

EOS for Users Site Report



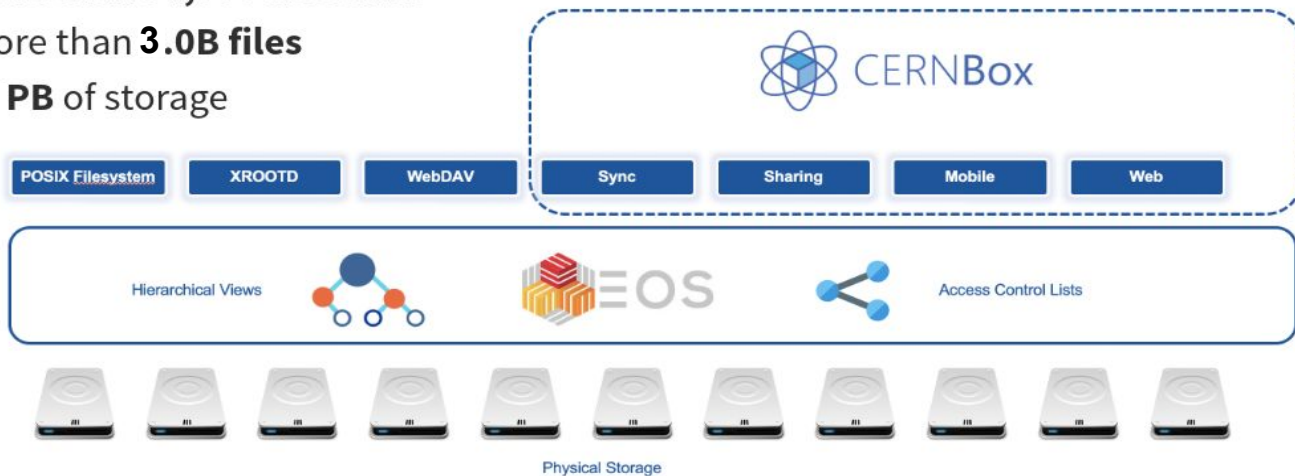
Emmanouil Bagakis

IT-SD-GSS, CERN
27 April 2023

CERNBox: Powered by EOS

CERNBox is the cloud sync and share service used at CERN.

- More than **37,000 accounts**
- More than **3.0B files**
- **20 PB** of storage



EOS for Users in Numbers

Production

EOSHOME

5 Instances
45 Disk Servers
10.9 PB Used
18.1 PB Total

EOSPROJECT

3 Instances
27 Disk Servers
5.5 PB Used
9.9 PB Total

EOSMEDIA

1 Instance
12 Disk Servers
3.3 PB Used
4.6 PB Total

QA

EOSHOME-CANARY

1 Instance
3 Disk Servers
13 TB Used
209 TB Total

Test

EOSHOME-DEV

1 Instance
3 Disk Servers
13 TB Used
550 TB Total

EOSHOME: Hardware

MGM + QuarkDB (x3)

