EOS workshop 2020

3rd to 5th of february @ CERN

platform for exchange between developers, users, sites and people interested in storage technology

disk - tape - cloud - sync & share - devops

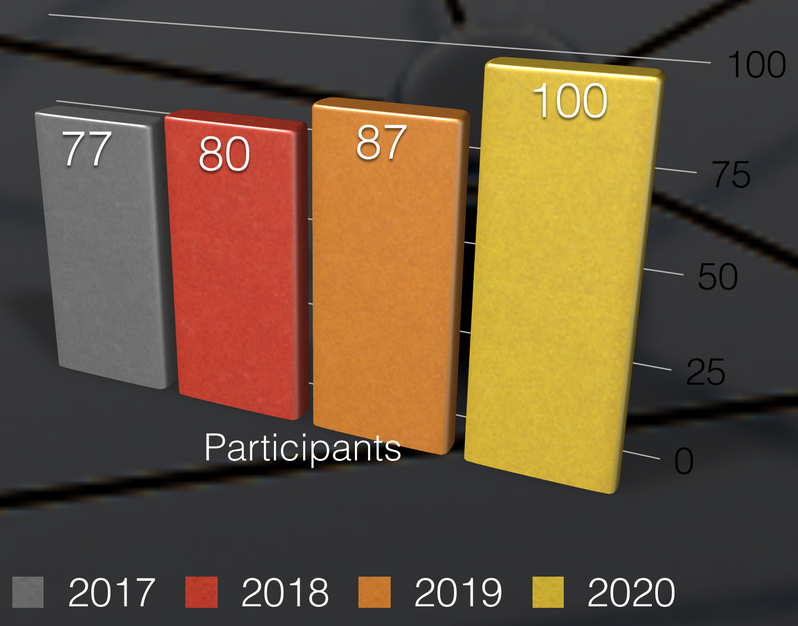
Welcome to the 4th **EOS** workshop

at CERN 3.-5.2. 2020

EOS WORKSHOP

Andreas-Joachim Peters
CERN IT Storage Group

Who is registered …



Our sympathies go to our chinese colleagues from IHEP which can only participate remotely to the workshop!

Main questions to ask

& hopefully answer

- **how to run EOS today**, how might it look in the future, how can it be used in the future, what is the **direction** of the project?
- **evolution** / (non-) achievements since last workshop
- what is provided, how do you use it, **what do you actually want/need**?
- support for **DAQ**, **Tier 1/2 & Tape storage**

Organisational Information

WORKSHOP ROOMS

Monday : 10am-5pm IT Auditorium (here) , **Bld 31** 3rd floor
Tuesday : 09am-1pm **Bld 31-S-028** available as work room for external participants
: 2pm - 5pm Tier1/2 & Online/Offline Session in **Bld 513-1-024**
Wednesday
: 10am - 4pm **Bld 513-1-024**

