



EOS Service @ CERN

Maria Arsuaga-Rios IT-SD-PDS



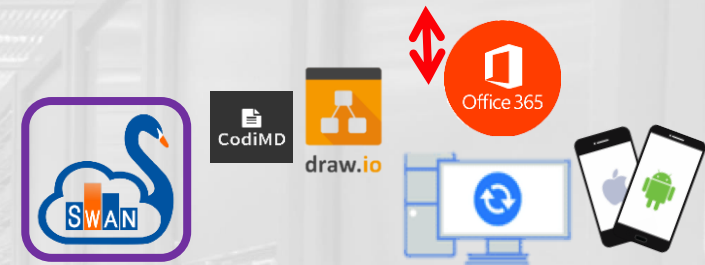
EOS Service @ CERN



General purpose services

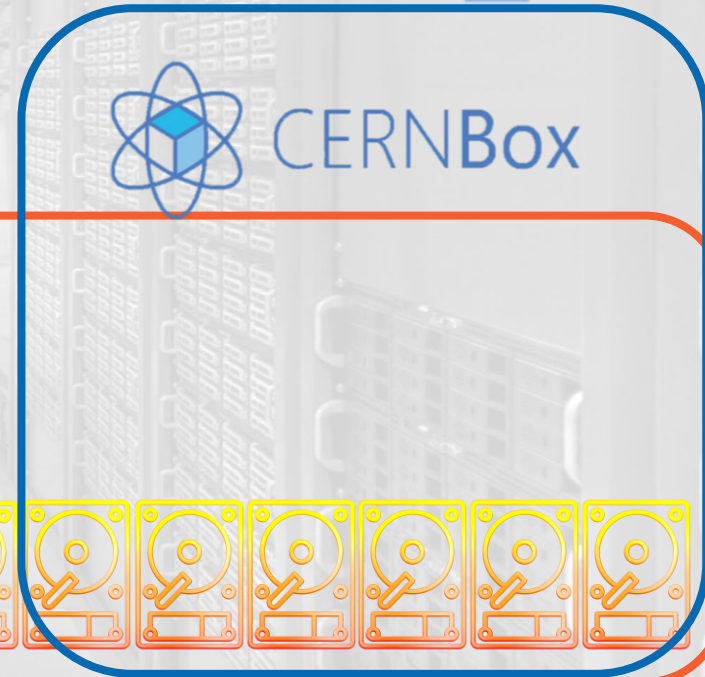
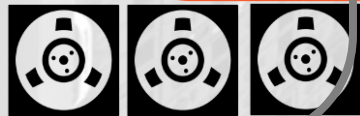


Physics services



Tape services

CERN
Tape Archive



EOS Service @ CERN

EOS instances

Cernbox: 9
(5 homes + 3 projects + 1 media)

Physics: 9
(6 LHC, 3 non-LHC,)

CTA: 8
(4 LHC, 2 non-LHC, 1 backup, 1 internal repack)

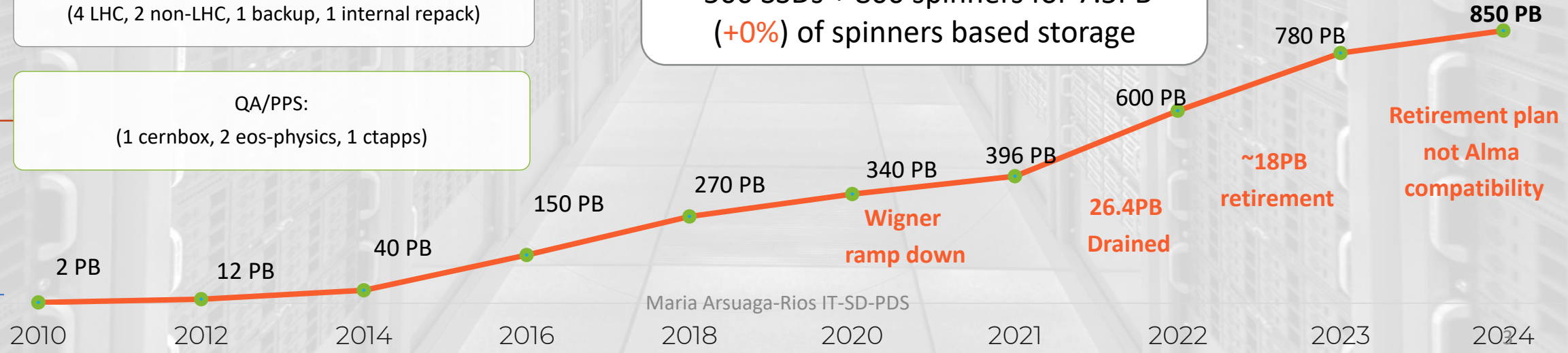
QA/PPS:
(1 cernbox, 2 eos-physics, 1 ctapps)

831 (+4%)
Storage Nodes

66 K(10%)
Disks

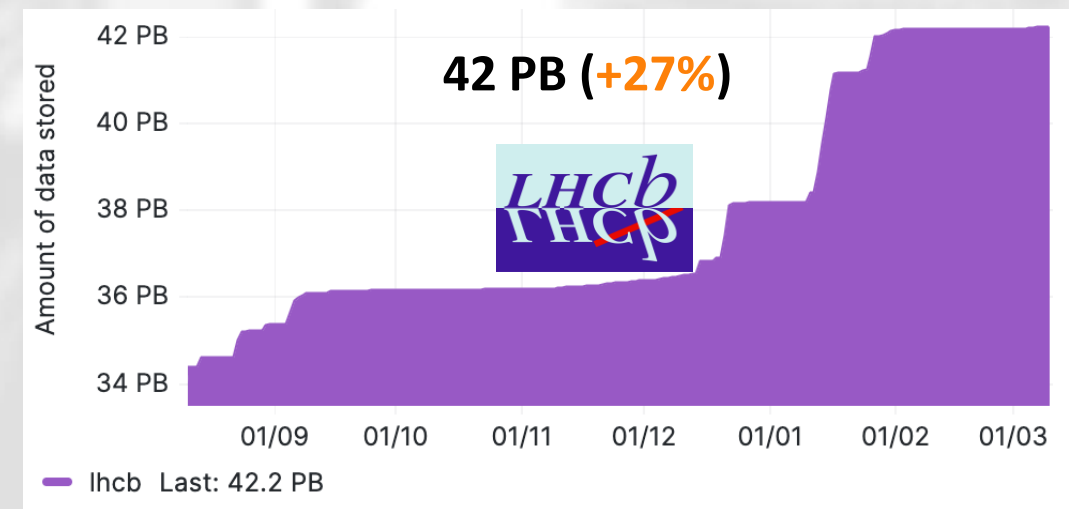
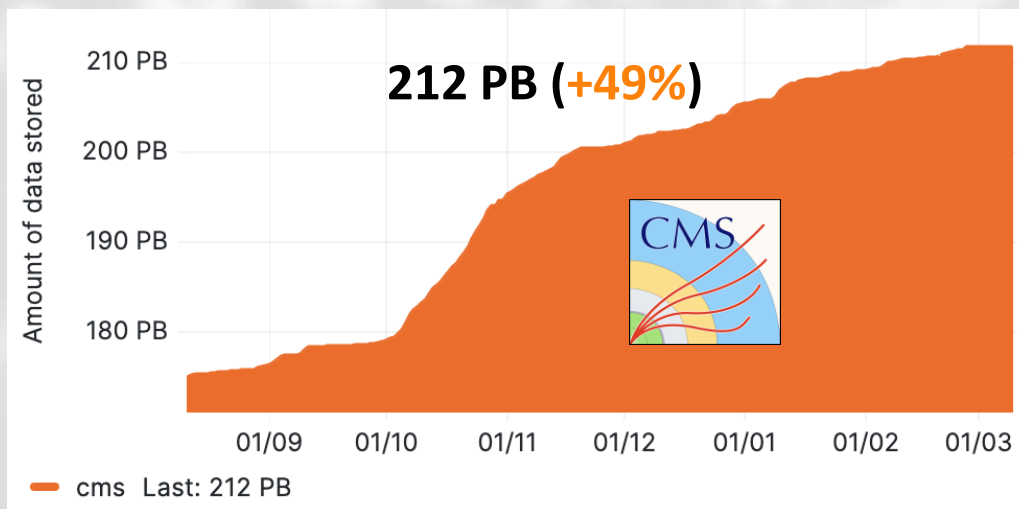
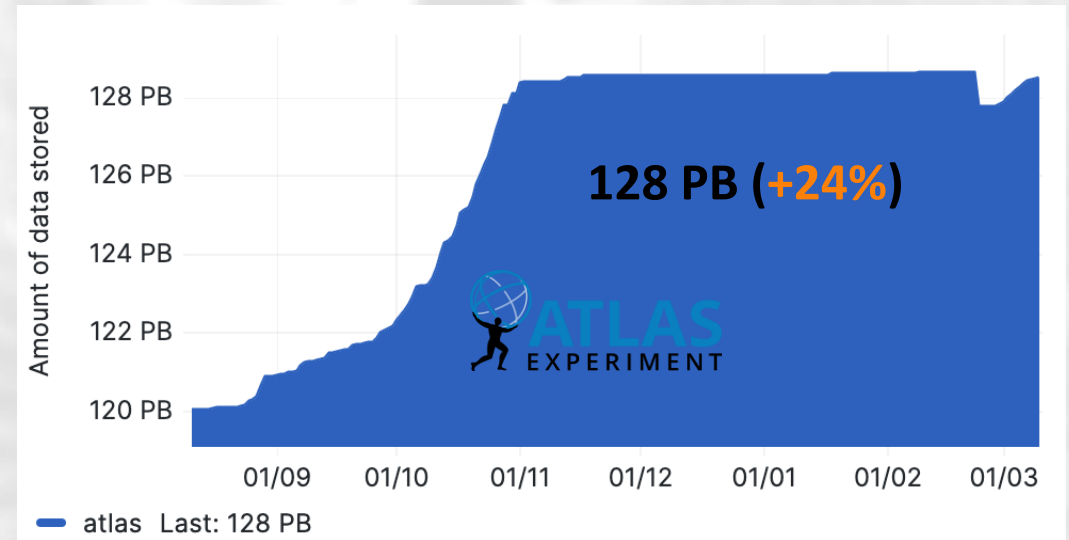
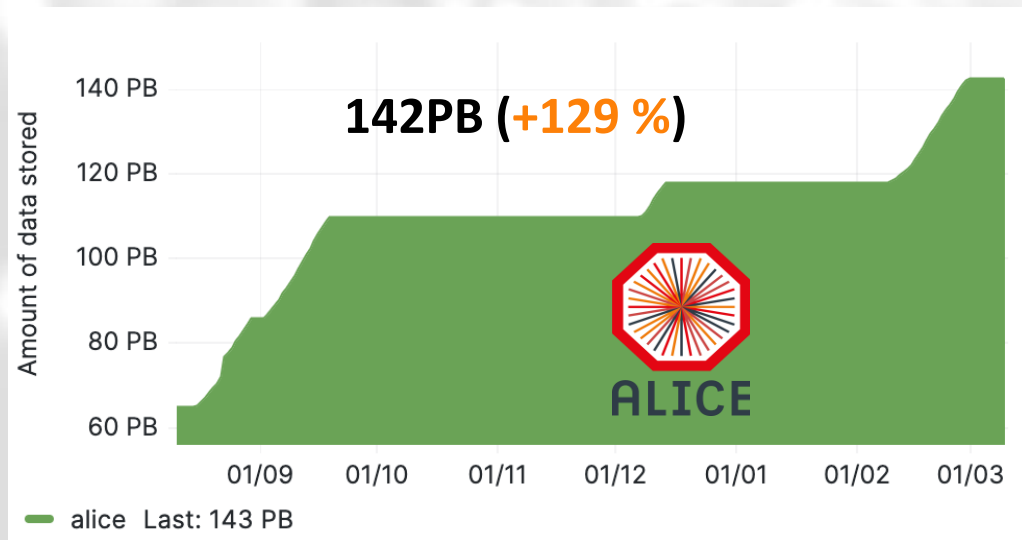
4 LHC:
500 SSDs + 800 spinners for 7.5PB
(+0%) of spinners based storage

