



# WLCG plans: presentation for SKA



Massimo Lamanna / CERN



May 16, 2017

SKA-WLCG workshop



# WLCG

Worldwide LHC Computing Grid

## Make LHC computing possible

Worldwide infrastructure (collaboration) open to all LHC physicists

Computing/storage resources at CERN: ~ 20%; 80% across about 200 sites worldwide

## Data Reconstruction

Goals: data quality and immediate access for analysis

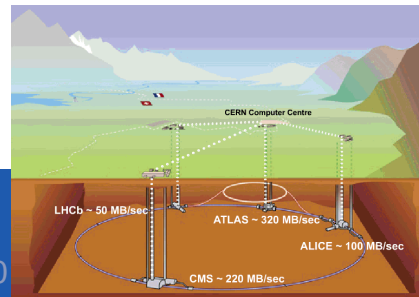
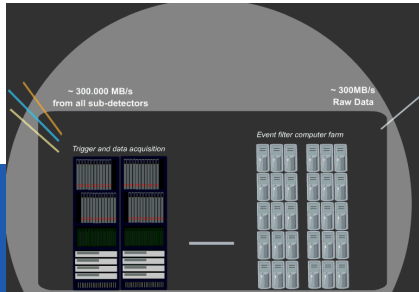
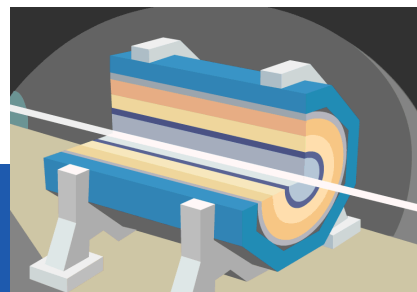
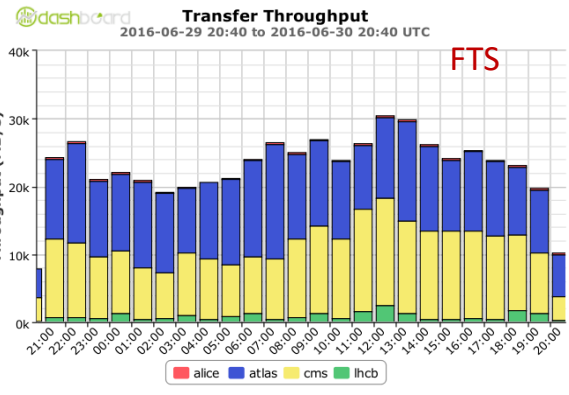
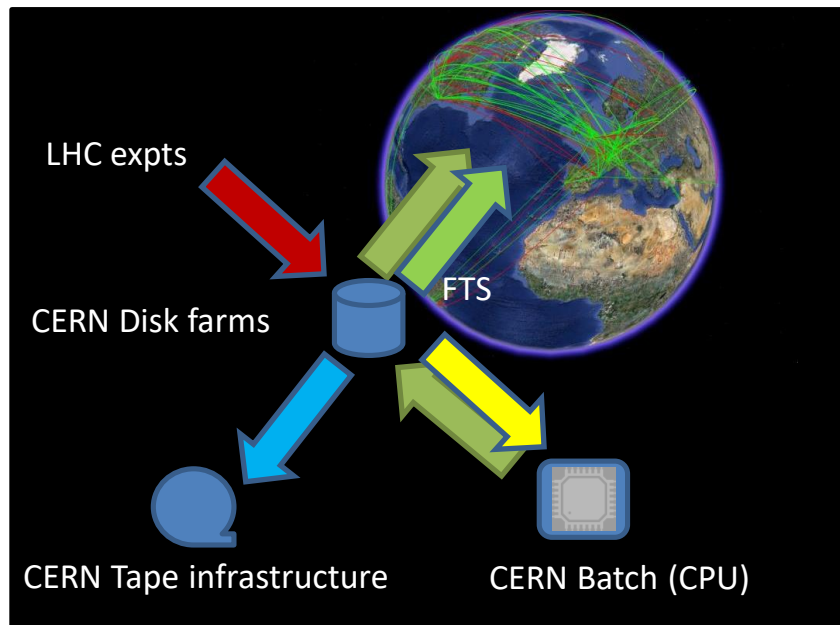
Organised activity dominated by heavy processing and replication (each expt: 1-8 GByte/s)

## Data Analysis

Goals: extract physics quantities (discovery)

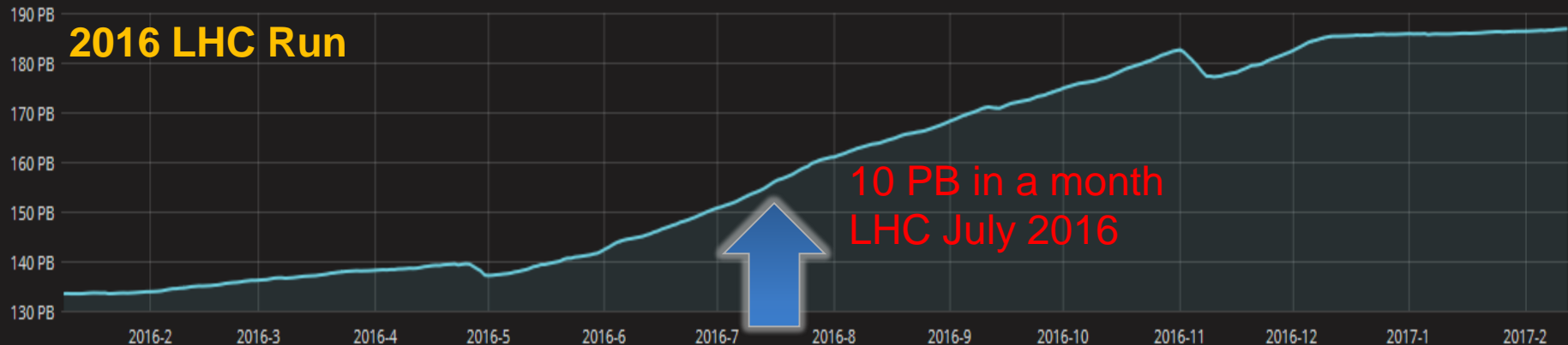
Individual activities dominated by event selection and sharing (thousands of physicists)

## (Detector) simulation



## Physics Data in CASTOR

**2016 LHC Run**



fileSize Current: 173.6 PB sizeOnTape Current: 186.8 PB

fileCount Current: 492 Mil

**Data archival at CERN  
(last 17 years)**



10 PB mark reached in 2008 (8 years)

May 16, 2017

SKA-WLCG workshop

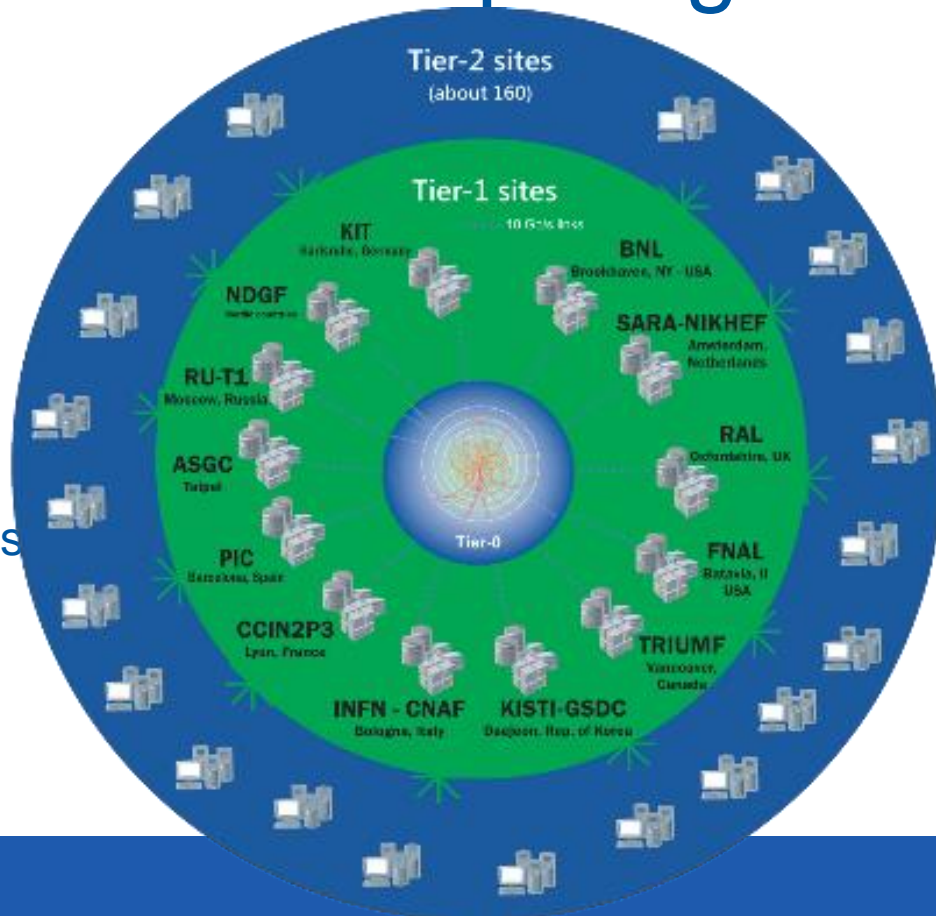


# Worldwide LHC Computing Grid

Tier-0:  
data recording,  
reconstruction and  
distribution

Tier-1: permanent  
storage, re-  
processing, analysis

Tier-2: Simulation,  
user analysis



~170 sites, 40 countries

~500k CPU cores

~1000 PB of storage

2+ million jobs/day

Multiple 10-100 Gb links

LCG:

Initial description: 2001

Tech. Design Report: 2005



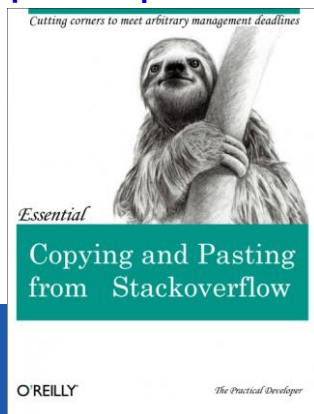
# Evolution does not stop here...



Low-impedance share of ideas  
to jump out the “submit-print-discuss” loop



“Agile” pick-up of new tools



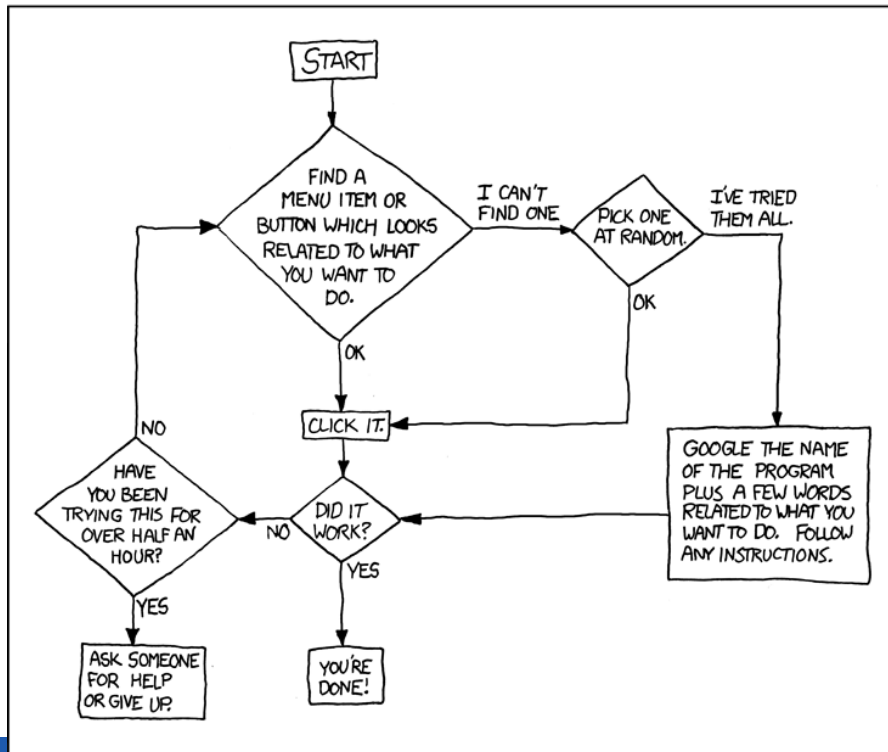
Heavy-duty tools made easy



Long-term  
reproducibility/usability

DEAR VARIOUS PARENTS, GRANDPARENTS, CO-WORKERS,  
AND OTHER "NOT COMPUTER PEOPLE."

