Massimo LAMANNA

for the CERN IT Storage group

XRootD

HTTPS

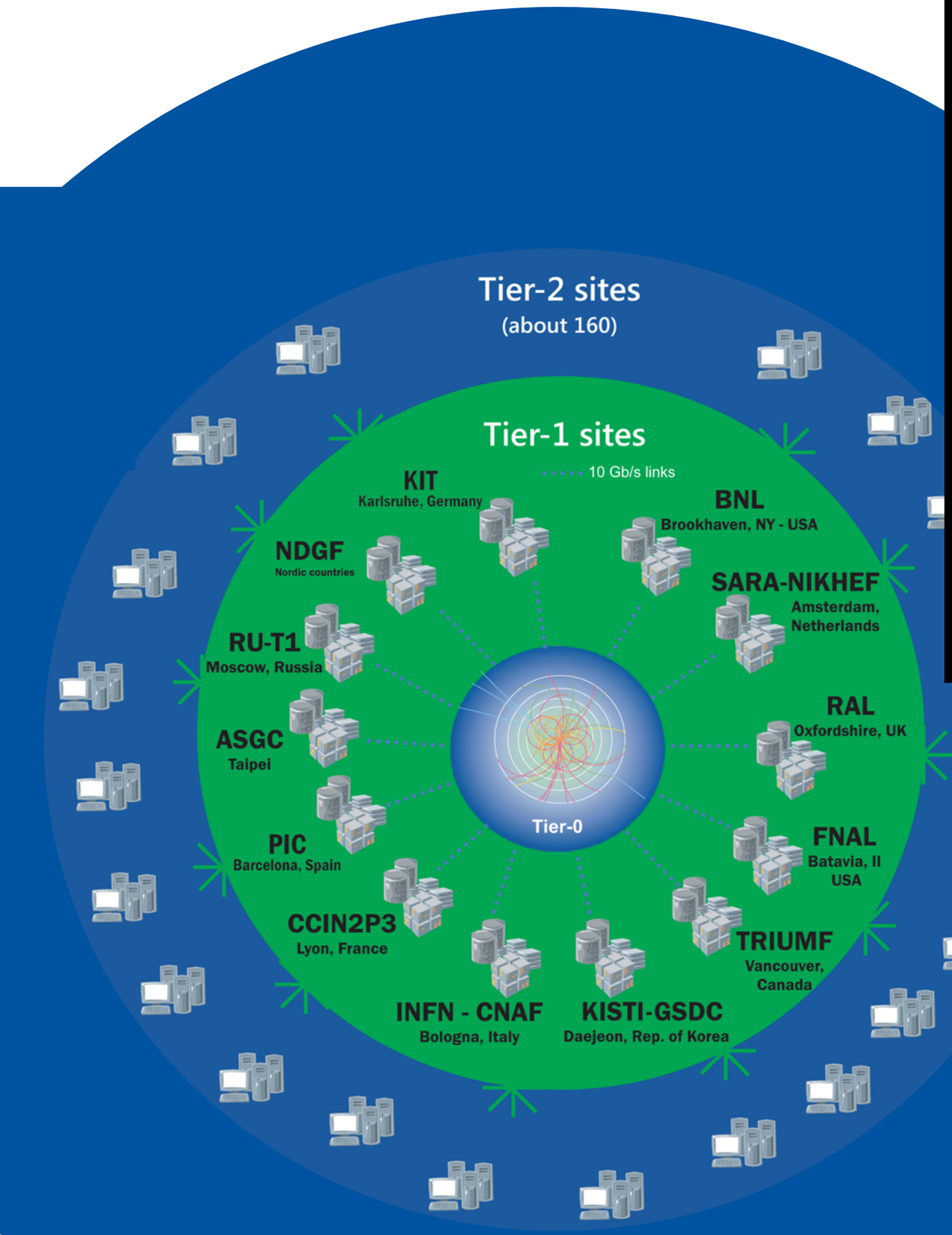
SRM

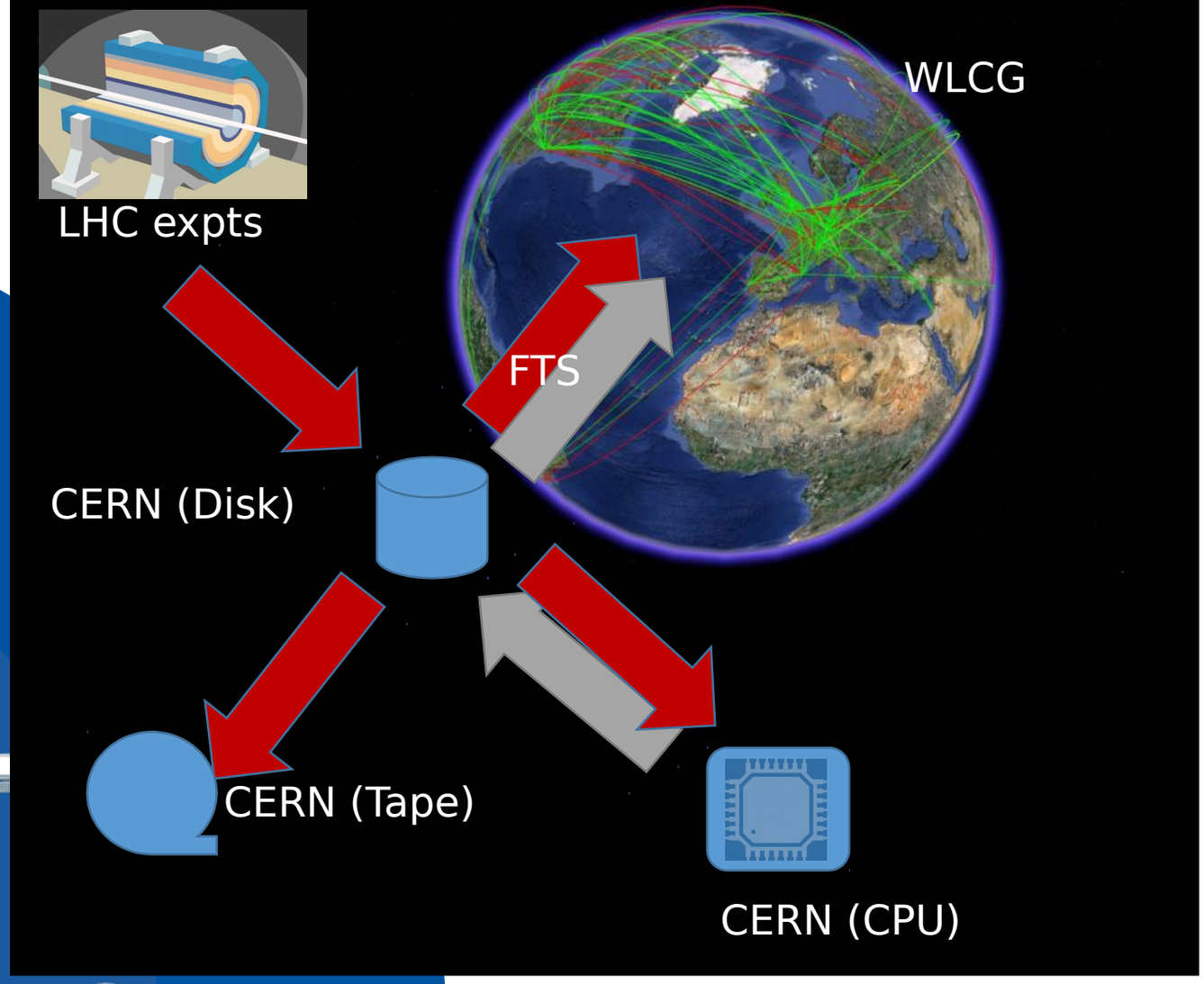
WebDAV

HTTP OwnCloud

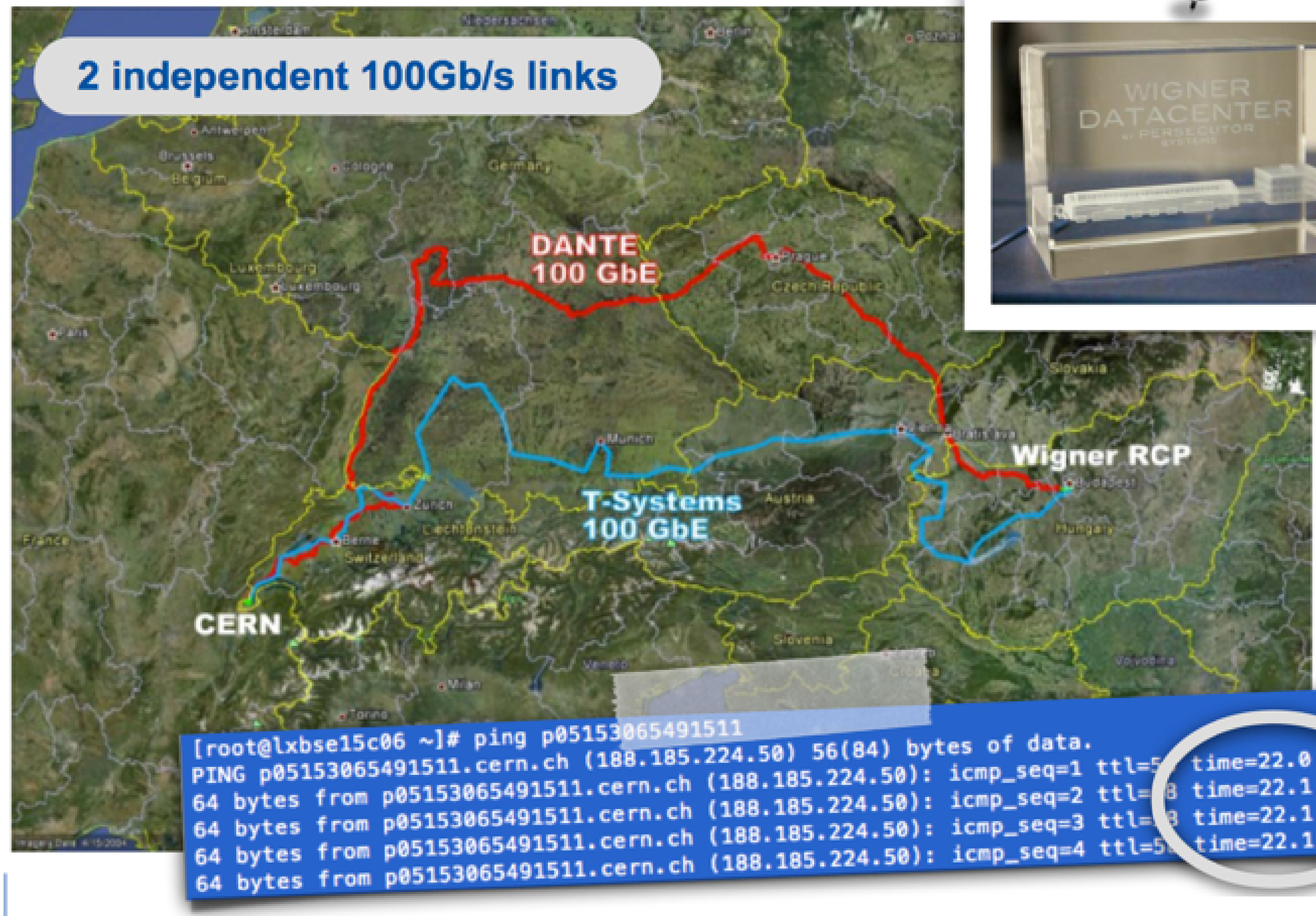
FUSE gridFTP

WLCG computing





