

<u>Home</u> » <u>People</u> » Exclusive Interview With Fred Moore, President, Horison Information Strategies

Exclusive Interview With Fred Moore, President, Horison Information Strategies

True believer in storage hierarchy with tape

By Philippe Nicolas | December 5, 2019 at 2:17 pm

Fred Moore is president of <u>Horison Information Strategies</u> based in Boulder, CO, launched in 1998. He started his long storage journey at Storage Tek where he stayed 21 years finishing as corporate VP of strategic planning and marketing. During recent Fujifilm Summit, we asked him questions about the current industry trends, the tape role, status and evolution and how he's seeing the future.

StorageNewsletter: The storage Industry has evolved a lot and rapidly during the last decade, what are your perspectives on this?

Moore: From a technology perspective, the storage industry has seen several key advancements and shifts in the past decade. We've seen much of the traditional HDD primary storage market move to SSDs as their price becomes more attractive. HDDs continue to increase capacity though at a slower pace than in the prior decade while HDD performance increases (IO/s) have been minimal if any which opened the door for SSDs. Tape has essentially disappeared from smaller data centers but is growing in large enterprise and hyperscale computing facilities. The evolution from centralized computing to the logical extremes of the network as the IoT expects to connect ~25 billion nodes by 2020 is placing unknown amounts amount of data capture and filtering at the far edges of the network. There is a shift of emphasis underway from storage CAPEX to OPEX (TCO) and it has recently accelerated as businesses realize initial costs can be dwarfed by operating costs.

SDS has increasing momentum as it offers relief for increasingly complex storage management tasks. The archival (store it forever) data explosion has encouraged using more cold storage solutions than ever before making tape in particular, and any other potential energy efficient technologies, very appealing from a total cost perspective. Data security was for years a data center problem, but today it is everyone's problem. There are many other shifts underway, but these capture much of the industry focus right now. In addition, the shift from many smaller data centers to fewer but much larger ones has really taken off. This is called the "Hyperscale Effect".

Tape has moved from backup to new applications including archive and even deep archive, how do you explain that, what is driving this shift?

The shift in tape usage from primarily a backup solution has been fueled by the rise in the requirements to retain data for much longer periods of time and the need for faster data recovery, and the favorable tape TCO. Several applications now see benefit from tape solutions. They include dormant big data waiting to be analyzed, archival data, legal compliance, GDPR, medical records, photos and images, e-mail history, unstructured files, scientific, video, movies, audio, documents, social media stockpiles, emerging cloud archive applications, video surveillance, remote data vaults, hyperscale, BC/DR are all good candidates for modern tape and offer market growth potential.

How do you see the role of tape today, and in the future?

Unlike HDDs and SSDs, tape only exists in the high end of the mid-range, enterprise and large-scale data centers. The vast number of single-user systems, personal appliances, PCs, and small data centers seldom use tape and therefore most people don't know much about tape having never seen it before. Tape is growing in the large-scale data center world. Given that 60-80% of all digital data is low-activity or archival, the total cost (TCO) and energy expense especially of keeping inactive data on HDDs is becoming cost prohibitive as data centers grow their storage environment. Like I have always said "if data is not being used, it shouldn't consume energy." If for some reason cost isn't an issue, then put all your data on SSDs.

What is the future of the tape format? Is the LTO the clear choice or other formats such Oracle or IBM a serious alternative?

The data center tape industry has settled on two distinct formats – the IBM TS11xx enterprise tape and LTO for everything else. LTO currently represents around 90% of total tape capacity shipped. Oracle, who owned the StorageTek tape business, is no longer developing their T10000 proprietary tape format as they exit the tape business. Sony, along with Fujifilm, are the two LTO media suppliers while Fujifilm manufactures enterprise media. I expect these two well-established formats to continue their roles for the next decade or more, as they are the clear standards with solid software support and few known technological limits.

Less players and less tape formats, is it result of the combination of HDD-based secondary storage with data reduction and even sometimes archiving with disk WORM models?

I don't think disk archiving with WORM has much to do with the tape consolidation. However, both the HDD and tape industry have witnessed significant consolidation. Today there is one tape drive manufacturer (IBM) and two tape media manufacturers (Fujifilm and Sony) to address the entire tape industry. There are several tape library suppliers. It's no surprise that

there are only 2 formats as this simplifies the buying decision greatly and helps format standardization. Note the HDD industry has only three manufacturers (Seagate, Toshiba and Western Digital). Fewer tape formats mainly resulted from having way too many incompatible formats, several proprietary file systems, and customers demanding open standards for tape.

