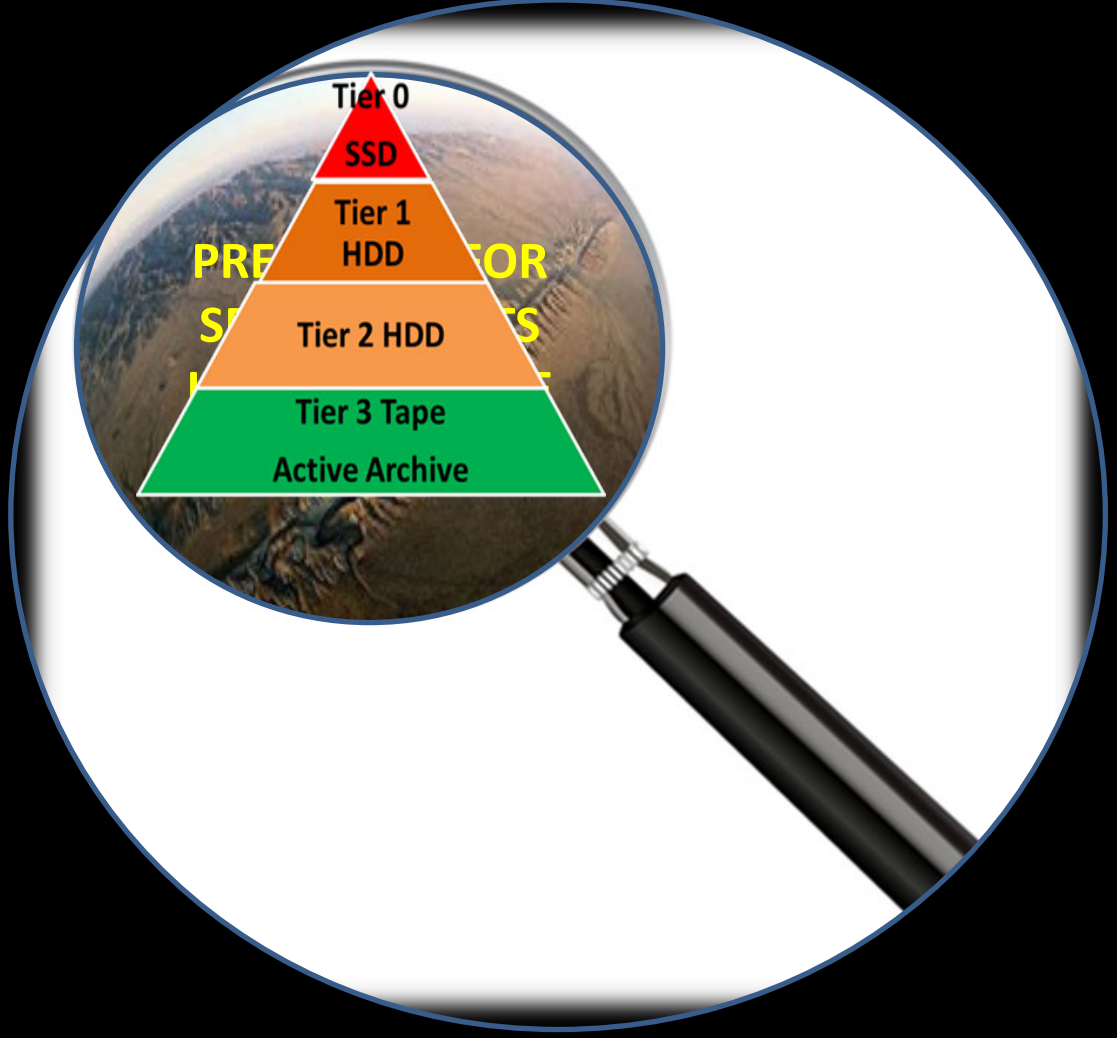
STORAGE OUTLOOK



Fujifilm 11th Annual Global IT Executive Summit

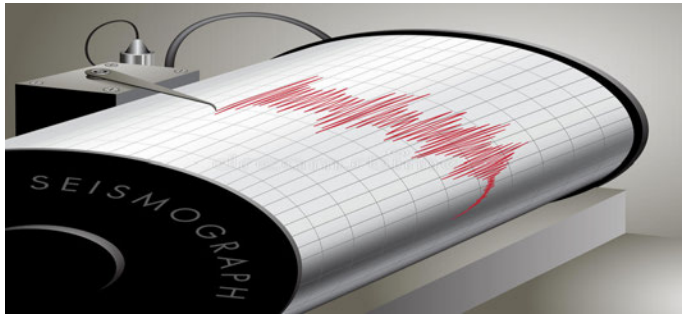
Oct. 19-22, 2019

San Francisco

HORISON
Information Strategies

Fred Moore
President
Horison Information Strategies
Horison.com

Seismic Shifts in the Storage and IT Industry

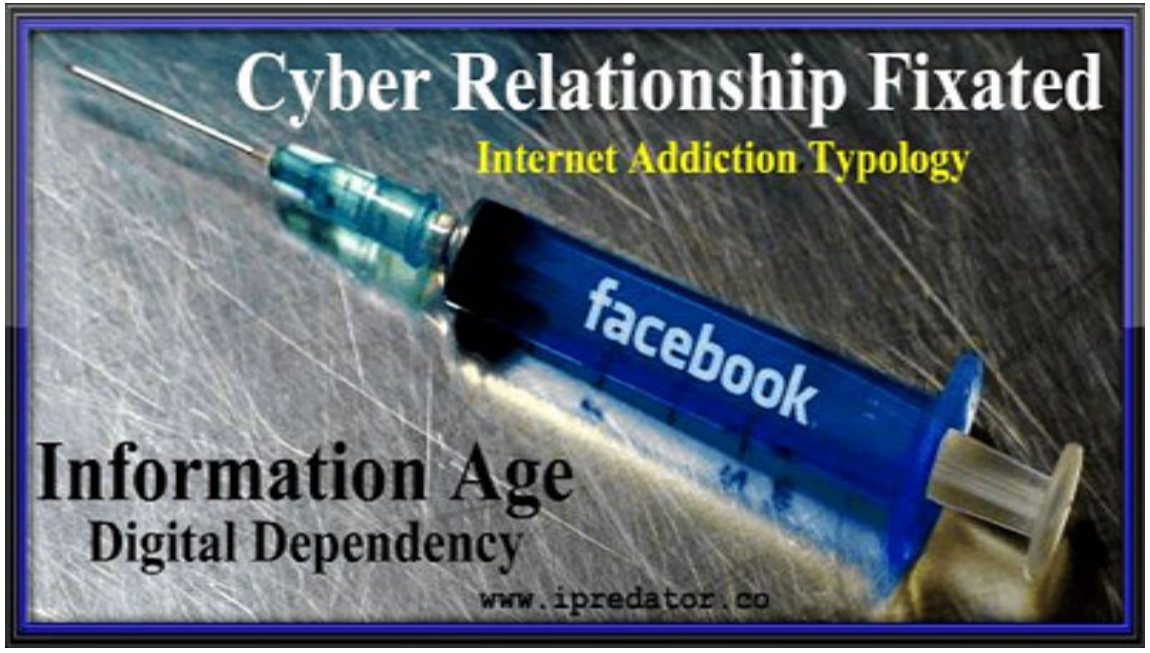


The word *shift* describes a disruption, something taking a new and significant direction, either physically or in thought. Shifts that have made fundamental changes include the Internet, iPhone, Uber and Airbnb.

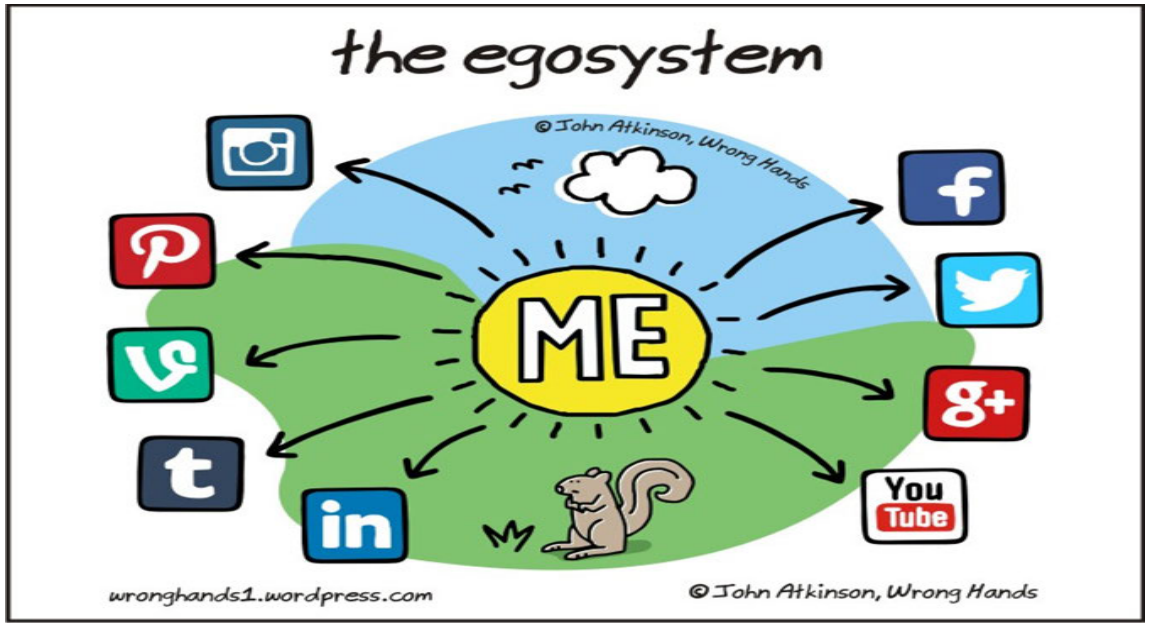
Key Storage Shifts:

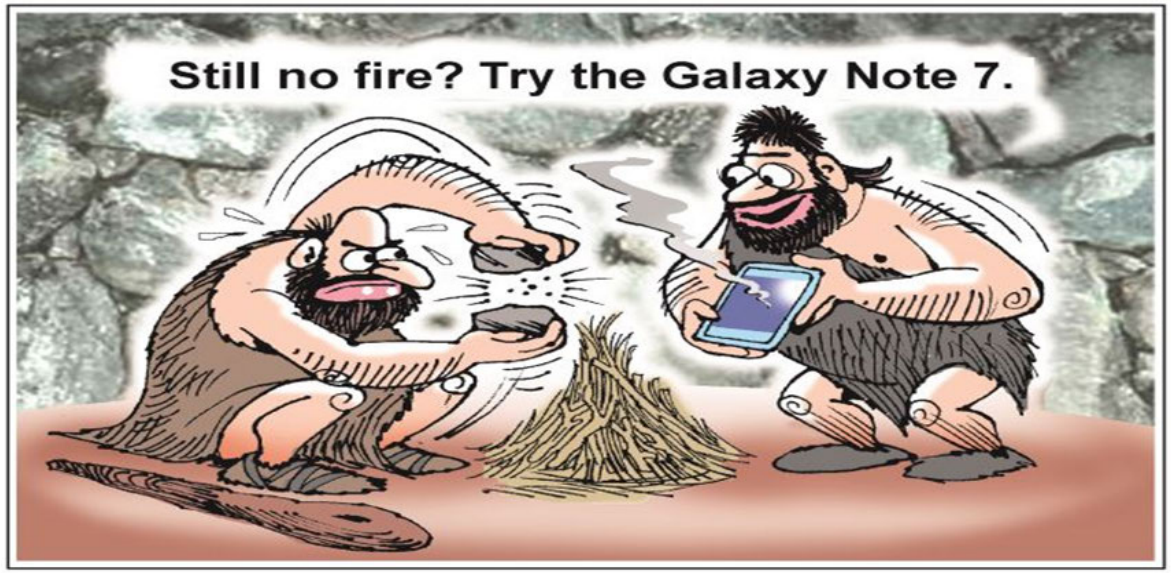
- **From many smaller data centers - to fewer - but much larger data centers (Hyperscale)**
- **From technology focus to data focus**
- **From HDDs to SSDs for primary storage**
- **From RAID to erasure coding as disk storage pools increase – *both* have limitations**
- **From storage CAPEX to OPEX (TCO) accelerates**
- **From centralized computing to the logical extremes of the network – the IoT (~25B nodes by 2020)**
- **From human to ML and AI based decision making**
- **To colder storage solutions as IT consumes ~7% of global electricity, forecasted to be 13% by 2030**
- **From security being an IT problem - to everyone's problem**
- **In social media from photograph collections to digital dependency**

Speaking of Digital Dependency…



iToons — Sunil Agarwal & Ajit Ninan





The Storage Hype Cycle

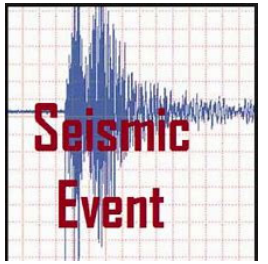
Where are These on the Hype Cycle?