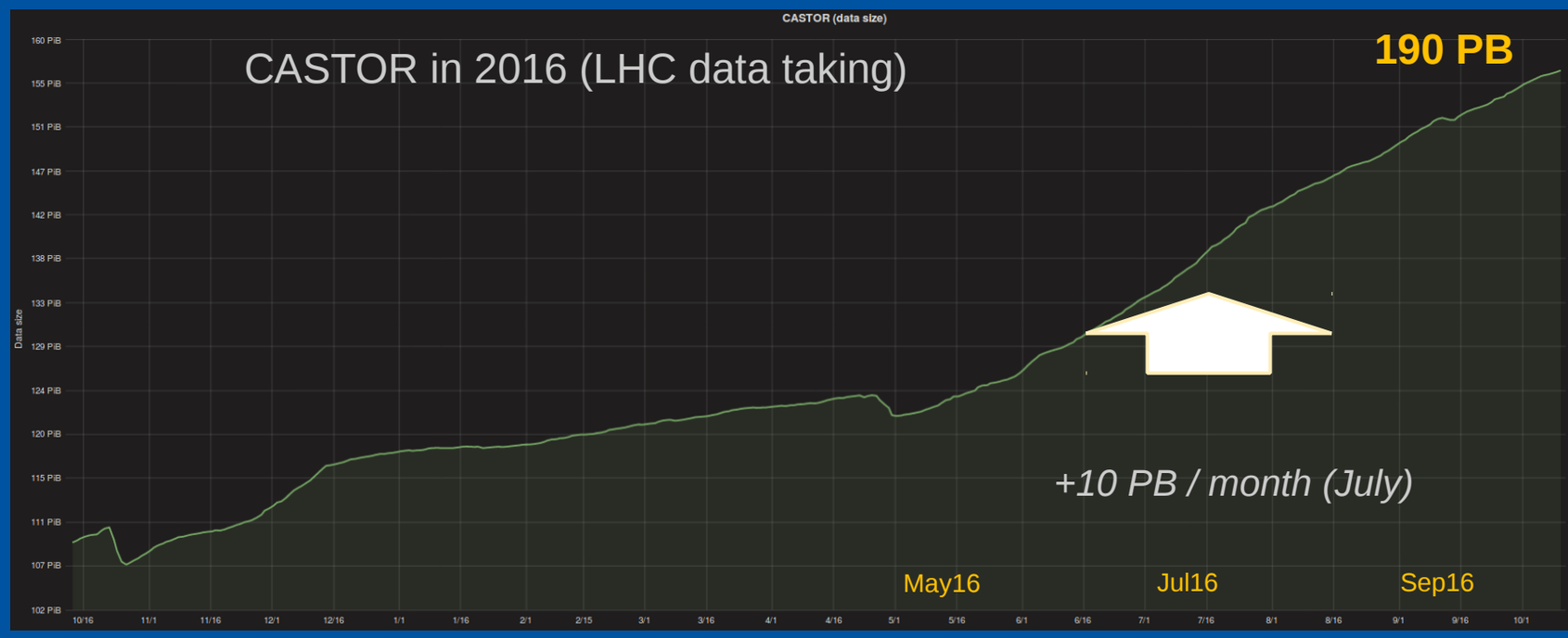
Wigner Computer Centre



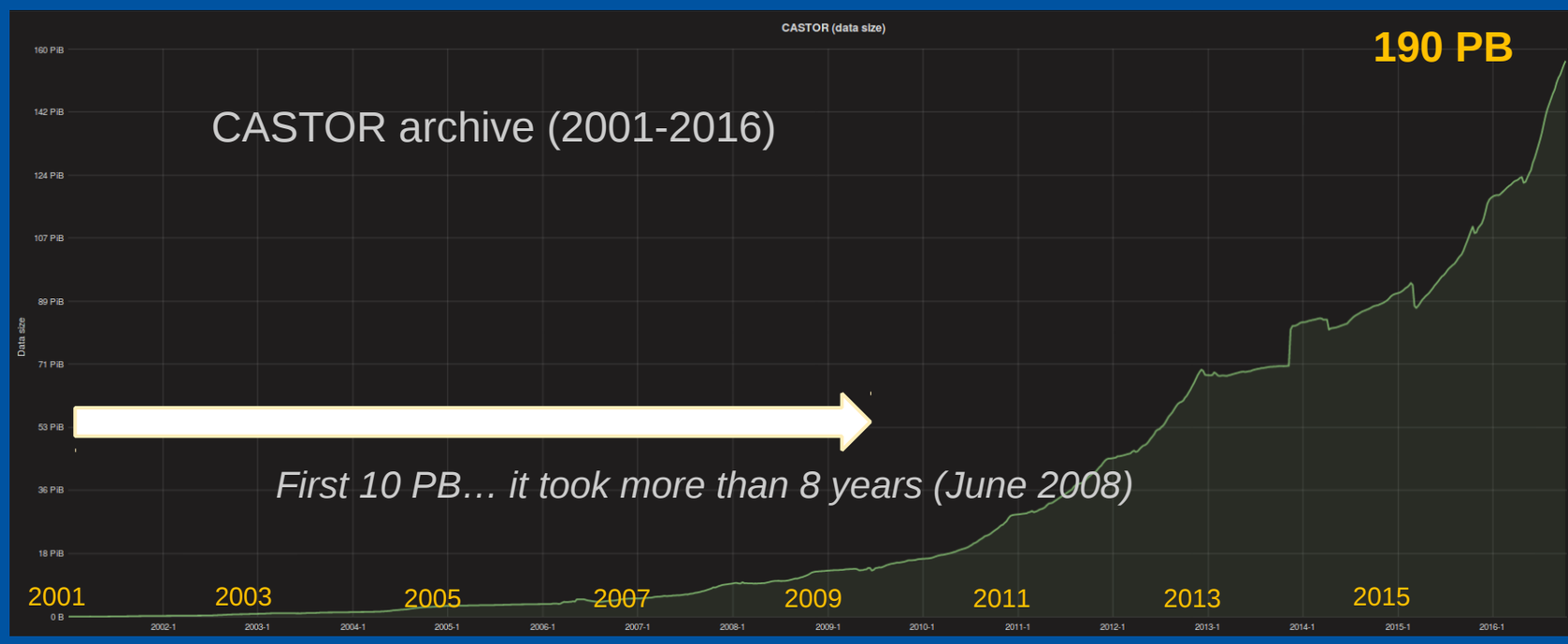
22ms latency

Data size (CERN tape archive CASTOR)

Record LHC data taking in July (and very high all across 2016)



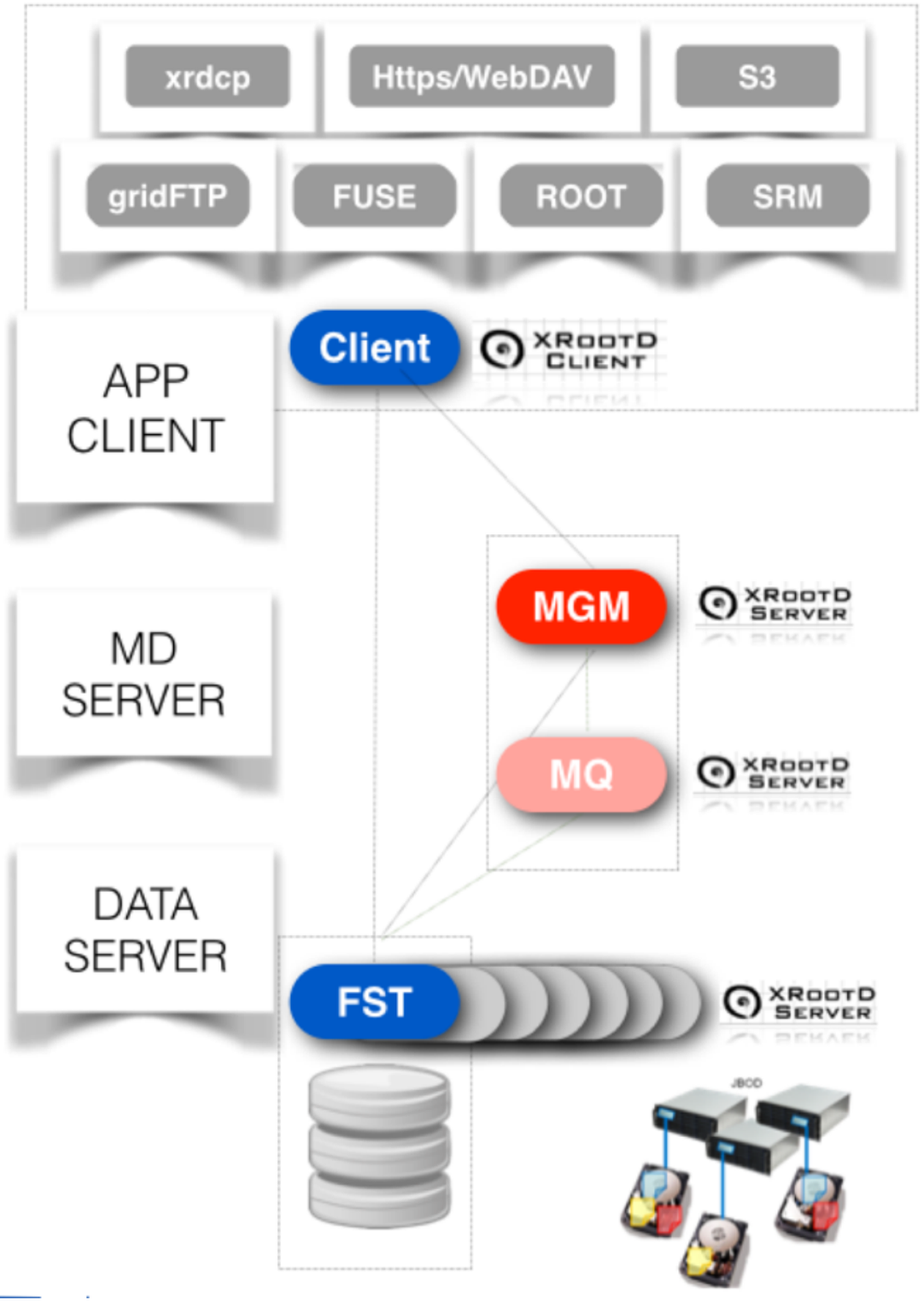
In **one** month accumulated the same amount of data as in the 2000-2008 period!



1000s of users

ATLAS (the largest LHC experiment) comprises 5,000 scientists from about 180 institutions around the world from 38 countries. It is one of the largest collaborative efforts ever attempted in science. Almost 1,200 doctoral students are involved in detector development, data collection and analysis. The collaboration depends on the efforts of countless engineers, technicians and administrative staff.