LTO arrived in 2000 and successfully replaced the older DDS, Travan, 4mm, 8mm, DLT and other aging formats as they presented too many troublesome operational issues including edge damage, stretch, tear, loading problems, and media alignment. Data centers grew tired of battling these issues. LTO gave users higher reliability and an open system tape format with an open system file system called LTFS resolving many of the tape dissatisfiers of the past

What about optical disk as Sony, IBM and Panasonic continue to develop and promote such approaches?

Optical disk has been attempting to gain traction in the data center mass storage market since the early 1980s but hasn't done it. The learning curve for optical recording has been slow and has fallen far behind the progress of the magnetic technologies in areal density, price, performance, capacity, throughput and reliability. A Blu-ray disc has a capacity of 50GB, HDDs and tape have a native media capacity of 20TB – 400x greater. Some other multi-layer optical products periodically appear but have struggled to make any impact. The amount of for optical media needed compared to an equivalent magnetic storage capacity makes the management task burdensome as storage requirements grow.

Today you will find optical disks (CD, DVD, Blu-ray) in your car and home for entertainment, but seldom in the data center. Long media life has always been a plus for optical and it reduces the resource consuming media re-mastering effort. However, this is a concept that never gives up. I am presently interested in a new multi-layer, multi-side approach using fluorescent disk technology which is currently under development offering modern enterprise data center capacities along with a major shift in the remastering capabilities. Stay tuned.

And what about RDX?

I've always liked the removable media, random access, mass storage opportunity having worked with some attempts earlier in my career – however none were successful! The RDX architecture uses removable HDDs and fills the remaining hole in the traditional storage hierarchy for random access-mass storage – but for smaller compute facilities. This removable, offline HDD approach improves on the TCO of online HDDs significantly. The RDX system currently scales to only 40TB, the capacity of two TS1160 tape cartridges explaining why RDX is confined to small systems. I would like to see the RDX architecture scale much higher to reach larger markets now that HDDs have a 20TB capacity. RDX visibility across the broad storage industry remains minimal.

HSM, tiering, and migration were for a long time associated with tape and later with secondary disk farms, what are the recent developments in that domain?

Today's tiering uses SSD, HDD and tape optimally and each technology tier has its advantages. In the past, the limitations of effective easy to use storage management software presented the biggest challenge to taking advantage of all the tiered storage benefits. As storage farms increase in capacity, the benefits of tiered storage soars. Management software is finally

realizing that and will get a boost from AI and ML over the next few years. This needs to happen as the sheer magnitude of managing millions of files has gone beyond the scope of user defined policies. Without advanced software, storage will become too labor intensive to effectively deal with. From 2011-2020, data was expected to grow $\sim 50X$ while the number of trained IT professionals was projected to grow just $\sim 1.5X$ meaning that software must step in as there simply isn't enough human resource. With over 60% of the world's data stored in the wrong place, the goal of tiered storage is to get the right data in the right place while optimizing storage expenditures.

You're a true believer of storage hierarchy to control the cost of the storage environment and makes if efficient, do you see any changes in such model?

Yes, I am a true believer in the storage hierarchy and the late IT expert Jon Toigo often said my career was based on a pyramid. I think we are poised to see advanced storage management software applications evolve that look at the usage of data rather than just the underlying hardware technology to manage data throughout its lifetime. Unlike the traditional four-tier technology hierarchy, this new model most likely defines a simplified 2-tier model with active data in one tier and everything else in the other, lower tier and allowing migrated data to be easily accessible to users.

How does cloud impact the tape world?

The cloud is becoming a large tape consumer. Since 60-80% of the world's digital data is essentially archival, low activity data, it makes no sense for cloud providers to incur the much higher TCO and energy expenses of spinning and cooling HDDs on a 7×24 basis. Tape has recently become a new cloud service offering for archive and deep archive (cold) data applications. Since tape spends most of its life in a rack or slot in a tape library, no energy consumption is required. Using HDDs for archival storage is a strategy – just not a very costeffective one – and the cloud providers now understand this as their storage requirements soar.

In addition, tape is an "air gap" technology. The tape air gap, inherent with tape technology, has renewed interest in backing up data on tape. It means that there is no electronic connection to the data on the removable tape cartridge therefore preventing malware and ransomware attacks. Disk systems remaining online 7x24x365 are always vulnerable to a cybercrime attack. The tape air gap adds a security bonus for cloud providers hosting valuable archives.

