MAS

USATLAS Meeting, March, 2020

Hironori Ito, Eric Lançon, Yinzi Wu, ATLAS ADC, RUCIO teams, etc…

BROOKHAVEN
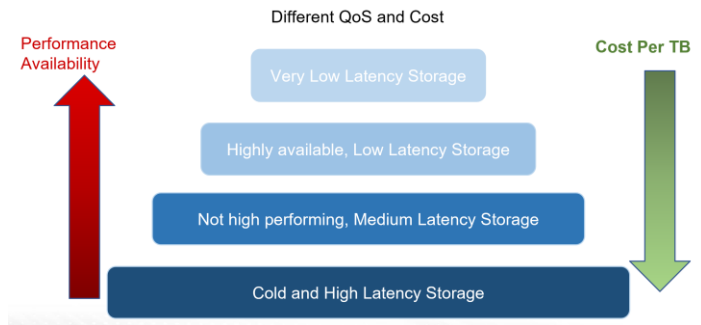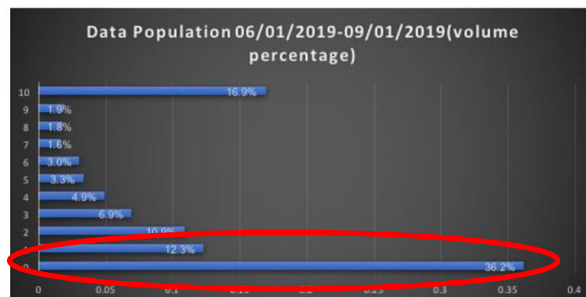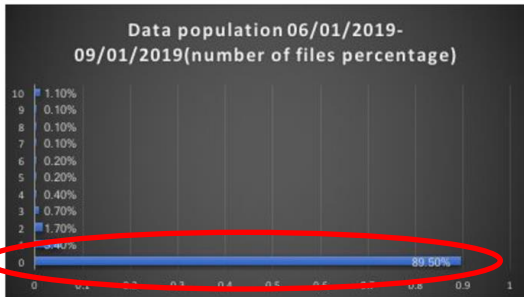NATIONAL LABORATORY