Architecture



- project since 2010
- production since 2011
- simple - license-free - JBOD hardware
- in-memory namespace
- strong security
  [server side security]
- many protocols
- quota, tunable QoS
- Dev&Ops @ CERN/IT

EOS native primary protocol is

1 XRootD

LAN/WAN opt.

- used for File IO, internal communication/replication, CLI
- currently all EOS disk servers world-wide visible

Secondary protocols

2,3 gridFTP-SRM

WAN opt.

- provided via gateway machines using VDT implementation + DSI plu
- provided via BestMan, FUSE mount & quota scripts

4 FUSE *not a protocol

POSIX flavor

- client: implemented over XRootD protocol
- no lock support, 95% compatible

5 HTTP

- currently provided via embedded HTTP server and overlay network

6 HTTPS

- provided via NGINX gateway, proxy module & add. Auth headers

7 WebDAV

5 http flavors

- implemented in embedded HTTP server
- incomplete implementation

8 S3

- implemented in embedded HTTP server
- incomplete implementation

9 OwnCloud HTTP

- implemented in embedded HTTP server
- compatible with OC http server

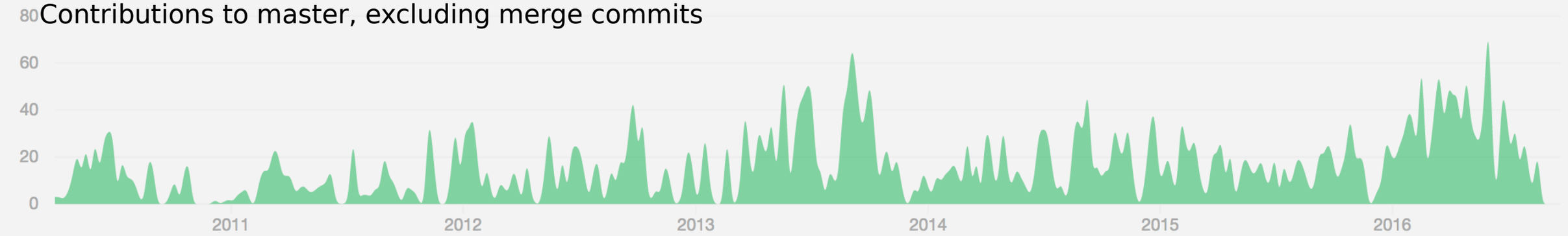
- used by experiments
  XRootD > gridFTP >> FUSE, HTTP(S)

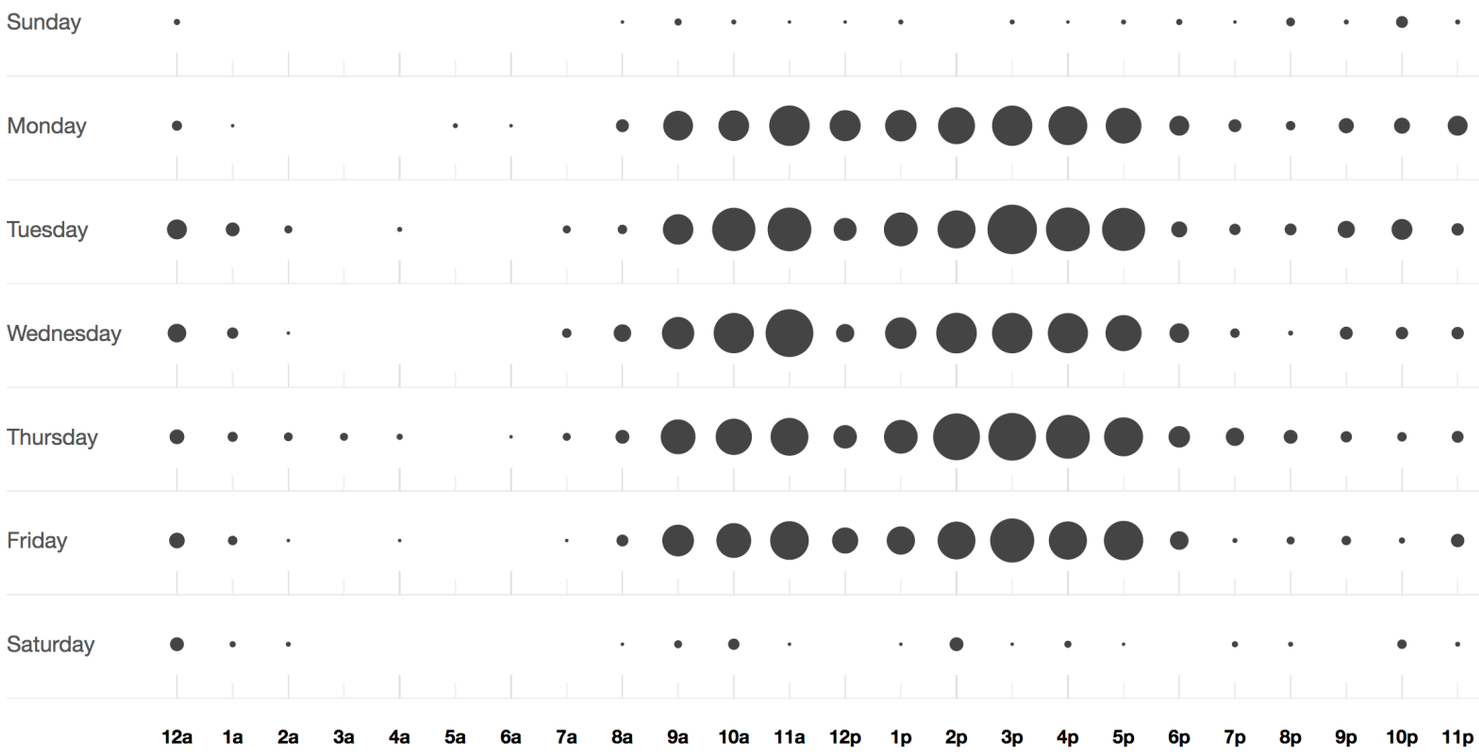
- XRootD most reliable and performant implementation in EOS