PLANS:
~20PB
+100PB

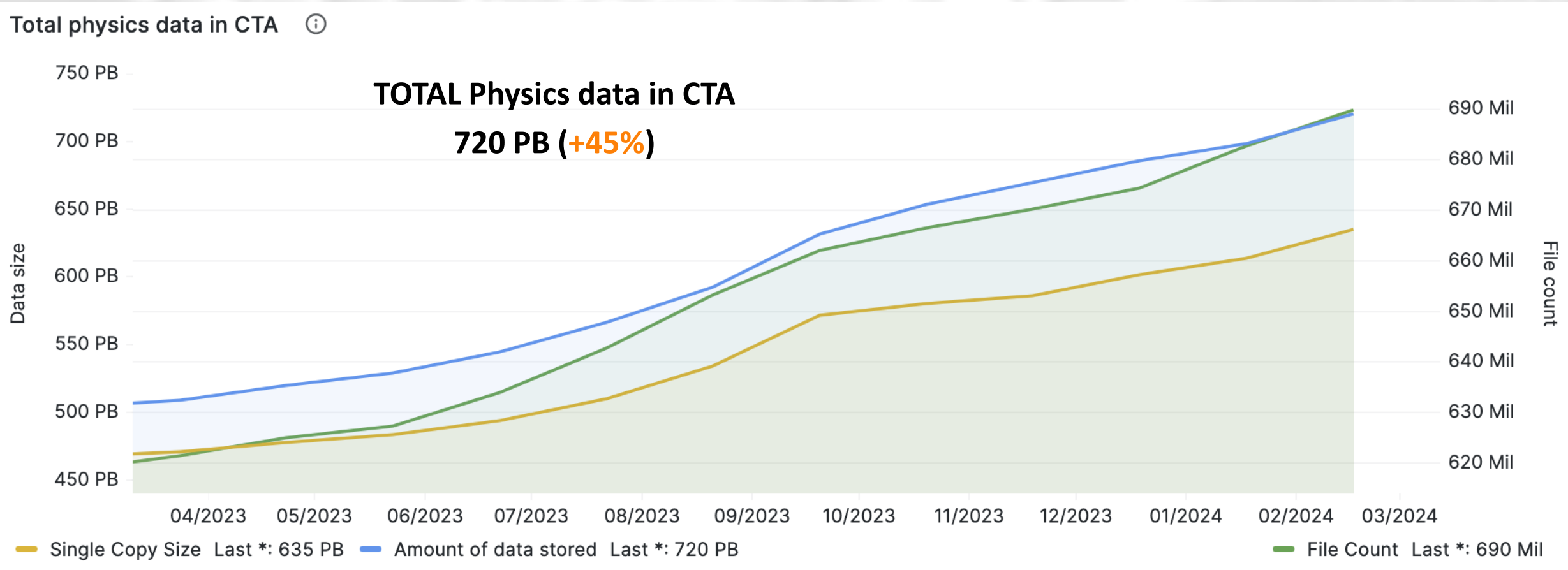


Tape Storage Capabilities

The primary purpose of CTA is to provide reliable, long-term archival storage of the custodial copy of the data from all of the physics experiments at CERN.



Tape Storage Capabilities



General purpose services



9

Instances

Homes, Projects and Media



3.37 Bill

Number of Files

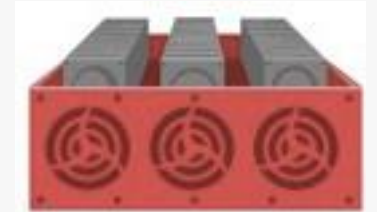
+13%



349 Mil

Number of Directories

+59%



33.47 PB

Total Space

+0.5%



Backup in a different technology

RESTIC

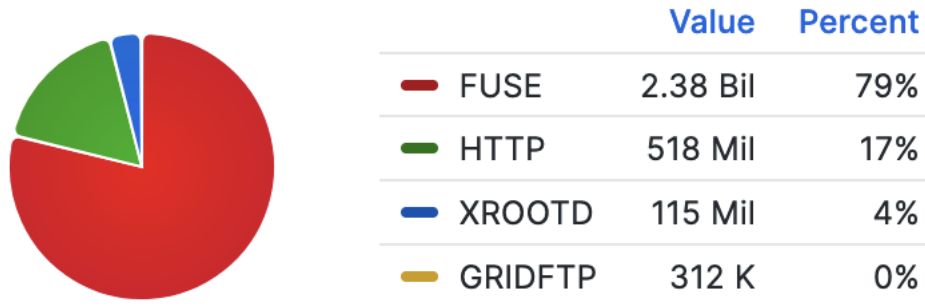


Maria Arsuaga-Rios IT-SD-PDS

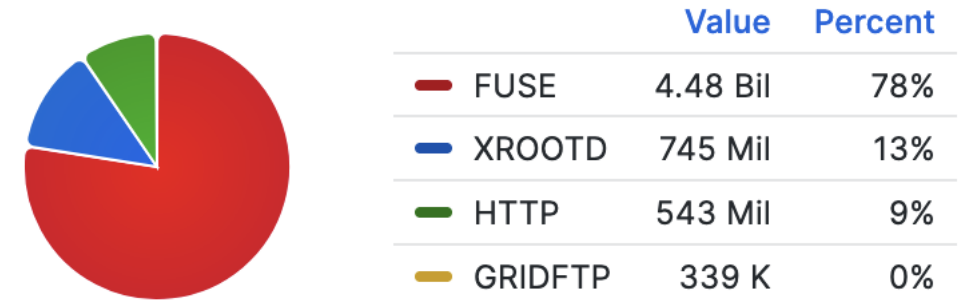
General purpose services



Ingestion: Amount of files written



Export: Amount of files read



**FUSE is the preferred protocol for writes and reads in CERNBOX (2023)
Even increasing its popularity by a 3% and 2% for writes and reads with respect to 2022.**