Expectations

Peak of Inflated Expectations

Slope of Enlightenment

Plateau of Productivity



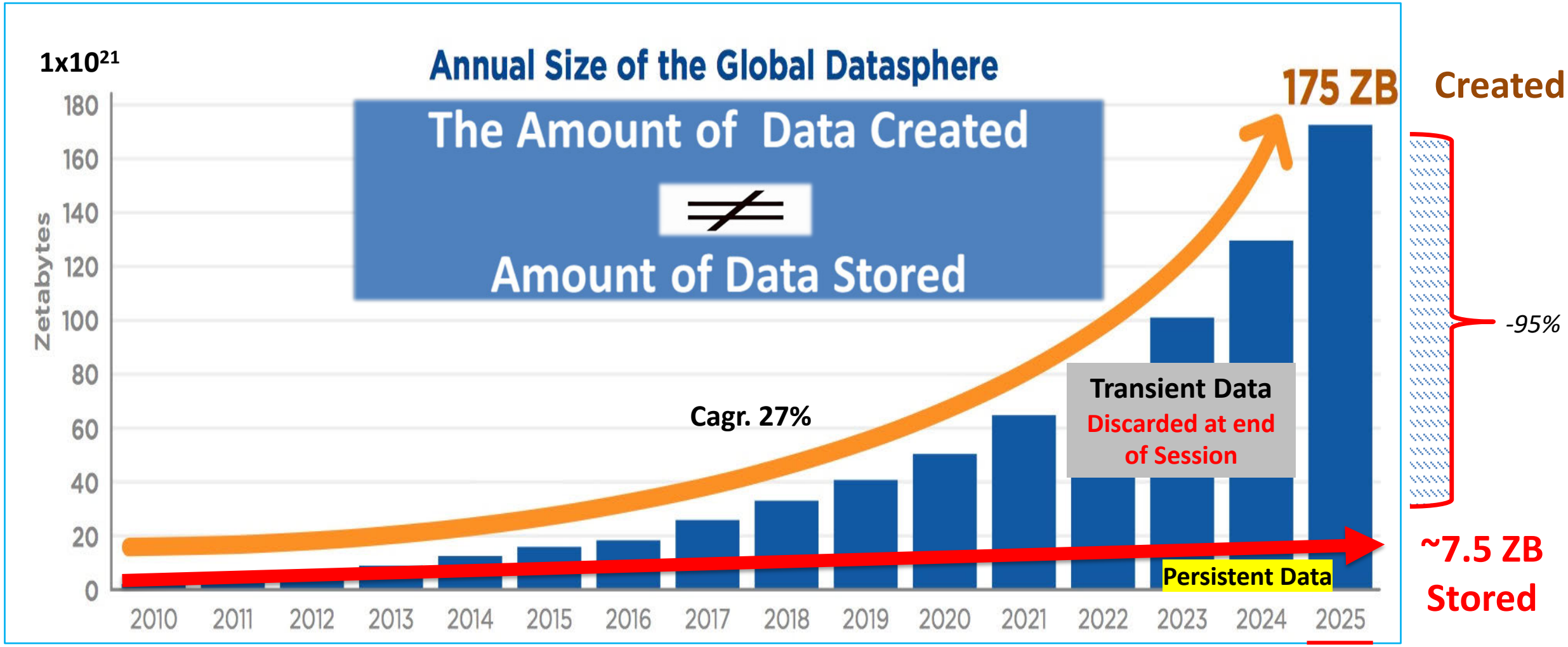
Innovation Trigger

Trough of Disillusionment

Time

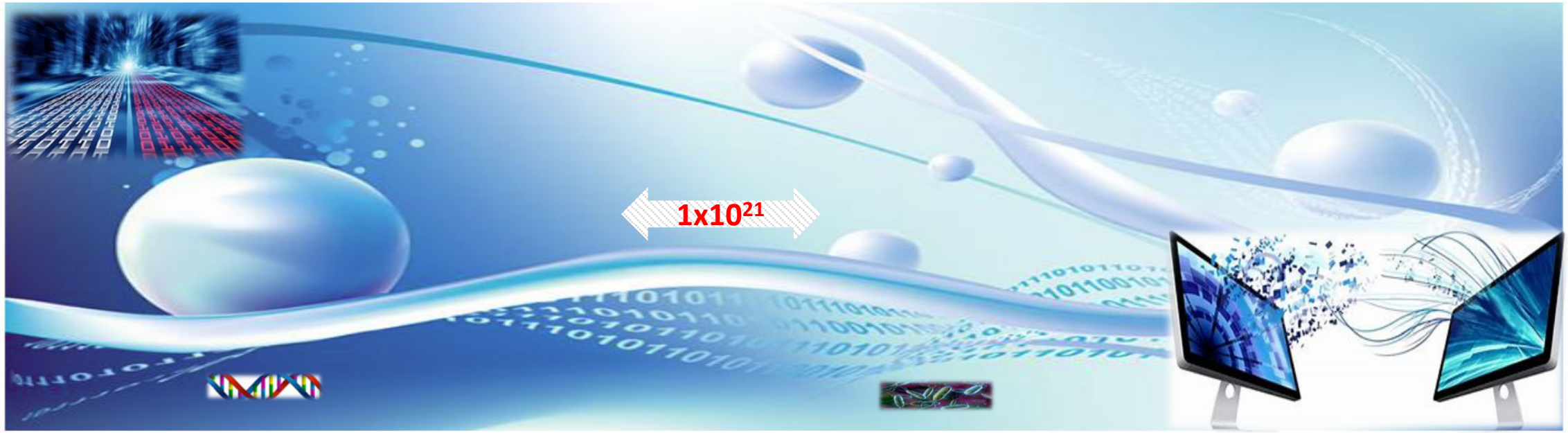
Global Datasphere Expansion is Never-ending

The Zettabyte Era Arrives – *However…*



Source: https://www.storagenewsletter.com/2018/11/28/global-datasphere-from-33zb-in-2018-to-175zb-by-2025/

The Zettabyte Era – How Big Is It?



One Zettabyte

- **Would hold the amount of data created by every living person on Earth "Tweeting" continuously for 100 years.**
- **Would fill 57.5 billion 32GB Apple iPads or 250 billion DVDs.**
- **Build the Great iPad Wall of China - at twice the average height of the original - 13,170 miles.**
- **The brain capacity of the world's first two hyper - intelligent humans.**
- **Build a 20-foot high wall around South America - 89,829.64 miles.**
- **The number of molecules in the original E. coli strand.**
- **Would fill 83.33 million LTO-8 (12 TB) cartridges.**

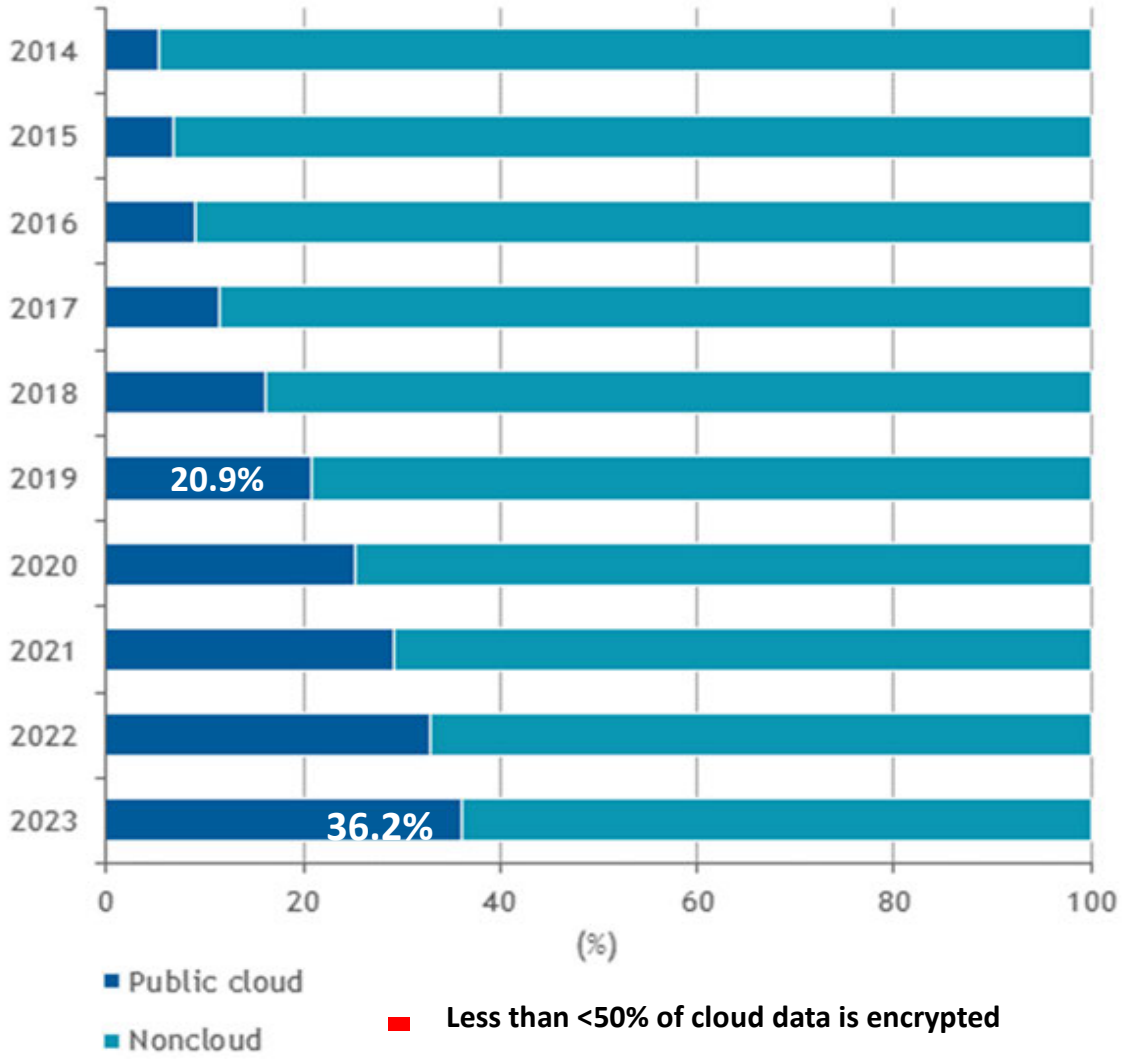
Global Datasphere Installed Capacity

Installed Capacity by Device Type

	PBs 2019	2021	2023	CAGR (%)
HDD	3,799,076	5,097,752	6,796,084	15.5
Tape	841,868	1,272,414	1,945,058	*23.2
Optical	521,927	507,116	501,076	-1
SSD	341,024	741,141	1,406,838	*44
NVM - NAND and NVM-other	369,447	660,543	1,029,060	*29.8
Total PB Installed	5,873,342 5.873 ZB	8,278,966 8.278 ZB	11,678,116 11.678 ZB	18.4

- SSD, NVM and Tape Fastest Growing Market Segments.
- HDD Remains Installed Capacity Leader at 65%.
- Hyperscale Driving Tape growth.