U.S. DEPARTMENT OF
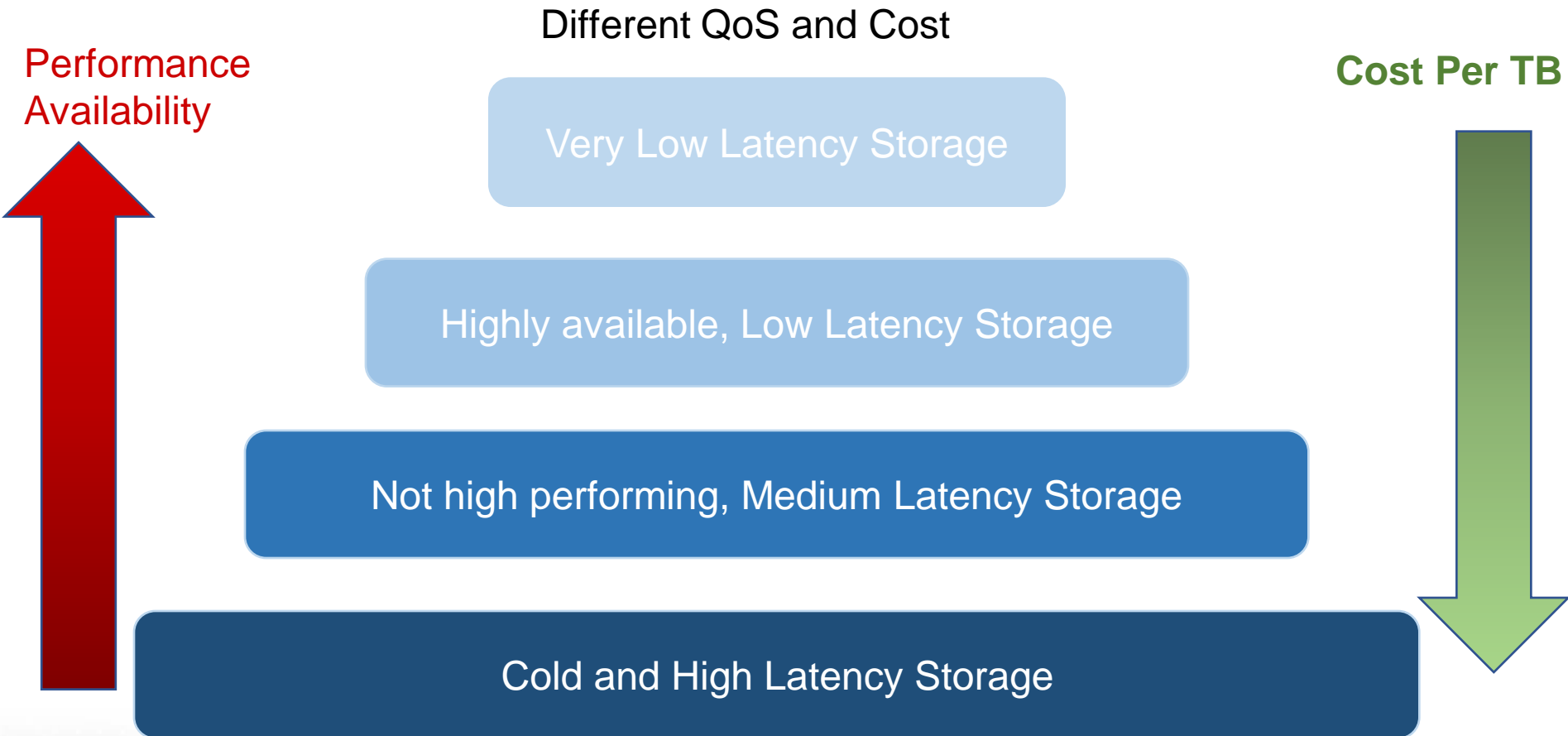**ENERGY**

BROOKHAVEN SCIENCE ASSOCIATES

Main Points

- Size of the on-line disk is finite.
    - The budget is finite and not keeping up with the rate of demand.

Access pattern of Datadisk files during 06/01/2019-09/01/2019.

Multi Layer Storage
Cost vs Performance

Different QoS and Cost

Performance
Availability

Cost Per TB

Very Low Latency Storage

Highly available, Low Latency Storage

Not high performing, Medium Latency Storage
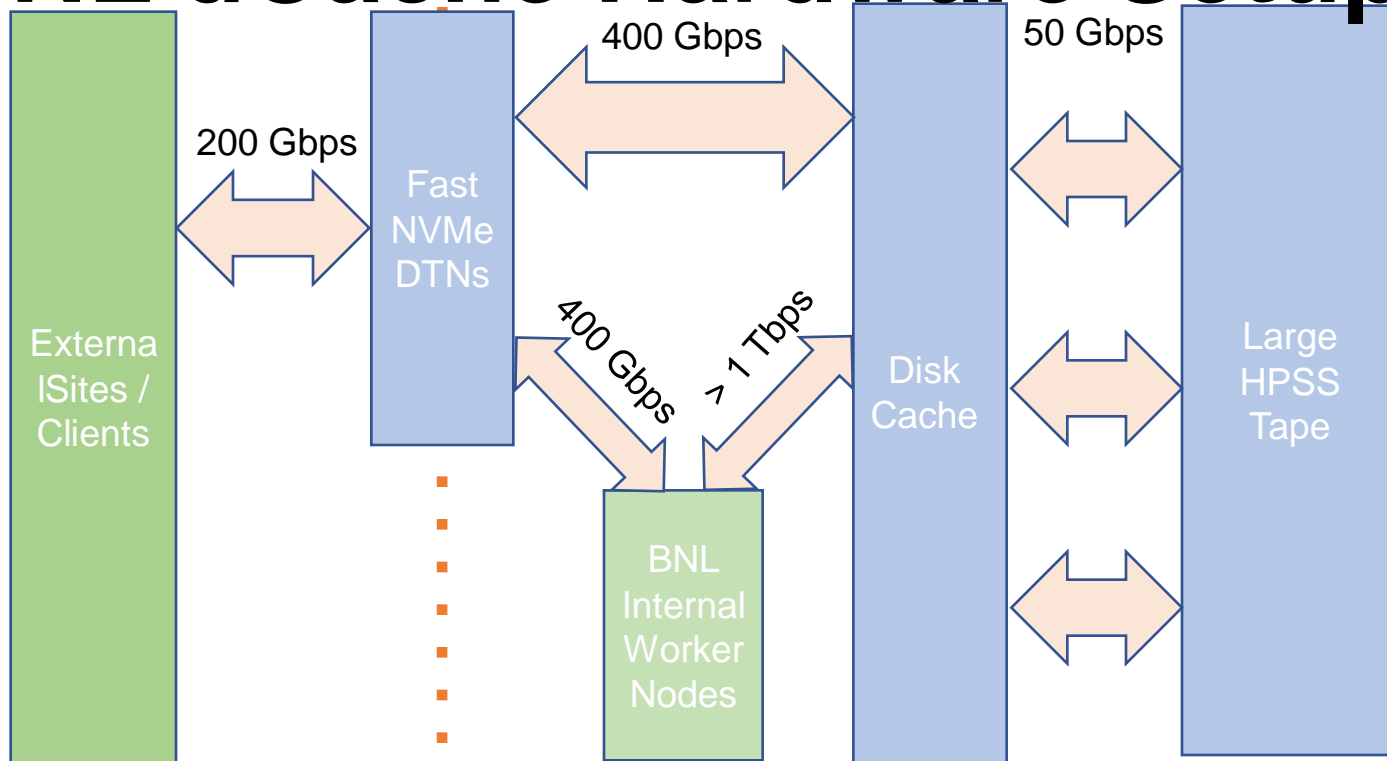
Cold and High Latency Storage

