CERNBox + EOS: End-user Storage for Science

Presenter:
Luca Mascetti

Authors & Co-authors:
Hugo Gonzalez Labrador, Massimo Lamanna, Luca Mascetti, Jakub Moscicki, Andreas-Joachim Peters, Elvin Sindrilaru

CERN - IT/DSS

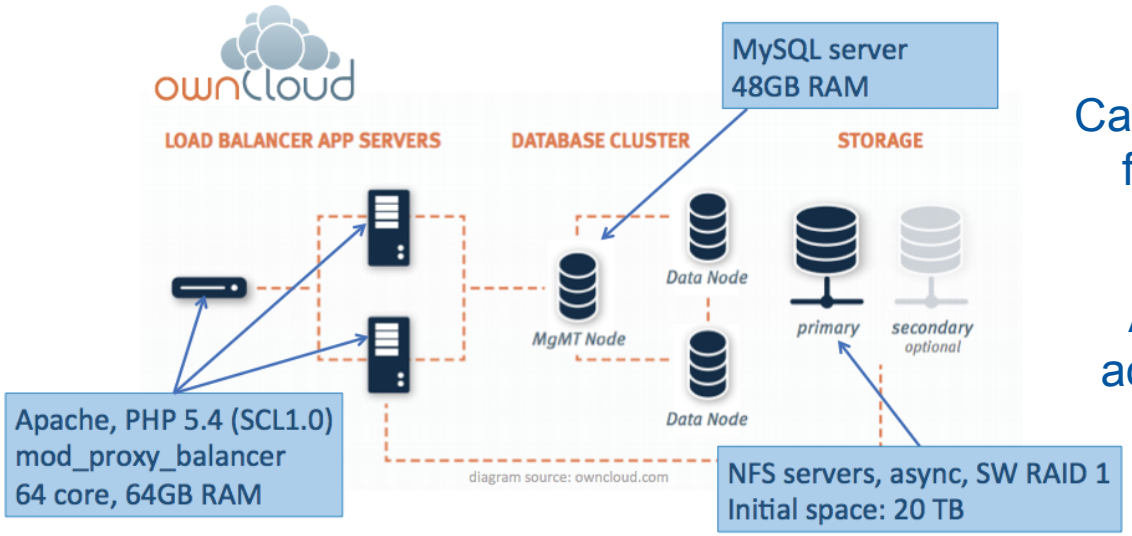
Why CERNBox?

- Competitive alternative to Dropbox for CERN users
  - 4500 distinct IPs contacting *.dropbox.com (daily)
  - SLAs: data availability and confidentiality
  - Archival and Back-up policies
  - Offline Data Access and Data sync across devices
  - Easy way to share files and folders with colleagues

CERN Disk Storage System

EOS	***140 PB (raw)***
CASTOR disk	***20 PB (raw)***
Ceph	***4 PB (raw)***

- We already manage large-scale online storage systems
- We started ownCloud evaluation and build prototype service



Can we integrate sync & share functionality with our main users workflow?

And being able to directly access the underlying data?