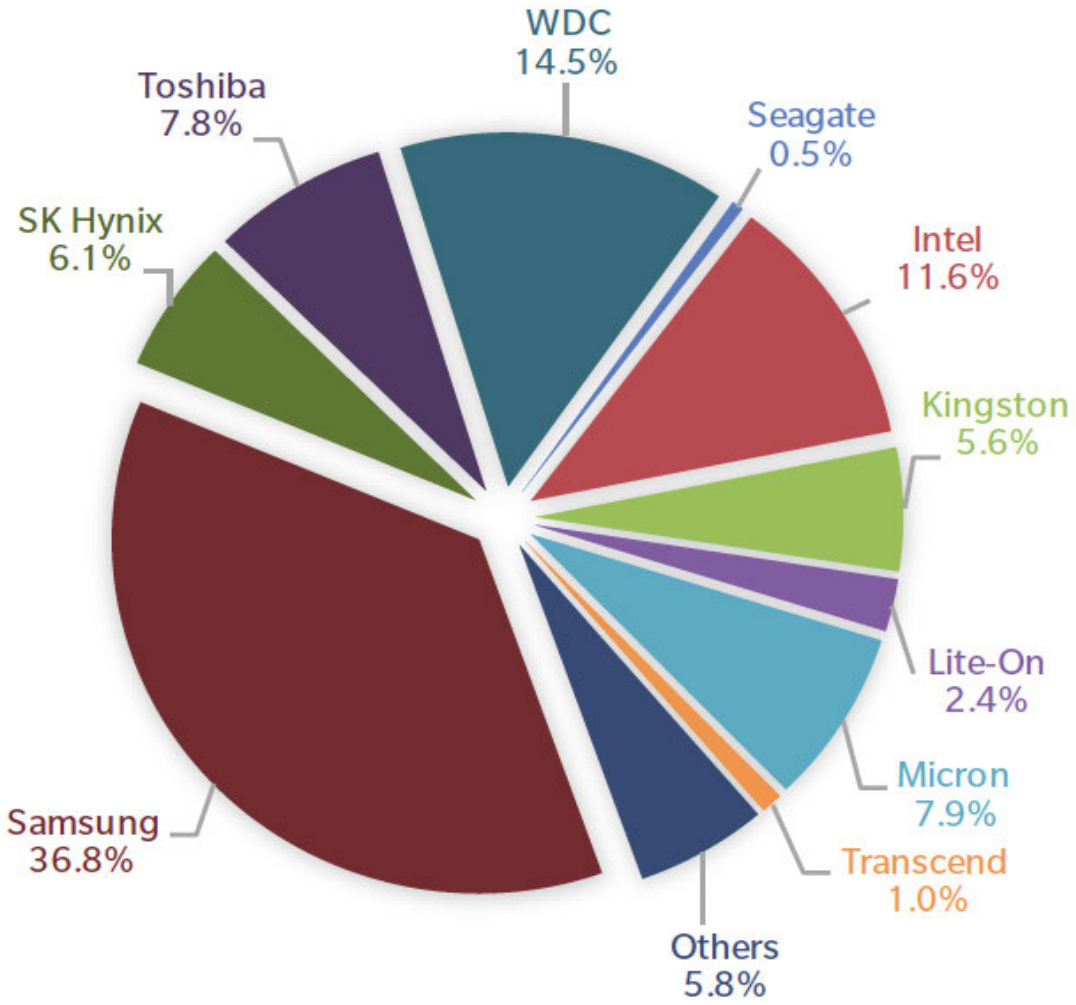
Installed Capacity by Cloud and Non-cloud



Less than <50% of cloud data is encrypted

Source: IDC Global Datasphere Report

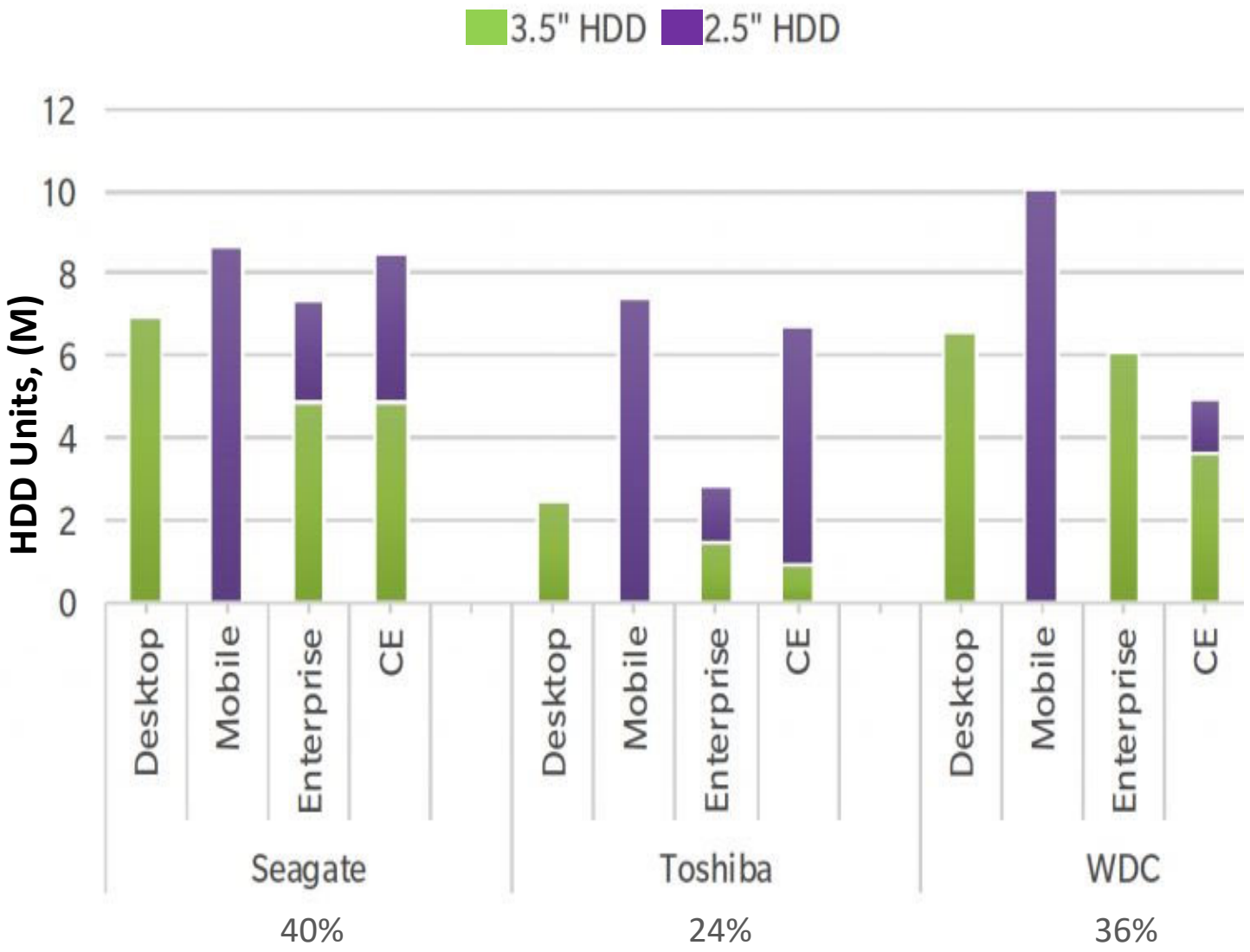
WW SSD Shipments - 2Q 2019



- **2Q 2019 SSD shipments rise mainly on Hyperscale demand and Cloud growth.**
- **NAND oversupply resulting in huge price reductions.**
- **3D NAND percentage grew to 90.5% of total SSDs shipped.**
- **31.59 EBs shipped in 2Q.**
- **Annual run rate total - 126.4 EBs**

Source: Trend Focus

WW HDD Shipments - 3Q 2019



- **3Q 2019 HDD shipments rise mainly on HSDC demand and Cloud growth.**
- **~83 Million HDDs shipped in 3Q. (Peak of 651M in 2010)**
- **240 EBs shipped in 3Q.**
- **Annual HDD run rate total - 960 EBs.**

Source: Trend Focus

Disk Utilization Profile

Why Are Storage Management Improvements Needed?

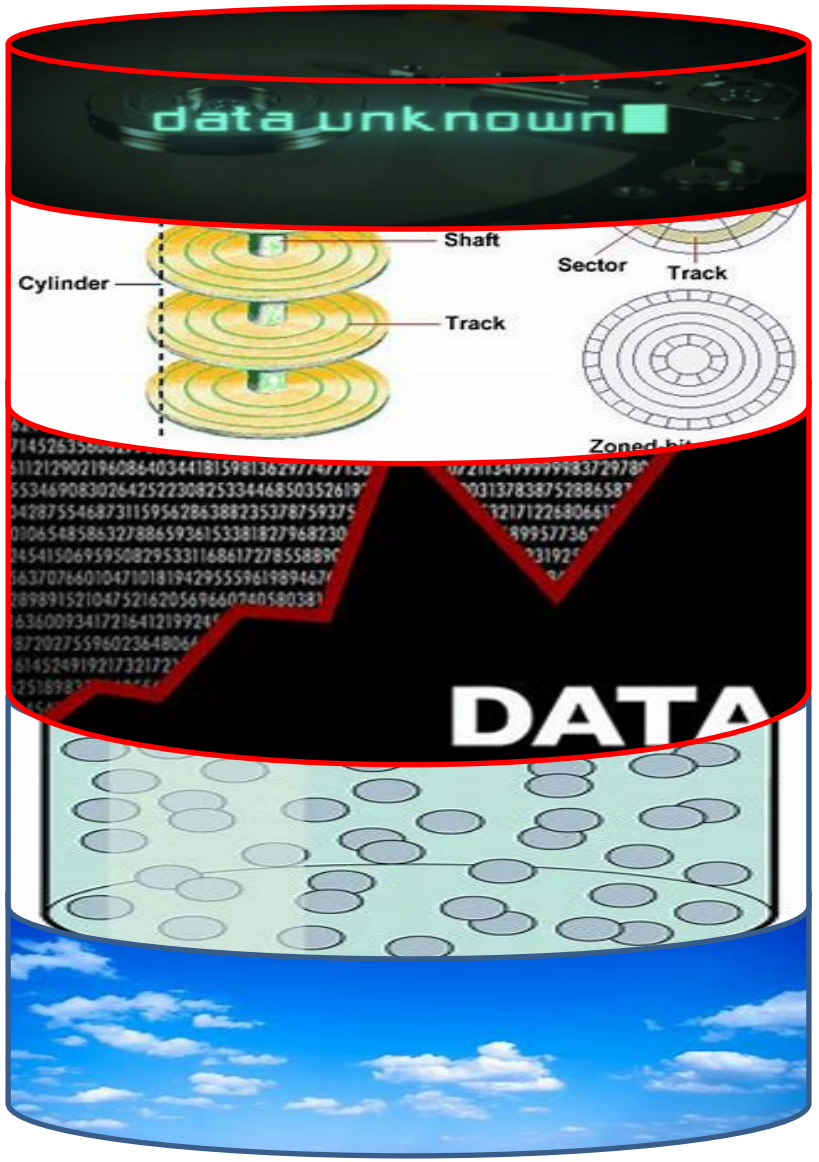
Invalid, Orphaned, and Unknown Files
5%

System Overhead, RAID, ECC, Control Fields
5%

Allocated and Used Live Data (Files, Blocks, Objects)
50%

Allocated and Unused (Gas - Over Allocated)
15%
Thin Provisioning

Unallocated and Unused – (Available Free Space)
25%



Data

Used Capacity
60%

- Allocated space often reduced to improve performance and reduce arm contention.
- Over 60% of live data is seldom accessed.