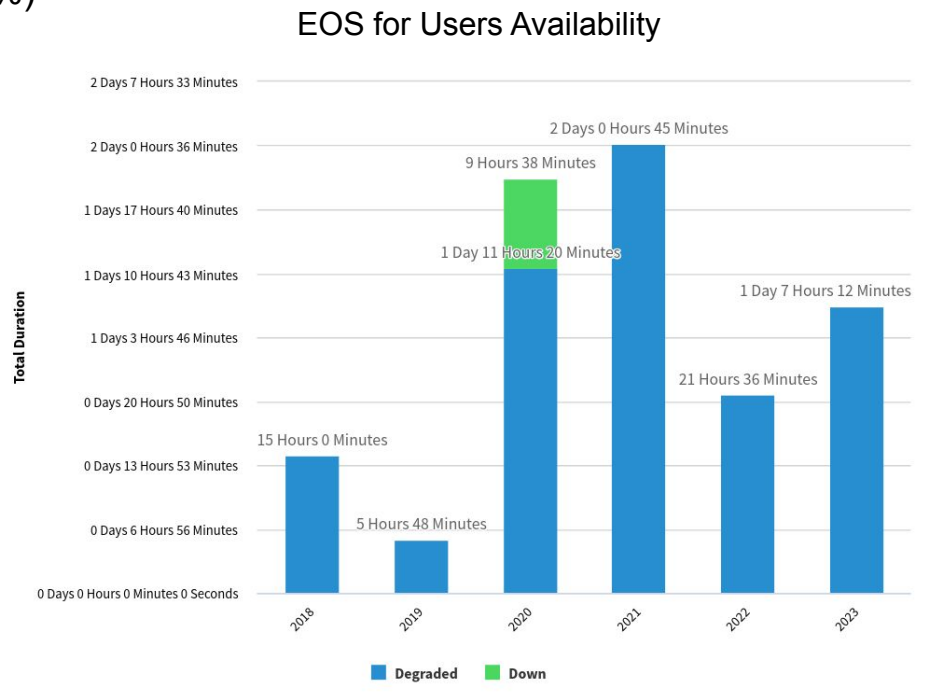
- CC7
- Intel Xeon, 32 cores.
- 192G/384G RAM
- 2 x 2T SSD (1 for QuarkDB)
- One node per switch
- 10G Network

FST

- CC7
- Intel Xeon/AMD Epyc, 20/32 cores
- 128G/256G RAM
- 48x6T/14T JBOD, 1 filesystem / disk.
XFS
- 10G/25G Network

Service Availability 2022

- **Available** (99.76%)
- **Degraded** (0.24%)
- **Down** (0.00%)



Service Growth I

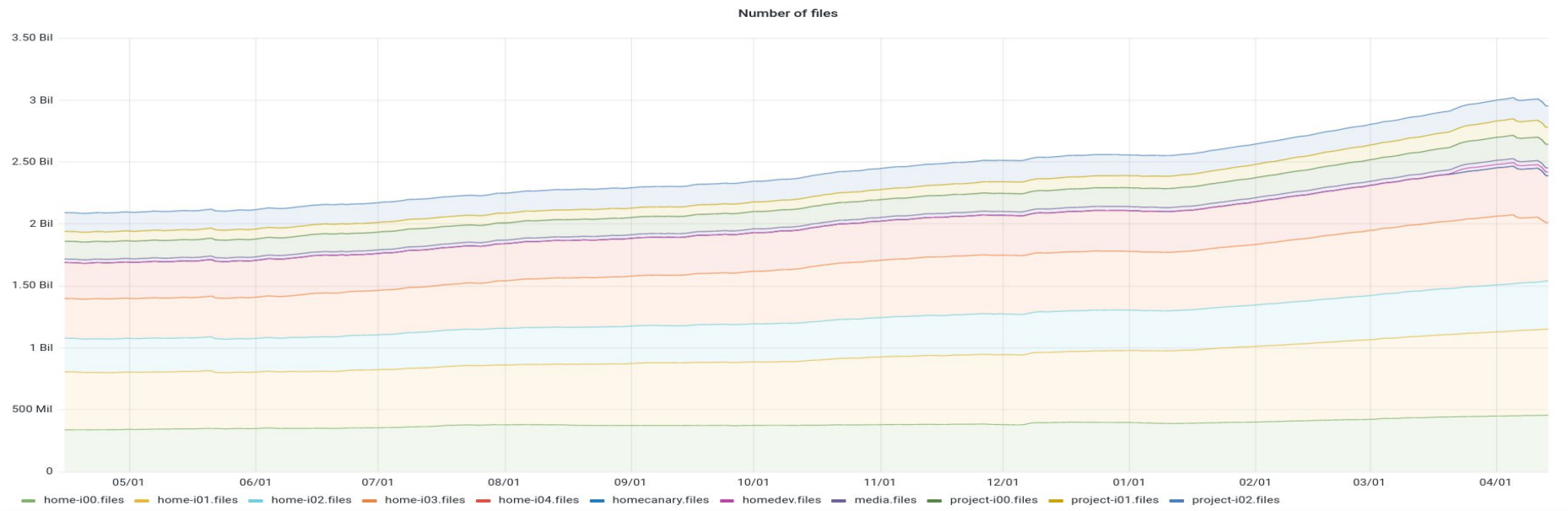
- **6PB** written since last year (+25%)