WE DON'T MAGICALLY KNOW HOW TO DO EVERYTHING IN EVERY  
PROGRAM. WHEN WE HELP YOU, WE'RE USUALLY JUST DOING THIS:



Tried this...

Got some (nice) lego blocks...



... you might want to try out!

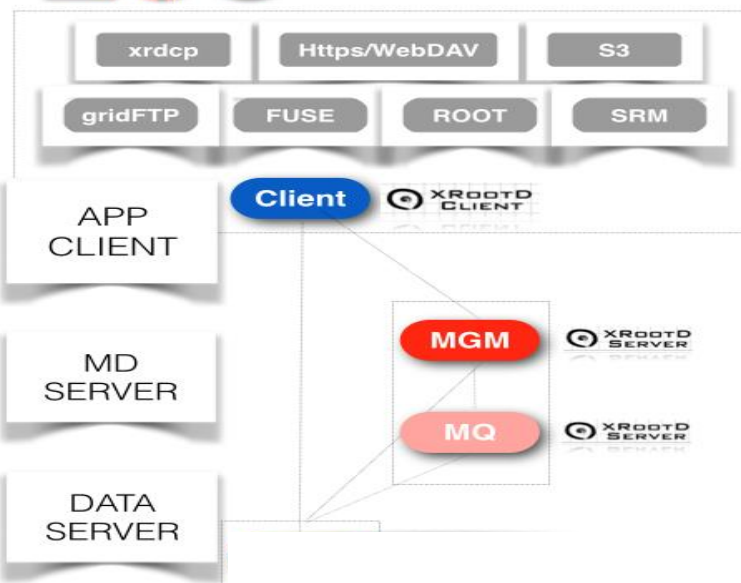
PLEASE PRINT THIS FLOWCHART OUT AND TAPE IT NEAR YOUR SCREEN.  
CONGRATULATIONS; YOU'RE NOW THE LOCAL COMPUTER EXPERT!

May 10, 2017

SKA-WLCG workshop



# Open Source Storage

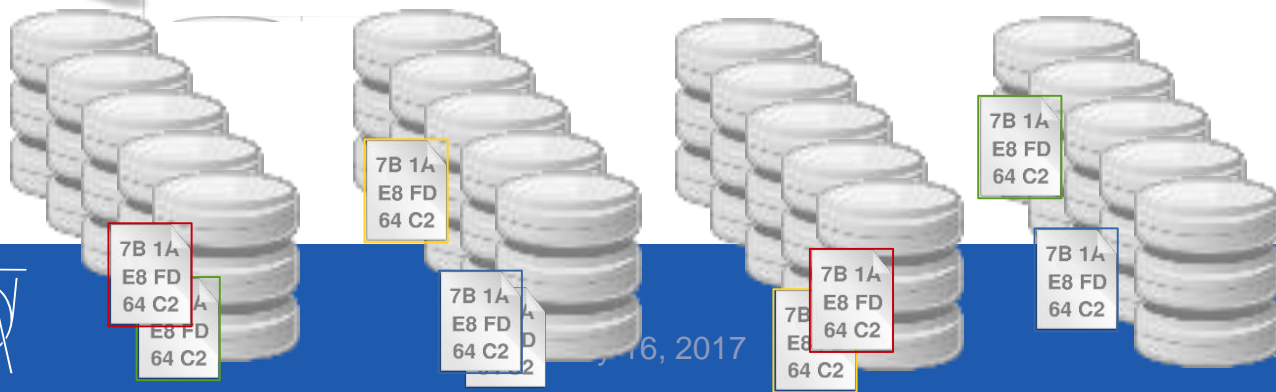


## › EOS: Large disk farms for physics and beyond

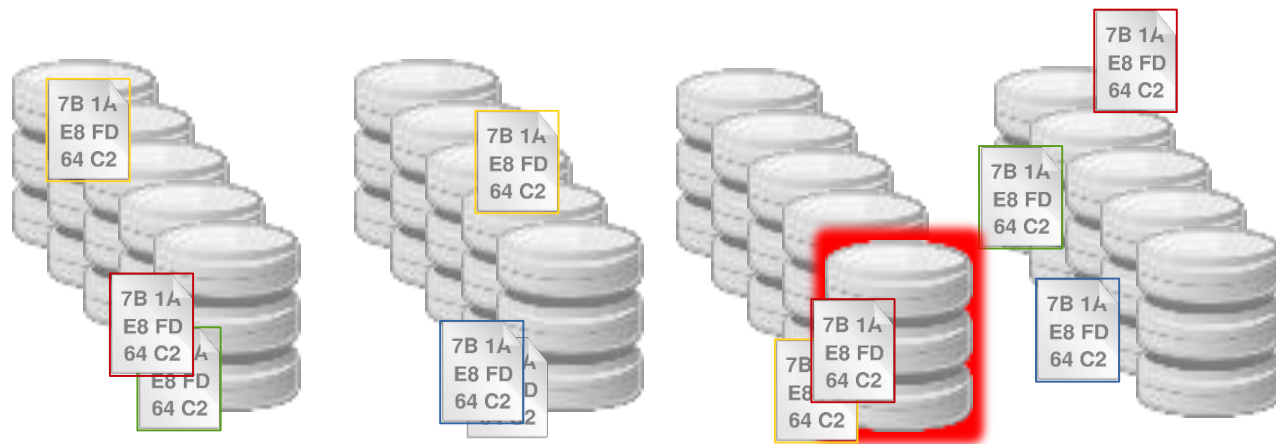
- Developed at CERN
- LHC: PBs for 100s/1000s independent scientists
- 200 PB JBOD installed (CERN installations)

## › Strategic points

- Distill **20+ years** of experience data management
- **Ultra-fast** name space
- Arbitrary level of data durability: cross-node file replication or RAIN on **commodity** hardware
- Large **protocol choice**



- Annual Failure Rate  $\sim 0.25\%$ 
  - $4 \cdot 10^4$  disks
- Better schemas?
  - Erasure code
    - RAID6 like
  - Ready but not deployed
  - Less overhead (cfr. RAID1 and RAID6)
  - A-priori also faster
    - Fragments go to clients from multiple nodes



NB: After 1 disk failures, N-1 sources available (N=10000)

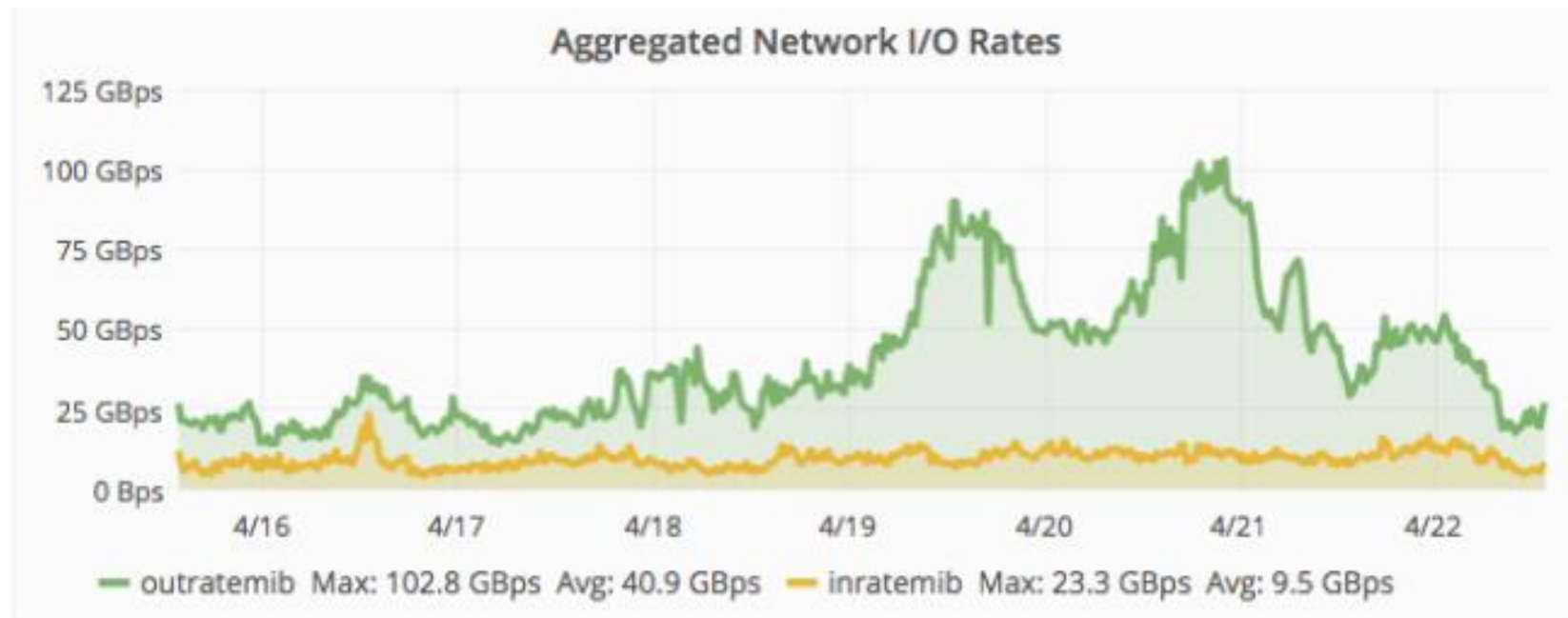
# EOS Service

Storage for physics

And for general storage (CERNBox: see later)

Twin computer-centre deployment

3 · 100-Gb links (~22 ms latency)



