News on XRootD & EOS

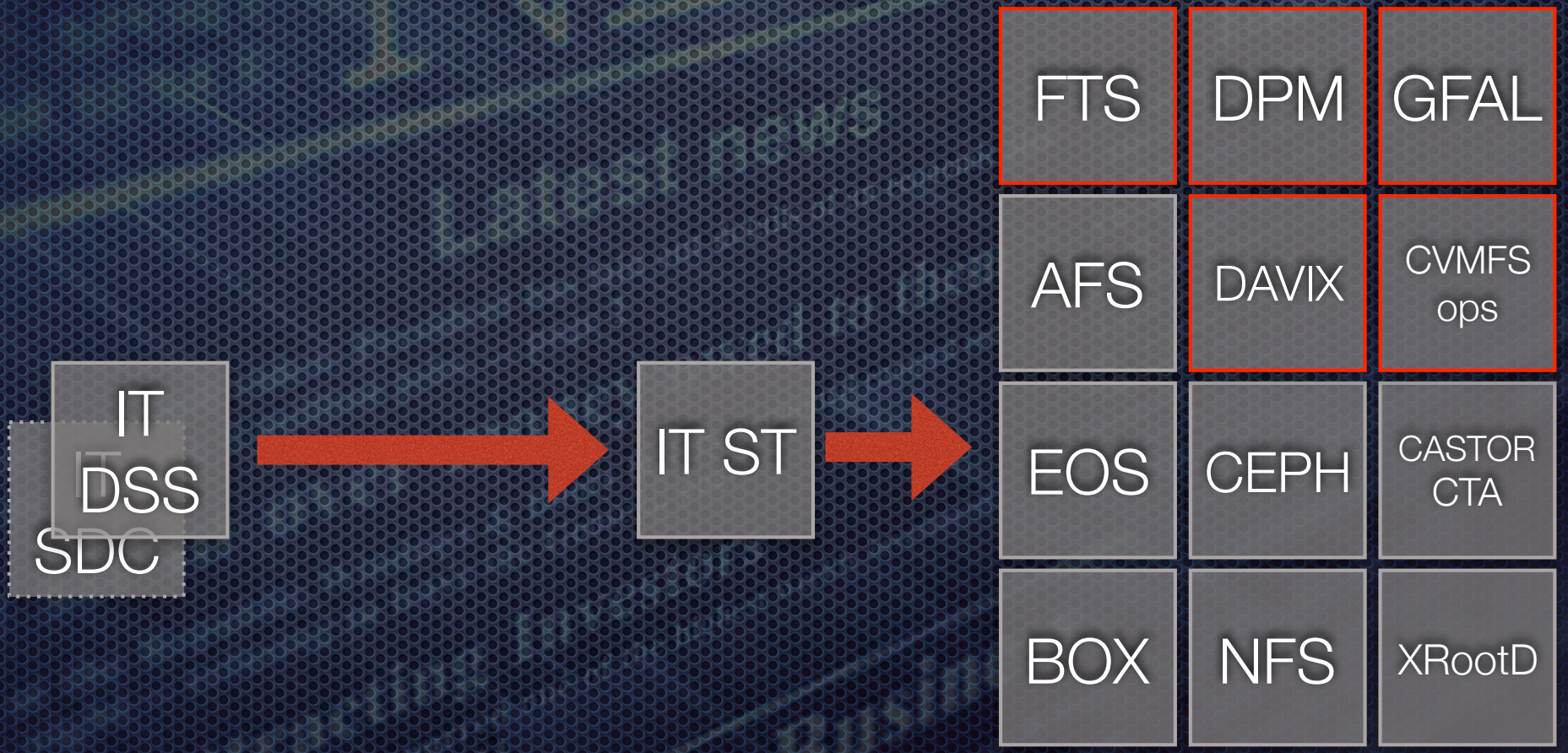
cern

Andreas-Joachim Peters for the IT-ST group / EOS project

17.04.2016

New Group Structure in CERN-IT

Data & Storage Services



NEWS
Latest news
Attracting Investors
Business
XRootD

XRootD

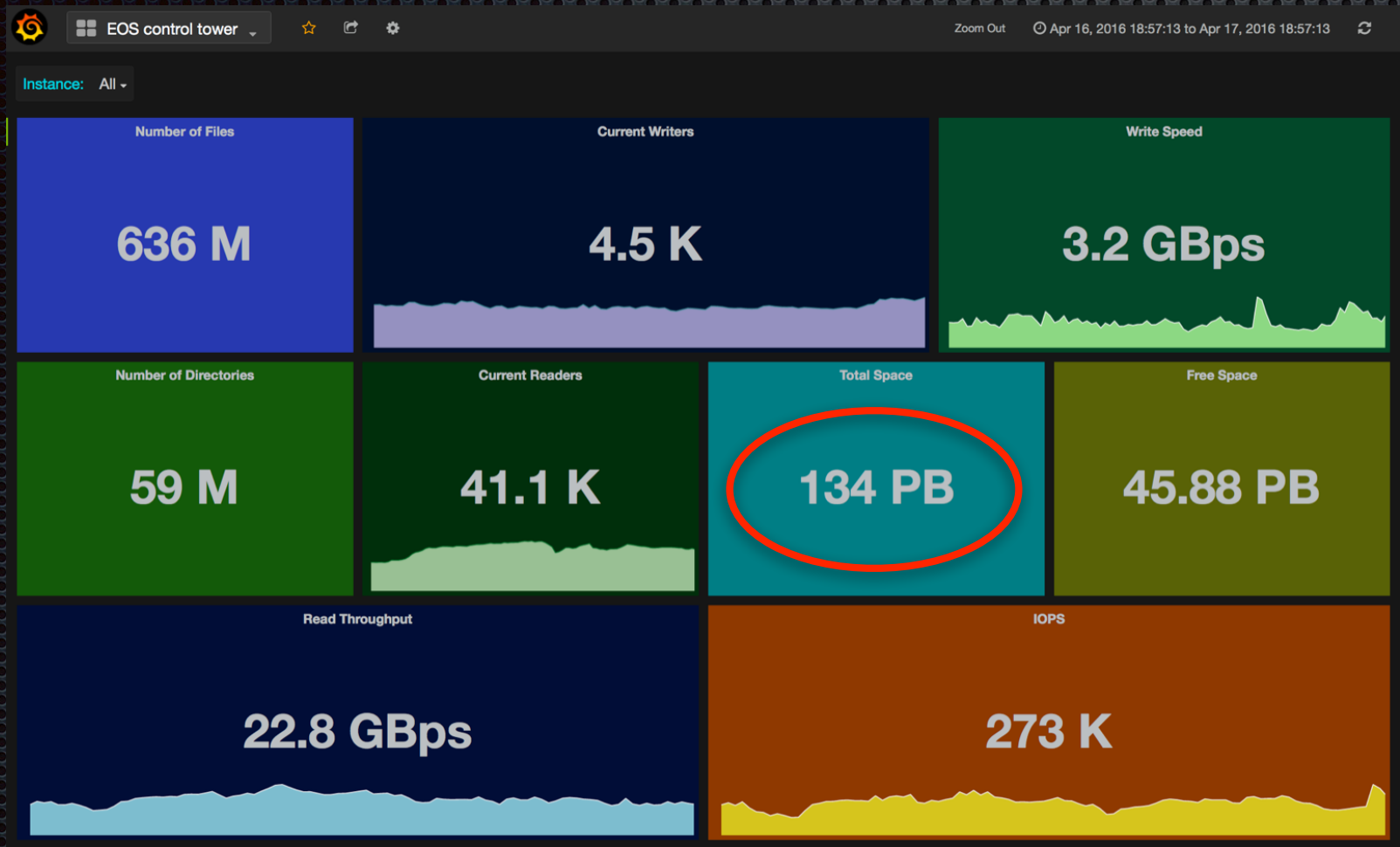
- **Latest Release v4.3.0 - highlights**
  - large IO buffer support
  - various bug fixes, file cache improvements
- **Development**
  - file meta-link support
  - multiple event loops for proxy server/proxy enhancements
  - segmented file transfer support
  - Workshop Tokyo 8.-10. November https://www.icepp.s.u-tokyo.ac.jp/index-e.html