EOS Data Volume Evolution



Service Growth II

- **800M** files written since last year (+30%)



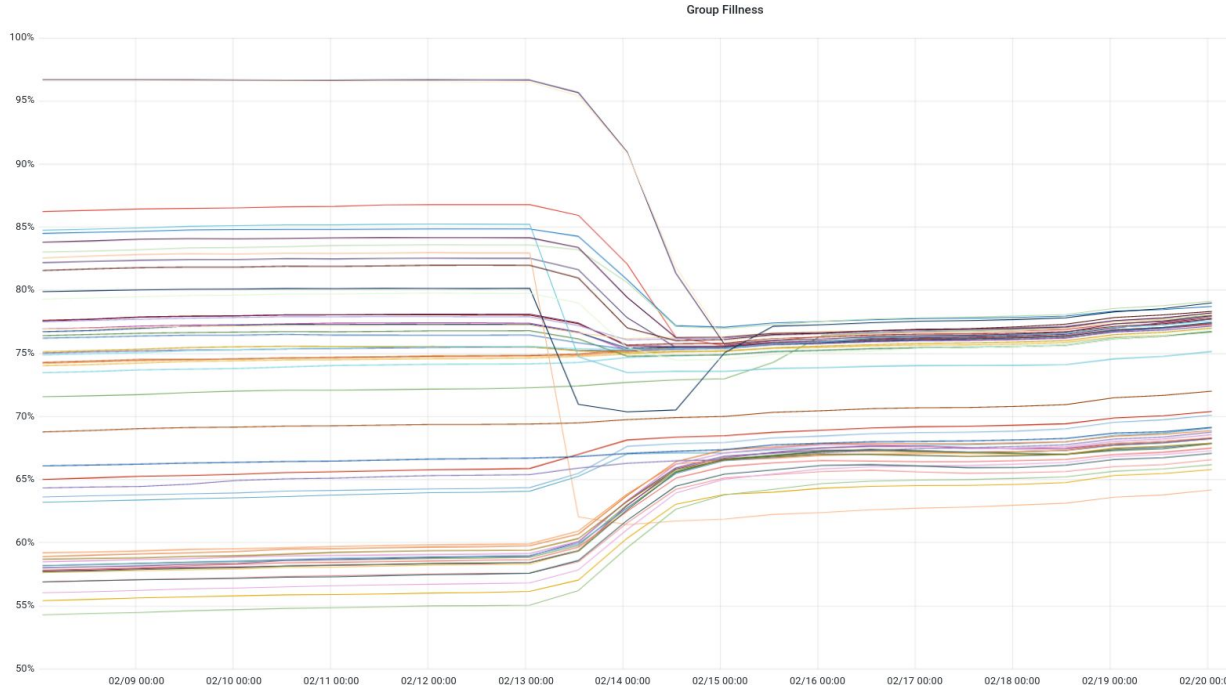
2022 Highlights I

- New **EOS stalling** system
 - Gently slow-down of user to the applied rates
 - If thread pool is exhausted, old system kicks in (Immediate rate limiting)
 - Work in progress

```
# sets the max number of threads to 4000.  
eos access set limit 4000 threads:max  
# sets the max number of threads for all users to 100.  
eos access set limit 100 threads:*  
# sets the max threads limit to 1 to user id 12345  
eos access set limit 1 threads:12345
```

