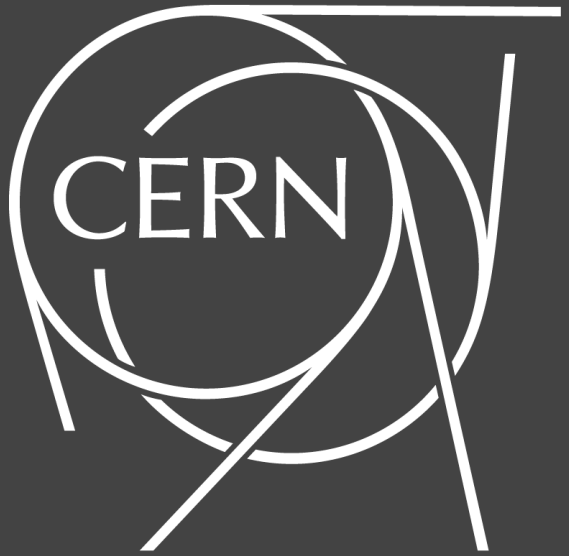

CERN

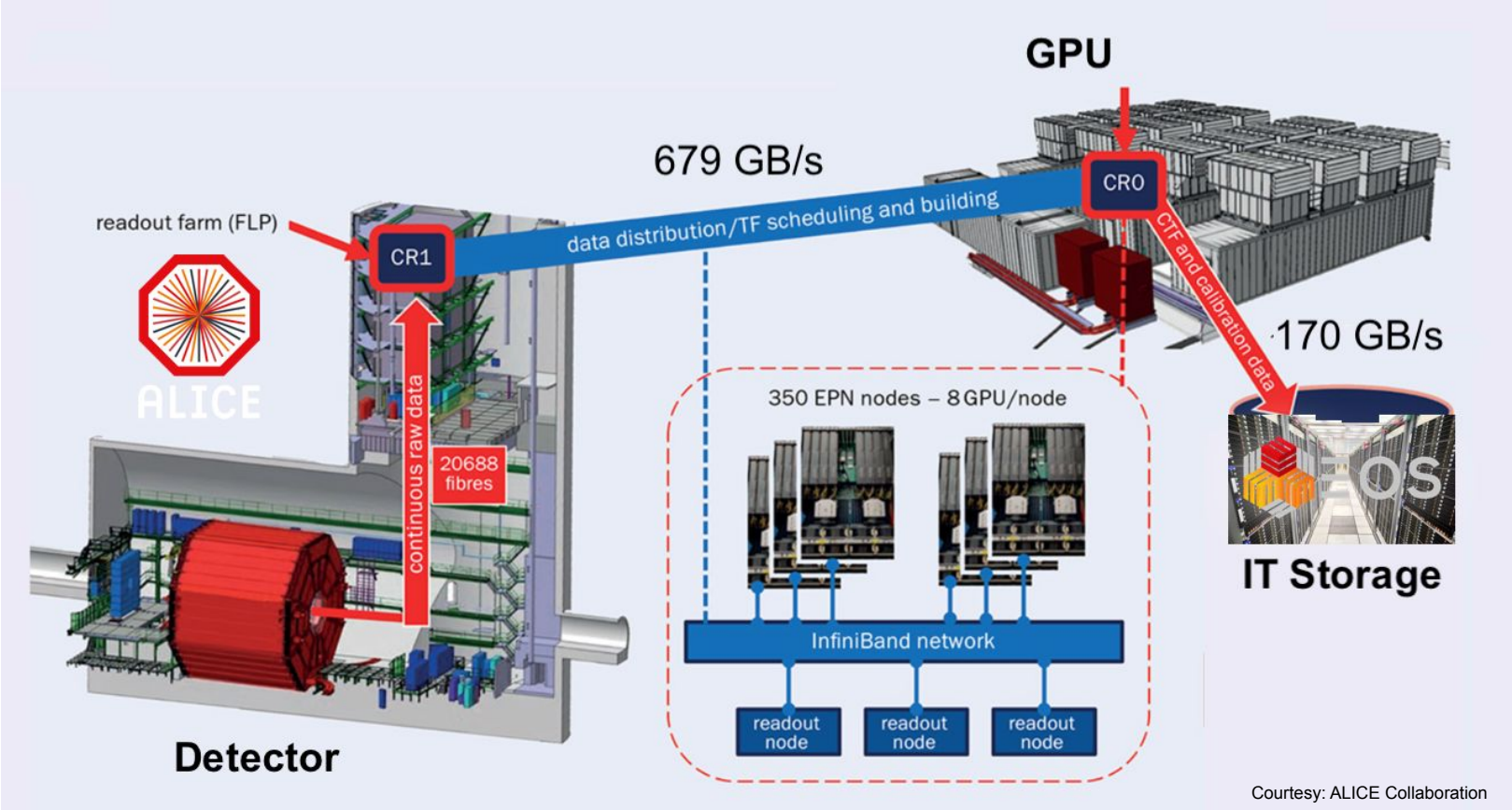
Next-Gen Storage Infrastructure for ALICE

