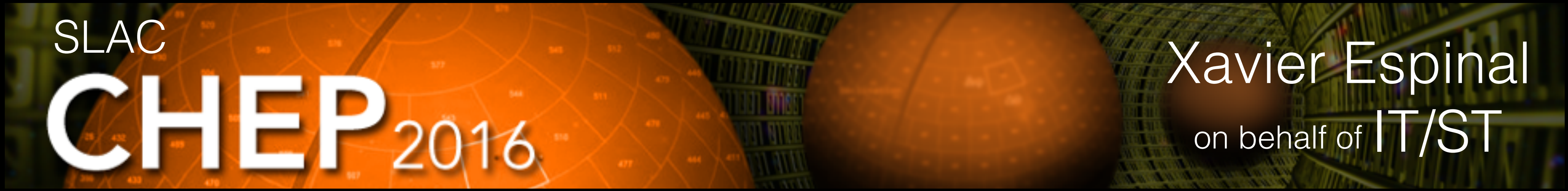
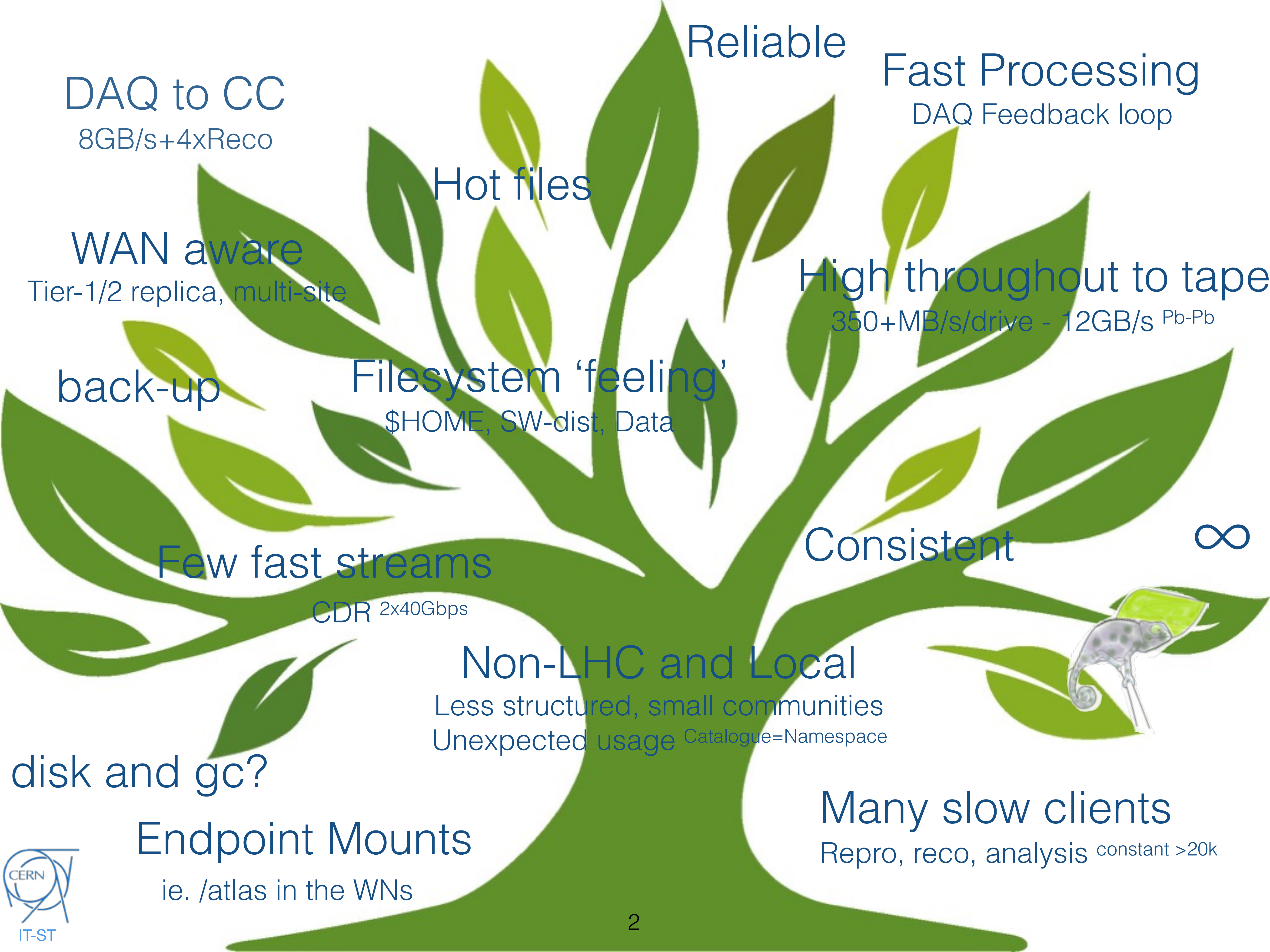
Data services for LHC computing



Xavier Espinal

on behalf of IT/ST



DAQ to CC

8GB/s+4xReco

Hot files

Reliable

Fast Processing

DAQ Feedback loop

WAN aware

Tier-1/2 replica, multi-site

High throughout to tape

350+MB/s/drive - 12GB/s Pb-Pb

back-up

Filesystem ‘feeling’

$HOME, SW-dist, Data

Consistent

∞

Few fast streams

CDR 2x40Gbps

Non-LHC and Local

Less structured, small communities

Unexpected usage Catalogue=Namespace

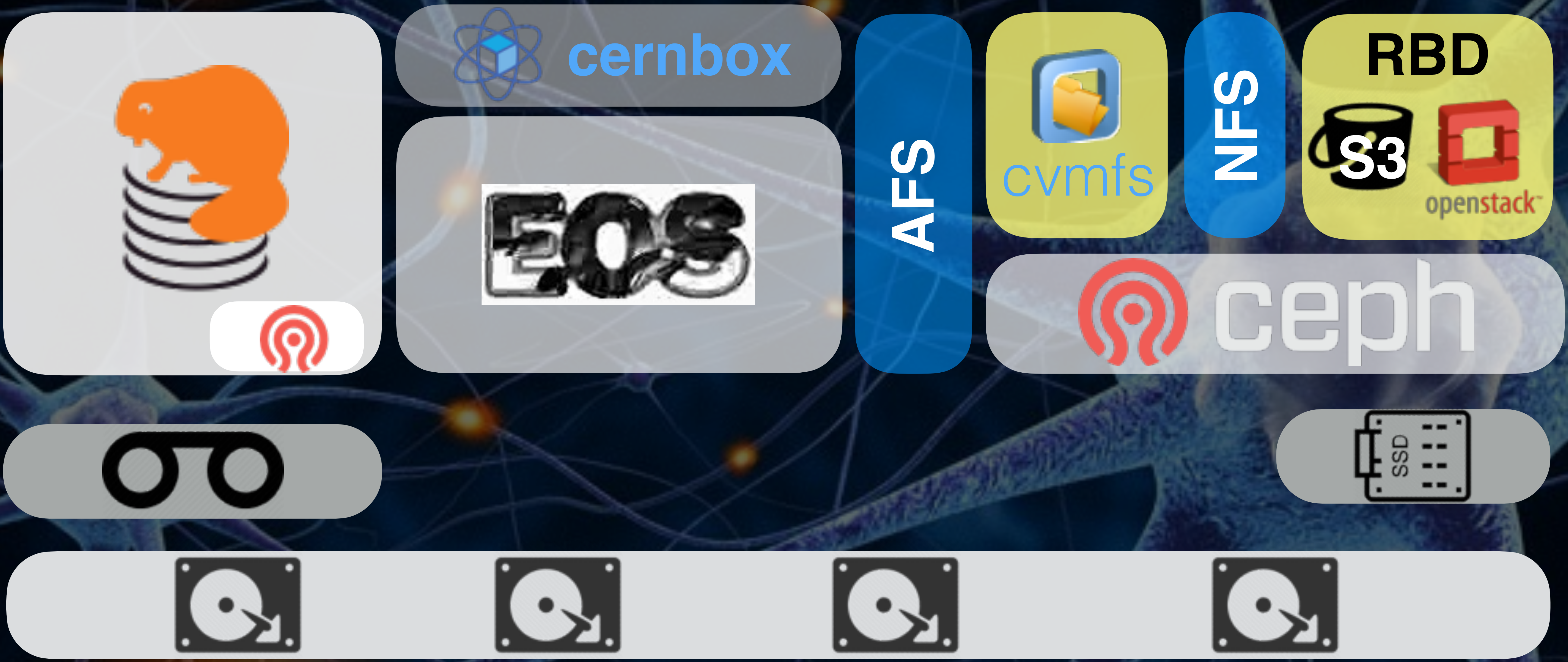
disk and gc?

Endpoint Mounts

ie. /atlas in the WNs

Many slow clients

Repro, reco, analysis constant >20k



cernbox

EOS

AFS

cvmfs

NFS

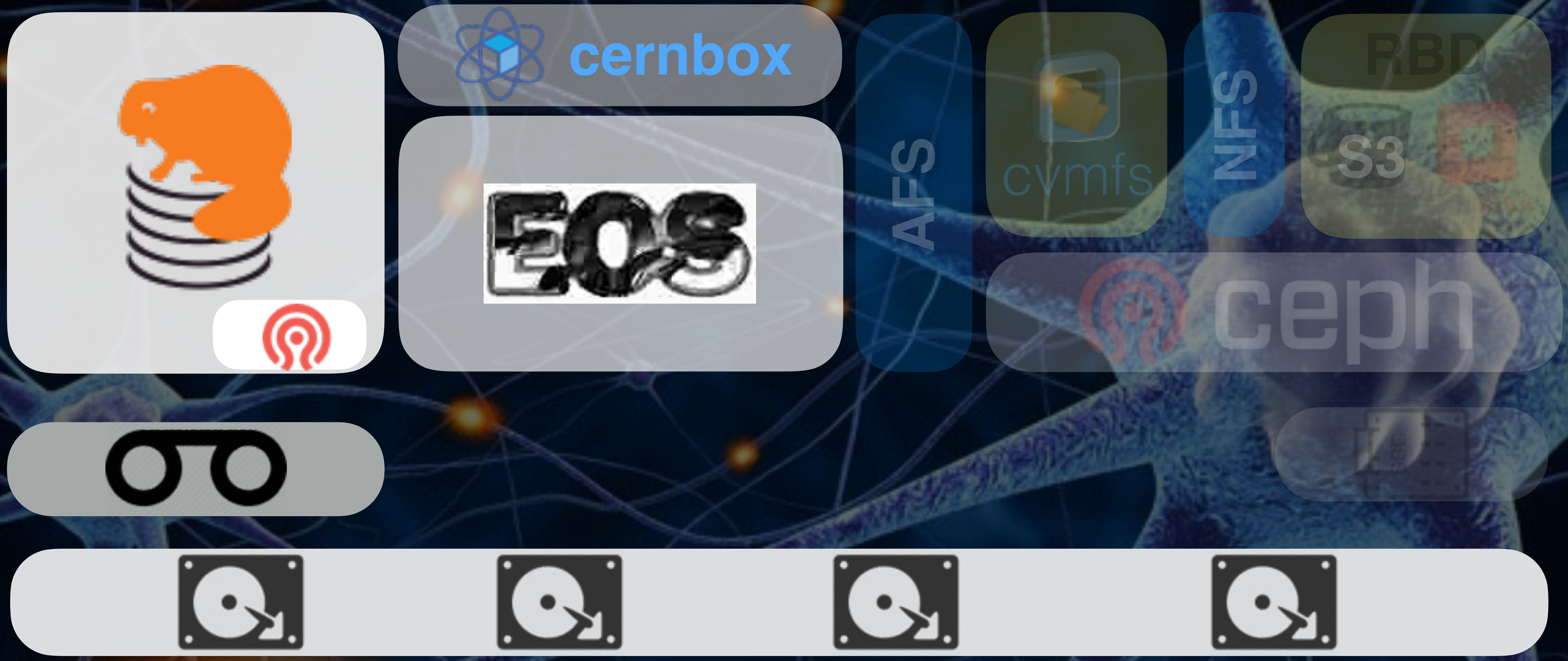
RBD

S3

openstack

ceph

SSD



cernbox

EOS

AFS

cvmfs

NFS

RBD

S3

openstack

ceph

SSD

CASTOR
CERN Advanced STORage manager

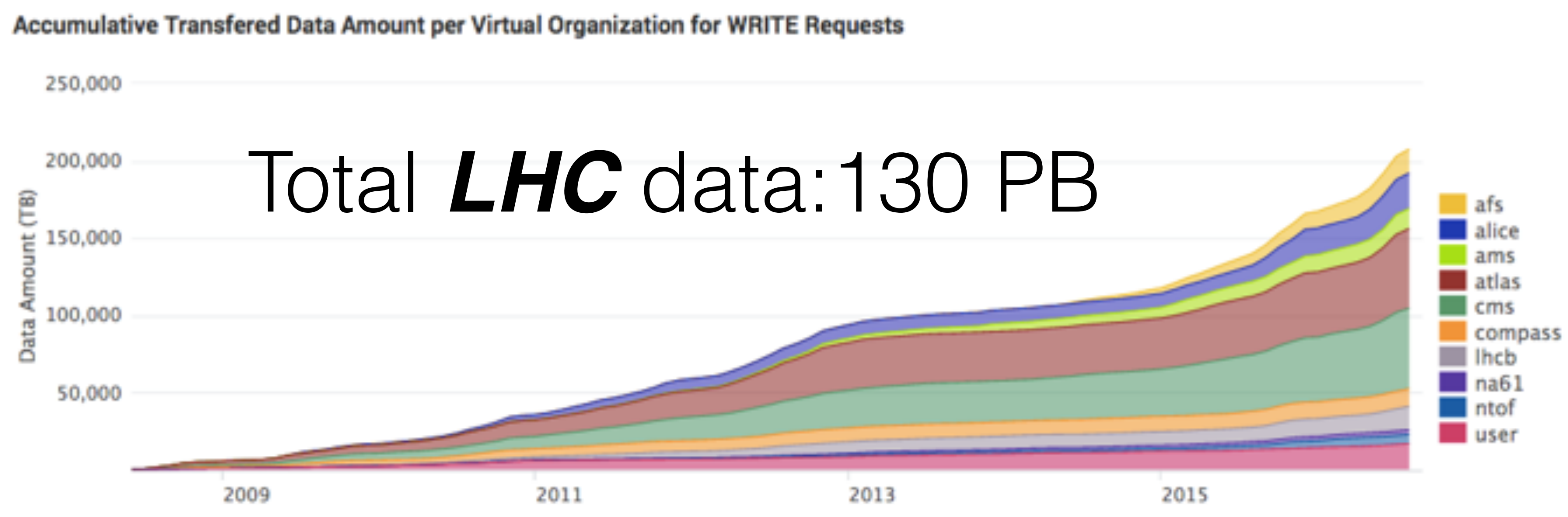
Evolved to
Tape oriented system
Key feature
Per stream speed

Biggest physics-repo worldwide 175PB and +500M files

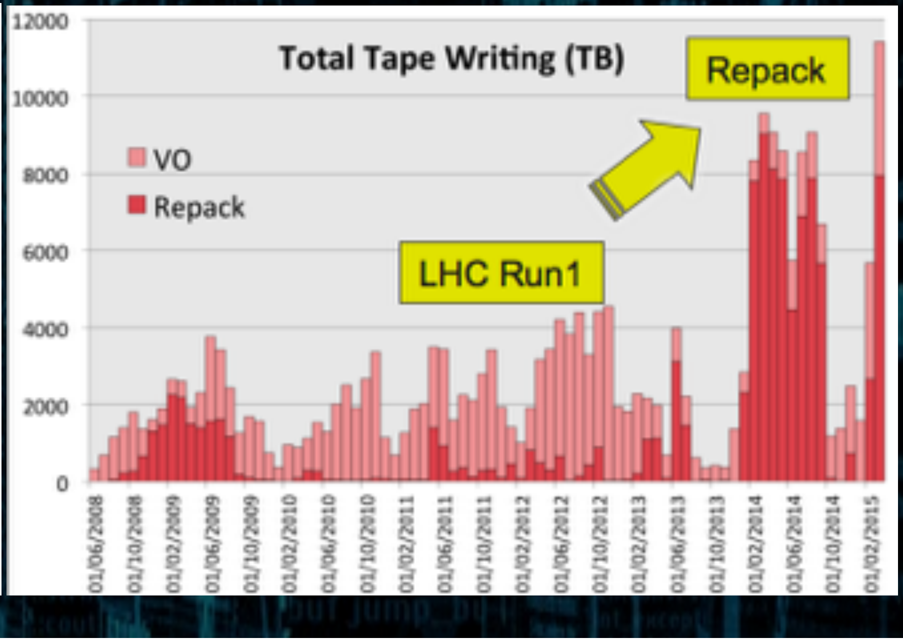
Towards a pluggable tape backend (EOS)

Cold by definition: hight throughput, high latency

Tape best technology for data repositories: TCO media power density and resilient/reliable very large disk caches nowadays