Lunch : 12:15 pm - 2pm Restaurant 2

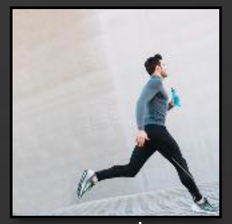
Organisational Information

Workshop Dinner Monday 7pm

Route de Meyrin 286, 1217 Meyrin

Workshop Dinner

How to get there …



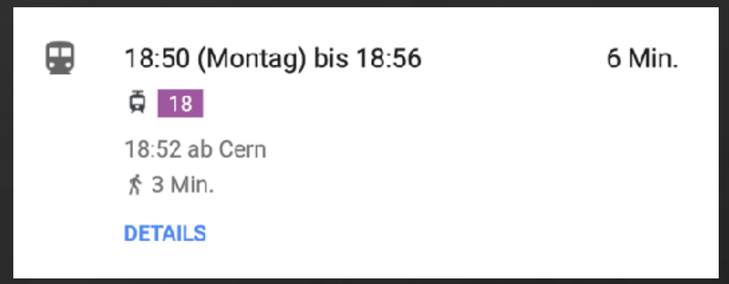
sportive

Workshop Dinner

How to get there …

relaxed

Tram 18 - 18:50 CERN



AdBlock

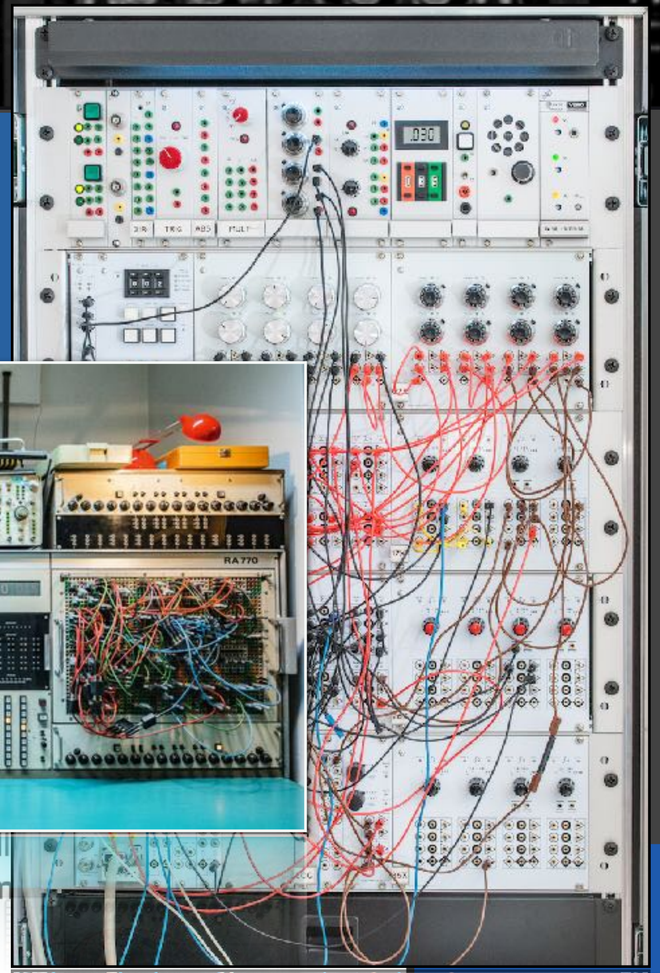
before we get going …

CERN Seminar

Friday 14.2. 11-12am IT-Auditorium

https://indico.cern.ch/event/881532/

Guest Speaker **Prof. Dr. Bernd Ulmann**



CERN Computing Seminar

Analog Computing - past, present, future?

by Prof. Bernd Ulmann

Friday 14 Feb 2020, 11:00 → 12:00 Europe/Zurich

31/3-004 - IT Amphitheatre (CERN)

Description

As classic stored-program digital computers are reaching physical and practica
suffer from problems like Amdahl's law, unconventional approaches to high-spe
more computing power at lower power consumption.One of these approaches
electronic models of the underlying mathematical equations. Largely forgotten
form of fully reconfigurable integrated circuits. Coupling these with traditional c
combine the best of two worlds, the programmability and vast program libraries available for digital computers as well
and low power consumption of analog computers. This talk briefly covers the history of analog computing, gives exam
applications and future developments.

About the speaker

Acknowledgements

- A tribute to the **INDICO team** for making organisation of such an event that simple with CERN INDICO software
- A special thanks to
  - the **IT department** and the **IT-ST** group management for the possibility to host this workshop
    - allows us to offer this workshop without any fee two coffee breaks
  - **EVERYONE** helping to make this workshop a success - **participation, presentations, discussions & organisation**!