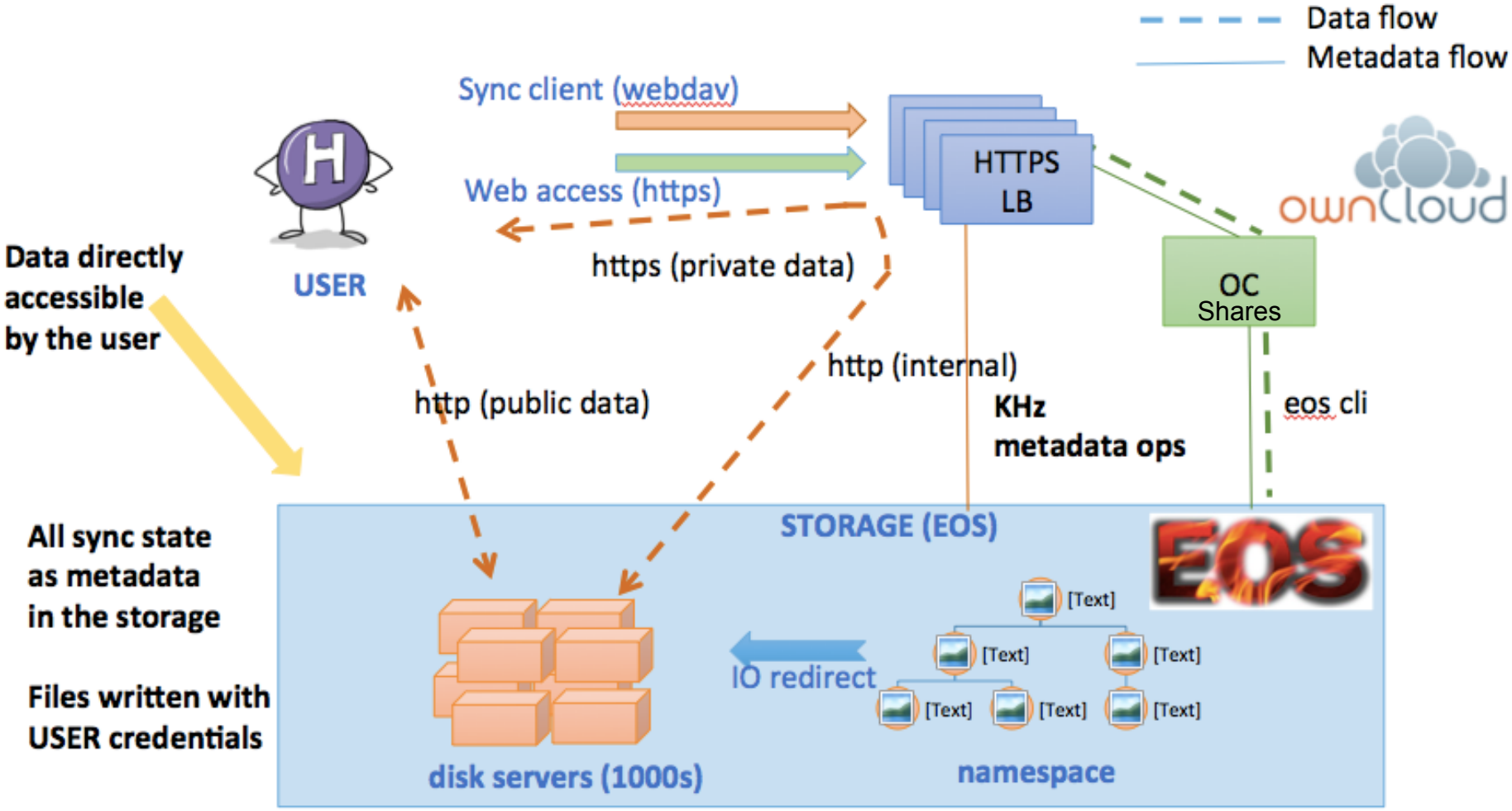
EOS Integration

- EOS offer “virtually unlimited” cloud storage for end-users
- Full solution compatible with **ownCloud clients**
  - we don’t want half-working CERN-specific solution
- Add a few features to EOS (or lifting restrictions)
- Remove dependency on the ownCloud Database
  - EOS has already a very fast i**n-memory namespace**
- Beef-up the webdav endpoint for ownCloud clients
- Integrate web-access and sharing functionality, trashbin, versions
  - Fusion between ownCloud model and EOS model
- Making EOS more robust
  - Lots, lots of testing of less stressed parts

