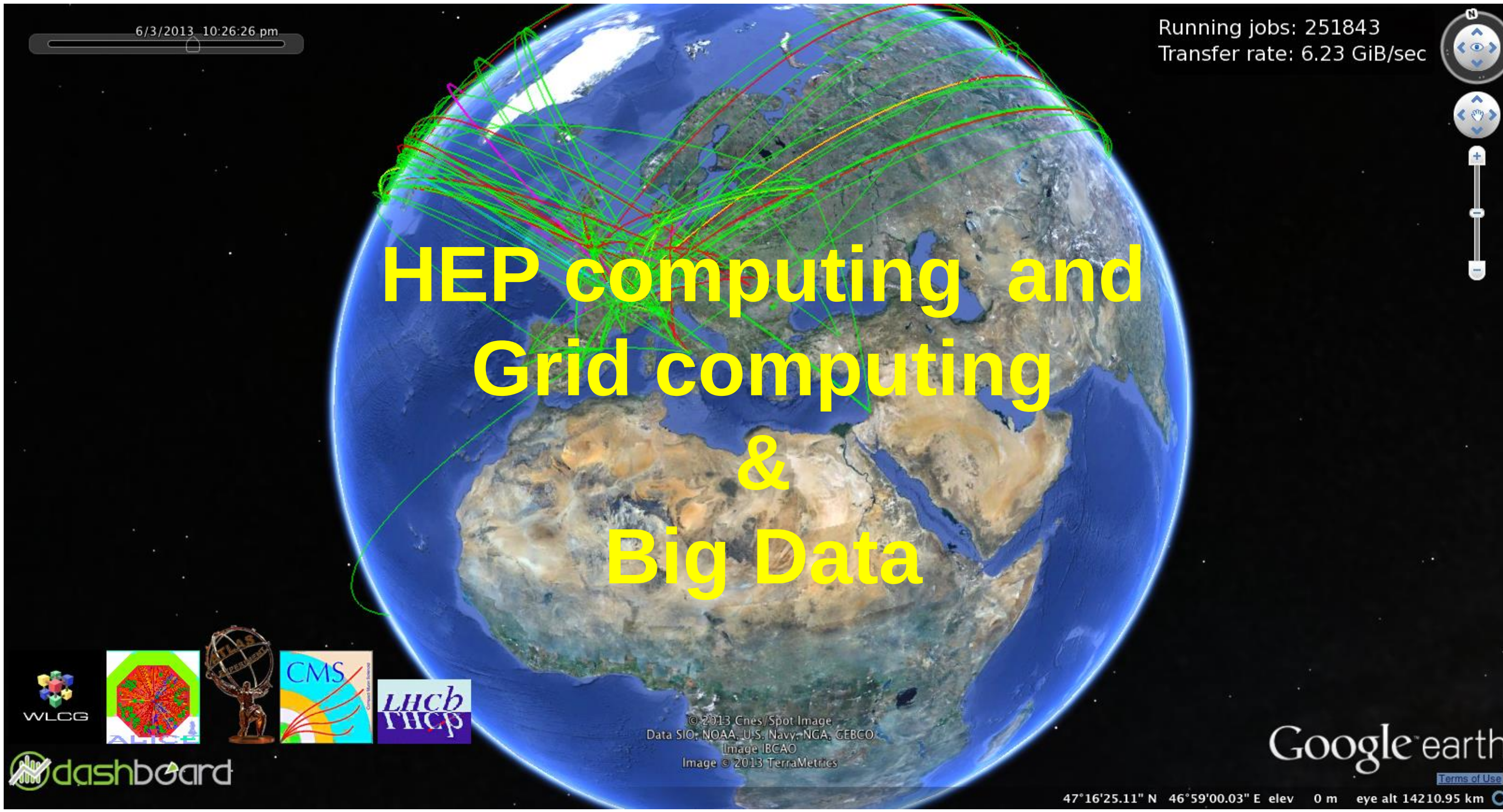
CERN IT Department

May 11th 2014
CC visit: Uni Trieste and Uni Udine



HEP computing and Grid computing & Big Data

Massimo Lamanna/CERN

IT department - Data Storage Services group

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

My objectives for today!

CERN IT Department

Why High-Energy Physics (HEP) needs computing?

Computing: CPU cycles and Data Storage

Lots of CPU, disks and tapes!

How can we deliver enough computing to HEP?

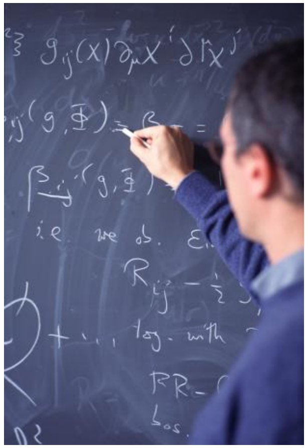
Any relevance beyond HEP?

Who is CERN IT? Staff, fellows and students!

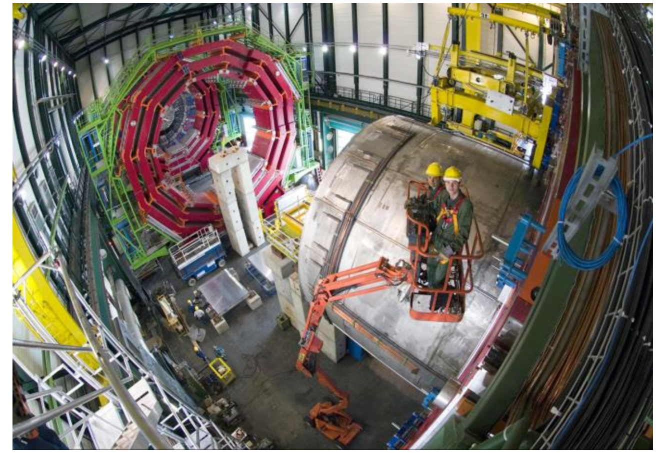
Have a look on the CERN CC

And of coursetrying to answer all your questions!

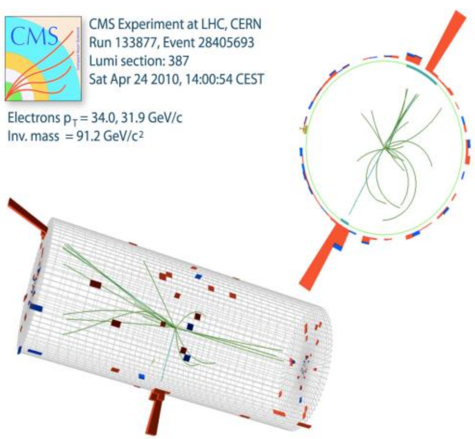
What are all these computers/disks/networks... for?

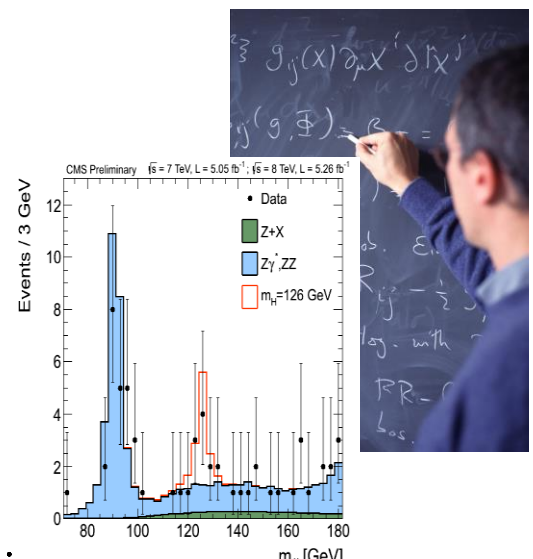


Theory and
(Detector) simulation



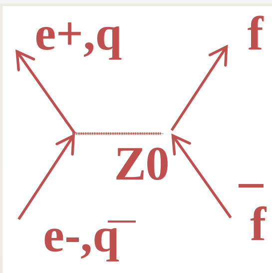
Data acquisition and
processing

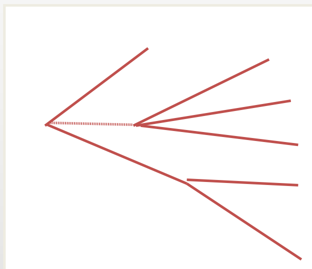


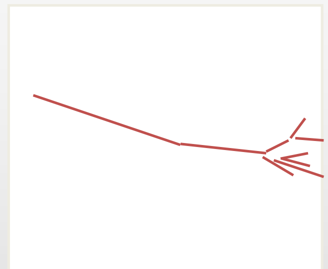


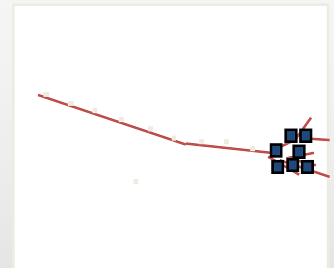
Data analysis and
Results interpretation

From Physics to Raw Data









2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

Basic physics

Fragmentation, Decay

Interaction with detector material
Multiple scattering, interactions

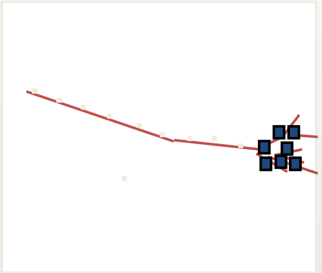
Detector response
Noise, pile-up, cross-talk, inefficiency, ambiguity, resolution, response function, alignment, temperature

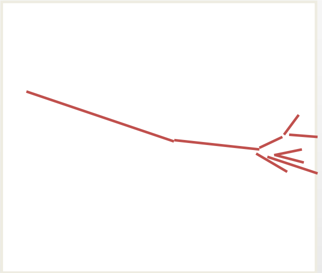
Raw data (Bytes)

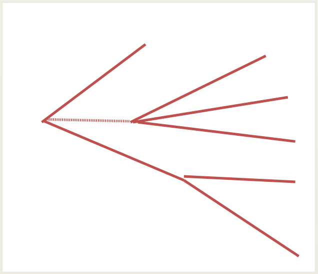
Read-out addresses, ADC, TDC values, Bit patterns

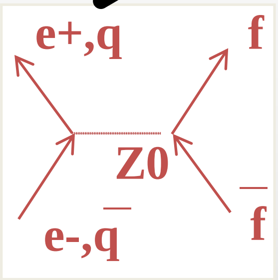
From Raw Data to Physics

2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142









Raw data

Convert to physics quantities

Detector response
apply calibration, alignment,

Interaction with detector material
Pattern, recognition, Particle identification

Fragmentation, Decay
Physics analysis

Basic physics

Results

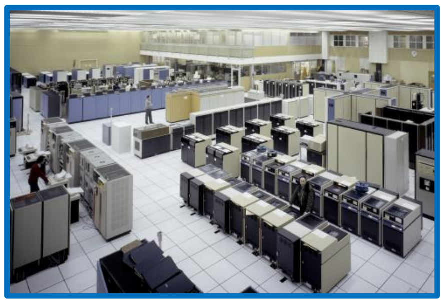
Reconstruction

Analysis

Simulation (Monte-Carlo)

(Innovation in) Computing in High-Energy Physics

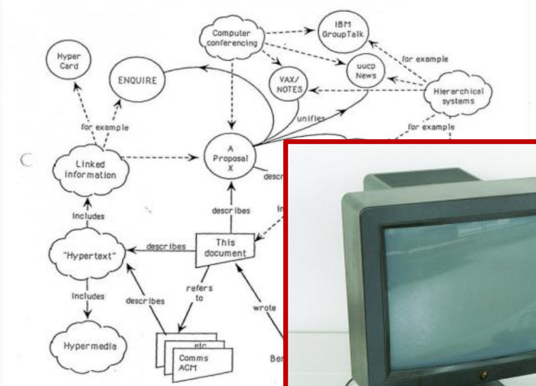
- Demanding science → Demanding computing

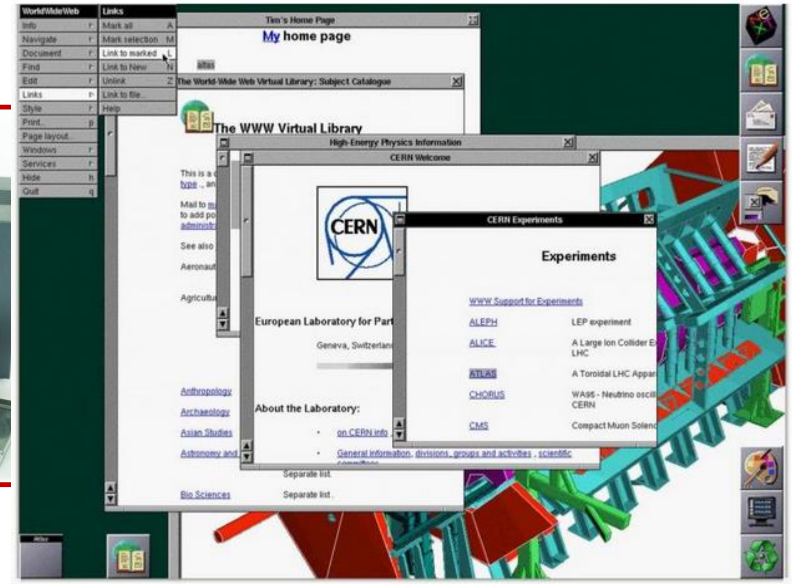


- Power usage

- Innovation
  - Web invention
  - Grid computing (LHC Computing Grid)

World Wide Web born at CERN
25 years ago





Screenshot of the original NeXT web browser in 1993 (Image: Berners-Lee/CERN)

Invented by/in/for the HEP community
Now at the hearth of our lives

License model: http://home.web.cern.ch/topics/birth-web/licensing-web

WLCG (Worldwide LHC Computing Grid)

- The Worldwide LHC Computing Grid (WLCG) collaboration was set up in 2002 for the Large Hadron Collider (LHC) physics programme at CERN. It links up national and international grid infrastructures.

- WLCG mission is to provide global computing resources to store, distribute and analyse the ~25 Petabytes (25 million Gigabytes) of data annually generated by the Large Hadron Collider (LHC) at CERN

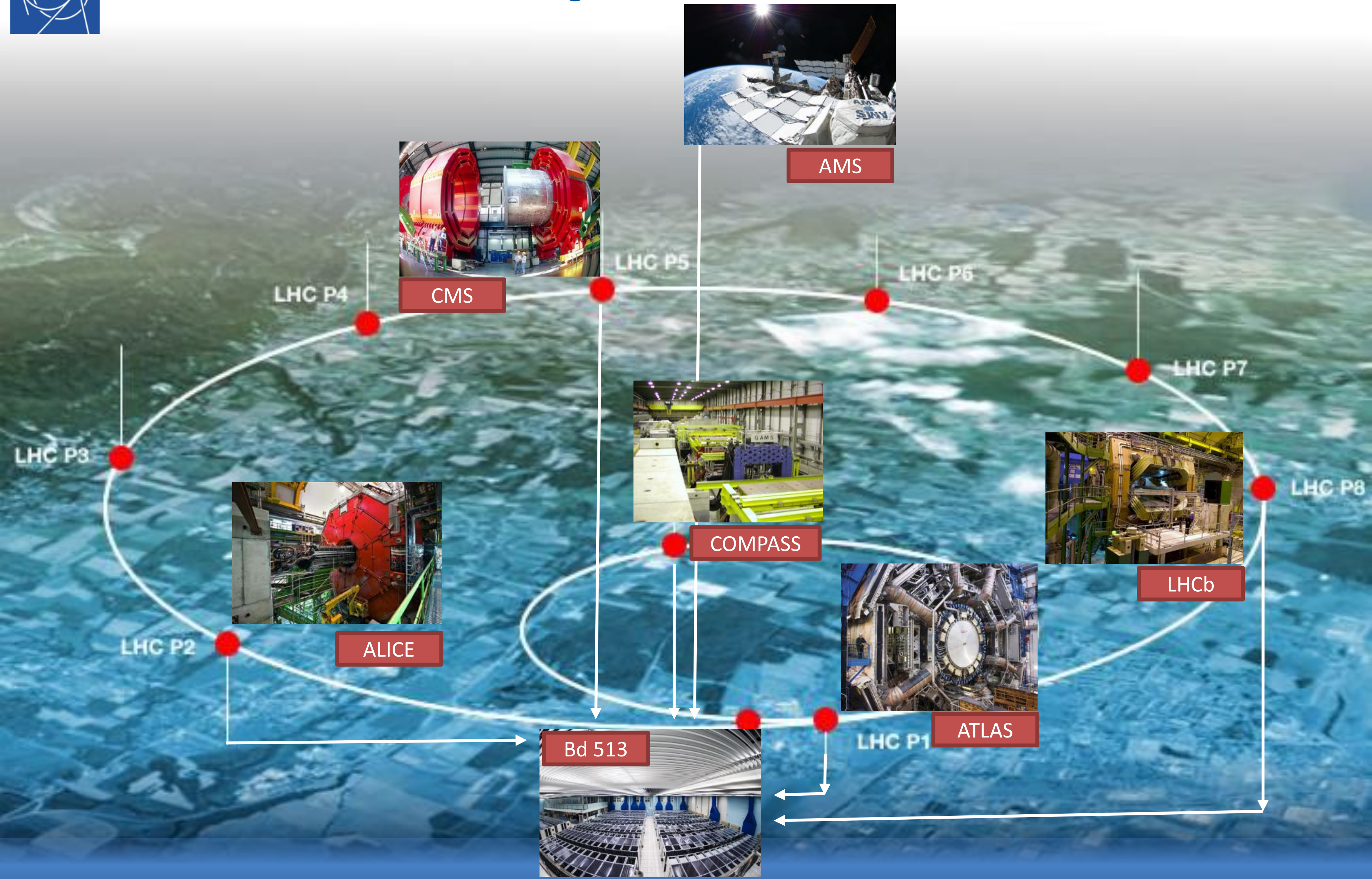
- Challenging points:
  - Large data volumes
  - Large computing needs (simulation and data reconstruction and analysis)
  - Very large community (including large number of computer centres and other institutes not necessary part of a LHC experiment)
  - Aiming to total:CERN resource ratio of the order of 5:1 or more (20%)

