THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠
Published Since 1993 Report #TCG2013017 September 12, 2013

Oracle's StorageTek T10000D Further Reduces the TCO for Storing Data on Tape

Analyst: David Reine

Management Summary

Each year when the holiday season arrives, we have to compare our list of friends and family to our budget. When the economy turns downward, we cannot be as generous as we once were. The idea of gifting everyone on our list with the latest *iPhone*, *iPad,* or other electronic gadget is no longer feasible. We have to keep the cost down to a reasonable level, so we resort to some old standbys: *Lego* kits, chess sets, and board games, for example; toys and games that are as popular as ever. Some toys, in fact, stand the test of time better than others, like the *Frisbee*. A Frisbee is a simple plastic disc that has been thrown back and forth between two people for generations. Well, not any more! Today, unlike the *Hula Hoop*, the Frisbee has graduated to the big leagues, *Major League Ultimate Frisbee*! An old technology has been adapted for today's upbeat, competitive society. And the Frisbee thrives.

Adapting older technologies to today's world is not unique to toys and games. For quite some time, we have been adapting older technologies for today's requirements in the enterprise data center. Where the Frisbee has persevered on the playing fields, tape has survived and continues to thrive in the enterprise data center. Where once magnetic tape held sway as the only storage medium for the enterprise, today it is in competition with disk (no, not the Frisbee), and other media, for functional supremacy in the data center in a variety of roles. In this age of Big Data, the need and costs to store petabytes of data for long periods of time continues to grow, seemingly without limits, as the IT staff strives to retain all of the information required to meet the needs of a variety of users, without breaking the data center budget. Tape had been a primary target for backups in the enterprise, but today, with the need to backup and recover recent data quickly, disk-based storage now dominates. However, we preserve data for a variety of other reasons, including disaster recovery and archiving. **The cost to retain all of the information required by the enterprise for a long, long time (perhaps decades) has reached the breaking point for many enterprises.** The data center staff continues to strive to reduce the total cost of ownership (TCO) for the IT infrastructure. In fact, many enterprises are outgrowing the physical dimensions of their data center(s). They have to get these costs under control. One way to do that is to reduce the cost for storing data, the cost of energy, and the cost of floor space, but how? The answer for many is tape.

Tape drives have been getting better, faster, more efficient, and more reliable every few years, like clockwork. The capacity of each tape cartridge also continues to grow. Take, for example, *Oracle's StorageTek T10000* family, which debuted in 2006 at 500 GBs of uncompressed capacity on a single cartridge. Today, the StorageTek T10000C can store 5500 GBs (5.5 TBs) on one cartridge, an eleven-fold increase. **With today's announcement of the *StorageTek T10000D*, Oracle has taken that capacity to a new level.** To learn more about the StorageTek T10000D, please read on.

IN THIS ISSUE

Big Data in the Enterprise Data Center

If there is one thing that we can say today with absolute certainty, it is that the storage of data, especially *Big Data*, is on an upward spiral, with no end in sight, as the need increases to preserve and protect that data for immediate retrieval and for years to come. In fact, the requirement for additional storage is doubling every 12-to-18 months; unfortunately the budget to support it is not keeping up.

What are the most significant issues for storage facing the IT staff of any enterprise data center today?

- Having enough capacity at the lowest TCO for the IT infrastructure.
- Having enough space within the data center to store all of that storage infrastructure.
- Having enough energy in the data center to drive that entire infrastructure and cool the data center environment.

We can break these issues down even further. Capacity is one thing; scalability is another. The data center staff must be able to scale the storage infrastructure into the petabyte range, at the lowest possible long-term cost, in order to ensure sufficient headroom for the coming years. The staff also must protect the investment already made in their storage infrastructure. Having all of this scalability, however, is useless if the storage infrastructure does not have the reliability and availability necessary to meet the information retrieval needs of a variety of applications, from mission-critical to business-critical, to Internet-facing. Beyond all of this, the deployed storage infrastructure must be easy to use, with detailed analytics available to diagnose any potential problem, and it must provide investment protection to the enterprise, something that is a problem with disk-based infrastructure. The data center cannot afford all of the costs involved in replacing a product that has become obsolete and reached end-of-life, especially if this is going to happen every three-to-five years.