Number of Files

1390 M

Number of Directories

108 M

Write Throughput

2.390 GBps

Read Throughput

24.5 GBps

Current Readers

57.7 K

Current Writers

9.5 K

Total Space

166 PB

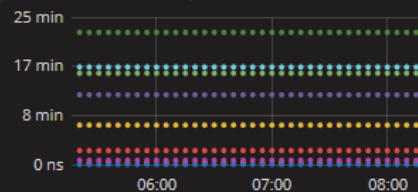
Free Space

43.81 PB

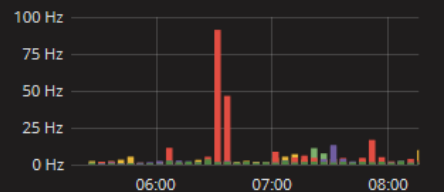
IOPS

374 K

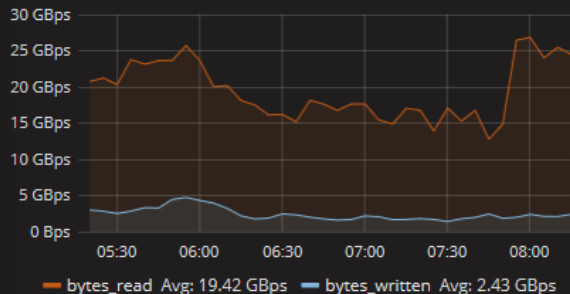
Namespace boot time



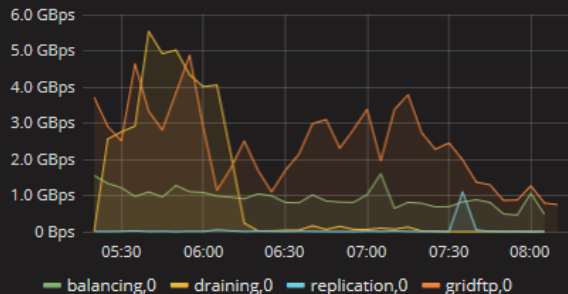
File deletion rate



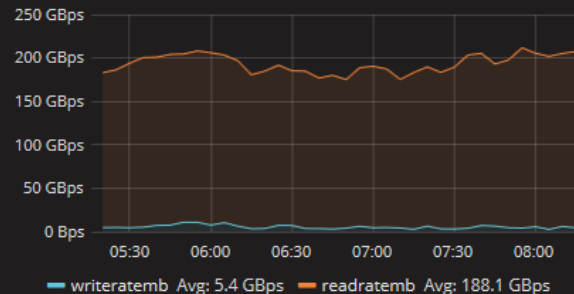
EOS Total IO



EOS Total IO internal



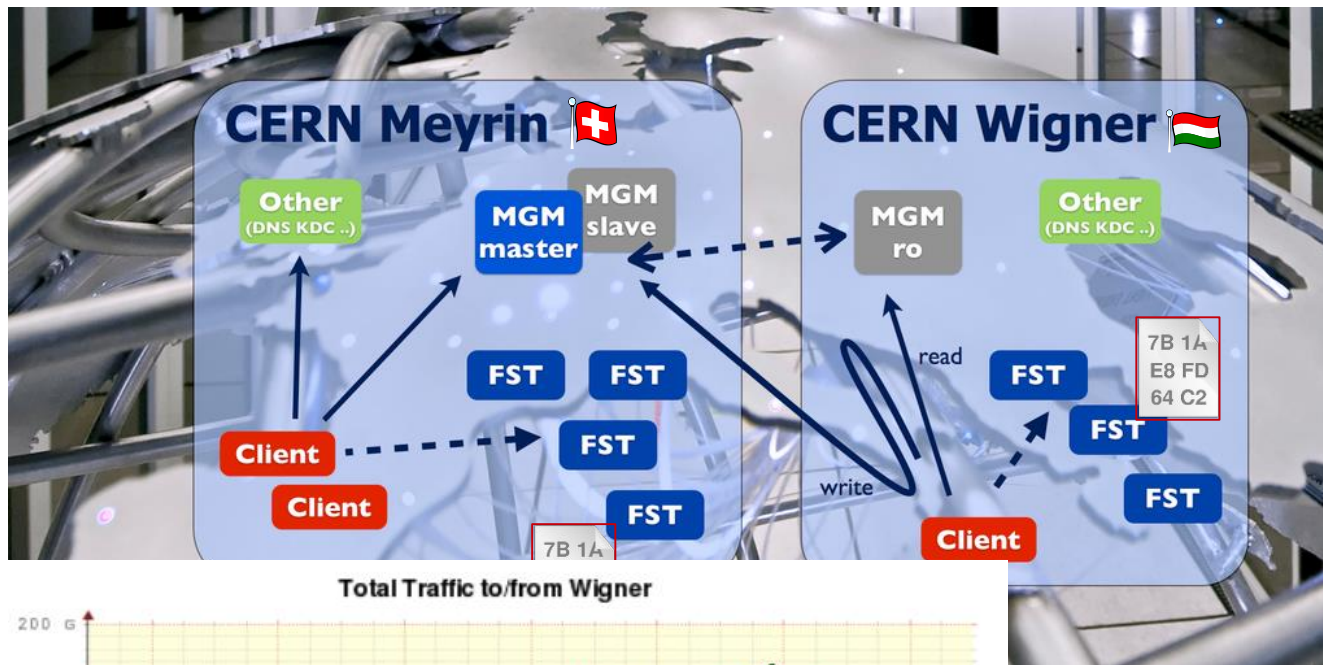
Aggregated Disk IO



May 16, 2017

SKA-WLCG workshop

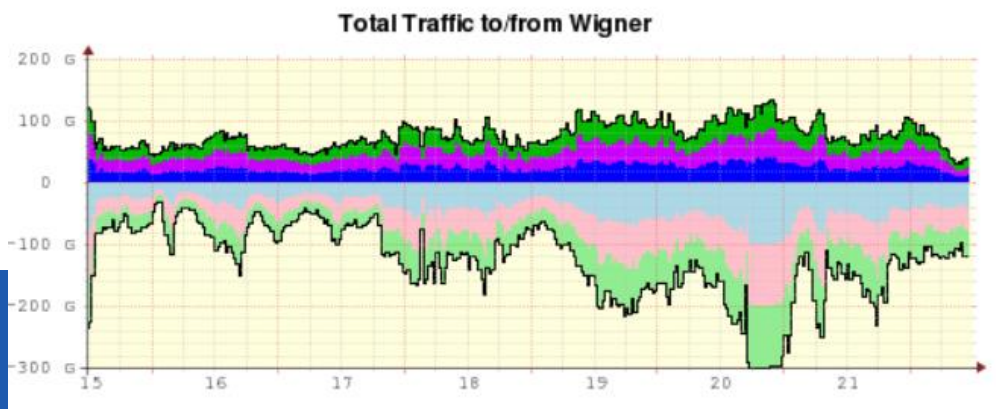
# Our “20-ms-large” computer centre



Geneva – Budapest  
 3 x 100 GB lines  
 ~ 22 ms latency (diff routes)  
 ~ 1000 km

Autonomic, Locality,  
 Business continuity

Certainly more complex  
 OK with 2 replicas, less interesting  
 with other erasure codes



# EOS evolution

- Resilient scalable catalogue well beyond 10B  
Now 1.3 B entries
- High-performance POSIX access (Fuse)
- Archival capabilities (CTA)
- Extended usage in production of erasure code  
Zero-operation mode  
Cheaper hardware not impacting quality of service
- Collaboration with external sites  
HEP sites: Russian cloud, IHEP in Beijing, ...  
Other sciences/activities: JRC and AARNET best examples
- Evolution of the WLCG  
Data federations

## R&D

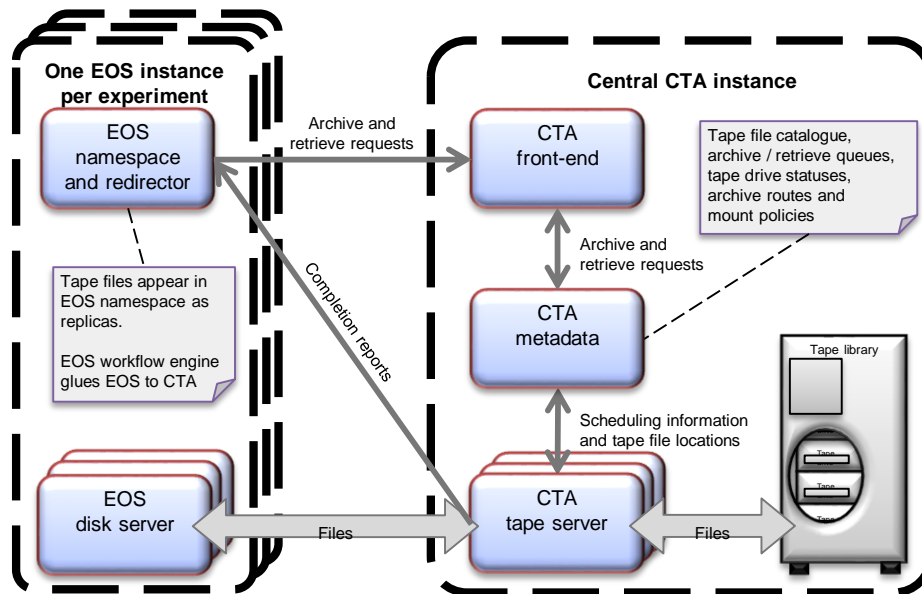
- CERN-IT extra-large disk server project
  - **8 x 24 x 6TB** disks connected to single front-end node [ 1.152 PB/node ]
    - capacity/performance ratio ?
    - OS limitations handling 192 disks ?
    - RAID vs. ZRAID vs. Software EC
    - which network IF ?
    - which CPU type ?
    - TCO evaluation



# CTA- CERN Tape Archive

## A tape backend for EOS

- Removes duplication between current MSS (CASTOR) and EOS: namespace, file access and protocols, disk cache management
- Thin scheduling layer on top of existing CASTOR tape software
- EOS drives life cycle for archiving/restoring files from/to tape
- Same tape format as CASTOR – only need to migrate metadata
- Under development, aimed for LHC Run-3



# Joint Research Centre (JRC)

## Science Service of the European Commission



### "Earth Observation & Social Sensing Big Data Pilot Project"

- The EU **Copernicus** Programme with the **Sentinel** fleet of satellites acts as a game changer by bringing EO in the Big Data era:
  - expected 10TB/day of **free and open** data
  - Requires new approaches for data management and processing
- Pilot project launched in January 2015
- Major goal: set up a central infrastructure for storing and processing of Earth Observation and Social Sensing data at JRC



Sentinel-1 (Credits: ESA/P. Carr)



Sentinel-2 (Credits: ESA/P. Carr)



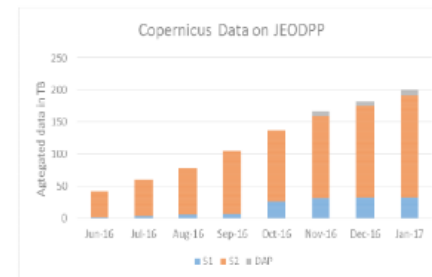
Sentinel-3 (Credits: ESA/J. Hu)

Joint  
Research  
Centre



### EOS set-up at JRC

- Installation and configuration at JRC with strong support from CERN storage team
- Current set-up:
  - 1.4 PB gross capacity
  - 10 FST nodes, each with one JBOD of 24x6 TB disks
  - Using replica 2
- Further extension planned
  - 2017: extend to ~6 PB gross