We used to say that disk is the new tape but with the cloud, it is now cloud is the new tape, do you agree?

I'm not sure if I would say cloud is the new tape yet, but tape dynamics have really changed for cloud services. Some cloud archive services are now exclusively tape based. Cloud providers strive to offer price competitive services that require an infrastructure that has lower cost and scales easily. Cloud providers lean heavily on HDDs, and SSDs to a lesser extent, but tape has joined them as the third technology. To be attractive to customers, the cloud market has to contain costs going forward.

Is the TCO still in favor of tape especially when we compare an entire tape library vs. a deduped disk array? And what about cloud comparison?

The most recent study I saw for tape TCO indicated LTO-8 tape had a TCO of about 30% of the

cloud for backup and archive applications. I don't focus much on the de-dupe TCO as it primarily serves the backup-recovery market where recovery time for critical data is the critical factor. De-dupe is excellent for fast recovery of small to medium sized files and reduces the size of the backup set but hasn't made an impact beyond the backup market or into archives for example. However, from a TCO perspective, de-dupe effectiveness and TCO increases as the number of duplicate copies of a file increase. TCO is currently one of the main selling points for tape whereas faster recovery is the best reason for de-dupe.

What about Hyperscalers? You published a paper about that explaining that it is the new Eldorado for tape vendors? Can you elaborate on that?

Hyperscale Data Centers are enormous computing facilities with typically over 400,000 square feet and number about 500 worldwide today. HSDCs are leveraging the many advantages of tape technology out of shear necessity to manage massive storage farm growth and long-term data retention challenges while slowing the number of disk drives needed. Keep in mind most digital data, and that includes cloud data, doesn't need to be immediately accessible and can optimally reside on tape subsystems for long periods of time before needed. Tape reduces the enormous HSDC energy consumption challenges, offers a lower TCO and provides air gap protection. HSDCs are a clear growth opportunity for tape and think it will be tough for HSDCs to grow storage pools without using tape to at least handle low-activity and archival data.

What do you predict for 2020? For tape landscape but also more globally for the industry? If data is the new currency, then storage is the new bank. As a result, the overall storage and tape landscape "must" continue to advance significantly. For tape, the good news is that many strides have been made in the past 10 years regarding price, performance, capacity, reliability and throughput – the bad news is that not many people know them. The overall low visibility of tape and timid market awareness approach continues to puzzle me as the tape industry doesn't promote itself or its advancements very effectively.

The reliability changes engineered from the past decade vaulted tape into the top spot in storage device reliability with a BER (Bit Error Rate) of 1×10^{19} , three orders of magnitude more reliable than HDDs at 1×10^{16} . Tape data rates have soared and reached 360MB/s for LTO-8 and 400 MB/s on the latest TS1160 enterprise drive, more than twice the data transfer rate of most HDDs. Tape offers two features (RAO for enterprise and TAOS for LTO) that order multiple requests to a tape cartridge to optimize physical tape movement typically improving file access time by as much as 50%. Improving tape access time has been a standing requirement. Not many people are aware about these facts? Why don't they know?

Demand for long term data preservation is soaring. Big data waiting to be analyzed is piling up. Legal data requirements are mounting daily. The IOT will create unknown amounts of data that may not be touched for several years. Without a significant effort to effectively educate the world what modern tape is really capable of, I expect tape capacities shipped will grow at traditional rates in the range 8-15% annually.

What advice would you give to the tape industry going forward?

Given that tape technology is in the best shape since its inception in 1952, and that most of the advancements go unnoticed, I would say the tape industry needs to promote itself much more

effectively than in the past. New tape products often appear without any public announcement forcing end users to dig through websites to per-chance discover what's going on. Proof of might found be in a recent published StorageNewsletter.com's article indicated the web site had published over twice as many articles on optical disk as on tape since 2007 yet optical has struggled to be a viable data storage solution for over 30 years. Too many have checked out of the tape industry remembering the older format issues, but today's tape is nothing like that of the past. Given the significant advancements tape has made, tape growth should be much higher. It's past time for the tape industry to stand up and be counted.

Share this news:



With all the daily news on the worldwide storage industry, StorageNewsletter is updated every day at 9AM in Chicago.

Subscribe to our free daily newsletter



Scale server storage easily using hardware you already own.



Copyright © 2019 StorageNewsletter is published by Micro-Journal