The two biggest culprits in the proliferation of data are backup (for disaster recovery, in particular) and archiving, essential components in any enterprise's data protection plan. Some enterprises might do a backup, save it forever, and consider it an archive. It is not. There are distinct differences between the two functions. First of all, backup is implemented to ensure *Business Continuity (BC*) to protect the day-to-day operations of the data center with the sole purpose of enabling the restoration of lost or corrupted data: applications, user files, email, etc. In addition to BC, the enterprise must protect itself from any number of disasters, both human and natural, i.e., terrorism, floods, earthquakes, hurricanes, etc. *Disaster Recovery (DR)* applies to the massive restore of server nodes and disk arrays, perhaps even an entire data center. While the IT staff might expect the recovery of a single application or files within minutes, perhaps even seconds, the restoration of an entire data center could entail many hours, or even days. It is important to understand that the target of the enterprise DR plan should be decoupled from as many sources of corruption as possible, in order to protect enterprise data from user error, software bugs, viruses, and malicious attacks.

Whereas backup protects the business, archiving often *is* the business, providing access to historical but important (valuable) content. No matter what industry the data center supports, there is a tremendous amount of historical data, often reaching exascale capacities that must be preserved, protected, secure and easily recovered, at the lowest possible cost. In healthcare, this could involve the preservation of patient records for decades, safely protected under the rules and guidelines of HIPAA. In a geophysical environment, this could be the preservation of research and exploration data for years while the industry discovers new, more economical ways to harvest oil and other mineral reserves. An archive is a permanent record of fixed content. It enables the collaboration and sharing of both mission- and business-critical data. In order to ensure the continued access to this data, the media and all of the data, must be validated for data integrity to ensure that any corruption is detected before that data is saved.

While the volume of data that needs to be protected continues to rise, the number of hours in a day, the number of hours in your backup window, do not. Shrinking backup windows have necessitated the development of new technologies to enable the safe and rapid backup and recovery of data. This includes technologies such as data deduplication to minimize the amount of data actually being stored. The security of online backups is always a concern, along with the need for scalability. Archiving has different, but similar concerns: the assurance of data integrity for decades, the accessibility of that data for many users operating on different platforms, in a variety of environments, and the issues of extreme data growth. There is an

additional, overriding concern, especially for archives – the data center staff must be able to migrate their storage technology from one generation to the next, to ensure data accessibility. Preserving data with no means to access it is useless.

Exactly how does the data center staff go about amassing sufficient storage to protect and preserve enterprise data for decades? First of all, management needs to understand that "one size does not fit all". It requires the correct mix of storage, both disk and tape, utilizing the strengths of each, to respond to both the Recovery Time Objectives (RTO) of the data center and the capacity needs, usually severely restricted by budgetary constraints. The data center must meet the BC needs of the enterprise today, along with enabling the preservation of rapidly expanding data reserves for years to come. In the event of a disaster, the enterprise must have a recovery plan. No recovery plan means no business!

It has been well established that costs are outpacing the budget and headcount of the typical enterprise data center, with the complexity of the solution increasing yearly. It has also been well established that tape has a significantly lower TCO than disk[1], especially when we are talking about petabytes, partially due to the energy requirements needed to keep petabytes of data on-line on spinning media. The question then becomes *how does (or should) tape fit into your data center?*

The Role of Tape with Big Data

Despite what you may have heard from all of the purveyors of disk arrays, tape is far from dead. In fact, the role of tape in the preservation of data for long-term storage, in addition to disk, has never been greater. In this age of big data, tape is bigger and better than ever, and continues to grow in terms of capabilities, capacity, functionality and performance to handle the demands being placed upon the IT staff to find additional capacity without adding to the financial burden. **The TCO of a tape library has been, and continues to be, significantly less than that of disk arrays configured for the same capacity and same use. The cost of tape media, in terms of cost per terabyte (TB), continues to fall as newer generations of existing technology are deployed.**

I am not saying that disk no longer has a role – it does. Disk continues to be the choice of many data centers as the home for speedy access to changing data and as the primary backup medium. When mission-critical data is lost, for whatever reason, it must be retrieved, and quickly. The cost in terms of lost business and lost reputation far exceeds the additional burden placed on the IT infrastructure budget for disk. However, it must be noted that in some cases, such as data deduplication or compression, tape may restore data faster than disk due to the re-expansion (undoing of deduplication) requirements.