EOS for Physics: Numbers



6 + 3

Instances



3.10 Bill

Number of Files

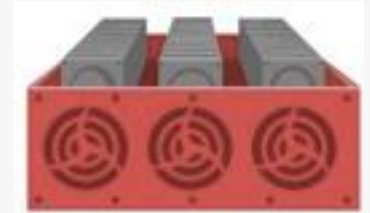
+14%



287 Mil

Number of Directories

+15%



715 PB

Total Space

+8%

| | Total space | Used space | Number of files |
|--------------|------------------|------------------|-------------------|
| ATLAS | 94.56 PB | 79.58 PB | 266 Mil |
| CMS | 102.97 PB | 77.67 PB | 240 Mil |
| ALICE | 116.73 PB | 107.32 PB | 840 Mil |
| LHCb | 70.14 PB | 44.43 PB | 1.13 Bill |
| Public & AMS | 134.25 PB | 108.97 PB | 589 Mil |
| ALICEO2 | 181.99 PB | 162.66 PB | 30.1 Mil |
| TOTAL | 700.64 PB | 580.63 PB | 3095.1 Mil |

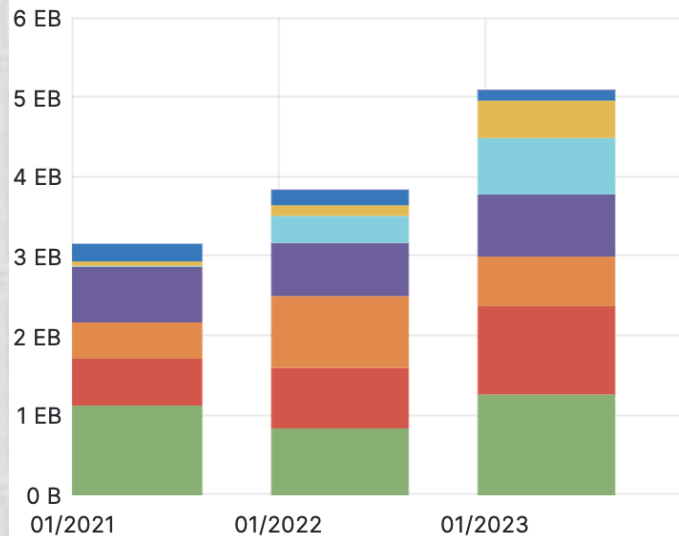
Production: ALICE, ALICEO2, ATLAS, CMS, LHCb, and Public
Dedicated: ALICEP2, AMS, EULAKE
Testing/Internal: Pilot, PPs, Backup (not included in the numbers)

EOS for Physics: Usage statistics



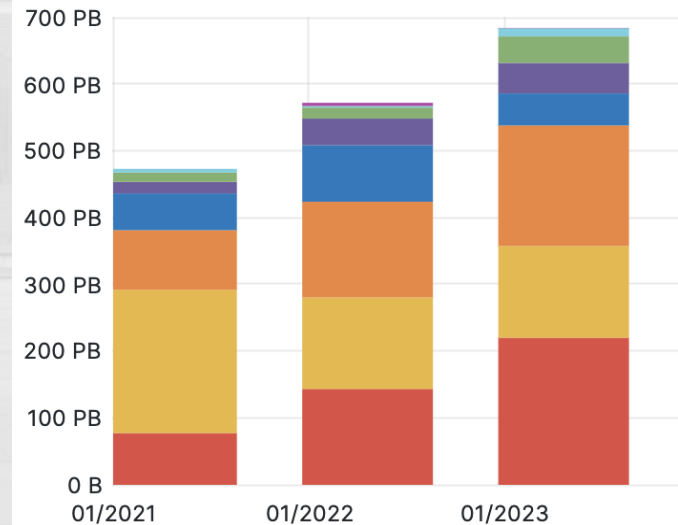
EOS served 5 Exabytes (+34%) of data from the physics instances and ~0.7 Exabytes (+16%) received

Export: Amount of bytes read



| | max | avg | current |
|---------|---------|---------|---------|
| alice | 1.27 EB | 811 PB | 7.10 PB |
| cms | 1.11 EB | 615 PB | 4.08 PB |
| atlas | 904 PB | 497 PB | 1.14 PB |
| public | 781 PB | 536 PB | 2.40 PB |
| ams02 | 711 PB | 267 PB | 229 TB |
| aliceo2 | 468 PB | 165 PB | 1.11 PB |
| lhcb | 223 PB | 138 PB | 1.34 PB |
| p2 | 4.03 PB | 1.47 PB | 251 MB |

Ingestion: Amount of bytes written



| | max | avg | current |
|---------|---------|---------|---------|
| cms | 221 PB | 111 PB | 412 TB |
| aliceo2 | 215 PB | 123 PB | 1.51 PB |
| atlas | 181 PB | 103 PB | 560 TB |
| lhcb | 84.7 PB | 47.2 PB | 1.22 PB |
| public | 45.7 PB | 25.7 PB | 97.2 TB |
| alice | 39.7 PB | 17.6 PB | 187 TB |
| ams02 | 12.4 PB | 5.08 PB | 16.4 TB |
| p2 | 4.80 PB | 1.74 PB | 251 MB |

EOS for Physics: Usage statistics



Ingestion is caused by more than just data taking!

CMS Ingestion: Amount of bytes written



| | max ▾ | avg | current |
|--------|---------|---------|---------|
| Others | 112 PB | 57.0 PB | 947 TB |
| CMST0 | 75.7 PB | 64.7 PB | 75.7 PB |
| Point5 | 32.9 PB | 31.6 PB | 32.9 PB |

ATLAS Ingestion: Amount of bytes written



| | max ▾ | avg | current |
|---------|---------|---------|---------|
| Others | 143 PB | 84.9 PB | 1.13 PB |
| Point1 | 20.8 PB | 11.7 PB | 728 GB |
| AtlasT0 | 17.3 PB | 9.75 PB | 122 GB |

CMS Ingestion: Amount of bytes written



| | |
|--------|-----|
| Others | 51% |
| CMST0 | 34% |
| Point5 | 15% |

Atlas Ingestion: Amount of bytes written



| | |
|---------|-----|
| Others | 79% |
| Point1 | 12% |
| AtlasT0 | 10% |

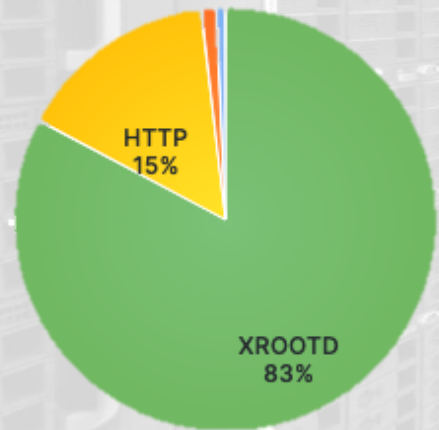
EOS for Physics: Protocol statistics



Most used protocol for writes volume: XRootD in 2023

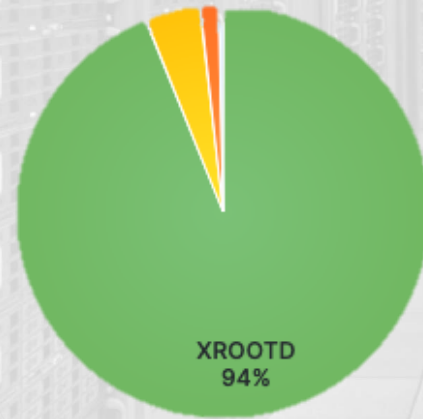
Total writes per protocol

2022



| | | |
|---------|----------------|--------------|
| XROOTD | Value: 477 PB | Percent: 83% |
| HTTP | Value: 88.5 PB | Percent: 15% |
| FUSE | Value: 6.20 PB | Percent: 1% |
| GRIDFTP | Value: 3.54 PB | Percent: 1% |

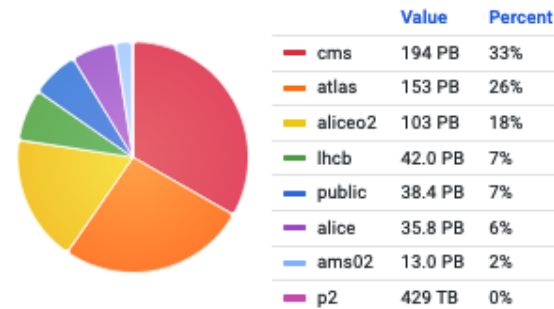
2023



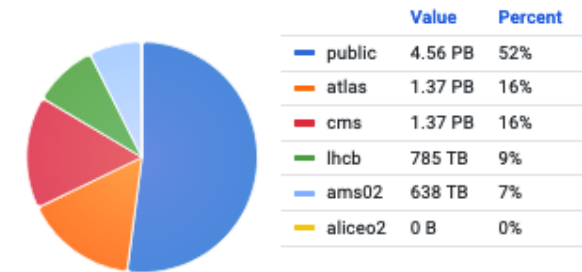
| | | |
|---------|----------------|--------------|
| XROOTD | Value: 629 PB | Percent: 94% |
| HTTP | Value: 28.3 PB | Percent: 4% |
| FUSE | Value: 9.30 PB | Percent: 1% |
| GRIDFTP | Value: 1.57 PB | Percent: 0% |

Total writes per protocol and instance

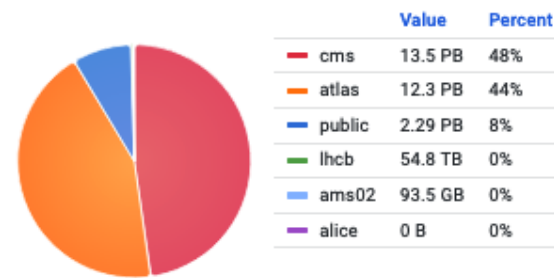
Ingestion: Amount of bytes written via XROOTD