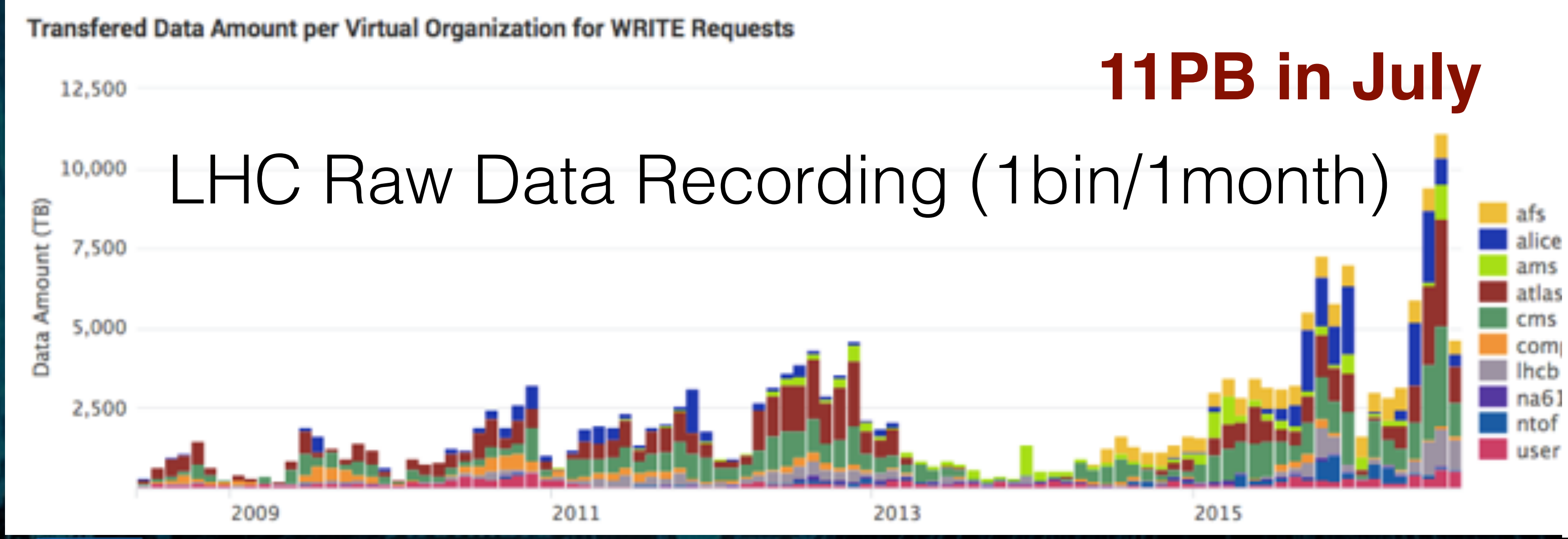

Accumulative Transfered Data Amount per Virtual Organization for WRITE Requests

Total ***LHC*** data:130 PB


Total Tape Writing (TB)

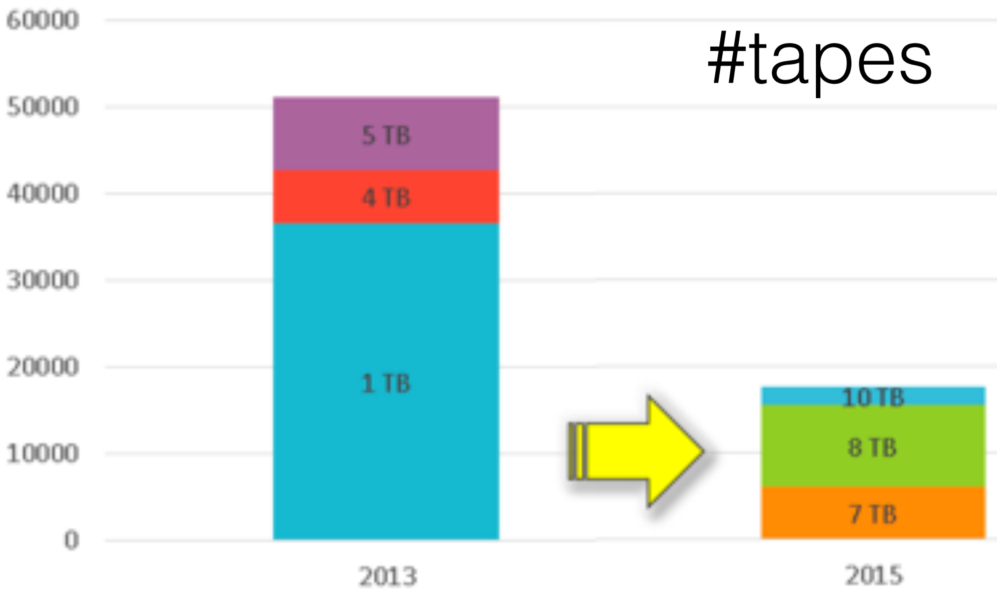
Repack

LHC Run1


Transfered Data Amount per Virtual Organization for WRITE Requests

11PB in July

LHC Raw Data Recording (1bin/1month)


#tapes

IT-ST

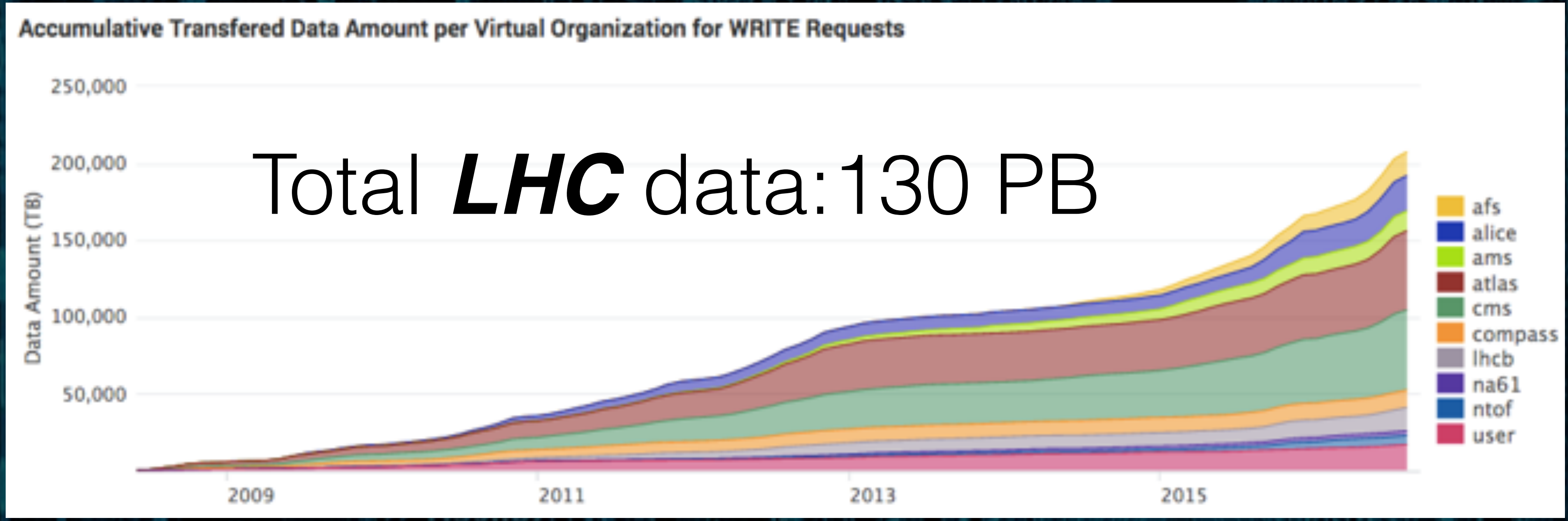
CERN Tape Archive

Biggest physics-repo worldwide 175PB and +500M files

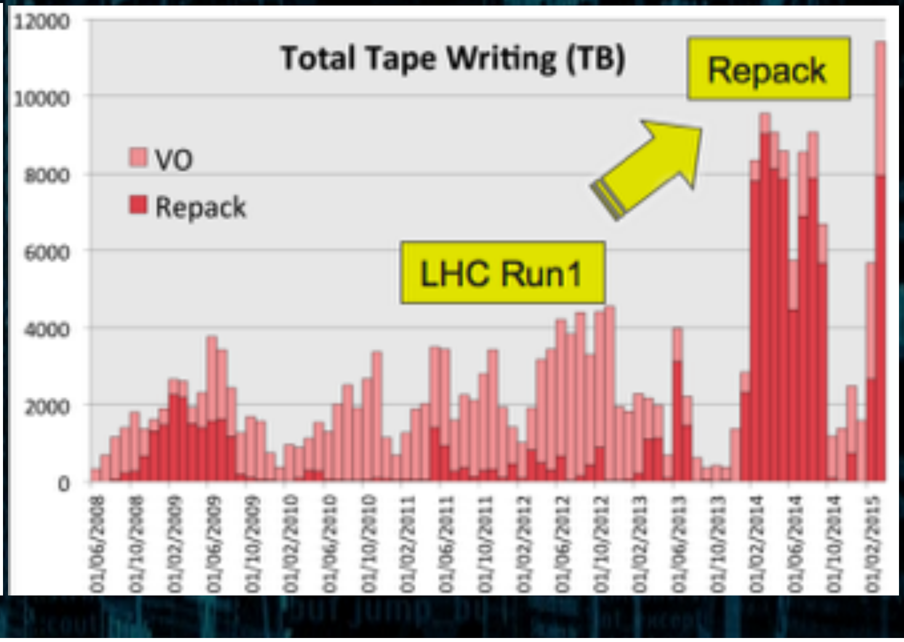
Towards a pluggable tape backend (EOS)

Cold by definition: hight throughput, high latency

Tape best technology for data repositories: TCO
media
power
density
and resilient/reliable
very large disk caches nowadays

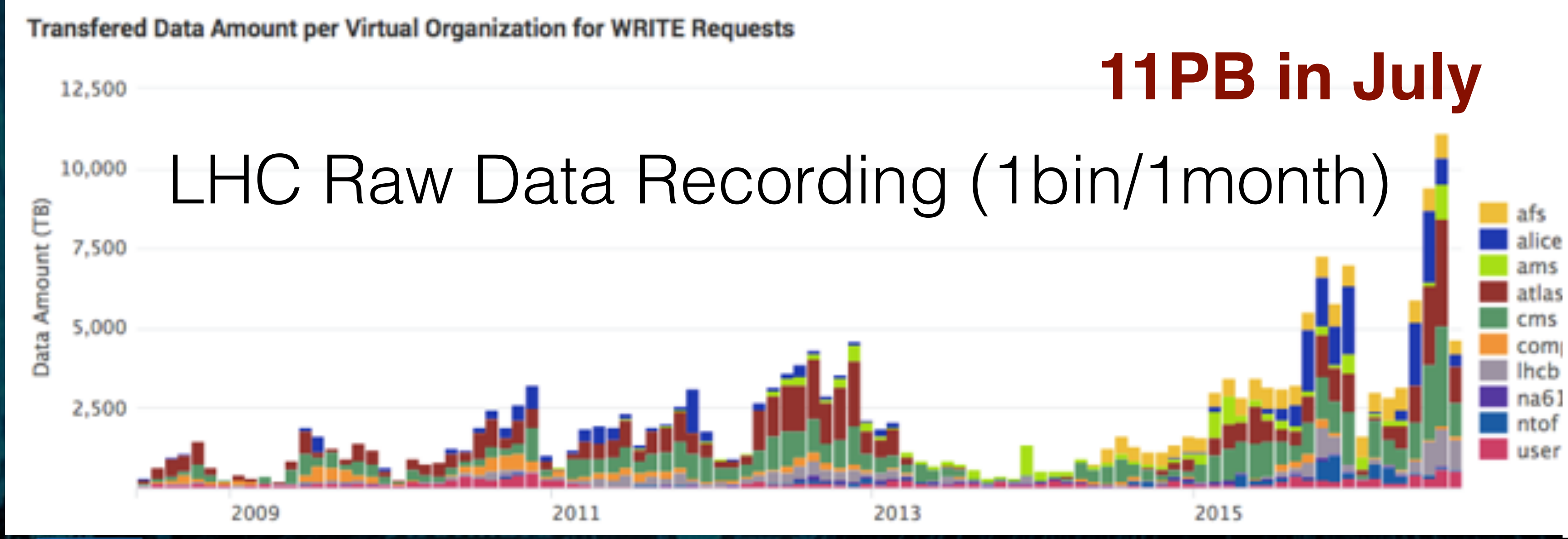


Total ***LHC*** data:130 PB



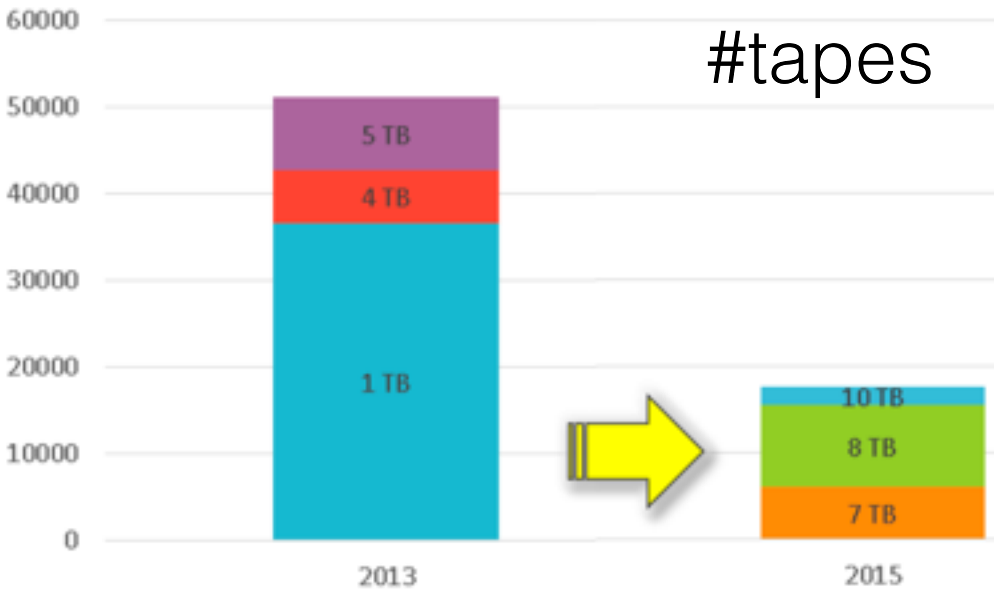
Repack

LHC Run1



11PB in July

LHC Raw Data Recording (1bin/1month)



#tapes

IT-ST

EOS now

October 2016

+1200

+45000

850M files

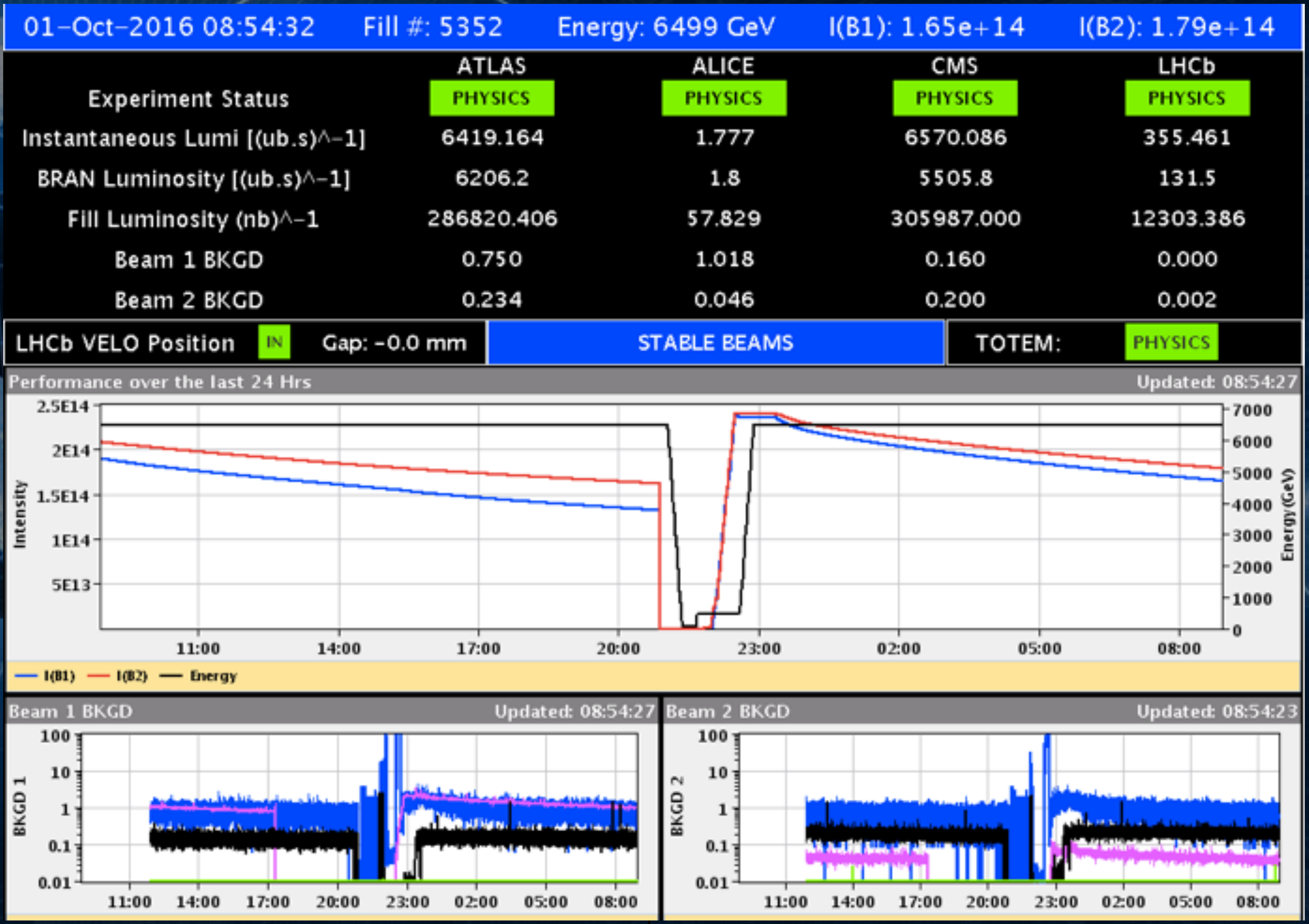
150PB

EB era

Easily scalable (#disk #servers)

Performant and manageable

LHC Main storage platform



01-Oct-2016 08:54:32 Fill #: 5352 Energy: 6499 GeV I(B1): 1.65e+14 I(B2): 1.79e+14