A. Burger and P. Soille (JRC)

May 16, 2017

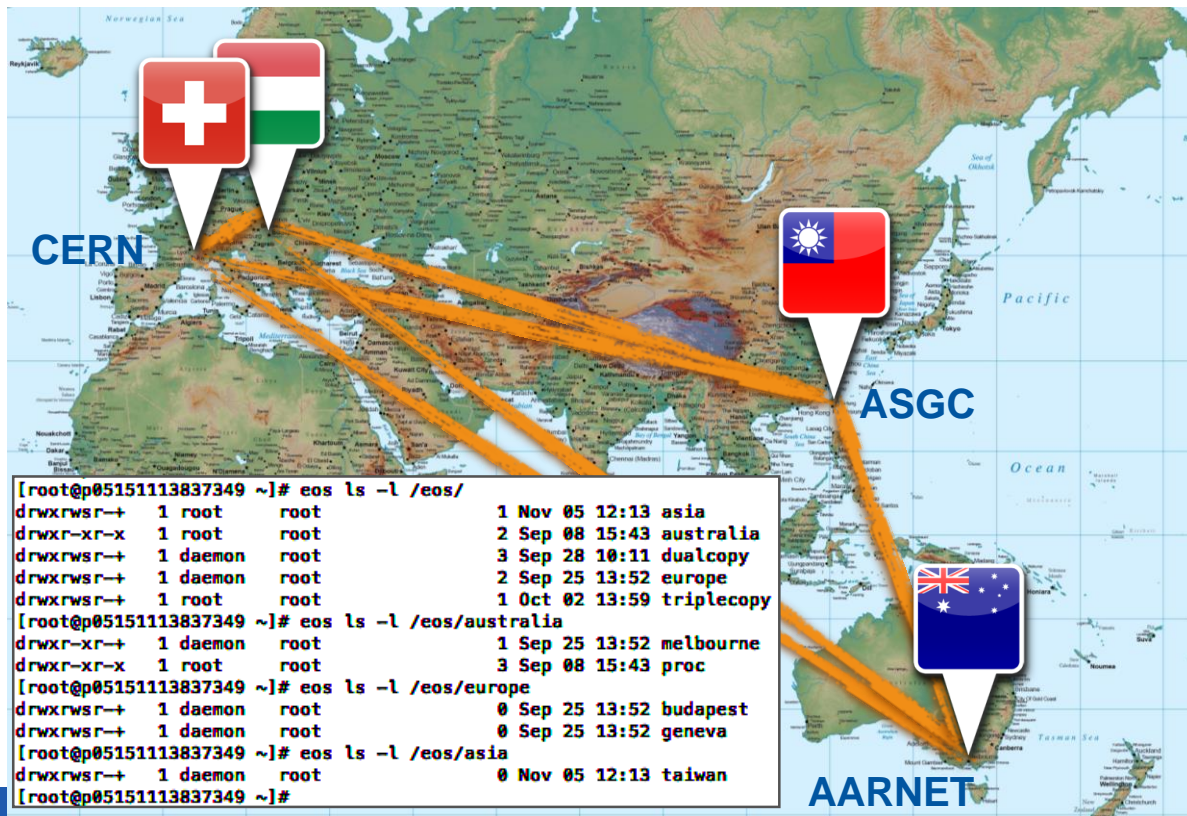
Joint  
Research  
Centre



SKA-WLCG workshop

# AARNET collaboration

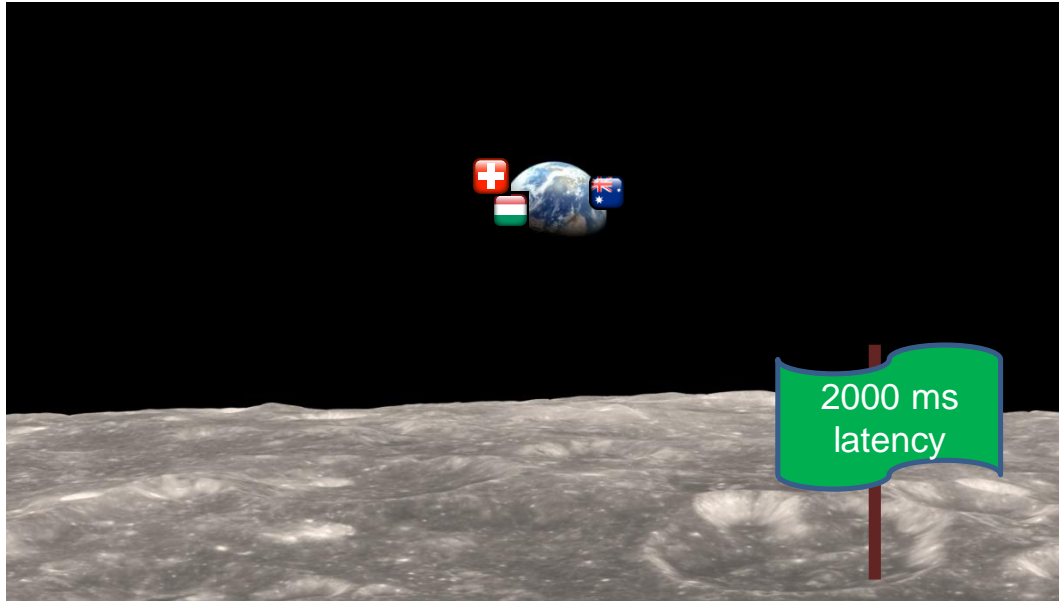
## Exploring the 300 ms region...



D. Jericho (AARNET), L. Mascetti (CERN),  
Asa Hsu (ASGC Taipei)



# Another factor of 10 is probably not needed, yet...





National Research Centre (NRC)  
"Kurchatov Institute"



Big Data Technologies Laboratory  
<http://bigdatalab.nrcki.ru/>



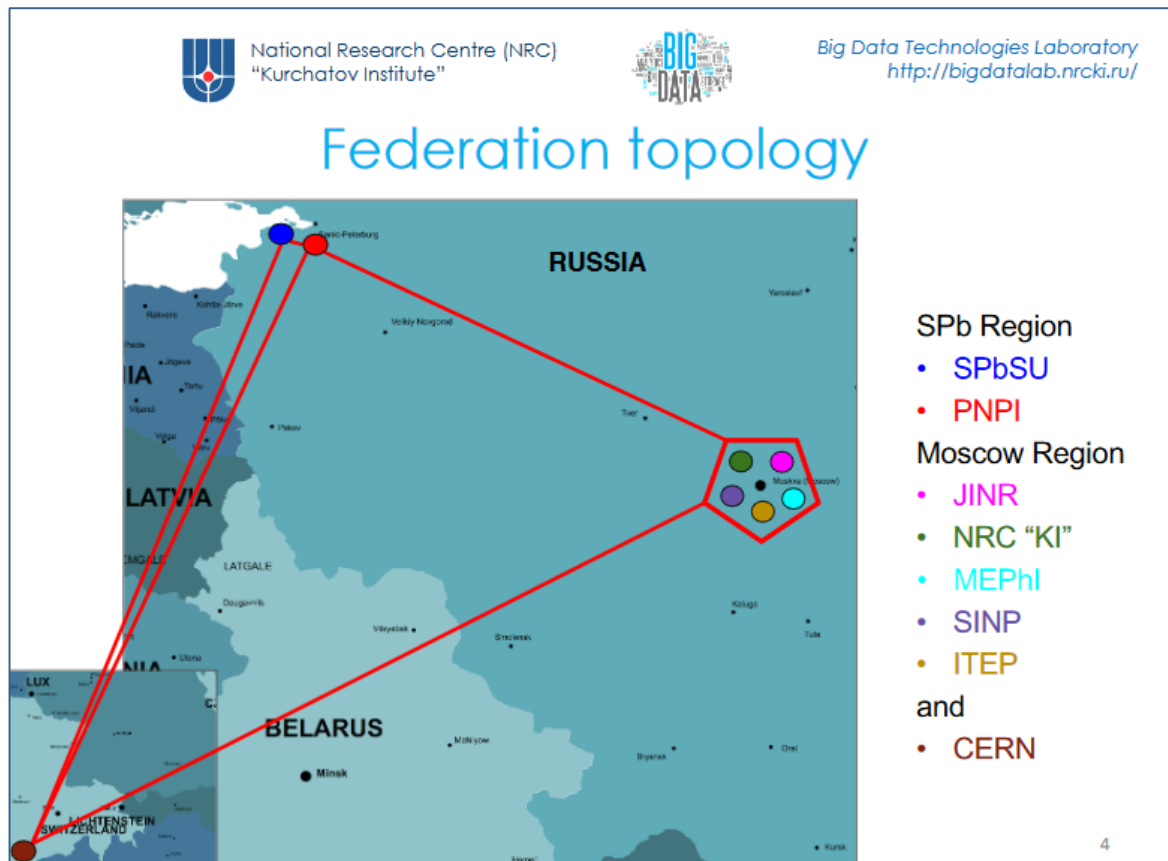
# Russian Federated Data Storage System Prototype

Andrey Kiryanov, Alexei Klimentov, Andrey Zarochentsev

on behalf of BigData lab @ NRC "KI" and  
Russian Federated Data Storage Project

- HEP communities
  - Collaboration
  - Complementarity
- Federation
  - Moscow area
  - St Petersburg area (CERN)
  - Sites from Russian Data Intensive Grid And WLCG site
- EOS workshop

A.Kyrianov et al.



# EOS AT 6,500 KILOMETRES WIDE

An Australian Experience  
David Jericho – Solutions Architect, AARNET

## SOLUTIONS WE HAVE TRIED



- Hadoop
  - MapR, Hortonworks, Apache official
- XtremFS
- Ceph
- GlusterFS
- pNFS
- OrangeFS

... and others

© AARNET Pty Ltd (3)



## SUCCESSES WE'VE HAD

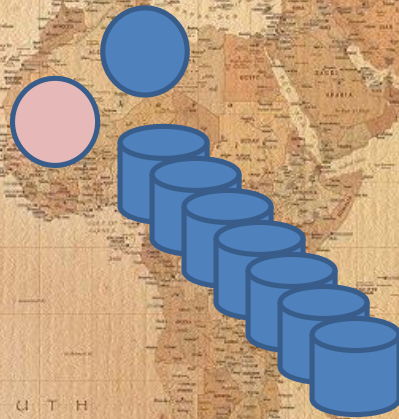
- IT WORKS!
  - Stable, server issues have been almost exclusively container related
  - Fast
- Obvious write latency penalty
  - Users don't notice
- Hello all, I know it's Monday...
  - CERN have been very responsive, THANKYOU!



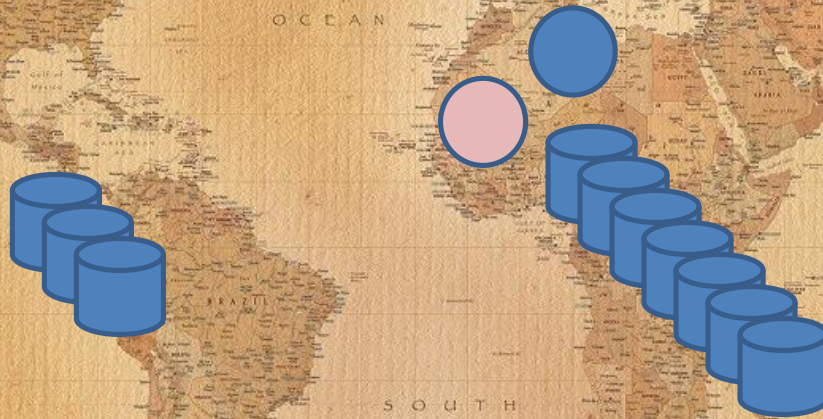
© AARNET Pty Ltd (3)



# Operations across a cloud



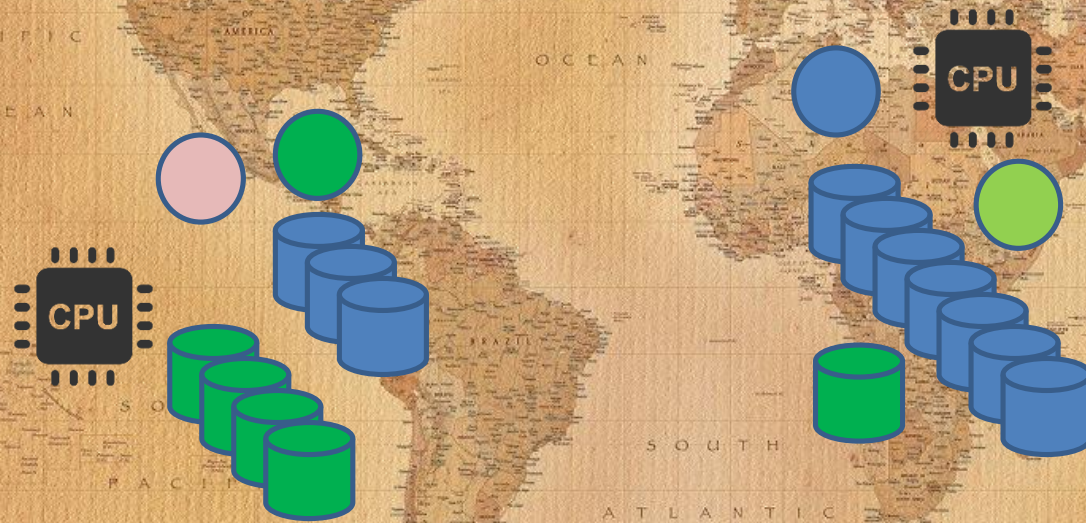
# Operations across a cloud



# Operations across a cloud



# Operations across a cloud



# EOS in DOCKER - 1 minute

- currently the docker scripts are only to get a **one machine instance** for testing
- the CERNBOX team is finishing a complete **dockerized CERNBOX**-like service package bundling EOS + OwnCloud
- prototyped a single host **ALICE docker** storage container with a pre-configured EOSALICE instance using the physical network inside the container
- interesting option to combine with **kubernetes** to simplify deployment in a storage federation .... integration is on the work plan ...
- if there is a broader interest, we can integrate the work of AARNET which is deploying EOS only via docker containers and add ALICE specifics



