STORAGE DEVELOPER CONFERENCE
SDC 22
Fremont, CA
September 12-15, 2022
BY Developers FOR Developers

A SNIA Event

24G SAS Advancements for Hyperscale Environments

Rick Kutcipal

Product Planning – Broadcom
Board of Directors – SCSI Trade Association

24G SAS Advancements for Hyperscale Environments

Opposing Trends in the Storage Ecosystem

Excitement and performance advancements driving large investments in flash

Growing capacity requirements on hyperscale architectures driving innovation in HDDs

Increased Investment and Innovation in HDD

Capacity optimized features

Specialized performance solutions

Advance security features

Power considerations

Storage SSD Industry Growth > HDD Vendor Consolidation

1991 — First production SSD

2005 — First SAS enterprise SSD

2007 — First PCIe enterprise SSD, 60 SSD manufacturers

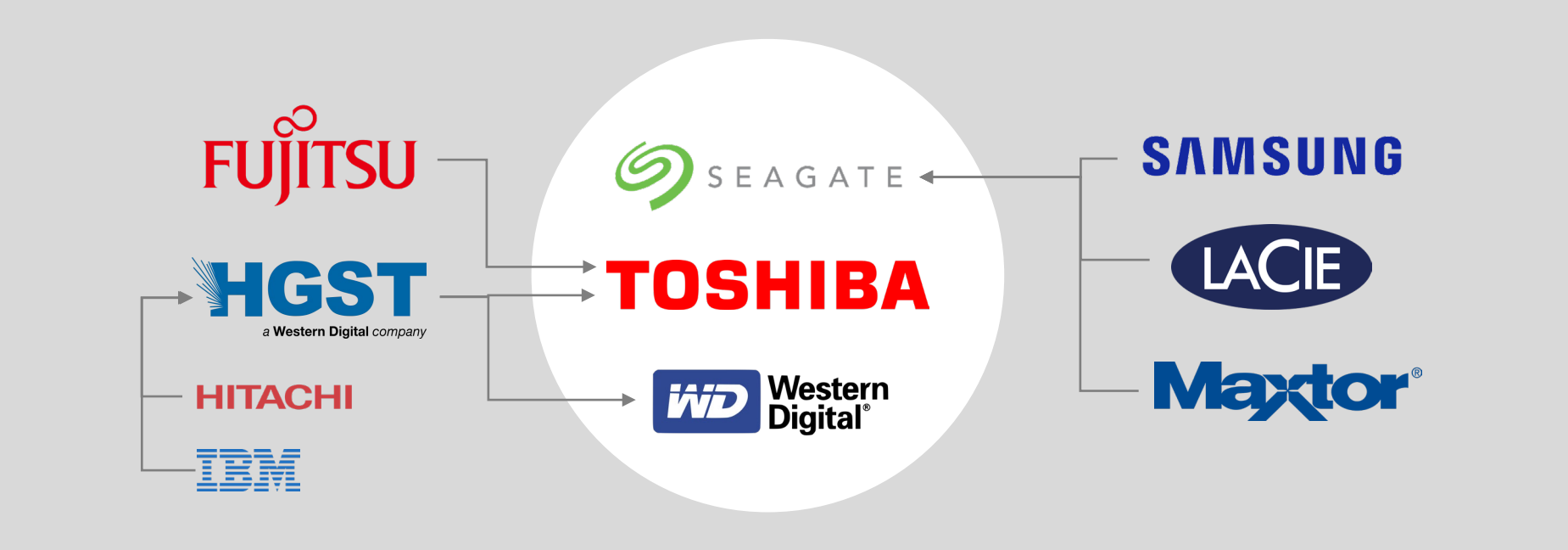
2008 — Approximately 100 SSD manufacturers