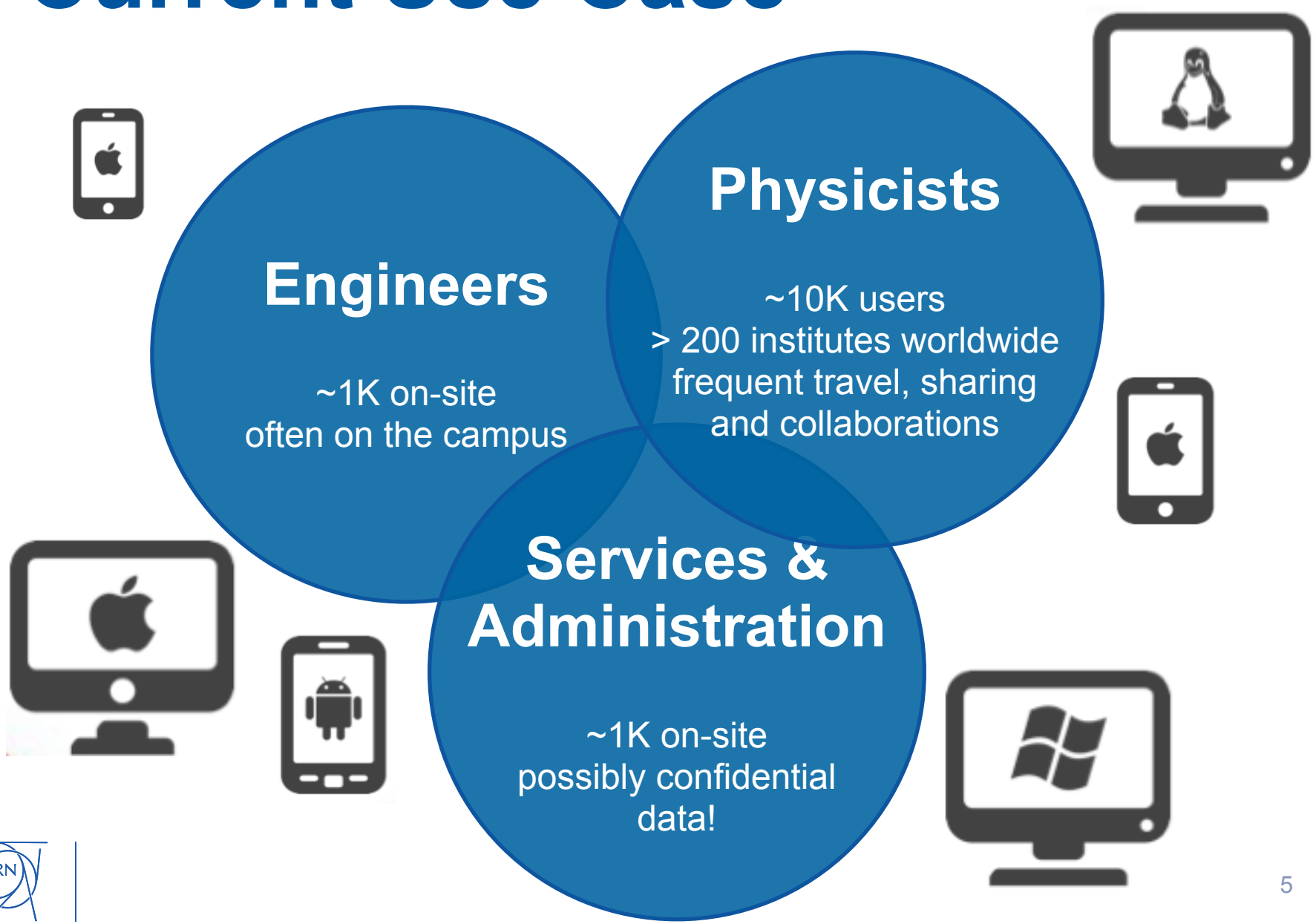
CERN

CERNBox and EOS

- CERNBox 2.0 architecture and EOS integration

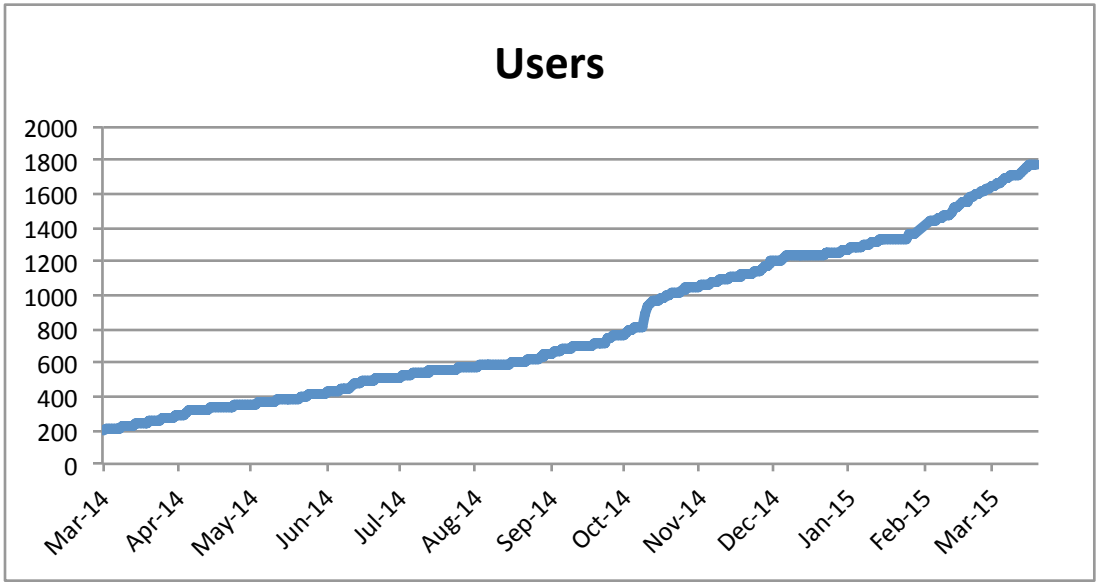


Current Use Case



