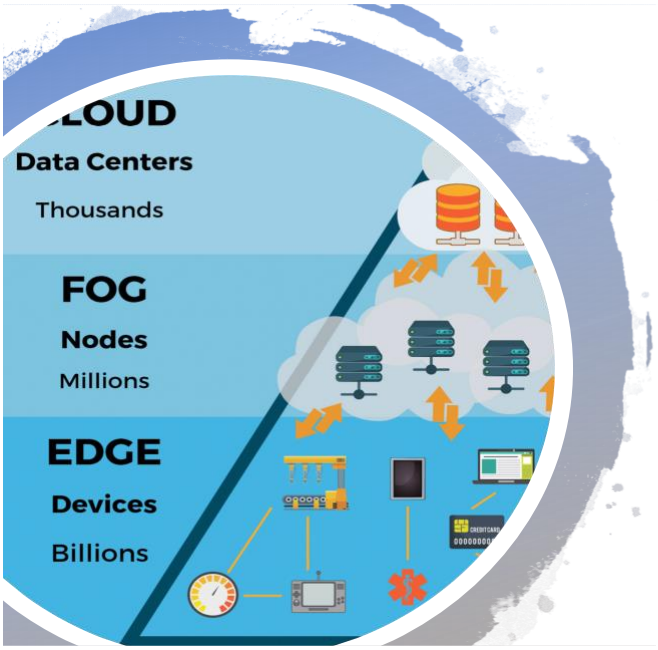


The Next Frontier for Tape

The Cloud, Fog, Edge and IoT

Horison Information Strategies
Fred Moore, President
www.horison.com

Introduction

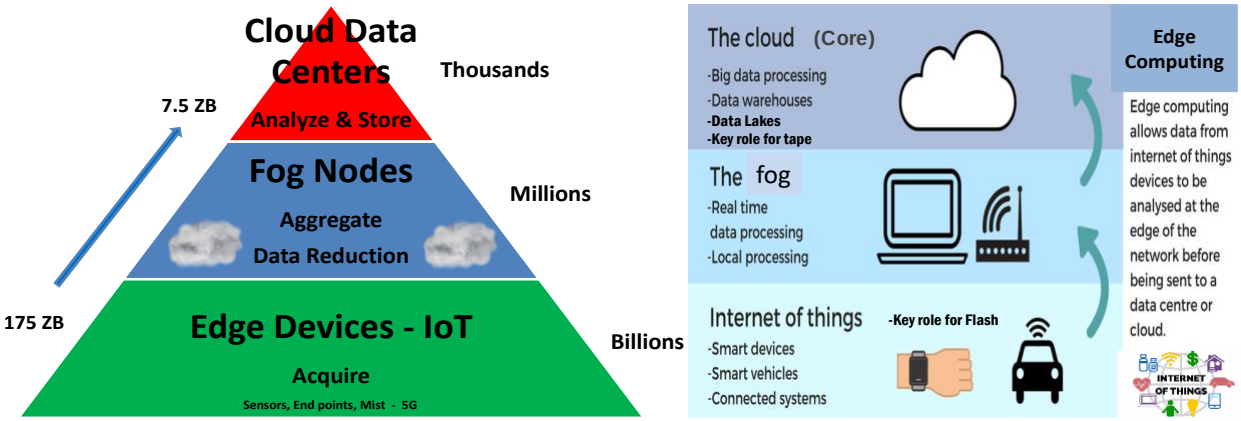
Since the first tape drives appeared in the early 1950s, tape has primarily served as a backup and archive device for HDDs. Times have changed as the magnetic tape industry has successfully re-architected itself to addresses many new, high growth cloud centric applications which are generating massive amounts of unstructured and archival data. IDC calculates that within five years, data intensive IoT devices, sensors and cameras will be creating as much as 79.4 zettabytes of data. Though most of this created data is transient or short-lived and won't ever be stored, much of the remaining data that is actually stored will live in archival status for indefinite periods of time awaiting analysis. This bodes well for future tape growth as archival data is piling up much faster than it is being analyzed.