XRootD

- **RADOS plug-in** developed at CERN by S. Ponce and integrated in CASTOR 2.1.16
  - evaluated for production early summer
  - continuous test running

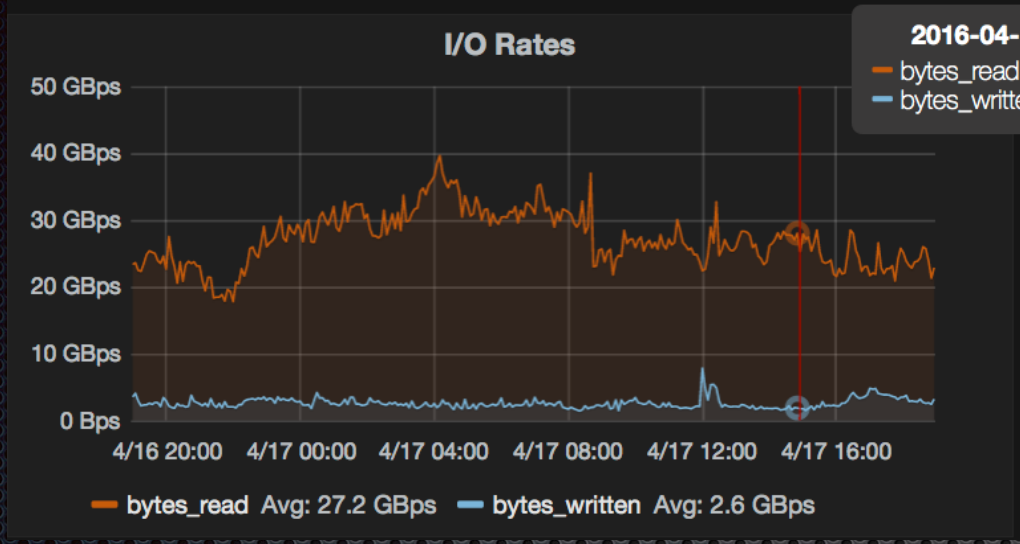
NEWS
Latest news
Economic Plan
Attracting Investors
EOS