- Historical notes:
  - Computing activities in the late 1990 (e.g. Monarc project: http://monarc.web.cern.ch/MONARC/)
  - “Grid” concepts (I. Foster and K. Kesselman) matched with our approach
  - Distributed approach “matches” our distributed user communities
  - Participation to multiscience Grid projects (notably EGEE 2004-2010)

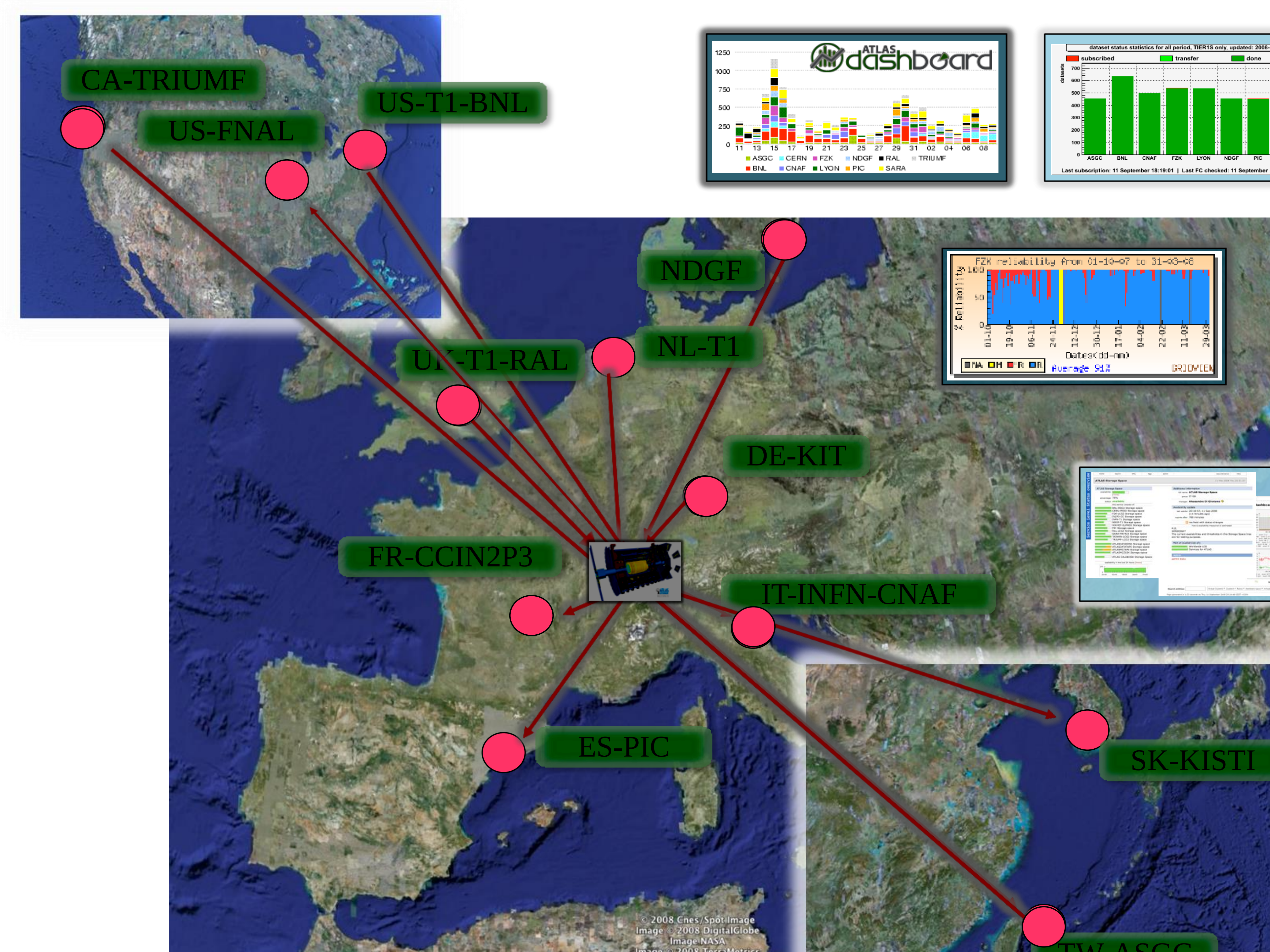
- WLCG now operates a distributed infrastructure with 200+ major cooperating computer centres worldwide and provides the foundation for all data processing activities. WLCG delivers access to LHC data and computing power for extensive data analysis to virtually every LHC physicists.

- Several tens of PB per year of data files are collected and distributed: experimental data (collected at rates up to several GB/s) supplemented by derived data (processing and filtering) and simulation data (generated on the same infrastructure). Depending on data type and usage, files are replicated on disk farms and tape repositories to ensure best access and durability.

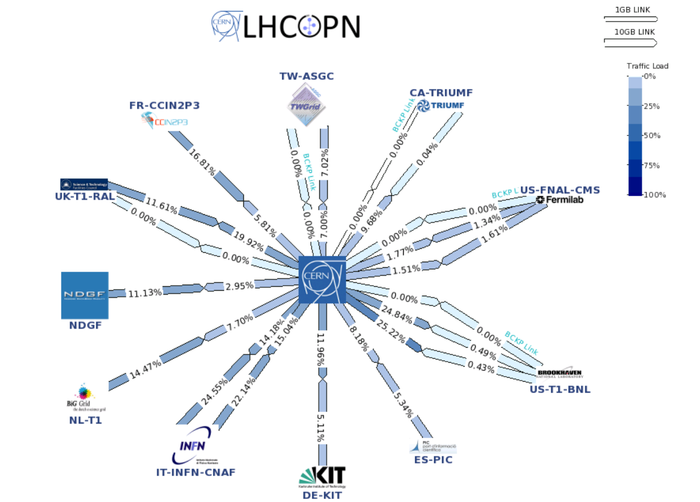
CERN Central Data Recording

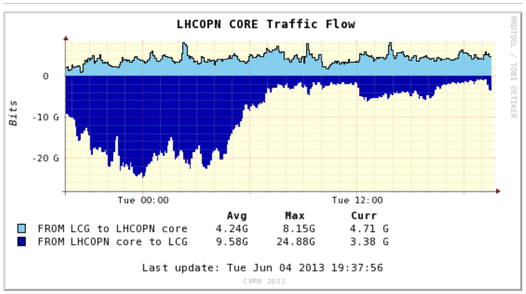


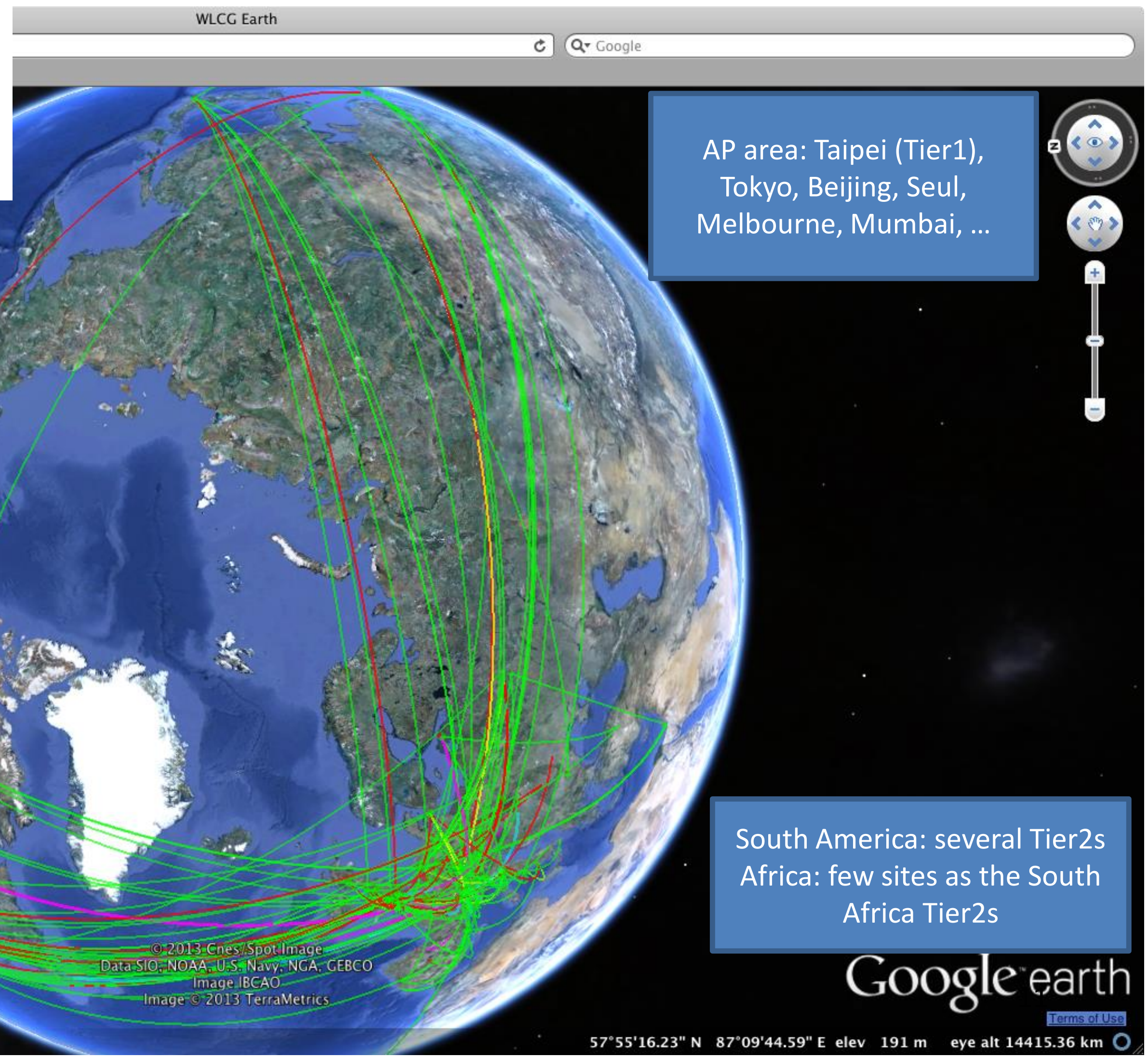
Typical data rates 100-1000 MB/s aver multiple 10 Gb fibres.
Large (and final) filtering takes place at the experiment (trigger filter farms). All data are then kept (RAW data)


CA-TRIUMF
US-T1-BNL
US-FNAL
NDGF
NL-T1
UK-T1-RAL
DE-KIT
FR-CCIN2P3
IT-INFN-CNAF
ES-PIC
SK-KISTI
ATLAS dashboard
ASGC CERN FZK NDGF RAL TRIUMF
BNL CNAF LYON PIC SARA
dataset status statistics for all period, TIER1S only, updated: 2008-
subscribed transfer done
ASGC BNL CNAF FZK LYON NDGF PIC
Last subscription: 11 September 18:19:01 | Last FC checked: 11 September
FZK reliability from 01-10-07 to 31-03-08
% Reliability
Dates(dd-mm)
Average 91%
GRIDVIEW
© 2008 Cnes/Spot Image
Image © 2008 DigitalGlobe
Image NASA

Intercontinental links **(data and jobs)**



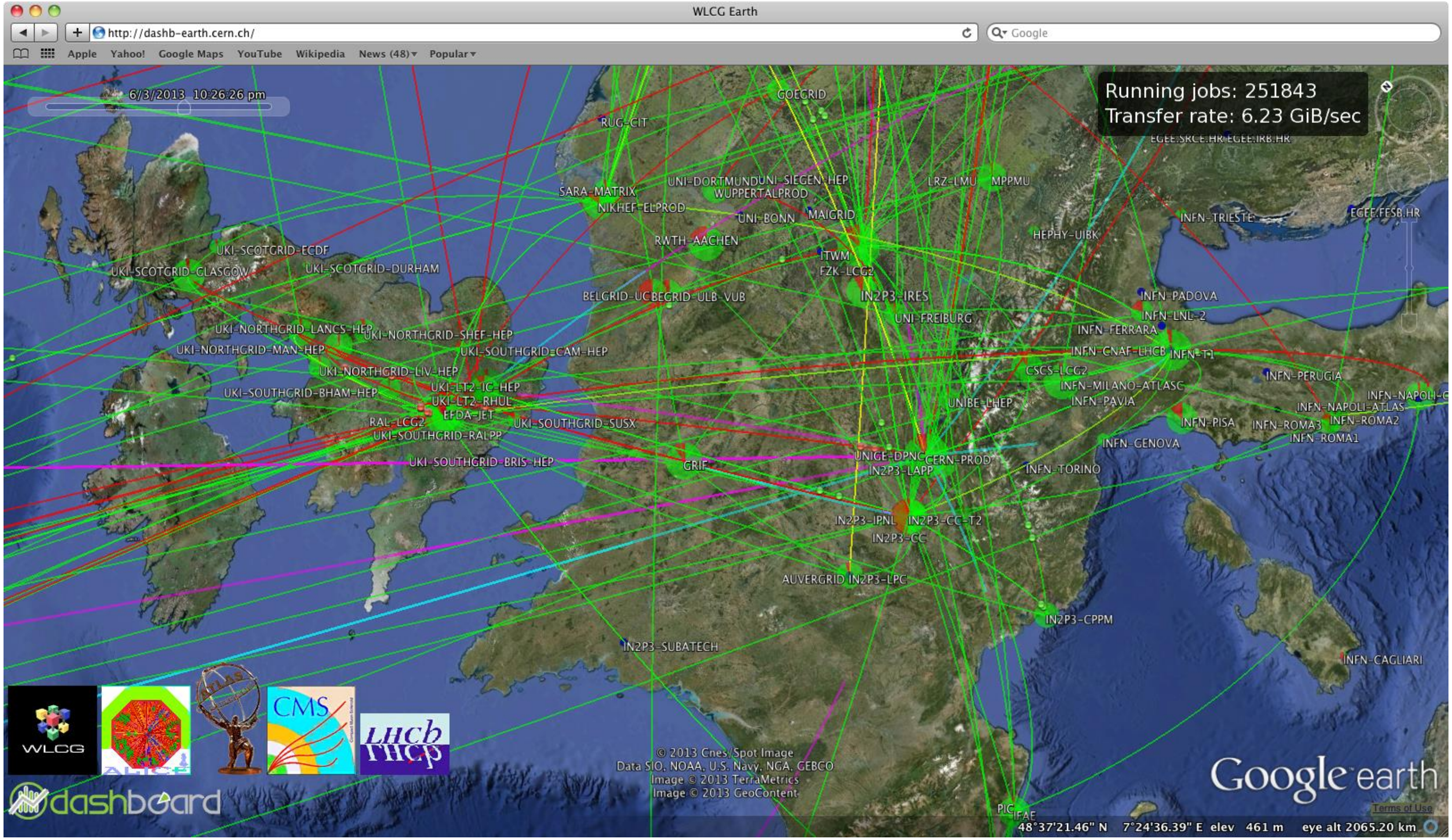




AP area: Taipei (Tier1), Tokyo, Beijing, Seul, Melbourne, Mumbai, …

North America: BNL and Fermilab (US Tier1), Victoria (Canada Tier1) and many Tier2 like Stanford, MIT, Wisconsin, Argonne, …