2010 — SSD market reaches $1B

HDD Vendor Consolidation

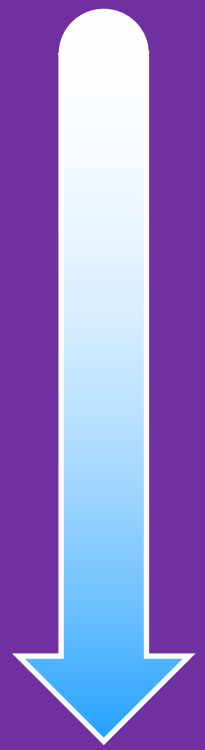
10 HDD vendors in 2000

3 HDD vendors today



Storage HDD Hyperscale Growth

Performance



Capacity Scalability

Data Centers began as simple arrays of servers to connect users to information

Focus on transactional performance

SSD were a natural fit

Today's Data Center collects massive amounts of data to be analyzed & accessed

Focus on maximizing the content available to a given user

HDDs, with a significant $/TB advantage over SSDs are the perfect fit for today's data center

Capacity needs scale every day

Every day, the world records over 413 Petabytes of video data

More video content gets uploaded in 30 days than what the major US television networks have broadcast in 30 years (Source: Wordstream)

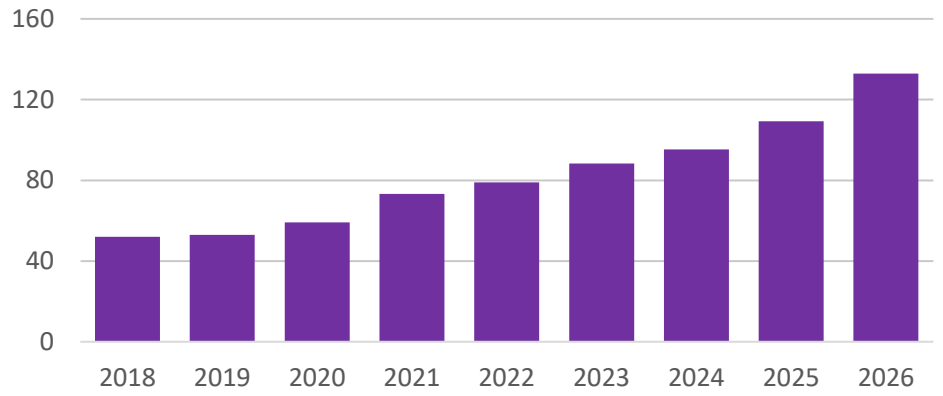
YouTubers upload about 720,000 hours of fresh video content per day

The Correction

89% of all storage consumed by hyperscale data centers in 2021 was HDD based – TRENDFOCUS

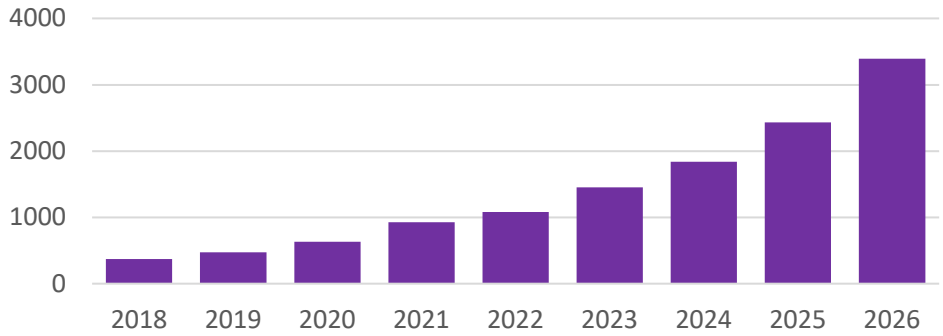
Source: TRENDFOCUS
Q1 '22 Quarterly Update

NL HDD Units (M)



2020-2026 Unit | CAGR **14.4%**

NL HDD Capacity (Exabytes)



2020-2026 Capacity | CAGR **32.2%**

Hyperscale HDD Technology Innovations

Shingled Magnetic Recording (SMR) – Improvements in arial density (capacity)

Hybrid SMR – Flexibility in SKU management

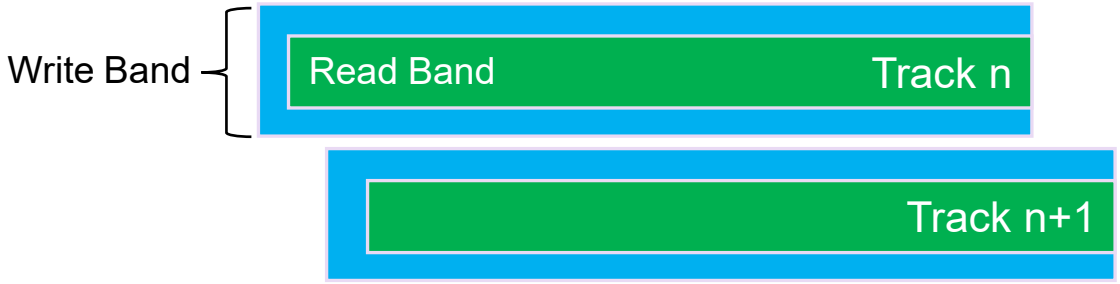
Repurposing Depopulation (DePop) – Data Center management