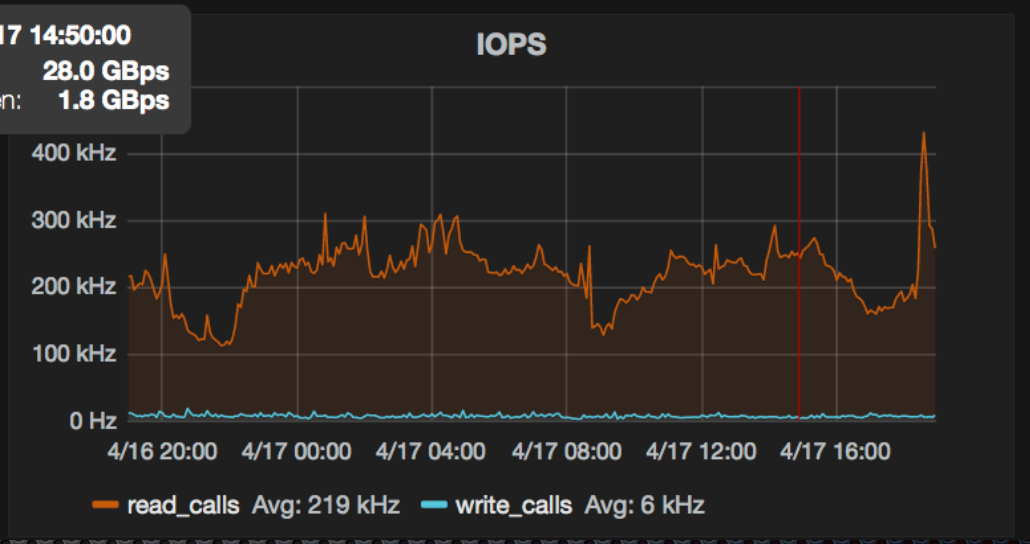
EOS @CERN in numbers

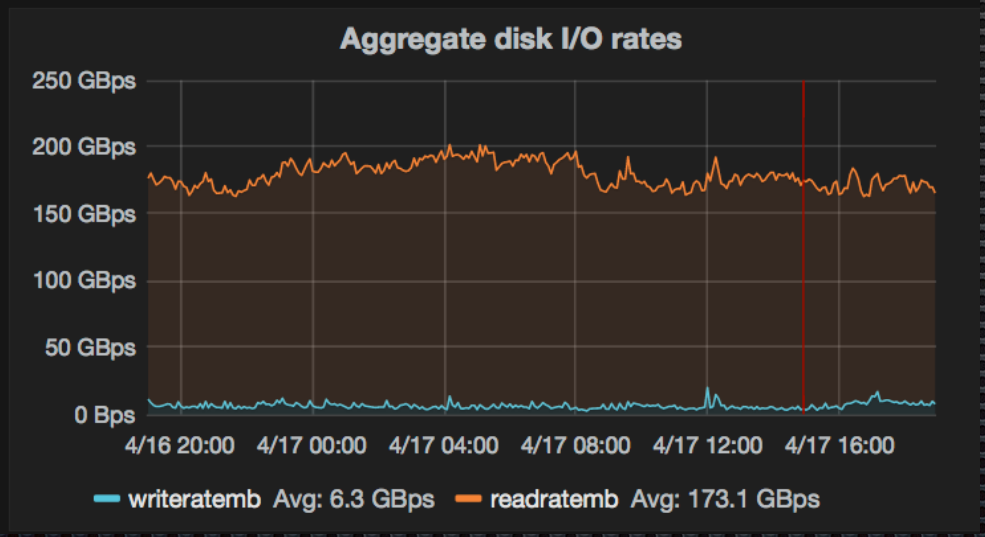


- **6 production** instances at CERN
- **3 pre-production** instances at CERN

EOS @CERN in numbers



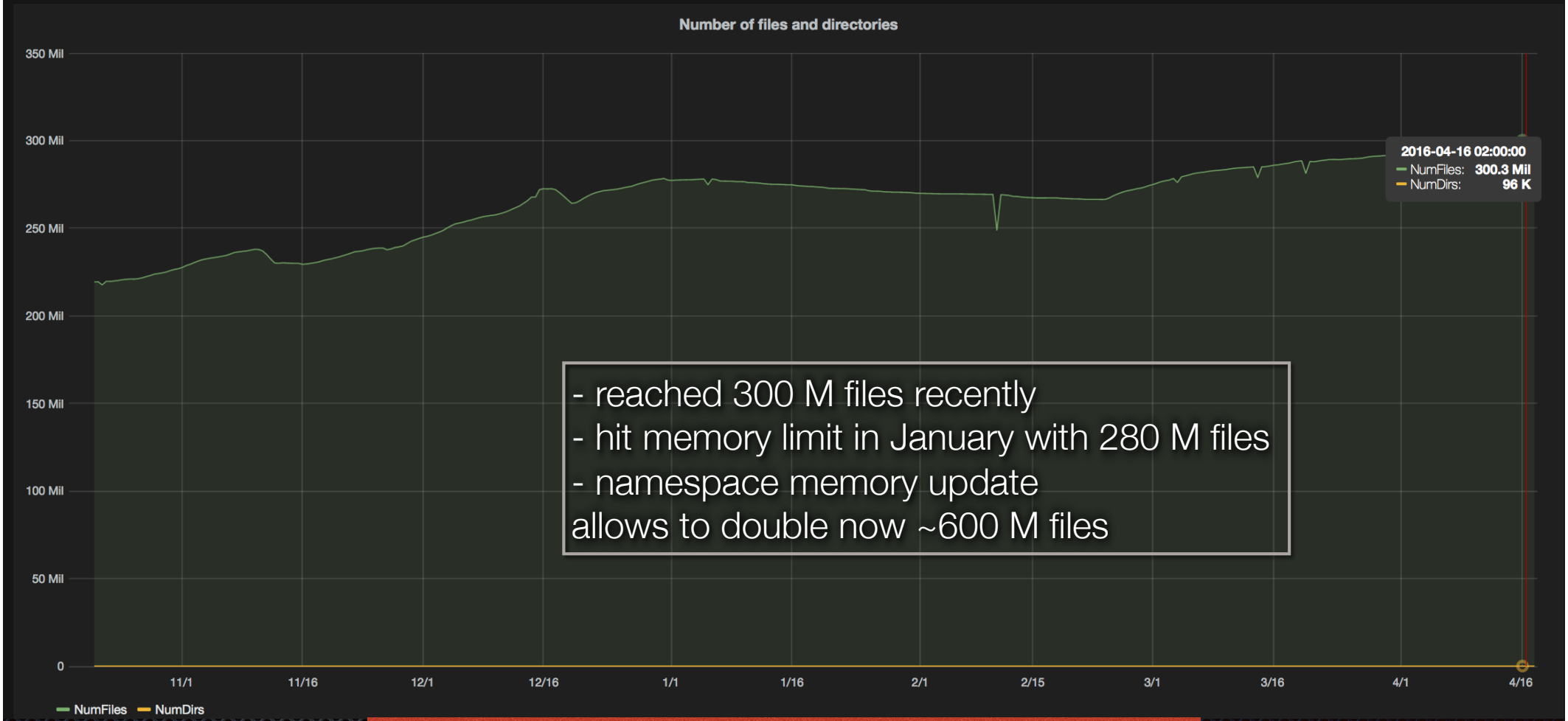




15 PB/day

read from disk for
client IO (10%) & scrubbing (90%)

EOS ALICE@CERN



- reached 300 M files recently
- hit memory limit in January with 280 M files
- namespace memory update