Project History

timeline

2009

CASTOR
XRootD
lustre
XRootD

castor-xrootd IF

lustre-xrootd IF

2010

architecture document

EOS Open Source Storage

2011

EOSATLAS
EOSCMS
EOSALICE

IPv4

v 0.2 AMBER

2012

EOSLHCB

v 0.3 BERYL

2013

EOSPUBLIC

2014

v 0.3.x AQUA MARINE

SEAGATE KINETIC

2015

EOSUSER

2016

2017

EOSMEDIA

IPv6

v 4.0 CITRINE

eXtreme DataCloud

2018

EOSALICEDAQ
EOSCTAATLASPPS

v 4.3 Quarkdb

RocksDB

EOSHOME

EOSPROJECT

2019

CERNBox

swan

wopi

Office

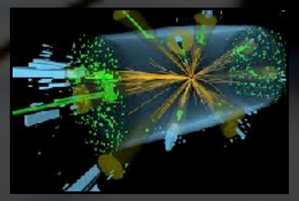
tape

EOS O2

2020

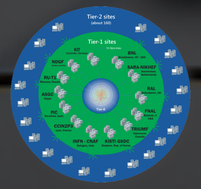
10 years

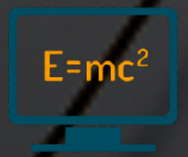
v5.0
2020



What is EOS used for …

- **disk storage**
  - raw data
  - analysis data
  - cernbox home & project spaces
  - cloudstore AARNet, Joint Research Centre JRC
  - Tier 2 & universities
  - online systems
- **tape storage cache**
  - Cern Tape Archive





Development Work Areas

- **namespace architecture** (MGM)
- **storage consistency** (FST)
- **filesystem access** (eosxd/ACLs)
- **tape integration** (CTA)
- **protocols/API** (ProtBuf,XrdHttp,GRPC)
- **tokens & authorisation**

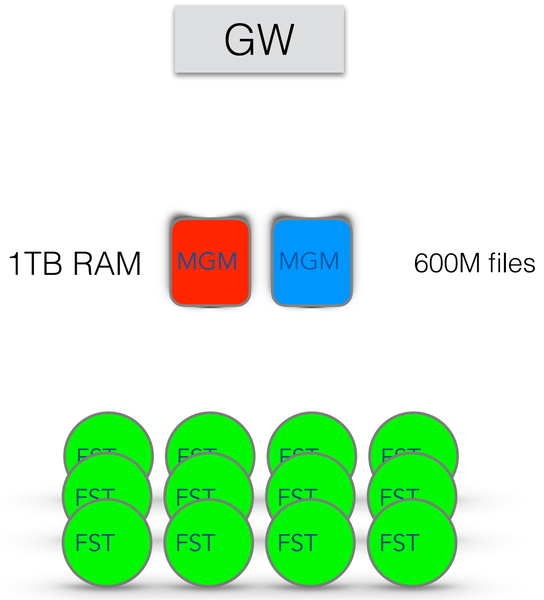
Architectural Evolution

EOS 2017		EOS 2019		
Master-Slave Architecture	→	**Active-Passive** Architecture	+	Service **Sharding**
stateful meta-data service		almost stateless meta-data service		scale-out meta-data performance