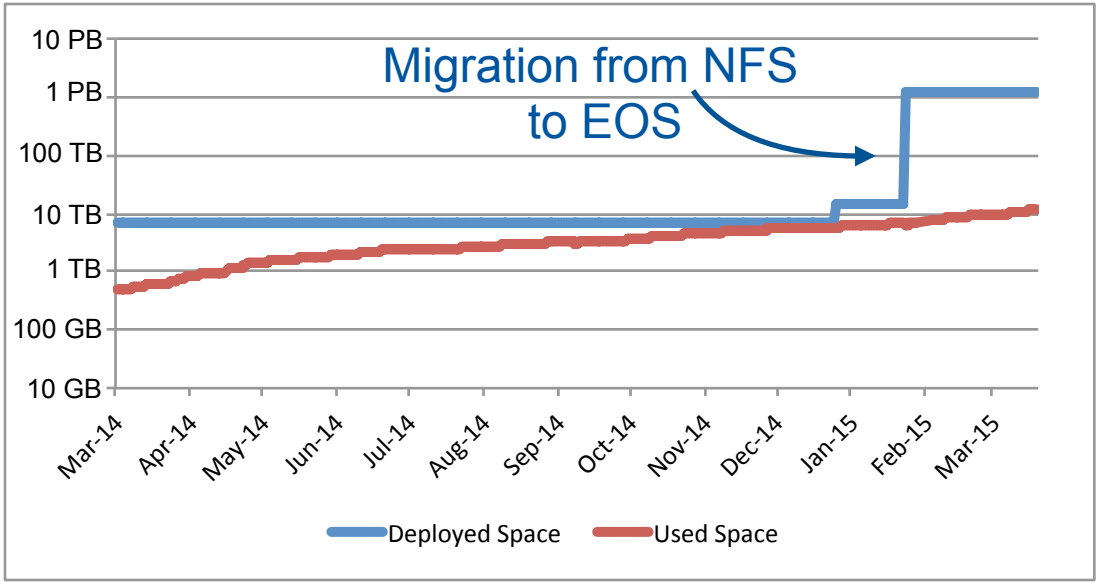
CERNBox Service Numbers

Users	*1877*
# files	*15 Millions*
# dirs	*1.3 Million*
Quota	*1TB/user*
Used Space	*37 TB*
Deployed Space	*1.1 PB*

