libraries such as the T950 and TFinity can hold the same number of tape cartridges that other libraries can, yet require only half the space or sometimes even less.

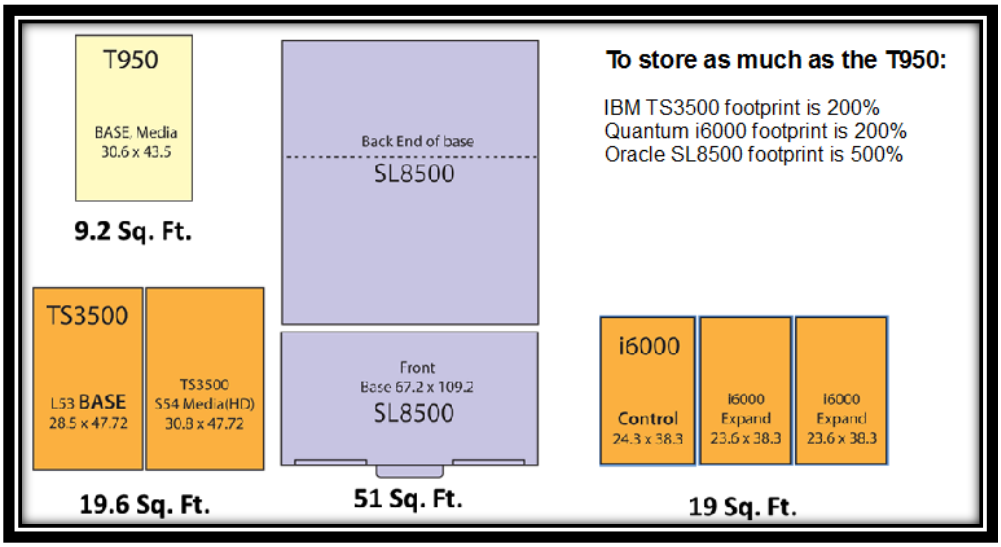

Figure 6: Comparison: Footprint holding 920 LTO tapes and 12 LTO drives [12]

[12] Sources: IBM and Oracle Tape Drive Data Sheets and Specifications documents.

Number of Robots

The EPH of a library may be similarly affected by the number of robots in use within the library. Obviously, a multi-robot library should be able to complete more exchanges in a given period than a single robot library can, given comparable robotics move speeds. However, every robot in a library might not have access to all tape drives and slots. A large number of robots delivering a high EPH figure may not mean much when one or some of those robots fail, if this prevents certain tapes or drives from being accessed at all.

Number of Robots and Spectra Logic

Only the IBM TS3500 and the Spectra TFinity provide full access to all tapes and all cartridges with both robots.

When it comes to multi-robot libraries, only four libraries are available. Of these, only the IBM TS3500 and the Spectra TFinity provide full access to all tapes and all cartridges with both robots. This means that when a robot fails, neither library will leave tapes or tape drives in a position in which they can't be accessed. Furthermore, because of its high density architecture, the TFinity provides the smallest form factor of any library allowing for the shortest robotics travel distance which helps improve performance.

Robot Speed

Robot move speed also influences EPH figures. The faster a robot can move horizontally or vertically, the more quickly it can cover given distances, thereby increasing EPH figures. Similarly, the faster the robot picker can pick tapes from slots or drives and put them into slots or drives, the more exchange cycles may be completed in an hour.

Robot Speed and the Spectra Logic High-Performance Transporter

Spectra Logic's new High Performance Transporter (HPT) improves performance time considerably, particularly for large library configurations that require longer robot moves in general. The new HPT reduces Terapack Picks and Puts from 6 seconds to 1 second per operation while also improving transit time between Terapack shelf location and drive location through the use of larger, faster robotic motors. The Terapack architecture is discussed later in this paper.

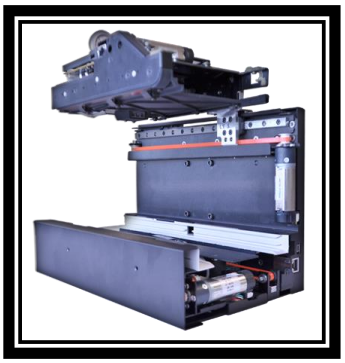

Figure 7: Spectra Logic High Performance Transporter

The Spectra Logic High Performance Transporter is expected to increase TFinity EPH to 7,500 per day by 1H 2016.

Library Use Cases, Drives, and Robotics

Finally, use cases or work flows can also affect library exchange performance. Some work flows call for very few exchanges because tapes will be in drives reading or writing for longer periods of time, thus reducing the number of exchanges called for. Alternative use case scenarios may require certain tapes be used repeatedly while others are not. In such situations, often used tapes may be stored closer to tape drives, reducing travel time and increasing EPH, while infrequently used tapes are grouped farther out since they are seldom needed.

