SDC
STORAGE DEVELOPER CONFERENCE
SNIA SANTA CLARA, 2016

Cold Storage: The Road to Enterprise

Ilya Kuznetsov
YADRO

Agenda

- Technical challenge
- Custom product
- Growth of aspirations
- Enterprise requirements
- Making an enterprise cold storage product

Technical Challenge

Initial requirements

- Extremely large storage solution – hundreds of PB, will scale up further
- Non-typical hyper-scale workload quite similar to data archive:
  - Write once, read unpredictably rarely
  - Latency sensitive
  - Sequential high bandwidth
  - Guaranteed data integrity required
- TCO is crucial factor to deliver the feasible solution
- Simplified features with no need for enterprise capabilities

Derived Requirements

- **Density**
  - Huge capacity with low footprint
- **Power consumption**
  - Appropriate media and tech choice to reach extremely low power profile and handle inbound data stream
- **Scalability**
  - Performance scalability is a must
- **Latency**
  - No moving parts, no robotic libraries
- **Data integrity**
  - Inline data verification on read/write

Media comparison

Ethernet HDDs

Pros:

- ❑ No low-level coding and firmware

Cons:

- ❑ New software stack
- ❑ No new products recently
- ❑ No improvements with density and speed
- ❑ No enclosures/switches on the market

Tape

Pros:

- ❑ Consumes energy only in active state
- ❑ Mature and reliable technology
- ❑ Relatively high throughput of 300 MB/sec

Cons:

- ❑ Huge load/seek time leads to huge latency

Blu-ray

Pros:

- ❑ Consumes energy only in active state

Cons:

- ❑ Quite low throughput
- ❑ 60–90 seconds load time
- ❑ Bigger footprint
- ❑ Low drive MTBF for continuous write

SSDs

Pros:

- ❑ Great density
- ❑ Outstanding reliability

Cons:

- ❑ Extreme pricing
- ❑ Wearing issues
- ❑ Restricted market availability for such volumes

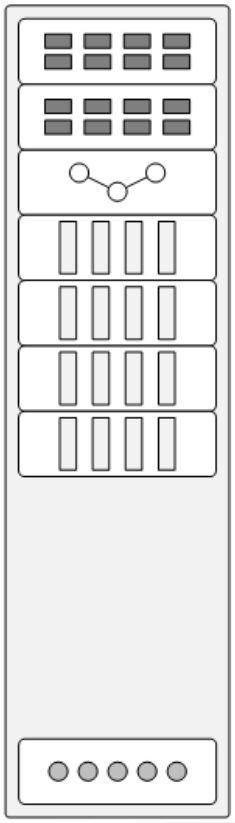
Custom product design principles

- Drives are powered off
- Drives get powered on for write/read purpose
- Maximum number of powered drives is 24 per shelf
- Dual parity block
- Enclosure w/o PSU (one shared PSU shelf per rack)
- Enclosure w/o moving parts
- No hot-swap for SAS HBA and expanders
- Low power profile leads to higher density
- Software features are simplified and workload specific
- No dedupe/tiering/compression/etc.

Custom product design decisions

- **Stop or idle**
  - Complete de-powering: lowest power, time to start
  - Spindle full stop: better power profile & OpEx
  - Idle: spindle slowdown, faster activation
- **Data protection**
  - Data integrity (hash checking)
  - Data mirroring/replication
  - RAID
  - Erasure coding

Custom product



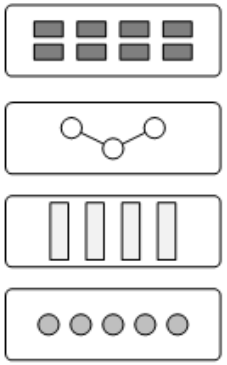
The solution consists of original design products and the minimal configuration is the following.

Redundant controllers
2 x YADRO OpenPOWER 4-Socket POWER8 systems (2U)
Each controller node supports up to 48 cores, 8TB RAM

Interconnect and cache
1 x YADRO PCIe Fabric & Cache Controller (3U)
Supports up to 128 NVMe SSD drives