	ATLAS	ALICE	CMS	LHCb
Experiment Status	PHYSICS	PHYSICS	PHYSICS	PHYSICS
Instantaneous Lumi [(ub.s)^-1]	6419.164	1.777	6570.086	355.461
BRAN Luminosity [(ub.s)^-1]	6206.2	1.8	5505.8	131.5
Fill Luminosity (nb)^-1	286820.406	57.829	305987.000	12303.386
Beam 1 BKGD	0.750	1.018	0.160	0.000
Beam 2 BKGD	0.234	0.046	0.200	0.002

LHCb VELO Position IN Gap: -0.0 mm STABLE BEAMS TOTEM: PHYSICS

Performance over the last 24 Hrs Updated: 08:54:27

Beam 1 BKGD Updated: 08:54:27

Beam 2 BKGD Updated: 08:54:23

IT-ST

EOS now

October 2016

+1200 🖥

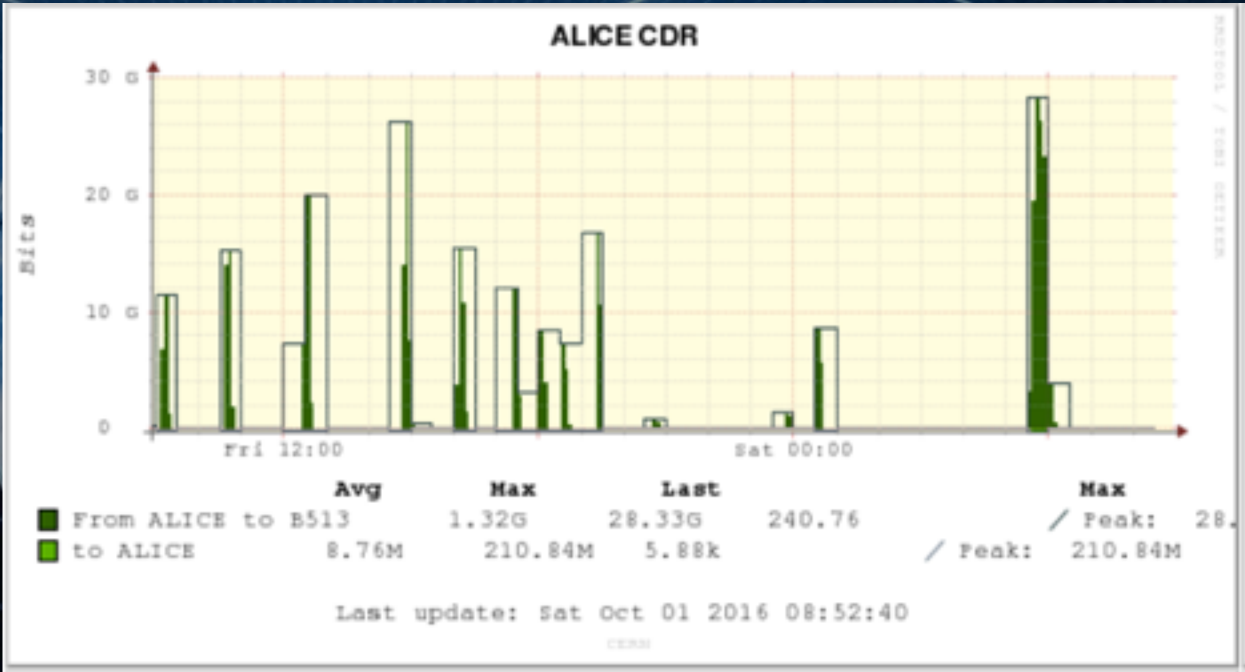
+45000 🖴

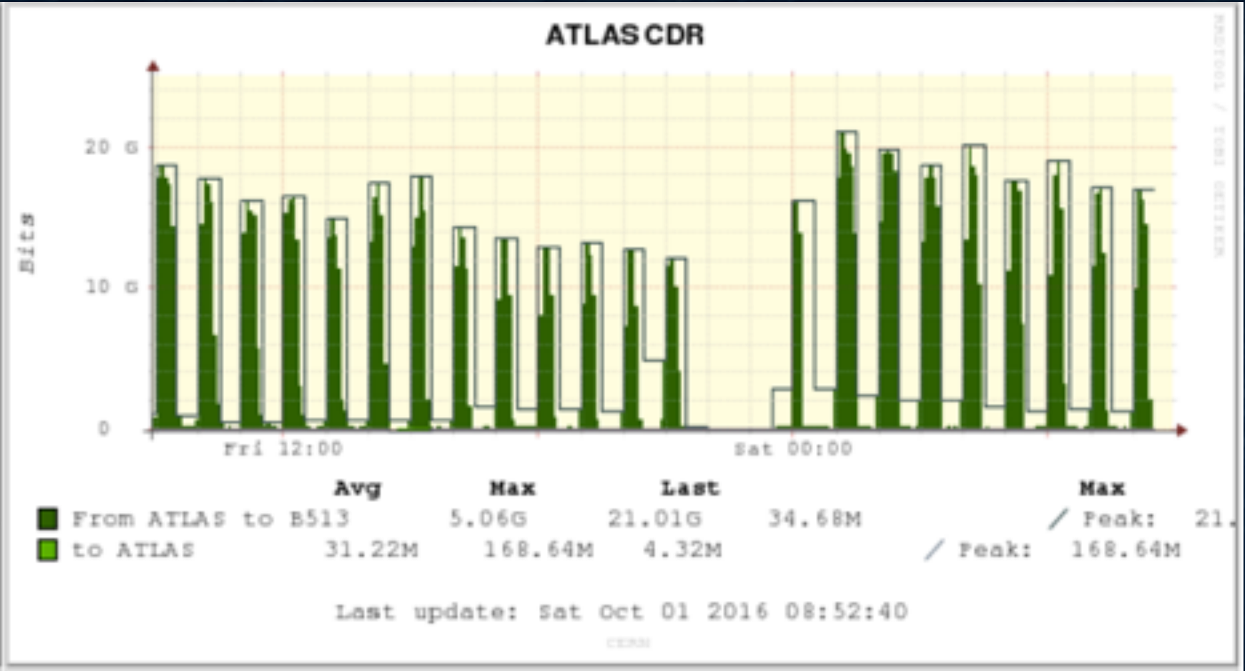
850M files

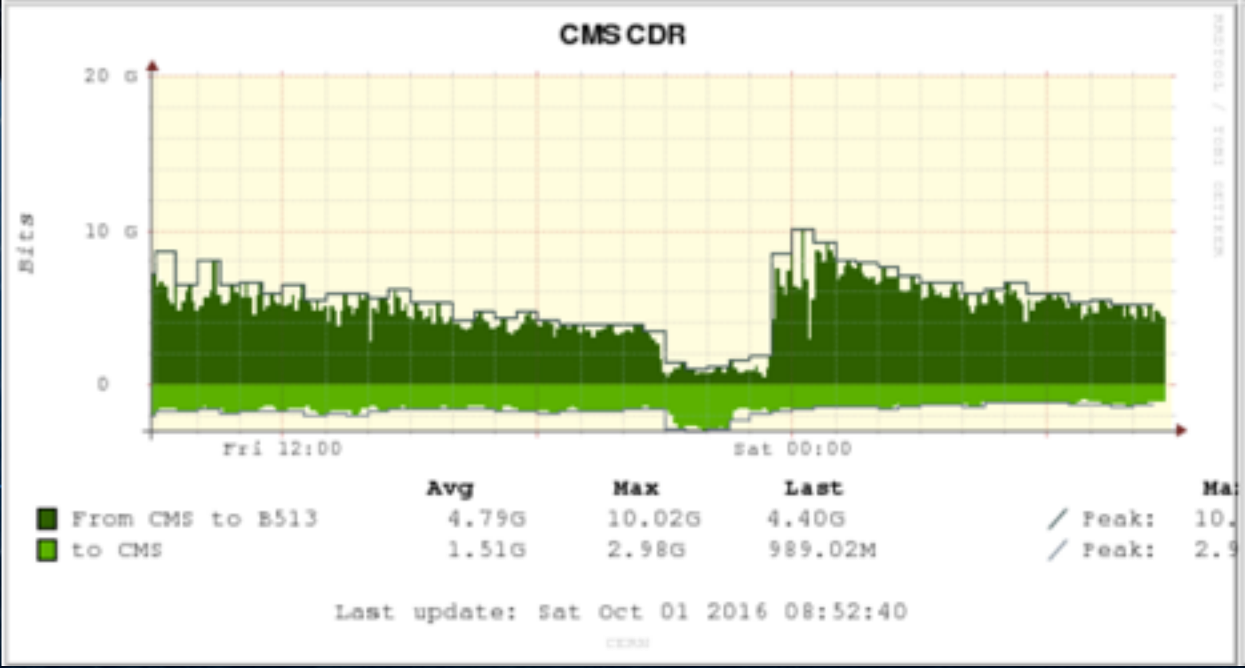
150PB 🖴

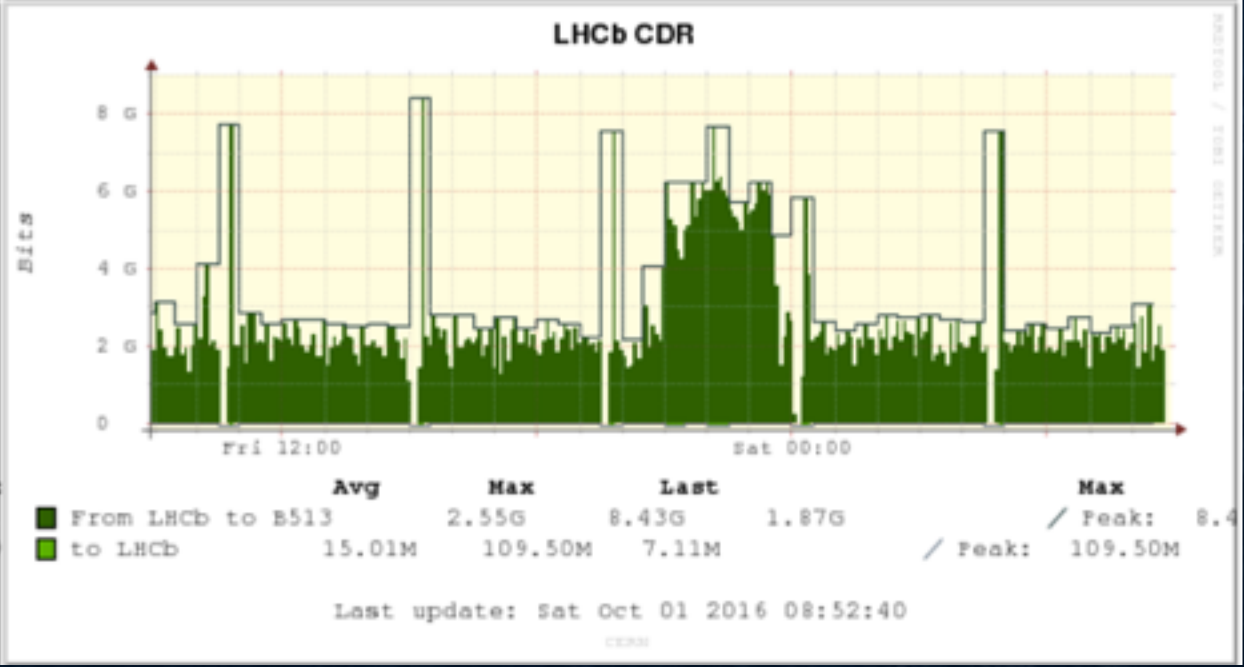
EB era

- Easily scalable (#disk #servers)
- Performant and manageable
- LHC Main storage platform









EOS now

October 2016

+1200

+45000

850M files

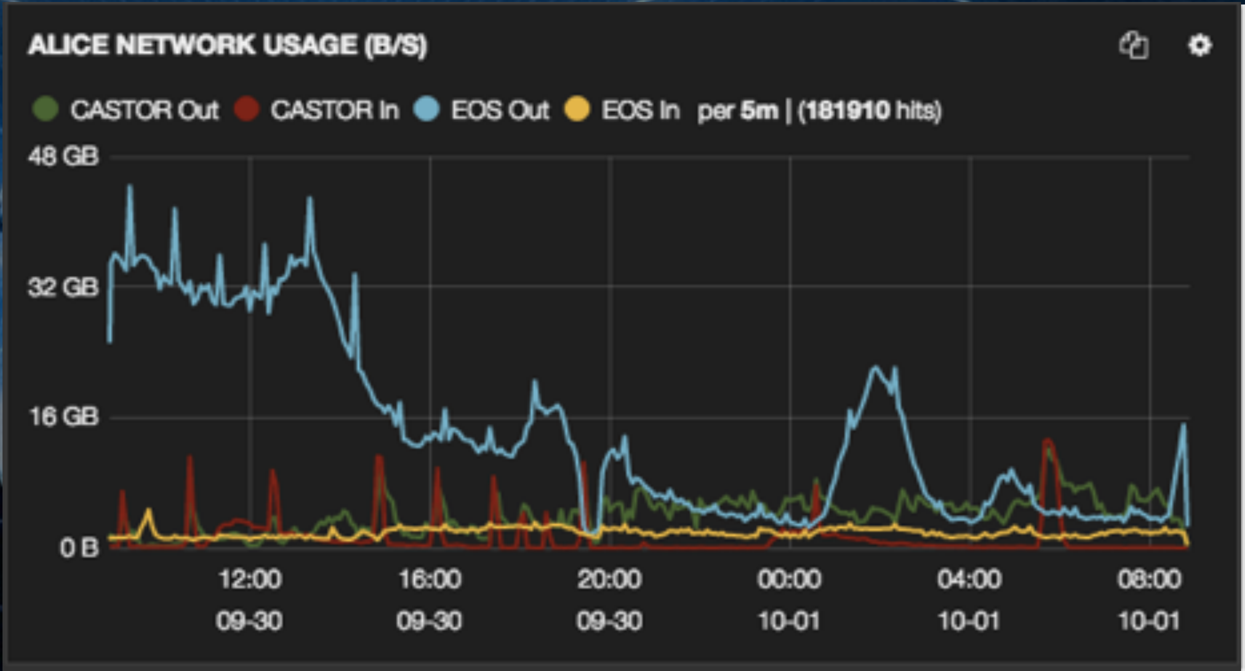
150PB

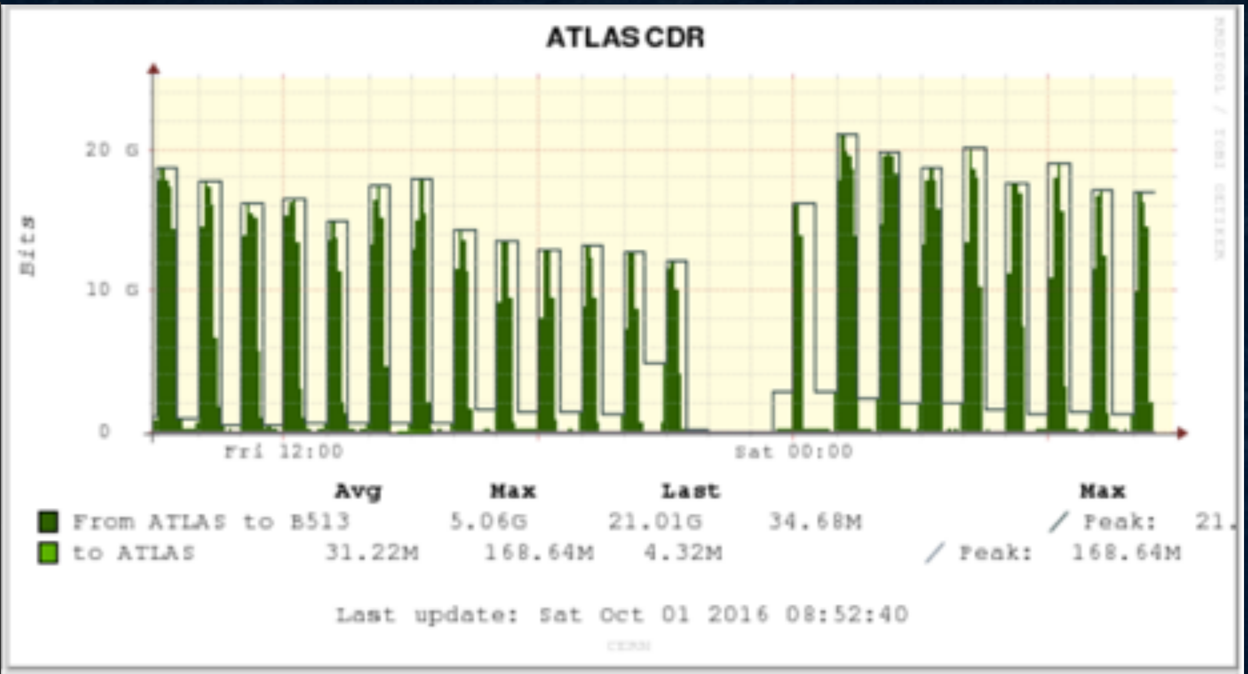
EB era

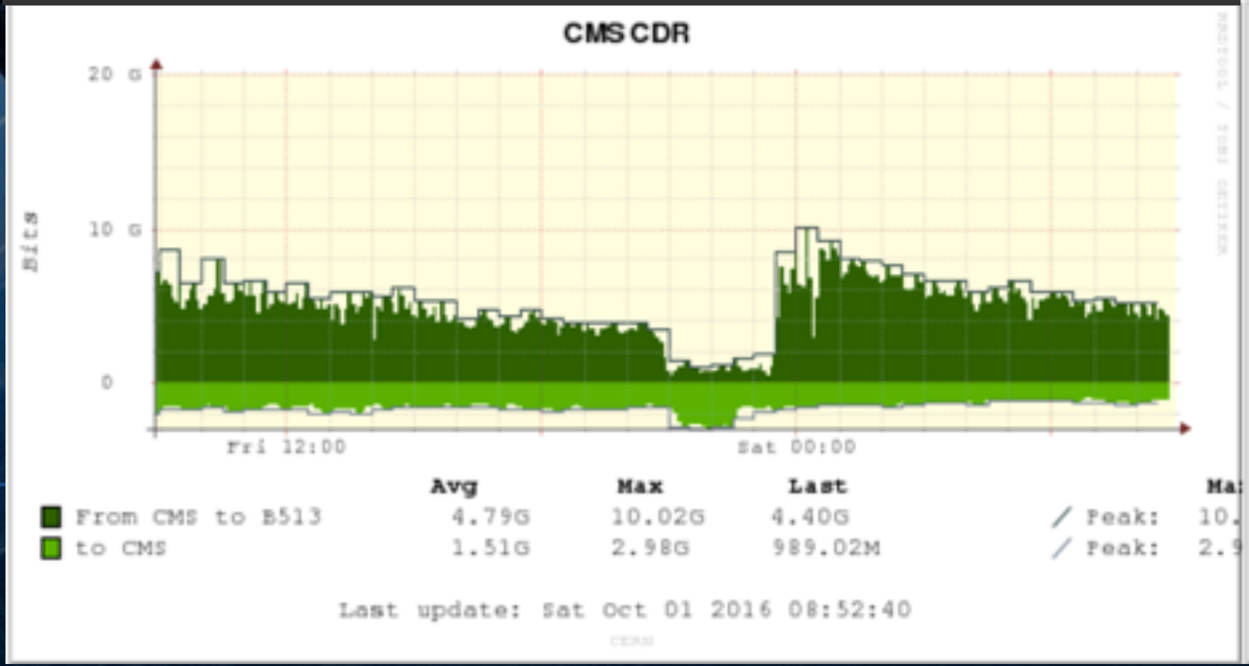
Easily scalable (#disk #servers)

