

# EOS Workshop April 2023

Notes for the ALICE USA T2 Meeting

Steve Moulton

May 15, 2023

ORNL is managed by UT-Battelle LLC  
for the US Department of Energy



U.S. DEPARTMENT OF  
**ENERGY**

# Highlights

- EOS5 Deployments are in progress
  - Not much penetration at T2 sites yet
  - 5.1.10.19 considered stable release
  - EOS 4 EOL (end of support) end of 2023
- All authentication/access control to be done via tokens (Macaroons) at the xrootd level. Tokens to expire after much shorter period. Certificates will remain in use for SSL only.
- QuarkDB now built on top of rocksdb (a Facebook project)

# Tokens

- Based on Macaroons
  - OAuth 2.0 technology
  - Distinct authentication and authorization tokens
  - Replaces Kerberos & Certificates
  - Will still need certificates to service SSL

# Monitoring

- EOS now has a Prometheus exporter
  - Exposes 144 EOS metrics
  - Alerting via Grafana
  - Parses `eos group ls -m` into Prometheus format
  - `./eos_exporter -eos -instance=<instance name>`
  - Speaker has Grafana dashboards he is willing to share
- EOS File Transfer Monitoring
  - `yum install eos-file-transfer-monitoring`
  - Builds data via periodic queries

# EOS 5 Client Rollout

- More focus on fuse
  - Significant performance improvements
- Uses Xrootd 5

# EOS other

- New stalling system
  - Gently slow down users to applied rates
  - If thread pool exhausted, old system (immediate rate limiting) kicks in.
- New group load balancer
- FSCK reenabled
- Xrdhttp replacing libmicrohttp (relief from vulnerabilities)

# Andreas' Presentation

- <https://indico.cern.ch/event/1227241/contributions/5332261/attachments/2636902/4562214/EOS%20Roadmap%20.pdf>

# EOS 5.2 Release



- **What we want to drop ...**

- libmicrohttpd
- LevelDB file meta data
- Old FSCK reporting
- Old Balancer
- Transfer Queues/Multiplexer
- MQ Daemon
- Internal HTTP browser JS

- **What we will target ...**

- HA without MQ
- MGM Latency Reduction ( replacing global mutexes with local mutexes )
- FLAT Scheduler
- REST API MGM ( summer project )
- Merge SHARE API and permission system homogenisation
- Possibly move to TPC processes instead of in-MGM multithreading with XrdCl
- FUSE Performance
- FST Gateway IO & scheduling for shared backends
- EC Updates (when XRootD range-clone/copy-on-write functionality is available)



# Xrootd 5.5.4 soon to be 5.6 then 6.0

- Code base moving from C++17 to C++20 in 2023
  - Big issue – namespace changes
  - Build changed to cmake
  - Uses GoogleTest for regression testing
- Xrootd team has 21 members
  - Of course some where other hats
- Improved and new plugins
  - Plugins used to provide functionality previously hand tooled.
- Very active development



# Cern Tape Archive

- An entire day was dedicated to CTA. I attended first intro session and determined that it had little relevance to us, though the KISTI changes are interesting and may be useful to other projects supported at ORNL

# Notes and Exceptions

- QuarkDB now frequently referenced at QDB
- RAIN is now also defined as RAID on two nodes as opposed to Erasure Encoding
  - As usual decision to do EC or RAIN is left to the sites
- Expected FST throughput 3 Gb/sec w/EC
  - 6Gb/sec w/out EC)
    - ORNL FSTs 5-8 seeing on the order of 2Gb/sec max throughput. Probably due to application mix.
- Much user commentary (!) on documentation quality particularly wrt QDB.
- CERN has settled on Alma Linux 8 as the base distro