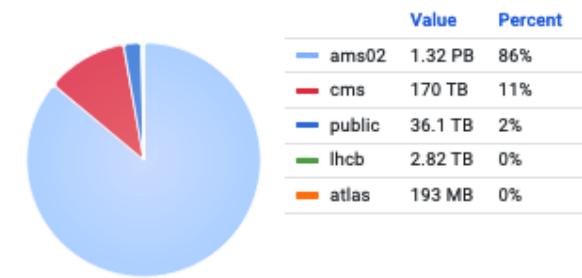
Ingestion: Amount of bytes written via FUSE



Ingestion: Amount of bytes written via HTTP



Ingestion: Amount of bytes written via GridFTP

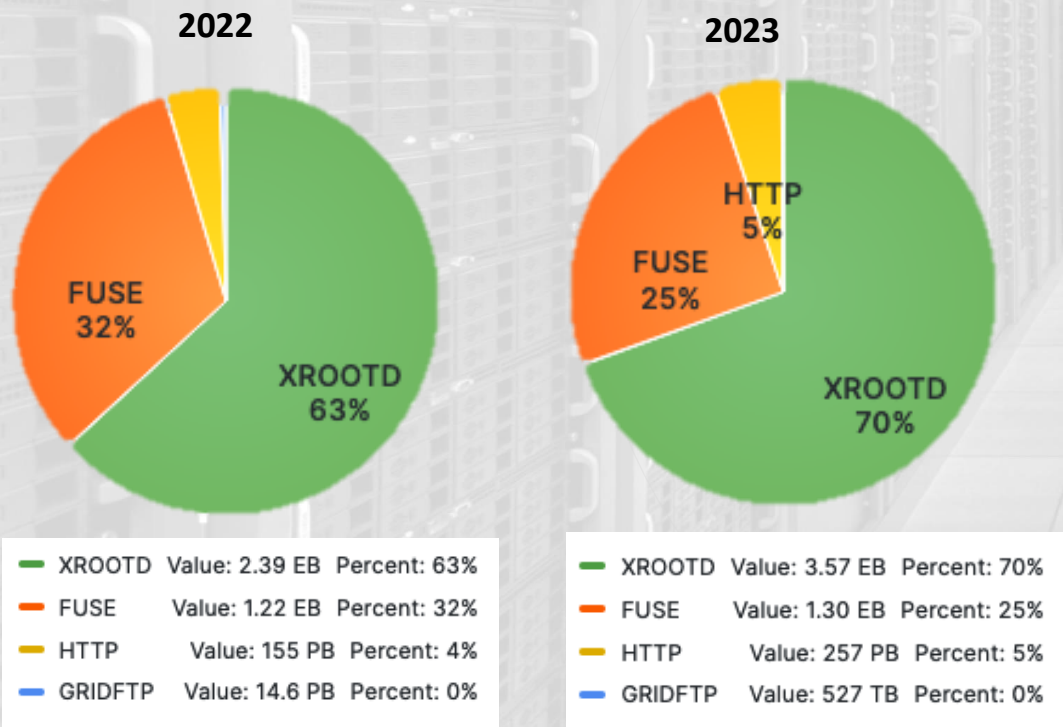


EOS for Physics: Protocol statistics

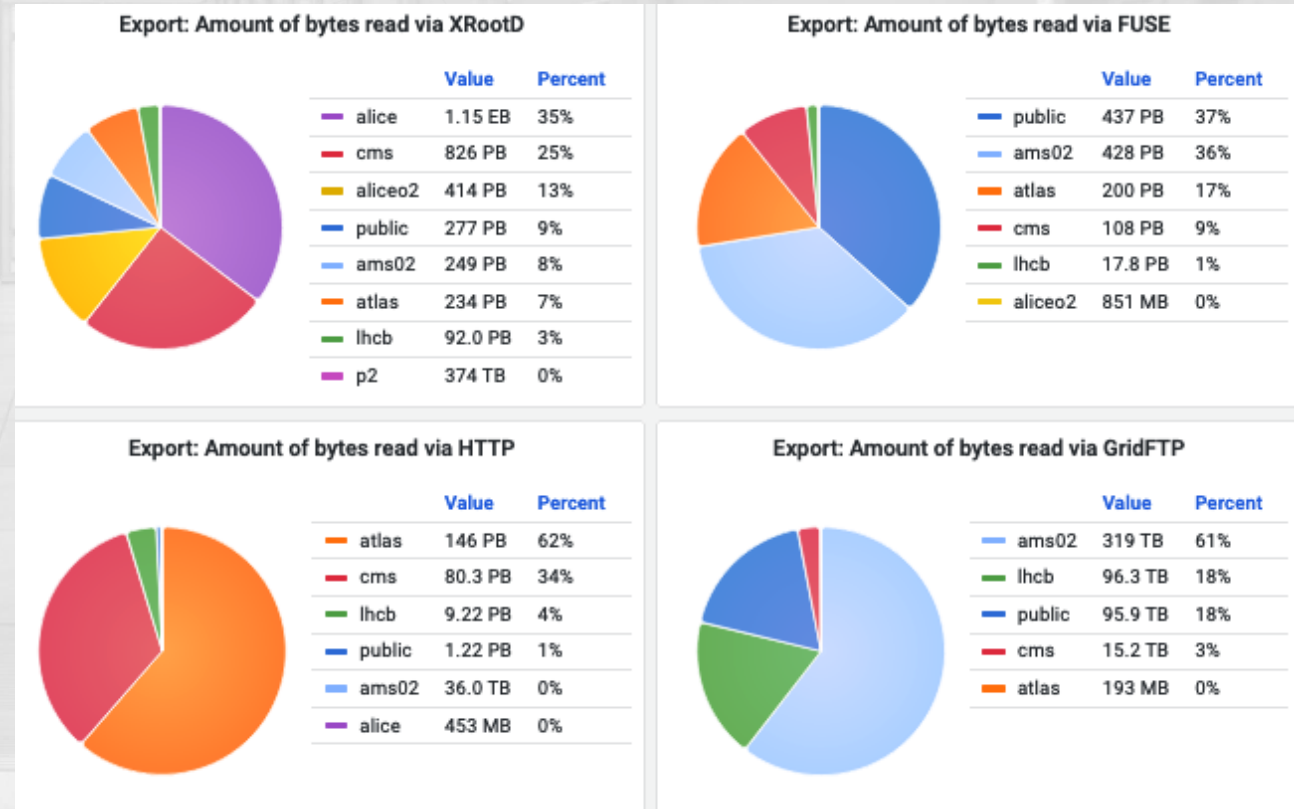


Most used protocol for reads: XRootD and FUSE in 2023.

Total reads per protocol



Total reads per protocol and instance



EOS for Physics: Protocol statistics



REPORT LOG EXAMPLE FROM CMS

```
log=7a9829c6-a774-11ee-a319-
a4bf0179624c&path=/eos/cms/store/group/phys_heavyions/subehera/d24_sub3f_0/PAMinimumBias5/RecoSkim2016_PbpMB5_d24_
sub3f_0_v10/231230_092528/0000/cumulants_3sub_f_525.root&fstpath=/data29/0006a368/1034f271d&ruid=118183&rgid=1399&t
d=nobody&host=st-048-
bb8d81c3.cern.ch&lid=1048850&fid=4350486301&fsid=26599&ots=1703982920&otms=997&cts=1703982921&ctms=174&nrc=0&nw
c=60&rb=0&rb_min=0&rb_max=0&rb_sigma=0.00&rv_op=0&rvb_min=0&rvb_max=0&rvb_sum=0&rvb_sigma=0.00&rs_op=0&rsb_mi
n=0&rsb_max=0&rsb_sum=0&rsb_sigma=0.00&rc_min=0&rc_max=0&rc_sum=0&rc_sigma=0.00&wb=62855914&wb_min=989930&w
b_max=1048576&wb_sigma=7507.81&sfwdb=0&sbwdb=0&slfwdb=0&slbwdb=0&nfwds=0&nbwds=0&nxfwds=0&nxbwds=0&usage
=71.72&iot=177.254&idt=50.128&lrt=0.000&lrvt=0.000&lwt=105.918&ot=9.298&ct=11.910&rt=0.00&rvt=0.00&wt=40.93&osize=0&cs
ize=62855914&delete_on_close=0&prio_c=2&prio_l=4&prio_d=1&forced_bw=0&ms_sleep=0&ior_err=0&iow_err=0&sec.prot=https&
sec.name=subehera&sec.host=[2001:1458:301:4e::100:89]&sec.vorg=&sec.grps=&sec.role=&sec.info=&sec.app=http
```

Ignoring internal operations in sec.app:

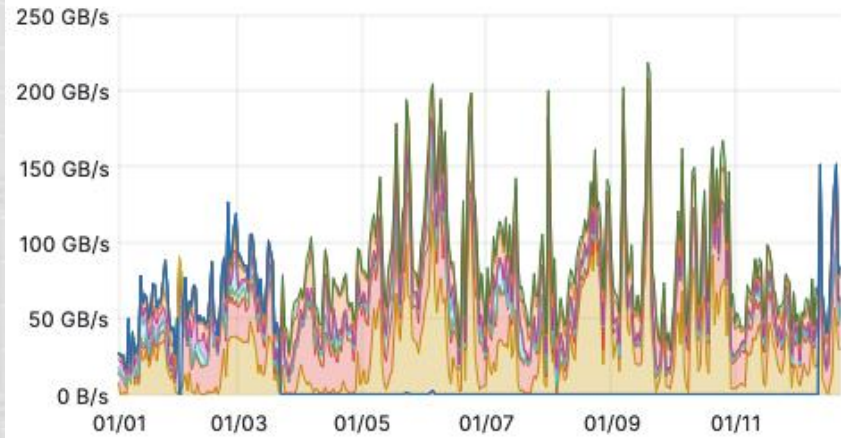
```
(?!eos/drain|eos/converter|eos/groupbalancer|eos/groupdrainer|eos/fsck|groupbalanc
er|drain|groupdrainer)(.*)
```

EOS for Physics: Network statistics



During 2023, we were able to achieve more than 200GB/s of ingest and more than 500GB/s of export.