Paving the Road Toward Hi-Luminosity LHC

Presenter: Dr. Andreas-Joachim Peters

October 19 - 25, 2024
CHEP 2024

The ALICE O^2 Use Case



Courtesy: ALICE Collaboration

The ALICE O² Timeline

ALICE & IT

First ALICE WP15 Meeting

Q4 2017

O2 WP15 Meeting

2018

Erasure Coding tests on DAQ instance

Q4 2019

Online/Offline Session 4th EOS workshop

Q1 2020

EOS tests+future WP15 Meeting

Q3 2020

Decision to host O2 in IT

O2 WP15 Meeting

EOS for ALICE 5th EOS workshop

Q1 2021

ALICE/IT Joint Validation Test Plan

ALICE/IT Joint Validation Tests

Q2 2021

LHC/IT Joint Data Challenge

June 2021

IT Hardware

Q2 2020

EOSALICEO2
10 Machine setup
test setup

Q1 2021

EOSALICEO2
26 Machine setup
20% of disks setup

Q3 2021

EOSALICEO2
74 Machine setup
90% of disks setup

10 GE

100 GE

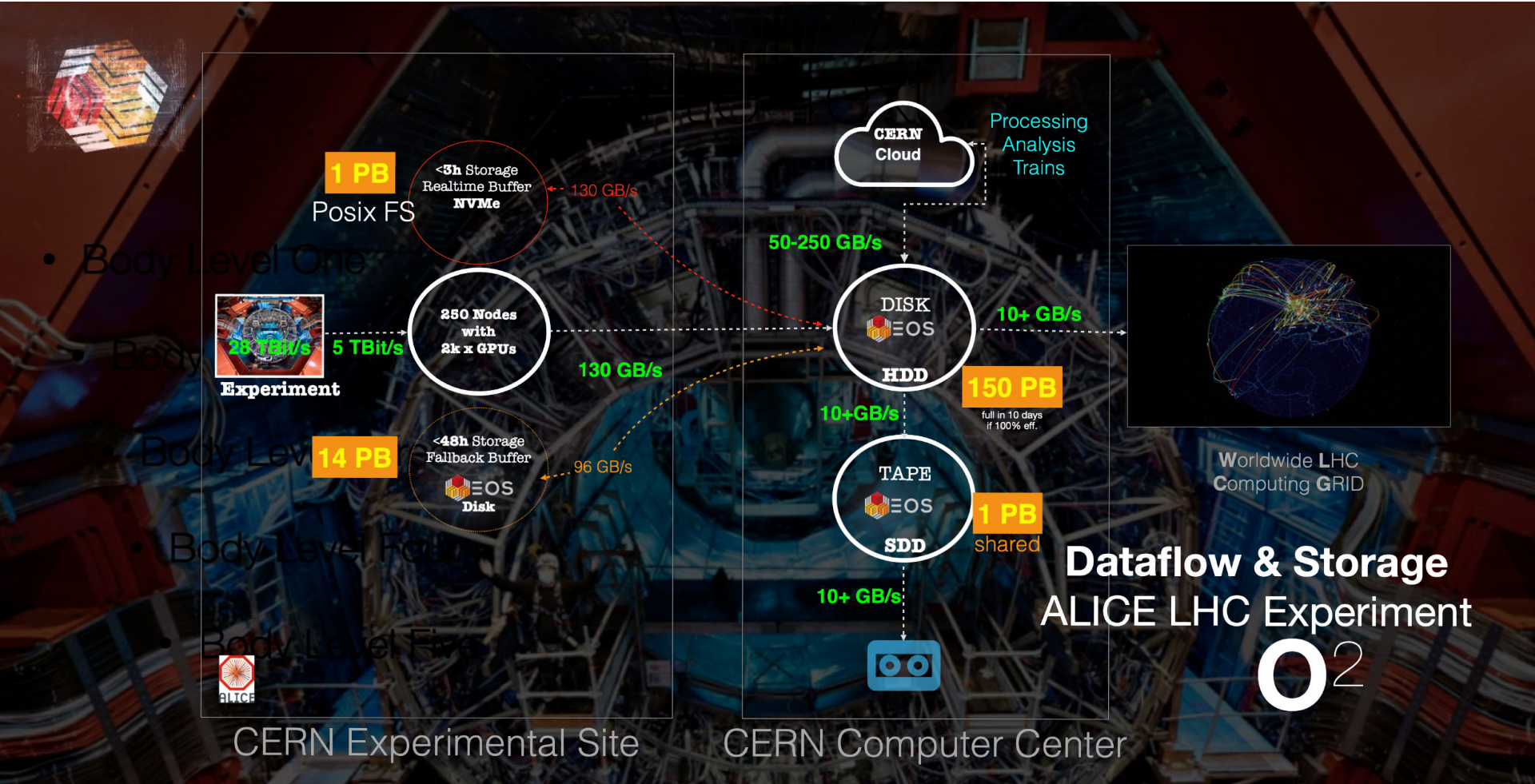
August 2023

EOSALICEO2
125 Machine setup
150% of disks setup

October 2022

EOSALICEO2
75 Machine setup
100% of disks setup

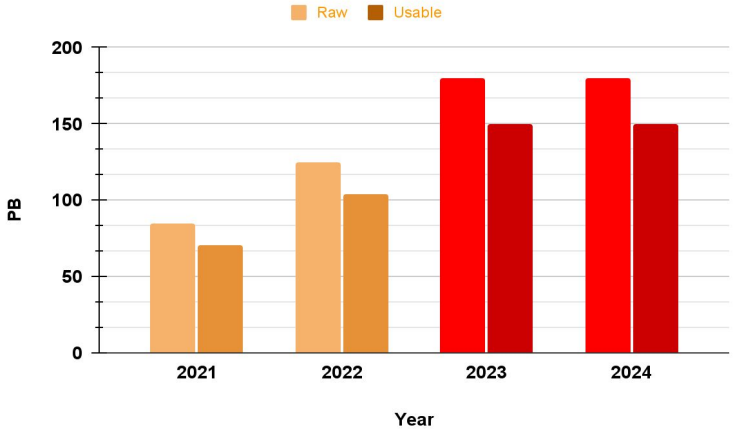
The ALICE O² complete Use Case



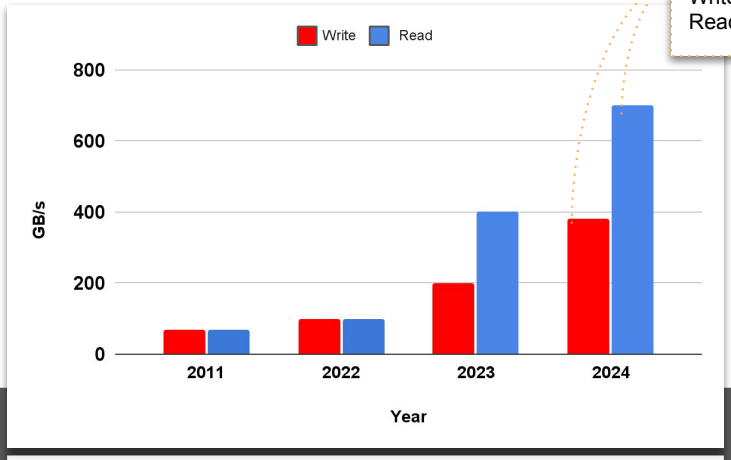
EOS as Storage System for ALICE O^2

EOS Open Storage - Open Source Physics Storage developed by CERN since 2011.