The Cloud Ecosystem is Transforming the Storge Landscape

The cloud, fog, edge and the IoT (the cloud ecosystem) are altering the way data is being generated, processed, and stored from anywhere on the globe. The *cloud* is becoming the new core data center and offers a landing zone for backup, disaster recovery and a staggering amount of archival and big data. *Fog* computing is an architecture that performs a substantial amount of computation and data reduction before moving data into the cloud. The *IoT* (Internet of Things) refers to the billions of physical devices on the *edge* that are connected to the internet, all using sensors to ingest and create enormous storage and compute demand for on-premise and cloud environments. The lifecycle of data from birth at sensors on the edge to the cloud is becoming the new data creation model. The looming challenge of storing and protecting this much data for long periods of time is forcing cloud providers to make significant investments in more cost-effective tiered storage and tape solutions. The future role that LTO will play as a key component in the overall cloud ecosystem is becoming more evident every day and represents the next frontier for tape.

The Edge, the Fog and Cloud

Computing and Storage Span Centralized Nodes to the Logical Extremes of a Network

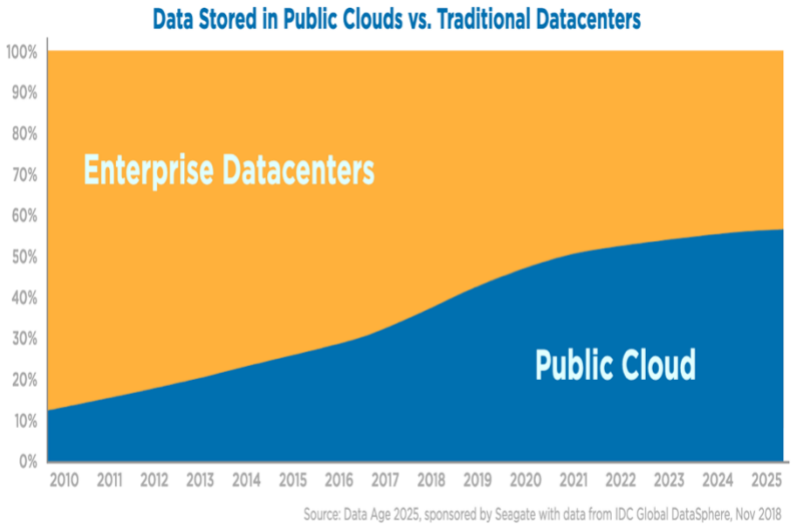


Source: Horison, Inc.

- **The Fog quickly aggregates/reduces IoT data before it reaches the cloud.**
- **Any device with computing, storage, and network connectivity (Hyper-converged) can be a Fog node.**
- **Most IoT data will be processed/reduced before being sent to a data center – new cybersecurity challenges!**

The Cloud

CSPs (Cloud Service Providers) operate data centers which are often enormous warehouse-like structures, often using millions of HDDs, SSDs and tape cartridges to store petabytes and even exabytes arriving via the internet from all over the world. The installed global storage capacity for public cloud compared to enterprise data centers is growing quickly and is projected to reach 50% of all installed storage capacity by 2025 (see adjacent chart).



It is important to note that clouds serve as a centralized home for many smaller infrastructures which take advantage of the benefits of compute and storage resources at scale. CSPs deal with data storage environments at magnitudes barely imaginable by most organizations and heavily emphasize infrastructure issues such as power consumption, carbon footprint and floorspace. Fortunately, LTO's advantages increase at scale, where its low acquisition cost, extremely low power consumption, TCO and ROI give it enormous advantages for long-term storage over an architecture built exclusively with HDDs. The economic benefits of using a tiered storage with tape provides the most cost-effective storage infrastructure available for CSPs.

Fog Computing

Transmitting data from IoT edge devices to cloud or core data centers becomes costly and time consuming in terms of network bandwidth and latency. Fog computing is a node architecture that serves numerous edge devices and sensors to carry out a substantial amount of computation, data reduction, interim storage, and communication before moving data from the edge to the core over the internet backbone. The fog nodes can be located anywhere between the data source and the cloud. Storing small amounts of data in the cloud is straightforward. Large data sets and blobs however, are highly dependent on how much available internet bandwidth an organization has, therefore data reduction in the fog is highly desirable as there is simply too much data trying to get to the cloud via too few pathways. The fog infrastructure is architected to filter only the data relevant to support specific processes or tasks.