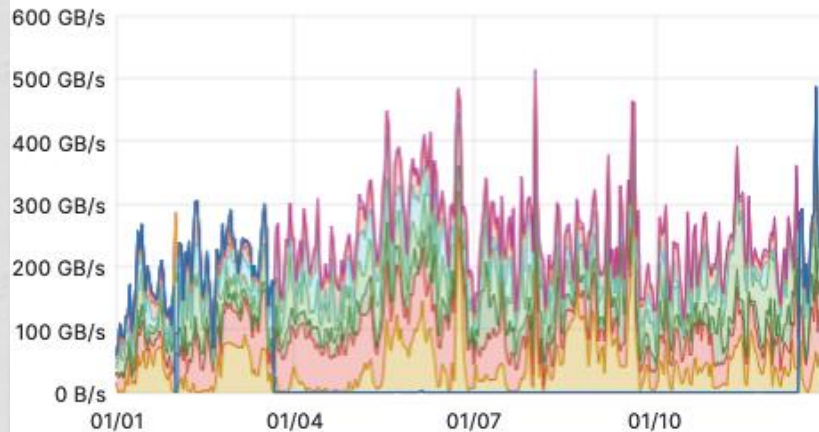
Cluster Network Rates (in)



| | max | avg | current |
|---------------------------|-----------|-----------|-----------|
| aliceo2.inratemib.erasure | 170 GB/s | 28.9 GB/s | 28.7 GB/s |
| cms.inratemib | 64.5 GB/s | 23.1 GB/s | 27.7 GB/s |
| alice.inratemib | 41.8 GB/s | 5.36 GB/s | 5.92 GB/s |
| ams02.inratemib | 38.8 GB/s | 3.89 GB/s | 814 MB/s |
| lhcb.inratemib | 35.7 GB/s | 3.88 GB/s | 1.32 GB/s |
| atlas.inratemib | 30.4 GB/s | 13.8 GB/s | 15.6 GB/s |
| public.inratemib | 20.6 GB/s | 4.60 GB/s | 4.53 GB/s |
| aliceo2.inratemib | 4.80 GB/s | 1.09 GB/s | 0 B/s |

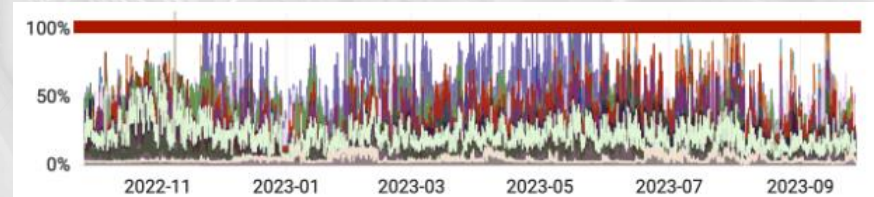
```
[root@eoscms-ns-ip563 (mgm:master
mq:master) ~]$ eos space ls -m
...
sum.stat.net.inratemib=17826
sum.stat.net.outratemib=62056
...
```

Cluster Network Rates (out)



| | max | avg | current |
|----------------------------|-----------|-----------|-----------|
| aliceo2.outratemib.erasure | 340 GB/s | 43.9 GB/s | 59.6 GB/s |
| cms.outratemib | 178 GB/s | 54.7 GB/s | 73.4 GB/s |
| public.outratemib | 173 GB/s | 30.0 GB/s | 45.5 GB/s |
| alice.outratemib | 157 GB/s | 47.7 GB/s | 13.6 GB/s |
| ams02.outratemib | 94.7 GB/s | 27.8 GB/s | 14.4 GB/s |
| atlas.outratemib | 66.2 GB/s | 28.1 GB/s | 19.3 GB/s |
| lhcb.outratemib | 34.9 GB/s | 6.65 GB/s | 3.37 GB/s |
| aliceo2.outratemib | 9.96 GB/s | 2.12 GB/s | 0 B/s |

Saturations on data outgoing traffic

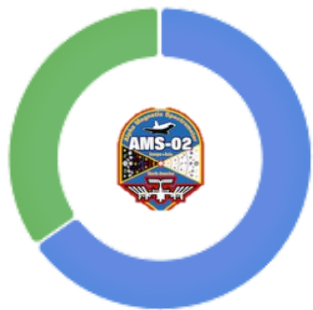


EOS for Physics: Space optimization



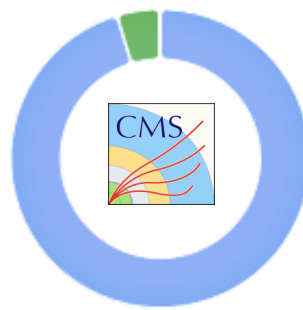
Thanks to the erasure coding layout we saved 47.5PB in 2023

[EOSAMS] Erasure Coding vs Replica percentage



| | Value | Percent |
|-----------|---------|---------|
| — replica | 14.7 PB | 65% |
| — raid6 | 7.80 PB | 35% |
| — plain | 643 kB | 0% |

[EOSCMS] Erasure Coding vs Replica percentage



| | Value | Percent |
|-----------|---------|---------|
| — replica | 38.2 PB | 95% |
| — raid6 | 1.82 PB | 5% |
| — plain | 153 GB | 0% |

[EOSALICEo2] Erasure Coding vs Replica percentage



| | Value | Percent |
|-----------|---------|---------|
| — raid6 | 134 PB | 100% |
| — replica | 79.4 GB | 0% |
| — qrain | 2.15 GB | 0% |
| — plain | 20.5 kB | 0% |

Space optimized



| | Value |
|--|---------|
| — ALICEO2 Savings with Erasure Coding 10+2 | 44.6 PB |
| — AMS Savings with Erasure Coding 8+2 | 2.34 PB |
| — CMS Savings with Erasure Coding 10+2 | 606 TB |

> eos space inspector -m

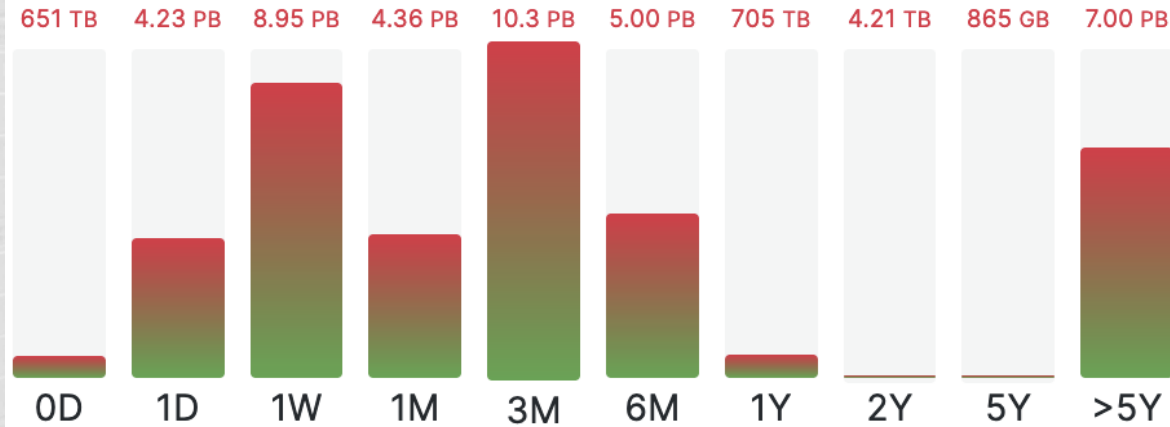
key=last layout=20640b42 **type=raid6 nominal_stripes=12** checksum=adler32
 blockchecksum=crc32c blocksize=1M locations=31971843 nolocation=998
 repdelta:+1=3 repdelta:+11=2 repdelta:-1=141 repdelta:-11=1 repdelta:-
 12=998 repdelta:-2=1 repdelta:0=2664183 shadowdeletion=8213
 unlinkedlocations=8814 **volume=1378893992629426** zerosize=6