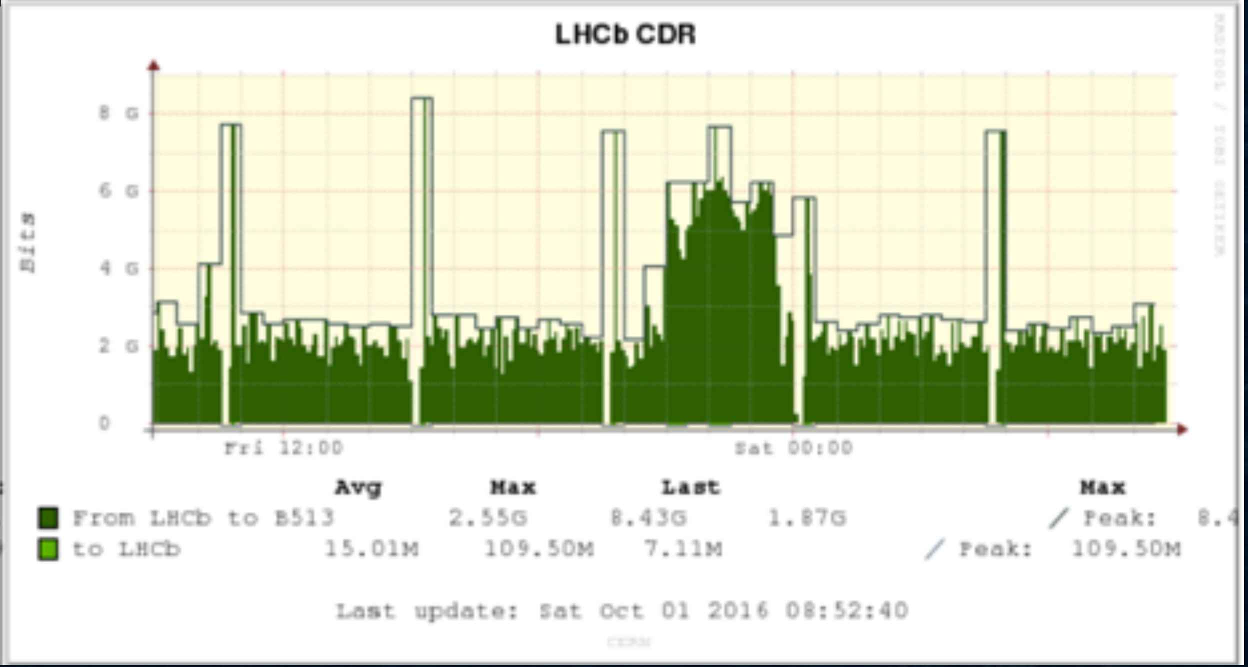
Performant and manageable

LHC Main storage platform









EOS now

October 2016

+1200

+45000

850M files

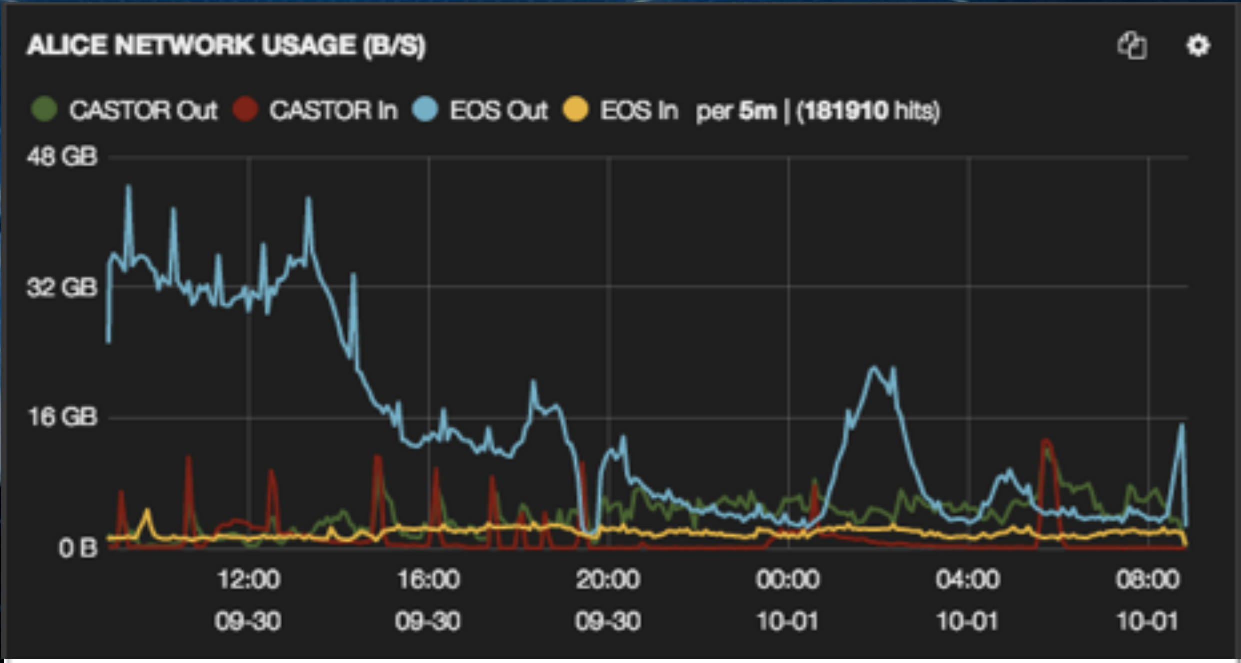
150PB

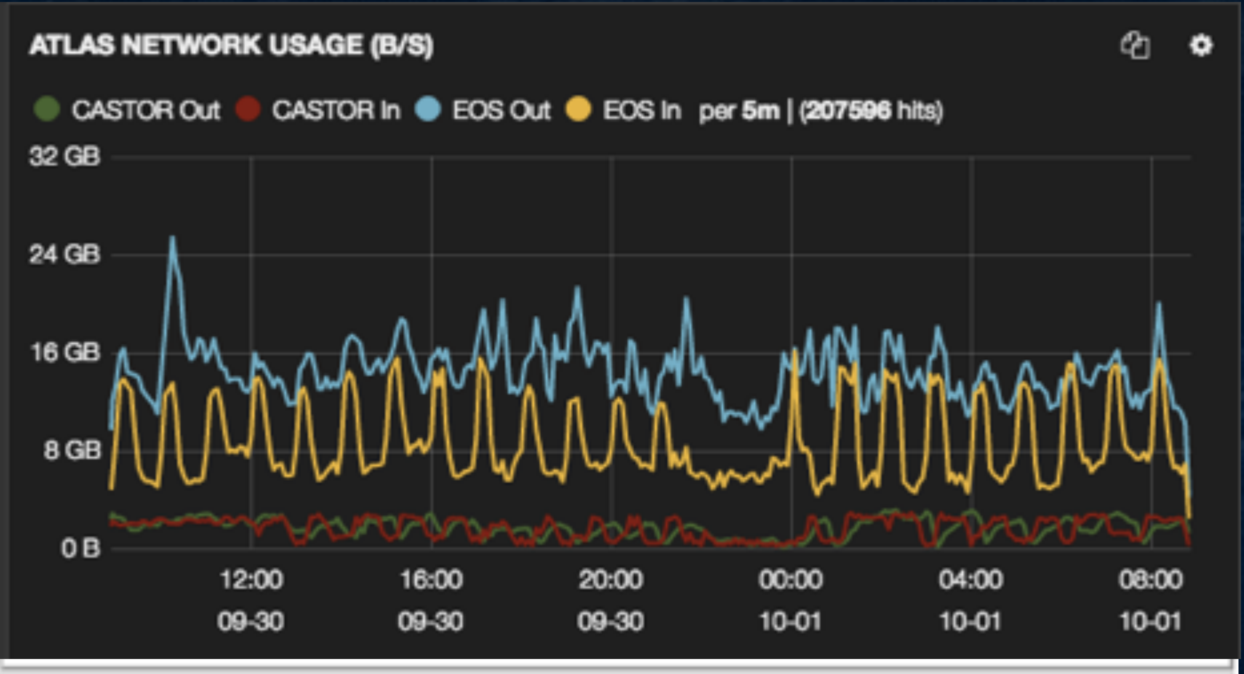
EB era

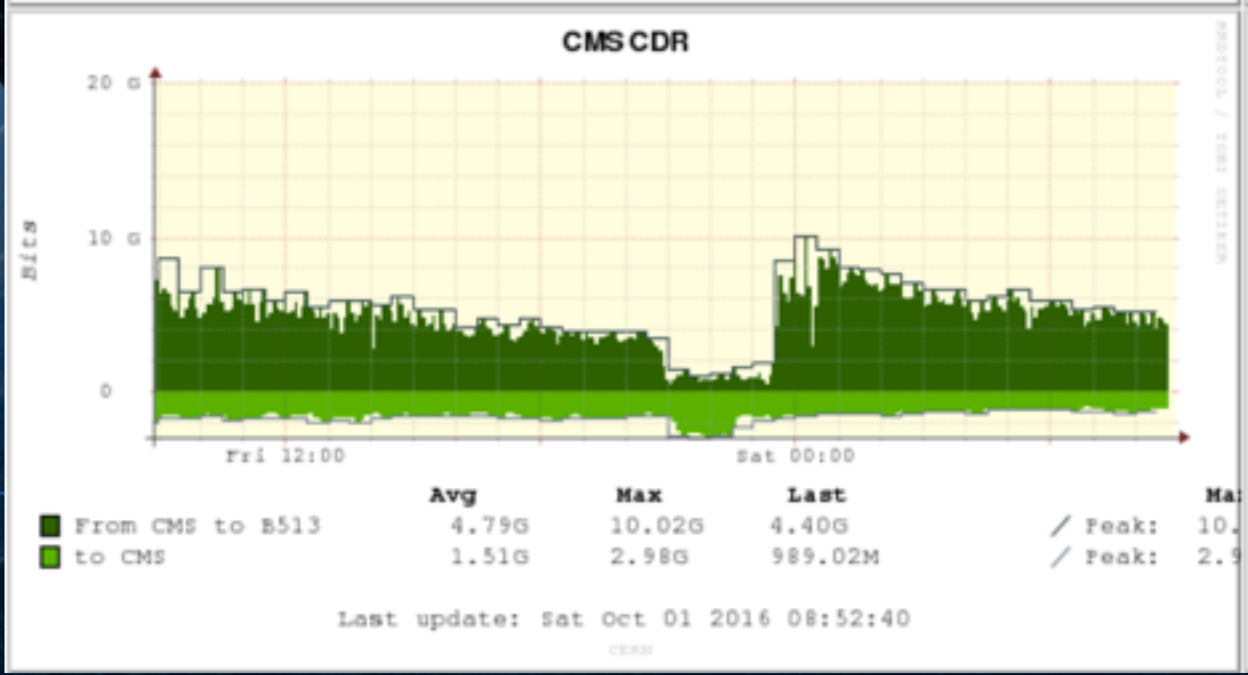
Easily scalable (#disk #servers)