No Data

Unused Capacity
40%

- Available for future use.

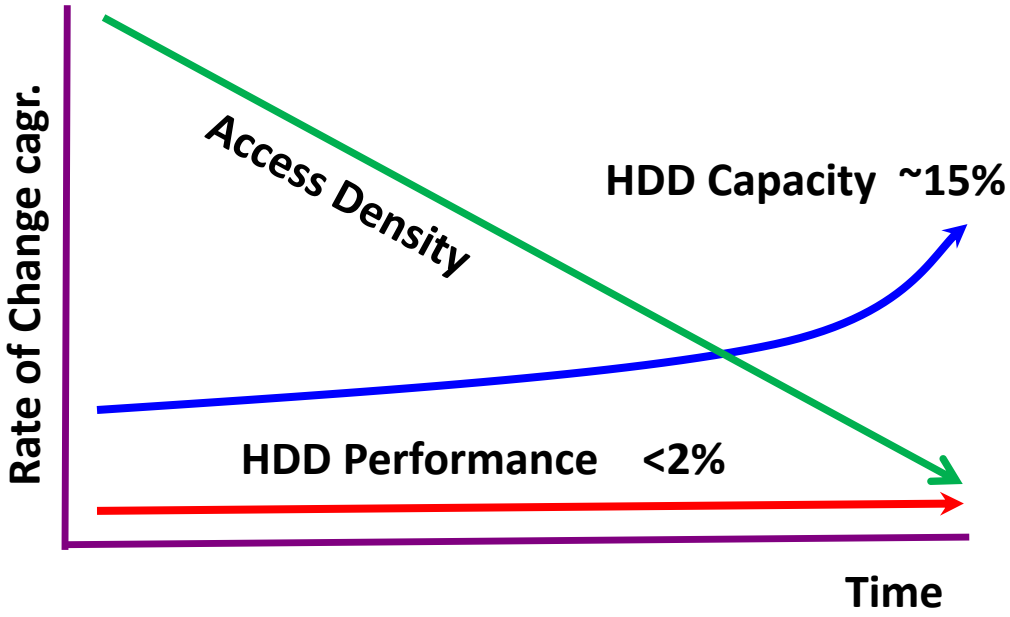
Source: Horison, Inc.

Can HDD Utilization Improve?

Performance Gains are Negligible for HDDs

$$\text{Access Density} = \frac{\text{IOPS}}{\text{TB}}$$

IOPS @ 10ms	HDD Cap. TB	Access Density
100	1.0	100
100	4.0	25
100	8.0	12.5
100	16.0	6.25



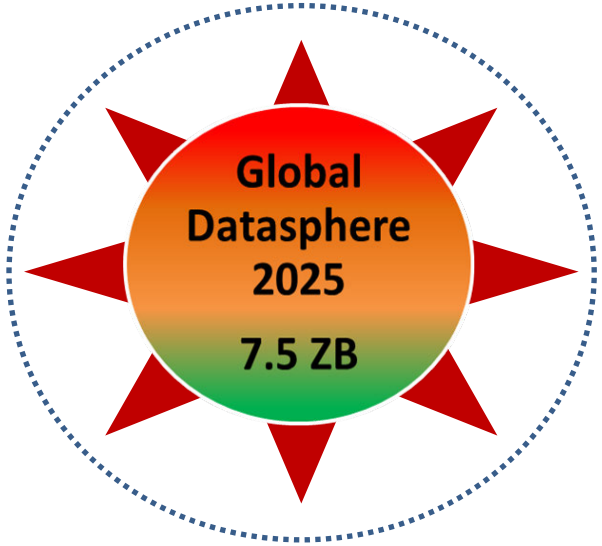
- HDD Performance (Speed) ***Not*** Scaling With HDD Capacity Growth or Server Speed
- Future HDD Performance Gains are Minimal - <u>if Any</u>
- Access Density Will Continue to Decline as HDD Capacity Increases
- Creates Additional Demand For SSD/NVM (Tier 0) Systems
- Results in HDD Capacity Reductions to Maintain Performance (Short Stroke) – Or Less Active Files
- Utilization ***Unlikely*** to Improve Without New Architecture and Storage Management Methods

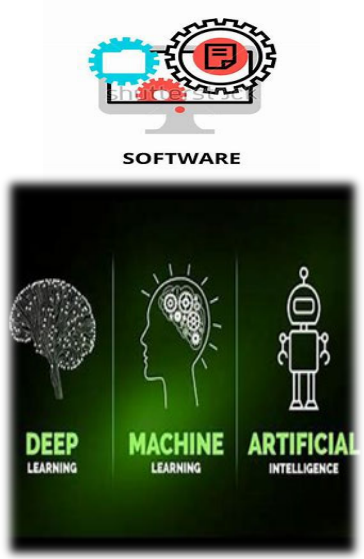
Source: Horison, Inc.

Global Datasphere by Data Class and Storage Tier

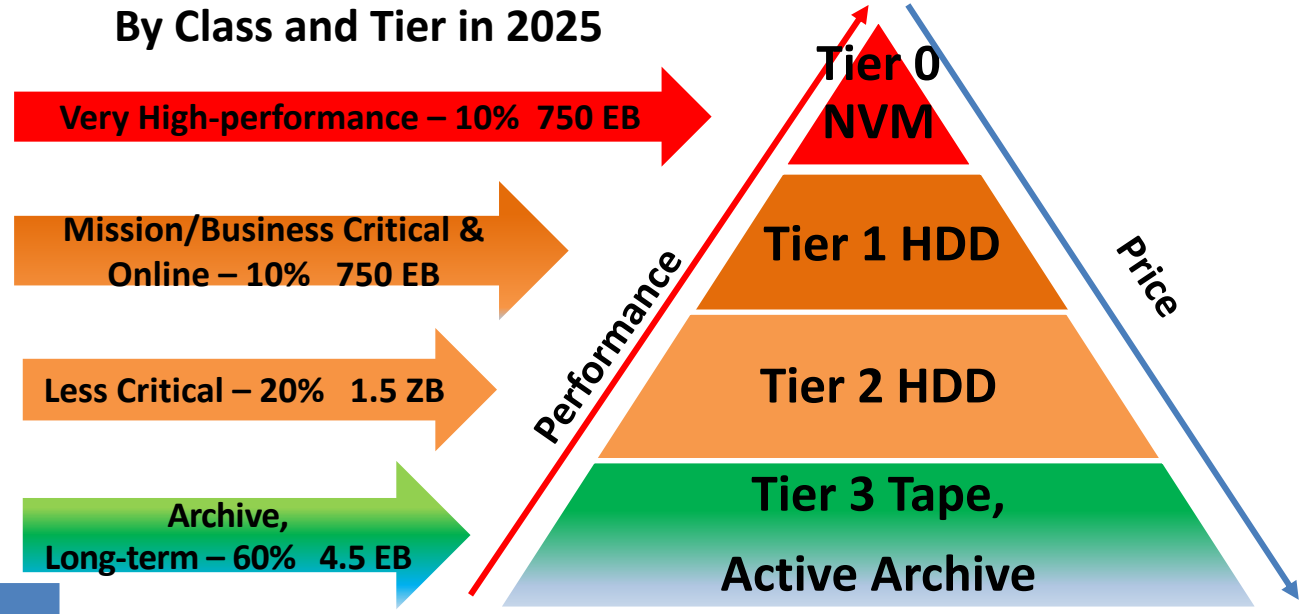
WW Digital Stored Data in 2025 – IDC est.

Security Force Field

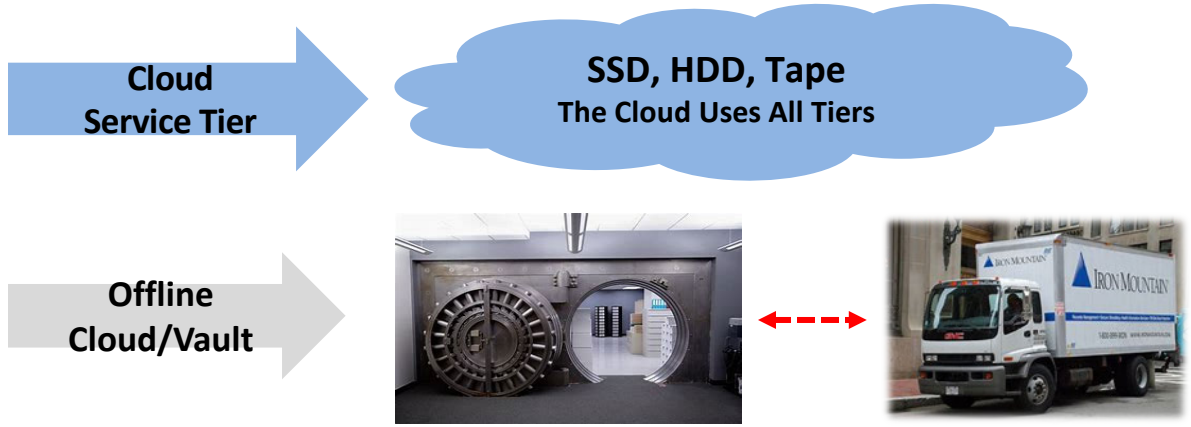




7.5 ZB If Optimally Stored By Class and Tier in 2025



Key - AI Targets for Storage Management Optimization	
Data Allocation	Allocate to best meet SLA's & manage space
Migration/archive	Move data to most cost-effective tier
Availability	Move critical data to hi-availability storage
Performance tuning	Move data to optimal tier to meet changing performance and response-time objectives
Deletion & eradication	Delete obsolete data, eradicate obsolete media
Predictive self-healing	Diagnose problems by comparing with prior errors