Multi-Actuator – Addressing the performance per capacity issue

Command Duration Limits – Improving performance without sacrificing latency

Shingled Magnetic Recording – SMR

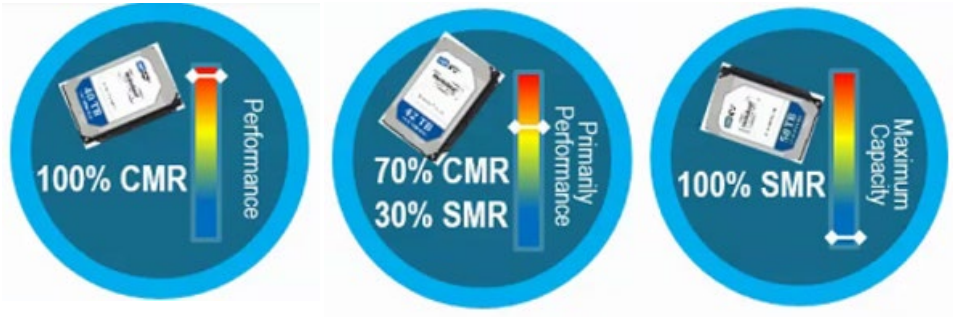
- SMR enables higher drive capacity by overlapping written tracks
- Tracks are organized into zones
- Requires sequential writes for optimal performance, no read performance impact
- Standards
  - ZBC – The 1st revision of Zone Block Commands for SAS
  - ZAC – The 1st revision of Zone ATA Commands for SATA
  - SAT-4 – The SCSI to ATA translation revision that defines the mapping between ZBC and ZAC
- 3 types
  - *Drive Managed – zones are managed by the drive, no host visibility*
  - *Host Aware – zones can be managed and visible to the host*
  - Host Managed – zones and workloads managed by the OS/application



Advancements in SMR

- Format Capabilities
  - Drive reports what it supports (SMR/CMR and sector size)
  - T10 (ZBC-2) – Format with Preset
  - T13 (ZAC-2) – Mutate + Set Sector Configuration
  - State of the standards
    - Both T10 and T13 are complete
    - SCSI to ATA translation does not exist today