Architectural Evolution

CERNBOX 2017

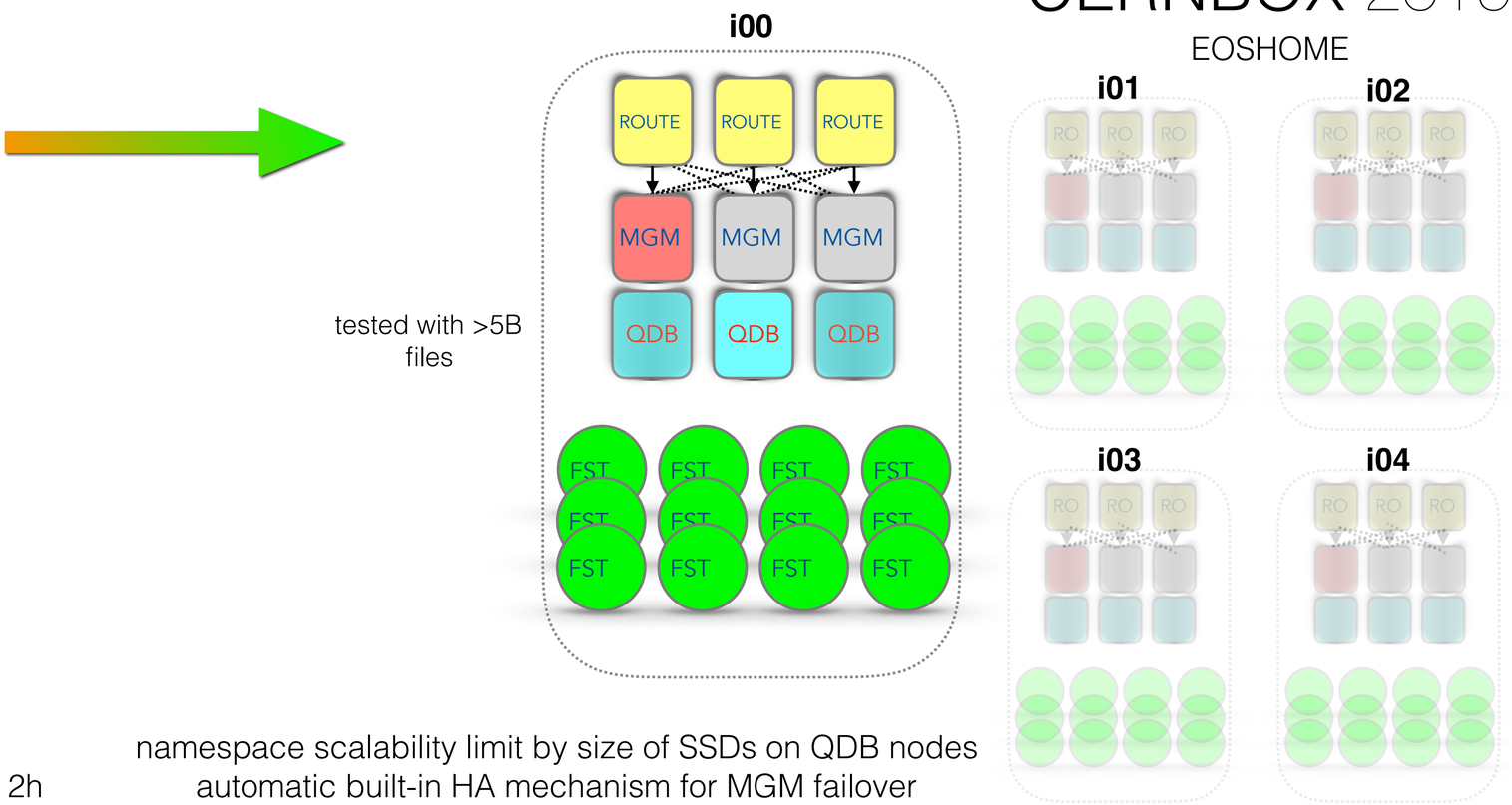
EOSUSER



at namespace scalability limit
availability constrained by infrequent long boot time of 2h

CERNBOX 2019

EOSHOME



namespace scalability limit by size of SSDs on QDB nodes
automatic built-in HA mechanism for MGM failover

