Tape Performance Accelerates

HORISON
Information Strategies

www.horison.com

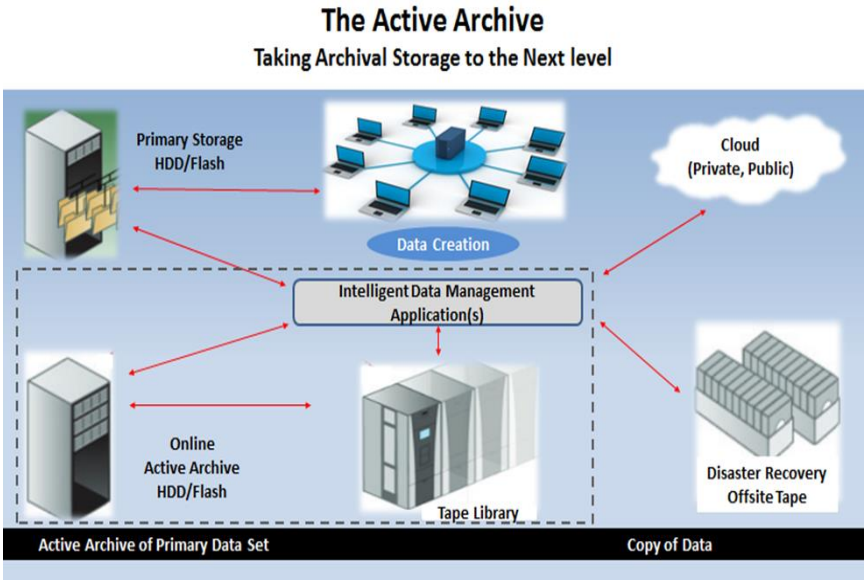
By: Fred Moore President

Introduction

With the arrival of Active Archive, much higher data rates, LTFS, RAIT, and RAO capabilities now in place, the tape industry is making significant strides in delivering much faster initial access times and throughput levels. For all the amazing technological progress made in the traditional data centers and cloud computing, the fundamental challenges of reliably transferring large files and volumes of data at high speeds to locations around the world continue to increase. Large amounts of unstructured object data, including text and binary data, archival data, images, videos, audio, documents, scientific data, and Blob storage are ideally suited to be stored and retrieved most cost-effectively on modern tape. These emerging trends present an enormous mass transit problem for digital data access and a potential impediment to future growth if it weren't for the tape industry's renewed focus on performance.

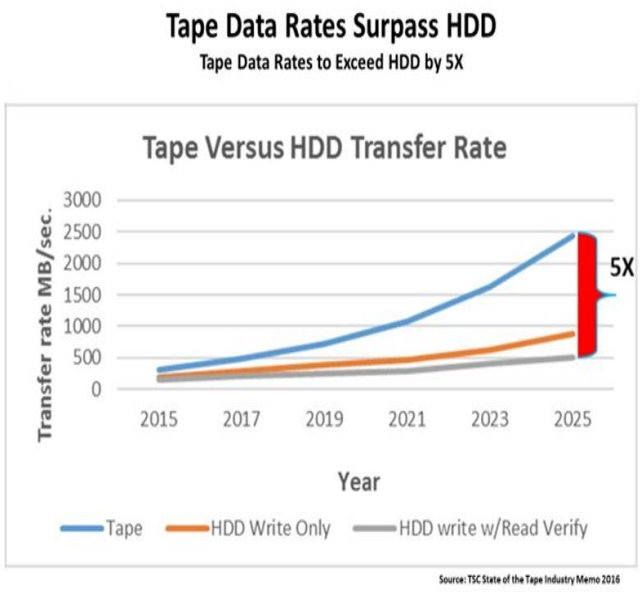
Active Archive

An Active Archive integrates SSD, HDD, tape, and the (public or private) cloud making it a special case of the popular tiered storage model dedicated to the archive function. The active archive greatly improves tape access time by using HDD or SSD as a cache buffer for a tape library. The active archive enables a high percentage of accesses to the tape subsystem to be satisfied online from SSD or



HDDs (the cache hit ratio) significantly improving access time to first byte of data. Intelligent data management software automatically migrates data between storage tiers based on user policies.

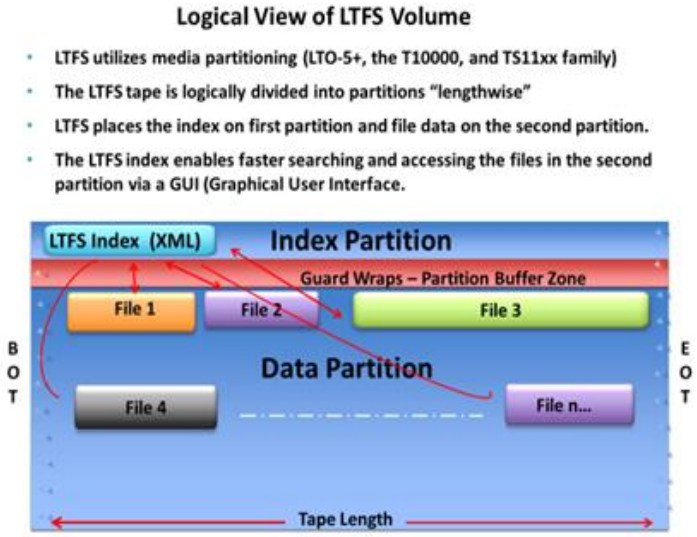
Tape Delivers the Fastest Data Rates Available