U.S. DEPARTMENT OF
ENERGY

BROOKHAVEN
NATIONAL LABORATORY

Things Done So Far at BNL

- BNL has setup one SE and one PANDA queue to test MAS concepts under the ATLAS workflow.

  - SE: BNLLAKE_DATADISK_DATALAKES under BNL production dCache

  - PANDA QUEUE: BNL_LAKE_UCORE

- BNL Test dCache

  - Testing the functionality of QoS in the new version

- BNL Production dCache

  - Allocate large read pools BNLLAKE_DATADISK_DATALAKES area backed by HPSS

- Identify the unused data in DATADISK

- Transfer of identified data to BNLLAKE via RUCIO

- Removal of transferred datasets from BNL DATADISK via RUCIO

BNL dCache Hardware Setup

External Sites / Clients

200 Gbps

Fast NVMe DTNs

400 Gbps

400 Gbps

> 1 Tbps

BNL Internal Worker Nodes

Disk Cache

50 Gbps

Large HPSS Tape

External : Internal

5

Mechanics of MAS

- Modify the dCache internal tables (SRM space file) to store the additional info because it only contains the dcache interal ID (CHIMERID)
    - PFN: reconstruct from name space by CHIMERAID
    - Access Time: obtained from the dCache billing database.
    - Dataset Name: obtained from RURICO catalog

Select datasets with the number of files larger than **N** bytes and not access more than **M** days except LOG and USER datasets.

NOTE: Not all files in the datasets are large than N bytes and not all files are non-accessed in more than M days

RUCIO "move" command to transfer DATADISK to LAKE

Analysis of the data transfers

- AS of Feb 27 2020

- total: 68M files, 18.5PB
- **Cut by size of files**
- not accessed 100 days: 55M files, 5.5PB
- Note; amount to be transferred is the **5.5PB – shown size**

- not access 100 days and less than 100MB, 51.3M files, 0.73PB
- not access 100 days and less than 200MB, 52.3M files, 0.87PB
- not access 100 days and less than 300MB, 52.8M files, 1.0PB
- not access 100 days and less than 400MB, 53.3M files, 1.2PB

- not access 100 days and less than 500MB, 53.6M files, 1.3PB
- not access 100 days and less than 1000MB, 54.1M files, 1.7PB
- not access 100 days and less than 1500MB, 54.6M files, 2.3PB
- not access 100 days and less than 2000MB, 54.9M files, 2.6PB
- not access 100 days and less than 3000MB, 55.0M files, 3.1PB

- AS of Feb 27 2020

- total: 68M files, 18.5PB
- **Cut by non-accessed days**

- not accessed 130 days: 53M files, 4.8PB
- not accessed 120 days: 54M files, 5.0PB
- not accessed 110 days: 55M files, 5.3PB
- not accessed 100 days: 55M files, 5.5PB
- not accessed 90 days: 56M files, 5.8PB
- not accessed 80 days: 57M files, 6.2PB
- not accessed 70 days: 58M files, 6.8PB
- not accessed 60 days: 58M files, 6.8PB