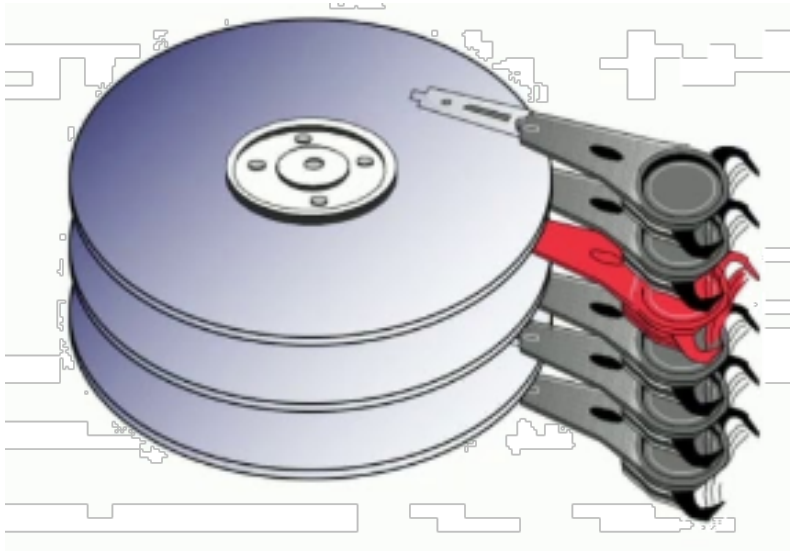
Example of Format Capabilities



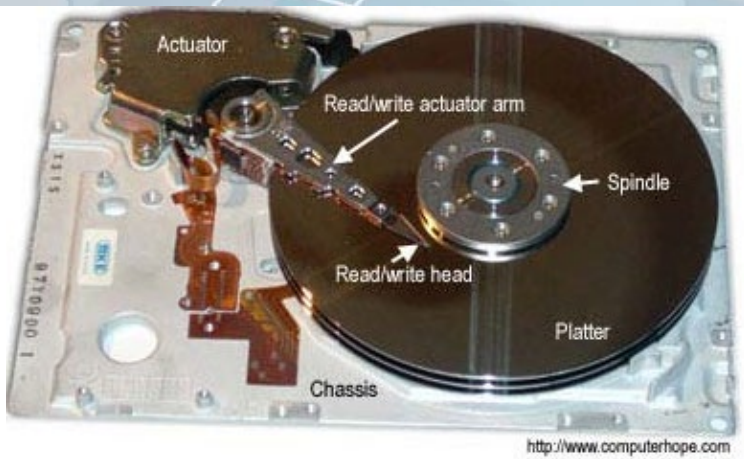
- Hybrid SMR
  - The capability to dynamically change zone type (SMR vs CMR)

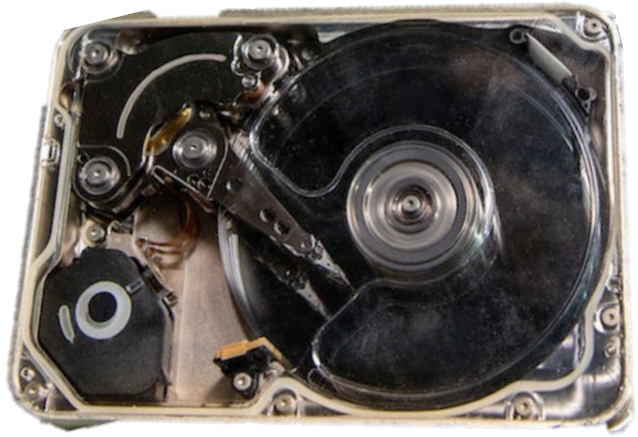
Repurposing DePopulation (RDP)

- Growing capacity of HDD presents challenges for large-scale Data Center deployments
  - Increased frequency of correctable errors increases tail latencies and decreases performance
  - Failing a drive results in vast amounts of disk capacity remaining offline until a failed unit can be replaced
  - Additional capacity needs to be provisioned to account for failure rate
  - Significant number of returned drives found fault is with a single failed head
- T10 and T13 standards bodies defined “Offline Logical Depop”
  - Capacity backed by the failed head is removed from namespace
  - Drive is reformatted at the lower capacity
  - Drive brought back online with the lower capacity
- Standard applies to both HDD and SSD, but HDD only today
  - Applies to “Physical and Storage Elements”
  - Example elements could include: Head, Actuator, Die, Flash Channel



Multi-Actuator Technology

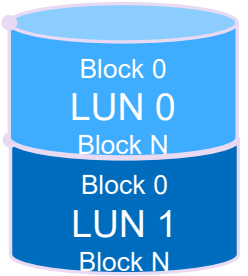




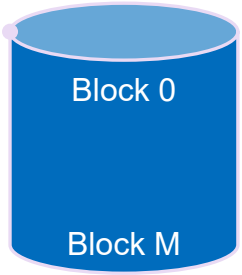
- Multi-Actuator SAS devices expose one LUN per actuator
  - Each LUN addresses unique storage on the device
  - Single LUN SAS DA drives are being explored
- SATA doesn't support multiple LUNs and must present a single device
  - Split LBAs
  - Striped
- Dual actuator a strategic technology to improve IOPs/TB in high-capacity HDDs
- Industry concerns
  - Power: Sizable increase over single actuator
  - Capacity: Dual actuator drives will have one less platter (room needed for additional mechanicals)