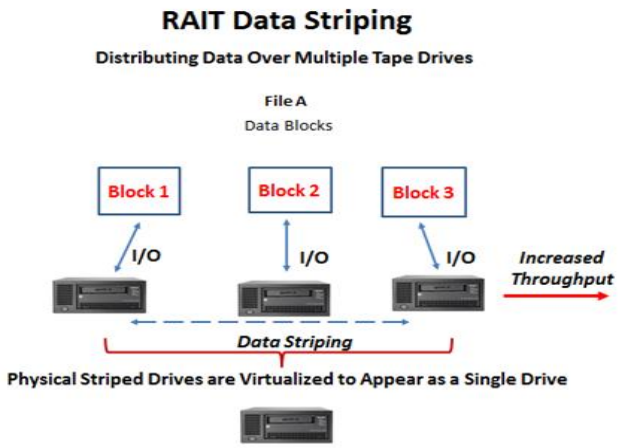
Tape capacities and data rates (tape throughput) are growing faster than all other storage technologies. When comparing native data rates, the enterprise TS1155 and LTO-8 tape drives both transfer data at 360 MB/sec. which is significantly faster than the typical 7,200 RPM HDD at ~160 MB/sec. Future tape transfer rate increases are even more promising projecting to yield tape data rates 5x faster than HDDs by 2025 with no fundamental technology limitations in sight. When architected with RAIT, the aggregate throughput of a tape subsystem is unprecedented.

LTFS Enables Faster File Access

To improve the access and interchange capabilities of tape, a new, long awaited open standard file system specification for LTO called LTFS (Linear Tape File System) became available with LTO-5 in 2010. Originally developed by IBM, LTFS provides an easier way to access and archive data to tape without the need for another backup software product. With the new dual partitioning functionality of LTFS, one partition holds the index and the other contains the content, allowing the tape to be self-describing. The metadata of each cartridge, once mounted, is cached in server memory. Metadata operations, such as browse directory tree structures and file-name search, are performed more quickly in server memory and do not require physical tape movement. LTFS recently announced connection with OpenStack Swift to enable movement of cold (archive) data for object storage to more economical tape and cloud storage for long-term retention. The faster file access capability provided by LTFS becomes more important as tape capacities continue to increase and the number of files per tape steadily increases. LTFS makes archiving and retrieving data easier than ever before for tape drive and library applications.



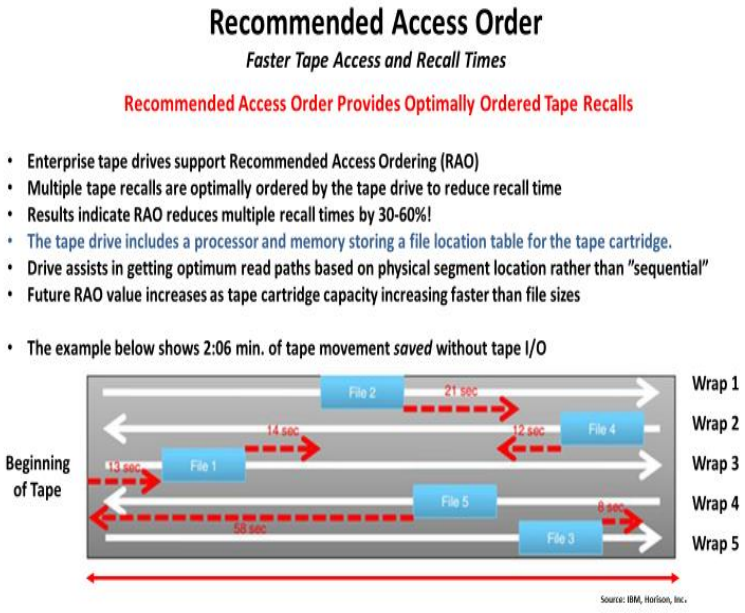
RAIT Improves Throughput and Offers Fault Tolerance



RAIT (Redundant Arrays of Independent Tape) significantly improves the throughput of large sequential files by creating multiple parallel data lanes into the tape subsystem and can also provide varying levels of fault tolerance for higher availability. Fault tolerance means providing a safety net for failed hardware by ensuring that the drive with the failed component can still operate while not impacting availability or increasing the chance of data loss. RAIT levels are implemented in software and depend on how many tape drives you have in an array, how critical drive recovery is, and how important it is to maximize tape performance. The only extra RAIT cost is the amount of space used for parity. RAIT provides data redundancy without needing to create multiple backup copies. Striping and parity are the keys to RAIT (and RAID for HDD) implementations. As files, objects, and unstructured data constantly grow, coupled with increasing emphasis on geographically dispersed DR activities, RAIT is poised to offer major reductions in data transfer time and recovery time.

RAO (Recommended Access Order) Improves Tape File Access Times

The RAO capability is available on enterprise tape drives for improving tape file access or recall times (time to first byte). Presently, files are written on tape in sequential order but are most often accessed (reading data) in random order. As tape capacities and therefore the number files on a cartridge continue to grow, file access times increase as contention mounts. The RAO determination is performed by the drive and produces an optimized list called "best access order" which provides the least amount of time that is needed to locate and read all concurrently requested files or data sets on a tape.



Summary

Tape performance is not standing still. Recent developments are positioning tape for the massive amounts of data that will need to be accessed, transferred and stored as the information age moves quickly to embrace the cloud, Big Data, the IoT, and increased regulatory requirements which archive almost everything indefinitely. Future roadmaps indicate that tape will continue to make major strides in capacity, performance, reliability, and total cost of ownership with relatively few limits in sight.