# Federations (rationale)

- Reduce the number of storage services to manage
- Bundle many small resources into bigger single resources
- Reduce operation effort (WLCG)
- Reduce integration effort from users (experiments)

# EOS Storage Federation

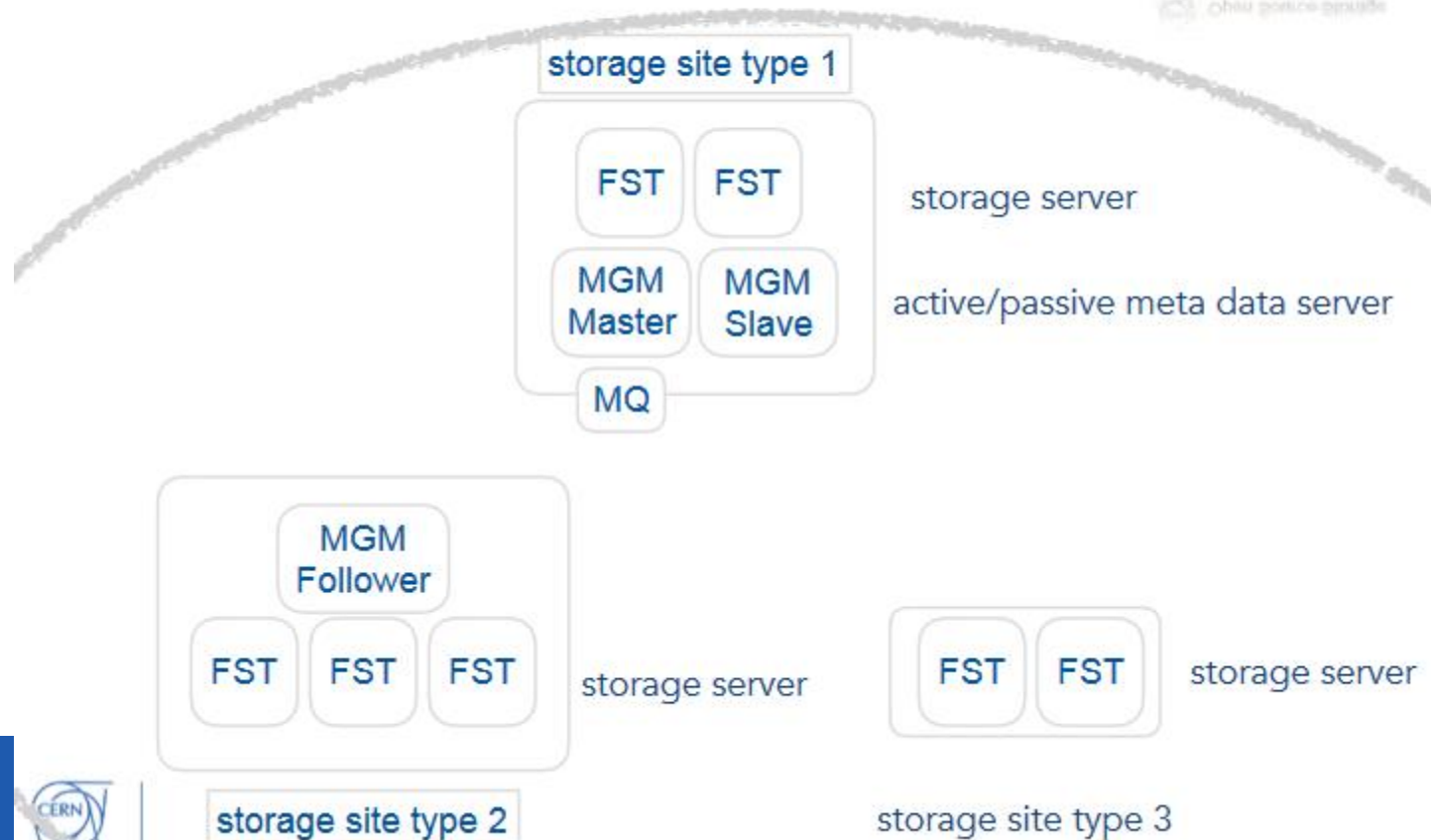
3 types of sites



Open Source Storage

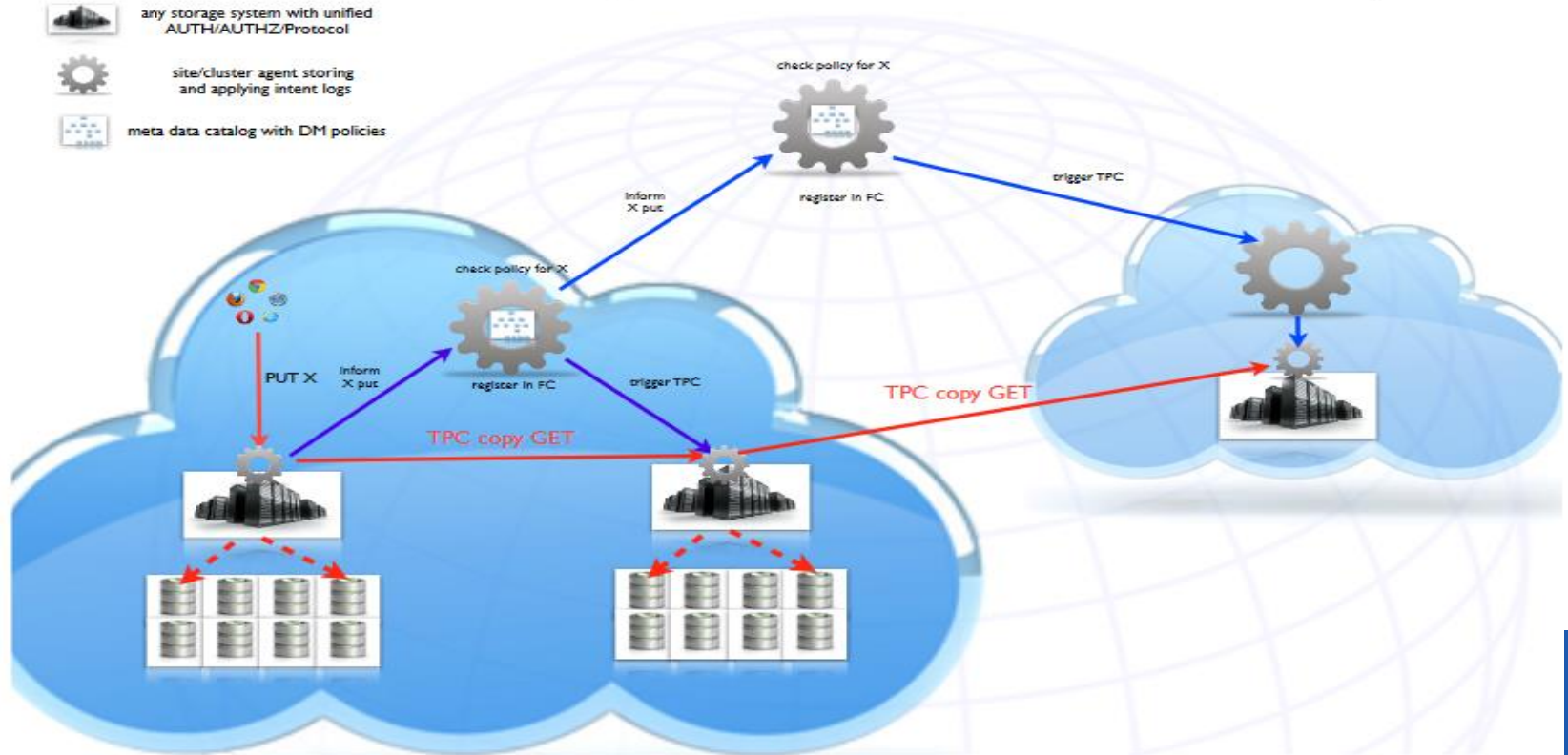


CERN



# Data federations

basic concept of event triggered data management



# 1<sup>st</sup> EOS workshop (February 2-3 2017)

## Participants



TECHNICKÁ  
UNIVERZITA  
V KOŠICIACH



Integration and support  
Disk technology

## External Collaborators



May 16, 2017

SKA-WLCG workshop



CERNBox

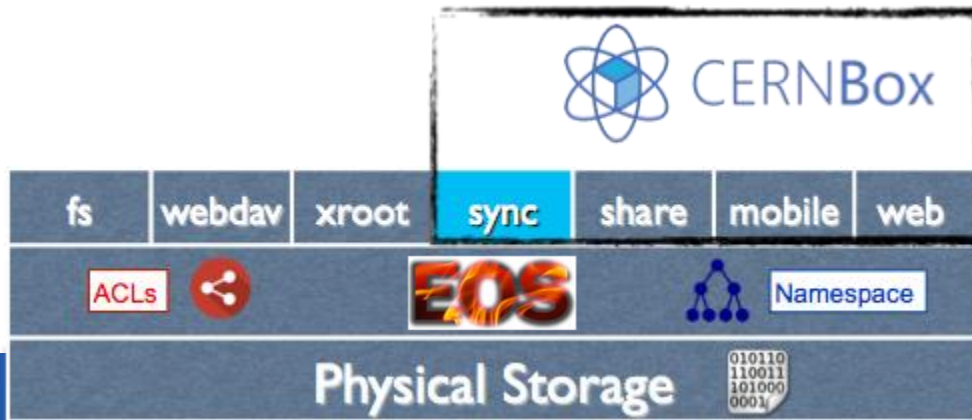


# CERNBox

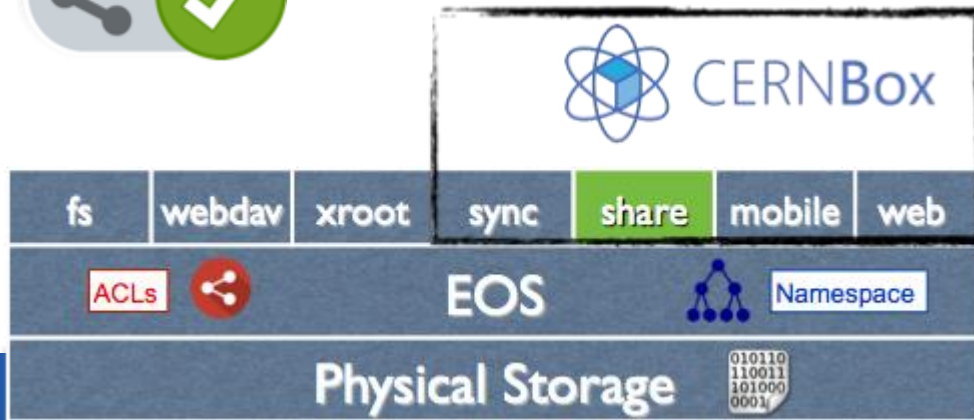
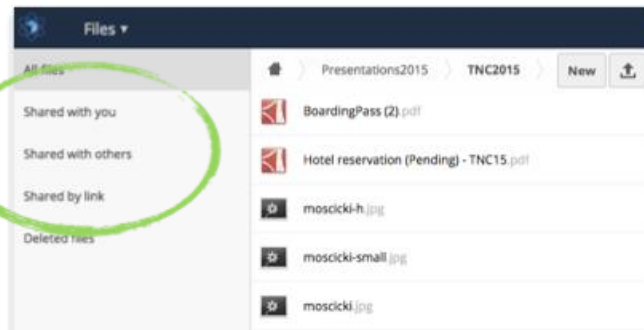