QuarkDB

https://github.com/gbitzes/QuarkDB

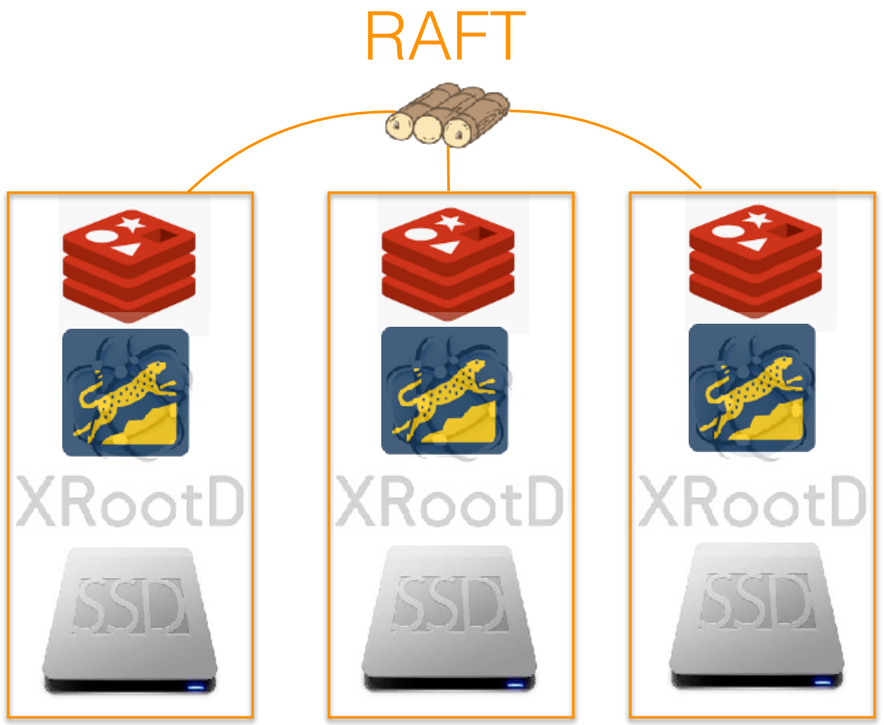
- **Introduction of QuarkDB as persistent KV store for namespace meta-data**
  - based on **REDIS** protocol, **RocksDB** & **RAFT** consensus algorithm
  - high-**available**, high-**performant**, **scalable**, low-**latency**
  - **extremely positive** production **experience**

C++ client library
https://gitlab.cern.ch/eos/qclient

QDB performance example: retrieve KV@200kHz

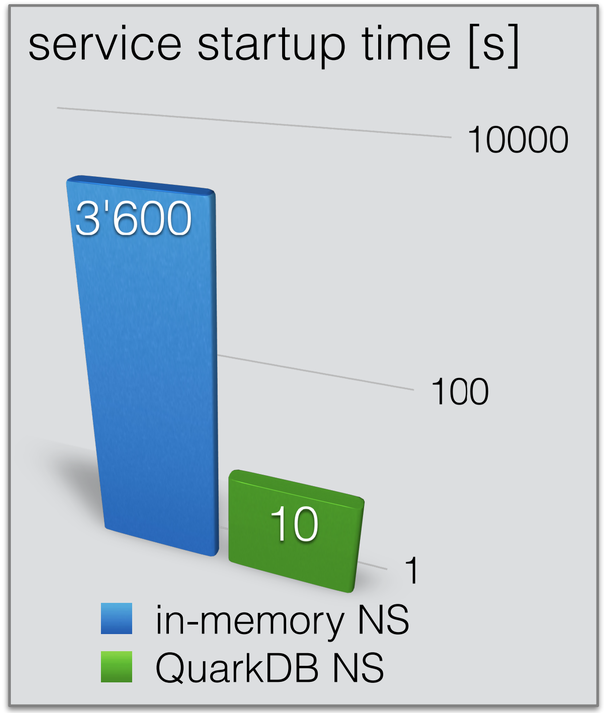
QDB api
- kv
- sets
- hashes
- pub-sub
- lease