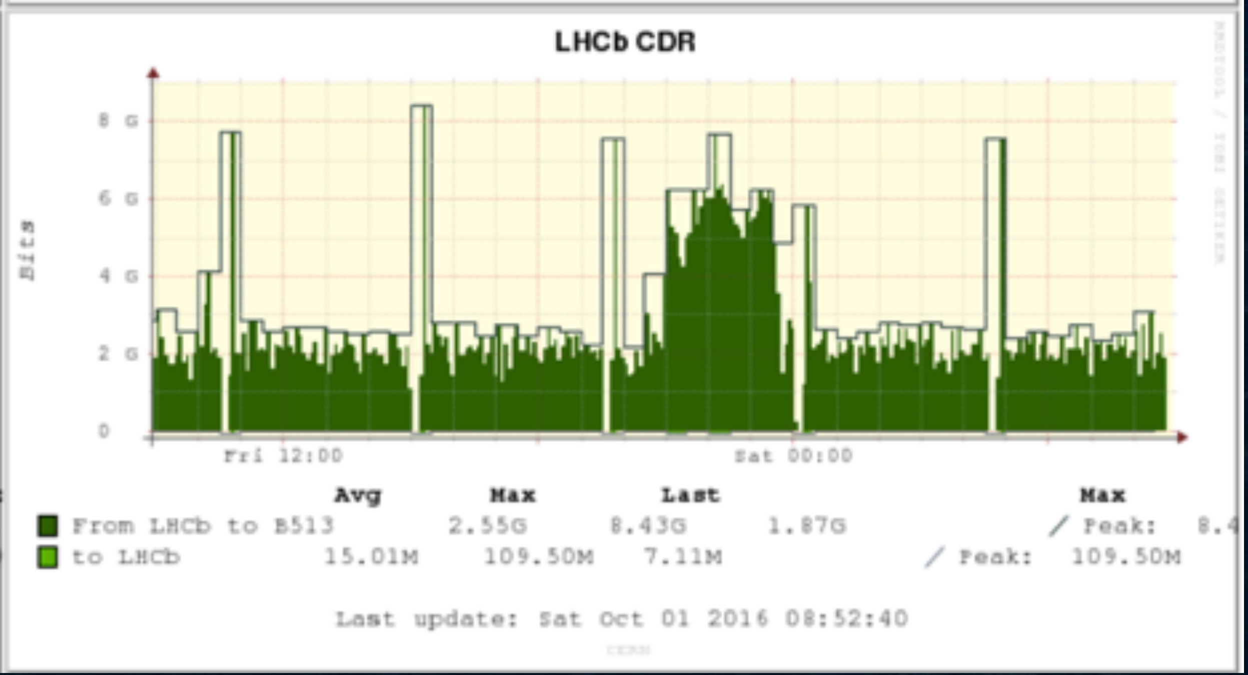
Performant and manageable

LHC Main storage platform









EOS now

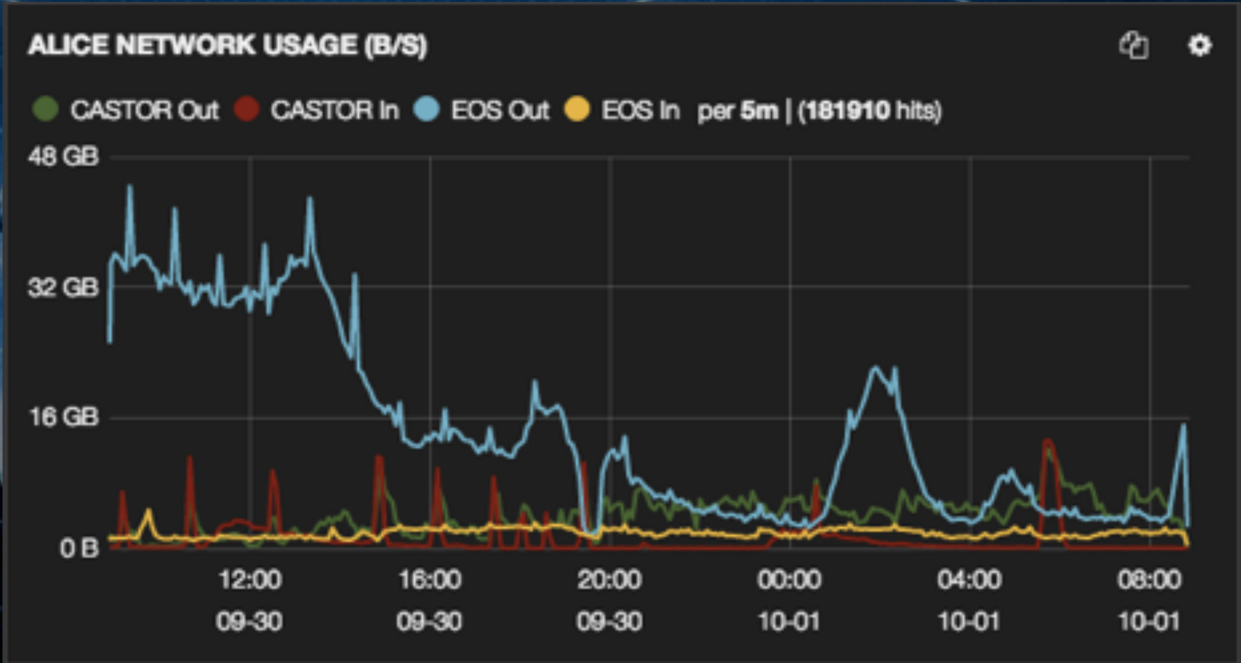
October 2016

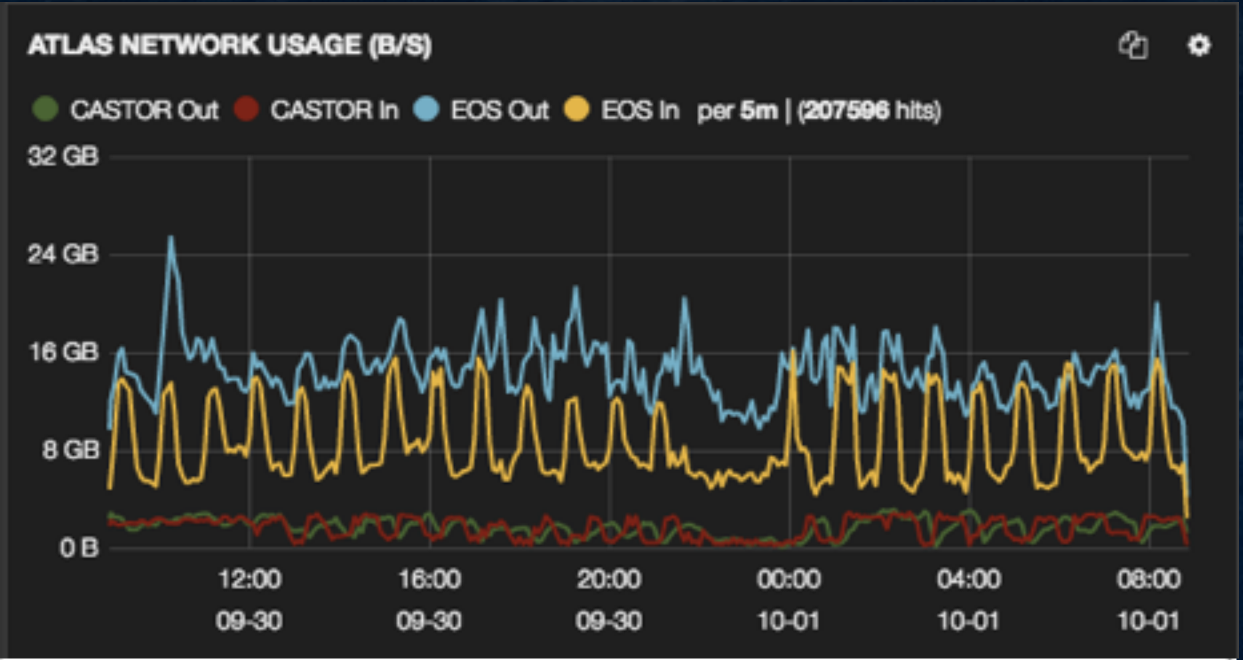
+1200
+45000

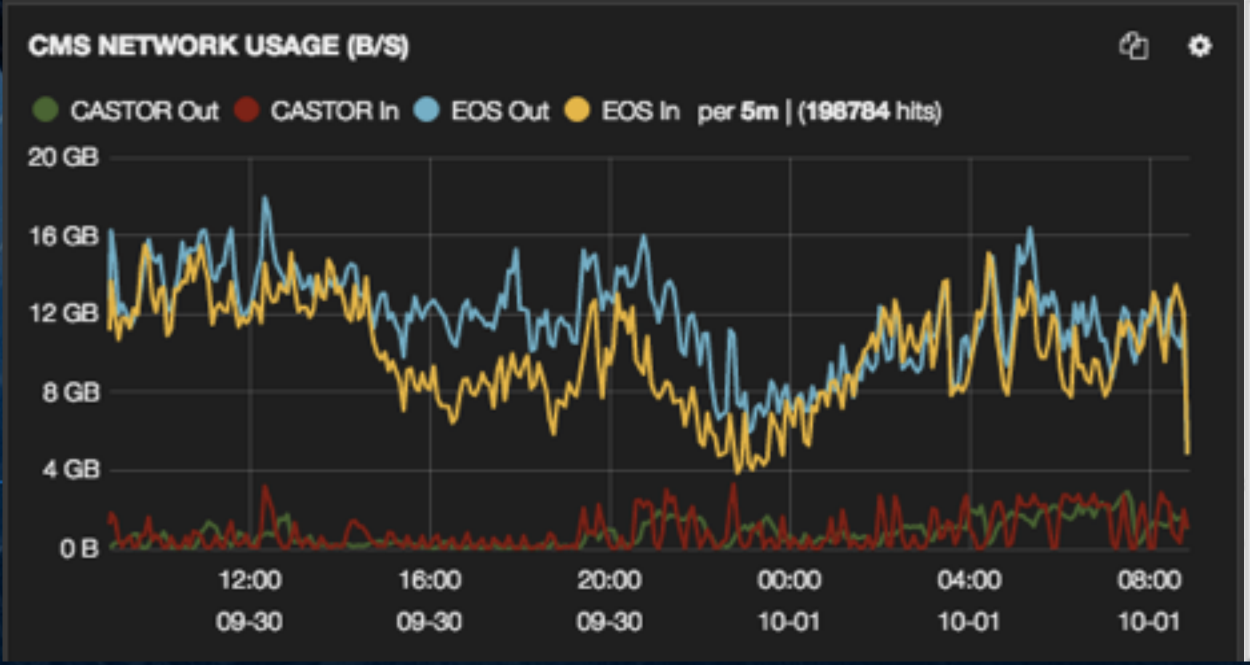
850M files
150PB

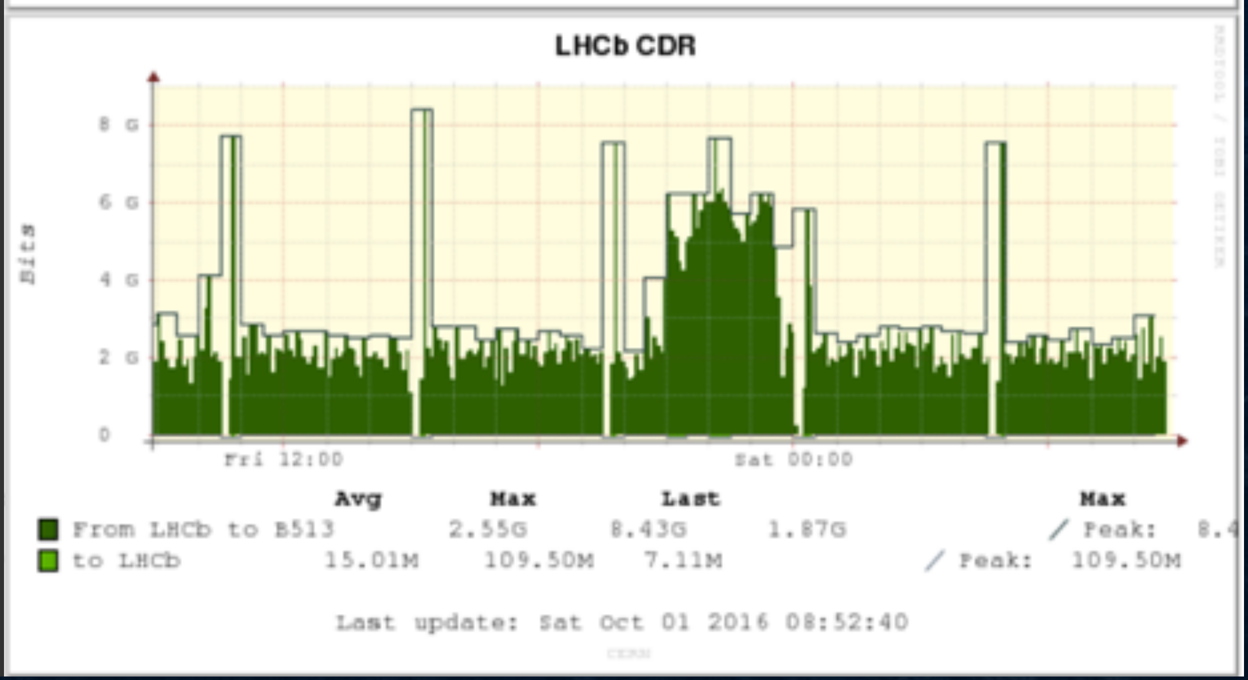
EB era

- Easily scalable (#disk #servers)
- Performant and manageable
- LHC Main storage platform









IT-ST

EOS now

October 2016

+1200 🖥

+45000 🖴

850M files

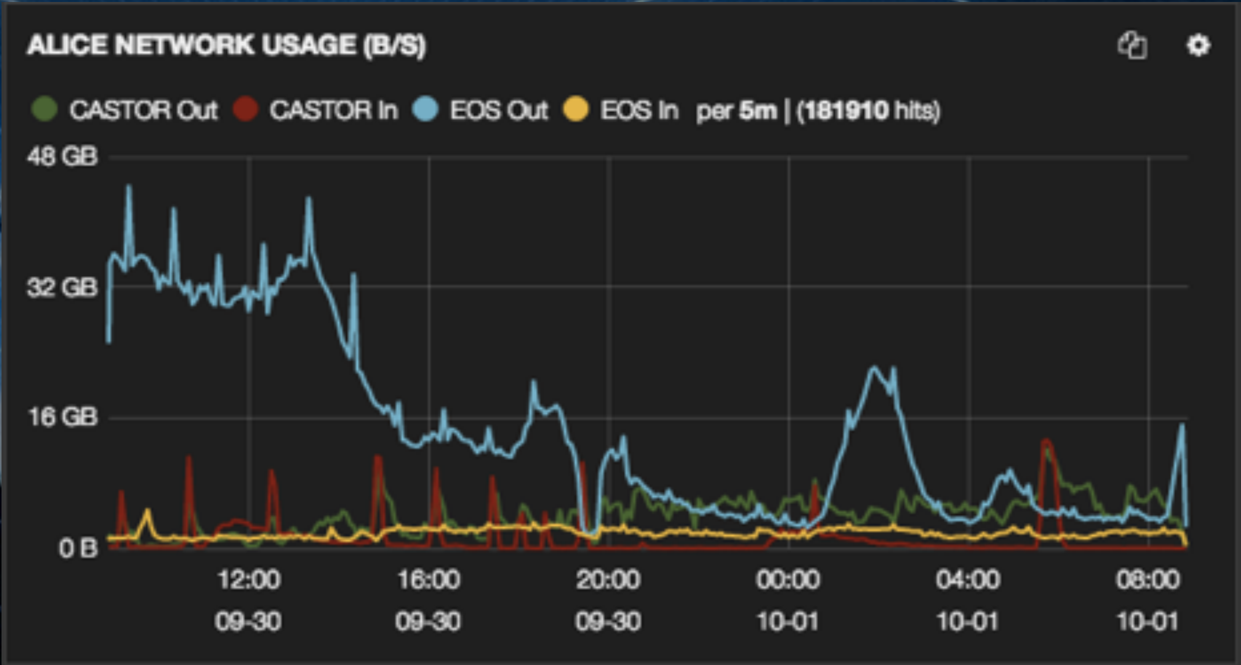
150PB 🖴

EB era

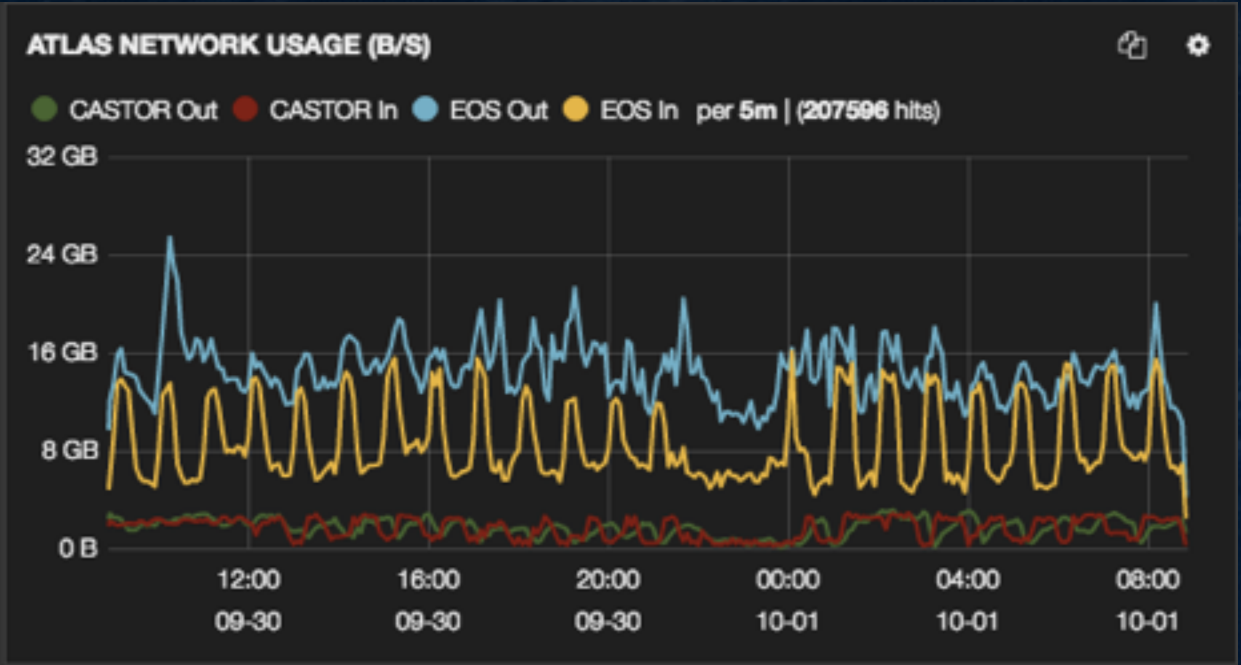
Easily scalable (#disk #servers)

Performant and manageable

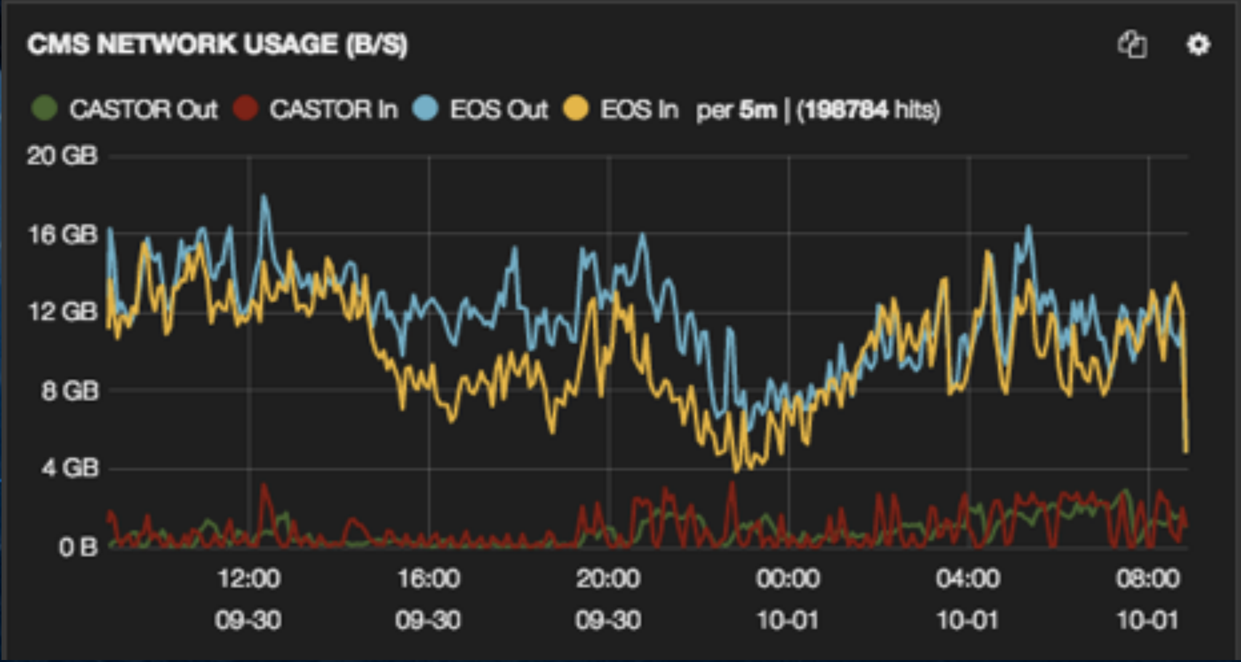
LHC Main storage platform



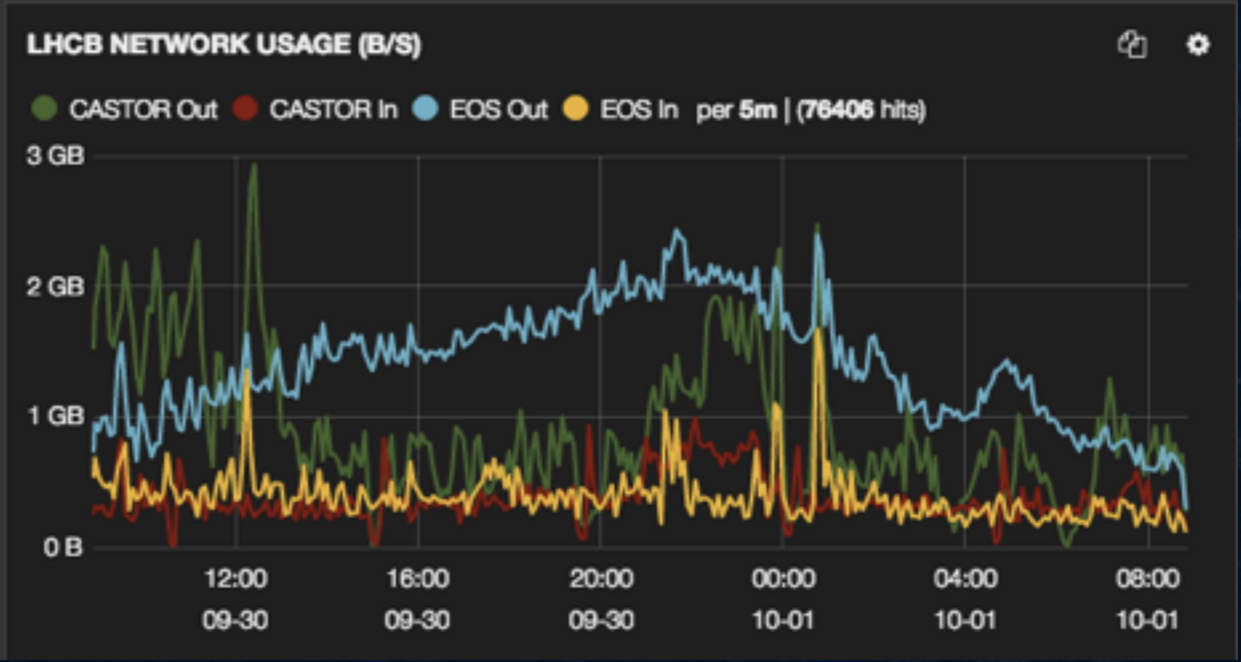
ALICE NETWORK USAGE (B/S)