The number of variables and their combinations that can affect EPH may be significant. As a result, it is good practice to fully understand those factors and how they combine when considering the exchange rate that may be required by a user and / or delivered by a library.

Enterprise libraries providing the best overall performance will deliver on all of these factors. For instance, the best enterprise libraries will be compact in nature which reduces travel time. They will have multiple robots working simultaneously, all of which are capable of accessing any tape slot or tape drive within the library. Robotics speed in all phases of operation, horizontal, vertical, and pick / put mode, will be high. And finally, they will be configurable, either at the library level, or by the host software, to store frequently used tapes close to tape drives for improved response times.

Different objectives can require different combinations of performance.

Different objectives can require different combinations of performance. For instance, a transactional environment may require a high rate of exchanges per hour order to access a lot of smaller files on different tapes within a short period of time. Conversely, a backup environment may be better served with a lower library EPH coupled with tape drives that have a high data transfer rate in order to meet backup window requirements since only a few drives will be loaded, but will run continuously to accept backup data. Then again, an archive environment, particularly an active archive, may do best with some combination of solid EPH and drive throughput metrics.

In the end, your data is likely to grow over time and your data management mission may change in which case performance mileage can vary. Because of this, purchasing the library with the best all-around performance can help meet your needs irrespective of changing requirements.

MEDIA HANDLING

In addition to the major metric involving EPH, many library users also want to know about the lesser known metrics. One of which has to do with media handling. Most libraries store tapes in individual slots along the vertical and horizontal access of a library wall. For any given tape move, import, or export, the robot will select only a single tape and move it to the requested location. For movements involving large numbers of tapes, those libraries will require a large number of individual robot moves to complete the entire move. The more moves required, the longer it will take to complete a job.

Figure 8: Standard tape slot configurations for Oracle SL8500 library

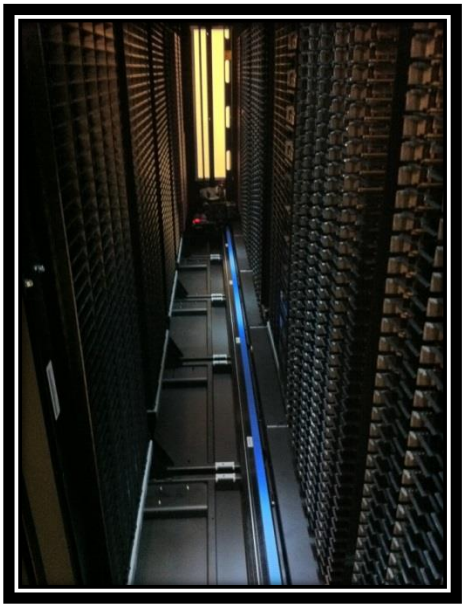

Figure 9: Standard tape slot configurations for IBM TS3500 library

Media Handling and Spectra Logic

Spectra Logic libraries are specially designed to handle media efficiently, with terapacks, bulk loading ports, and the benefits of terapack cases.

Spectra Logic Terapacks

Enterprise libraries from Spectra Logic employ a Terapack design that houses tapes within portable drawers which are recessed within the library's walls. This Terapack architecture provides two distinct advantages:

1. Utilization of the three-dimensional space within a library wall (versus standard 2-dimensional library walls where cartridges are only stored one level deep) resulting in smaller libraries with much greater density.
2. Allowing a robot to move up to 10 tapes at a time (instead of the standard approach of only 1 tape at a time) for internal movements to and from drives in the drive bay or for import / export purposes.

Figure 10: Spectra Enterprise Library Terapack Architecture

Figure 11: Spectra TS1140 and LTO tape terapacks

The net result of the architecture is that more work gets done in fewer and shorter moves. Fewer / shorter moves generally mean less time required to respond to and complete requests as well as lower power consumption which is a valuable side benefit to the increased performance.

Spectra Logic Bulk Terapack Access Port (TAP)

Media handling performance is very important for customers who are required to import or export large numbers of tapes to or from their library on a regular basis. Standard libraries require the customer to place tapes, one after another, into an I/O (In / Out) slot repeatedly until all tapes are imported. These ports are scattered across multiple frames. If imports / exports require only a handful of tapes at a time be transitioned, then individual tape movement like this may be acceptable. However, for customers who are often faced with importing or exporting tens or even hundreds of tapes at once, especially across multiple frames, doing so one tape at a time is a painfully time-consuming process.