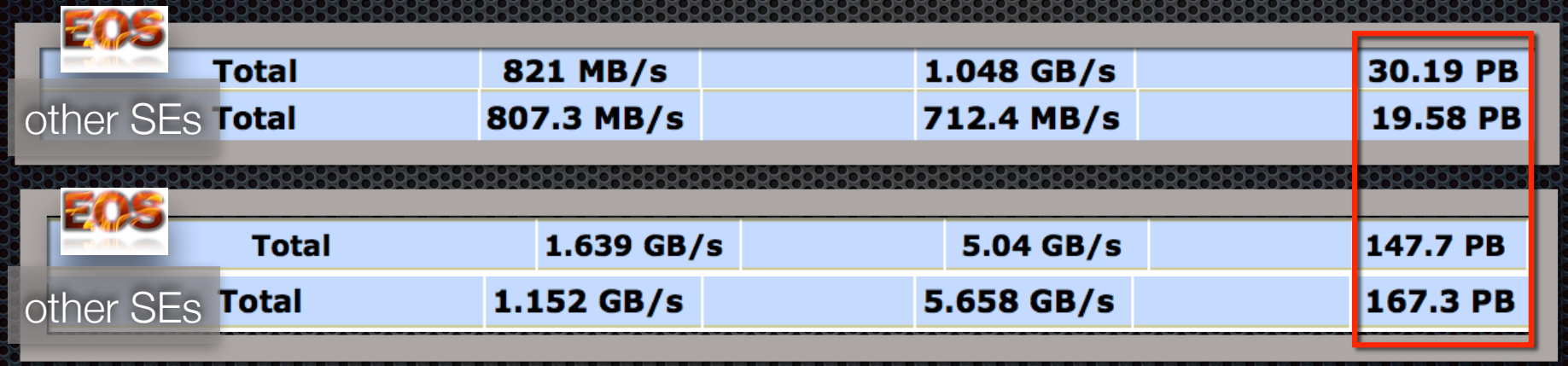
allows to double now ~600 M files

MGM memory 256 GB => 512 GB

EOS ALICE@CERN

1 year average 2015

data written

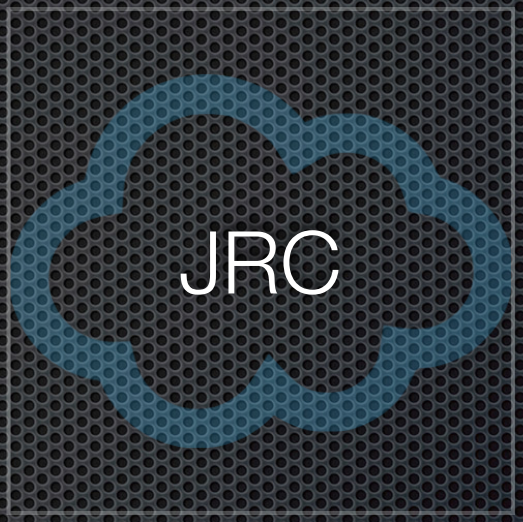


EOS	Total	821 MB/s		1.048 GB/s	30.19 PB
other SEs	Total	807.3 MB/s		712.4 MB/s	19.58 PB

data read

EOS	Total	1.639 GB/s		5.04 GB/s	147.7 PB
other SEs	Total	1.152 GB/s		5.658 GB/s	167.3 PB