SA SAS HDD

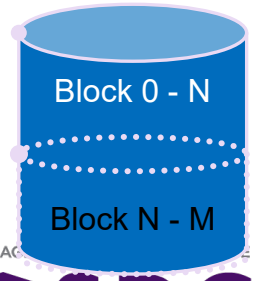
MA SAS HDD



SA SATA HDD

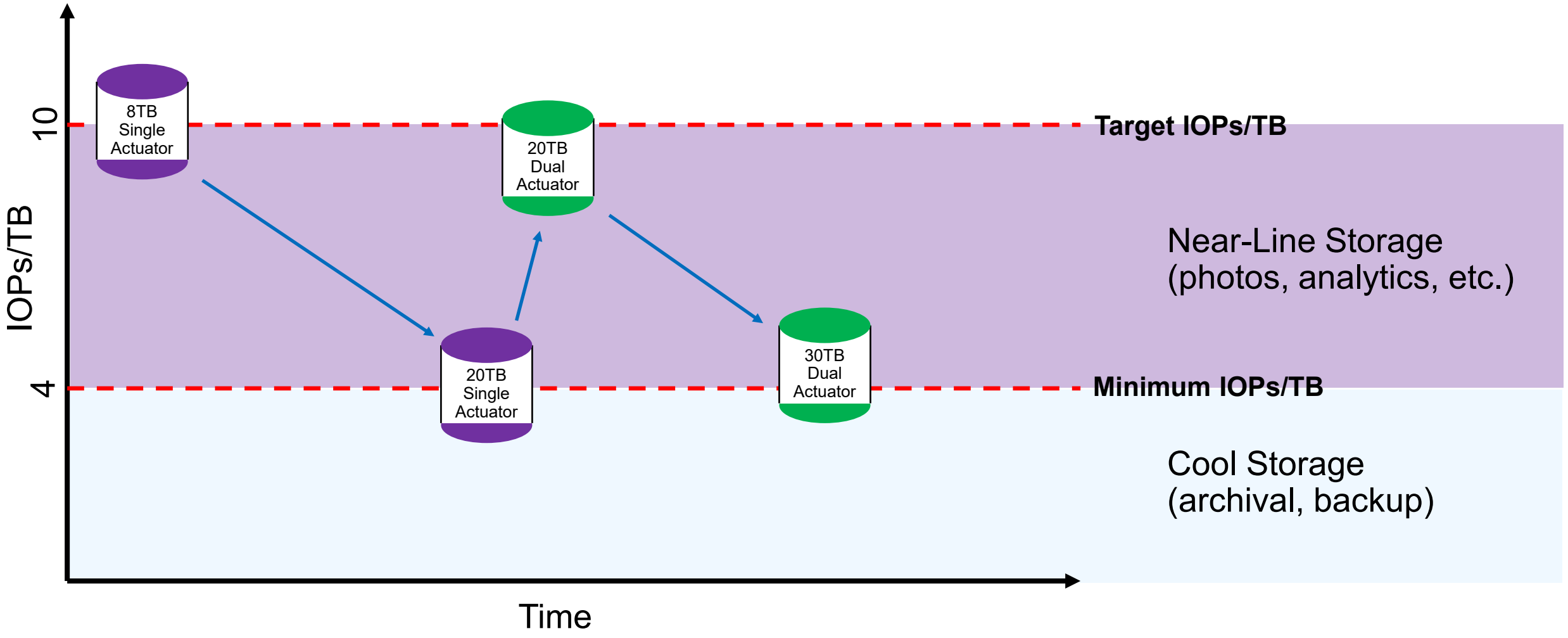


MA SATA HDD*



* Split LBA example

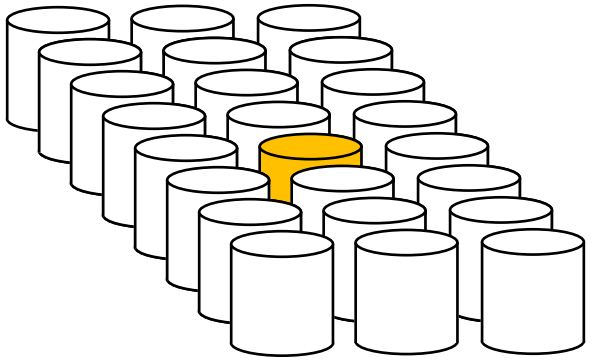
Multi-Actuator HDD Maintains IOPs/TB as Capacity Scales