Spectra Logic Enterprise Libraries and their Terapack design make import / export operations much easier and faster, saving valuable data center operations resources. By using the tape Terapacks to put tapes into a library or take them out, the user is able to make those tape moves at rates of up to 10 at a time instead of 1. For instance, using the Bulk Terapack Access Port (TAP) on a T950 or TFinity library, a user can import or export up to 140 LTO tapes (126 TS1140 tapes) within minutes and do so from a single library frame. In the case of TFinity, when two Bulk TAP frames are in operation, up to 280 LTO tapes (252 TS1140) can be imported within 11-12 minutes.

Figure 12: TFinity Bulk TAP (left) and T380 with single TAP

Even for smaller enterprise libraries such as the T380, the Terapack is used for importing / exporting tapes in 9 (TS1140) or 10 (LTO) tape increments making tape moves into and out of smaller libraries exceedingly quick.

Spectra Logic Terapack Cases

After tape Terapacks have been exported and are ready for transport, having an easy, secure means to move those cases becomes important. Unlike competing products in which tapes may be dumped loosely into a box, Spectra Logic provides a hard sided, padded tape carrying case . The carrying case makes it easier to securely move tapes once they are outside the library and ensures the tapes are stored cleanly given the Terapack dust covers that fully enclose the tapes.

Figure 13: Terapack carrying case (left) and Terapack with dust cover

Code Upgrades

One performance characteristic of libraries that users often overlook – until after they've purchased the library and put it into use – is the length of time and the administrative overhead required to upgrade firmware for the library and / or the tape drives housed within the library. In worst case scenarios the user will have to take the library off line, wait for an extended period of time to download the new code, install it in the library, ensure everything has updated correctly, and adjust the configuration. Depending on the size and complexity of the library, that procedure can take many hours or even days. As a result, the library is out of service for some time and may preclude users from accessing their data. In a 24 x 7 world, maintenance outages of this nature may not be acceptable.

Code Upgrades and Spectra Logic

Spectra Logic code upgrades have two key systems: push code and code staging.

Spectra Logic Push Code

Enterprise libraries from Spectra Logic include a feature known as Push Code. The Push Code feature allows libraries that are connected to the internet to automatically access Spectra Logic code distribution servers and check the code level of the library and tape drives against the most current code available from Spectra Logic. (*This feature will be generally available in the fourth quarter of calendar year 2014).* If the code on the library or drives is not the most recent, the library will automatically download the current code from Spectra Logic and notify the user of its availability. At that time, users will be able to apply the code at their discretion. This method of code delivery ensures that users always be aware that they have the most current available code, with the newest features and improvements, whenever they want it.

Spectra Logic Code Staging

In addition to the automatic code downloads, Spectra Enterprise libraries will also pre-stage code across the library server and tape drives as a background task. *(This feature will be generally available in the first quarter of calendar year 2015)*. When the library user decides it's time to apply the code, it can be done quickly, in part or in whole. Since the code has been previously distributed as a background task, the need for downtime to distribute code to all points within the library is eliminated, thereby speeding up the process. This expedited code stage is particularly important for configurations that include a large number of tape drives.

Conclusion

The performance of a tape automation system can be broken into many components, any one of which may become the gating factor in a system solution selection. Time to data access, read / write performance of tape drives to robotics exchanges per hour, and media handling characteristics of the library, all affect library performance to varying degrees. Furthermore, the way in which the library will be deployed can also affect system performance.

Given the number and combination of factors that influence overall system performance, tape library users should strongly consider all the factors in total and weigh those factors across both midrange and enterprise tape system offerings from multiple vendors. Such a comprehensive analysis of the factors involved and the solutions available ensures that the selection will achieve or exceed requirements for the library, and that data processing and user requirements will be met currently and for years to come. Doing anything less leaves the user at the mercy of suboptimal performance and the possibility of organizational damage due to slow response times or excessive down time. TS1140 / TS1150 tape technology enterprise tape drives and Spectra Logic Enterprise Libraries lead the tape industry in addressing these issues and delivering customers maximum performance.

Deep Storage Experts

Spectra Logic develops deep storage solutions that solve the problem of long term storage for business and technology professionals dealing with exponential data growth.

Dedicated solely to storage innovation for more than 35 years, Spectra Logic's uncompromising product and customer focus is proven by the largest information users in multiple vertical markets globally.

Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of deep storage—including archive, backup, cold storage, cloud, and private cloud.

For more information, please visit http://www.spectralogic.com.

SPECTRA

DEEP STORAGE EXPERTS

303-449-6400 • 800-833-1132 • (Fax) 303-939-8844 • 6285 Lookout Road • Boulder, CO 80301 USA • spectralogic.com