- Starting point: Dropbox-like service
  - Cloud synchronisation service
  - Just the starting point!
- Innovative way to offer storage
  - Sync and share from ownCloud GmbH
  - EOS as a back-end (all LHC data!)
  - New way to interact with your data
- Strong interest
  - In HEP: here! Interesting meeting yesterday
  - Broader scientific/university community

# Access Methods: Sync



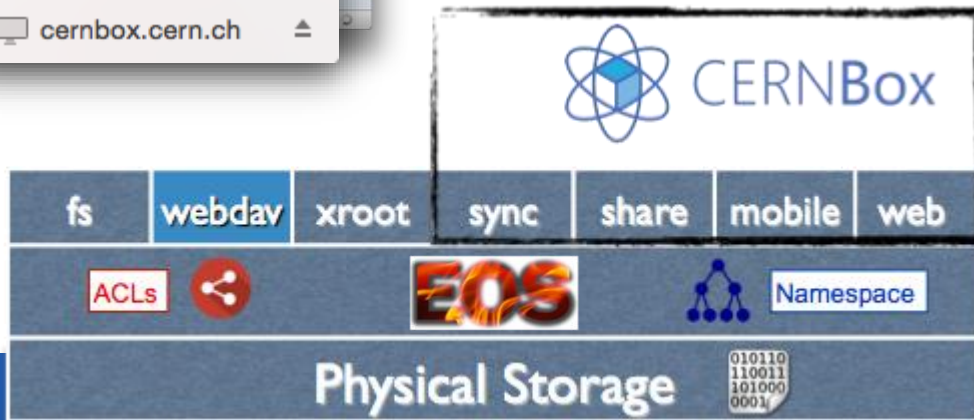
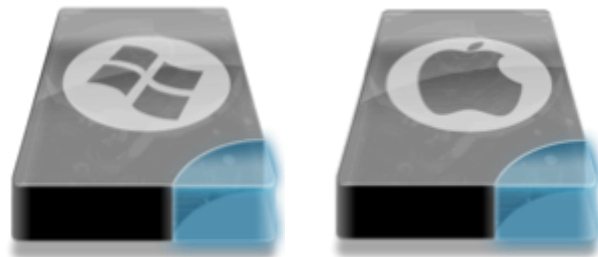
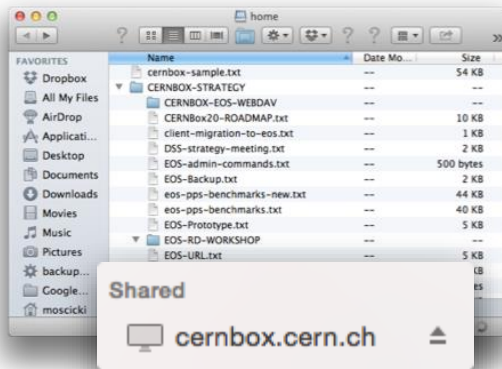
# Access Methods: Sharing



# Access Methods: Mobile & Web



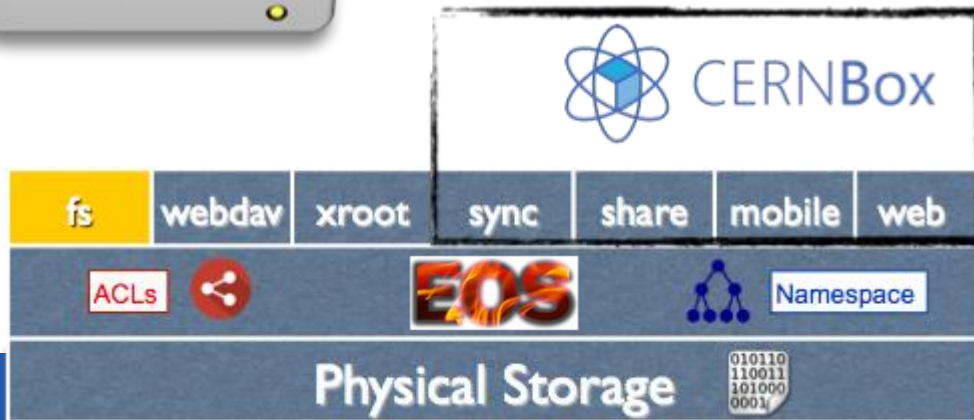
# Access Methods: WebDAV



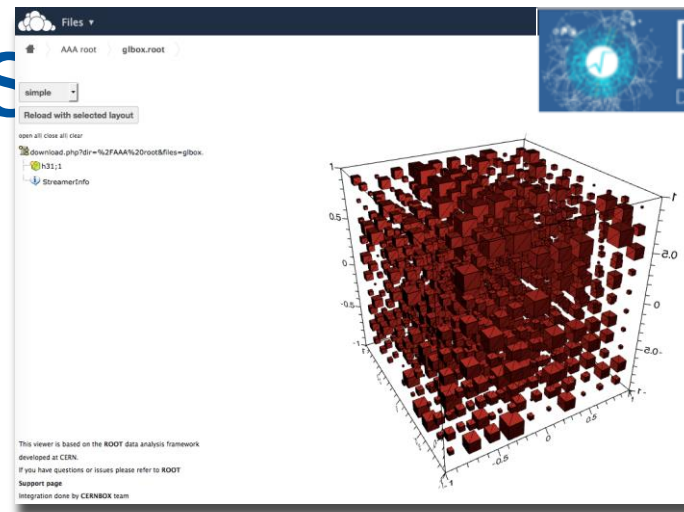
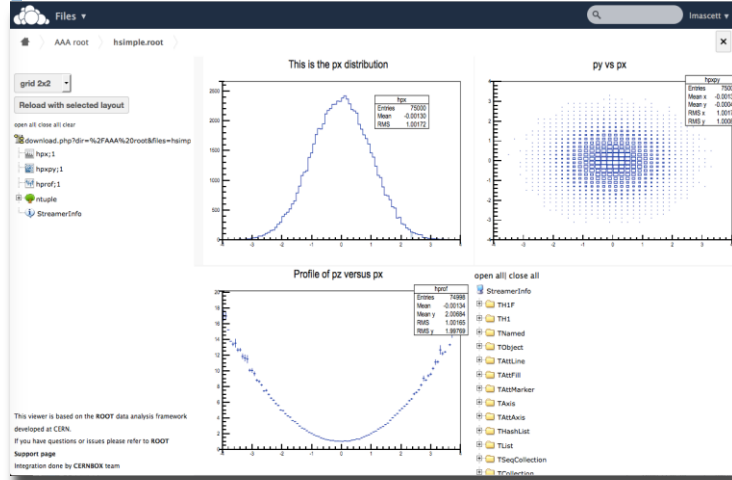
# Access Methods: FUSE



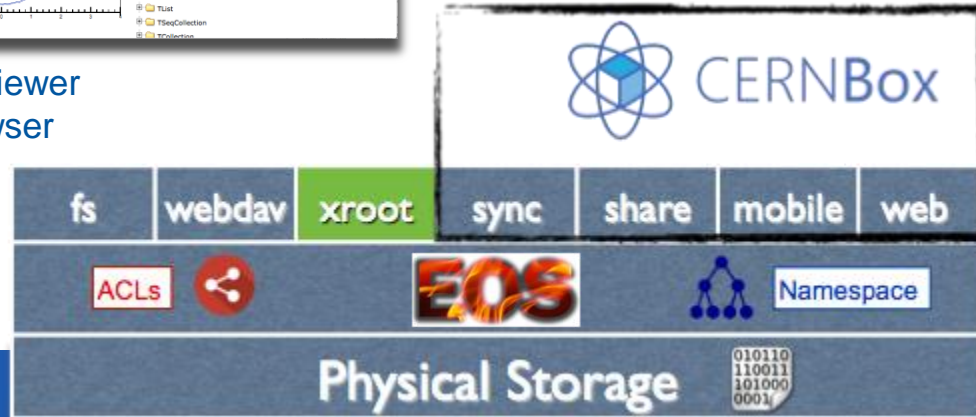
```
[lmascett@lxplus2015 ~]#  
[lmascett@lxplus2015 ~]# df -H -t fuse  
Filesystem      Size  Used Avail Use% Mounted on  
eosuser         506T   70T  437T   14% /eos/user  
eosatlas        36P    17P   20P   45% /eos/atlas  
eosalice        20P    11P   8.5P   57% /eos/alice  
eoscms          28P    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      8 Jan 26 18:18 debug  
drwx-----, 1 lmascett c3      8 Dec 11 09:43 download  
drwx-----, 1 lmascett c3      8 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
```



# Optimised access



Embedded ROOT viewer  
in CERNBox browser



May 16, 2017

SKA-WLCG workshop

# 3<sup>rd</sup> Cloud Services for Synchronisation and Sharing (CS3)

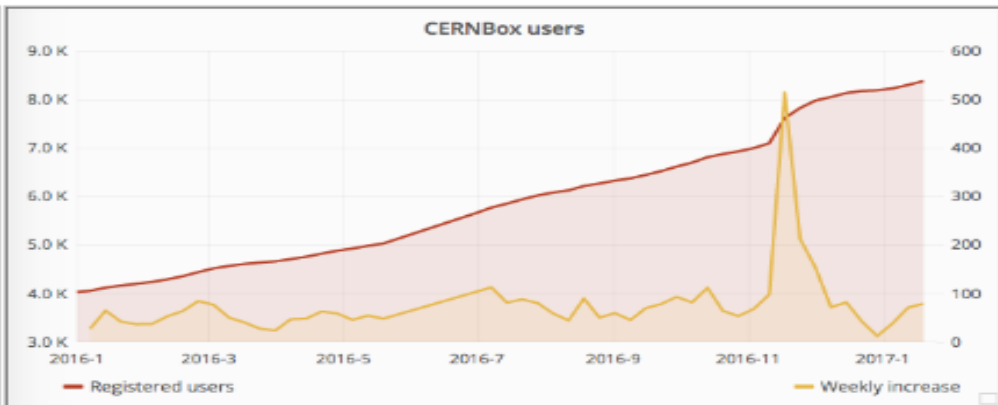
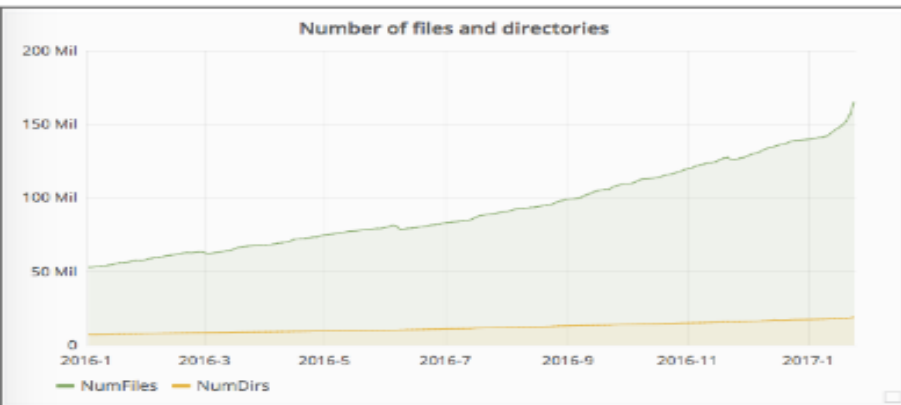
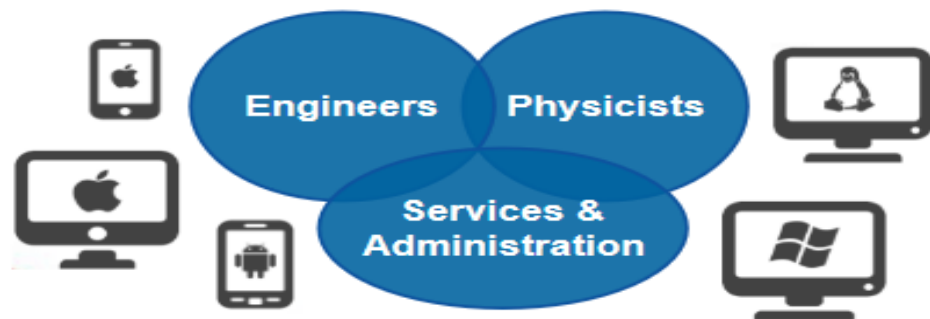
Novel applications, cloud storage technology, collaborations