ATLAS NETWORK USAGE (B/S)



CMS NETWORK USAGE (B/S)



LHCB NETWORK USAGE (B/S)

CERN made for LHC experiments needs, but…

Adaptable
Catering with different uses

Data processing

User Analysis

LHC Data Recording

Sync&Share

Community storage

CERNBOX

Collaborate

Share

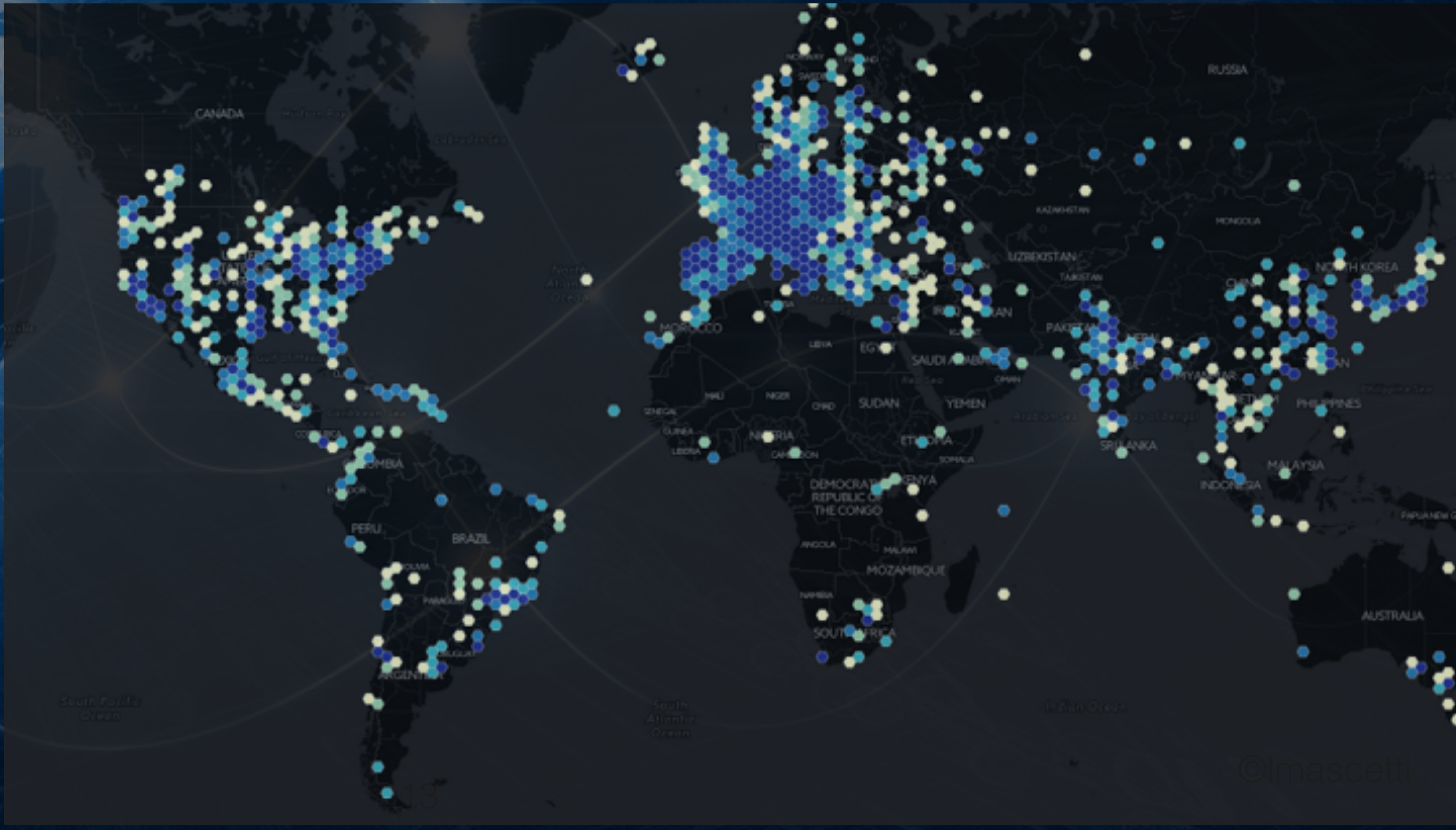
Offline work

Sync

SWAN (Jupyter)

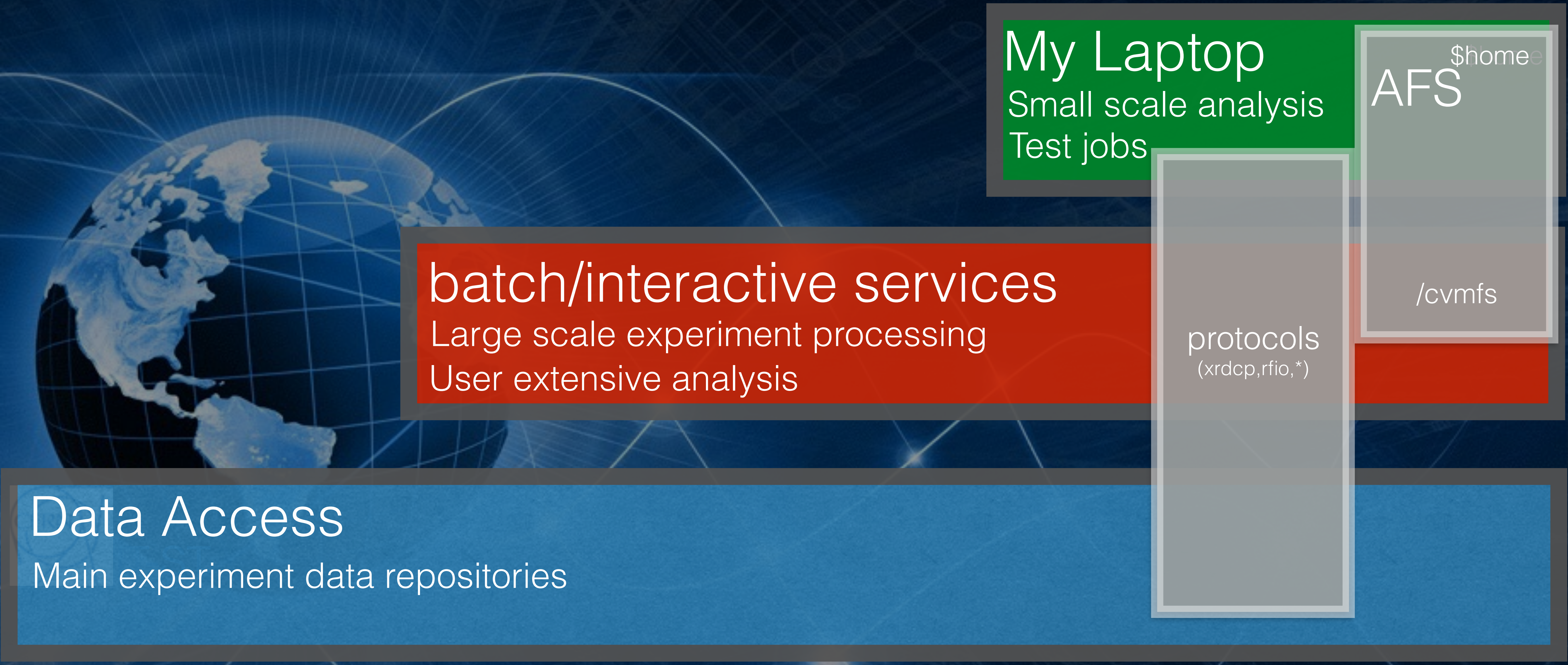
CERNBox

Users	6000 (Δ^{7d}=60)
#files	110M (Δ^{7d}=250K)
#dirs	14M
Quota	1TB/user
Used Space	240TB
Deployed Space	1.5PB



Goals

- Make data access easy
- Make analysis simple
- Facilitate Science



My Laptop

Small scale analysis
Test jobs

AFS

$home

/cvmfs

batch/interactive services

Large scale experiment processing
User extensive analysis

protocols
(xrdcp,rfio,*)