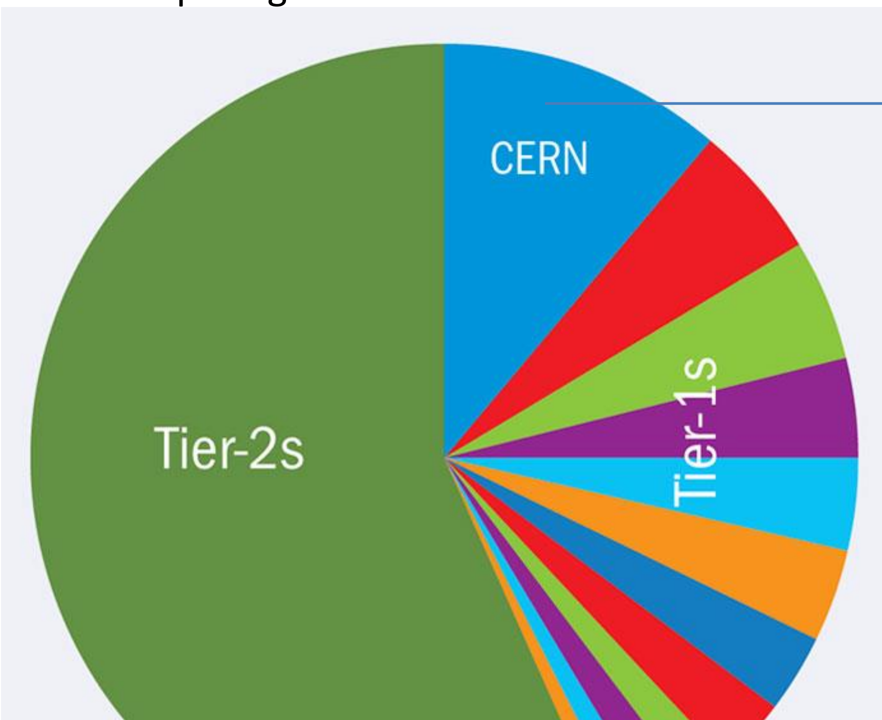
South America: several Tier2s
Africa: few sites as the South Africa Tier2s



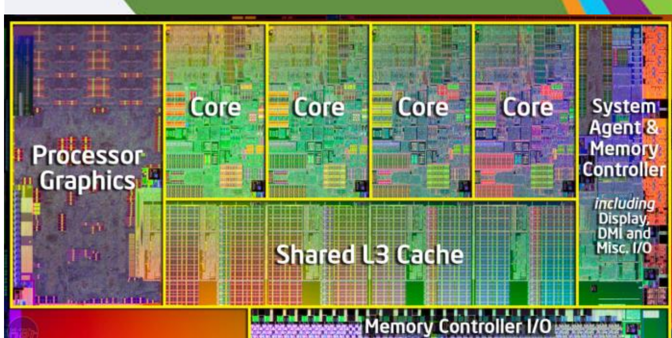
- Data distribution (e.g. CERN distributes RAW data to Tier1s)
  - 12 centres on-line
  - Tier1s being built in Moscow area (@Dubna and @Kurchatov Institute)
- (Large) processing campaigns on preplaced data (e.g. Reprocessing)
- Download small sample for local analysis? … but
  - Scale out with grid jobs (user executable dispatched where data are)
    - Using file parallelism (data set → list of files → list of independent jobs)
- Federating storages
  - Recall data `on the flight’

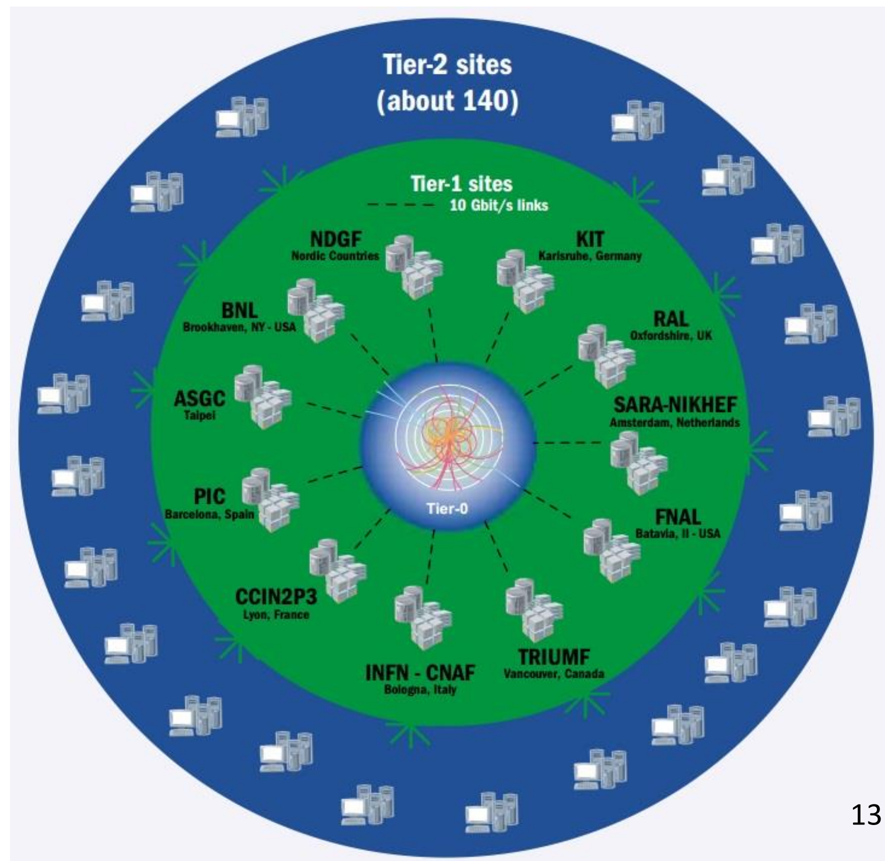
Grid at work (computing power)

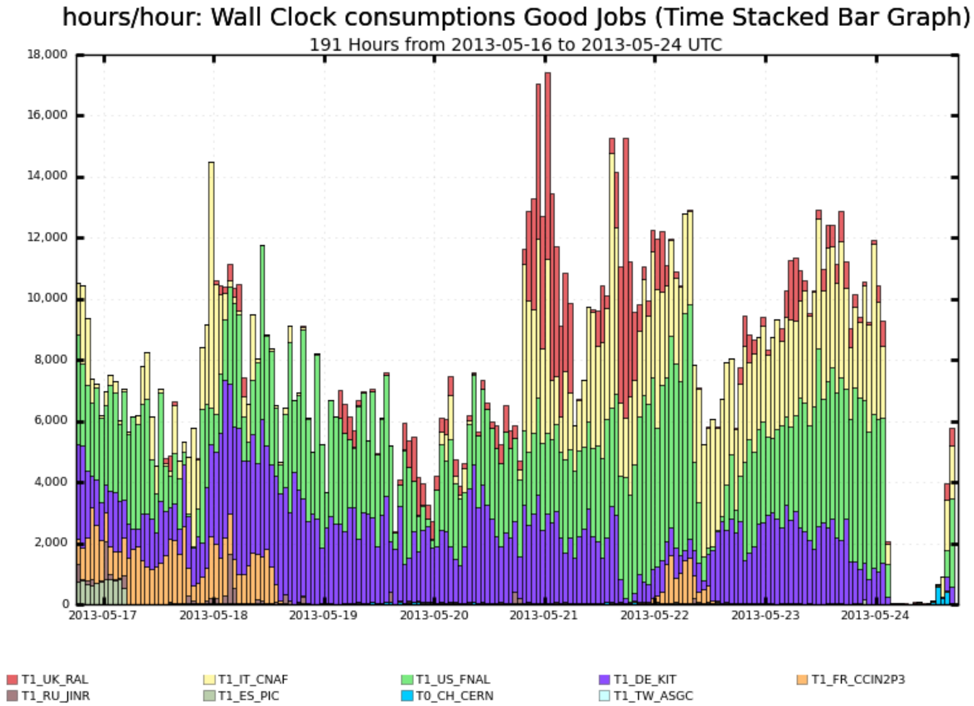
CPU contributions across the worldwide LHC Computing Grid

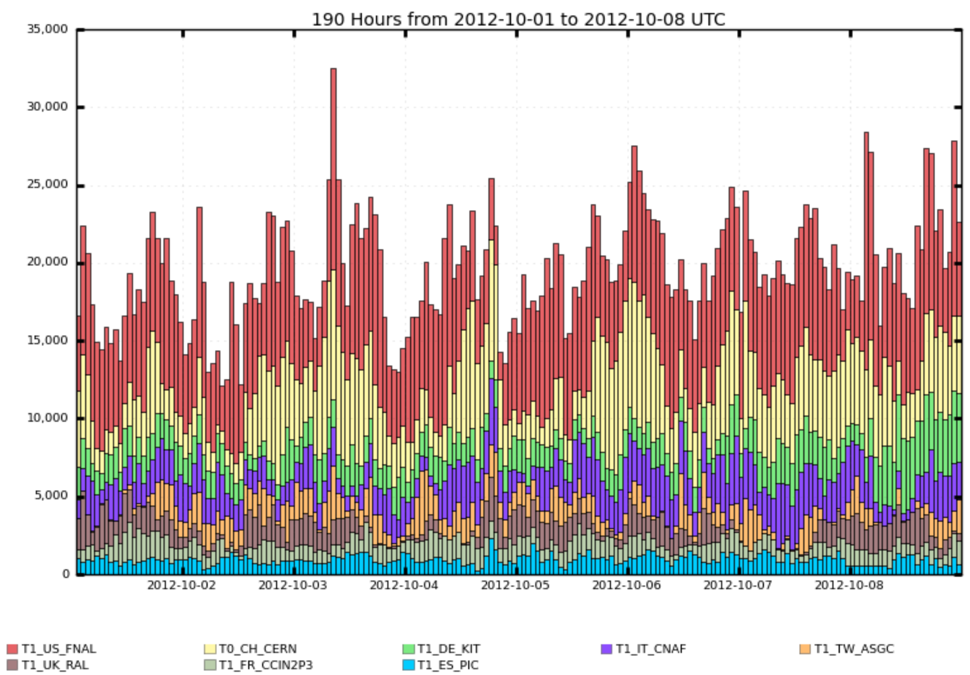


See next slides: ~80,000 cores







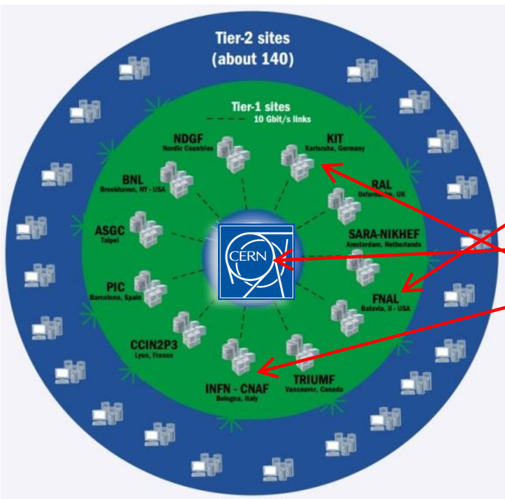


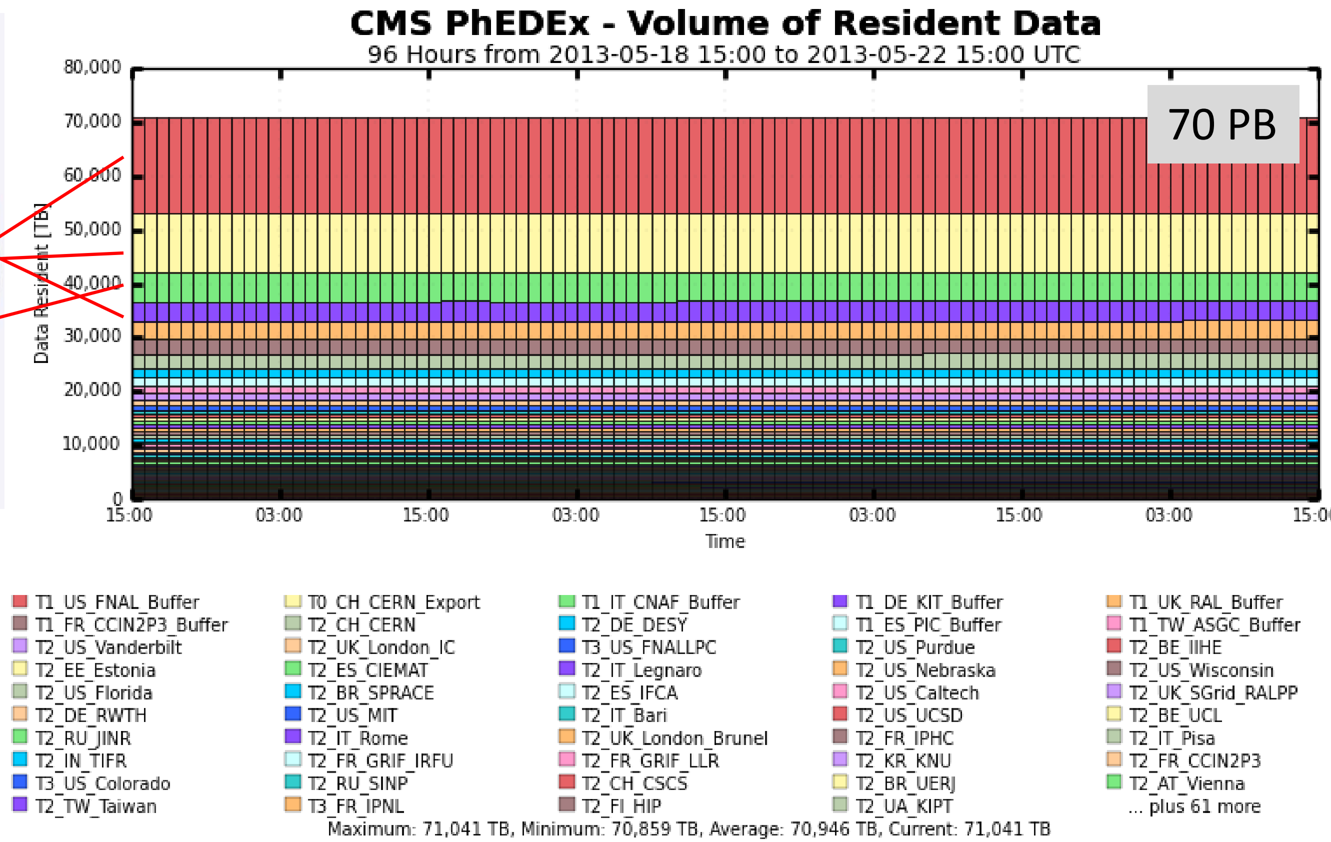
CMS processing: wall clock consumption

- Tier0 and Tier1 processing
  - Top/Bottom: particle collisions off/o
  - Sizeable even if no data taking: continuous reprocessing
- Reconstruction activities
  - RAW → reconstructed objects
  - Organised processing
  - Output for physicists analysis
    - They can access RAW data if needed
    - Final analysis more efficient on the files containing the “reconstructed objects”

The grid never sleeps…

Data on the Grid (CMS experiment)



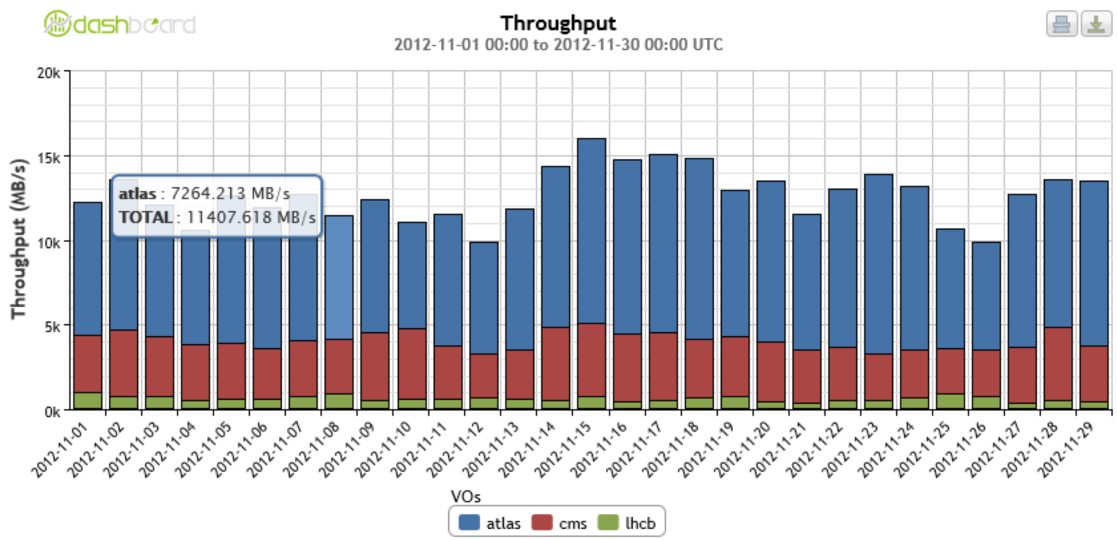


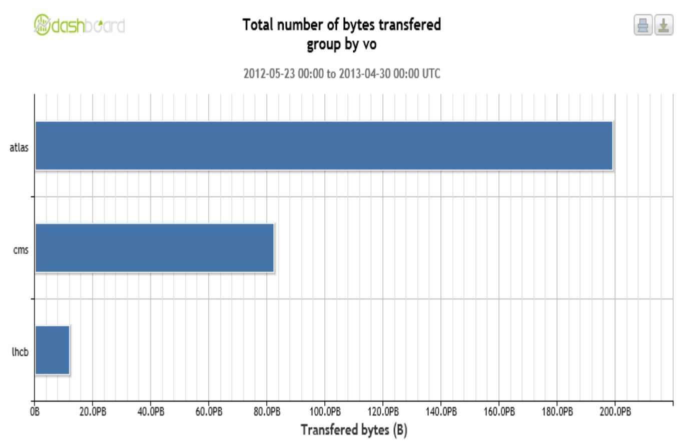
Much more data than RAW!