EOS for Physics: Space Optimization

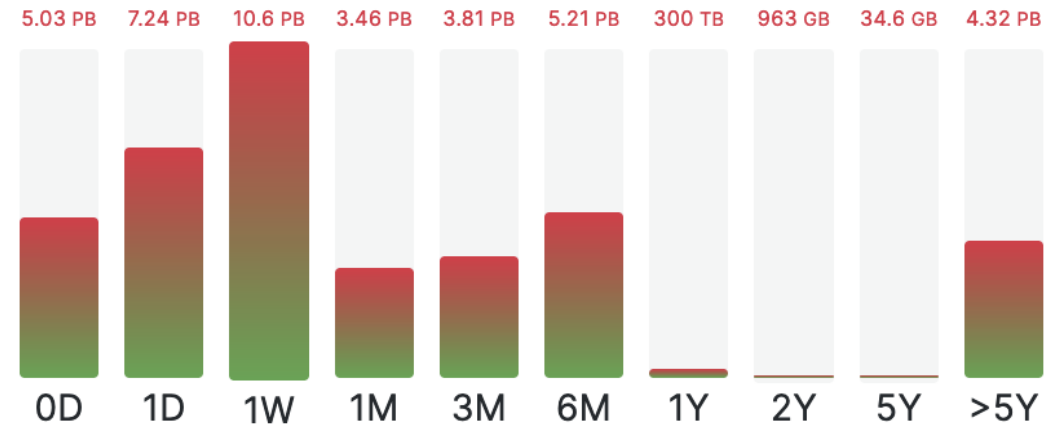


New metrics: Identify areas that have not been accessed as prospects for tape migration

[EOSATLAS] Access Time Volume



[EOSCMS] Access Time Volume



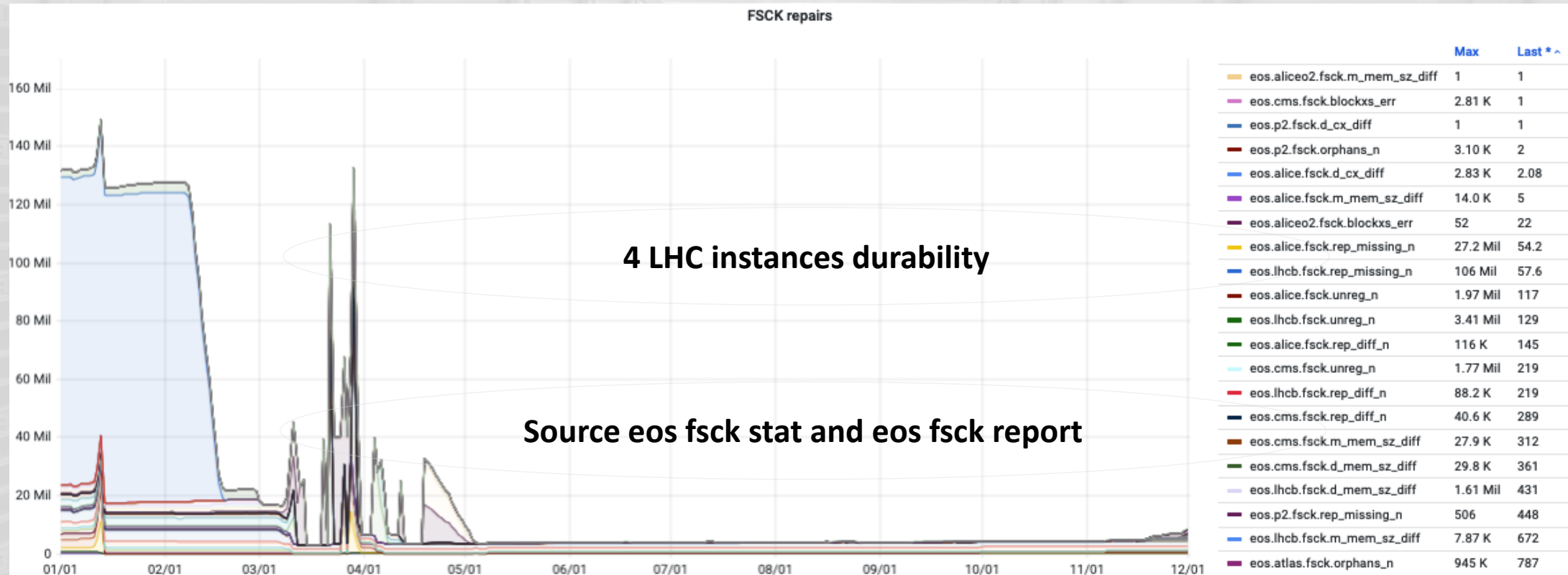
Enable/disable the file inspector [default=off]
space config <space-name> space.inspector=on|off

```
> eos space inspector -m  
key=last tag=accesstime::volume bin=0 value=5033318851015405
```


EOS for Physics: FSKK statistics



99.9955% decrease inconsistencies in the 4 LHC instances

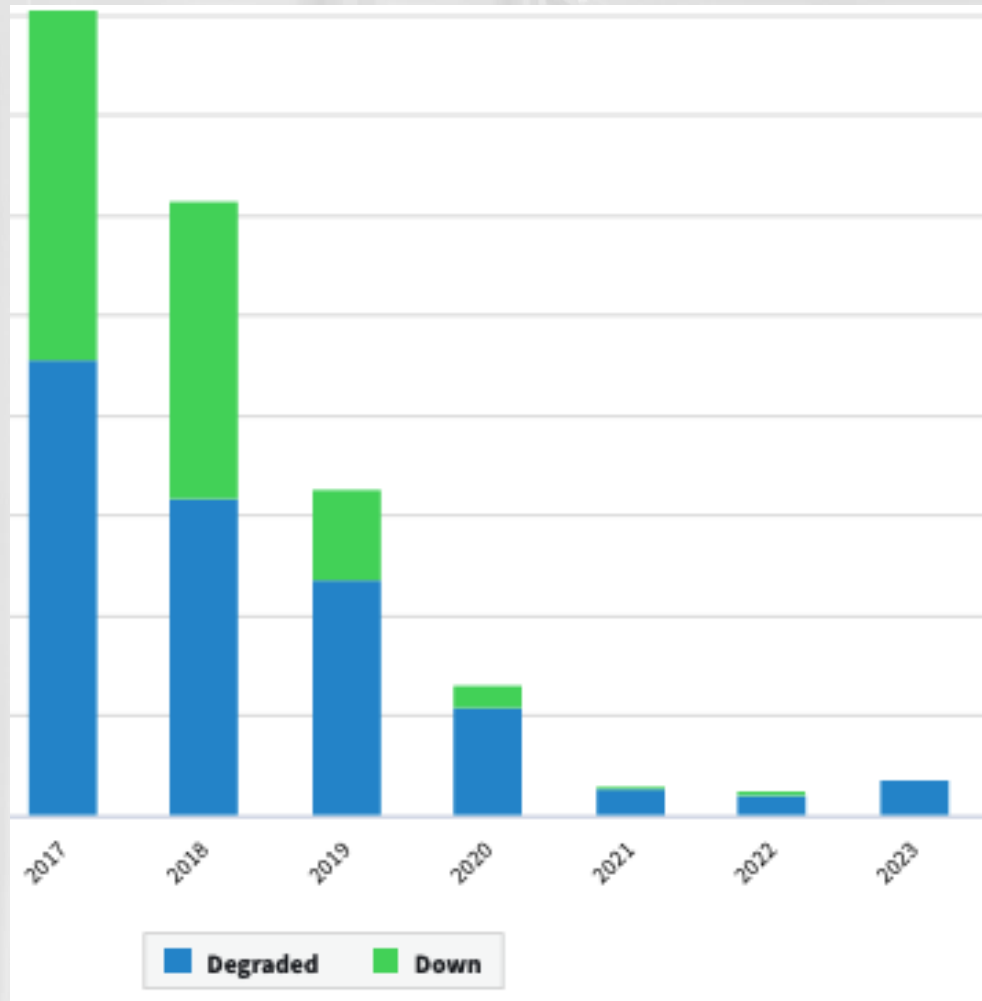


EOS for Physics: Availability



11 days -

Total duration (sum*)