Amsterdam January 2017 \*\*\* 120+ participants \*\*\* 15+ companies



# CERNBox Service Numbers

	Jan 2016	Jan 2017
<b>Users</b>	<b>4074</b>	<b>8411</b>
<b># files</b>	<b>55 Million</b>	<b>176 Million</b>
<b># dirs</b>	<b>7.2 Million</b>	<b>19 Million</b>
<b>Used Raw Space</b>	<b>208 TB</b>	<b>806 TB</b>
<b>Deployed Raw Space</b>	<b>1.3 PB</b>	<b>3.2 PB</b>





# CERNBox Clients

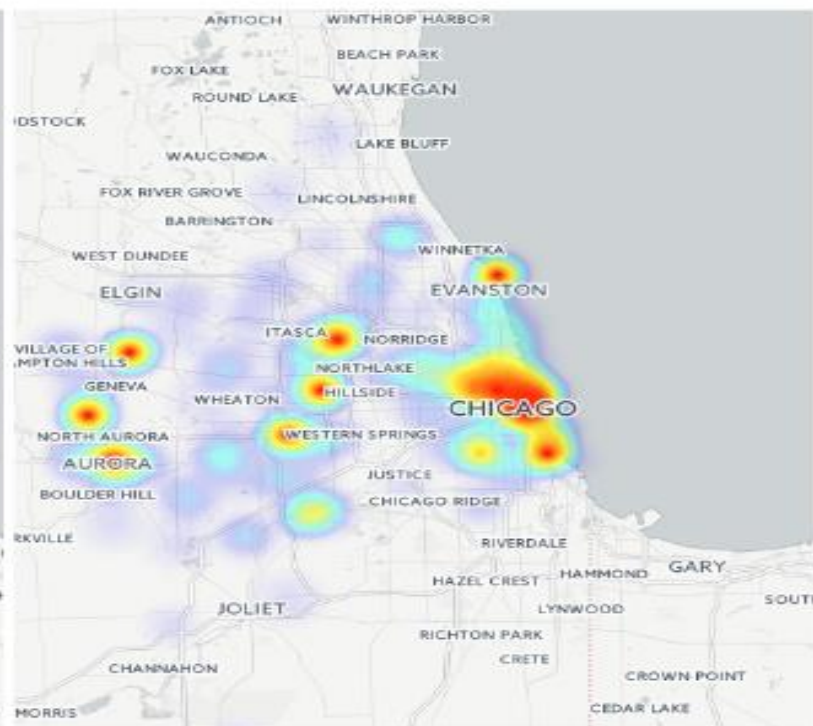
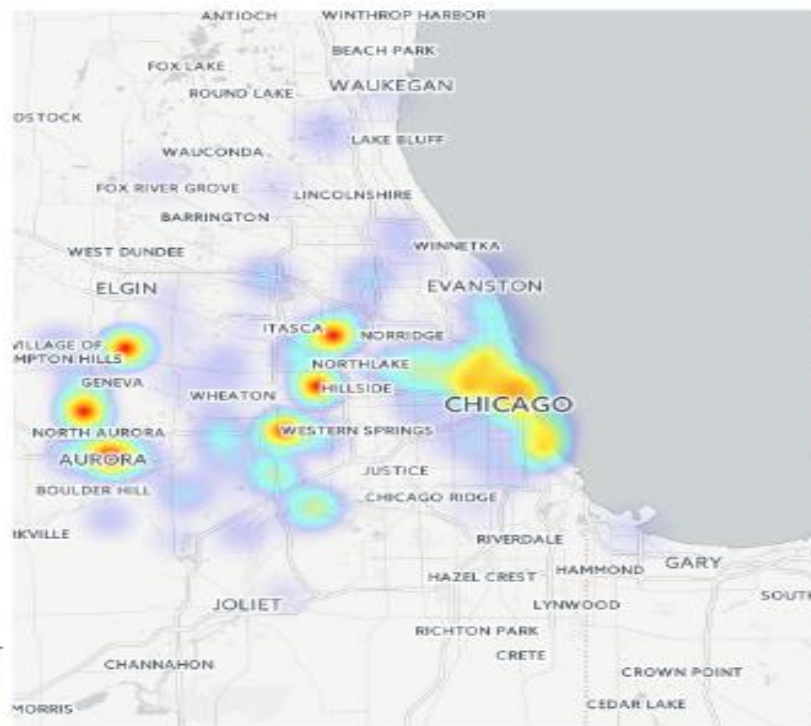


© OpenStreetMap contributors, © CARTO



# 38th INTERNATIONAL CONFERENCE ON HIGH ENERGY PHYSICS

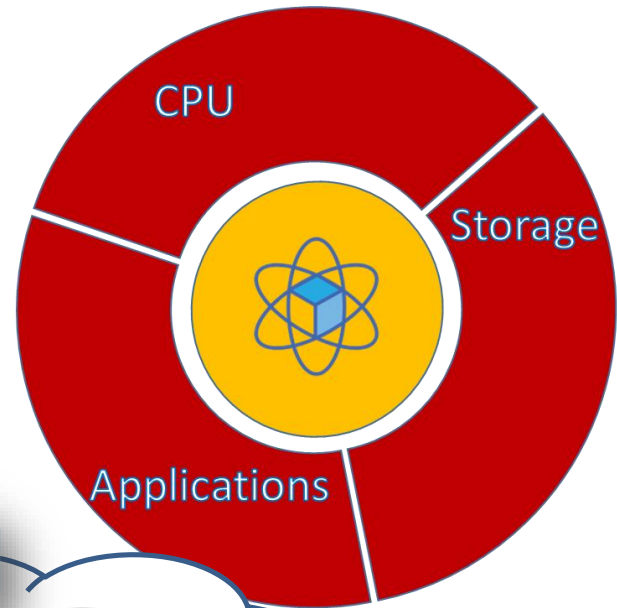
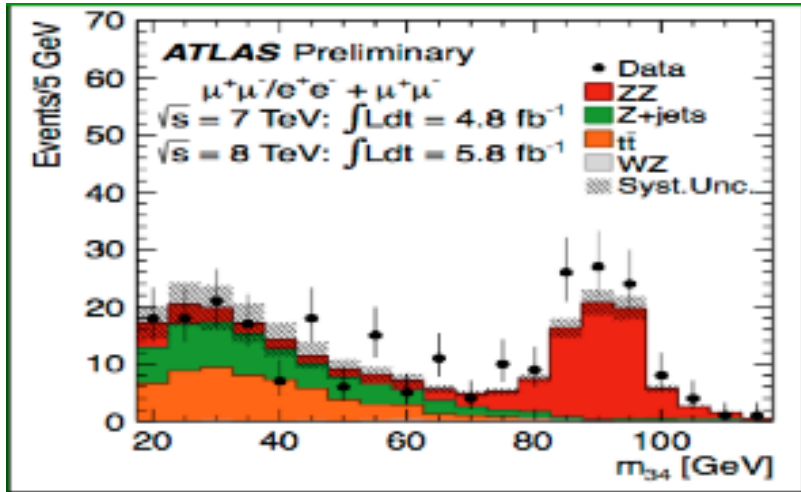
AUGUST 3 - 10, 2016  
CHICAGO





# Cloud analysis: SWAN project

with CERN Physics Department



Lots of activity in previous projects with several Russian groups, notably with V. Korenkov (JINR Dubna)

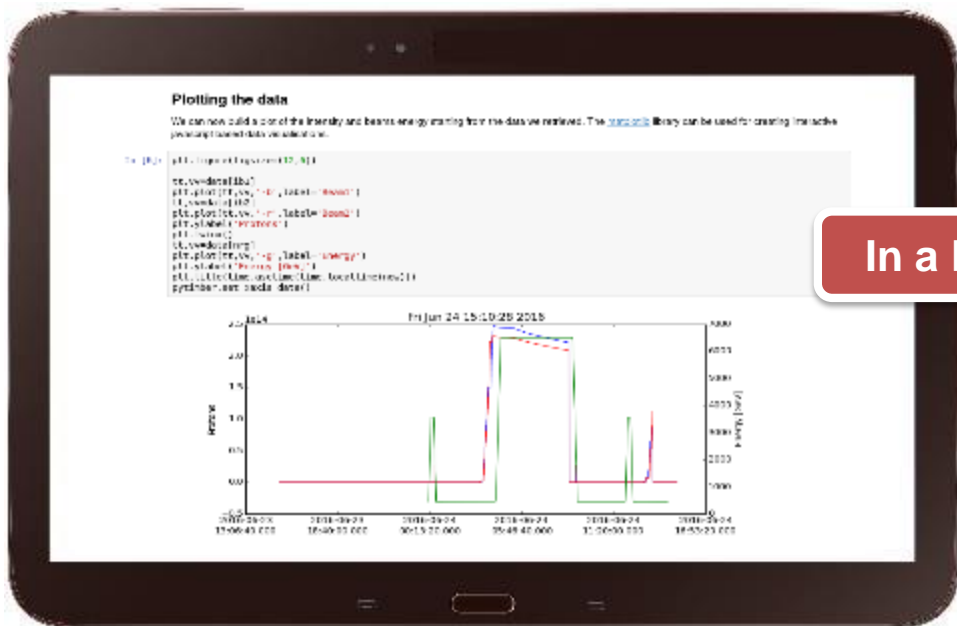


**ROOT is the CERN data analysis framework: <http://root.cern.ch>**



# Interface: The Notebook

**Jupyter Notebook:** A web-based **interactive computing** interface and platform that combines **code**, **equations**, **text** and **visualisations**



In a Browser



# Interface: The Notebook

Text

Code

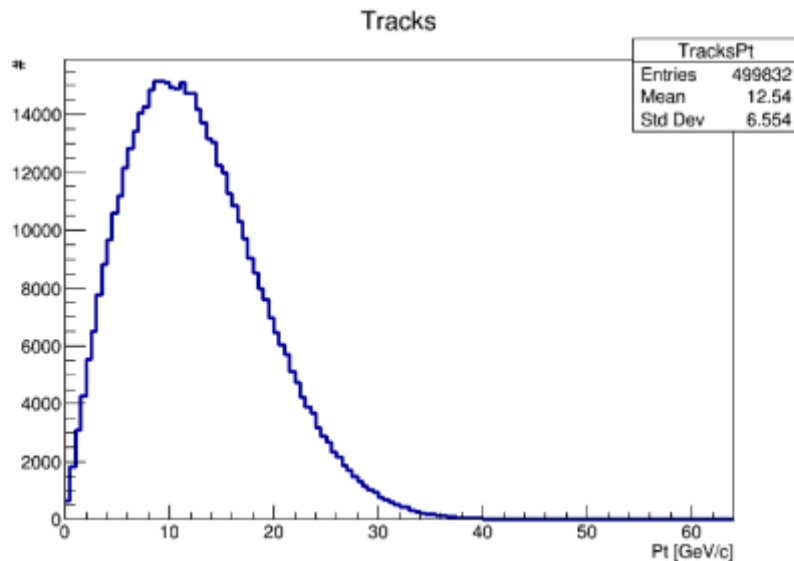
Graphics

## Access TTree in Python using PyROOT and fill a histogram

Loop over the TTree called "events" in a file located on the web. The tree is accessed with the dot operator. Same holds for the access to the branches: no need to set them up - they are just accessed by name, again with the dot operator.

```
In [1]: import ROOT

f = ROOT.TFile.Open("http://indico.cern.ch/event/395198/material/0/0.root");
h = ROOT.TH1F("TracksPt", "Tracks;Pt [GeV/c];#", 128, 0, 64)
for event in f.events:
    for track in event.tracks:
        h.Fill(track.Pt())
c = ROOT.TCanvas()
h.Draw()
c.Draw()
```





# CERNBox as Home

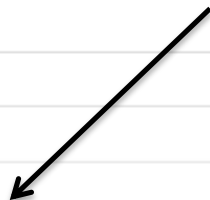
[Control Panel](#)[Logout](#)[Files](#)[Running](#)[Clusters](#)

Select items to perform actions on them.