Source: Horison, Inc. The Digital Universe IDC

Inside the Storage Tiers

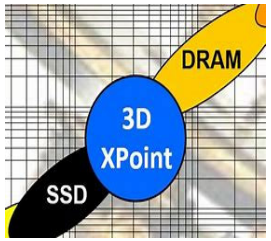
The Physical View – Technology Focus

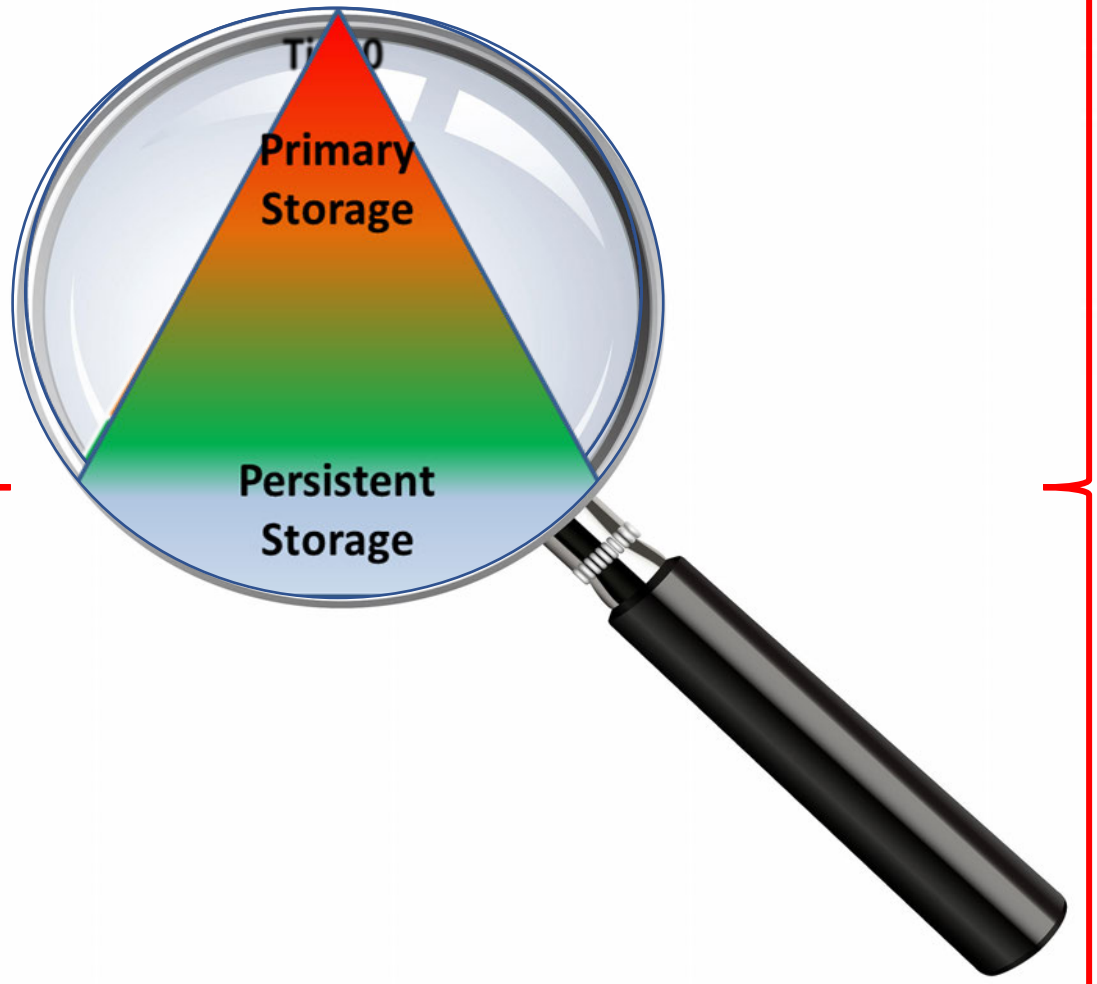
Source: Horison Information Strategies

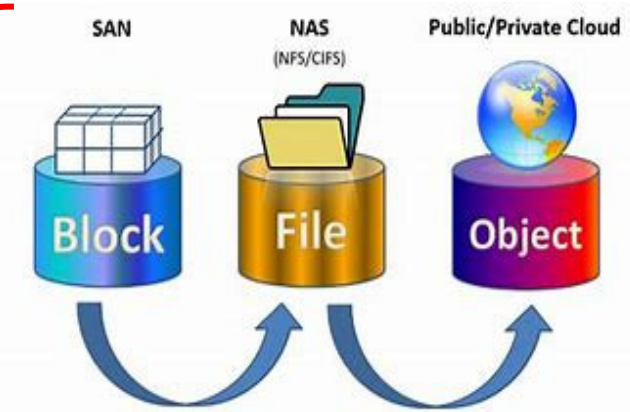
Shifting the Focus From Hardware to Data

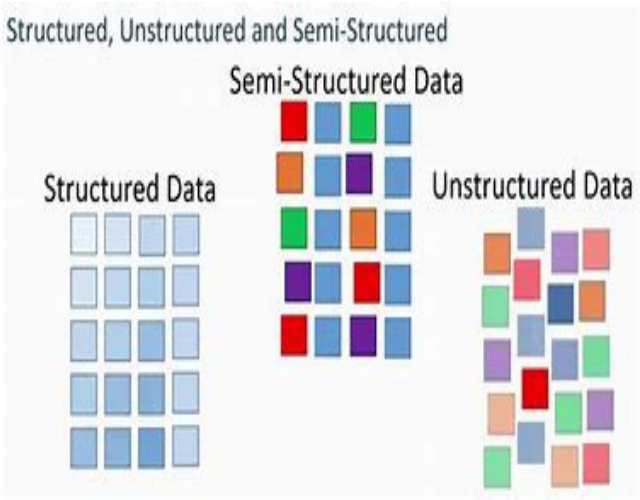
The Logical View of the Storage Hierarchy

From Hardware ⟶ **To Data**









Source: Horison, Inc.

Data Transformation Fueling Future Growth

Shifting to Higher Density Formats

Unstructured Data

- **Constantly Pushing Compute, Bandwidth and Storage Architecture Limits.**
- **New Formats, Architectures and Security Needed as Storage Density Increases.**

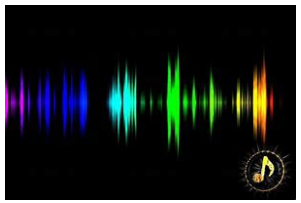
Traditional Data

58674322
98323456
20419335
49567053

Numbers:
5 KB / record

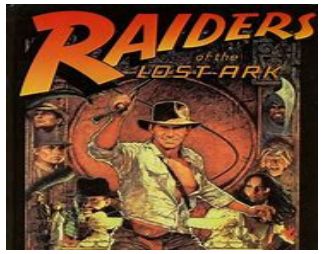
Tape continues to expand its offerings and reach and has been fueled by more than a decade of adv..

Text:
500 KB / record



Audio:
5 MB / song

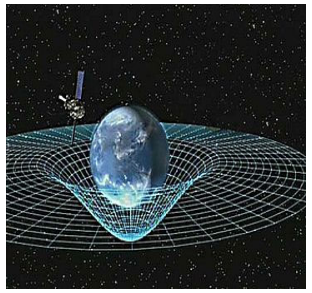
Images: 2D
2 MB / picture



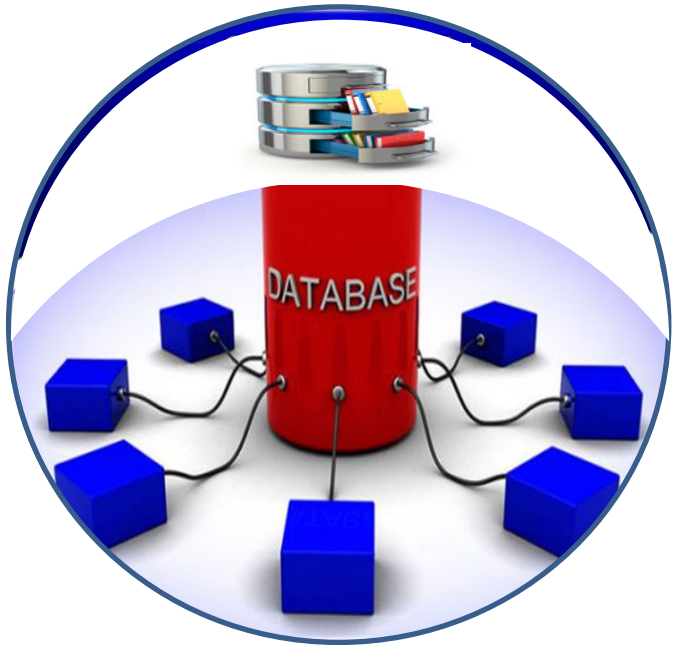
Video:
5 GB / movie

Higher-Res: 3D
50+ GB / object

4D..
Motion Vector:
100's GB / capsule



Structured Data



Source: Horison, Inc.

Structured and Unstructured Data Insights

Examples

Structured and Semi-structured data (SSD, HDD)

Highly organized, semantically tagged, formatted and easily searchable in relational databases. ***Amount beginning to increase*** with metadata & tags for "smart" archival data (Big Data analysis).

Structured Data

Unstructured Data

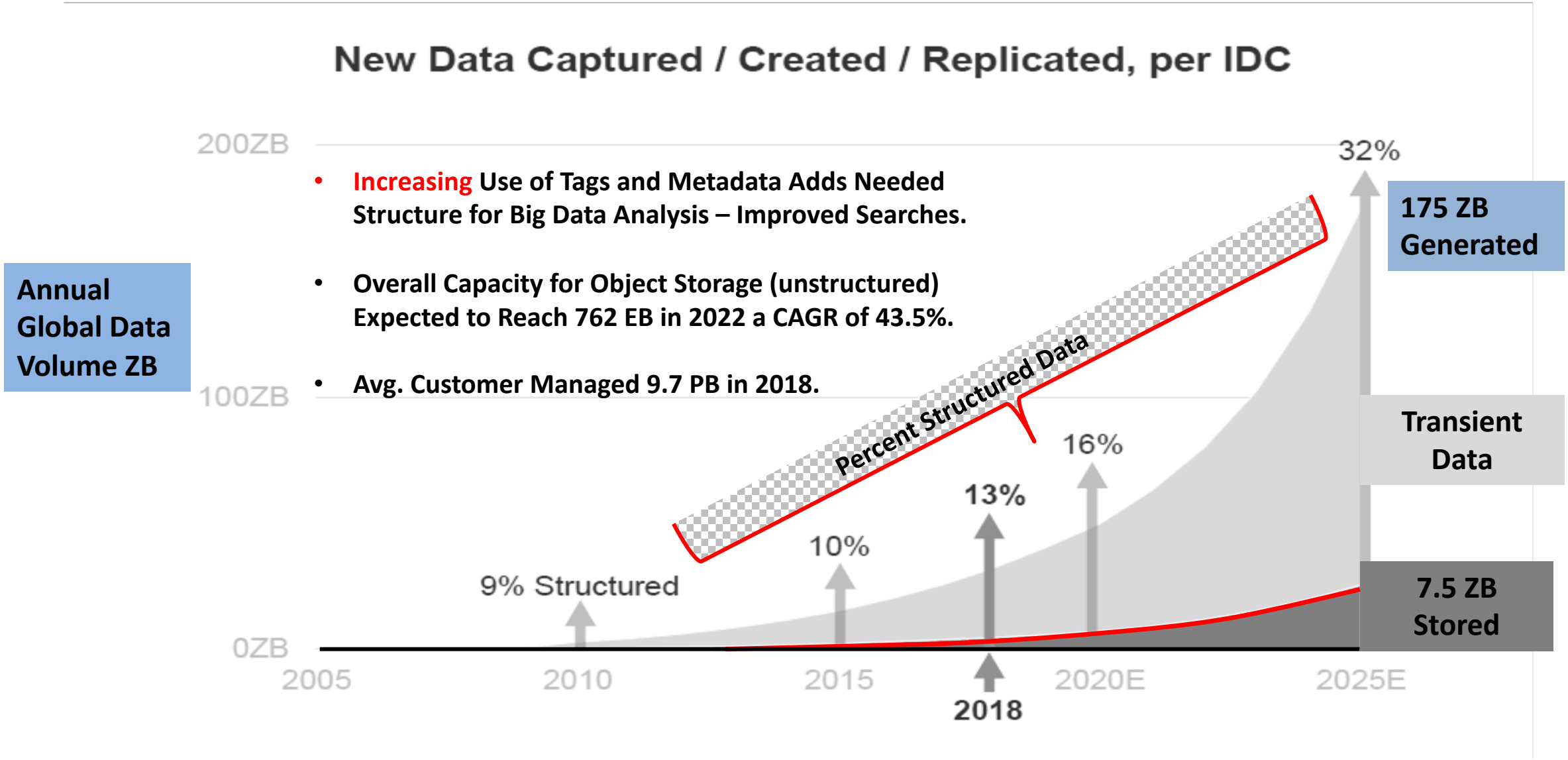
Unstructured data (Tape, HDD)

No pre-defined format or organization making it much more difficult to collect, process, and analyze. Unstructured data isn't suited to high IOPs or transaction processing applications.