- no new protocols foreseen
  deployment change on the horizon
  all external traffic in- & out via XRootD gateways

- planned to provide XrdCl plug-ins to
  enhance features and performance
  loaded by client according to plug-in policies

Development: key numbers





GITHUB punch card

master: 181k LOC

production: 150k LOC

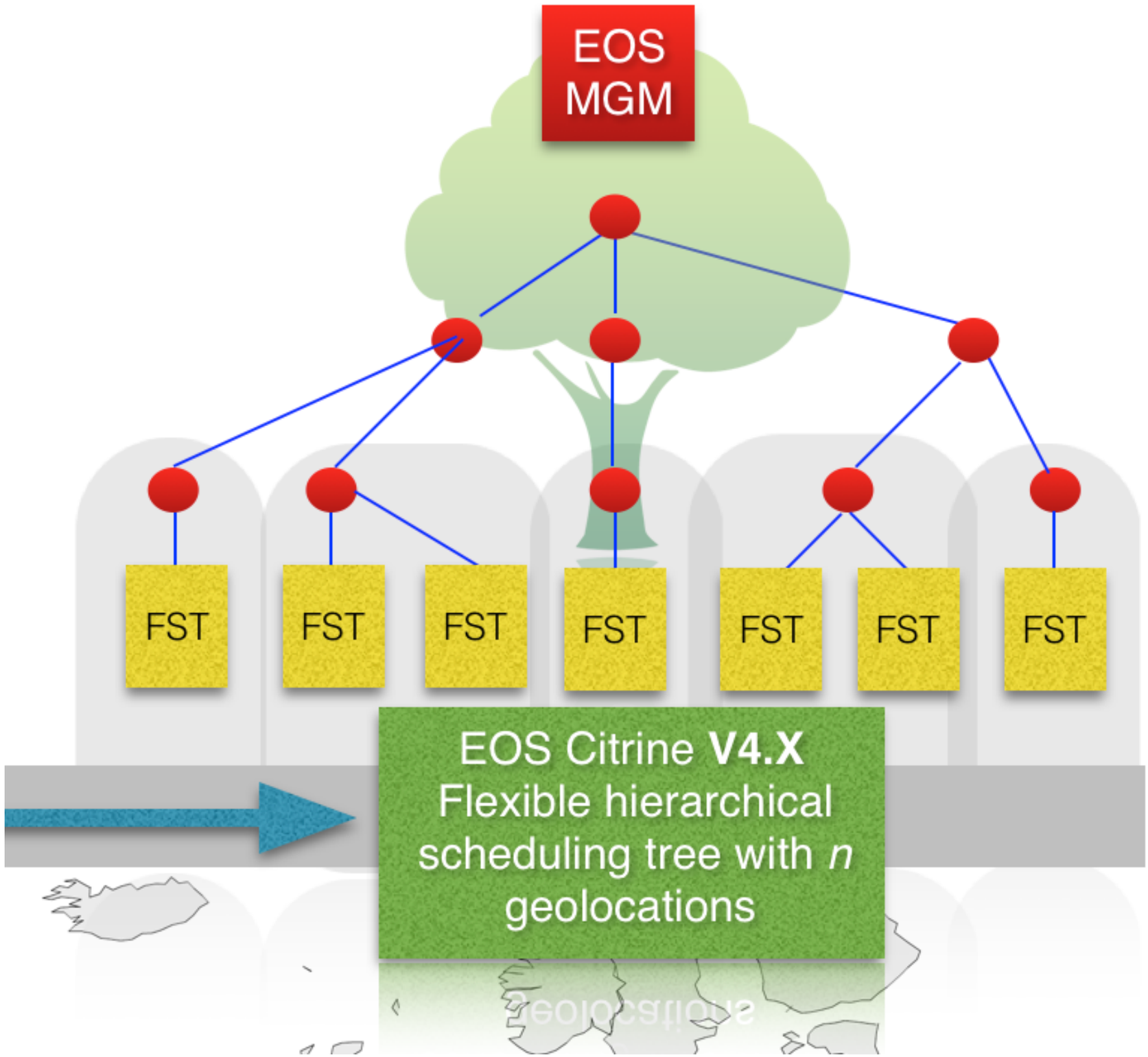
excluding external projects
jerasure/sqlite/gf-complete