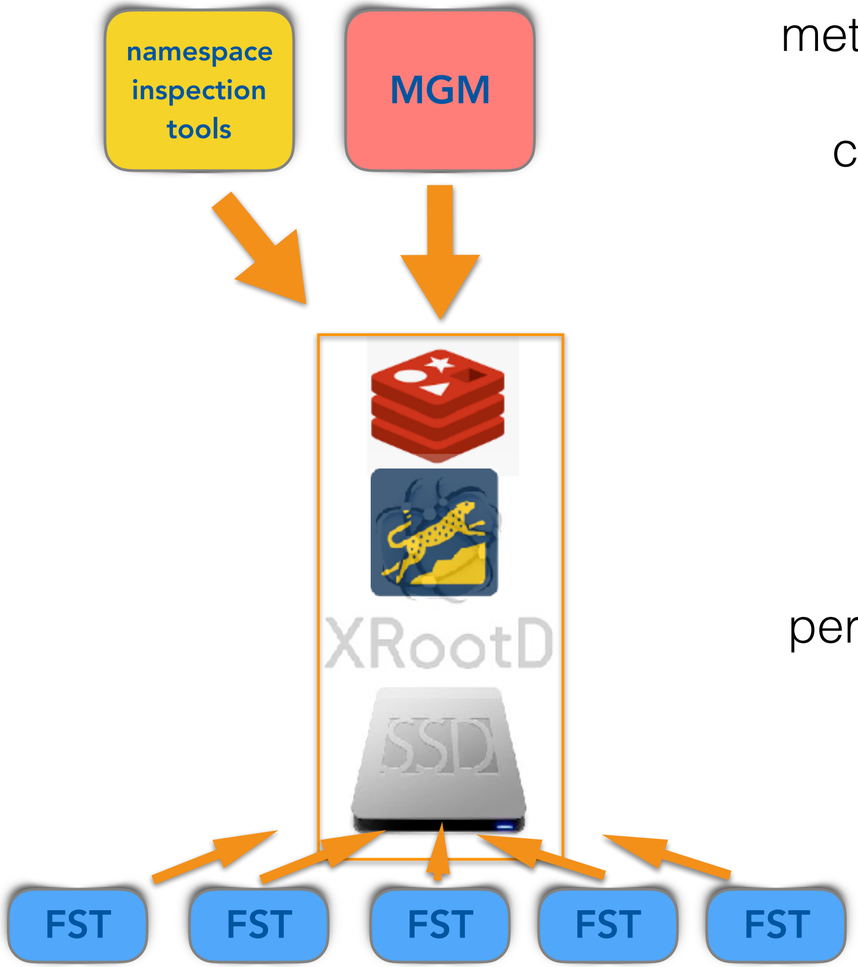
QuarkDB



QuarkDB Namespace

- service **startup time** was major **source** for service **downtime** for in-memory namespace