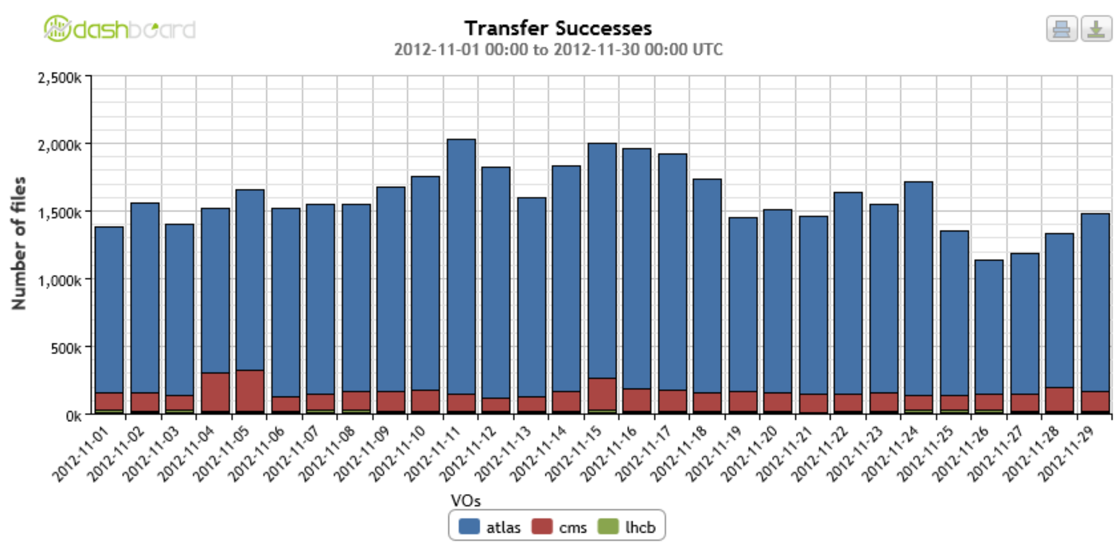
- Primary data from collisions
  - + their replicas
- Different levels of reconstructed and filtered data sets
  - + their replicas

LHC run 1 (2012) data (transfers)

- Data taking is part of the full picture
  - Primary data
- Replication, multi-step analysis and simulation “generates” more data
  - And network traffic (data placement)

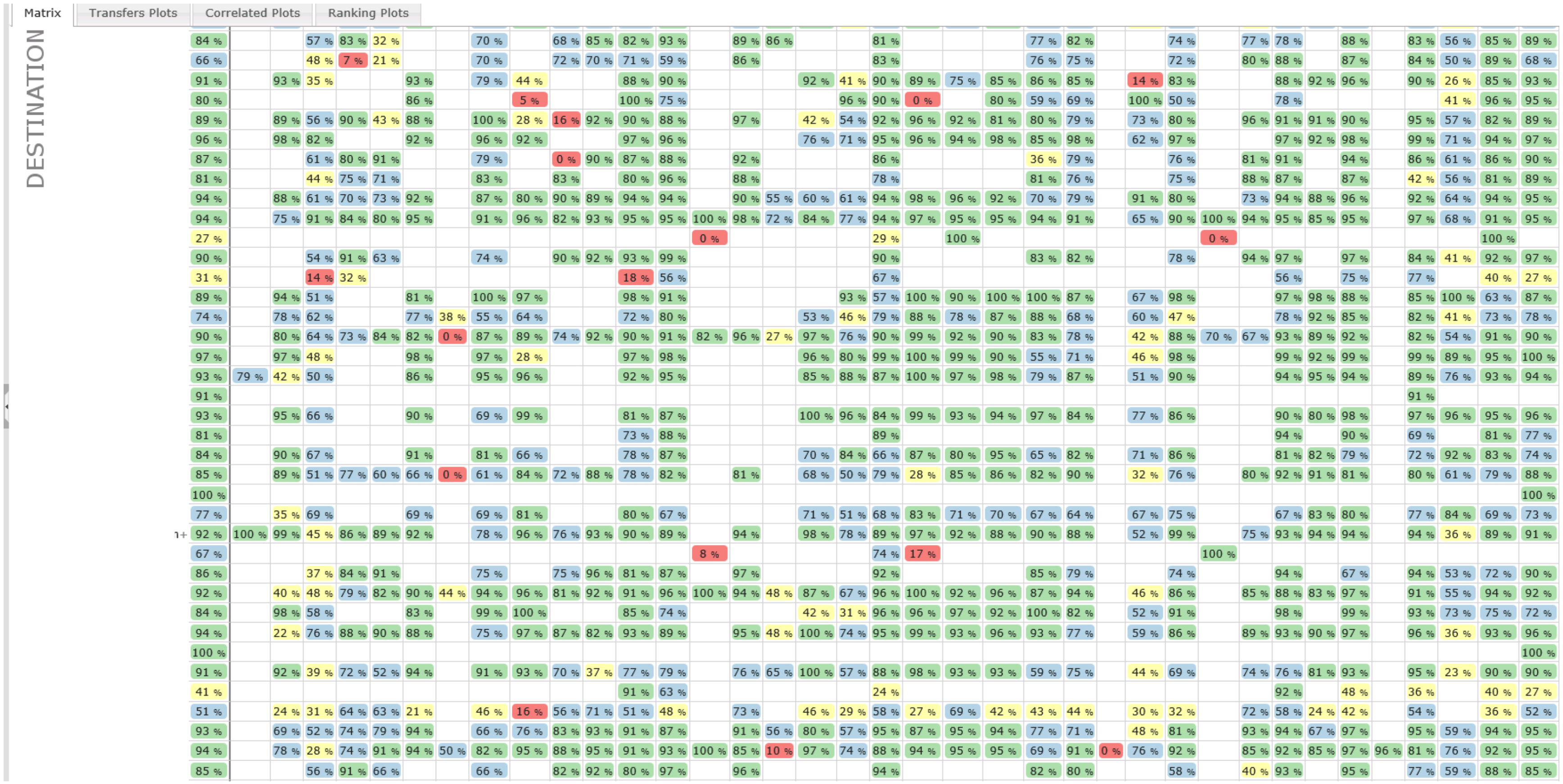




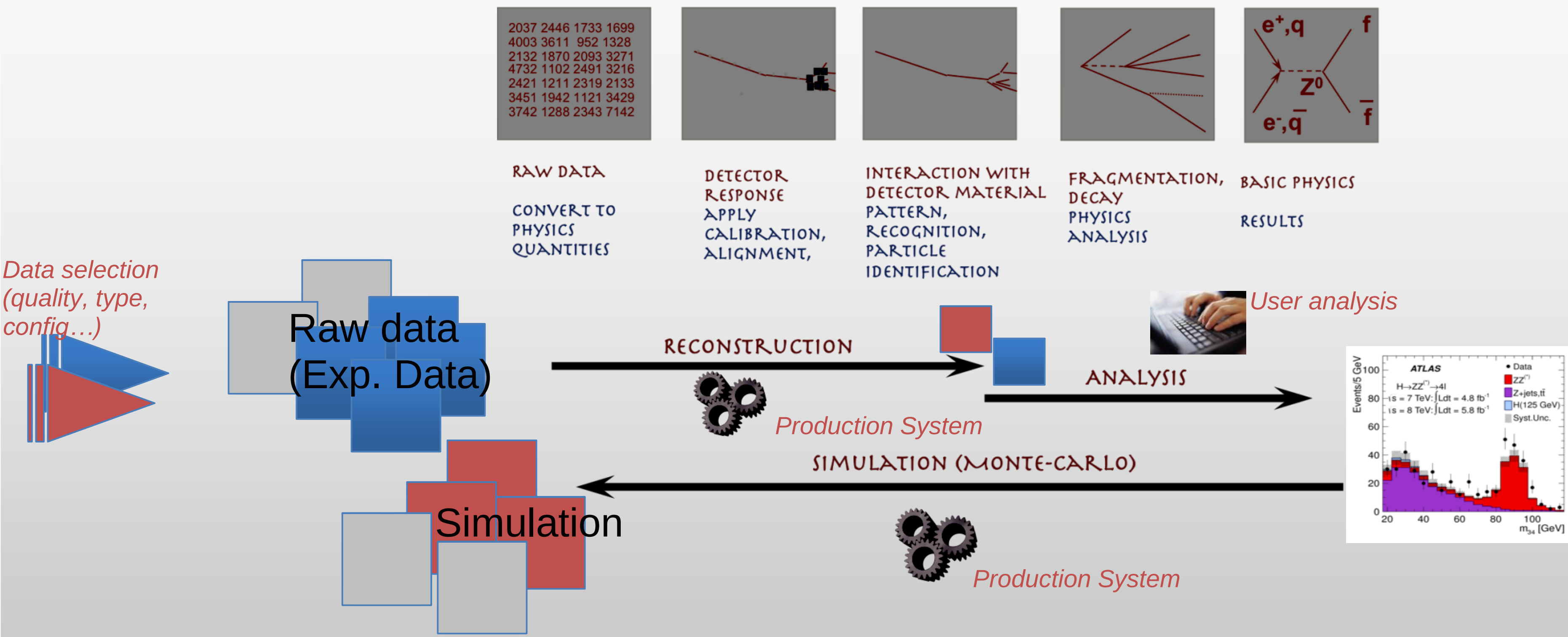


Data handling is an operational challenge

Source destination correlation matrix



Analysis flow (user view)



But how this is done *in practice*? Of course we need CPUs, disks, networks etc..
The problem is orchestrating hardware resources, software and humans :)

"All" data are stored in files (aggregated as "datasets" = collections of files). Only a small fraction of data in real DBs (e.g calibrations). This is one characteristics of HEP computing.

Toy model of user analysis:

2011 data; quality; run conditions… to be analyse it with myprog.exe

- Central services
- Regional services
- Tier2/3 (data on disk + CPU)
- Execution site

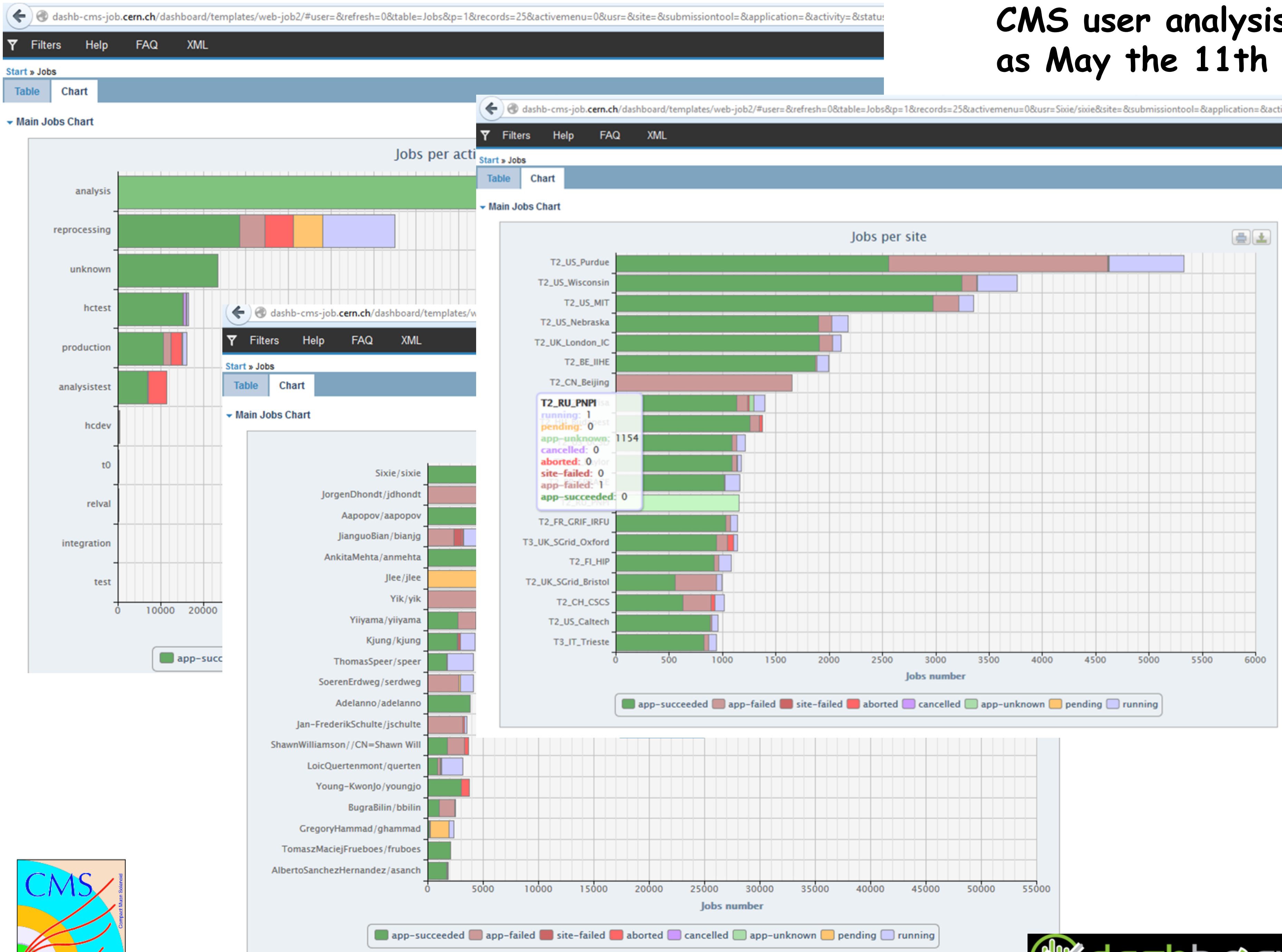
$[f_1, f_2, \ldots f_n]$

Federation/location: f_j is in $[site_1, site_2, \ldots site_m]$

Executed myprog.exe(/storage/data/2011/run4/f_j)

In general (Execution site) is from $[site_1, site_2, \ldots site_m]$ but not always(flexibility, robustness, efficiency)

CMS user analysis as May the 11th



CCF at CERN

Part of a larger farms

Eventually decommissioned

Still in use

ACID in Trieste (INFN)