Capacity Evolution



Bandwidth tested



Performance Capacity Ratios 2024
Write: **2.5 MB/s per TB**
Read: **4.6 MB/s per TB**

IO since **2021**

Bytes Read	**1.44 EB**
Bytes Written	**644 PB**
Files Read	**2.7 Billion**
Files Written	**6.98 Billion**
XRootD Protocol	**100%**

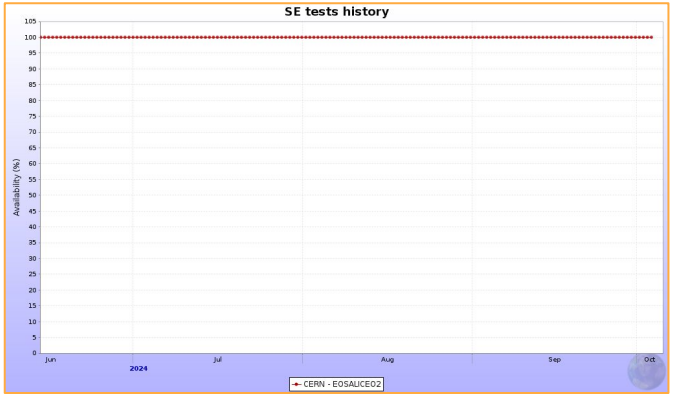
EOS as Storage System for ALICE O^2

100% Availability Run-3 2024

SE tests history

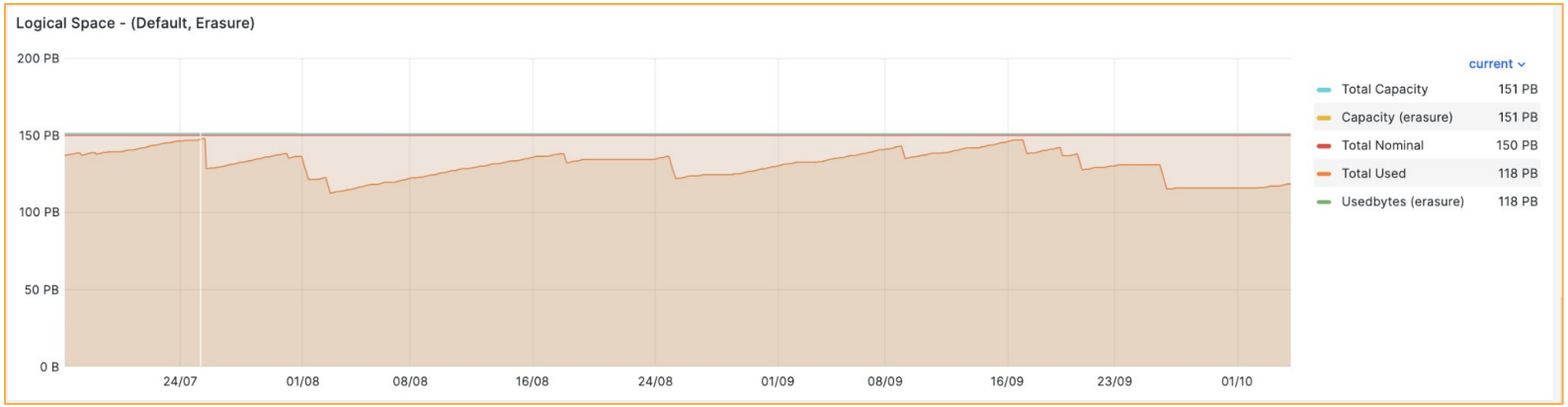
	Series	Last value	Min	Avg	Max
1.	CERN - EOSALICEO2	100	100	100	100
	Total	**100**		**100**	