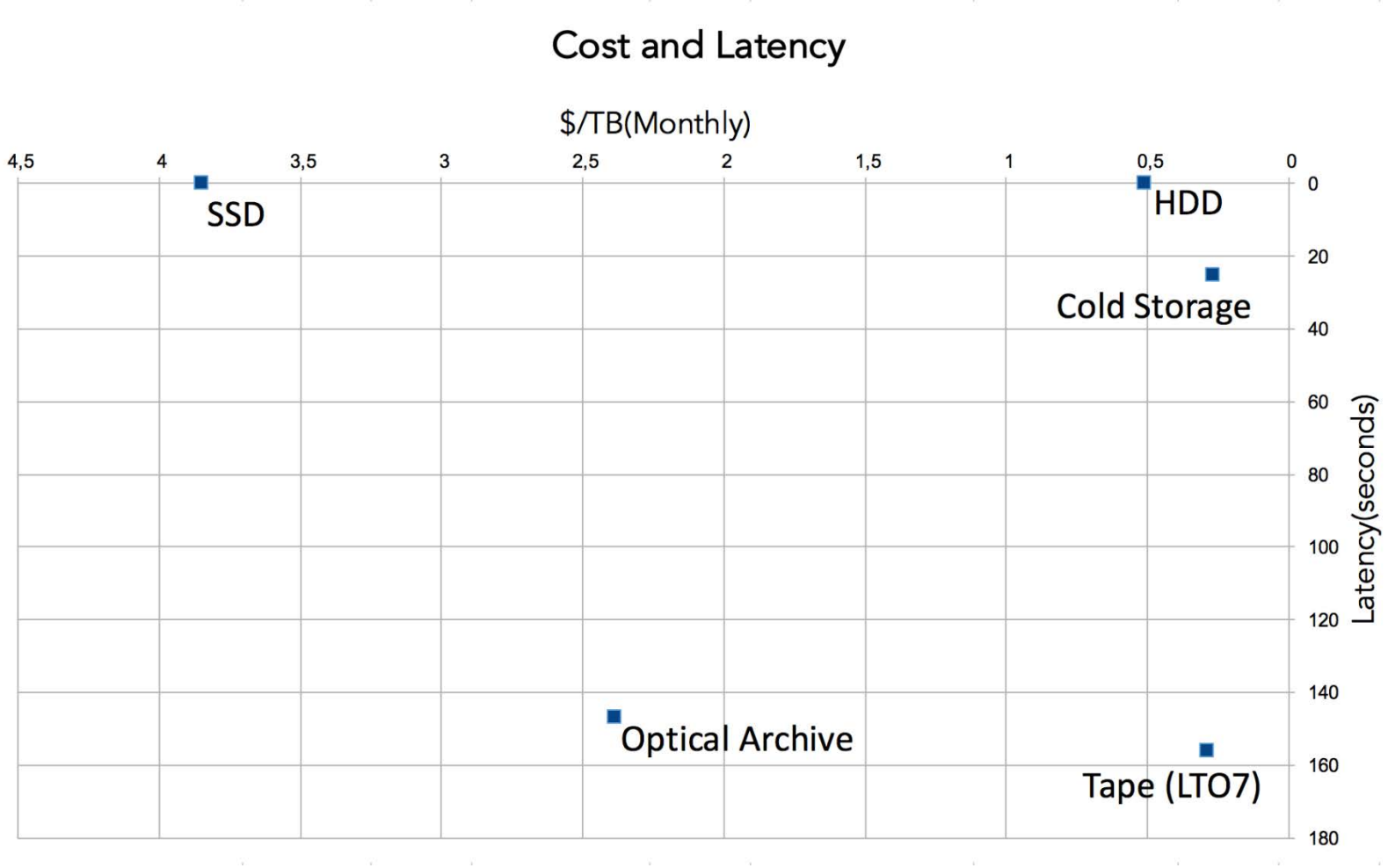
Cold Storage Shelf
4 x YADRO Cold Storage Shelf (4U)
Supports up to 128 SAS 3,5” drives

Power Supply Units Shelf
1 x YADRO PSU Shelf (1U) in N+1 configuration



In minimal redundant configuration storage raw capacity is up to 4PB with DRAM/SSD cache with total power consumption up to 3 kW .

Custom Cold Storage vs other media



This chart shows monthly cost per 1 TB based on 5-years TCO calculation for 100PB raw storage, including the costs for acquisition of storage media and associated hardware (except storage system controllers) and including the costs of electricity and cooling, assuming write once and read once for the full storage volume over the 5 year period. Deployment, maintenance and other costs are excluded.

The Road Goes On Forever

and the party never ends

Let’s turn our thing into an enterprise cold storage system:

A cold storage for any data, not just frozen

In our concept cold storage can be used for wide variety of workloads far beyond cold data delivering the common enterprise features with remaining outstanding TCO efficiency.

Challenges

- **Latency tackling approaches**
  - Applications behavior patterns, profile tiering
  - Metadata must go to Flash or DRAM/NVRAM cache
- **Disk start/stop cycle limitations**
  - Modern disks have ~300k start/stop cycles (165/day for 5 years)
- **Availability**
  - No single point of failure
- **Rebuild time**
  - Drive recovery takes a lot of time, whole drive group have to stay active

Extra requirements

- **Placement policies**
  - Cache and priority tags for data serialization.
- **Caching/Tiering**
  - Huge NVRAM-based write-back cache
- **Rebuild time**
  - Drive recovery takes a lot of time, whole drive group have to stay active
- **Complex management**
  - Unified HW&SW management, enterprise systems integration, monitoring and reporting

R&D directions

- **Drives optimized for cold storage**
  - Fast spin-up
  - Low RPM operation

Summary

- **Enterprise features**
  - Object storage
  - Thin provisioning
  - Caching/Tiering
  - Snapshots
  - Any media back-end support
- **Density**
  - 9PB per rack
- **Power consumption**
  - Depends on activity, 0 to 4kW per rack
- **Latency**
  - 20ms for typical scenarios and balanced configuration

Thank you!