Still in use

Close collaboration with CERN IT

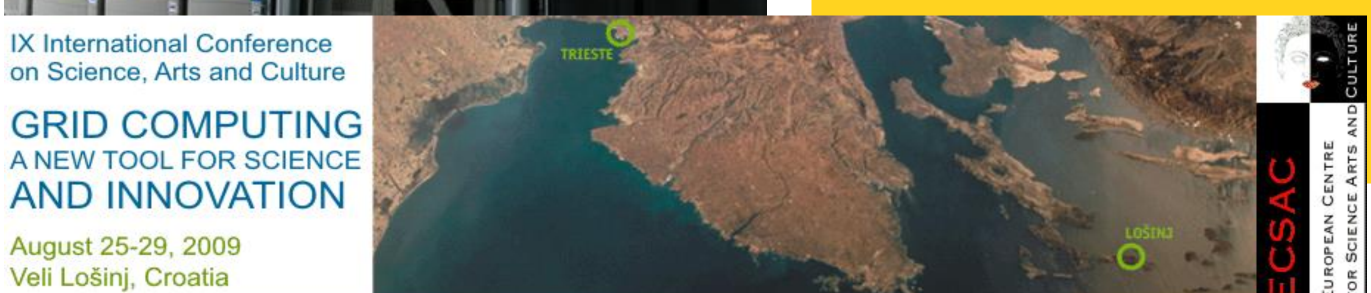
COMPASS was the first/among the first to:

- Migration to Linux PC farms
  - CHEP2000
- CCF at CERN
- ACID in Trieste

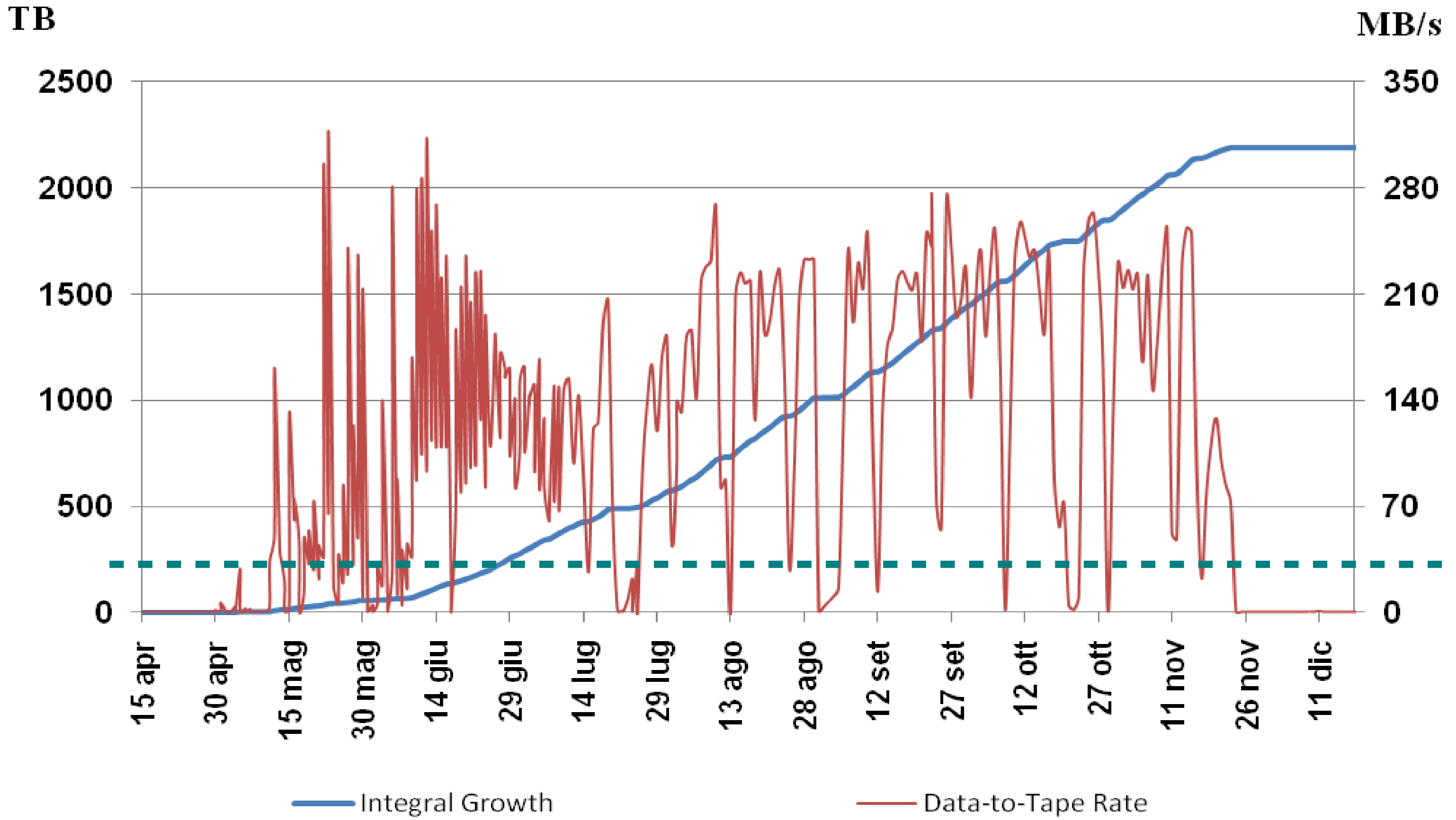
- Use of Objectivity/DB at the 10^2 TB scale

- First production user of CASTOR

- Deliver and operate CORAL (C++ framework) for data processing
- And the algorithmic parts, notably the RICH reconstruction



Run 2010 - Collected data to CASTOR



Design
Value (35 MB/s)

5th of May 2014 CASTOR statistics (tape files only)

Total size: 87.337 PB
Total number of files: 262.601 M

Top-6 size (by group)
30.436 PB with group ATLAS (id 1307)
17.266 PB with group CMS (id 1399)
13.888 PB with group COMPASS (id 1665)
9.616 PB with group ALICE (id 1395)
6.585 PB with group LHCb (id 1470)
1.839 PB with group zf (id 1018)

Zoom to CERN

https://sls.cern.ch/sls/dcbynum.php

Service Level Status overview

Home | Search | KPIs | Data Centre by numbers | Tags | Admin | Documentation | Help

Data Centre by numbers

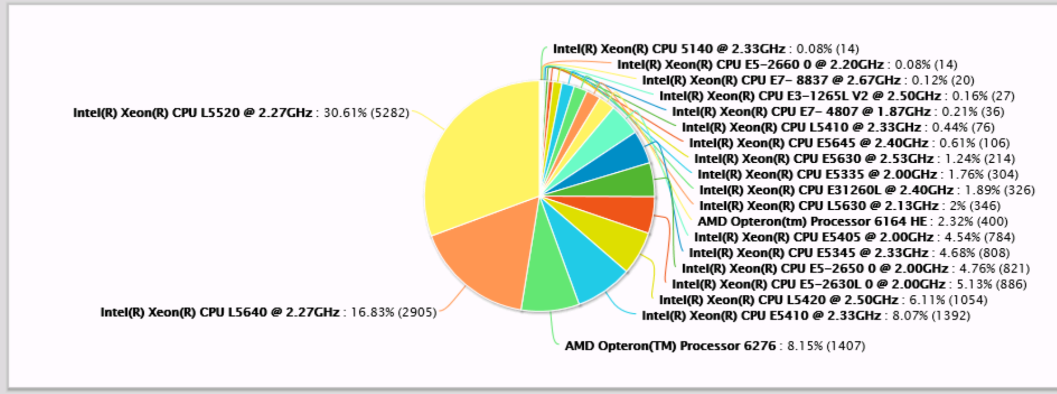
10 May 2014 Sat 19:52:55

CERN's TSM Backup Service		Machine Inventory	
Total Data on Tape (PB)	7.56	Number of 10GB NICs	3,220
		Number of 1GB NICs	19,331
		Number of cores	99,067
		Number of disks	80,562
		Number of memory modules	68,362
		Number of processors	18,633
		Number of servers	10,809
		Total disk space (TiB)	128,849.37
		Total memory capacity (TiB)	349.20

SLS by CERN IT/CF

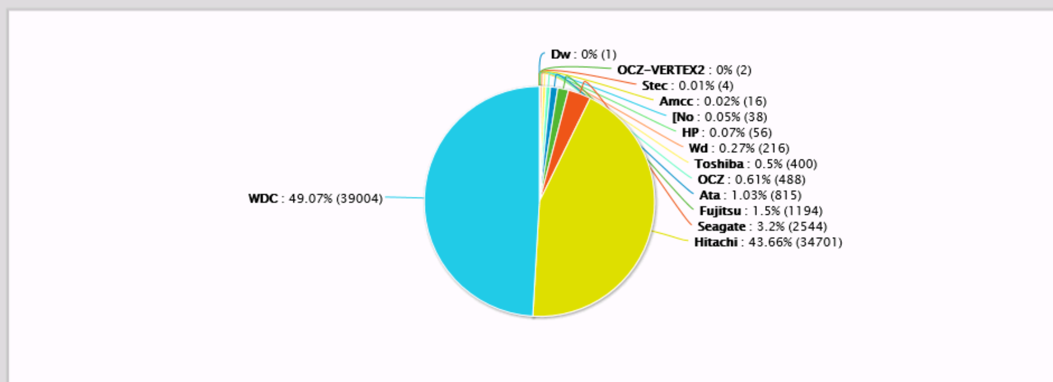
Graphs

CPU type distribution

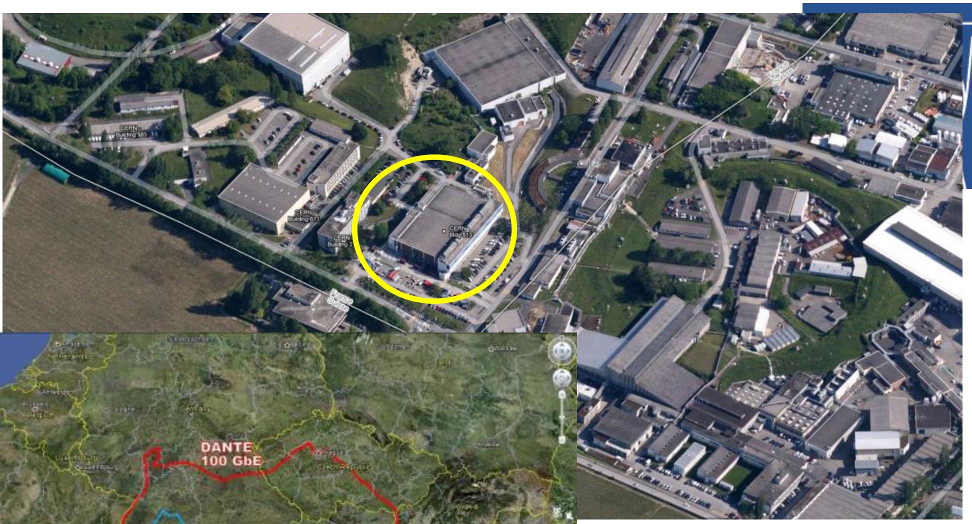


Download as csv

HDD vendor distribution



CERN Computer Centre

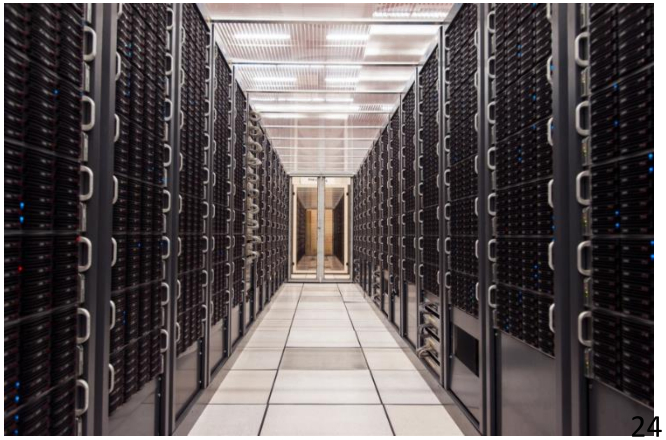


CERN computer centre:

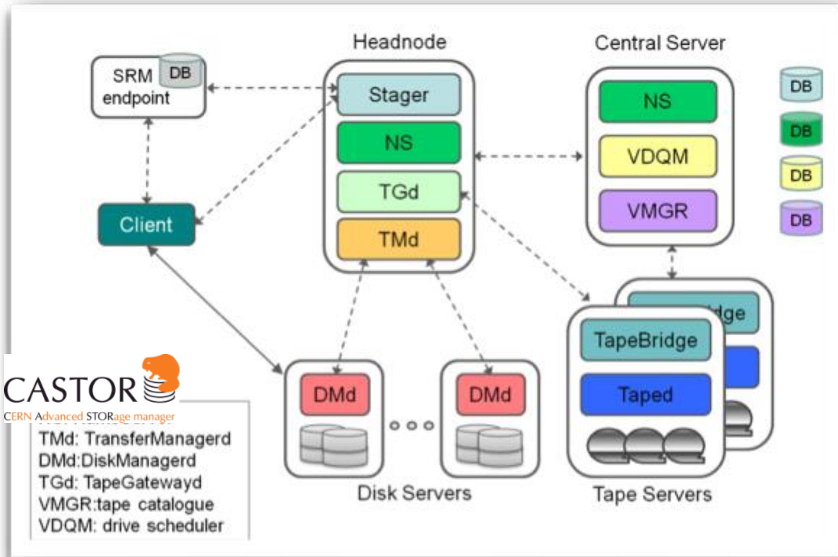
- Built in the 70s on the CERN site (Meyrin-Geneva)
- ~3000 m^2 (3 main machine rooms)
- 3.5 MW for equipment
- Est. PUE ~ 1.6

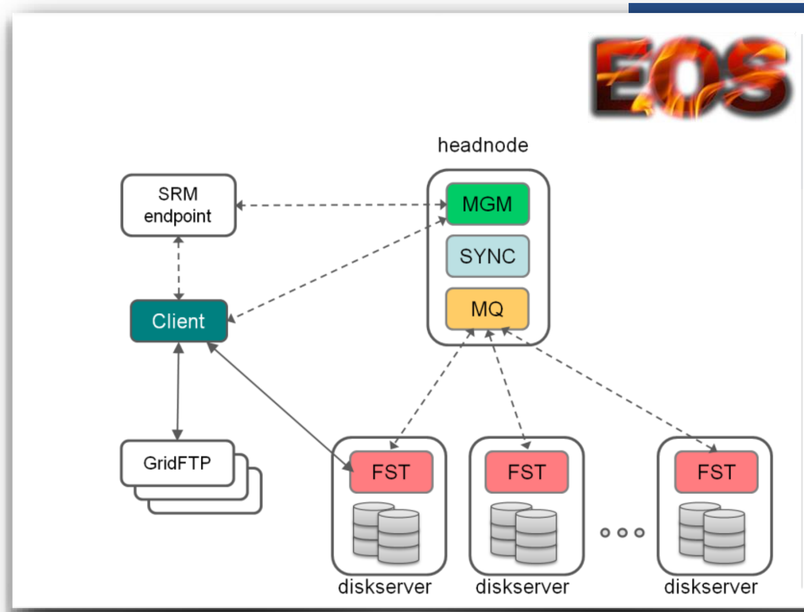
New extension:

- Located at Wigner (Budapest)
- ~1000 m^2
- 2.7 MW for equipment
- Connected to the Geneva CC with 2x100Gb links (21 and 24 ms RTT)
- **Another room ~20 ms away…**