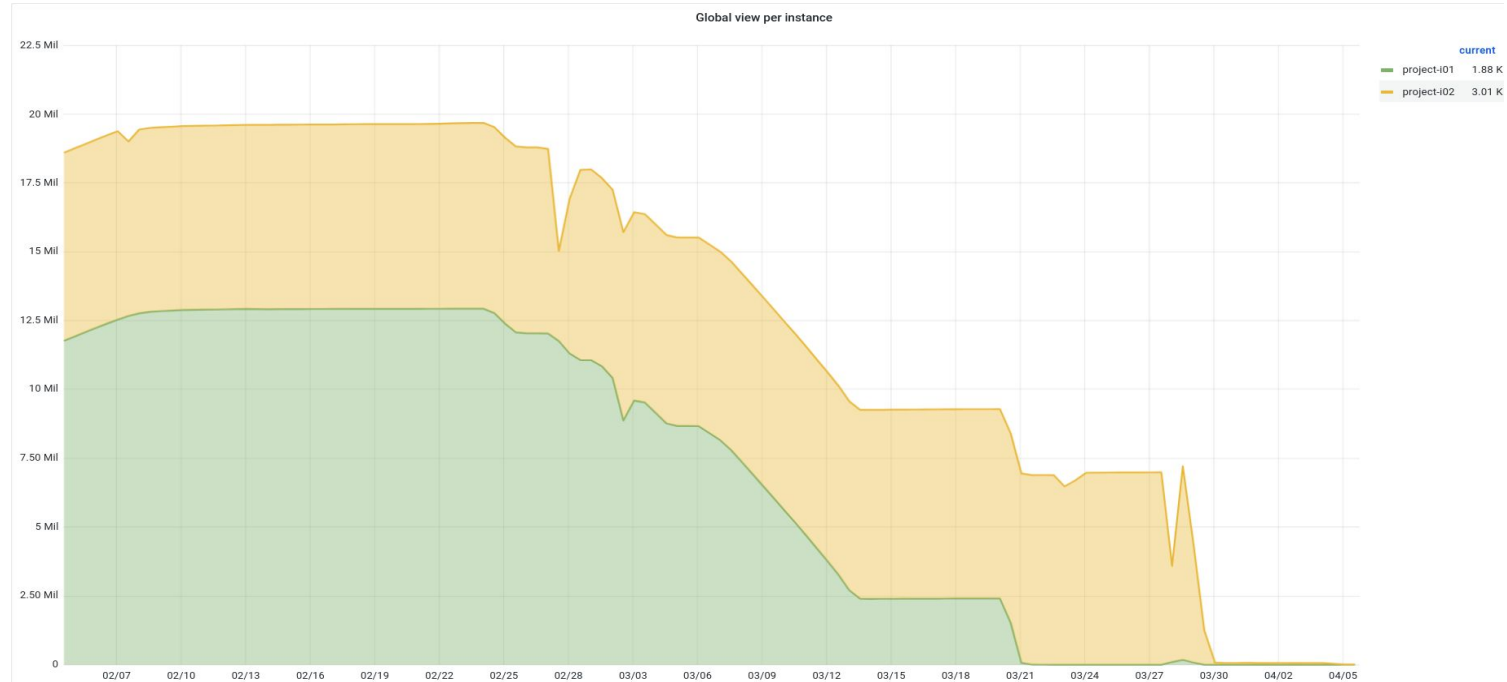
2022 Highlights II

- New **Group Balancer** in Production ([Presentation](#))



2022 Highlights III Durability

- **FSCK** tool re-enabled on EOSPROJECT ([Presentation](#))