Availability during run June to October 2024 - measured by ALICE



Availability during run June to October - measured by ALICE

Capacity Usage during run June to October - operating between 75% and 98% full



EOS as Storage System for ALICE O^2 - Value & Failure Rates

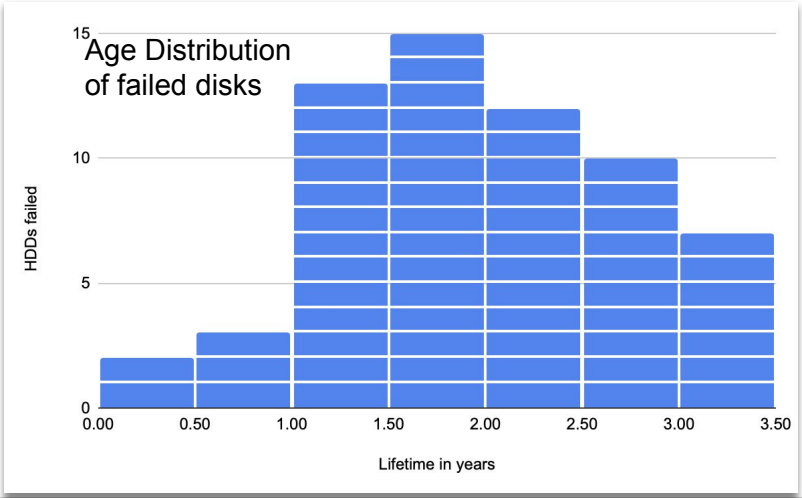
Virtual value created if O^2 would be market price cloud storage **97 M$**

Cost-Matrix	TB*Years	Avg-Drive-Hours	Tot-Drive-Hours	Cloud$-Replica	Cloud$-Erasure
eosaliceo2	466.41 K	23.59 K	281.63 M	58.30 M$	97.17 M$

Azure Calculator Google Calculator AWS Calculator

Model: 20$ / TB*Month
(with 3 year subscription Azure is cheaper!

64 disks removed in population of **12k** in **10 month** = **Failure rate** 0.64%



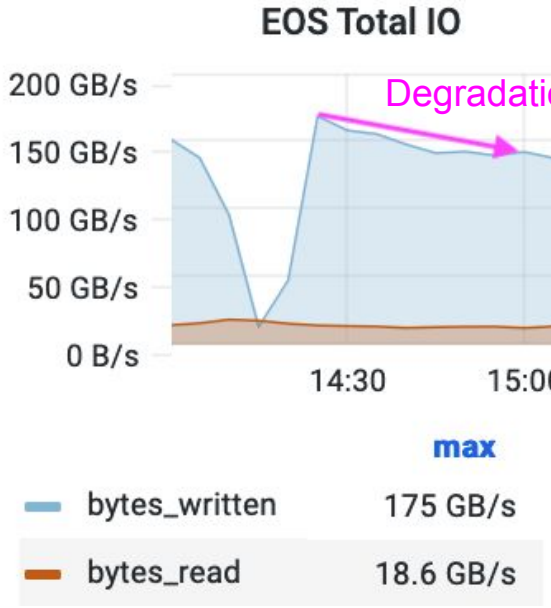
EOS as Storage System for ALICE O^2 - Run-3 Validation Tests

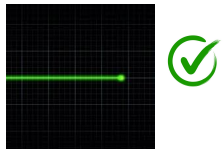
- **O^2 behaviour** in synthetic benchmarks **differs** from production usage with data coming from EPN
  - Subtle stream behaviour - fixed timing
- **Dedicated load tests** with dataflow from **EPNs** end of 2023
  - Scope of the testing
    - target rate of **170 GB/s** injection
    - operate instance at high occupancy > 90% capacity usage
    - operate instance with 50% of drives but same number of storage nodes
      - to investigate feasibility of a possible capacity reduction
- **Synthetic Testing** ingesting data from 100GE clients inside IT
  - Scope of the testing
    - Measure maximum stable write performance at high occupancy > 90% capacity usage

ALICE O^2 running 93% full with 10% of HDDs at 100%

December 2023

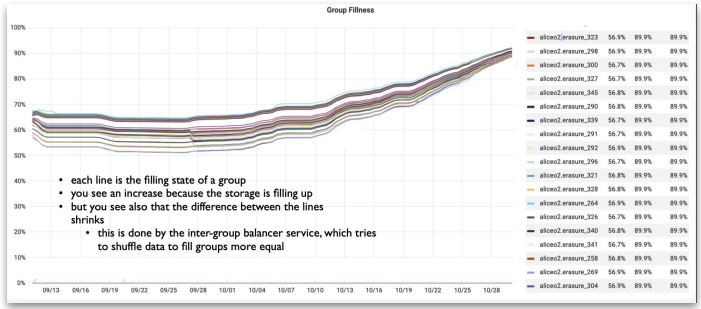
Full chain test from EPNs to O^2





Rebalancing

170 GB/s stable for 1h



- Running with a very full instance bears the risk, that part of the disk population is full at close to 100%, **write bandwidth decays** due to unusable scheduling groups
- **Rebalancing** full instances **stabilizes performance** because even at 99% all HDDs can still write data