CASTOR and EOS





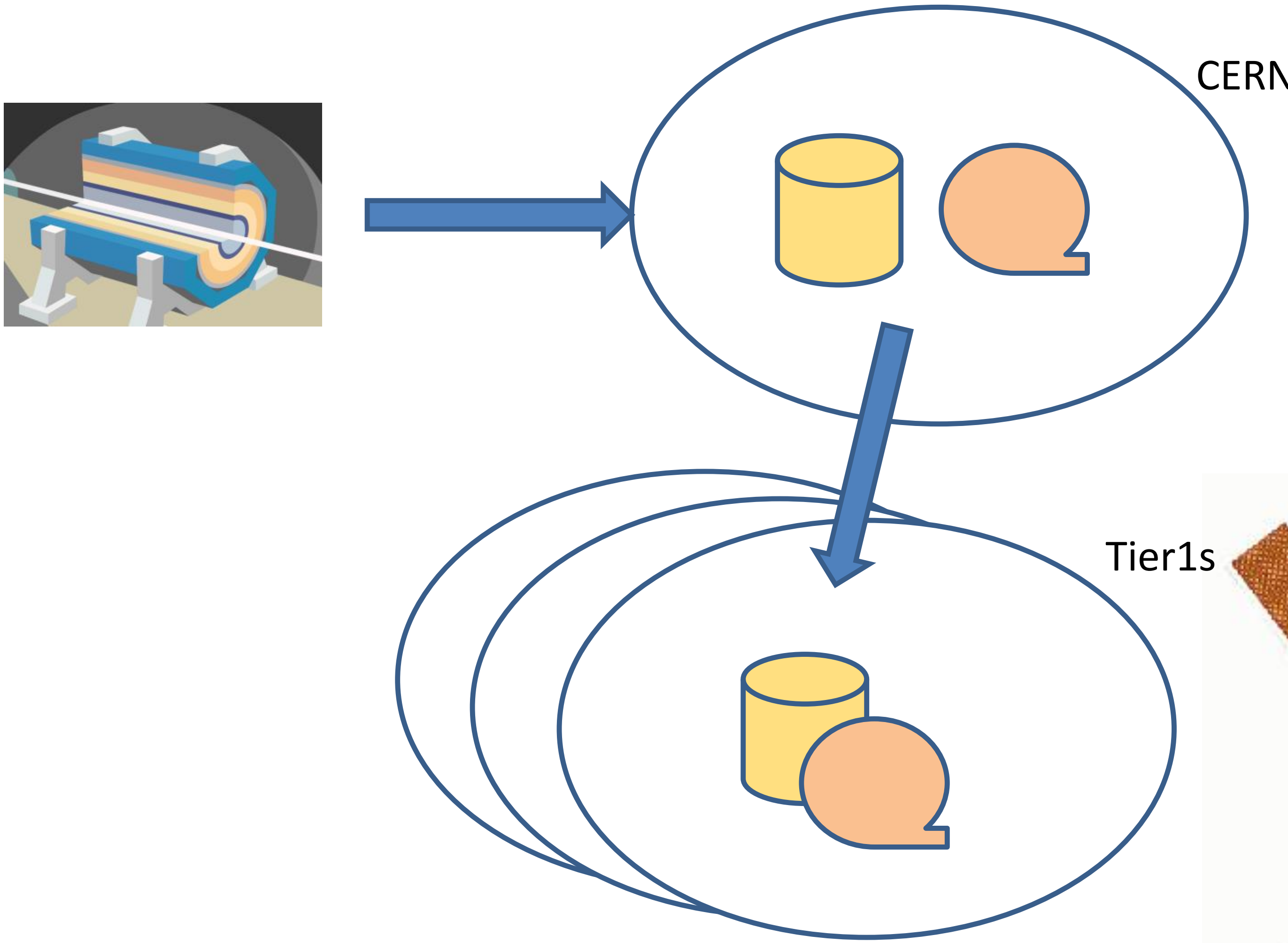
CASTOR and EOS: developed at CERN (IT department)

Workhorses for physics data (archival, reconstruction, analysis and distribution)

CASTOR and EOS are using the same commodity HW

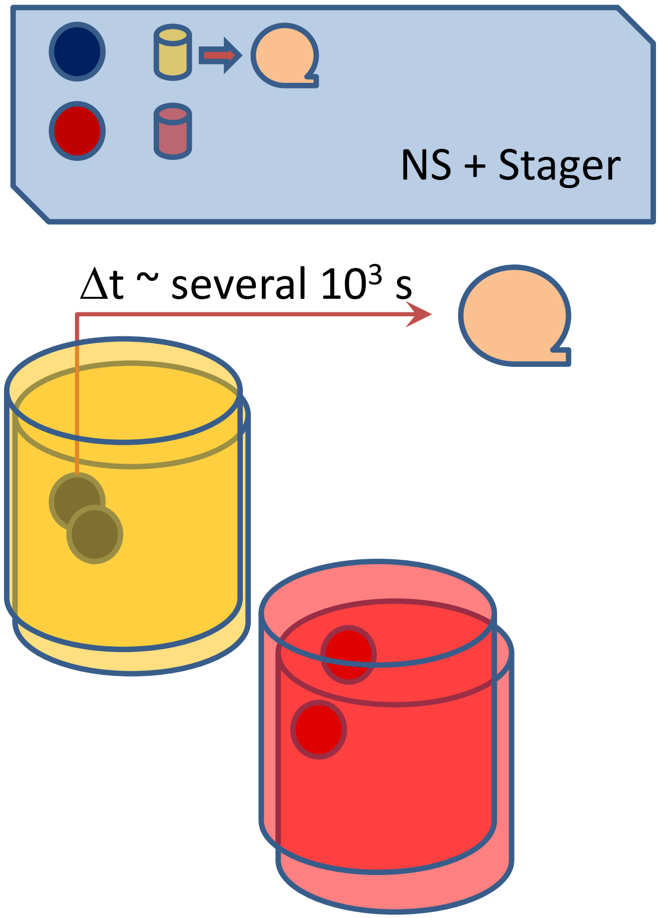
- **RAID-1 for CASTOR**
  - **2 copies in the mirror (same box, multiple copies possible)**
- **JBOD with RAIN for EOS**
  - **enforce replicas to be on different disk servers**
    - **tunable number (usually 2 replicas)**
    - **Erasure coding**
  - **“Arbitrary” configurable durability**
  - **Disaster recovery: cross-site replication**

Data durability

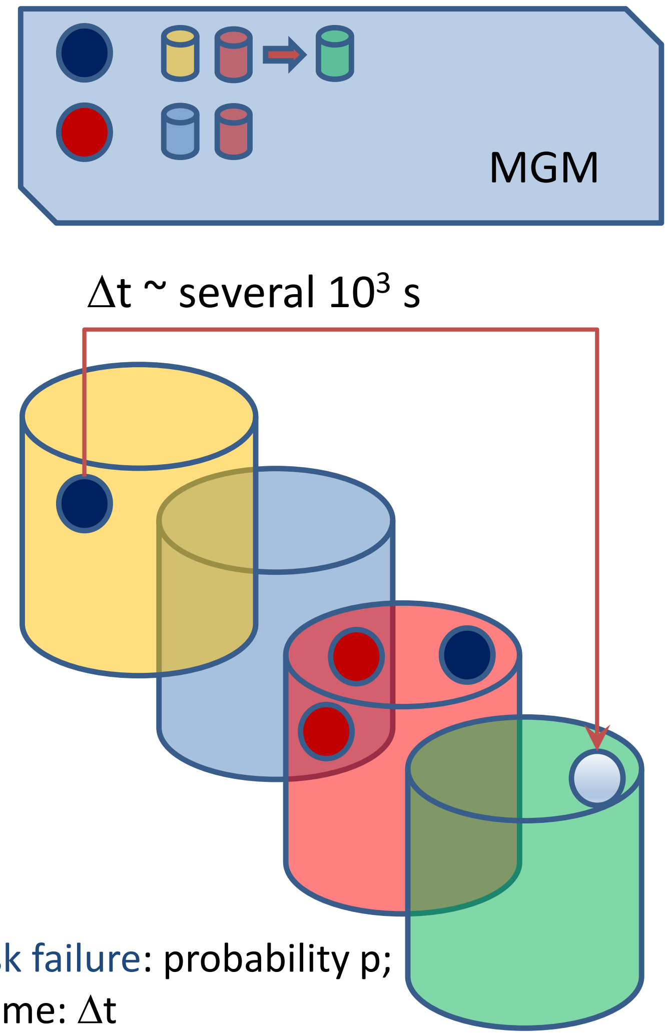


Simplified schema: 1 copy + 1 “distributed” copy (2 tape copies)
Scalability + data durability (disaster recovery)

Data durability (2)



Single disk failure: probability p;
Double failure: p^2 effect; overlapping failure during healing time $\approx \Delta t$; $\cap(disk_x, disk_y)$



Single disk failure: probability p;
healing time: Δt
Double failure: p^2 effect; overlapping failure during healing time $\approx \Delta t$; $\cap(disk_x, disk_y)$

CERN data management (physics data)

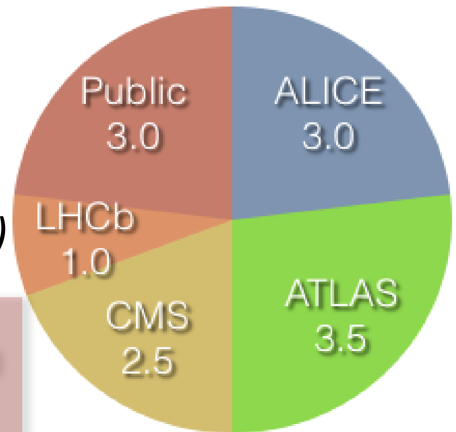
CASTOR
CERN Advanced STORage manager

Total installed disk capacity: 13.0PB
~600 disk servers

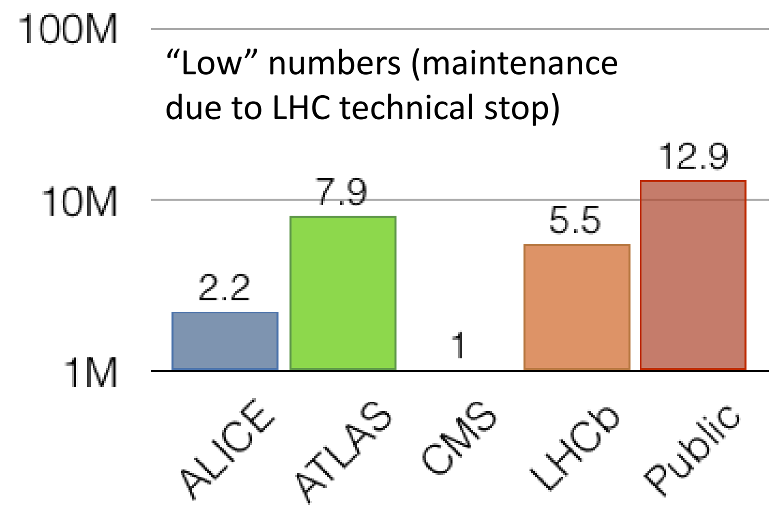
Mainly used as disk front-end (includes LHC data repository)

Unique function across LCG:
- Primary repository for LHC data (raw)
- Source of prompt reconstruction
- Distribution point for LHC data (raw)

Installed (usable) disk capacity



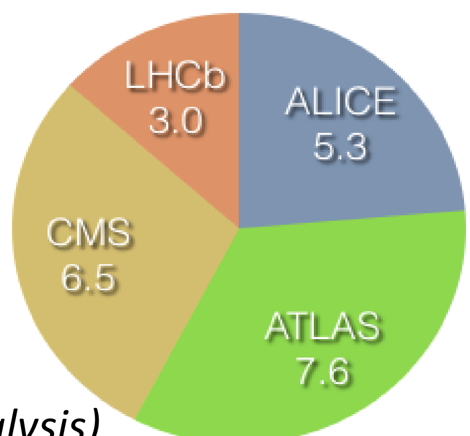
Number of Staged Files



EOS

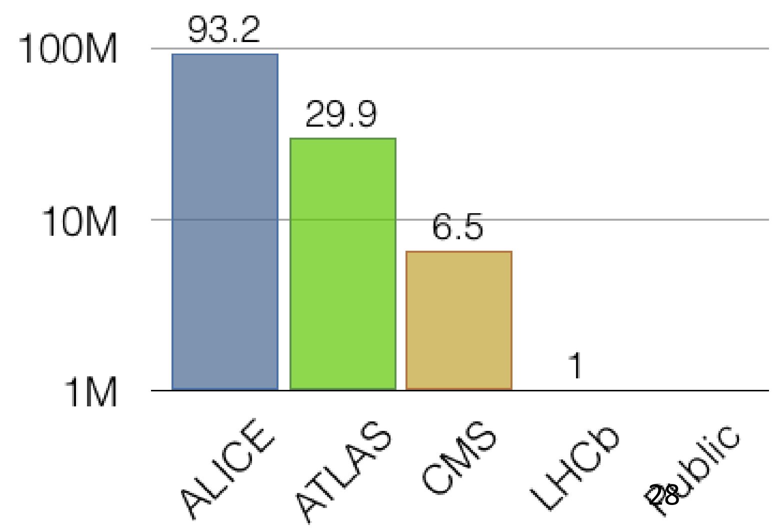
Total installed capacity: 25PB
~900 disk servers

Installed (usable) capacity



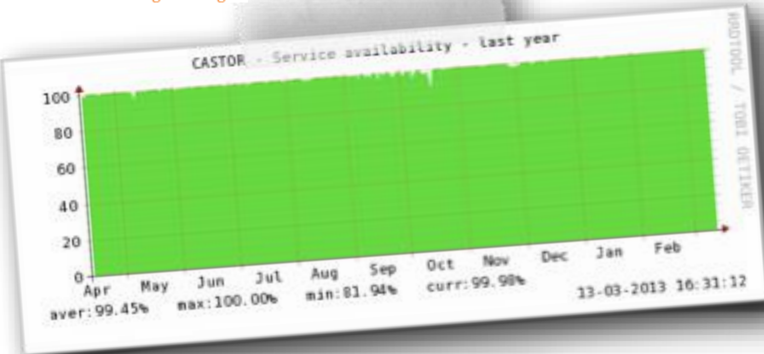
Disk-only data (analysis)