Source: Horison, Inc.

The Shift to Structured Data

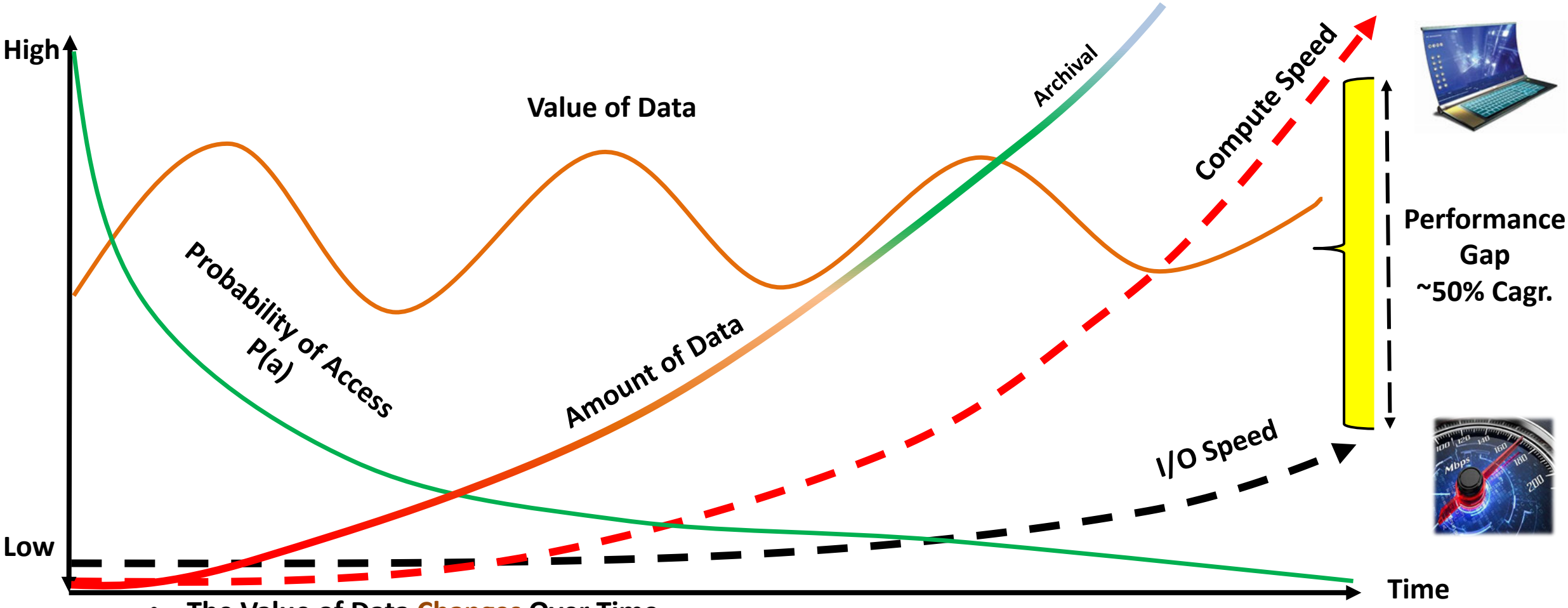
The Greatest Potential for AI is Unlocking Unstructured and Archival Data



Source: IDC Global Datasphere Report

Data Lifecycle Behavior

Understanding Access Patterns Optimizes Storage Management Over Time

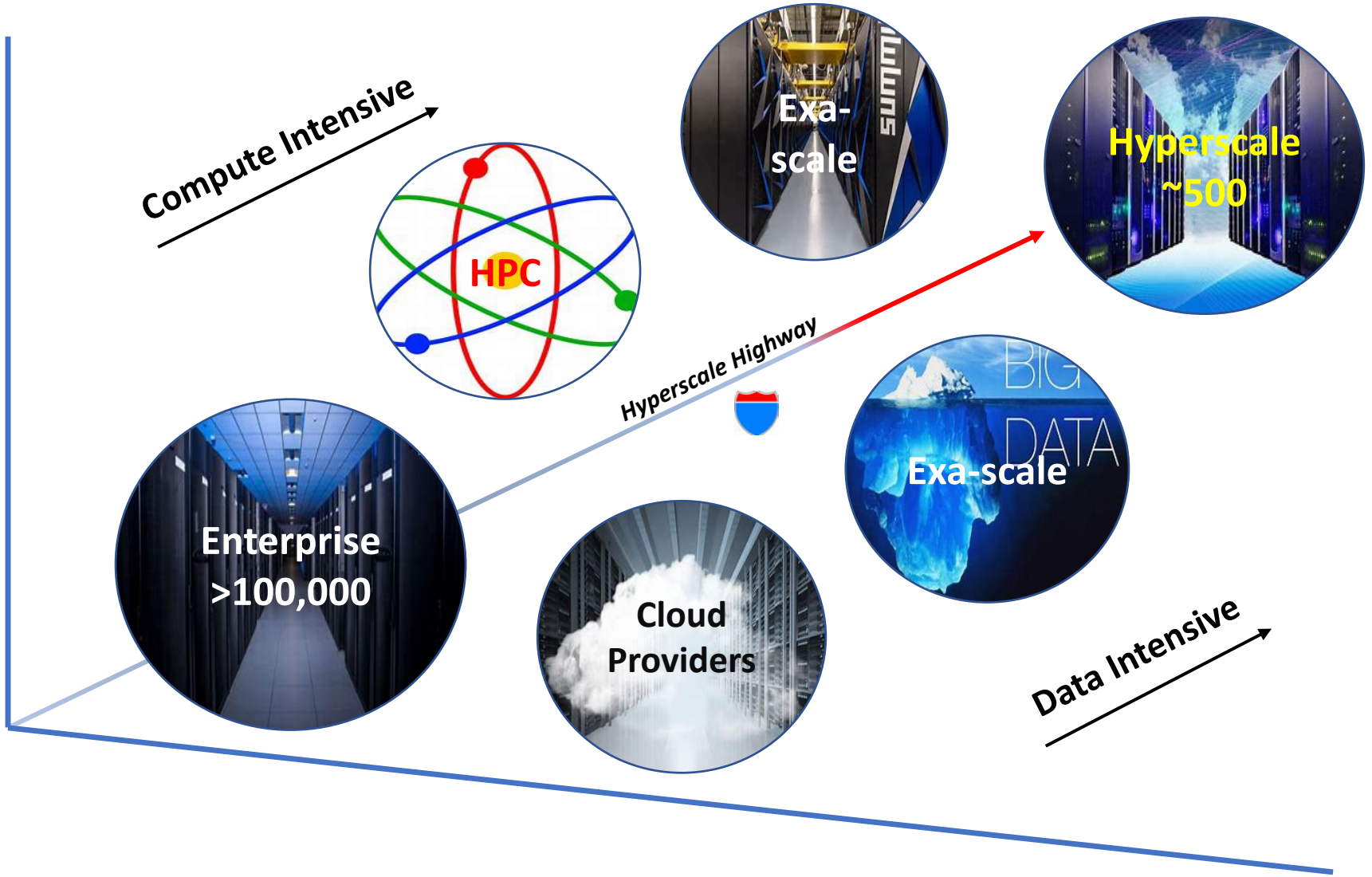


- The Value of Data Changes Over Time
- The Probability of Accessing Data Decreases Over Time
- The Amount of Stored Data Increases Over Time (~30% cagr.)
- Compute and I/O Speeds Diverging Over Time

Source: Horison, Inc.

The Ascent to Hyperscale

The Fastest Growing Data Center Segment

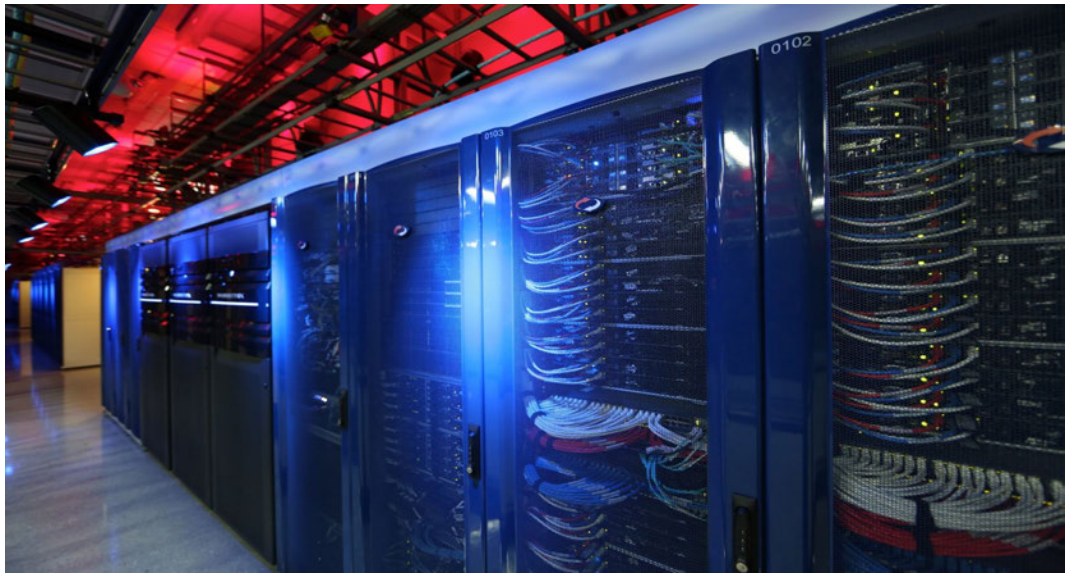


Hyperscale Data Center	Massive scaling
Hyper-scale lite	The next wave of Hyperscalers
Exascale Computing	one exaFLOPS, or ($1x10^{18}$)
Exascale Storage	One ($1x10^{18}$) exabyte of storage
HPC	Compute Intensive
Cloud	Private, public, hybrid
Enterprise	Large footprint, many apps

Source: Horison, Inc.