ALICE O^2 Reducing capacity from 12.000 to 6.000 HDDs

- We evaluated the possible case to still run at **170 GB/s** writing while we shrink the O^2 instance to half capacity (to reduce cost for ALICE)
  - still to have only half the space is a problem on a different level

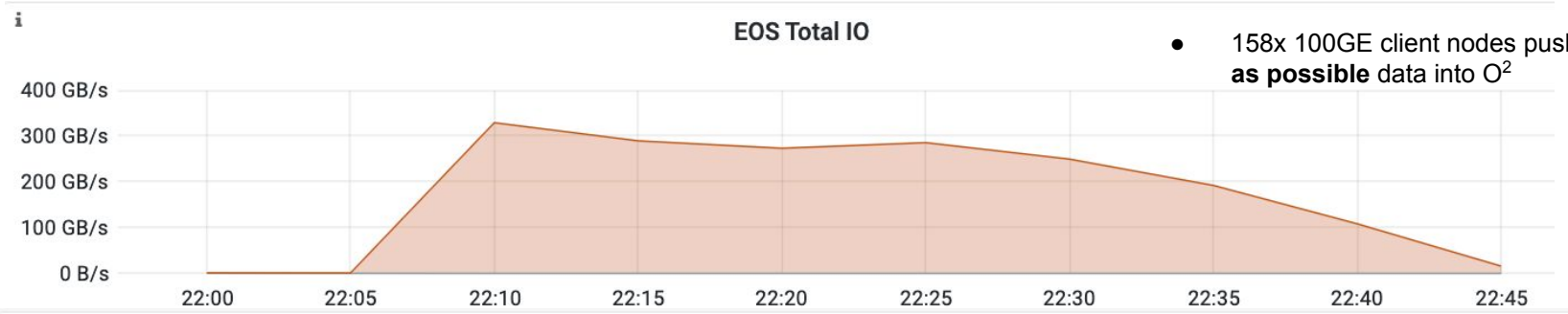
The reduction has **no impact if the number of storage nodes is not reduced**

- essentially the network connectvity = bandwidth is as before
  - No HDD bottleneck with 6.000 HDs at 170 GB/s (~30 MB/s per disk [+20% for EC])

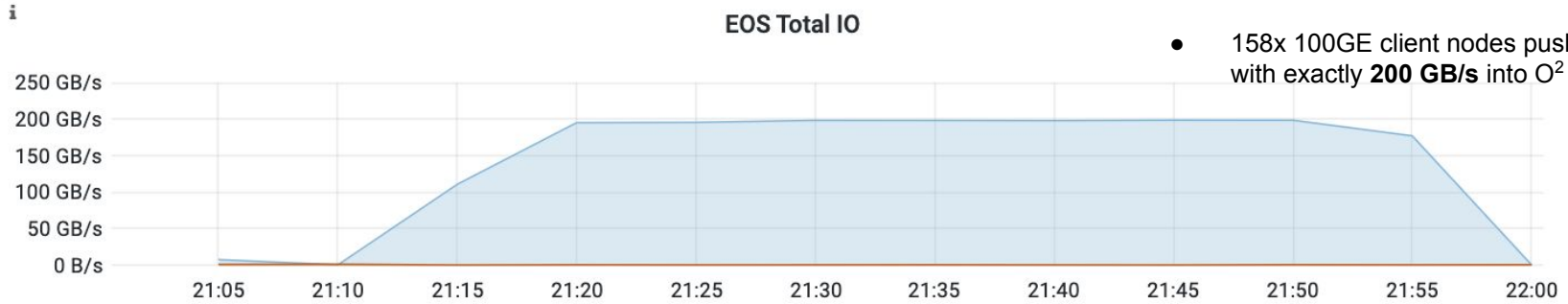
The reduction has also **no negative impact if half of the nodes are removed** if the remaining disks are well balanced

- The network connectivity is now only 50%, same for HDDs performance, but the required 170GB/s are less than 50% of the maximum performance of 380 GB/s with all nodes

