Special Report on Fujifilm Summit 2019 – San Francisco

Tape still alive for hyperscale data centers

By Philippe Nicolas Storage Newsletter November 5, 2019 at 2:23 pm

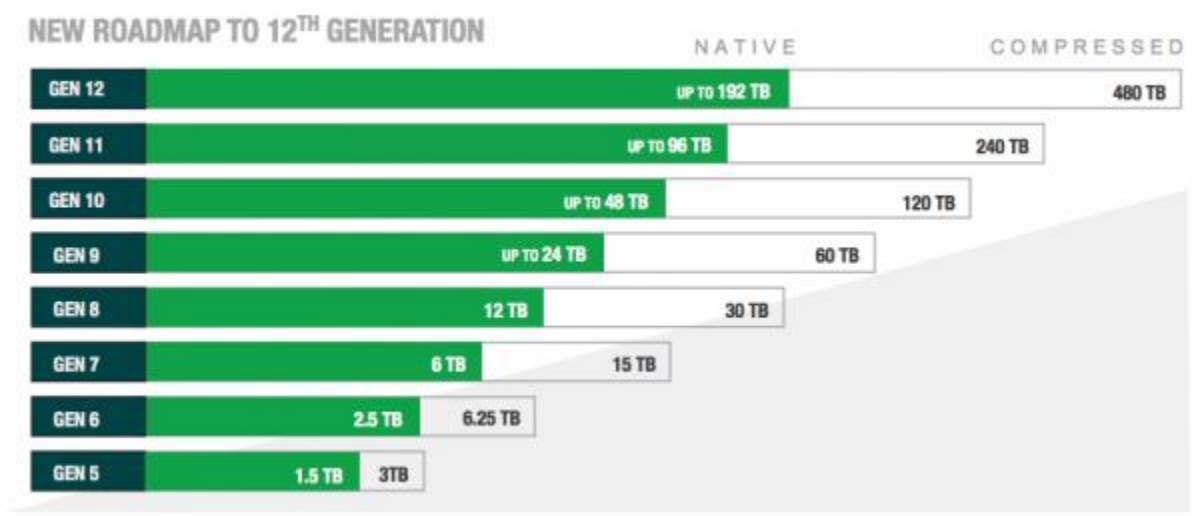
FUJIFILM Recording Media USA organized recently its 11th Summit in San Francisco, CA, supported by IBM, Spectra Logic and StrongBox Data Solutions with participation of Quantum, HPE, Western Digital, Microsoft Azure, AWS, and Twist Bioscience plus end-users and the participation of Fred Moore, CEO, Horison Information Strategies.

We count approximately **110 attendees** as the event works by invitation. The theme of this conference was *Preparing for seismic shifts in the storage landscape*. And the legitimate question could have been: is tape dead? what is the future of tape? or is cloud the new tape? among others.

With cloud, flash, HDD and tape technologies present and evolving on the market at different pace, the role of the tape has changed for a few decades and this event tried to position the role of the new tape promoting some advantages that still exist for this media. The reality is that tapes move down in the data storage hierarchy moving from a backup media many years ago to a deep archive play.

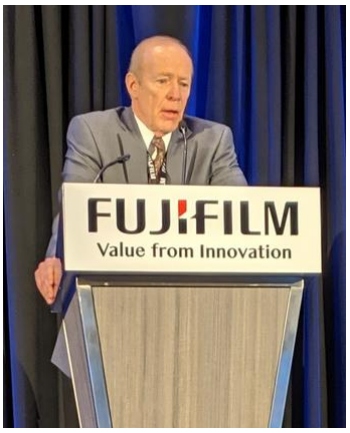
Peter Faulhaber, president and CEO, Fujifilm Recording Media USA, started the conference with the state of the company and the tape industry. As a corporation, Fujifilm represents 78,000 employees generating $21.9 billion of revenue with $1.5 billion of annual R&D budget. Faulhaber introduced the explosion of data and insisted on the 60% of data to be archived, positioning the tape as the preferred archive device, with only 8% of this volume effectively stored on tape based on current tape shipments. So where the rest is going? This is what is called dark data exposed but not protected.

Fujifilm shared also the **LTO roadmap** with interesting capacity and speed to sustain market demands and explosion of data volumes but no timeline is indicated. The direction taken by Fujifilm R&D is to introduce new tape film built with strontium ferrite to reach a new level of areal density to offer 400TB per cartridge, 33x times more than LTO-8. Faulhaber was also awarded by HPE for passing the 100 million LTO tapes sold.



Illustrating the in-progress data deluge reaching 175ZB in 2025 according to IDC with 7.5ZB of persistent data, Moore, long time StorageTek leader, took some time to articulate the need for a tiering storage approach coming from a multi-tier approach used for many years, still active and used by some users, but very complex to maintain over time,

to a new model with only 2 tiers, essentially a primary tier and a persistent tier. In other words, users consider one tier for production and one or more secondary tiers for persistence but no extra tiers behind the tier-2. Each of 3 tier-2 elements provide its own attributes such latency, capacity, redundancy or access methods to list a few. We all remember the ILM market wave early 2000 where the idea was to align the data value to the storage cost.