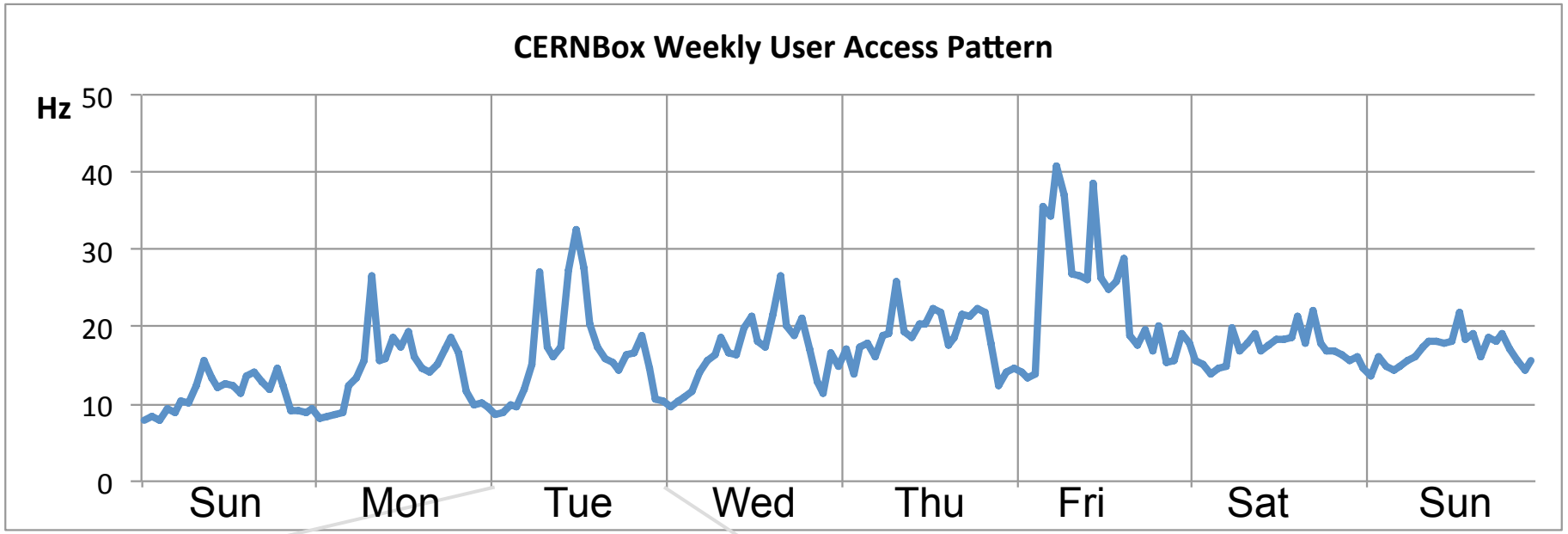


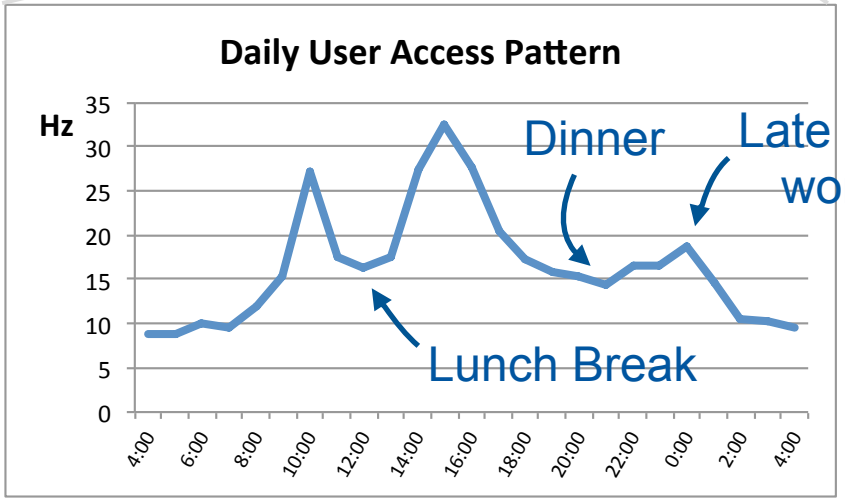
EOS offers "virtually unlimited" cloud-storage for our end-users

The EOS installation at CERN is around 140 PB with the primary role of storing physics data

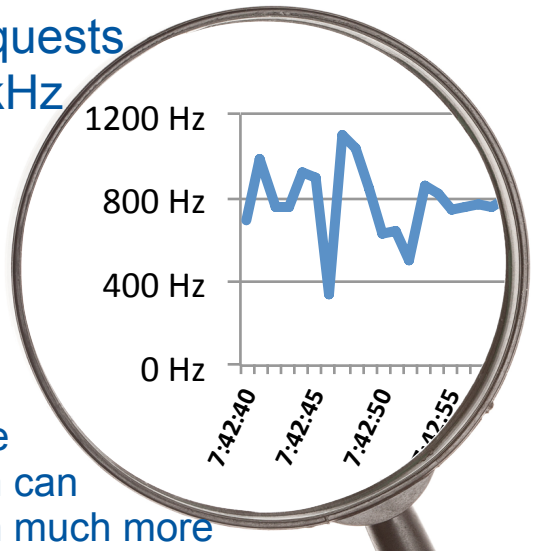


Current System Usage





Peak Requests at 1.1kHz



and the system can sustain much more

“Box Community”

Workshop organised at CERN very successful
- more than 80 participants from all over the world
- very interesting presentations and lots of ideas