[Upload](#)[New ▾](#)

<input type="checkbox"/>	<input type="text"/>	
<input type="checkbox"/>	ACAT 2016	
<input type="checkbox"/>	CHEP 2016	
<input type="checkbox"/>	cmsdata	
<input type="checkbox"/>	CSC	
<input type="checkbox"/>	ExampleDir	
<input type="checkbox"/>	IMLmeeting	
<input type="checkbox"/>	mylibs	
<input type="checkbox"/>	node_modules	
<input type="checkbox"/>	other	
<input type="checkbox"/>	ROOT-Primer	

Same content as in [cernbox.cern.ch](http://cernbox.cern.ch)





# Notebook Galleries



SWAN

Interactive Data Analysis, In the Cloud.

Home

Galleries

FAQ

Talks and Publications

Back

ROOT Primer

Accelerator Complex

Machine Learning

Apache Spark

## Basic Examples

This is a gallery of basic example notebooks: click on the

Open in SWAN

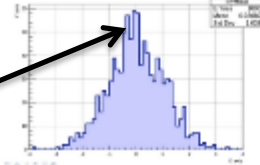
Many of the notebooks are ROOTbooks, based on the ROOT framework. To know more about ROOT, visit [root.cern.ch](http://root.cern.ch).

### Simple ROOTbook (Python)

Python is a really flexible and easy-to-use language for data analysis and visualization. In this notebook, we will use Python to analyze a dataset and create a histogram.

```
from ROOT import *
import sys
```

My Histogram



Open in SWAN

### Simple ROOTbook (C++)

ROOT is a powerful framework for data analysis and visualization. In this notebook, we will use ROOT to analyze a dataset and create a histogram.

```
#include <ROOT/ROOT.h>
using namespace ROOT;
```

My Histogram



Open in SWAN

Click on the image for a static visualisation

Example notebooks at [swan.web.cern.ch](http://swan.web.cern.ch)

Click on the blue ribbon to open them in SWAN!

### Fitting

```
from ROOT import *
import sys
```

This notebook shows how to fit a dataset with a Gaussian function using the ROOT framework.

Open in SWAN

### Simple I/O

```
from ROOT import *
import sys
```

This notebook shows how to read and write data using the ROOT framework.

Open in SWAN



# SWAN Use Cases

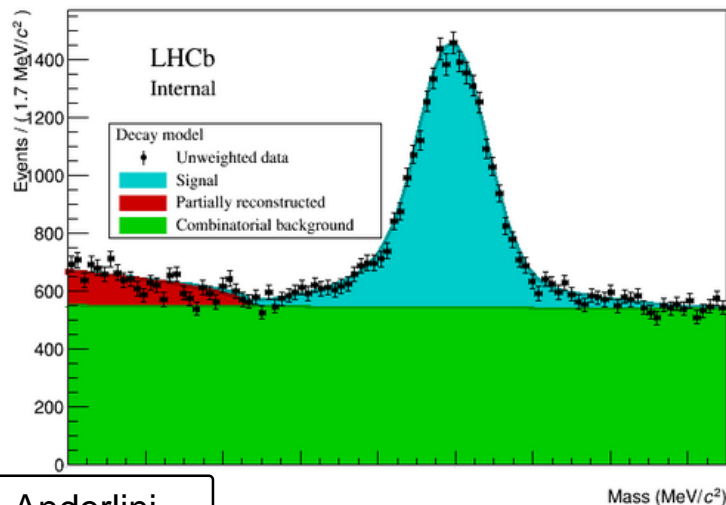
```
title = { "model": "Signal" , "pdfBkg" : "Partially reconstructed" , "cmbBkg": "Combinatorial background"}

for (component, color) in [ ("model",kCyan), ("pdfBkg",kRed), ("cmbBkg",kGreen)]:
    model.plotOn (frame, LineColor(color+2) , DrawOption('L') , Components(component), LineWidth(5))
    model.plotOn (frame, FillColor(color+1) , DrawOption('F') , Components(component), LineWidth(0), Name("P"+component)
    ))
    leg.AddEntry ( frame.findObject ("P"+component), title[component] , "P" )

data.plotOn ( frame, MarkerColor ( ROOT.kBlack ) )
frame.Draw()
Graphics().lhcbMarker(0.2,0.8, "Internal")

leg.Draw()
ROOT.gPad.Draw()
```

Results coming  
from real data!  
(published now)



## Physics Analysis

Rare B meson decay in LHCb

- Read data from EOS
- Setup complex fit
- Document and inspect results



- SWAN as platform for outreach
  - Introductory course about experimental HEP for future high school teachers

## Particle open data teaching (Hiukkasfysiikan avoin data opetuksessa)

**Lähdetäänpä tutkimaan!**

Lähdetään seuraavaksi tarkastelemaan, miten pseudorapiditeetin vaikutus mittatarkkuuteen voidaan havaita CMS-ilmaisimen keräämän oikean datan avulla. Käytetään CMS:n vuodelta 2011 kerättyä dataa [1], josta on valittu 10851 törmäystapahtumaa (events) tiedostoon "Zmumu\_Run2011A\_massolla.csv". (Karsinta on suoritettu koodilla, joka on avoimesti saatavilla osoitteessa <https://github.com/tpmccauley/dimuon-filter>.)

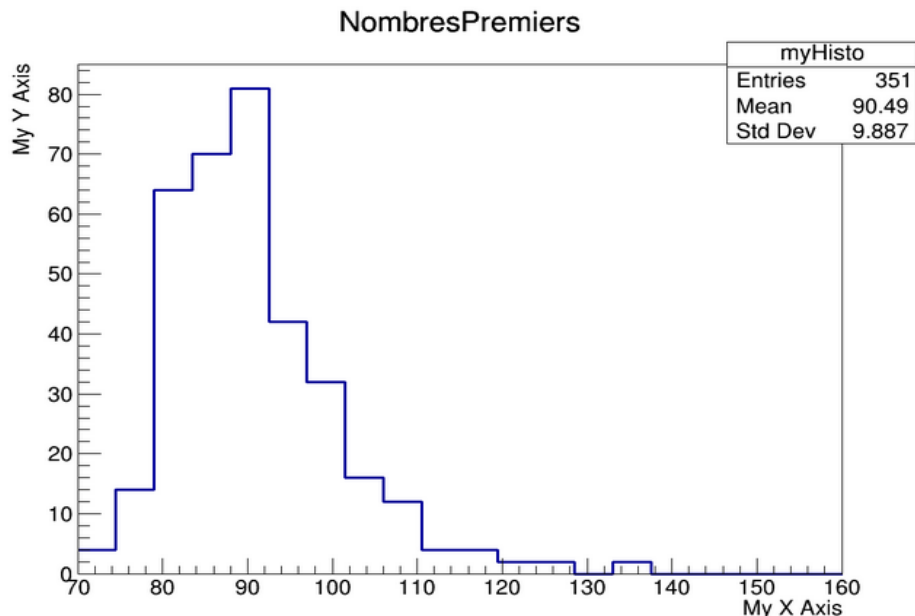
Tiedostoon on valittu niitä törmäystapahtumia, joissa syntynyt Z-bosoni on hajonnut myoniksi  $\mu^-$  ja antimyoniksi  $\mu^+$ . Ilmaisim on havainnut nämä myonit ja mitannut niiden liikemäärät.

**P. Rikkilä**



```
In [138]: import ROOT
htemp = ROOT.TH1F("myHisto", "NombresPremiers;My X Axis;My Y Axis", 20, 70, 160)
for i in range(len(data)):
    d = data[i][0]
    htemp.Fill(float(d))
c = ROOT.TCanvas("myCanvas", "myCanvasTitle", 1024, 768)
htemp.Draw()
c.Draw()

TROOT::Append:0: RuntimeWarning: Replacing existing TH1: myHisto (Potential memory leak).
TCanvas::Constructor:0: RuntimeWarning: Deleting canvas with same name: myCanvas
```



## Mano S. (14 years old), K12 student

- Approaches programming for the first time
- Verifies numerically what he learned at school
- Shares results with his supervisor and classmates

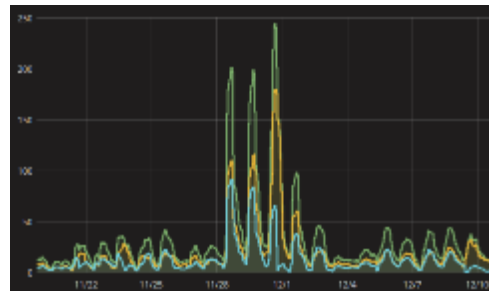


- Practical Statistics for Particle Physics Analyses

<https://indico.cern.ch/event/545212/>

- CERN Summer Student Program: ROOT

<https://indico.cern.ch/event/536772/>



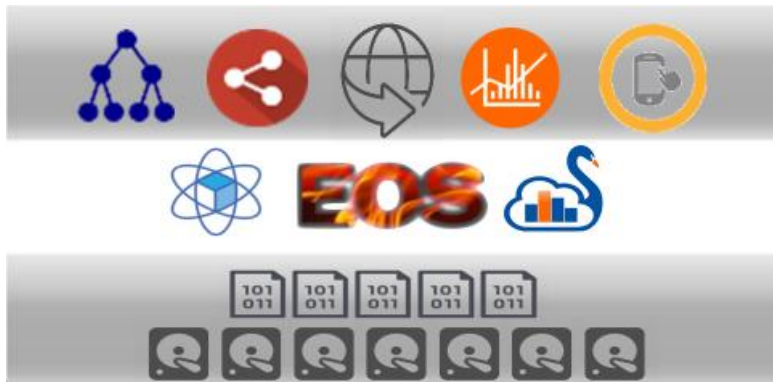
- CERN School of computing: Parallelization lectures

<http://indico.cern.ch/event/502875/>

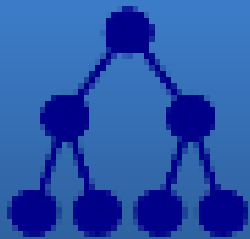
- Data Science @ LHC Workshop, Multivariate analysis tutorial

<http://indico.cern.ch/event/395374/>

# Summary



- **Solid foundations**
  - 200 PB LHC disk infrastructure
    - Steadily growing!
  - HEP collaborations
- **Strategic partnership**
  - HEP computing evolution
  - Cloud storage enables new use cases
    - and new ways to work and to collaborate





[www.cern.ch](http://www.cern.ch)