The second point presented by Moore was the ascent of hyperscale data center and their growing need to protect data in a very vast geo-distributed IT environment for each of them. With 500+ data centers of that nature existing on the planet, extreme ones represent 18+ times the soccer field size to give an idea. Associated with this hyperscale storage need, these players started to add intelligence with AI and ML to go beyond classic methods.

Seeing hyperscale players at this kind of event could be a surprise for some readers but tape is key component of their data protection strategy over long-term. It explains why Google, AWS, Azure, LinkedIn and Facebook participated in this conference being user of tapes and having strong challenges around deep data archiving.

Microsoft Azure explains that they used commodity hardware or designed systems powered by ASICs or FPGSa with software developed by them with open source components. Of course, at scale, erasure coding is mandatory and used in many places. Azure and AWS confirm their difficulty to find on the market products that fit their need as hyperscalers really redefine vendors' directions.

The other dimension is the automation and orchestration and the small number of people needed to manage such environments boosting petabyte/individual. AWS insisted on the price of its S3 Glacier Deep Archive at $12k/PB/year and to continue to offer such good price, they need attractive tape specifications at a good $/TB.

HyperScalers insisted on this TCO of technologies putting pressure on vendors projecting potential big, very big contracts. They expect tape vendors to deliver on the LTO roadmap as current LTO-8 has 12TB native and HDD reaches soon 20TB and keep a compelling $/TB. And to cite one HyperScaler vendor present at the event: *"If tape continues to deliver on its promise of low cost and high density, it will continue to see more adoption in hyperscale datacenters"*. LTFS was also mentioned several times as a key integration point in large environments with its standard and mobility aspects.

Twist Bioscience delivered an interesting presentation about DNA Storage with a few key messages: DNA is very stable as we can *read* information thousand years later solving the data persistence or durability perpetual challenge with the notion of data permanence, 14 atoms to store a bit instead of 10^{14} atoms for tape or disk to store the same bit, very low energy and footprint. For some of you who remember their advanced biology course, DNA is a molecule composed of 2 chains to form a double helix, and these two strands use nucleotides built with 4 nucleobases C, G, A and T. DNA storage will use a 2-bit model mapped to these letters and 00 is associated to A, 01 is G, 10 C and 11 T during the synthesize process. In DNA storage the media generated is the information as data is synthesized. And the beauty is 1g can store 1EB, each human being its own datacenter…

Nathan C. Thompson, CEO and founder of Spectra Logic Corp., leverages the new tier-2 storage model with primary and perpetual storage to promote its new tiering solution named StorCycle, a new HSM-like software, avoiding completely the rest of its storage device portfolio. It integrates multiple back-end devices such tape libraries, object storage or a secondary NAS. The product offers two flavors of files when data are migrated: html-links and symlinks. Spectra dropped also its ArticBlue and Verde products replaced by a comprehensive updated BackPearl product line: Converged Storage System, NAS and Object Storage Disk.

From StrongBox Data Solutions, Inc., the other software presentation delivered by **Floyd Christofferson**, CEO, insisted on the metadata role in unstructured data management and their 4 different types: file system, rich application, external and user-created meta data. Leveraging a deep meta data knowledge, wide storage support and comprehensive policy engine, users can build a efficient data centric storage environments with cost of storage aligned with data access needs.

IBM has presented its TS1160 offering 16Gb FC, now 10GbE and soon over RoCE v2.

Western Digital Corp. spoke about zoned storage, flash and SMR, nothing about ActiveScale, WD's object storage platform, even if the speaker, **Stefaan Vervaet**, promoted the platform for a long time coming from Amplidata, acquired by WDC in 2015 for $310 million. Wow! As WD suffered on this system and platform directions, the company has decided to sell Tegile assets to DDN – good for DDN – but ActiveScale is still orphaned.

Tape TCO has its own presentation and a few others covered that aspect as well but as usual comparison is based on tape vs. disk and not tape library vs. secondary disk array. Things like data reduction (compression and de-dupe) is not covered for disk and absence of refresh for cloud mode when you compare with on-premise approaches. Fujifilm provides an interesting tool available here. The topic of energy was mentioned as a key advantage in favor of tape with media mobility but we regret energy savings approach was not covered from some vendors as it represents a real alternative used at Hyperscaler as well. We invite the reader to check Pelican project papers from Microsoft.

We read that tape system scales by adding tape cartridges and disk by adding disk drives but the reality is that adding tapes without drives is useless as you reduce tape/drive ratio and parallelism. The truth is to keep the ratio tape/drive confirming what all people know, a tape cartridge needs a tape drive and disk is a drive by itself with media included. Tape doesn't need power but both tape and disk drives require power.

Optical was not covered except a few questions from the audience. We understand why but as a market observer it would have been interesting to articulate comparison between these two models as optical storage carries two key values of tapes: passive nature of the device and mobility.

We wish to see as well next time some presentations around indexing especially content indexing as it is a fundamental component of a deep archive model as it is almost impossible to read back thousands or millions of tapes just to do that. This indexing has to be realized in the ingest process and facilitates the location of the right data searched. Also, a section about compliance would have its place covering GDPR and CCPA for instance.

This was an event established as a reference for tape industry.