Data movement from DATADISK to LAKE
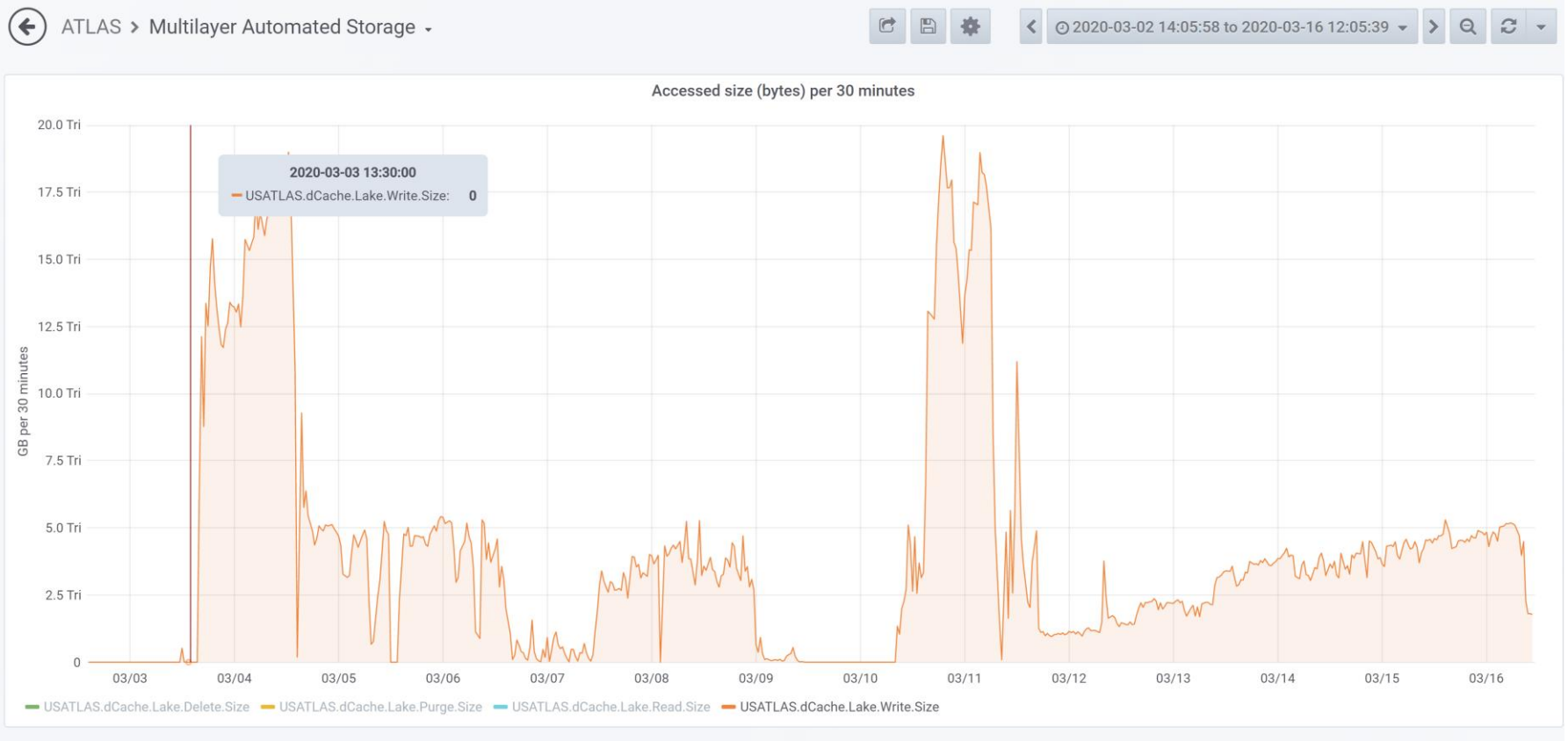
DDM Monitor