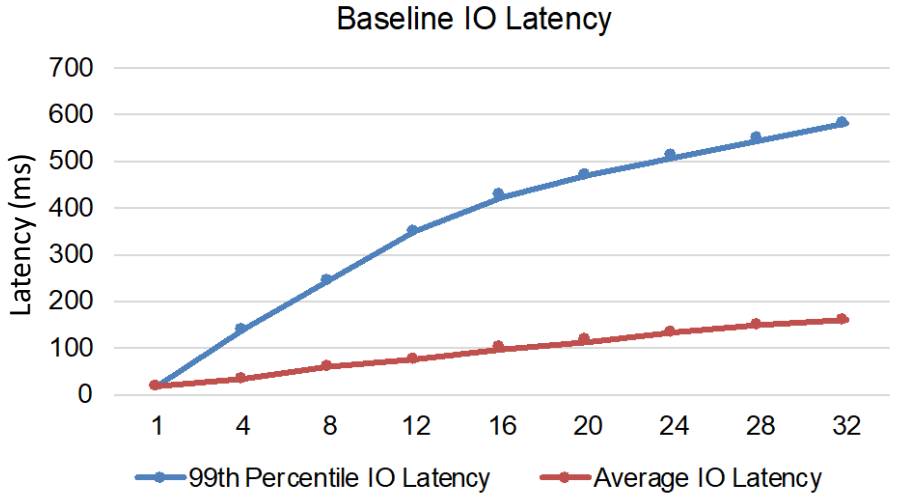
IOPs/TB calculated based on RR QD1 4k IOPs

Command Duration Limits

- HDD tail latency is important in large data centers
  - For implementing different service level agreements
  - For overall system performance – the aggregate system performance is throttled by the drive with the longest access time
- OCP published ”Cloud HDD – Fast Fail Read” in 2018
- In 2019, T10 introduced Command Duration Limits and proposed it to SPC-6



One Read to One Location



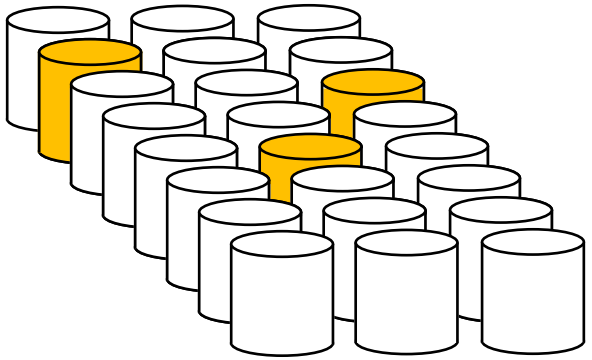
Latency data source:

Damien Le Moal, SDC21 - *Be On Time: Command Duration Limits Feature Support in Linux™*

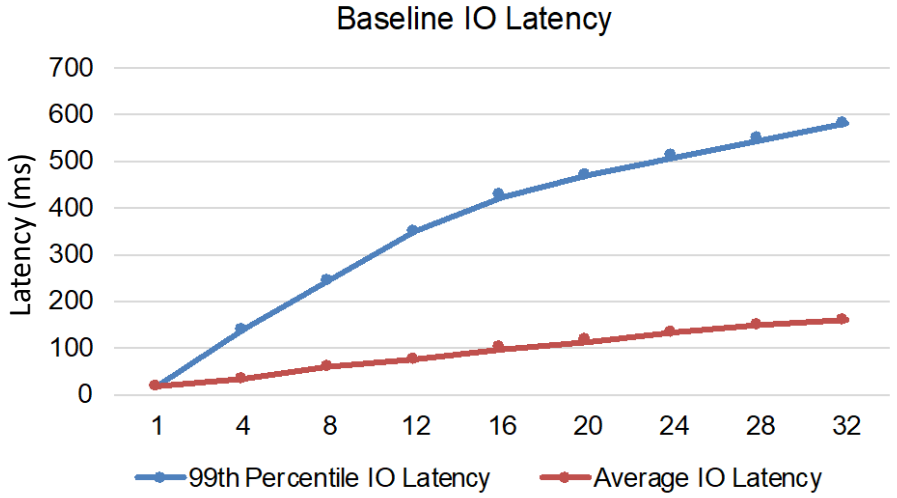
Tail Latency Grows w/ Queue Depth

Command Duration Limits

- HDD tail latency is important in large data centers
  - For implementing different service level agreements
  - For overall system performance – the aggregate system performance is throttled by the drive with the longest access time
- OCP published ”Cloud HDD – Fast Fail Read” in 2018
- In 2019, T10 introduced Command Duration Limits and proposed it to SPC-6