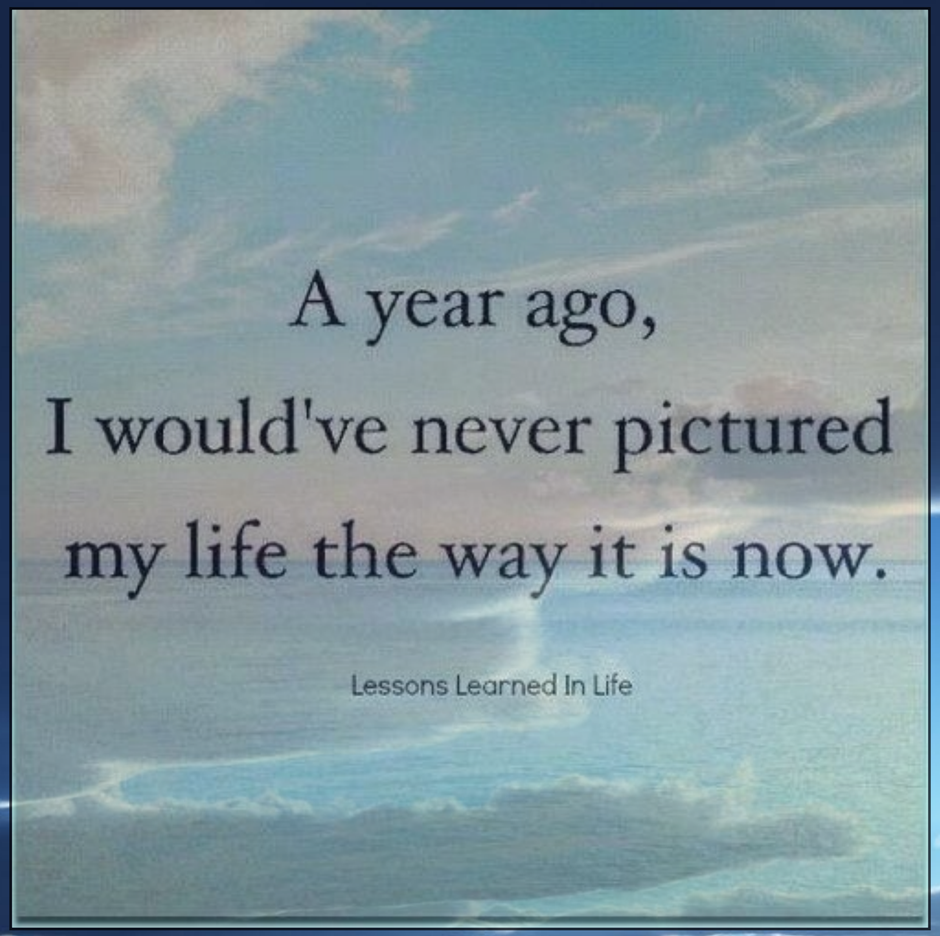




meta-data service **MGM** stateless with configurable cache

meta-data persistency with **QuarkDB**

storage server **FSTs**



Fictional quote from an EOS+QuarkDB operations person doing a restart …

2019++

Tape Integration

EOS + Tape = EOSCTA

integrated support for tape into EOS file on tape=offline replica

- loose service coupling between EOS and CTA via protocol buffer interface & notification events - everything is synchronous
- no SRM, using XRootD protocol only - integrated with FTS

high disk capacity

EOSATLAS

TPC

low disk capacity

EOSATLASCTA

short file lifetime

Cern Tape Archive

Operation Model

Protocol Support

GRPC support with token and x509 support

- mapping applications identity using GRPC token=>(uid,gid) or DN=>(uid,gid) mapping

Namespace interface

- metadata injection - used for Castor=>CTA meta-data migration
- mkdir|rmdir|touch|rm|unlink|ls|find|rename|symlink|setxattr|chown|chmod|acl|token|create-version|list-version|purge-version with streaming support for large responses

HTTP(S) support with token and x509 support

- using XrdHttp and external handler

HTTP TPC / XRootD with delegation support & WLCG tokens

- using default proxy server in front of EOS instances on gateway machines
- using SciTokens library

S3 support with MINIO gateway

- via plug-in for MINIO developed by AARNet - *currently not deployed at CERN*

General Directions for 2020

consolidation of new architecture, **improvement** of reliability & consistency and **optimisation** of internal storage services to profit from QuarkDB

support **HTTP** eco-system: provide **GRPC** as MD API, HTTPS/**DAV** as Data API for front-end CERNBOX

establish/support **tokens** for applications and GRID access

focus on **erasure coding**

- **pre-defined conversion policies** for files from/to EC layouts
- **light-weight object storage** for sequential access & archiving use-cases - client-driven e.g. native XrdCl support

EOS

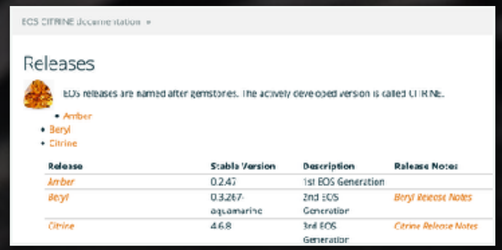
Web Page https://eos.cern.ch

GIT Repository https://gitlab.cern.ch/dss/eos

Community Forum https://eos-community.web.cern.ch/

email: eos-community@cern.ch

Documentation http://eos-docs.web.cern.ch/eos-docs/



Support email: eos-support@cern.ch

Enjoy the workshop!