Edge Computing and the IoT

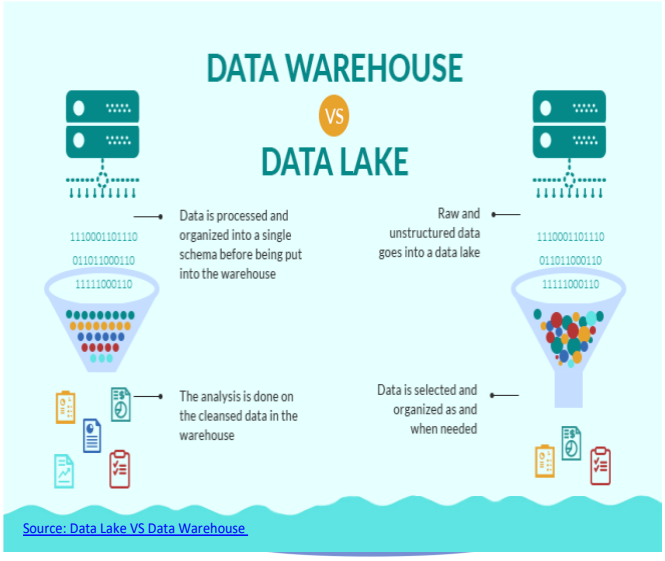
At its basic level, the edge brings computation and data storage closer to the IoT sensors where data is being gathered, rather than moving the data to the fog or a central location that can be thousands of miles away. This is done so that real-time data does not suffer long latency issues impacting an application's performance. For example, autonomous cars use edge computing to make the time-critical decision about stopping a self-driving car in real time. A remote weather station will use sensors to measure temperature, wind speed, humidity and barometric pressure transmitting data that may be used immediately for safer flight navigation. Security cameras might have a proximity motion sensor using sound and video to initiate an immediate response. Some of this IoT data will be short lived and erratic, other sensors might create huge amounts of streaming data traffic, like a video surveillance camera. Edge computing and IoT sensors will greatly benefit from the faster performance with the arrival of 5G wireless transmission which can reduce round-trip-time by up to two orders of magnitude for many critical internet applications. For example, the global video surveillance storage market is expected to grow from $7.5 billion in 2020 to $10.2 billion by 2025 while generating much of its data at the edge. A typical autonomous vehicle (an edge device) produces approximately 40 TB of data every 8 hours it is used. This rising tide of IoT data presents immediate bandwidth and storage challenges *and* enormous storage opportunities.

Data Lakes, Swamps and Oceans

A data lake is part of the cloud-fog-edge-IoT ecosystem serving as a centralized storage reservoir. Data lakes normally store newly ingested IoT and edge data at any scale as unstructured data in its native or raw format without incurring the overhead of adding structure or descriptive metadata. The data lake often serves as a data archive containing data awaiting future analysis and can foster data overindulgence. Having too much unorganized data creates complexity and can create a *data swamp* lacking the necessary context to make it useful. You can build a smart data lake by applying metadata at a later time making it easier to derive business insights and value. Since data lakes contain vast amounts of archival data, using tape libraries for data lakes becomes compelling from a TCO, ROI and security perspective as tape can deliver unlimited capacity scaling by simply adding cartridges rather than adding more expensive HDDs. A data lake can physically reside in the cloud, or on premises within an organization's data center. If your data lake grows and becomes a *data ocean*, consider the largest tape libraries will be able to contain over 55,000 LTO-9 18 TB cartridges (when LTO-9 becomes available) storing over an exabyte of data.

Data Lakes – A Reservoir for Future Use

A data lake is a large storage repository that holds a vast amount of raw data in its native file or BLOB format until it is needed (cold data).
Data lakes are often distributed over multiple nodes rather than in a fixed, structured data warehouse.



DATA WAREHOUSE	vs.	DATA LAKE
structured, processed	DATA	structured / semi-structured / unstructured, raw
schema-on-write	PROCESSING	schema-on-read
expensive for large data volumes	STORAGE	designed for low-cost storage **Optimal for tape**
less agile, fixed configuration	AGILITY	highly agile, configure and reconfigure as needed
mature	SECURITY	maturing
business professionals	USERS	data scientists et. al.