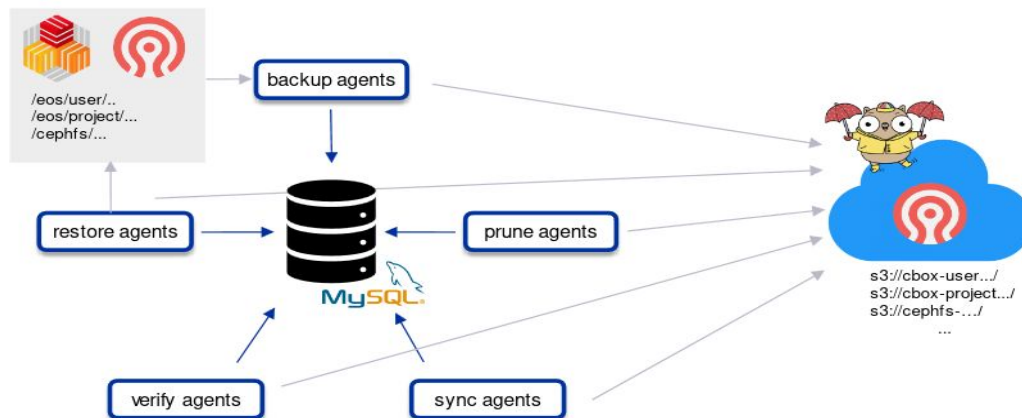
2022 Highlights IV Backup

cback backup system

- Daily snapshots for CERNBox users and projects spaces
- Stored in Ceph buckets, outside of Meyrin datacenter
- Based on restic ([Presentation](#))

cback in numbers

- ~40.000 backup jobs
- ~700M files processed every day
- 5.3 PB of backed up files



2022 Highlights v Monitoring/Logging



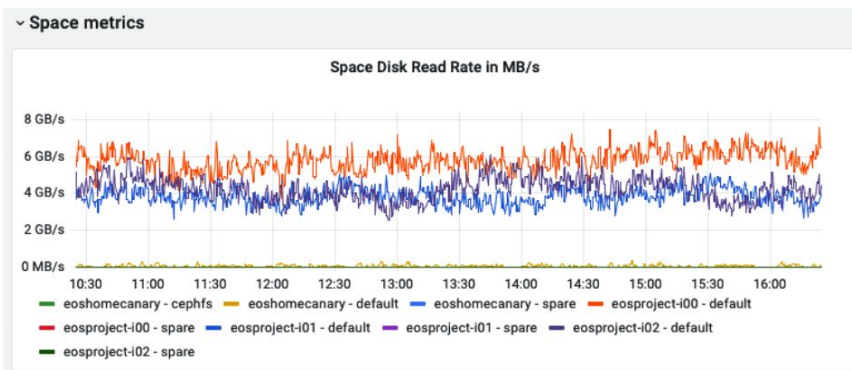
EOS Prometheus exporter

- Exposes metrics from EOS machines to prometheus server
- Filesystem, Node, Space, Group, Namespace collectors
- Visualization in Grafana
- Advanced visualization via **PromQL** ([Presentation](#))



Loki

- **Promtail** daemon collects MGM/FST and ships them to a [Loki](#) instance
- Faster investigation by searching through multiple nodes' logs in one query
- Monolithic & on k8s (scalable) architectures available



Log labels `{eosinstance="eosproject-i00", eoscomponent="fst"}`

Time ☒ Unique labels ☐ Wrap lines ☒ Dedup

Common labels: `fst eosproject-i00 st-project-rg46-47953.cern.ch eos` Limit: 2000 (64 returned) Total bytes processed: 26.5 kB

```
> 2021-02-28 19:07:38 210228 19:07:38 time=1614535658.496987 func=Report level=INFO logid=static.....
86host=st-project-rg46-47953.cern.ch&fid=318181276&fxid=12f70f9c6&fsid=278&del_ts=1614535657&del_tns=879947909&dc_ts=1614534
> 2021-02-28 19:07:38 210228 19:07:37 time=1614535657.879678 func=_rem level=INFO logid=19d96cca-7104-11eb-9cbd-2c600c20da98
> 2021-02-28 18:57:36 210228 18:57:36 time=1614535056.362519 func=MgmSyncer level=INFO logid=static.....
> 2021-02-28 18:56:37 210228 18:56:37 time=1614534997.442512 func=Report level=INFO logid=static.....
0&path=/ :xt6fstpath=/data34/00007c4a/12f70f9c6&guid=
.00&rv_op=0&rvb_min=0&rvb_max=0&rvb_sum=0&rvb_sigma=0.00&rs_op=0&rsb_min=0&rsb_max=0&rsb_sum=0&rsb_sigma=0.00&rc_min=0&rc_m
&csz=4240909&delete_on_close=0&sec.prot=https&sec.name=rpcgif&sec.host=cbox-lbproxy-05.cern.ch&sec.vorg=6sec.grps=6sec.ro
> 2021-02-28 18:56:37 210228 18:56:37 time=1614534997.073479 func=rmnleteHandler level=INFO logid=static.....
```