In addition, as we have seen in several cloud examples recently, there always seems to be a failsafe copy of backup data stored on tape, in a remote location, as the ultimate last line of defense to protect the enterprise when the primary backups are lost. There is tremendous significance in the portability of tape media and the capability to store backups and archives in remote locations, not the least of which is ensuring that your backups and archives are decoupled from your IT infrastructure, which can protect data from local corruption.

Tape, however, has multiple roles to play. In addition to its role in support of disk for the backup of non-mission-critical data, tape is used in many data centers to provide a lower cost alternative to disk for nearline storage. Tape is also a popular target for DR sites where the somewhat slower speed of the recovery is offset by the lower cost and high reliability that tape provides. **The most popular role for tape today, however, remains as an archival target for mission- and business-critical data across a wide array of industries.** From entertainment to health care to legal, tape provides the data center with a low-cost, scalable, portable vehicle to preserve and protect the information which is the life-blood of the enterprise, but may not need to be accessed for months, or years, mitigating the need to keep it on more expensive disk media. With tape, all that the IT staff needs to do to increase capacity is to add more cartridges to the tape library rather than procuring more disk devices. With tape, this is easy and the cost per TB is low. It should be noted that some tape technologies permit the reuse of media from one generation to the next, protecting the investment that the data center already has made.

[1] See the issue of *The Clipper Group Calculator* dated May 13, 2013, entitled *Revisiting the Search for Long-Term Storage – A TCO Analysis of Tape and Disk*, and available at http://www.clipper.com/research/TCG2013009.pdf.

In addition to capacity, with tape, the performance also is scalable. Simply add more drives to the library, non-disruptively, to increase parallel access to enterprise data. These drives also must have interface flexibility, enabling communication with all of the servers as the network architecture evolves. Simplicity also is enabled with the availability of *LTFS*[2] to the tape architecture, providing drag-and-drop file-access functionality for ease-of-use.

With tape gaining an ever-strengthening position within the enterprise data center, the question moves from *"Do we use tape?"* to *"Which tape do (or should) we use?* Many advocates of tape have been impressed with the open systems qualities, capacity, performance, and reliability of *LTO* Tape Drives and Media. However, in some enterprise data center environments LTO tape does not have the scalability, throughput, and reliability offered by what is often referred to as *Enterprise Tape*[3]. One example is the enterprise offering from IBM which has persevered for decades in a steady climb through what had been perceived to be capacity ceilings.

Another example is Oracle (neé StorageTek), a vendor that has been providing enterprise data centers with the cutting-edge technology needed for greater tape capacity, higher performance, more reliability, and cost effectiveness for generations. These qualities lead to fewer cartridges and drives being required, leading to smaller libraries, a smaller footprint and the ability to consolidate the data on multiple older cartridges onto a single cartridge – all in order to improve storage efficiency. In fact, many data centers have adopted enterprise tape for their open systems data and servers, as well[4].

Oracle's *StorageTek T10000 Family* of tape drives has been leading the way in improving the capacity and throughput of enterprise tape, and continues in that quest as every industry leader should. Originally introduced in 2006, the StorageTek T10000 provided the enterprise with the capability to store up to 500 GBs of data on a single cartridge. In 2008, the *StorageTek T10000B* doubled that capacity to 1000 GBs, or 1 TB. **Most recently, in 2011, Oracle took a giant step up the capacity ladder with the *StorageTek T10000C*[5], increasing capacity five-fold to a standard 5 TBs on a single cartridge (5.5 TBs with Oracle's *Maximum Capacity feature*[6]), which (until today) remained #1 in capacity over both LTO-6[7] and the IBM *System Storage TS1140 Tape Drive*[8], the other enterprise tape drive in the room**. Now, only two years later, Oracle has one-upped itself with the announcement of the *StorageTek T10000D*.

Oracle's StorageTek T10000D Tape Drive

It should go without saying by now that Oracle has a firm grip on technology leadership within the tape storage industry. The announcement of the StorageTek T10000C proved that. With a maximum native capacity of 5.5 TBs, the StorageTek T10000D not only has more than 3.5 times the native capacity of LTO-5 (1.5 TBs), but it also has more than twice the native capacity of the recently announced LTO-6 (2.5 TBs). In addition, IBM could not even match the standard 5TB capacity of the T10000C with their later announcement of the 4 TB IBM *TS1140* drive.