Number of Files

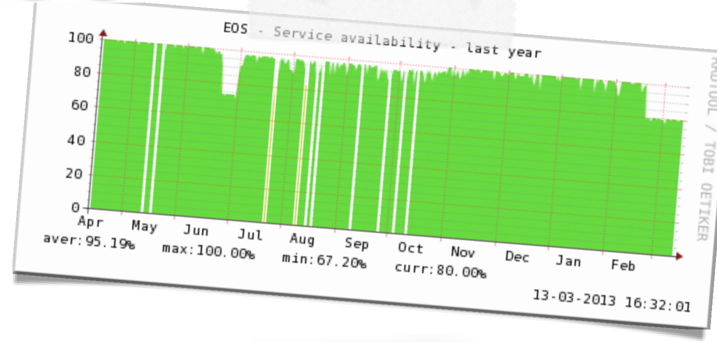


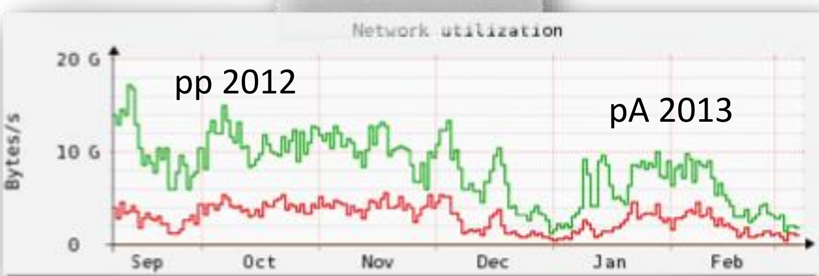
Availability and I/O performances

CASTOR
CERN Advanced STORage manager

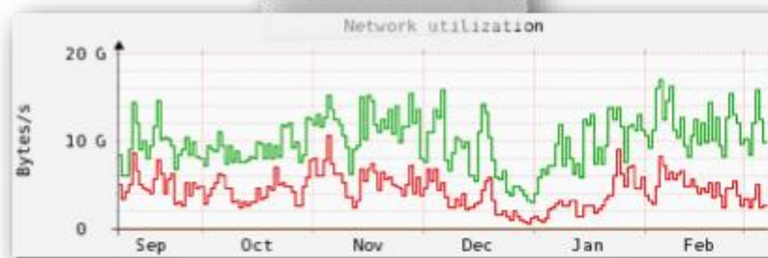


EOS

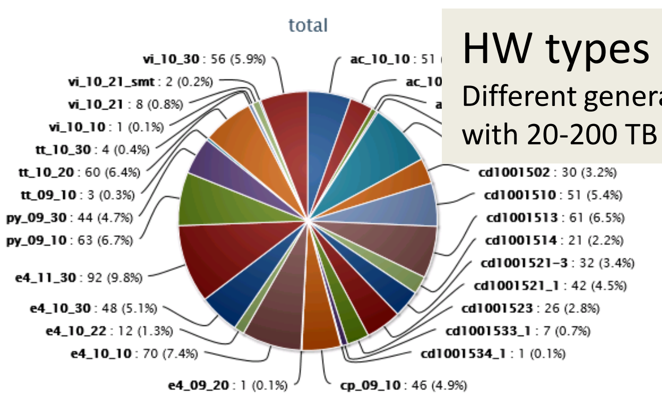




Data taken driven

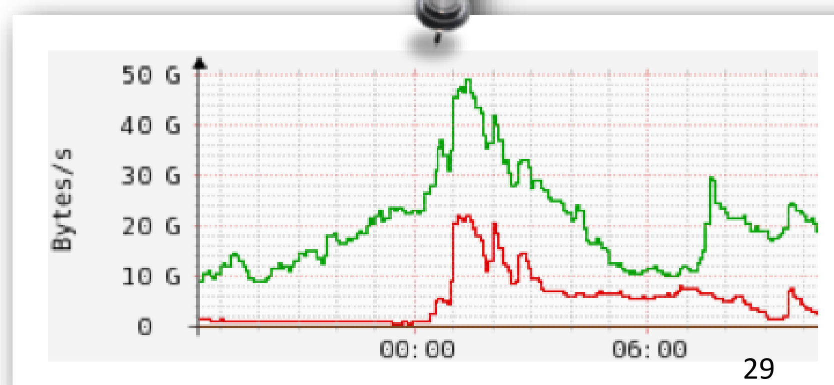


Analysis driven



HW types

Different generations of PC boxes with 20-200 TB disks space



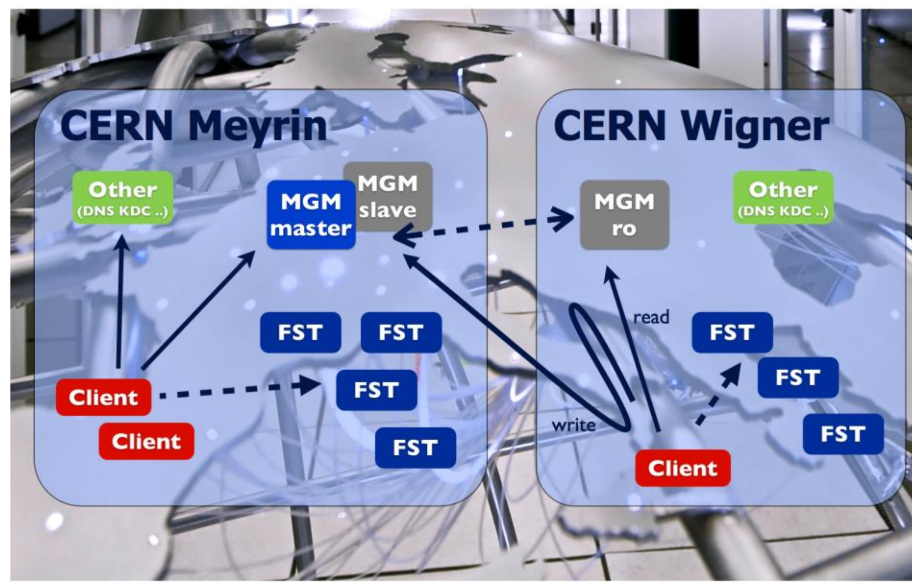
EOS 2-site deployment

Multiple copy

- Geolocalisation
- 1 in Meyrin (CERN Geneva), 1 in Budapest (Wigner)
- Whenever possible

Multiple MGM deployment

- Complex deployment to guarantee low latency
- Xroot protocol: redirect clients
- Single master MGM
- Most operations can be done on R/O MGM
- Transparent redirection



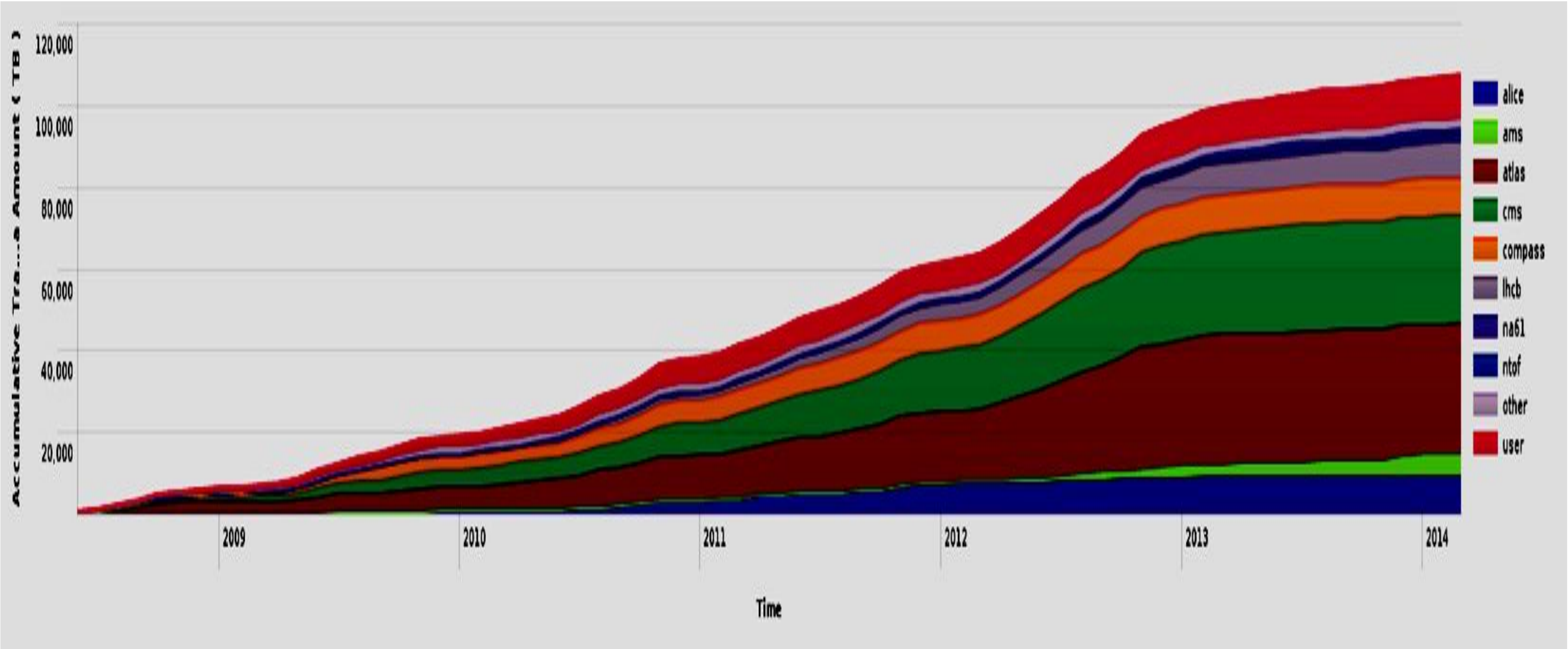
CASTOR (tape archive)

Data:

- ~90 PB of data on tape; 250 M files
- Up to 4.5 PB new data per month
- Over 10GB/s (R+W) peaks

Infrastructure:

- ~ 52K tapes (1TB, 4TB, 5TB, 8TB)
- 9 Robotic libraries (IBM and Oracle)
- 80 production + 30 legacy tape drives



DSS/FDO weekly report (16 Mar 2014)

ACRON service

Instance	Jobs	Efficiency (*)	Unique user/host
@cern	908.9 K	96.0 %	761.0

AFS service

Instance	Capacity	Files	Δ	Size	Δ
@cern	992.0 TB	2.3 G	2.6 M	292.5 TB	-2.5 TB

CEPH service

Instance	Capacity	Objects	Δ	Size	Δ
@cern	2871.8 TB	7.6 M	0.2 M	28.6 TB	1.2 TB

FILER service

Instance	Capacity	Files	Δ	Size	Δ
@cern	192.0 TB	213.7 M	-0.1 M	40.6 TB	0.2 TB

CASTOR service

Instance	Files	Δ	Size	Δ	OnTape	Δ
@cern	326.0 M	0.4 M	89.1 PB	31.0 TB	79.0 PB	97.0 TB

EOS service

Instance	Files	Δ	Size	Δ
@cern	166.5 M	3.0 M	20.4 PB	85.4 TB
alice	90.2 M	0.5 M	4.4 PB	44.3 TB
atlas	59.9 M	0.0 M	6.9 PB	-33.4 TB
cms	9.2 M	0.2 M	5.1 PB	-9.8 TB
lhcb	4.6 M	2.0 M	3.2 PB	80.6 TB
public	2.6 M	0.2 M	0.8 PB	3.7 TB

Other main services:

- TSM

Internal/beta services:

- Hadoop (*)
- CERNBOX (cloud sync)

(*) Initiated by a Technical Student (S. Russo) from TS+UD uni (ATLAS experiment)

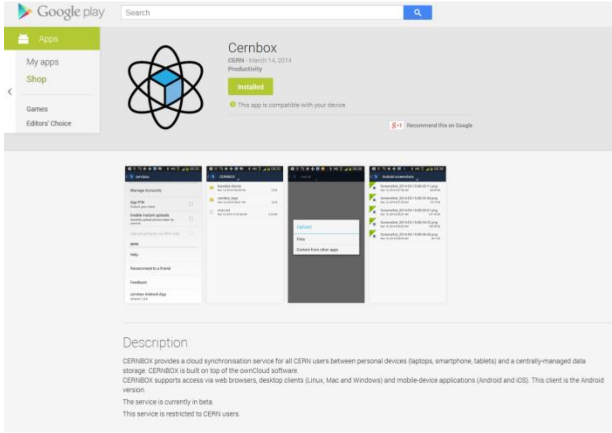
New end-user services?

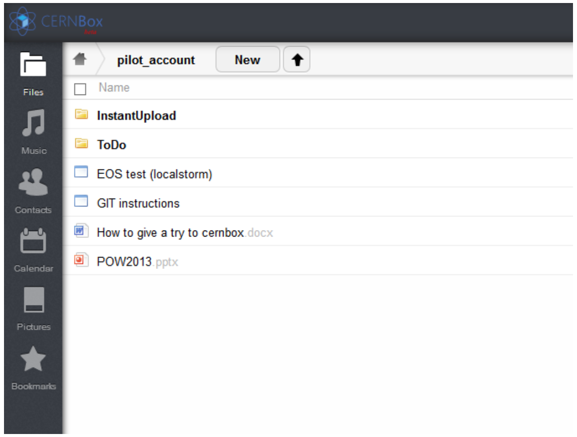
CERNBOX to enter beta soon

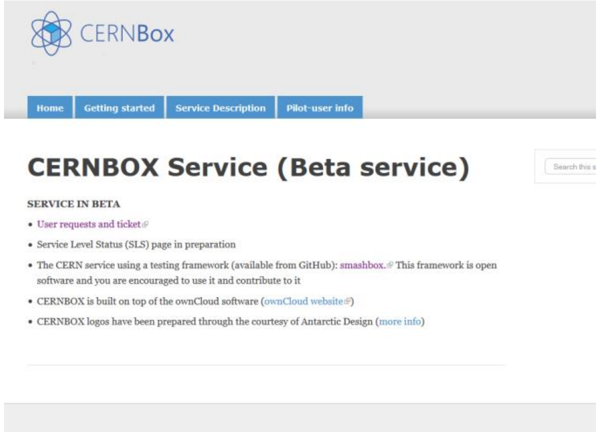
CERNBOX provides a cloud synchronisation service for all CERN users between personal devices (laptops, smartphone, tablets) and a centrally-managed data storage.

Coherent data handling (with our other services): SLA, ACL, curation...

CERNBOX is built on top of the ownCloud software. CERNBOX supports access via web browsers, desktop clients (Linux, Mac and Windows) and mobile-device applications (Android and iOS).



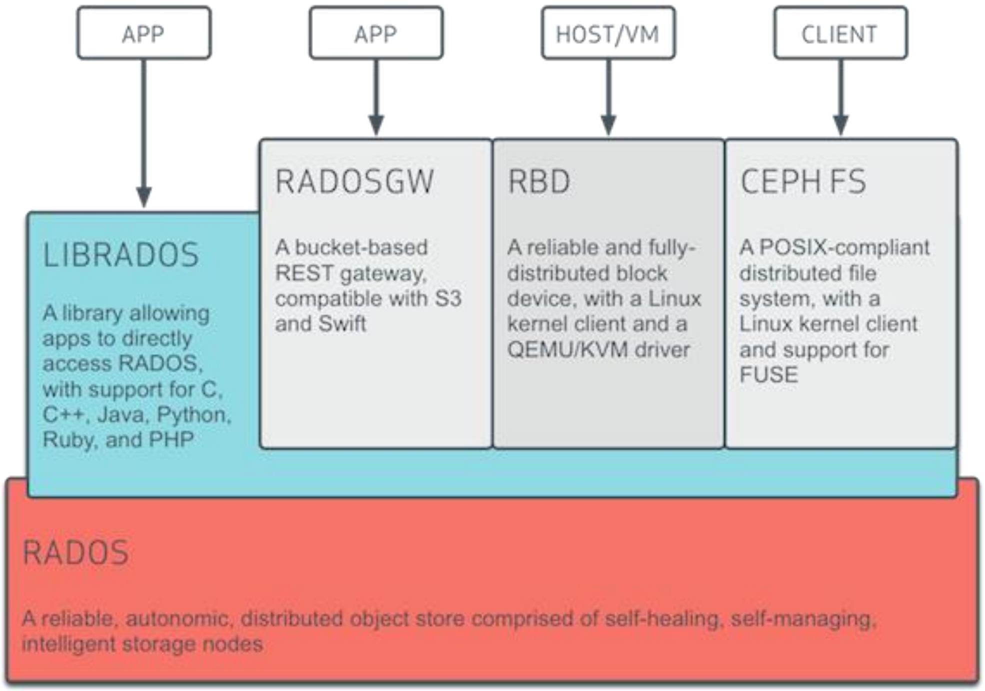




http://owncloud.com

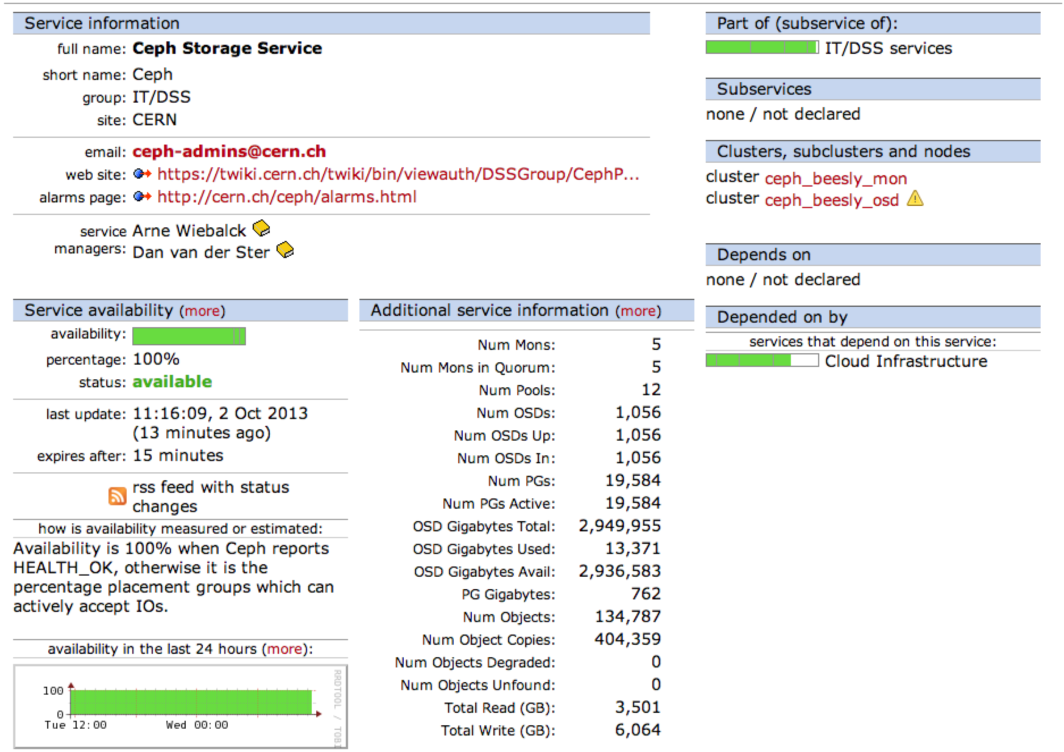
Ceph Architecture and Use-Cases

- Open-software storage solution
- Scalable from 10s to 10,000s of machines, from Terabytes to Exabytes (10^3 PB)
- Fault tolerant with no SPOF, uses commodity hardware
- Self-managing, self-healing