70 tags in production branch
34 tags in master branch

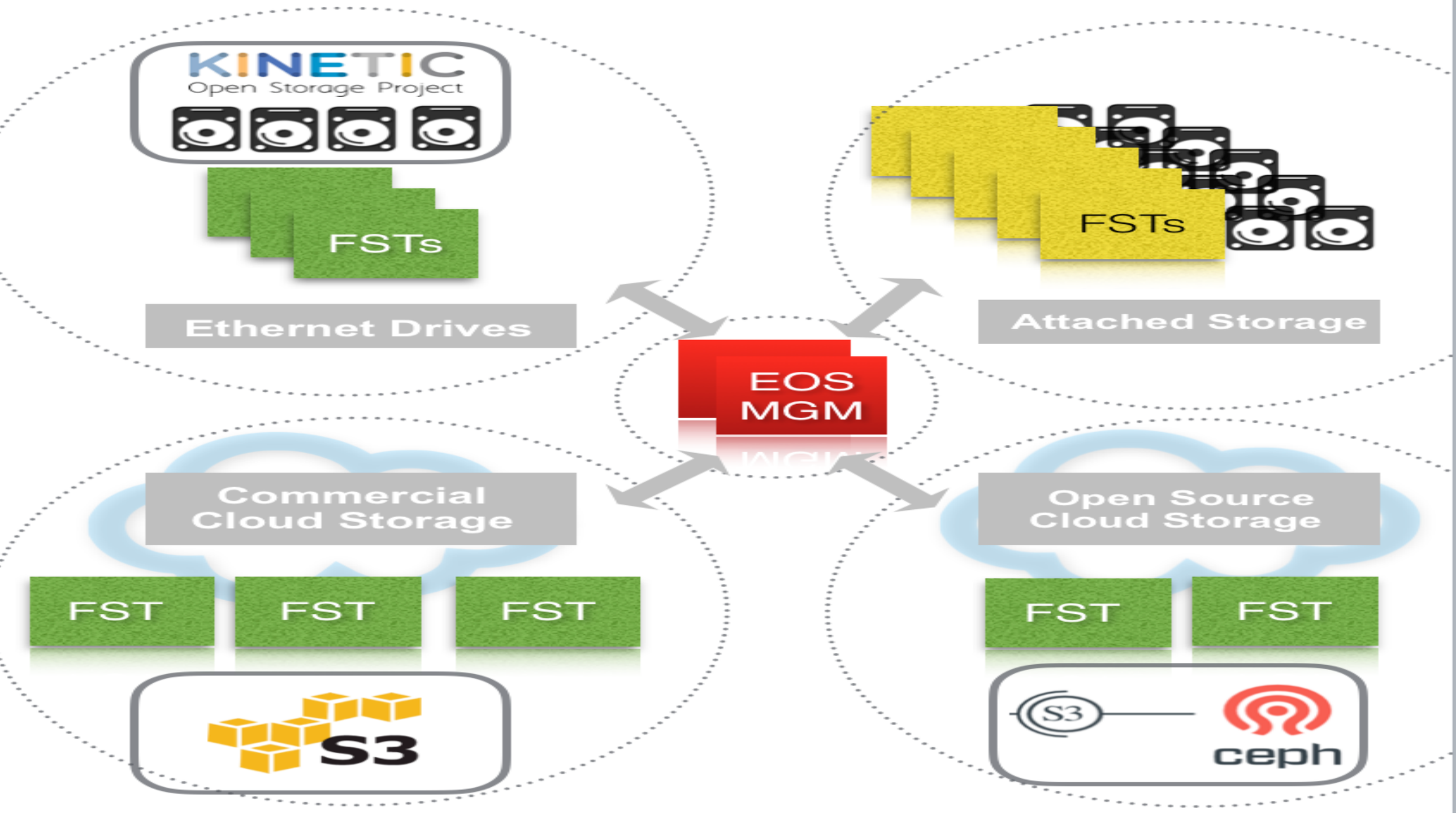
8 contributors

Open Source Storage

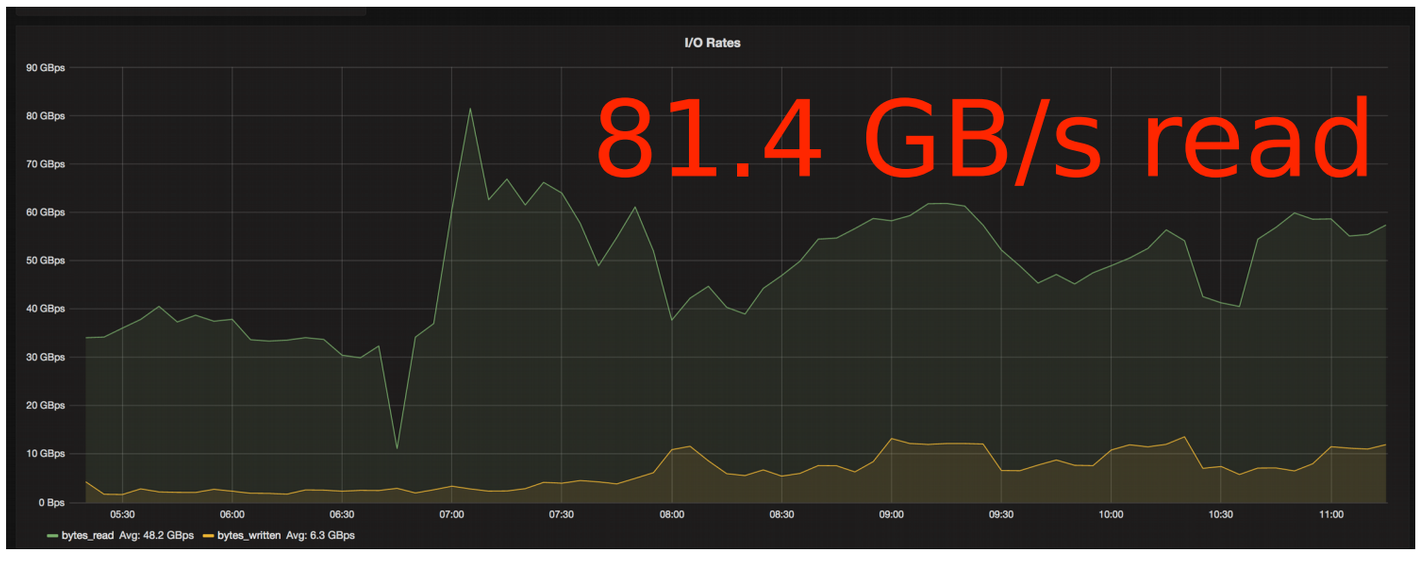
Large-scale (WLCG)