Hyperscale Data Centers Arrive - in a BIG Way

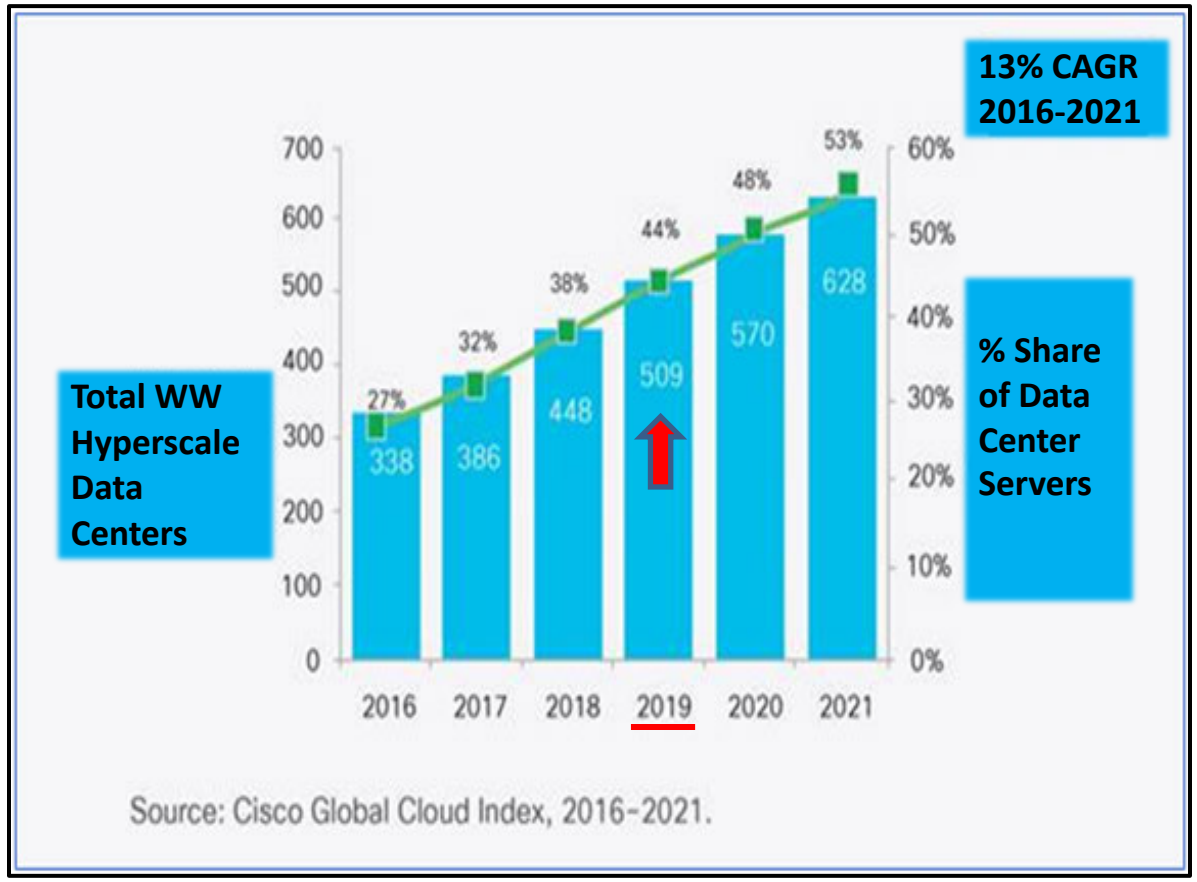
Shift Toward *Fewer* - but Much *Larger* Data Centers



- A Hyperscale Data Center (HSDC) is an enormous distributed computing environment.
- Massive infrastructure - over 400,000 ft^2, largest is >1.1 million ft^2 (= 18.3 soccer fields).
- HSDCs scale compute and storage from PBs to EBs independently – and fast.
- Designed with "self-healing" redundant components – if a failure - workload moves to another server.
- Using RAID or replication protection for most active data.
- Using Erasure Coding protection for large objects and archives where slow recovery performance is not an issue.
- Extreme energy consumption and carbon footprint challenges.
- Tape usage increasing and ***will be critical*** to enable HDSC growth and manage costs.

Source: Horison, Inc.

The Hyperscale Market Profile



- HSDC Cloud Providers - Amazon, Google, IBM and Microsoft collectively control more than half of the WW cloud infrastructure service market.

- HSDC Non-Cloud Providers are primarily focused in either the
  - US 44% (Apple, Twitter, Facebook, eBay, LinkedIn, Yahoo,…)
  - or China 8% (Tencent, Baidu,…)

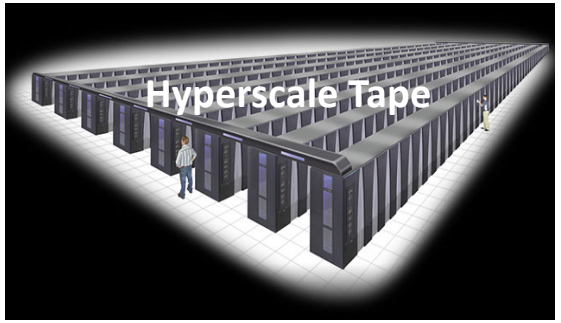
- The four biggest cloud providers (Amazon, Google, IBM and Microsoft) operate the largest footprints.

- Each has at least 45 data center locations WW.

- Global data centers consumed ~416 terawatts (3%) of the total electricity consumed last year, nearly 40% more than the entire United Kingdom.

Hyperscale Leverages Tape for Growth

Energy and Carbon Footprint Issues Loom for HSDCs



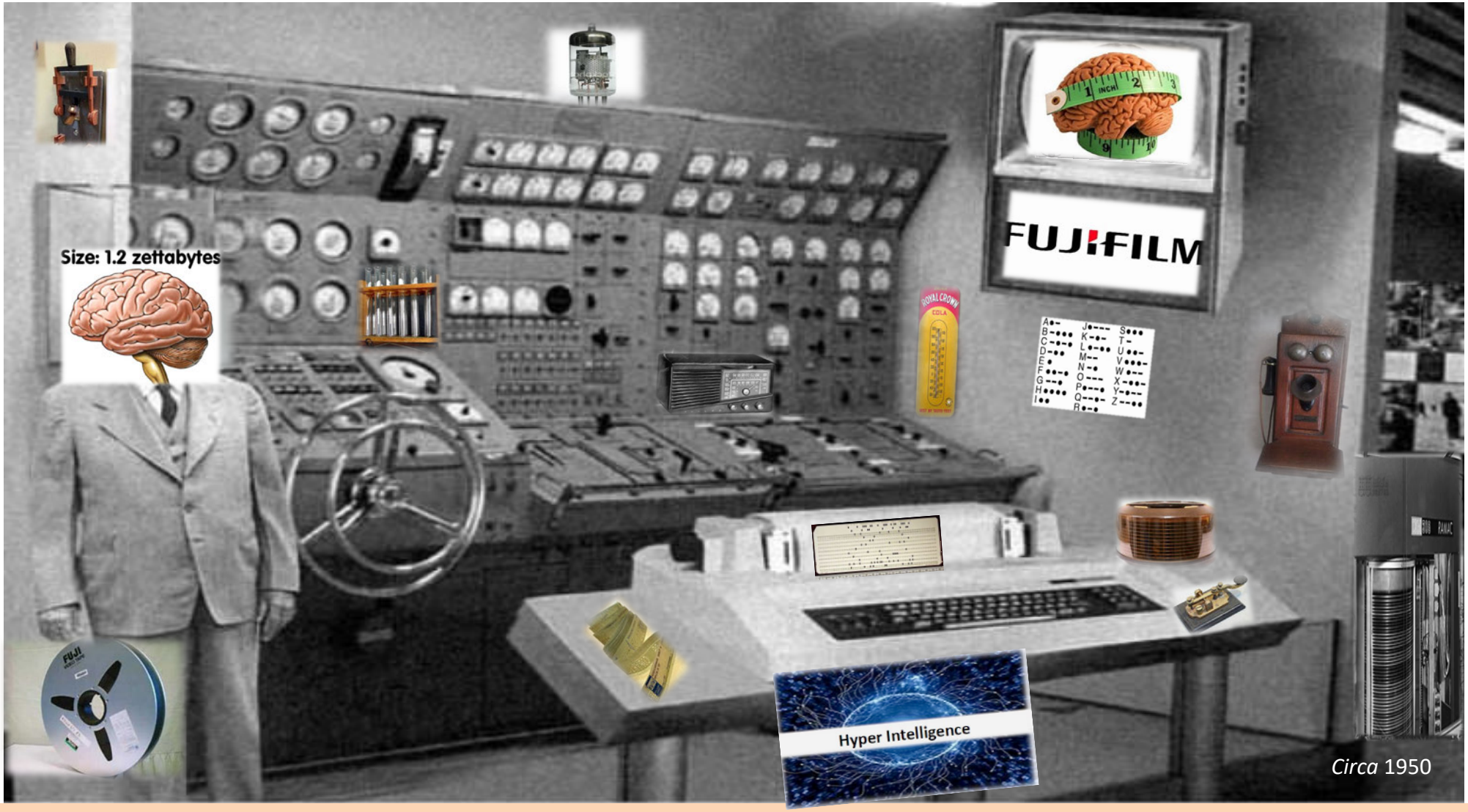
By YE 2020	Hyperscale Data Centers Will Have	YE 2017
~570	Total HSDCs	386
47%	Of all data center servers	21%
68%	Of all data center processing power	39%
53%	Of all data center traffic	34%
57%	Of all data stored in data centers ~4.2 ZB	49%

- Ex: If all HSDC data (~4.2 ZB) is stored on HDDs, 281.4 million 15 TB HDD's and ~1.7 billion watts would be required.
- For 40% of data stored on HDDs, 675.2 million HDD watts (@6watts/drive).
- For 60% on tape, ~67.8 million tape watts (1/15th of HDD).
- Total energy savings ~ 945 million watts (megawatts) if tape used.

For HSDCs - physically scaling capacity beyond EB levels will be nearly impossible without tape to store less-active data.

Source: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html, Energy estimates from Horison, Inc. www.horison.com

The First Hyperscale Data Center

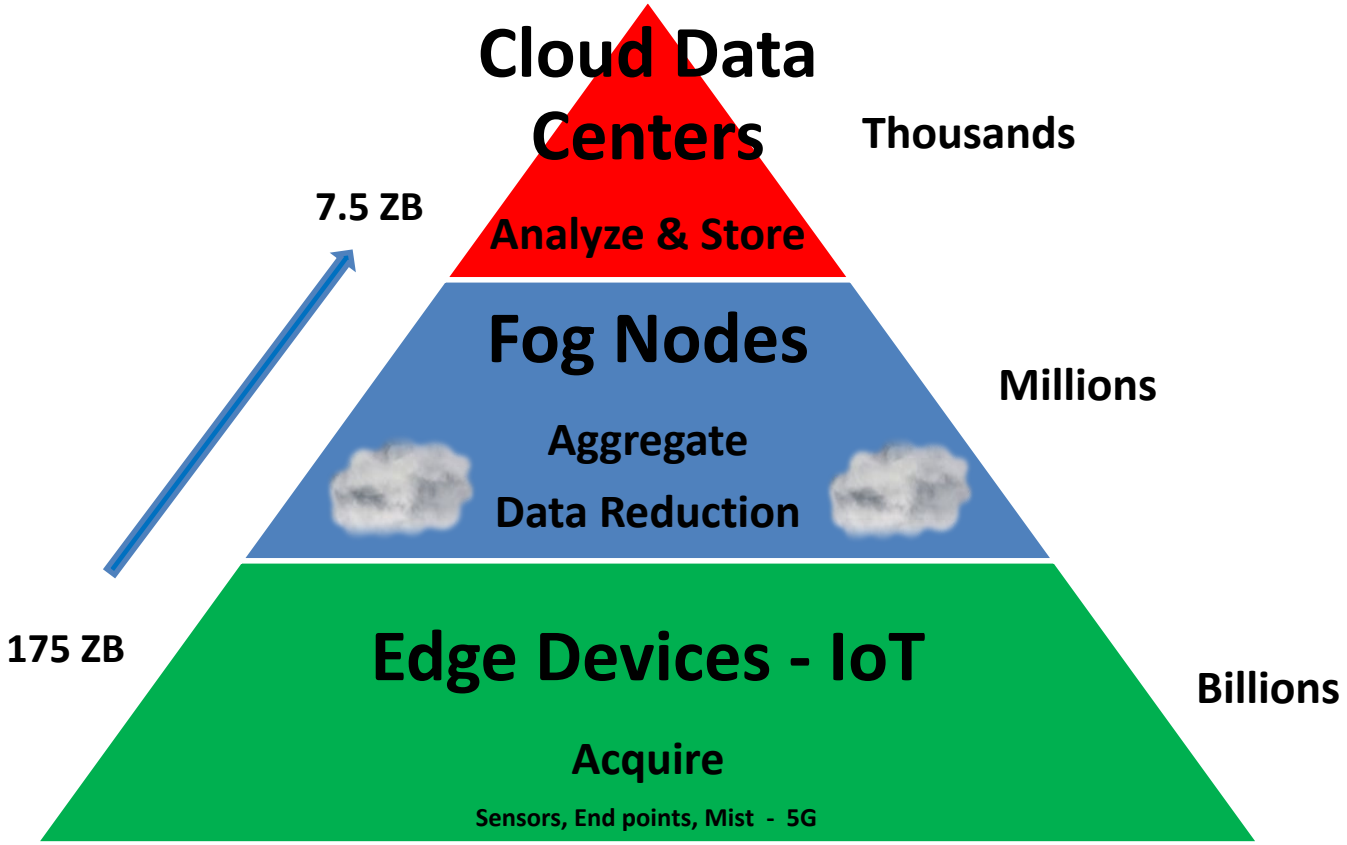


Two famous and hyper-intelligent computer scientists created this prototype to illustrate how the first Hyperscale Computer could look like in the year 2000. Also the scientists readily admit that "the computer will require not yet invented technologies to work, but 50 years from now scientific progress is expected to solve these problems. With a nautical steering-wheel mouse, a vivid, long-life 21" RCA CRT monitor & console, and a teletype print interface running TOS and the Fortran language, the computer will be very powerful and easy to use".