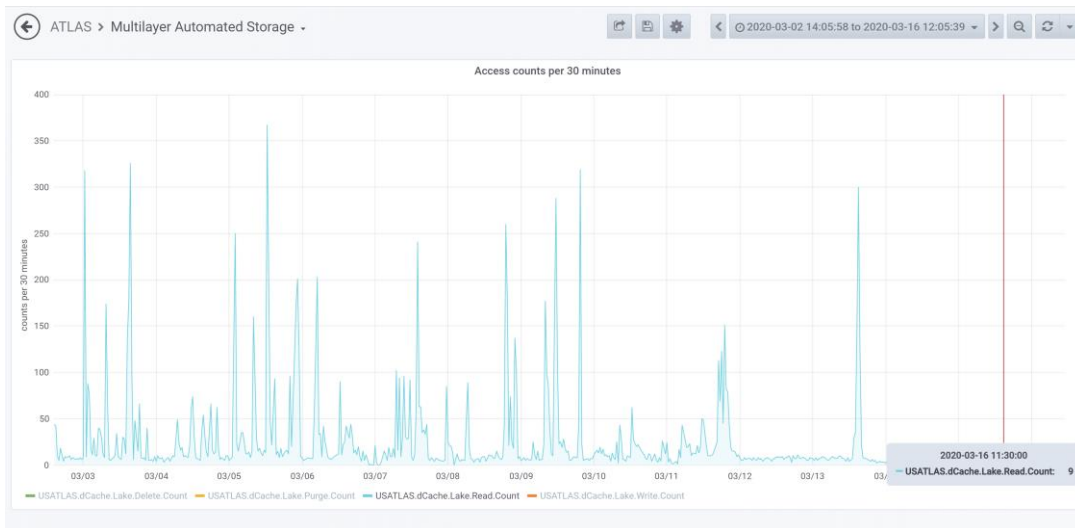Over **2PB** of the data being transferred
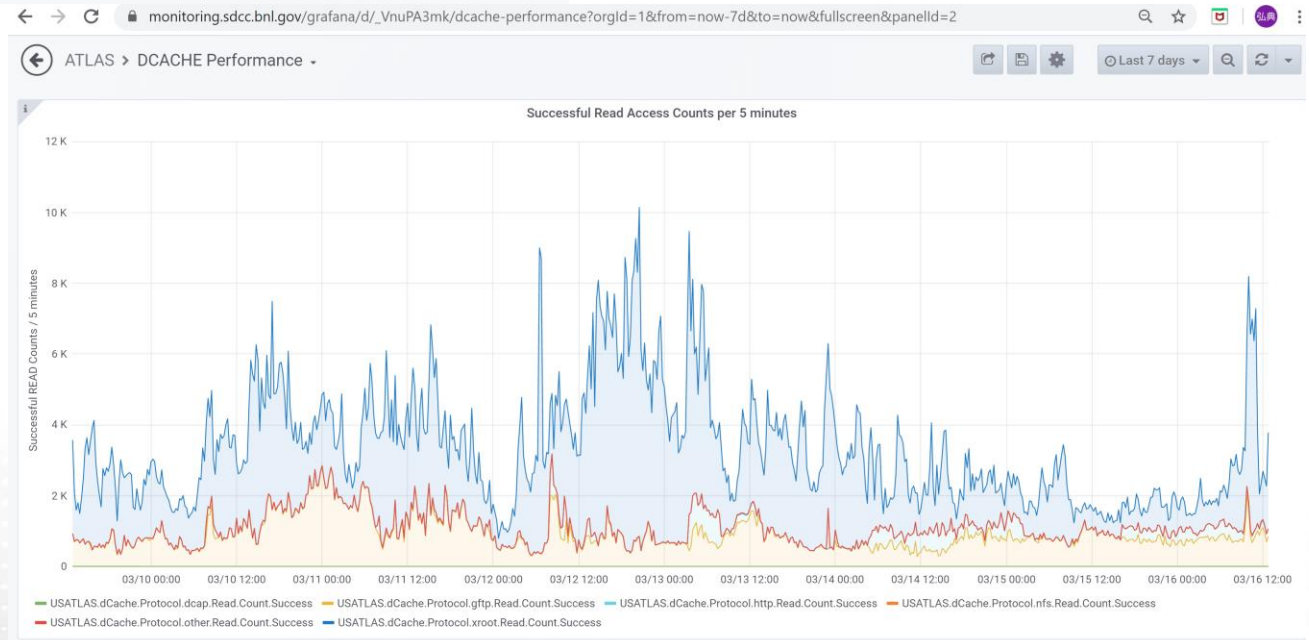
What happens to DATADISK

Access to the relocated data

Access to Lake files per 30 minutes

Access to all dCache by 5 per minutes

Conclusion

- The data access for files in LAKE is very small.

  - The data can be safely transferred without change in the ATLAS operation.

- RUCIO will immediately fill the DATADISK with the other data.

  - Change in the some rules(?) to reduce pledged amount in DATADISK?