Data Access

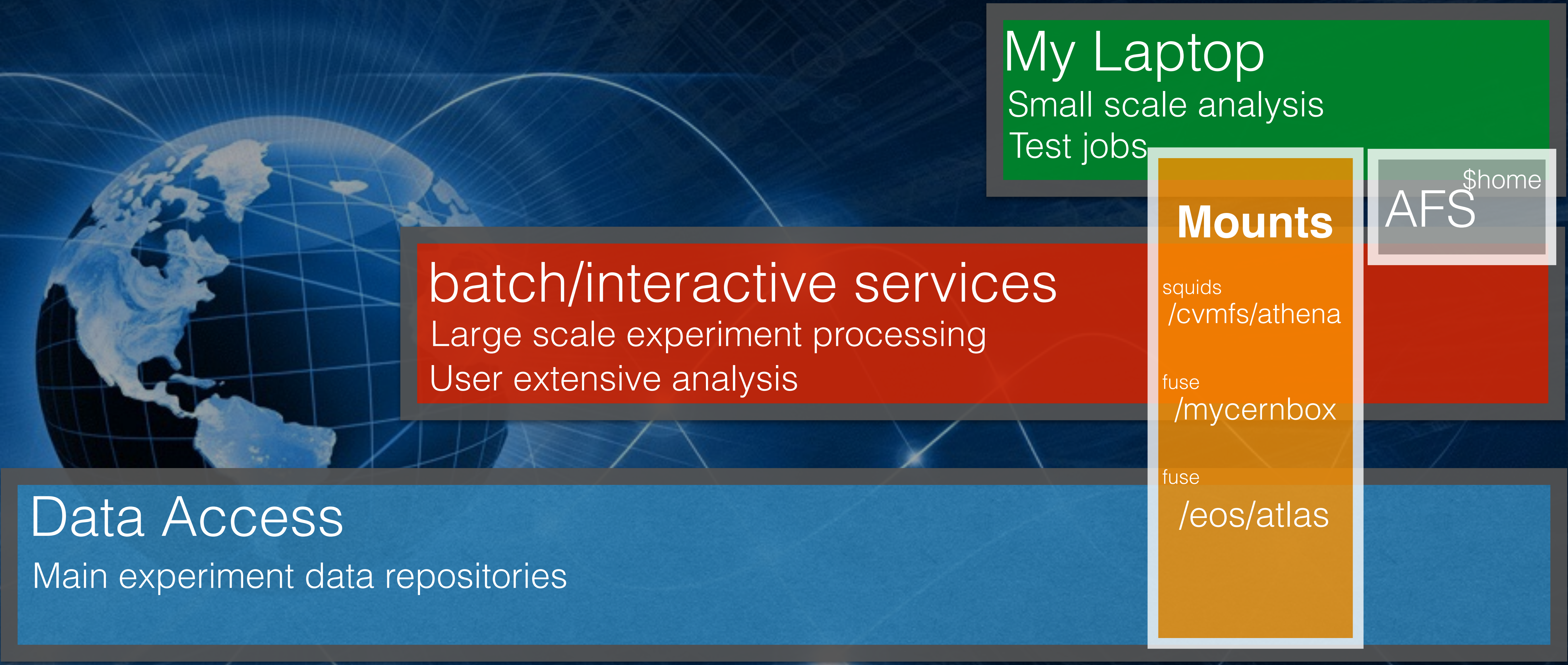
Main experiment data repositories

Goals

Make data access easy

Make analysis simple

Facilitate Science



My Laptop

Small scale analysis

Test jobs

Mounts

squids

/cvmfs/athena

fuse

/mycernbox

fuse

/eos/atlas

$home

AFS

batch/interactive services

Large scale experiment processing

User extensive analysis

Data Access

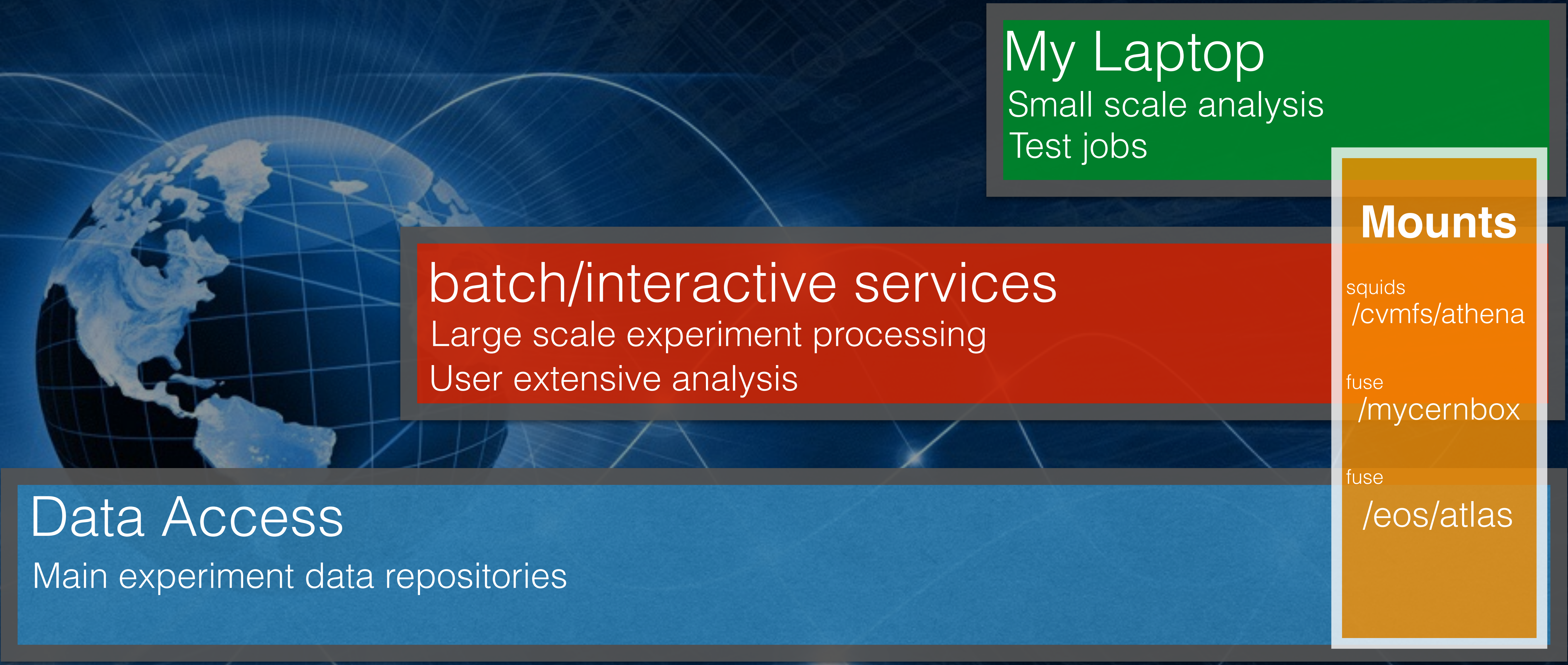
Main experiment data repositories

Goals

Make data access easy

Make analysis simple

Facilitate Science



My Laptop

Small scale analysis

Test jobs

batch/interactive services

Large scale experiment processing

User extensive analysis

Data Access

Main experiment data repositories

Mounts

squids /cvmfs/athena

fuse /mycernbox

fuse /eos/atlas

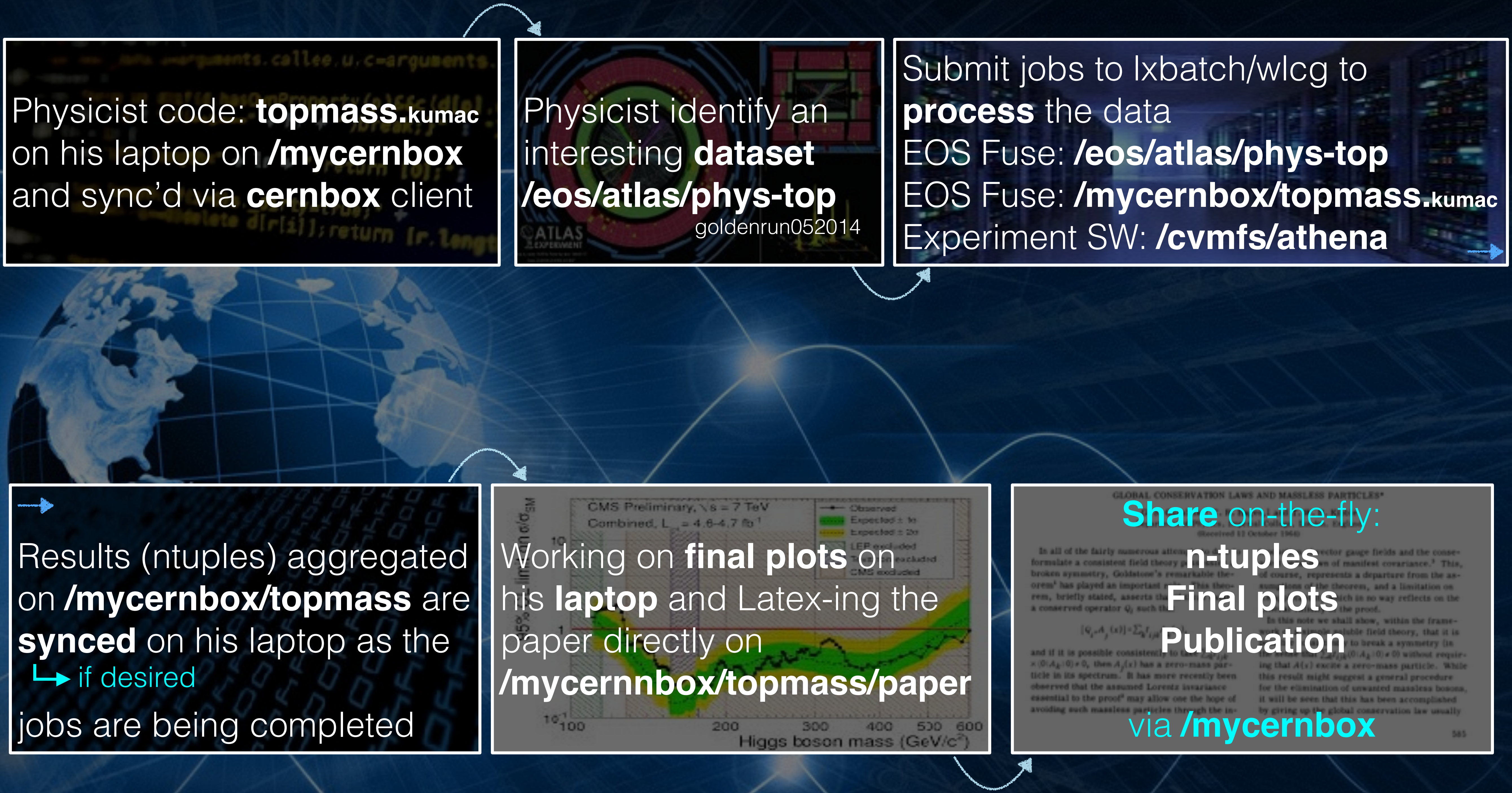
EOS CERNBOX does *“your files”* /cernbox/jdoe
EOS ”*experiment*” does *“big data”* /eos/lhcb
Different QoS, different patterns, overlaps

Goals

Make data access easy

Make analysis simple

Facilitate Science



Physicist code: **topmass.kumac** on his laptop on **/mycernbox** and sync'd via **cernbox** client

Physicist identify an interesting **dataset** **/eos/atlas/phys-top**

goldenrun052014

Submit jobs to lxbatch/wlcg to **process** the data

EOS Fuse: **/eos/atlas/phys-top**

EOS Fuse: **/mycernbox/topmass.kumac**

Experiment SW: **/cvmfs/athena**

Results (ntuples) aggregated on **/mycernbox/topmass** are **synced** on his laptop as the

↳ if desired

jobs are being completed

Working on **final plots** on his **laptop** and Latex-ing the paper directly on **/mycernnbox/topmass/paper**

Share on-the-fly:

n-tuples

Final plots

Publication

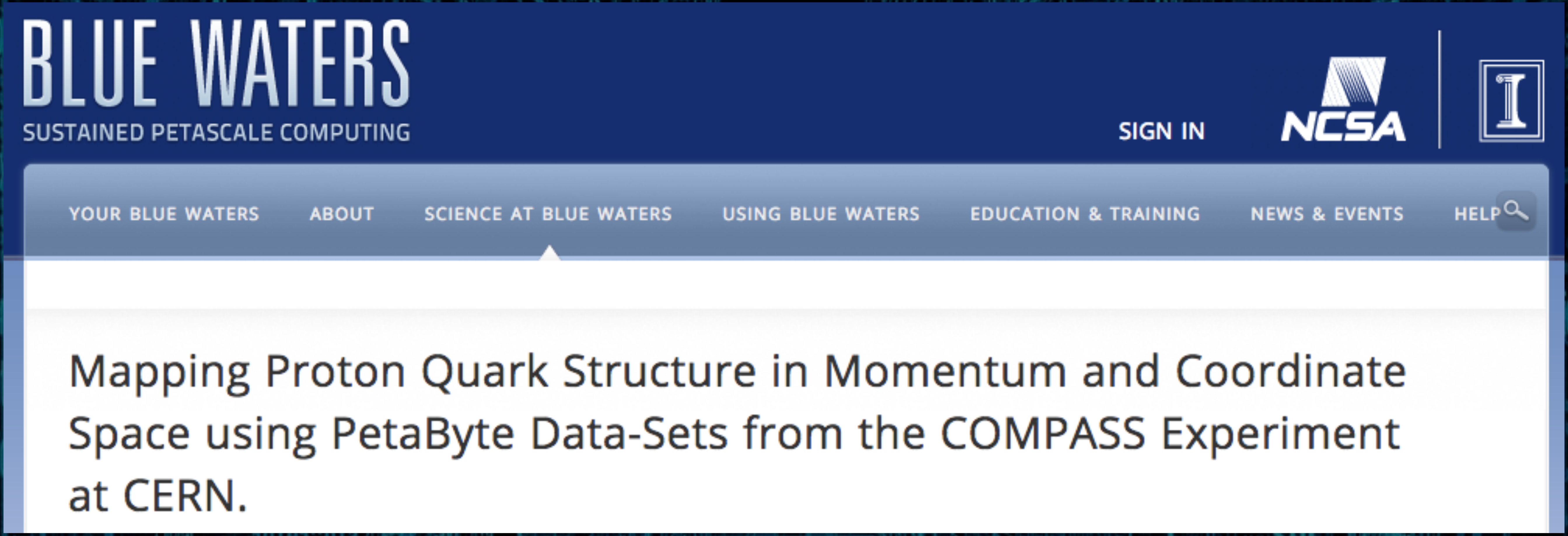
via **/mycernbox**

EOS is the enabling technology binding all this

Multi QoS Access patterns Protocols Redundancy

Goals

Make data access easy
Make analysis simple
Facilitate Science



Mapping Proton Quark Structure in Momentum and Coordinate Space using PetaByte Data-Sets from the COMPASS Experiment at CERN.

Interfaced **CERN storage services** with **Blue Waters NCSA** using **WLCG's FTS3** to manage the data workflow

Open door for HPC environments to link with our HTC and Distributed Computing expertise

Goals

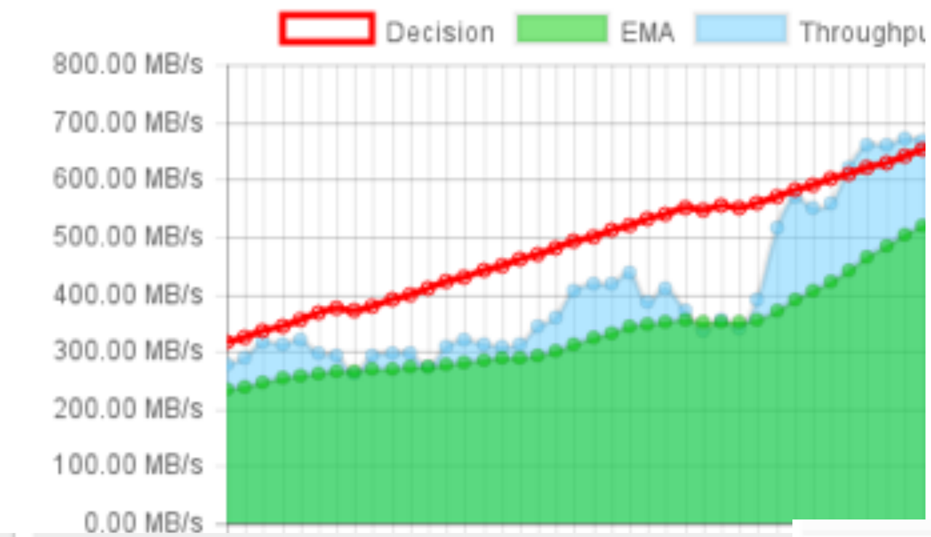
Make data access easy
Make analysis simple
Facilitate Science

COMPASS

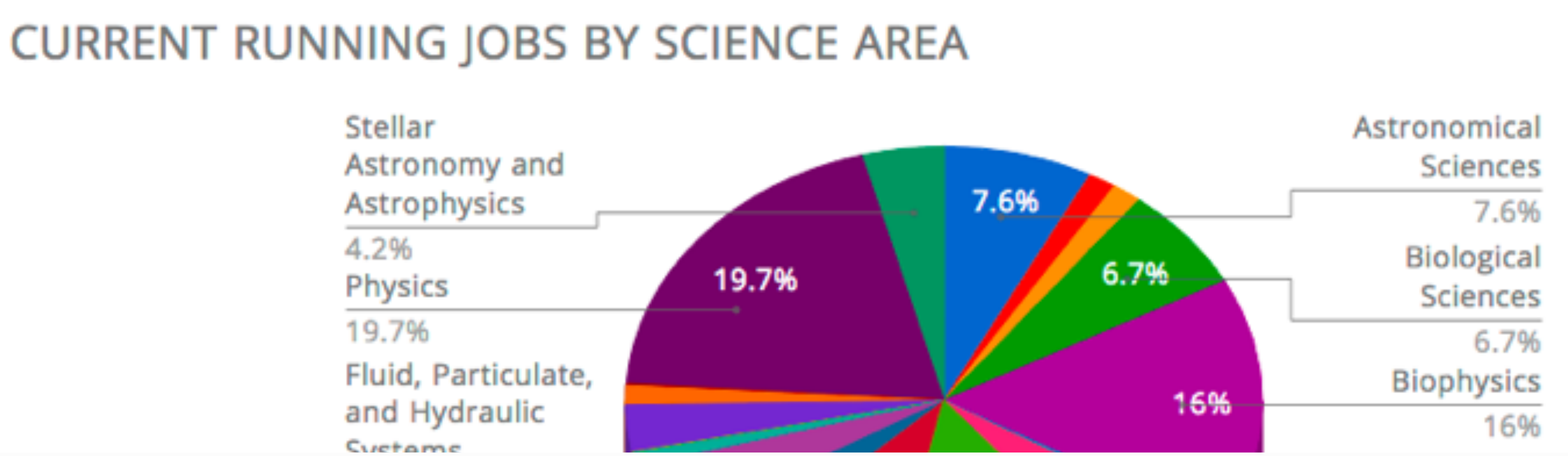
WLCG

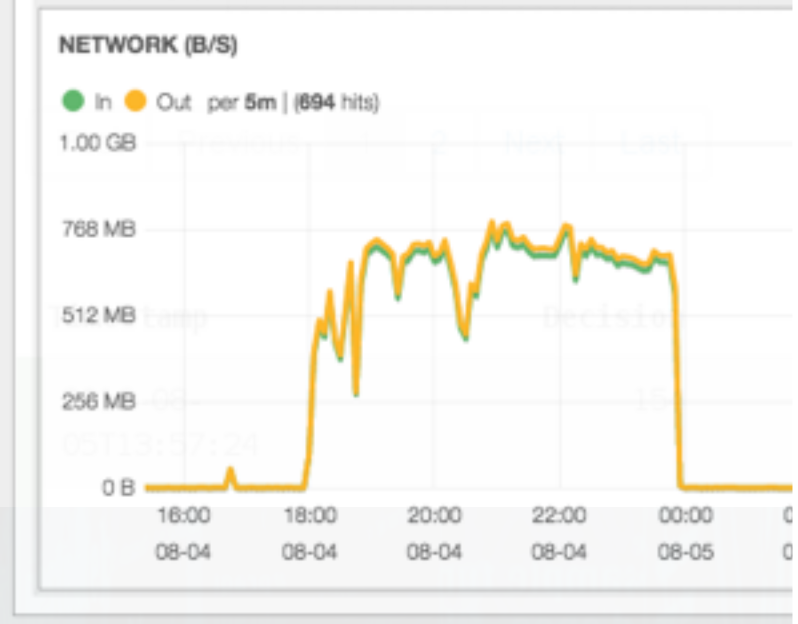
BLUE WATERS
SUSTAINED PETASCALE COMPUTING

Details for srm://castorpublic.cern.ch → gsiftp://ie15.ncsa.illinois.edu



CURRENT RUNNING JOBS BY SCIENCE AREA

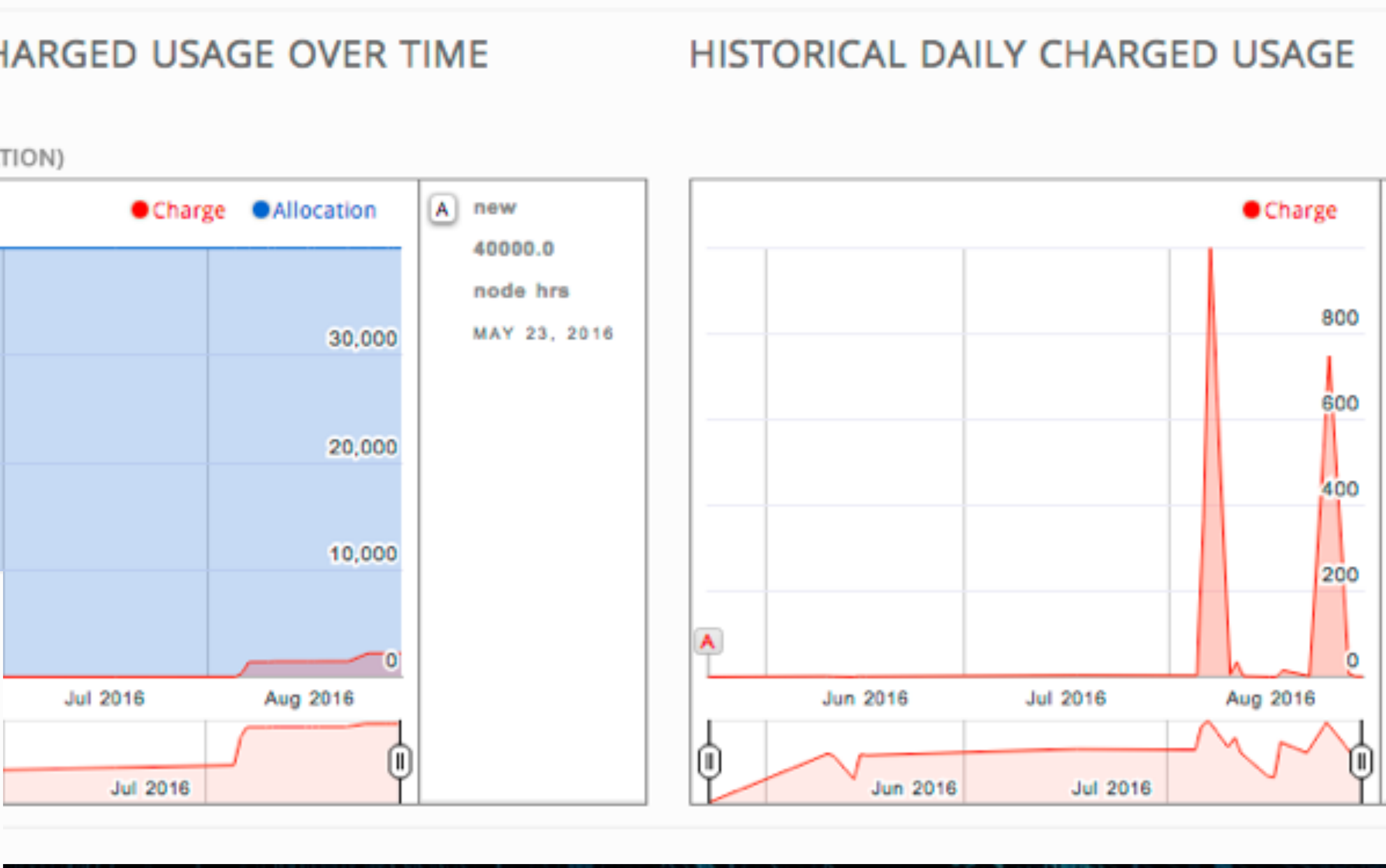




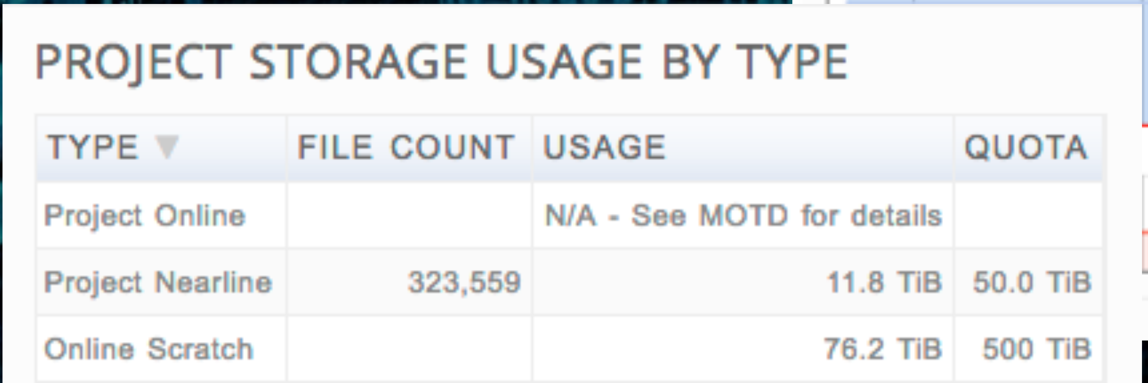
HISTORICAL CHARGED USAGE OVER TIME

HISTORICAL DAILY CHARGED USAGE

(81 DAYS UNTIL EXPIRATION)