Scientific American.

The IoT, Edge and Fog – The Next Frontier

Shifting IoT Computing Away From Centralized Nodes to the Logical Extremes of a Network

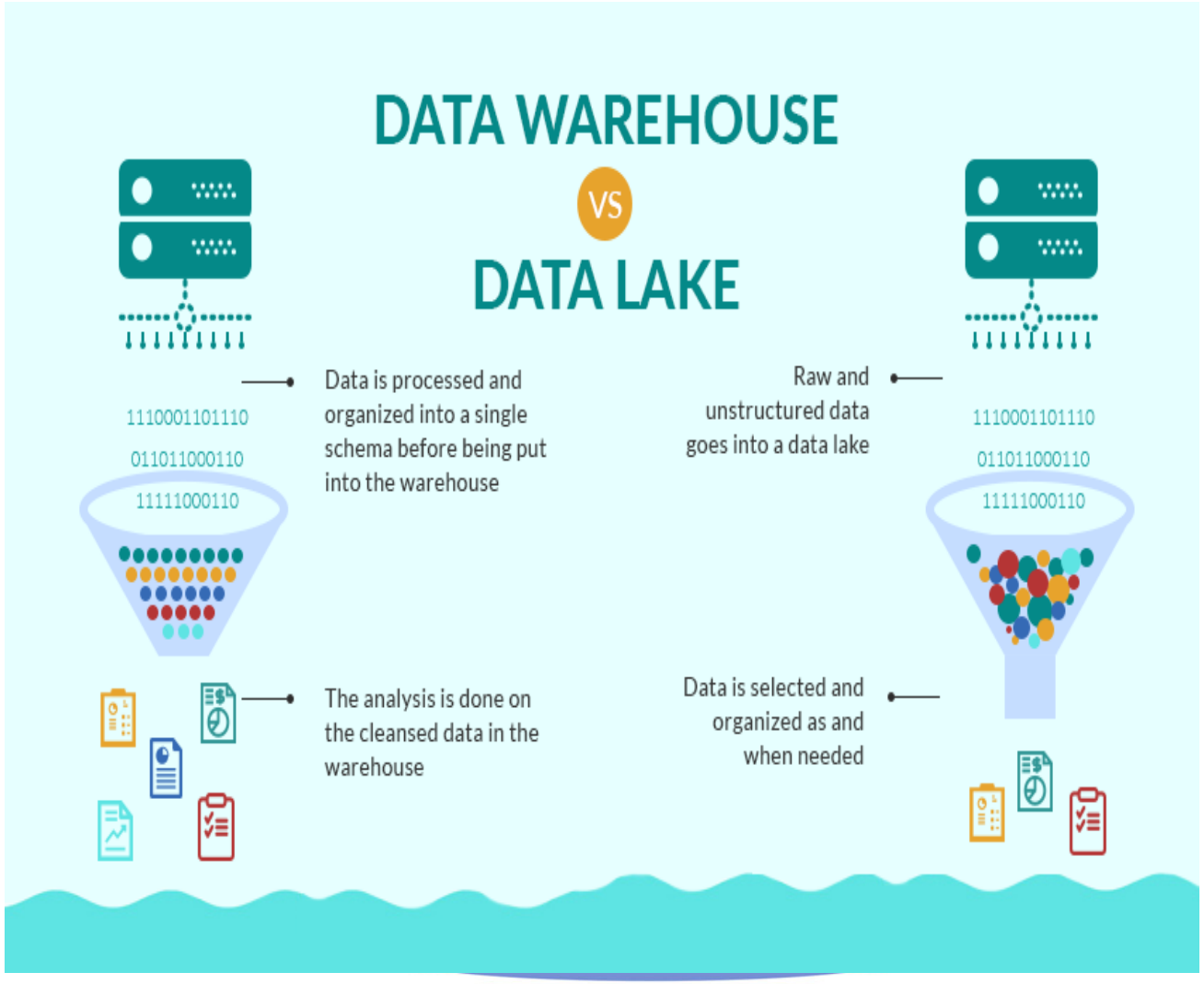


- The Fog quickly aggregates/reduces IoT data before it reaches the cloud.
- Any device with computing, storage, and network connectivity (Hyper-converged) can be a Fog node.
- Most IoT data will be processed/reduced before being sent to a data center – new cybersecurity challenges!

Source: Horison, Inc.

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, object or BLOB format until it is needed (**cold data**).
Data lakes are often distributed over multiple nodes rather than the fixed, structured environment of a 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.

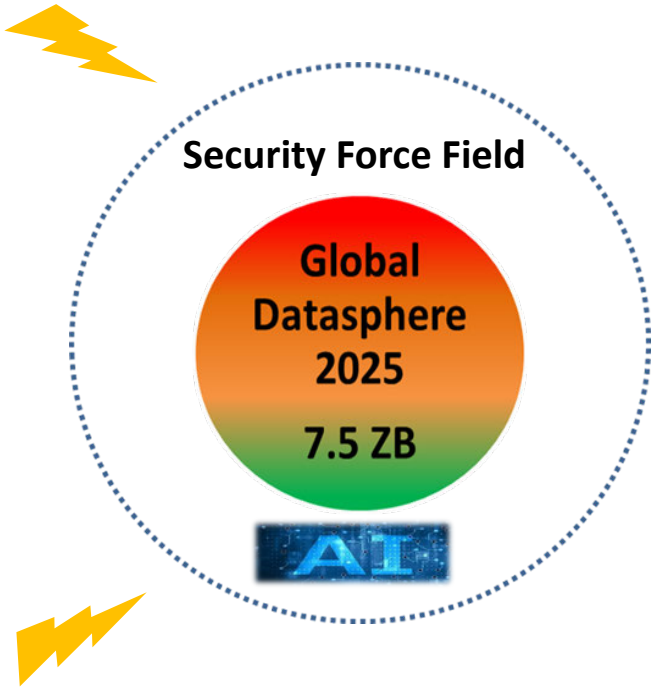
BLOB - A Binary Large Object (Unstructured)

Security Is Not An Option

Shifting From an IT Problem - to Everyone's Problem

EQUIFAX BREACH
- 143 MILLION AMERICANS
- NAMES, ADDRESSES
- SOCIAL SECURITY NUMBERS

Attack Forces
Cybercrime, malware
Natural disasters
Software corruption
Human error
Terrorism, theft
Hardware failure
Energy outages



"The identity I stole was a fake!
Boy, you just can't trust people these days!"

What is the Value of Digital Data at Risk?

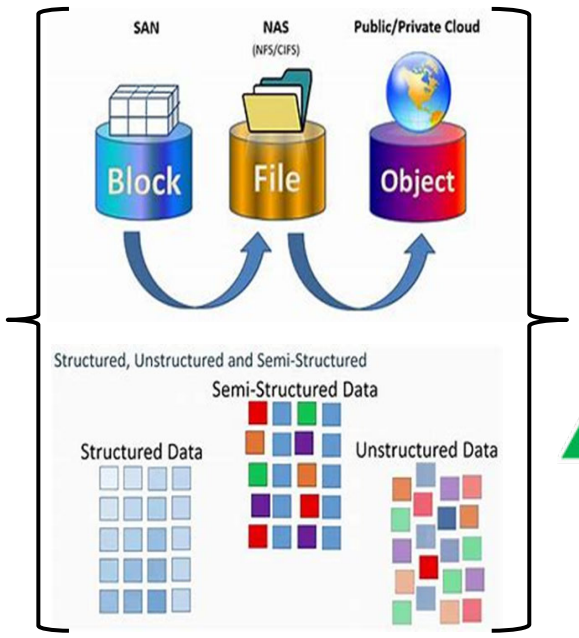
- Equifax, with over 800 million individual consumers and more than 88 million businesses worldwide, suffered a data breach in 2017 of 143 million users.
- Equifax faces a class action lawsuit up to **$70 billion** representing the perceived value of the data at risk.

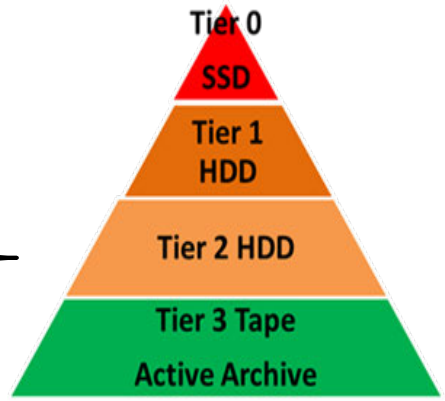
The Shift to Intelligent Storage Is Underway…

For 2020 and Beyond - *It's All About the Data*

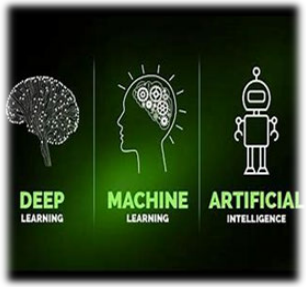
Global Datasphere 2025 Created ~175 ZB

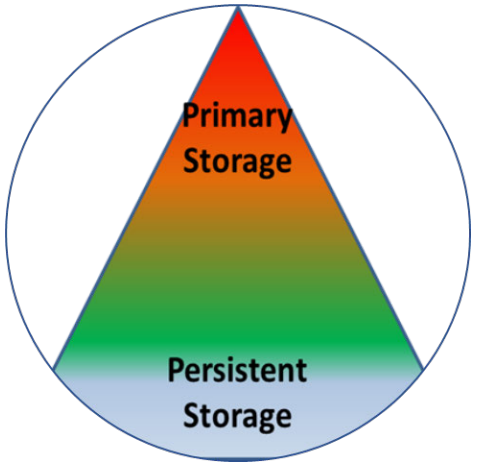
Global Datasphere 2025 Stored ~7.5 ZB





Specialized Clouds





Source: Horison, Inc.


Quantum Computer
HYPERSPACE
SINGULARITY
Hyper Intelligence
Yottabyte^n...
SEISMIC SHIFTS
Understanding the
FOURTH DIMENSION
WELCOME TO THE FUTURE
Vision 2020