EOS @outside HEP



Distributed Storage
in Australia

OwnCloud

aarnet
Australia's Academic
and Research Network

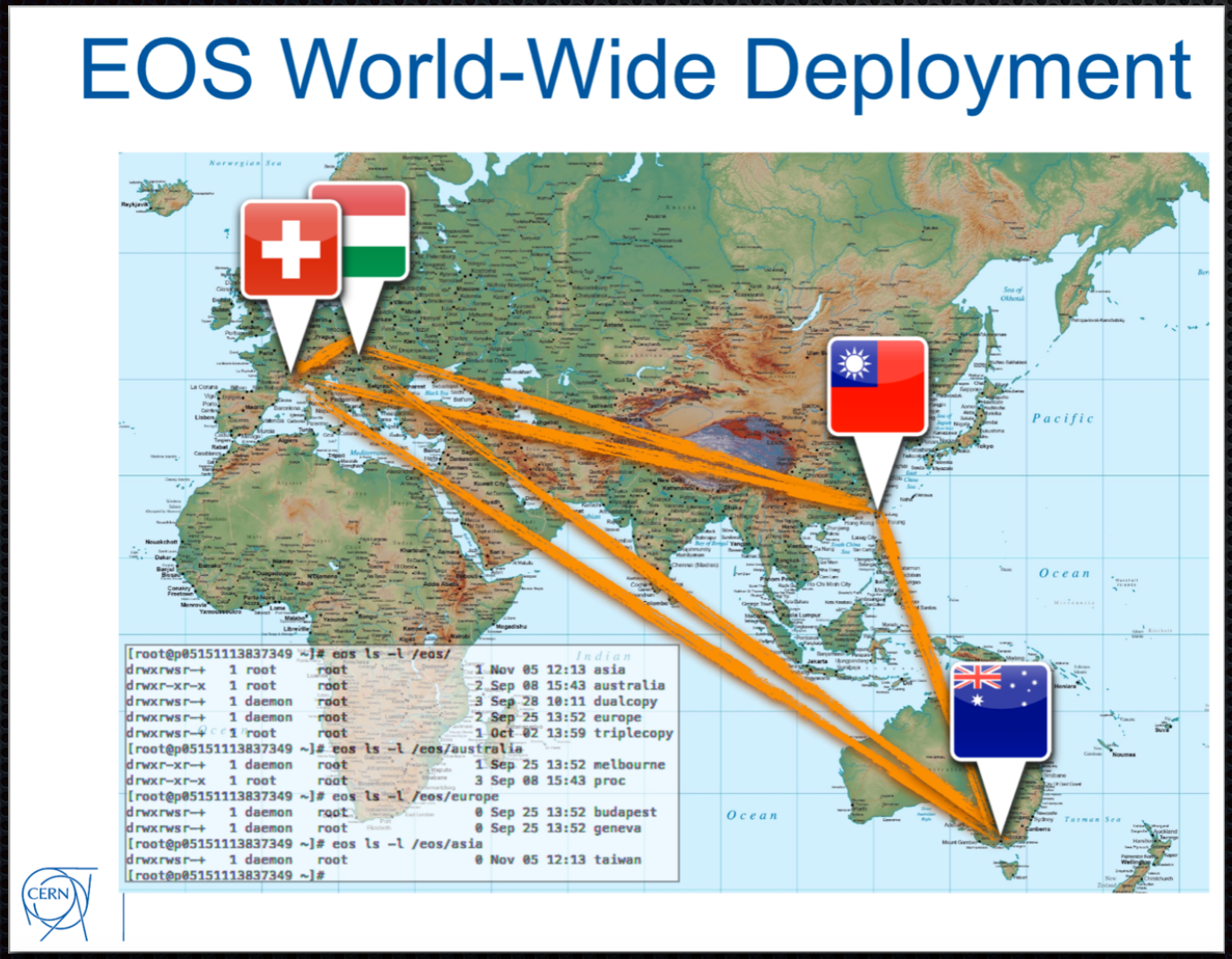
Earth Observation
& Social Sensing

BigData pilot project

Joint Research Centre (JRC)
European Commission

Distributed EOS

Test Deployment with 300ms latency



disk server in
Geneva
Budapest
Taipei
Melbourne

EOS web

new web portal eos.cern.ch

CERN Accelerating science

Sign in Directory

eos.web.cern.ch

Large Disk Storage at CERN

CERNBox

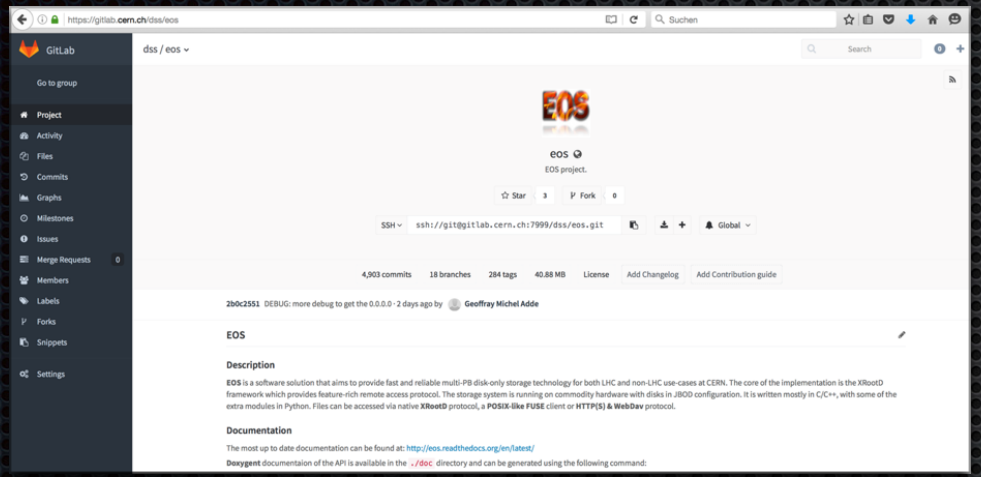
About | CERN Service | DOCs | EOS Resources | News

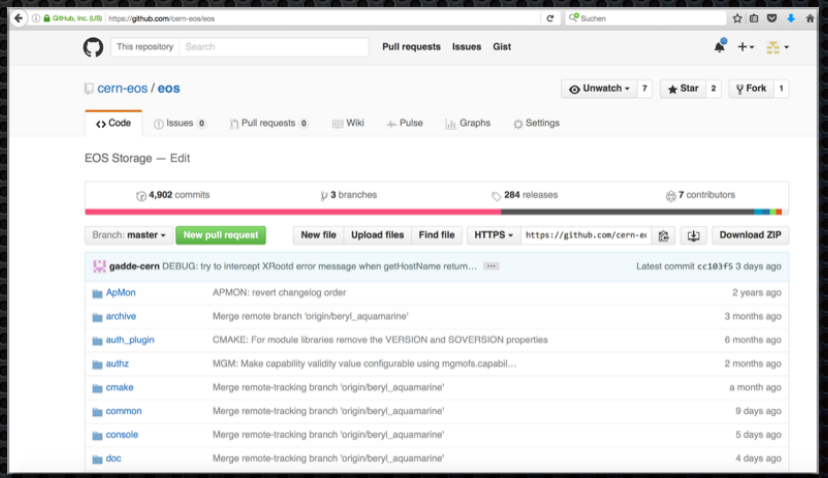
Welcome to CERN's EOS!

Tier-0 disk storage system for LHC

EOS code

new GIT repositories





gitlab.cern.ch/dss/eos

for users with
CERN account

connected to JIRA buck tracker

github.com/cern-eos/eos

for non CERN users

GitHUB issue tracker

EOS repository

new YUM repositories

Production Branch

Aquamarine

https://dss-ci-repo.web.cern.ch/dss-ci-repo/eos/aquamarine/tag/

Aquamarine-Dependencies

https://dss-ci-repo.web.cern.ch/dss-ci-repo/eos/aquamarine-depend/

Citrine

https://dss-ci-repo.web.cern.ch/dss-ci-repo/eos/citrine/tag/

Developement Branch

Aquamarine-Dev

https://dss-ci-repo.web.cern.ch/dss-ci-repo/eos/aquamarine/commit/

Citrine-Dev

https://dss-ci-repo.web.cern.ch/dss-ci-repo/eos/citrine/commit/

EOS support

CERN best effort

CERN Mailing Lists:

eos-support@cern.ch

- any question you might have around EOS deployment,configuration or operations
- issues/bug reports

eos-community@cern.ch **open subscription**

- announcement of releases, community events, exchange and other communication

EOS support

Commercial Support via **COMTRADE**

COMTRADE is CERN Openlab partner to provide easy EOS installation, documentation and a support contract for institutions, if this is a baseline requirement

Example: JRC - currently in preparation

Development Focus

EOS fuse

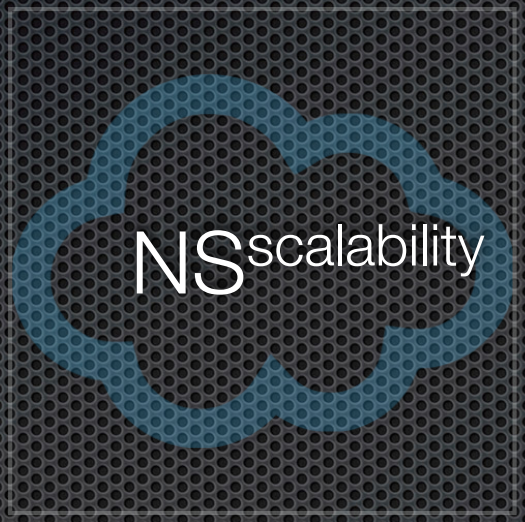
Motivation: provide FS access
Goal: deploy **/eos** in all lxplus/lxbatch

EOS CITRINE provides *kerberized* FUSE filesystem like AFS

- initially FUSE implementation relativly slow
- first round of refactoring of FUSE code done
- huge performance improvements - second round planned

compile XRootD - old 18 min => new~**2 min**
touch/s - old **4-25/s** new **80/s - 250/s**
wIOPS - **3k/s**
rIOPS (after write) - **300k/s**
seq write rate - **300 MB/s**