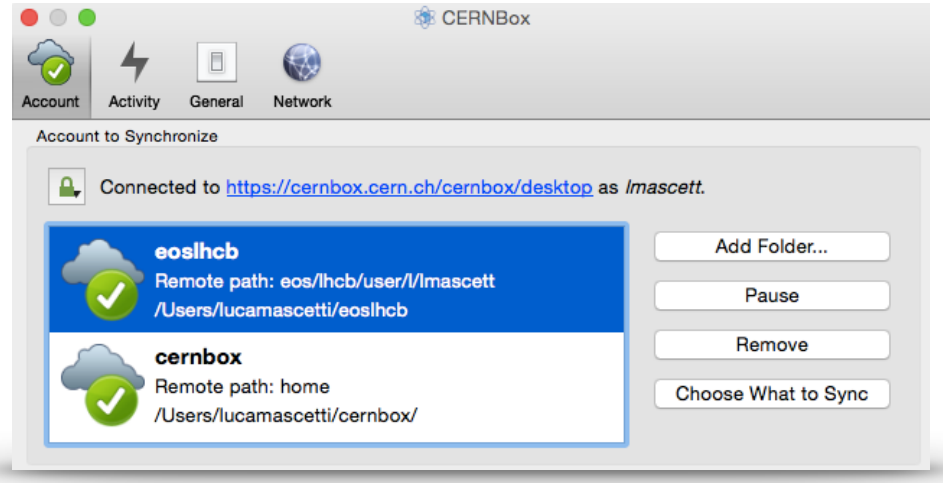
We want to do it again!

Plan to organise another workshop on the topic

Tomorrow's Features

- Direct access to EOSUSER (and not only…)
  - not only own cloud sync client
  - xroot, fuse, http/WebDAV
- Access to Physics Data
  - synchronise experiment's data
- Direct access from lxplus and batch
  - sync from your laptop and run!
  - sync results back

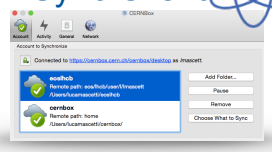
```
[lmascett@lxplus2015 ~]#
[lmascett@lxplus2015 ~]# df -H -t fuse
Filesystem      Size  Used Avail Use% Mounted on
eosuser         506T   70T  437T  14% /eos/user
eosatlas         36P   17P   20P  45% /eos/atlas
eosalice         20P   11P  8.5P  57% /eos/alice
eoscms           28P   14P   15P  49% /eos/cms
eoslhcb          13P  7.6P  4.6P  63% /eos/lhcb
eospublic        16P  5.8P   11P  36% /eos/public
[lmascett@lxplus2015 ~]#
[lmascett@lxplus2015 ~]# ls -lc /eos/user/l/lmascett/
total 6644
drwx------. 1 lmascett c3       5 Dec 10 15:58 CERN
drwx------. 1 lmascett c3       0 Jan 26 18:18 debug
drwx------. 1 lmascett c3       0 Dec 11 09:43 download
drwx------. 1 lmascett c3       0 Oct 31 18:24 pdf
drwx------. 1 lmascett c3       1 Dec 11 09:44 personal
drwx------. 1 lmascett c3       8 Dec 10 12:11 pictures
```



Future Directions

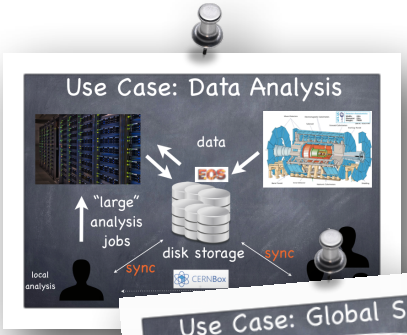
Laptops, PCs
& Mobile devices

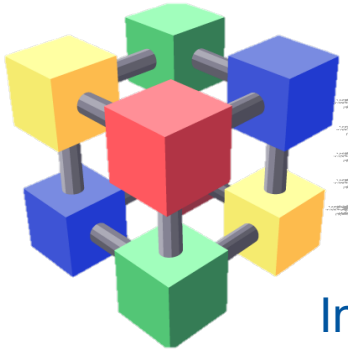
CERNBox
Sync Client



EOS (CERN)

Batch
Integration





Grid
Integration

going to “Grid Home Directory”?

www.cern.ch