ALICE O^2 Unexpected Performance Degradation



- 158x 100GE client nodes pushing **as fast as possible** data into O^2



- 158x 100GE client nodes pushing with exactly **200 GB/s** into O^2

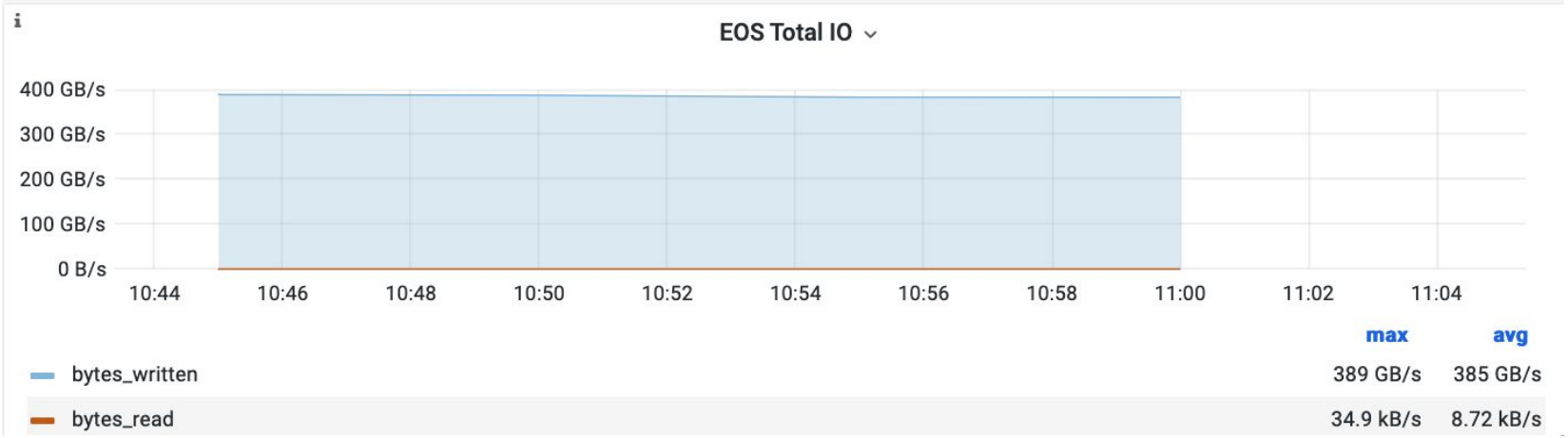
- After tests with EPNs from ALICE we went back to synthetic testing
  - running with identical file sizes and client nodes we measured that the write bandwidth fell 25% short of the expectation and degraded over time (upper plot)
- When we artificially limited the clients to push not more than 200 GB/s
  - it worked perfectly (lower plot)

ALICE O^2 Unexpected Performance Degradation

- An investigation revealed that one of 125 server nodes was experiencing **high packet loss under load** and causing dropped packets in the cluster!
  - faulty interface got replaced
  - We added packet loss monitoring natively into EOS

```
stat.net.tx_dropped             := 0
stat.net.tx_errors              := 0
```

- Final result of synthetic tests: **385 GB/s** writing into O^2



ALICE O^2

- **ALICE O^2 is**
  - the largest single storage system at CERN - **150 PB**
  - the most stable physics storage system at CERN - **100% availability June to October**
  - the only 100% erasure coding physics storage with the smallest overhead **+20%**
  - the most performant storage system at CERN **380 GB/s wr/rd + 700 GB/s rd**/direct
  - the most cost-effective storage system at CERN - still performant when 98% full
  - **really great!**
- **Lessons learned**
  - in a large distributed storage system a single hardware issue can lead to significant and unpredictable performance loss - you always learn something in production!
- **Future Outlook**
  - O^2 is a prototype for **Run-4** storage systems online & offline environments
    - the performance capacity ratio has to be followed with care with ever growing HDD capacities
      - Possible mitigations
        - 400G or bonded 200GE
        - Less disks per front-end [96 => 60 HDDs] at the cost of having more front-end nodes

Thank You!

Authors

Abhishek Lekshmanan (CERN)
Andreas Joachim Peters (CERN)
Cedric Caffy (CERN)
David Smith (CERN)
Elvin Alin Sindrilaru (CERN)
Gianmaria Del Monte (CERN)
Guilherme Amadio (CERN)
Ioanna Vrachnaki
Luca Mascetti (CERN)
Maria Arsuaga Rios (CERN)/
Volodymyr Yurchenko (National Academy of Sciences of Ukraine (UA))

A special thanks to the ALICE team for the successful collaboration!