EOS scalability

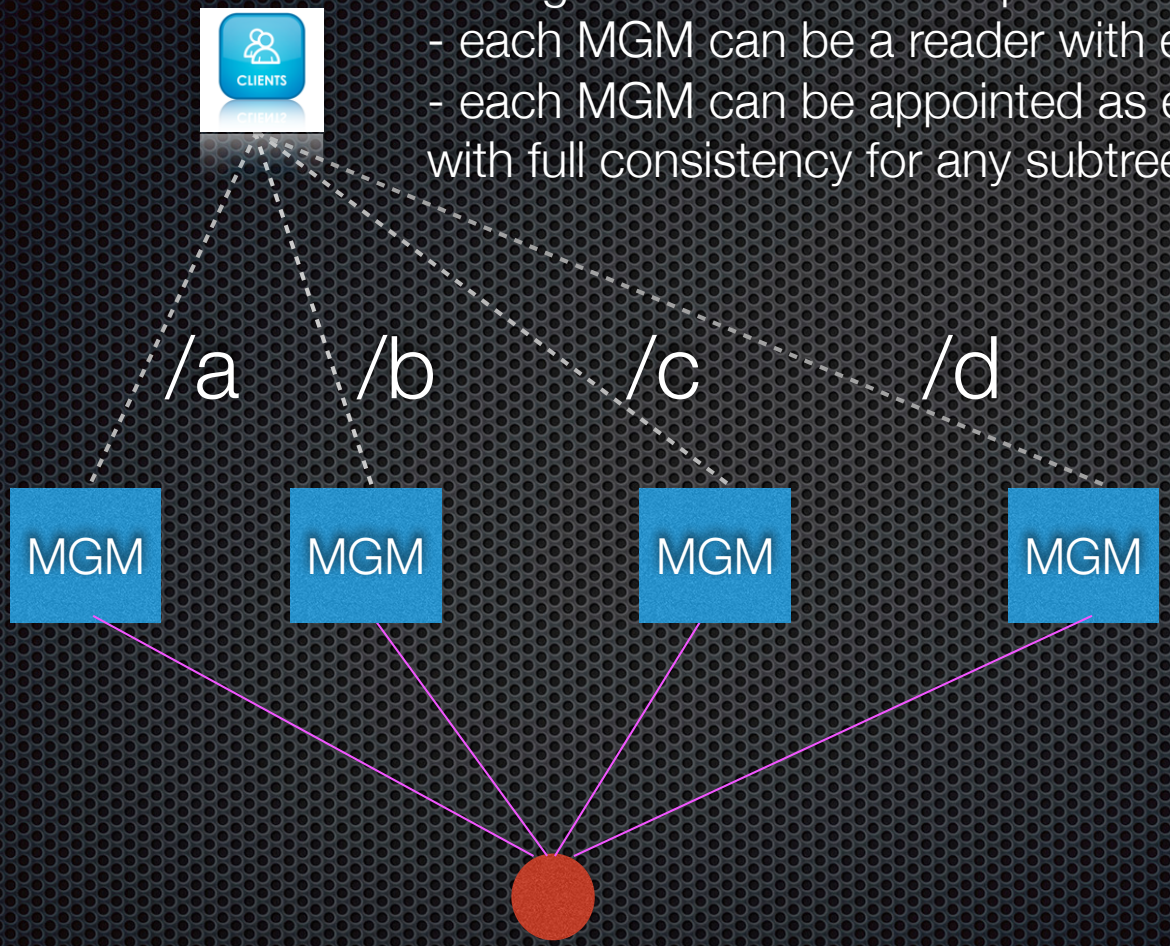


Motivation:
- operation point for in-memory NS has been reached in EOSALICE
- decouple meta-data service and NS persistency

Goal:
- provide multi-master scale-out implementation with identical low latency

multi master MGM

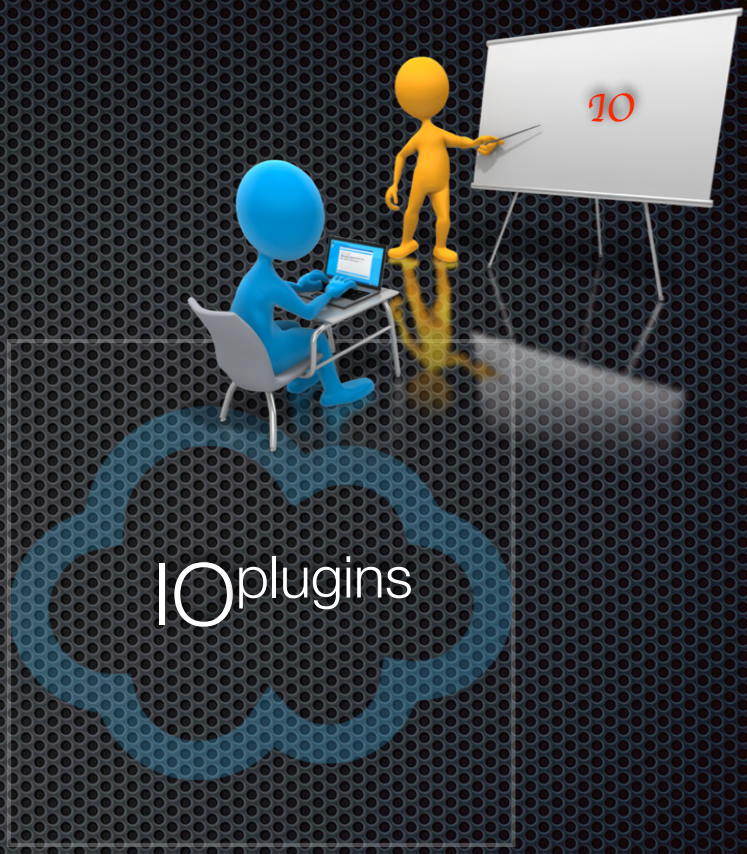
- assignment of current responsible MGM for each subtree
- each MGM can be a reader with eventual consistency
- each MGM can be appointed as exclusive reader/writer with full consistency for any subtree



- new persistency format
- no boot time
- much less memory - LRU in-memory cache of meta data
- transparent service migration

back-end HA KV store e.g. REDIS, ARDB … or XRootD!