With no one else to top, Oracle has broken through their own ceiling with the announcement of the 8 TB T10000D with up to 8.5 TB using the maximum capacity feature[9], almost 2.5 times the capacity of an LTO-6 cartridge and more than twice the capacity of a TS1140 cartridge. (Oracle also is introducing a 1.6 TB sport tape cartridge for accelerated access.) **The StorageTek T10000D is the highest capacity, lowest TCO solution available.**

---

[2] Linear Tape File System.

[3] See The Clipper Group Captain's Log entitled *Enterprise Tape for Archival Storage? – Why This Just Might Make Sense* dated March 31, 2013, and available online at http://www.clipper.com/research/TCG2013005.pdf.

[4] See The Clipper Group Captain's Log entitled *Ten Reasons Why You Should Consider Enterprise-Class Tape for Open Systems Storage* dated July 13, 2011 and available online at http://www.clipper.com/research/TCG2011025.pdf.

[5] See **The Clipper Group Navigator** entitled *Oracle Fulfills Commitment – StorageTek T10000C Takes Leap Ahead* dated January 31, 2012, and available online at http://www.clipper.com/research/TCG2011003.pdf.

[6] The Maximum Capacity feature ensures that an older cartridge can be rewritten in full to a new cartridge.

[7] See **The Clipper Group Navigator** entitled *Magnetic Tape Turns 60 – The IT Industry Receives Another Gift* dated July 12, 2012, and available online at http://www.clipper.com/research/TCG2012015.pdf.

[8] See **The Clipper Group Navigator** entitled *IBM's New Enterprise Tape Extends Data Retention Capabilities and Lowers the Cost of Data Protection* dated June 6, 2011, and available online at http://www.clipper.com/research/TCG2011021.pdf.

[9] And 8 TBs without that feature.

What does this mean in the real world? An enterprise with T10000D drives/media deployed can preserve 1PB of data with only 118 cartridges (in native format without compression), as opposed to the 400 cartridges that would be required for an LTO-6 environment or even worse, the 670 cartridges required by LTO-5. In fact, if the enterprise is currently on LTO-4, the data center would require more than 1250 cartridges. **The StorageTek T10000D provides the data center with a tremendous opportunity to consolidate its tape library environment and reclaim slots for future growth without buying more frames.** By the way, the T10000D is compatible with both the SL8500 and SL3000 tape libraries, as well as single and dual rackmounted tape drive offerings.

In addition to the giant leap upward in capacity, Oracle has also improved the throughput of the T10000D with a native sustained data rate of 252 MB/second for uncompressed data[10], slightly higher than the 240 MB/second achieved with the T10000C. This is 58% faster than the LTO-6 drive at 160 MB/second, and even exceeds the performance of the IBM TS1140. This means that the data center with T10000D drives can achieve a throughput of 500MB per second with only two drives as compared to four drives required for LTO-6. This has been implemented through dual 32-channel recording heads that ensure a robust write performance while minimizing tape speeds. Oracle also has exhibited leadership with the availability of both a 16 Gbps native Fibre Channel interface and a 10 Gbps native FCoE interface.

As with the T10000C, the T10000D has been designed for enterprise-class reliability to support 24x7 operations as well as full AES-256 bit data encryption, in conjunction *with the Oracle Key Manager* or *Data Path Key Management* for in-line processing with no performance impact. In addition, the T10000D has a WORM capability with the *StorageTek T10000 T2 VolSafe* cartridge for the simple and scalable security of data, with no performance degradation. Also, with Oracle's *Safe Guide* tape path, only the non-data side of the tape comes in contact with the drive.

StorageTek T10000D advanced error correction provides the reliability required to write 100 times more data than with LTO-6 before encountering an uncorrectable bit error. Oracle also has implemented an integrated data integrity validation capability to detect corrupted data from the host prior to being written to tape.

The T10000D also has a variety of media validation levels to ensure data integrity.

- ***Basic*** – which takes 1-2 minutes to validate the cartridge media region;
- ***Standard*** – which takes 15-30 minutes to validate two upper and two lower data wraps on tape;
- ***Complete*** – which takes 9 hours to validate the entire contents of the cartridge without needing to decrypt or decompress the data; and
- ***Complete+*** – which requires 9 or more hours to validate the entire contents plus the DIV CRCs.