2022 Highlights VI Security

File Permissions Consistency

- Frequent namespace checks
- Ensure files within private directories have correct permissions (e.g. not public)
- Mostly “ancient” cases, no new cases enabled by current infrastructure



eos-sec-logs

- Sending eos clients' actions logs to security
- Read, write, rename, deletion operations
- Assist malware/anomaly detection



2022 Highlights VII Evaluation

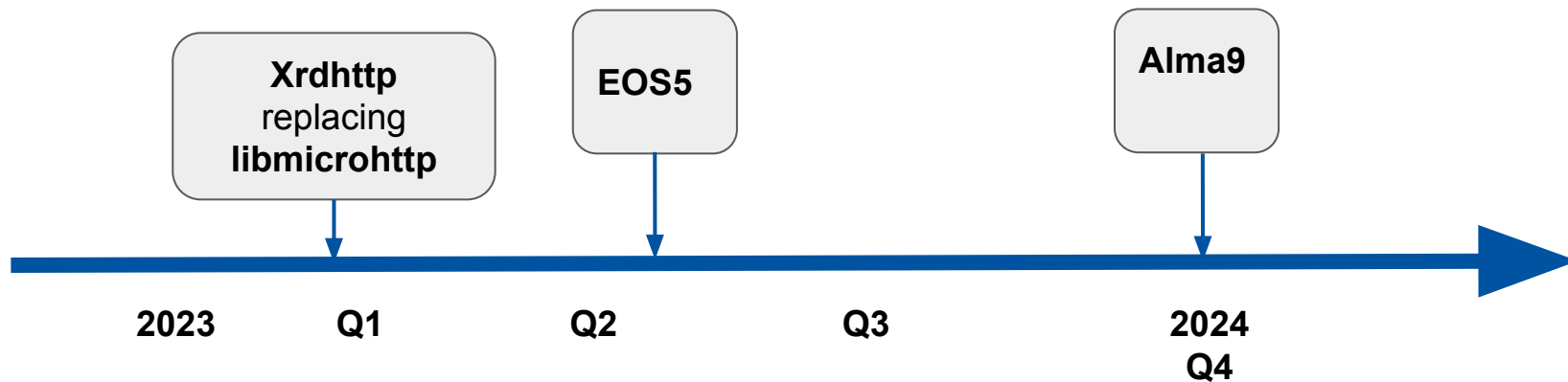
- Evaluation: **EOS** on **CephFS** (Halted)

Replace **FST-local HDD** with a **CephFS** mount

- Initially expected benefits were not conclusive to justify the project
 - Performance concerns:
 - Latency spikes still present in our testing
 - a. Note: Improvements on EOS produce a more stable latency (e.g. not using levelDB)
 - Architecture concerns:
 - Not ideal: Durability handled on two levels. EOS complexity isn't reduced.
 - Risk concerns:
 - Namespace vs disk metadata discrepancy could be a potential durability issue
 - Data availability concerns (FST becomes a single point of failure)
 - Ceph dependency
 - Cost concerns:
 - Majority of files are very small (< 100KB). Erasure coding becomes expensive



2023 EOS for Users: Roadmap



Thank you!
Any questions?

Special thanks to:

- *Roberto Valverde Cameselle*
- *Jan Iven*
- *Hugo Gonzalez Labrador*