EOS Physics Outages

Improved service availability on downtime incidents
Mayor deployment of EOS 5.2

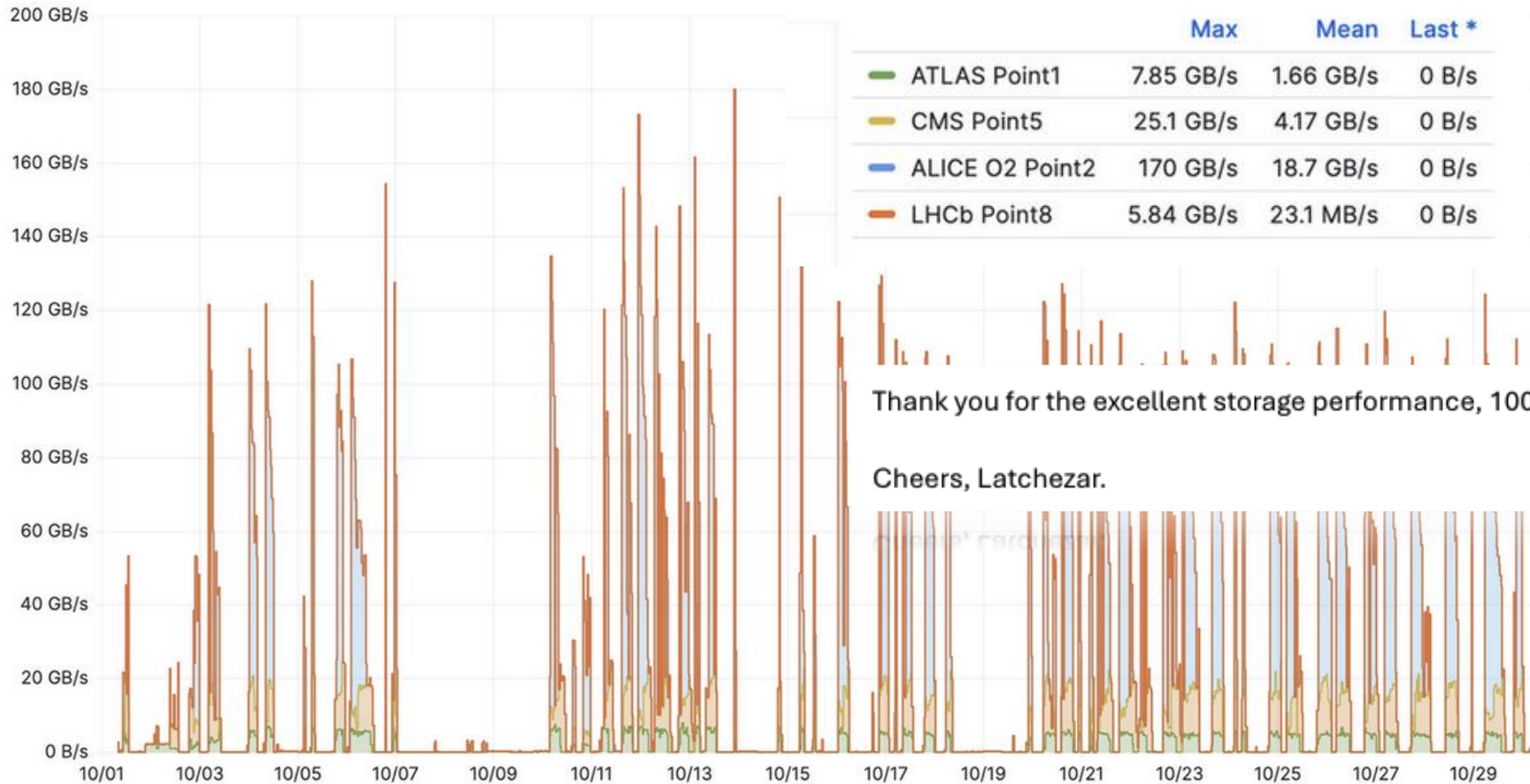
* this is summing all availabilities of all the eos physics instances, it is not a service availability

EOS for Physics: Activities – Heavy-Ion



During Heavy-Ion run (1st of October to 30th of October): ~63 PB got written for 14 million files.

Data Taking Throughput



Thank you for the excellent storage performance, 100% availability!

Cheers, Latchezar.

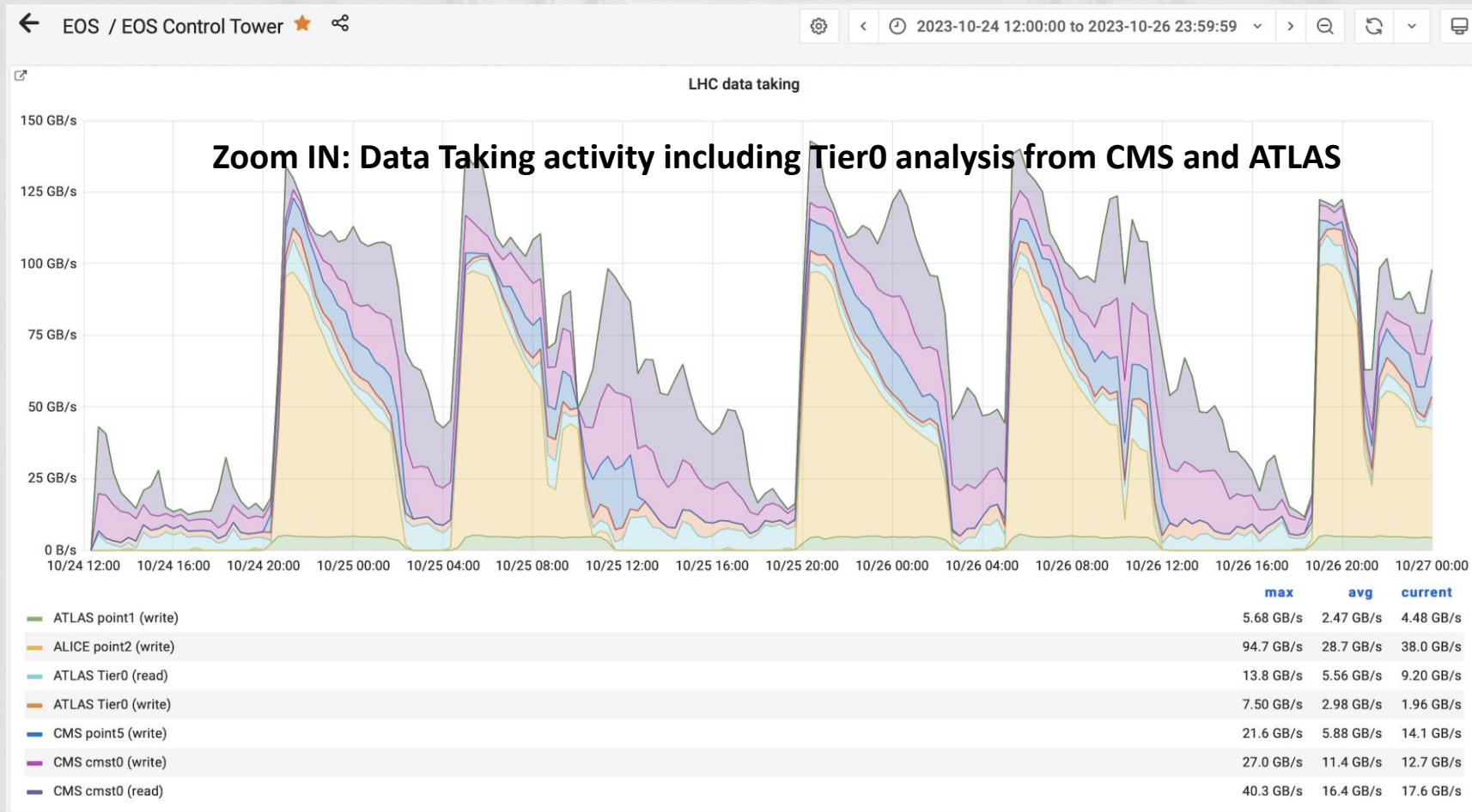


ALICE

EOS for Physics: Activities – Heavy-Ion



During Heavy-Ion run (1st of October to 30th of October): ~63 PB got written for 14 million files.



EOS for Physics: Activities – DC24



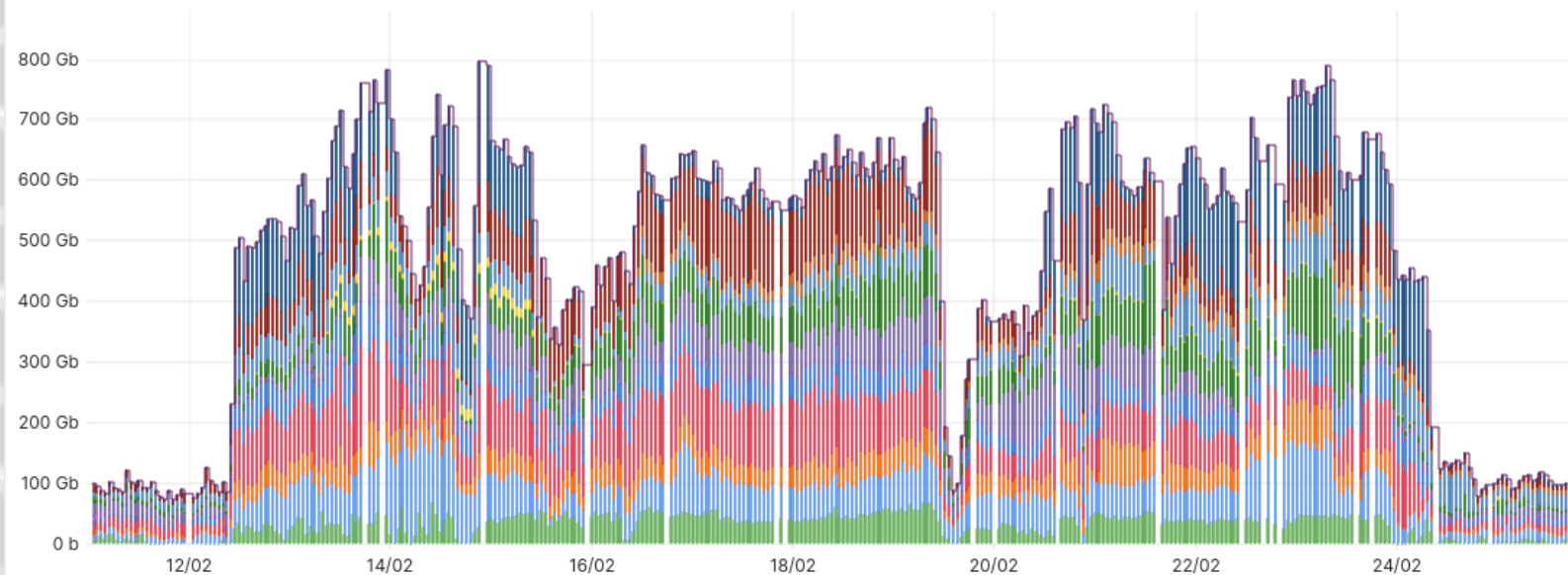
When? From 12/02/2024 to 23/02/2024

Contributions from EOS perspective:

1. Ensuring EOS stability, availability, and performance for seamless exports to Tier1
2. Deployment of HTTP tokens in all our instances
3. Deployment of scitags usage test for CMS

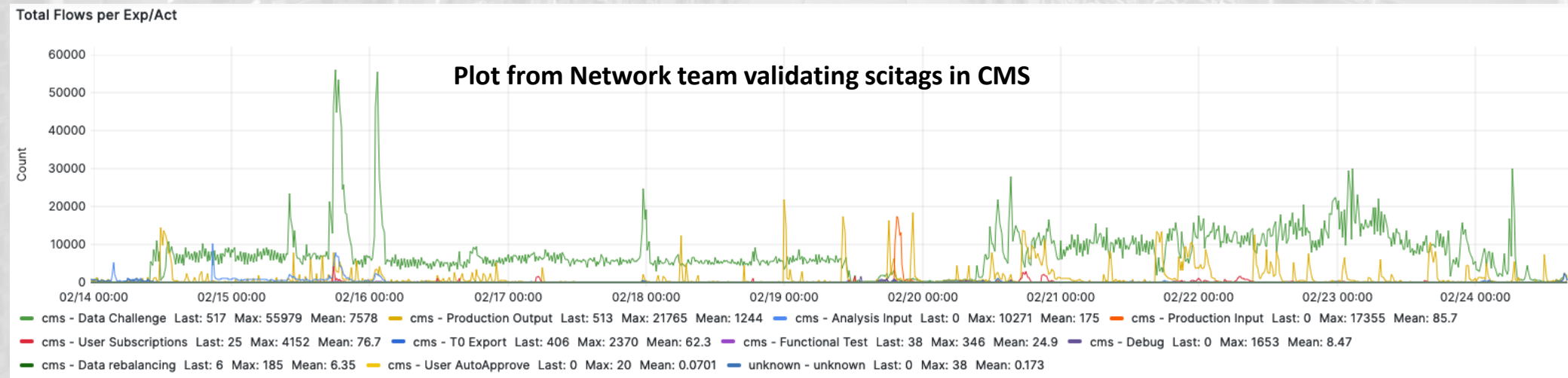
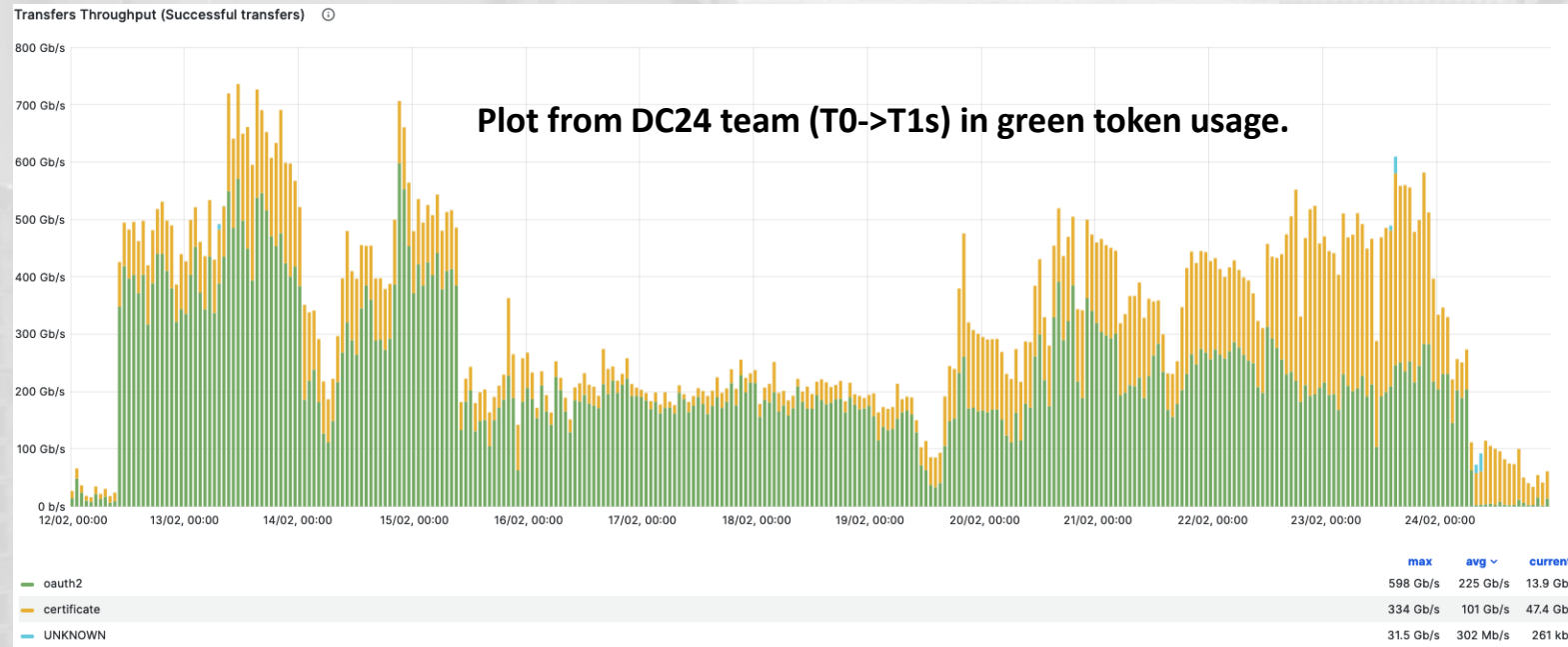
Plot from IT-CS team

LHCOPN Total Traffic (CERN → T1s)



| Name | Mean | Max |
|-----------------------|---------|---------|
| Outgoing CA-TRIUMF | 31.9 Gb | 69.6 Gb |
| Outgoing CN-IHEP | 4.15 Mb | 101 Mb |
| Outgoing DE-KIT | 55.4 Gb | 143 Gb |
| Outgoing ES-PIC | 30.7 Gb | 80.2 Gb |
| Outgoing FR-IN2P3 | 66.1 Gb | 169 Gb |
| Outgoing IT-INFN-CNAF | 34.9 Gb | 79.3 Gb |
| Outgoing KR-KISTI | 1.74 Gb | 11.4 Gb |
| Outgoing NDGF | 39.3 Gb | 110 Gb |
| Outgoing NL-T1 | 45.6 Gb | 138 Gb |
| Outgoing-PL-NCBJ | 2.63 Gb | 17.4 Gb |
| Outgoing RU-T1 | 36.0 Gb | 73.9 Gb |
| Outgoing UK-RAL | 10.8 Gb | 36.0 Gb |
| Outgoing US-BNL | 59.8 Gb | 148 Gb |
| Outgoing US-FNAL | 57.4 Gb | 204 Gb |
| Total | 472 Gb | 797 Gb |

EOS for Physics: Activities – DC24



EOS for Physics: Service Operations Highlights



Successful Storage Operations during LHC Run 3, Heavy Ion and DC24

Successful deployment of tokens for HTTP transfers

Deployment of scitags for network marking on time for DC24 in EOSCMS

Consolidation of durability for EC

Deprecation of NGINX HTTP gateways and soon GridFTP ones 😊

Ongoing migration to ALMA 9nodes in all LHC instances



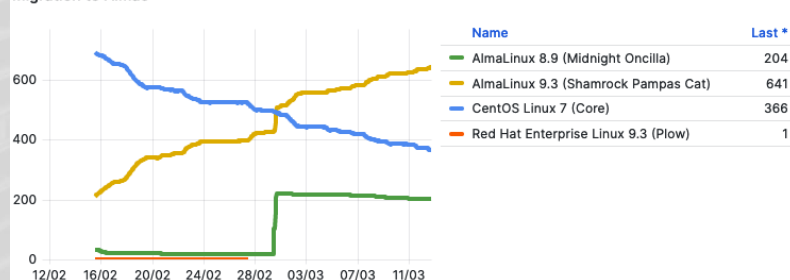
Configuration commissioning

- Adopting cerneosserver module for all (cernbox, eoscta and eosphysics)
- Simplify configuration management across storage teams

Monitoring commissioning

- Ensuring high data resolution moving to Prometheus and Thanos with S3 backend.
- Ensured the accuracy plots, quality communication and error/issue detection.
- Different monitoring views (urgent view, service evolution, reports logs)

Migration to Alma9



File descriptors



Example: Running out fds detection

Reducing xrd.timeout in xrd.cf.mgm

-xrd.timeout idle 86400
+xrd.timeout idle 120

Main Plans 2024

Focus on Quality Assurance

- ROTA implementation approach
- Enabling focus on software/operations for devops and knowledge transfer per physics workflows.

Hardware Monitoring and operations scripts commissioning

- Introduce a tool for basic emergency analysis of the instance to speed up debugging
- Improve pre-production testing with continuous load with different configurations and access patterns.

Improve Operations Efficiency

- Disk replacements, hardware types, etc.
- Simplify operations scripts management

Maintaining RUN 3 Success