Large-scale (WLCG)



Operations: key numbers



- #file/#dir ratio **11:1**
- disk read/write **5.3:1**
  24 GB/s avg. read
  4.5 GB/s avg. write
- peak rates
  1 MHz read IOPS
  81.4 GB/s read
- background scrub
  avg. **165 GB/s** (100PB/week)
- **1143** storage nodes

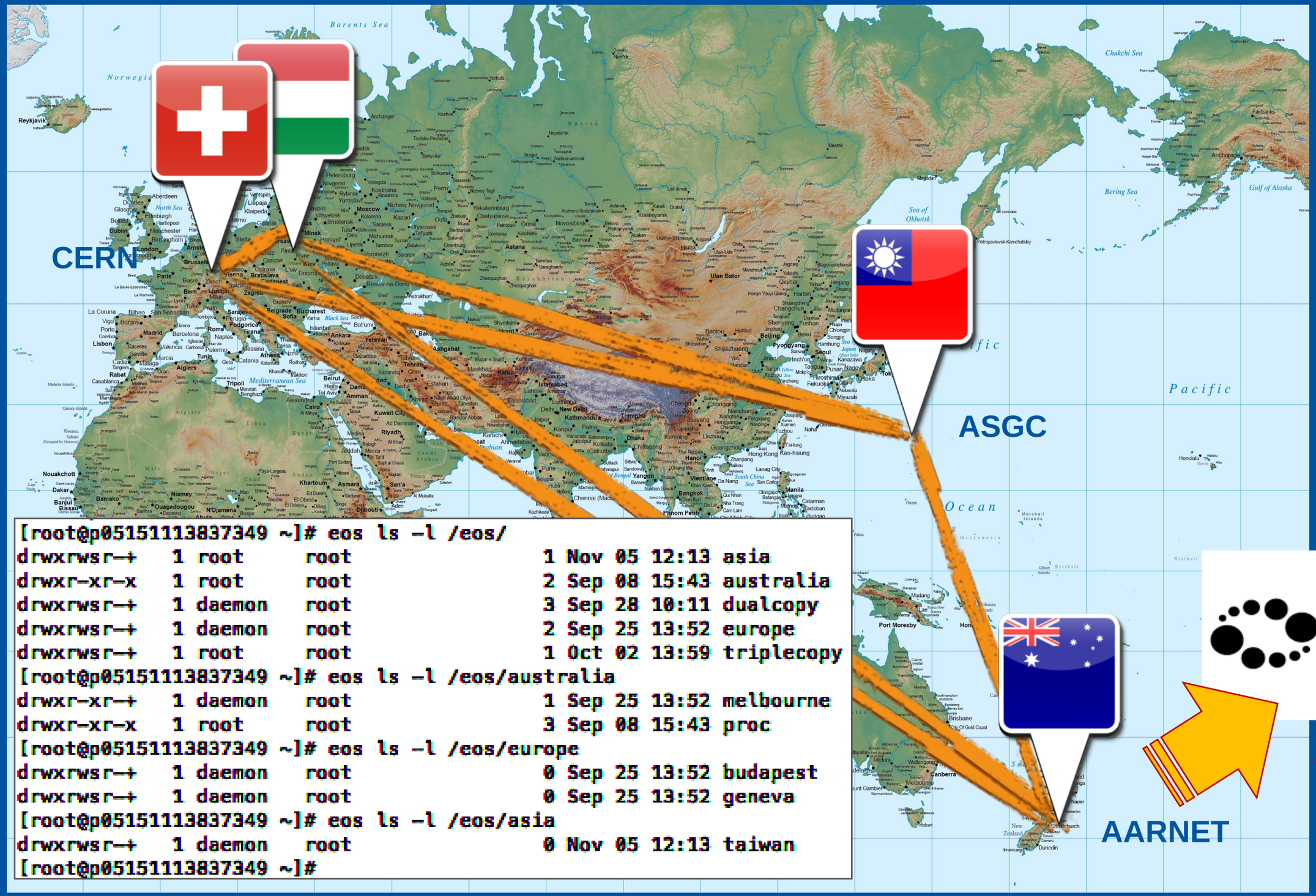
~200 PB disk space

1y growth rates: +85% more files +50% more space used
-14% less disks to manage (37k) +33% more disk space

External instances: FNAL, SASKE, Subatech, SINICA, RRC Kurchatov, UNAM, Kolkatta, LBL, JRC (EU), AARNET, ...

R&D - EOS World-Wide Deployment

aka “exploring the 300 ms region”...




```
[root@p05151113837349 ~]# eos ls -l /eos/
drwxrwsr-+    1 root      root                1 Nov 05 12:13 asia
drwxr-xr-x    1 root      root                2 Sep 08 15:43 australia
drwxrwsr-+    1 daemon    root                3 Sep 28 10:11 dualcopy
drwxrwsr-+    1 daemon    root                2 Sep 25 13:52 europe
drwxrwsr-+    1 root      root                1 Oct 02 13:59 triplecopy
[root@p05151113837349 ~]# eos ls -l /eos/australia
drwxr-xr-+    1 daemon    root                1 Sep 25 13:52 melbourne
drwxr-xr-x    1 root      root                3 Sep 08 15:43 proc
[root@p05151113837349 ~]# eos ls -l /eos/europe
drwxrwsr-+    1 daemon    root                0 Sep 25 13:52 budapest
drwxrwsr-+    1 daemon    root                0 Sep 25 13:52 geneva
[root@p05151113837349 ~]# eos ls -l /eos/asia
drwxrwsr-+    1 daemon    root                0 Nov 05 12:13 taiwan
[root@p05151113837349 ~]#
```

Questions?

www.cern.ch