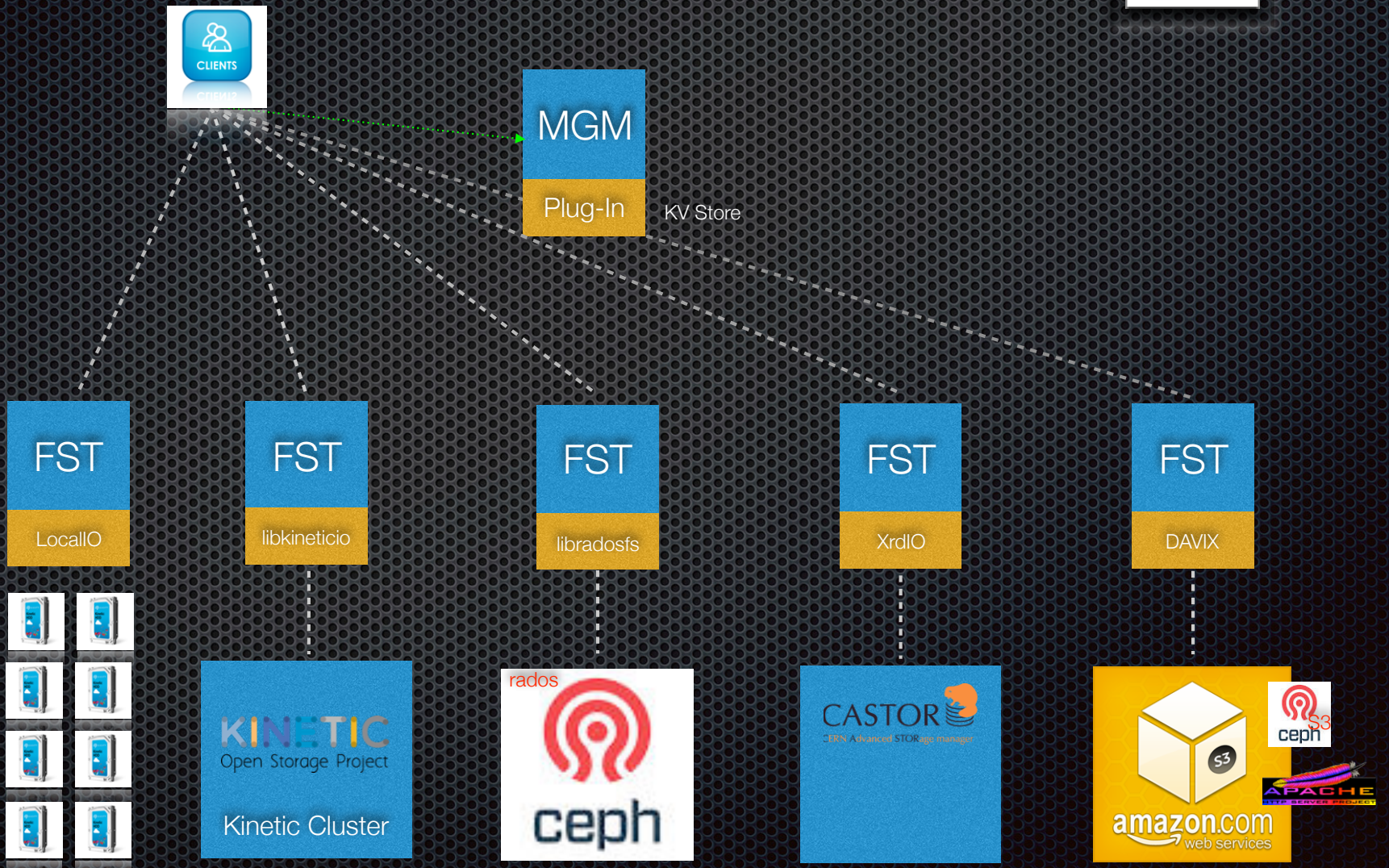
EOS IO plug-ins



Motivation: leverage state-of-the art storage technology
Goal: provide integration for external storage systems like Kinetic, Amazon, CEPH, HTTP based , XRootd based

EOS IO plug-ins

PLUG-IN EXAMPLES



EOS kinetic

- Kinetic disc support in EOS in pre-production test
  - **EOSGENOME** test instance (1PB) imported 300 TB of genome data from Amazon storage
    - stored on kinetic discs with RS (16,4)
    - stored on conventional JBOD with 2 replica for comparison
  - kinetic discs **virtually attached** to existing disc server
    - 100% IO compatible, random read-write, checksumming etc.
    - kinetic disc cluster attached as single EOS mount point
  - second kinetic pool with 2 PB (SMR) expected this year
  - **GUI development** s for administration planned for summer 2016

EOS tape

Cern Tape Archive

Motivation: CTA = CASTOR without disk pools

Goal: integrate CTA project with EOS as disk front-end

- added work flow engine to EOS
- provides event triggered actions

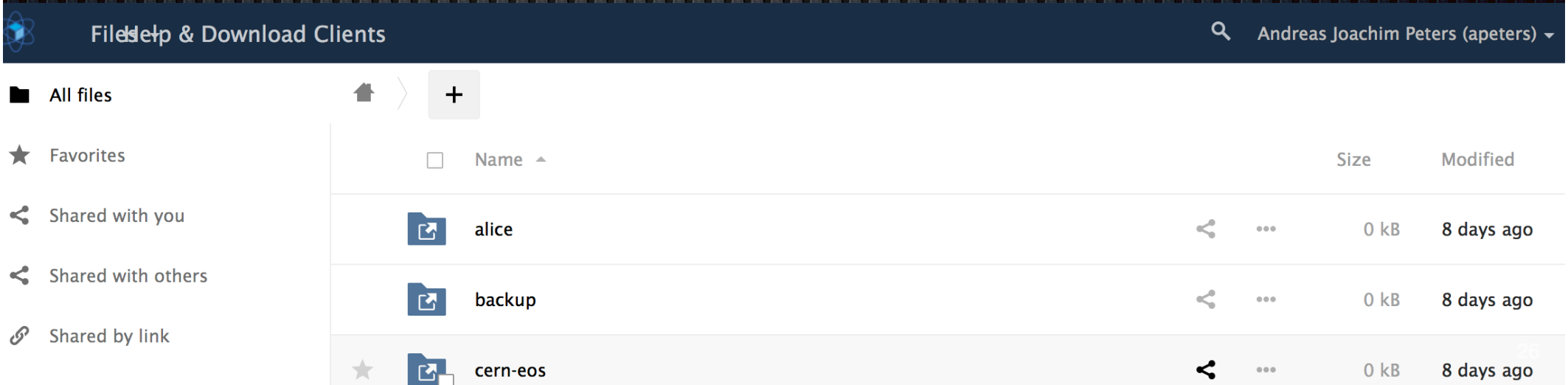
e.g. *if a new file arrives, store it on a tape*
or *make me a tar ball of all my root files*

EOS CernBOX & Swan

Motivation: dropbox, new physics desktop
Goal: better Owncloud, Jupyter integration …

CBox^extensions

… new CERNBOX web interface with faster rendering



http://swan.web.cern.ch/

SWAN

Interactive Data Analysis, in the Cloud.

Home | Useful Info | Gallery | FAQ | Contact Us | Talks and Publications

The Swan Service

SWAN (Service for Web based ANalysis) is a platform to perform interactive data analysis in the cloud.

- Analyse data without the need to install any software
- Jupyter notebook interface as well as shell access from the browser
- Synchronise your user space locally with CERNBox
- Access experiments' and user data in the CERN cloud

Don't know how to start? Visit the Gallery and the FAQ!