One Read to Three Locations



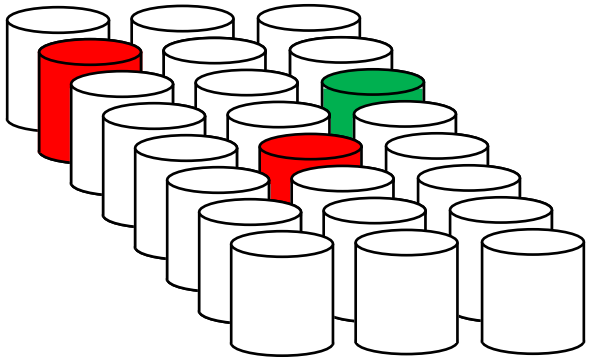
Latency data source:

Damien Le Moal, SDC21 - *Be On Time: Command Duration Limits Feature Support in Linux™*

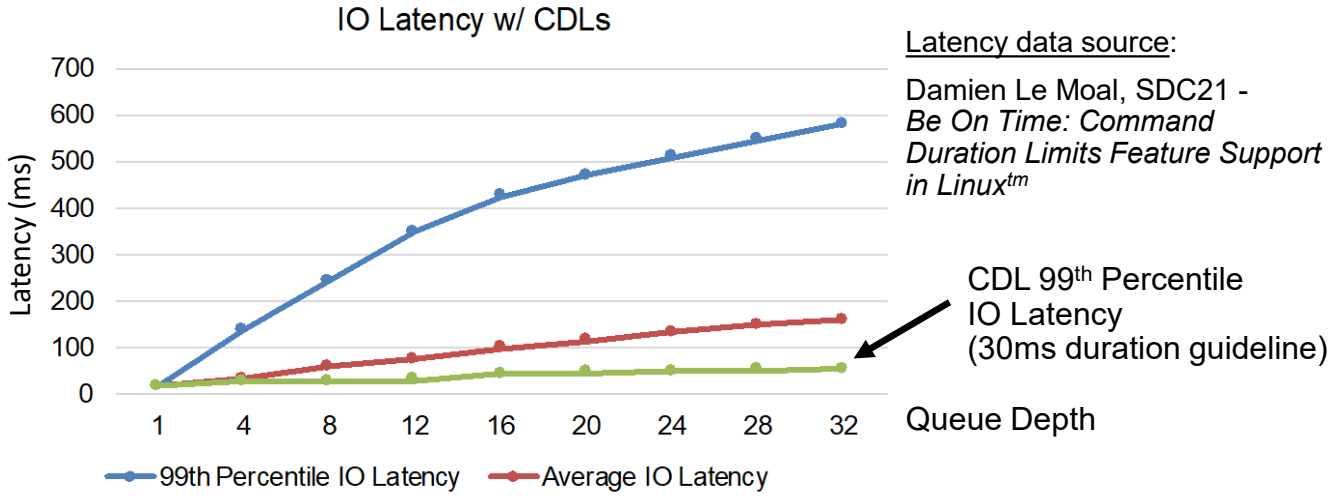
Tail Latency Grows w/ Queue Depth

Command Duration Limits

- HDD tail latency is important in large data centers for 2 reasons
  - For implementing different service level agreements
  - For overall system performance – the aggregate system performance is throttled by the drive with the longest access time
- OCP published ”Cloud HDD – Fast Fail Read” in 2018
- In 2019, T10 introduced Command Duration Limits and proposed it to SPC-6



One Read to Three Locations



CDLs Greatly Improve IO Tail Latency

24G SAS Advancements for Hyperscale Environments

Opposing Trends in the Storage Ecosystem

Excitement and performance advancements driving large investments in flash

Growing capacity requirements on hyperscale architectures driving innovation in HDDs

Increased Investment and Innovation in HDD

Capacity optimized features

Specialized performance solutions

Advance security features

Power considerations

Follow STA

THANK YOU

Please take a moment to rate this session.

Your feedback is important to us.