BLOB - A Binary Large Object (Unstructured)

Tape Storage in the IoT Environment

Although many CSPs initially established their storage services based entirely on HDDs, it has not taken long for them to recognize a critical need for a much less expensive, reliable and more secure storage solution. Edge devices and their applications generate massive amounts of data, filtered and reduced in the fog, that can quickly become historical and may be stored untouched for years in a data lake or the cloud awaiting future analysis. This archival data pile up has hastened the use of LTO in many cloud ecosystems to reduce costs, address cybercrime threats with offline storage via the tape air gap, and to preserve and discover the untapped value of archival data. However, even after fog reduction, the remaining amount of data stored is CSPs is staggering and must be securely stored and kept available for future reference or analysis. Evolving in-synch with the cloud ecosystem, the magnetic tape industry has successfully re-architected itself delivering compelling technologies and functionality including steady cartridge capacity increases with 2.5x compression, the lowest TCO, a higher ROI, a much longer replacement timeframe than HDDs and faster data transfer rates than any previous tape or HDD. Far from being "either/or" technologies, cloud archives, data lakes and tape are very complementary. It is no surprise that cloud providers are embracing tape to build and expand their archive services. Organizations that reflexively dismiss tape are missing an opportunity to markedly optimize their storage infrastructure, especially on a cost-effectiveness basis. The LTO Value Proposition chart for the cloud ecosystem below highlights many of tape's benefits.

LTO Value Proposition

Function	Key Benefits of LTO for the Cloud Storage Ecosystem
Price/TCO/ROI	Tape Has the Lowest Acquisition Price ($/TB), Lowest TCO and Highest ROI of any Storage Device.
Access Time and Performance	Much Improved - Active Archive, Fastest Data Rates, RAIT, Smarter and Faster Robotics, Time to 1st Byte Features RAO, TAOS, and LTFS Reduce Access Times.
Capacity	LTO-9 Cartridge Capacity Max. @18TB (45 TB compressed) with ~200x More Surface Area Than an 18 TB HDD. (Note: when LTO-9 becomes available.) Lab Demos Indicate Future Cartridge Capacities Can Reach up to 580 TBs.
Data Density (Floorspace)	The Largest Capacity Robotic Libraries Can Store an Exabyte ($1x10^{18}$) in 2.558 PB per sq. ft. (391 sq. ft.) Using 55,900 LTO-9 18 TB Cartridges.
Scalability	Tape Easily Adds Capacity by Adding Media, HDDs Add Capacity by Adding Drives. Some Tape Libraries Can Scale Beyond an Exabyte Native Capacity.
Data Integrity and Durability	Tape Uses a Read-While-Writing Process That Checks Data as it is Written on Tape. HDD Writes Require Full Verify Essentially Resulting In a 50% Data Rate Reduction. Durability and Replacement Timeframe Exceeds HDDs and SSDs.
Carbon Footprint	Tape Has the Lowest Carbon Footprint of any Storage Solution.
Portability	Tape Media is Easily Portable During Electrical Outage or Site Disaster, HDDs are More Difficult to Move.
Cybersecurity	Tape Air Gap Prevents Cybercrime Attacks, Strong Defense Against Malware.

Summary

Tape has the greatest potential of any existing storage technology to meet the data growth demands that are fueling the cloud era. Tape technology will not replace HDDs or SSDs, but it has significantly expanded its position as a highly cost-effective complement to SSDs and HDDs for the foreseeable future due to its high reliability, higher capacities, faster data rates, security, low cost, easy scalability and significantly lower carbon footprint. The role tape serves in today's modern data centers is quickly expanding and tape momentum will increase as data growth continues on an explosive trajectory across many new applications and workloads pushing clouds and many hyperscale and hyperscale-lite data centers to take advantage of tape's benefits.

LTO is well balanced between cost of technology and market requirements for steadily increasing capacity by keeping the cost of capacity low while continually improving the density of data. This is expected to continue for the foreseeable future. These steady and rich technology advancements have set the stage for LTO to be the most cost-effective storage solution for the enormous storage challenges that the next frontier presents, whether on-site, at a remote location, on the edge, or in the cloud.

Sponsored by the LTO Consortium

Linear Tape Open (LTO) Ultrium is a high capacity tape storage solution developed and continually enhanced by Hewlett Packard Enterprise, IBM and Quantum and promoted by the LTO Program. Linear Tape-Open LTO, the LTO logo, Ultrium and the Ultrium logo are registered trademarks of Hewlett Packard Enterprise, IBM and Quantum in the US and other countries.

About the author

Horison Information Strategies is a data storage industry analyst and consulting firm specializing in executive briefings, industry seminars, market strategy development, whitepapers and research reports encompassing current and future storage technologies. Horison identifies disruptive and emerging data storage trends and growth opportunities for end-users, storage industry providers, and startup ventures.