EOS distributed

Motivation:

- simplify storage usage for experiments - less entities

Goals:

- provide EOS for regional clouds and cloud federations
- support !CERN usage
- replication useful for low-volume data, not volume data like raw etc. - require global replication policies

WLCG Demonstrator coming …

https://indico.cern.ch/event/467565/

EOS development team

Core Team

A.J.Peters 80% core : CERNbox

E.A. Sindrilaru 50% core : ns : infrastructure : XRootD

G. Adde 100% core : cheduler : gridftp : secure FUSE

R&D/XRootd

M. Simon 20% FUSE R&D / rados

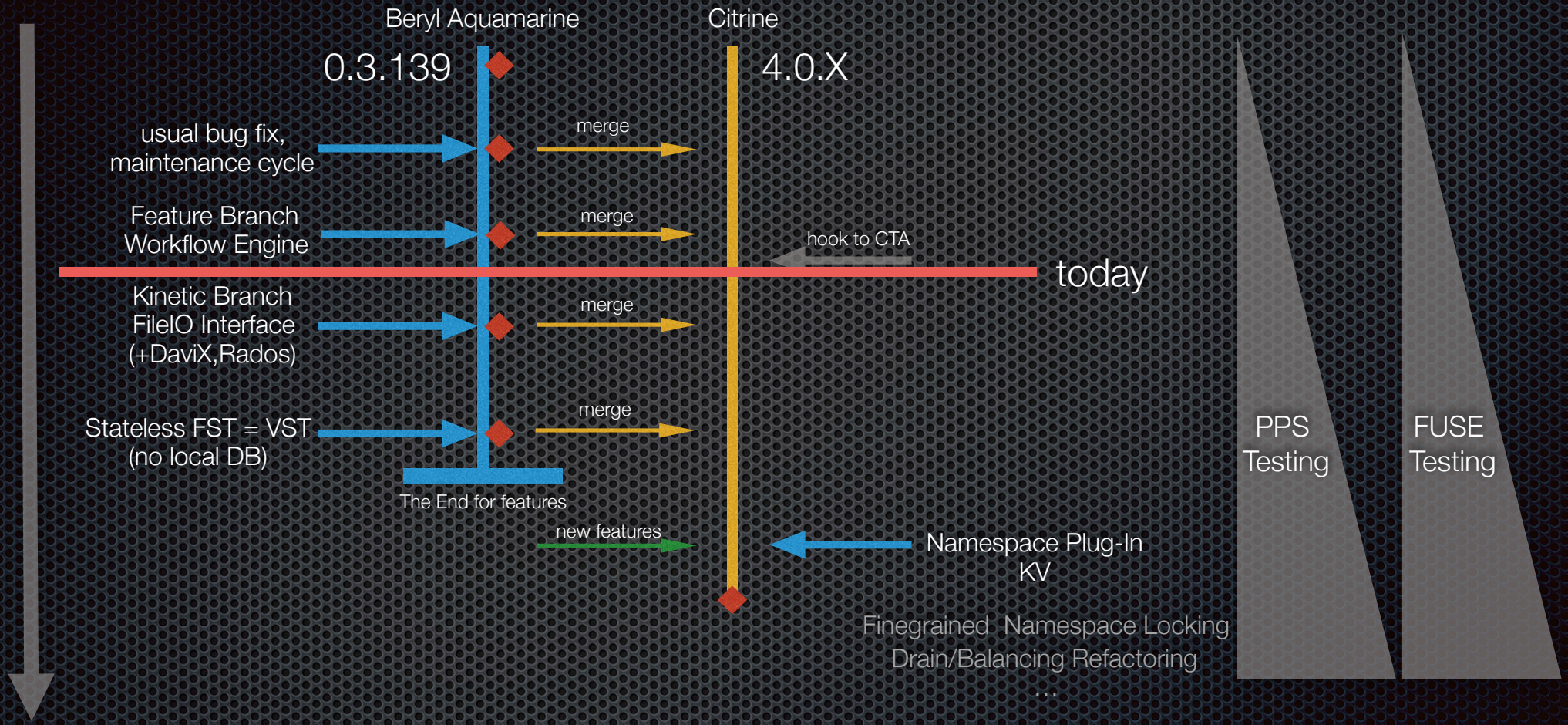
Seagate OpenLab Associate

P. Lensing - 100% libkineticio

EOS releases

- EOS **Aquamarine** (current 0.3.166 for LHC run)
  - used by most installations
  - if you use RAIN **please update** to >= 0.3.152 see GITHUB
- EOS **Citrine** >=4.0.17
  - */eos* lxplus client & PPS instance
  - if you want/need XRootD4 (IPV6), you can use Citrine !

EOS release plan



- **EOS software on very good track**
  - very stable production version (MGM uptime >> month)
    - no observed bugs in XRootD 3.3.6 since > year
  - new momentum by new communities joining
    - huge progress as mounted filesystem - still way to go to target AFS usage
  - a lot of innovation
    - cloud interfaces (Amazon S3, CERNBOX)
    - kinetic interfaces
    - object storage interfaces
    - scale-out namespace
    - workflow & tape interfaces
  - open software with very short release cycles
    - warmly invited to join as developers, storage provides or users

thank you
tenki
hvala
ขอขอบคุณคุณ
takk
спасибо
kam sah hamnida
dank je
הדות
dhanyavadagalu
tack
дзякуй
dziękuję
bedankt
blagodaram
mèsi
xièxie
tanemirt
gracias
rahmet
enkosi
mochchakkeram
trugarez
djere dieuf
arigatô
danke
kop khun krap
faafetai lava
manana
diolch
akun
shukriya
ありがとう
kia ora
dankon
děkuji
ačiū
grazzi
mamnun
gràçie
kiitos
spas
barka
dhanyavad
gratias ago
tau
शुक्रिया
dankie
gràcies
sulpáy
tapadh leat
chnorakaloutioun
teşekkür ederim
bayarlalaa
σας ευχαριστώ
obrigada
mahalo
didi madloba
chokrane
rahmat
sagolun
murakoze
taiku
sukriya
obrigado
dakujem
misaotra
welalin
najis tuke
terima kasih
mercé
ارك ش
köszönöm
asante
grazie
謝謝
mersi
sobodi
nandri
mauruuru
merci
nanni
vinaka
matondo
cảm ơn bạn
go raibh maith agat
paldies
ngiyabonga