Oracle has improved both the magnetic and dimensional stability of the media, which will minimize magnetic and dimensional changes over time that could affect readability. With a 2GB tape buffer, the T10000D improves the efficiency of writing data from slower streaming servers – an implementation that offers one-for-one host performance matching regardless of sudden changes in the host data rate or compression ratio. For the case when servers write small files to tape rather than streaming data continuously, Oracle has implemented two advanced features to ensure streaming performance across the entire range of file sizes: *StorageTek File Sync Accelerator*, which eliminates backhitches (repositioning to an earlier point on the tape) by using the buffer, spare wraps on tape, and multiple tape speeds, and *StorageTek Tape Application Accelerator*, which eliminates backhitches and enables streaming writes of small files by converting tape marks to buffered tape marks. These features make the T10000D immune to sudden changes in the host data rate or compression ratio, no matter how many backhitches might be required. All of this enables a T10000D drive to write 1GB files 3.3 times faster than an LTO-6 drive, and eight times faster for 25GBs of 13MB files (according to Oracle).

With advanced features to protect the media and your data, fewer cartridges, and faster drives, the StorageTek T10000D requires less run time and, as a result, less wear, providing the data center with a more reliable environment.

In another example of leadership, Oracle has enabled the StorageTek T10000D with the extended capability to read back three

[10] 800 MB/second for compressed data.

generations of T10000 technology, as compared to two generations with LTO and IBM technologies. This encompasses the T10000 T1 media from the T10000A and T10000B, as well as the T10000 T2 media from the T10000C technology. In addition, the T10000D drive writes to T10000 T2 media at the full new capacity to protect the investment that the data center has already made in Oracle hardware, increasing the capacity of existing media by 60%. Furthermore, *Oracle's StorageTek Virtual Storage Manager (VSM)* dynamically reclaims tape capacity when files are deleted, a feature that is not available on other tape drives.

The T10000D also supports the *Library Edition of LTFS.* This ensures drag and drop simplicity as if the files were on disk or flash, allowing the data center staff to consolidate multiple older cartridges onto a single cartridge. LTFS also permits the staff to retrieve data directly from the media without requiring a backup or archive application. Oracle also has improved StorageTek Tape Analytics for ease-of-use, proactively managing the entire tape storage system from a single interface. By leveraging intelligent analytics, Oracle has simplified tape management, enabling a worry-free deployment today, with peace-of-mind when the data center needs to grow the storage environment tomorrow.

Cost Advantages of the StorageTek T10000D

At $30,000 per drive and about $272 per 8.5 TB cartridge, the StorageTek T10000D is being introduced with the same costs as the StorageTek T10000C, even the same warranty, except with about 60% higher capacity per cartridge, lowering the cost per TB, plus higher throughput. This enables the deployment of smaller libraries, encompassing less data center floor space, lowering the TCO of the storage infrastructure. In fact, a *StorageTek SL3000 Modular Tape Library* with a 1PB T10000D configuration capable of storing 2 TBs per hour has a 19% cost advantage per TB over a StorageTek SL3000 Modular Tape Library with LTO-6 media deployed, plus a 41% smaller footprint.

Conclusion

In trying to create an environment that provides the most economical, reliable, and efficient solution to the data center's storage growth, you need to address the advantages of both disk and tape. **Disk may continue to provide the rapid response time that your applications require to meet committed backup SLAs; however, tape provides the low-cost capacity that your management demands for long-term storage to stay within budgetary guidelines.**

Oracle's StorageTek T10000D tape drive and media provide the most economical storage solution with the most functionality to lower the TCO of the storage infrastructure. With up to 8.5 TBs of capacity and 252 MB/second of throughput, the T10000D is the industry leader in scalability for both capacity and throughput. Plus, if your data can be compressed, the capacity per cartridge is 21.25 TBs, with a typical 2.5 times compression ratio.[11]

The T10000D provides the data center with the tools it needs to consolidate storage into a smaller, more efficient library. It has the reliability that your enterprise demands for 24x7 operation and, with Oracle's tape analytics and advanced, intuitive tools, is simple to use and easy to manage.

Oracle has provided the enterprise data center with backward compatibility for media to preserve the investment that has already been made in tape, and it has presented a roadmap for the next generation with tape capacity in excess of 12 TBs per cartridge. **If your mission is to overcome the data center's biggest storage challenges, then the StorageTek T10000D may be the answer.** Check it out!

[11] This is the same "average" ratio that is quoted for LTO-5, LTO-6, and TS1140 cartridges.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twenty-first year, is an independent publishing and consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

David Reine **is a Senior Contributing Analyst for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.