PROJECT STORAGE USAGE BY TYPE



TYPE ▼	FILE COUNT	USAGE	QUOTA
Project Online		N/A - See MOTD for details	
Project Nearline	323,559	11.8 TiB	50.0 TiB
Online Scratch		76.2 TiB	500 TiB

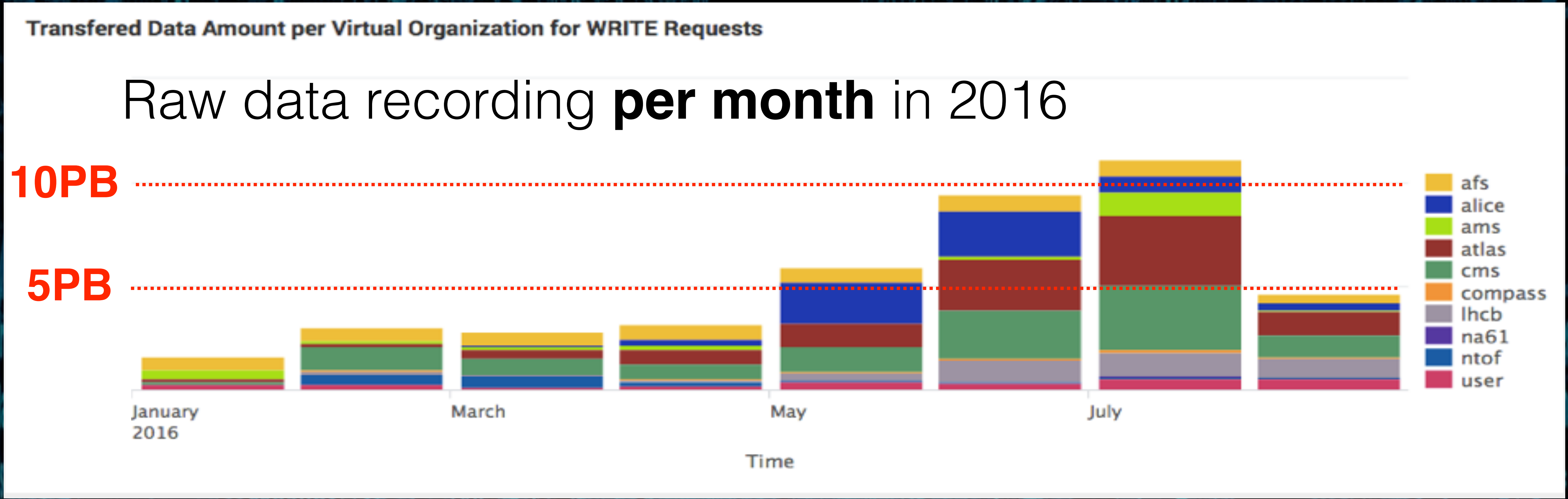
Pushing boundaries

Raw data recording for Jan [illegible] -August

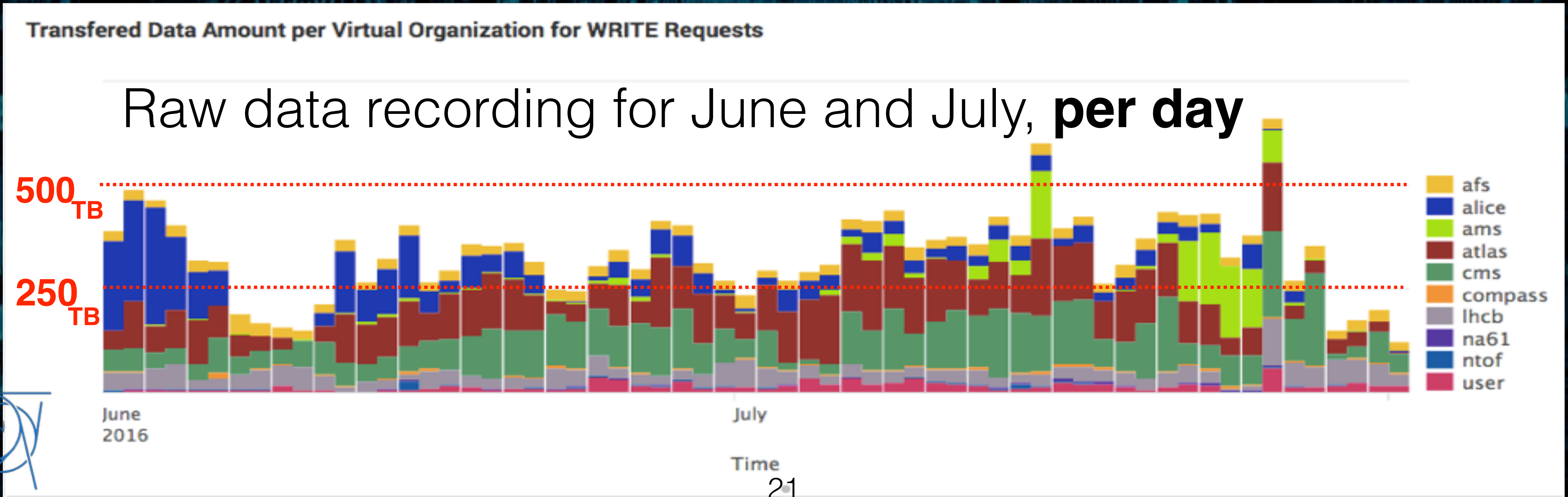
~~La Vie en Rose~~

Pushing boundaries

LHC running at **full speed** before ICHEP 2016, **unprecedent** amounts of data



Raw data recording **per month** in 2016

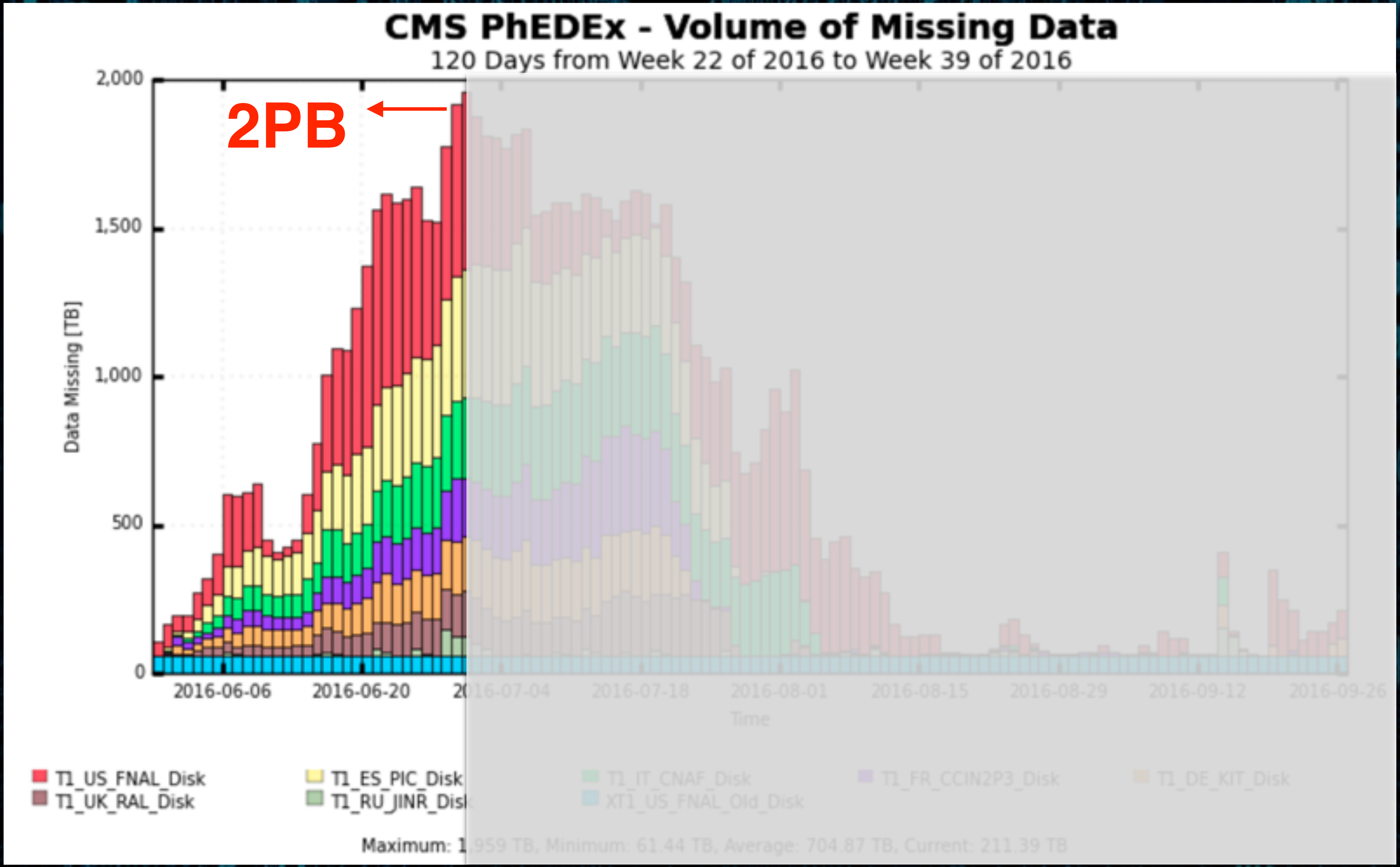


Raw data recording for June and July, **per day**

Pushing boundaries

LHC running at **full speed** before ICHEP 2016, **unprecedent** amounts of data

Systems **reaching limits** and not exporting data fast enough from Tier-0 to WLCG



WOLFGANG PETERSEN
Das Boot

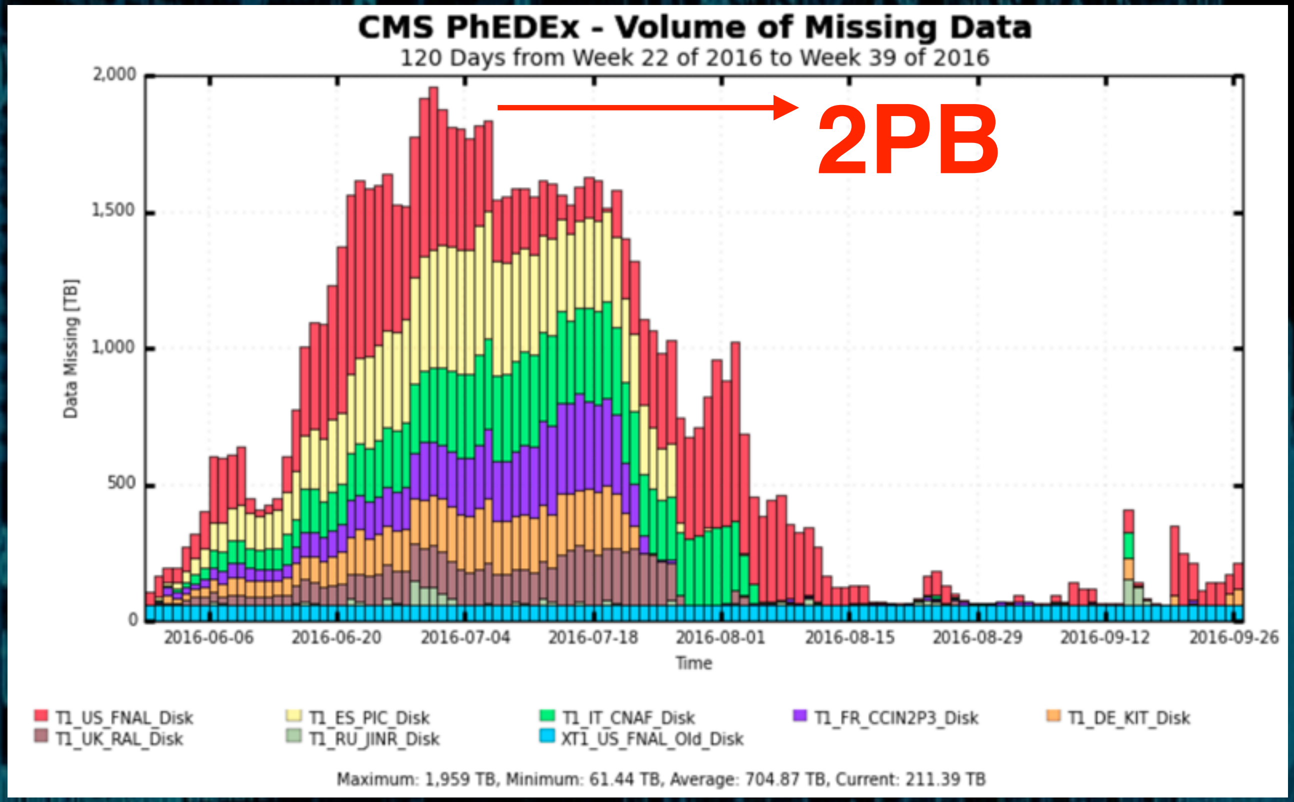
©

WOLFGANG PETERSEN
Das Boot

©
REGIE WOLFGANG PETERSEN
Das Boot

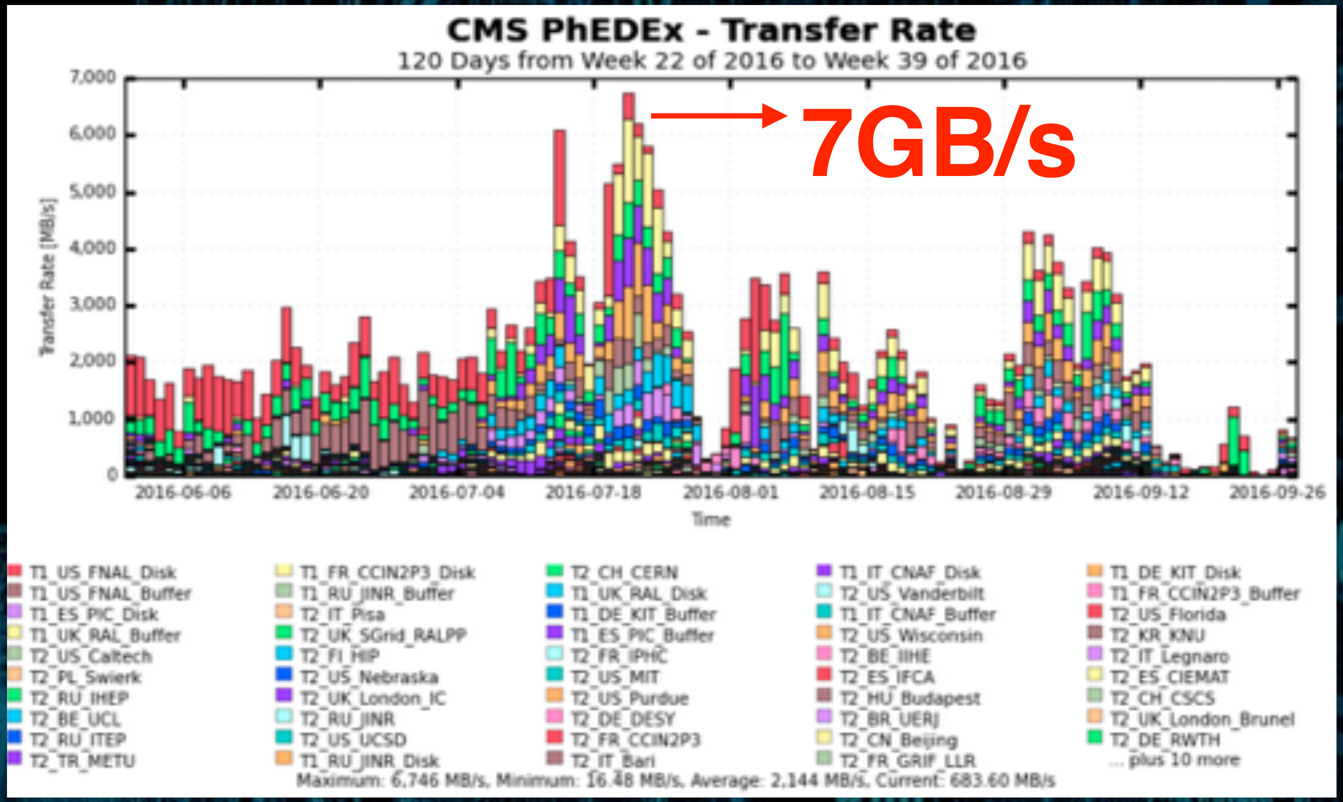
Pushing boundaries

CMS accumulated backlog



2PB

CMS Tier-0 to T* data export speed



7GB/s

Issue **solved after one week**. Many investigations, many actions and many experts from Storage, Network, Experiments and FTS worked together to identify the issues and solve them.

Goals
summary

Ensure a coherent development and operation of storage services at CERN for all aspects of physics data

Keep developing and operating Storage Services for Physics at the highest level

Communicating
Understanding
Delivering

Keep the ability to adapt and react fast

Problem/solution
Ask/Implement
In-house knowhow

Evaluate and investigate evolutions in technologies for better service/$

More for less
Operational costs
New applications

Envision new models on data mananagement and analysis

Sync&Share
SWAN
LHC@myPC

CERN

More on **CERNBOX**:

CERNBox: the data hub for data analysis (J.Moscicki) - Poster session

Total LHC data: 130PB

More on **SWAN**:

SWAN: a Service for Web-Based Data Analysis in the Cloud(D.Piparo/E.Tejedor)12th/Oct-11:45 (SierraB)

More on distributed EOS distributed:

Global EOS: exploring the 300-ms-latency region (L.Mascetti) - Poster session

More on **Cern Tape Archive (CTA)**:

An efficient, modular and simple tape archiving solution for LHC Run-3 (S. Murray) - Poster Session

From Physics to industry: EOS Outside HEP (XE)