Deploying a 3-PB prototype

- Fully puppetized deployment
- Automated machine commissioning and maintenance
  - Add a server to the hostgroup (osd, mon, radosgw)
  - OSD disks are detected, formatted, prepared, auth'd
    - Also after disk replacement
  - Auto-generated ceph.conf
  - Last step is manual/controlled: service ceph start
- Mcollective for bulk operations on the servers
  - Ceph rpm upgrades
  - daemon restarts




```
# df -h /mnt/ceph
Filesystem
Size  Used Avail Use% Mounted on
xxx:6789:/  3.1P  173T  2.9P   6% /mnt/ceph
```

Beyond HEP

Ground floor exposition

Sharing experience with other activities (e.g. Biology)
Planning strategic partnership with other sciences (HelixNebula)
Collaboration with technology providers (OpenLab)

A European cloud computing partnership: big science teams up with big business

HELIX NEBULA THE SCIENCE CLOUD

32

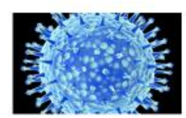
CERN openlab in a nutshell

- A science – industry partnership to drive R&D and innovation with over a decade of success
- Evaluate state-of-the-art technologies in a challenging environment and improve them
- Test in a research environment today what will be used in many business sectors tomorrow
- Train next generation of engineers/employees

and outreach to new

31

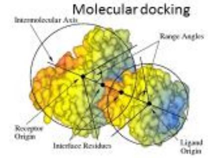
Flu outbreaks and other diseases

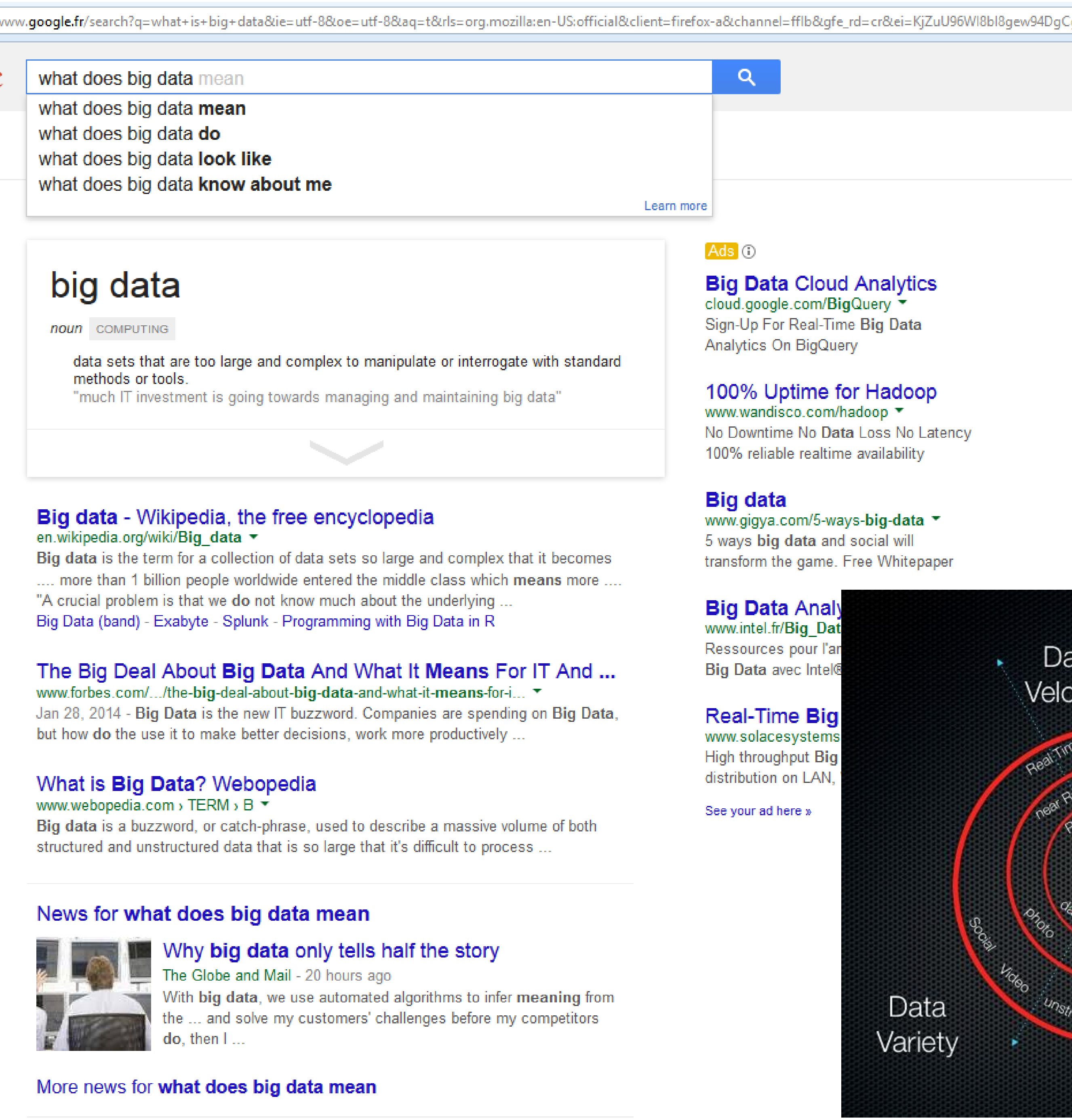


- Basic idea:
  - Compute how a given chemical interacts with a protein (e.g. belonging to a virus)
  - High affinity means the chemical is a potential drug against the virus
- In silico (i.e. use your PC):
  - Scan millions of chemicals (~10 minutes per chemical-protein pair)
    - With 1,000 PCs, 1 docking per second
  - Good candidate given to biologist (verification longer -and more expensive- than in silico docking)
    - In practice, you enrich the initial sample saving time
    - (and money)
    - Essential to fight to pandemic (H5N1) or
    - neglected diseases (like Malaria)
- WISDOM collaboration
  - Malaria
  - H5N1 (Bird Flu)

egee Enabling Grids for E-sciencE

Molecular docking




www.google.fr/search?q=what+is+big+data&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a&channel=fflb&gfe_rd=cr&ei=KjZuU96WI8bI8gew94DgCg

what does big data mean

what does big data **mean**
what does big data **do**
what does big data **look like**
what does big data **know about me**
Learn more

big data
noun COMPUTING
data sets that are too large and complex to manipulate or interrogate with standard methods or tools.
"much IT investment is going towards managing and maintaining big data"

Big data - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Big_data**
Big data is the term for a collection of data sets so large and complex that it becomes .... more than 1 billion people worldwide entered the middle class which **means** more .... "A crucial problem is that we **do** not know much about the underlying ...
Big Data (band) - Exabyte - Splunk - Programming with Big Data in R

The Big Deal About **Big Data** And What It **Means** For IT And ...
www.forbes.com/.../the-**big-deal-about-big-data-and-what-it-means**-for-i...
Jan 28, 2014 - **Big Data** is the new IT buzzword. Companies are spending on **Big Data**, but how **do** the use it to make better decisions, work more productively ...

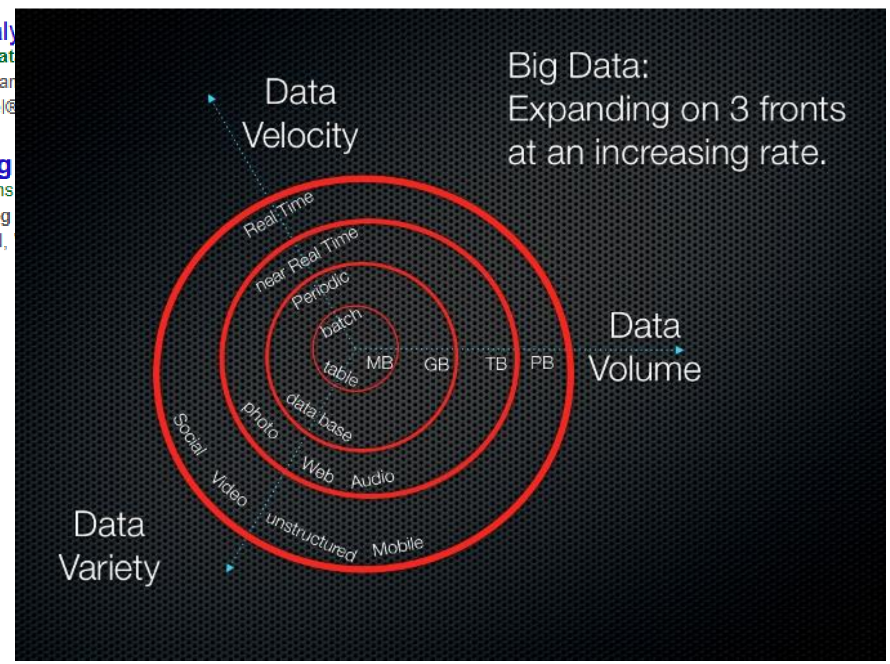
What is **Big Data**? Webopedia
www.webopedia.com › TERM › B
Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process ...

News for **what does big data mean**

Why **big data** only tells half the story
The Globe and Mail - 20 hours ago
With **big data**, we use automated algorithms to infer **meaning** from the ... and solve my customers' challenges before my competitors **do**, then I ...

More news for **what does big data mean**

Big Data



The 3 Vs in HEP

Volume

- For sure at somewhere between 10 and 100 PB you are getting big

Variety

- # of different technical solutions, operational procedures, users and <u>workflows</u>?
- # of different types (>3? >30?)

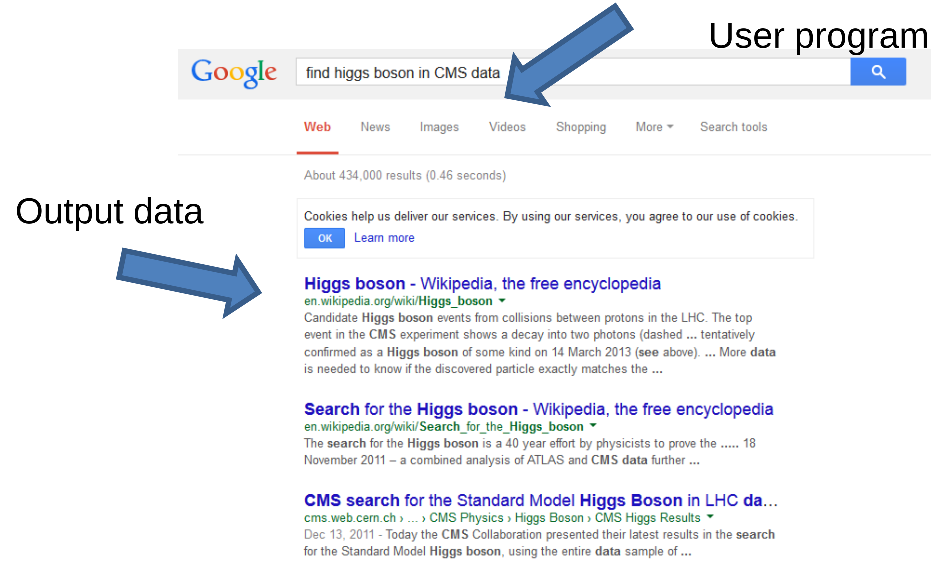
Velocity

- Δvolume/Δt in 1 PB/month (disk installation)
- Last 2 years of EOS@CERN
- Δvolume/Δt in 4 PB/month (recording, processing and redistributing)
- LHC running (end of pp Run1)

But do not forget:

- Max time you can be down!
- Max integrated time in a month!
- # of human interventions per day!

...and we want to read high **volume** of data at high **velocity** in a **variety** of <u>ways</u> ☺



∀ event in 2012 CMS AOD …

```
for(ev.toBegin(); !ev.atEnd(); ++ev, ++ievt){
        edm::EventBase const & event = ev;
        // break loop if maximal number of events is reached
        if(maxEvents_>0 ? ievt+1>maxEvents_ : false) break;
        // simple event counter
        if(outputEvery_!=0 ? (ievt>0 && ievt%outputEvery_==0) : false)
          std::cout << "  processing event: " << ievt << std::endl;

        // Handle to the muon collection
        edm::Handle<std::vector<Muon> > muons;
        event.getByLabel(std::string("muons"), muons);

        // loop muon collection and fill histograms
        for(std::vector<Muon>::const_iterator mu1=muons->begin(); mu1!=muons->end(); ++mu1){
          muonPt_ ->Fill( mu1->pt () );
          muonEta_->Fill( mu1->eta() );
          muonPhi_->Fill( mu1->phi() );
          if( mu1->pt()>20 && fabs(mu1->eta())<2.1 ){
            for(std::vector<Muon>::const_iterator mu2=muons->begin(); mu2!=muons->end(); ++mu2){
              if(mu2>mu1){ // prevent double conting
                if( mu1->charge()*mu2->charge()<0 ){ // check only muon pairs of unequal charge
                  if( mu2->pt()>20 && fabs(mu2->eta())<2.1 ){
                    mumuMass_->Fill( (mu1->p4()+mu2->p4()).mass() );
```

Complex output created by the program(s) …

Our "Big Data"

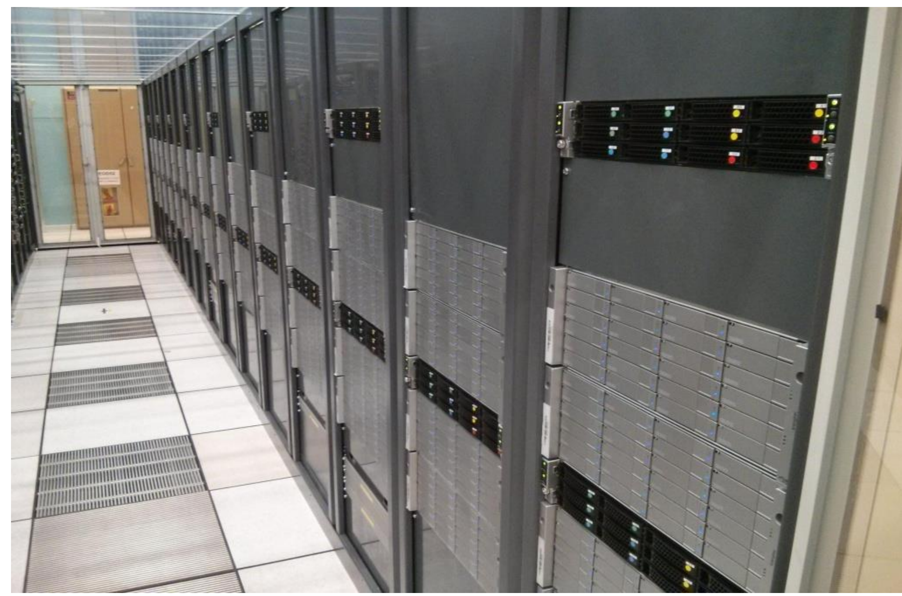
Physics Data on CASTOR/EOS

- LHC experiments produce ~10GB/s 25PB/year

Service	Size	Files
AFS	290TB	2.3B
CASTOR	89.0PB	325M
EOS	20.1PB	160M

User Data on AFS & DFS

- Home directories for 30k users
- Physics analysis development
- Project spaces for applications



Service Data on AFS/NFS/Filers

- Databases, admin applications

Tape archival with CASTOR/TSM

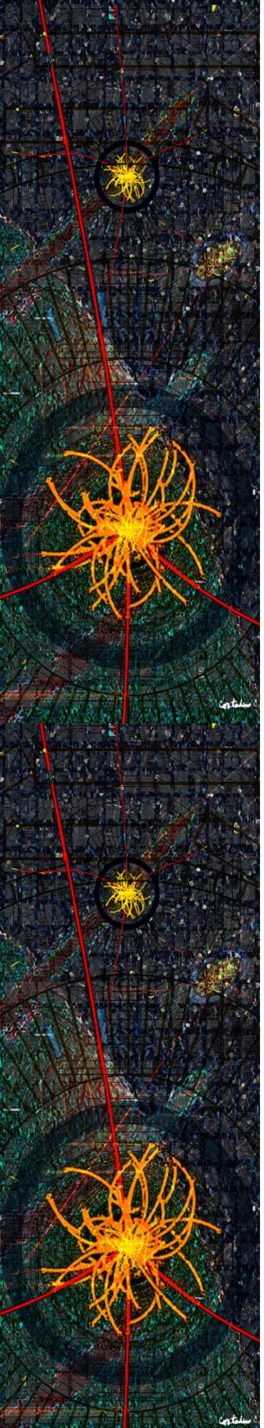
- Physics data (RAW and derived/additional)
- Desktop/Server backups

Data growth

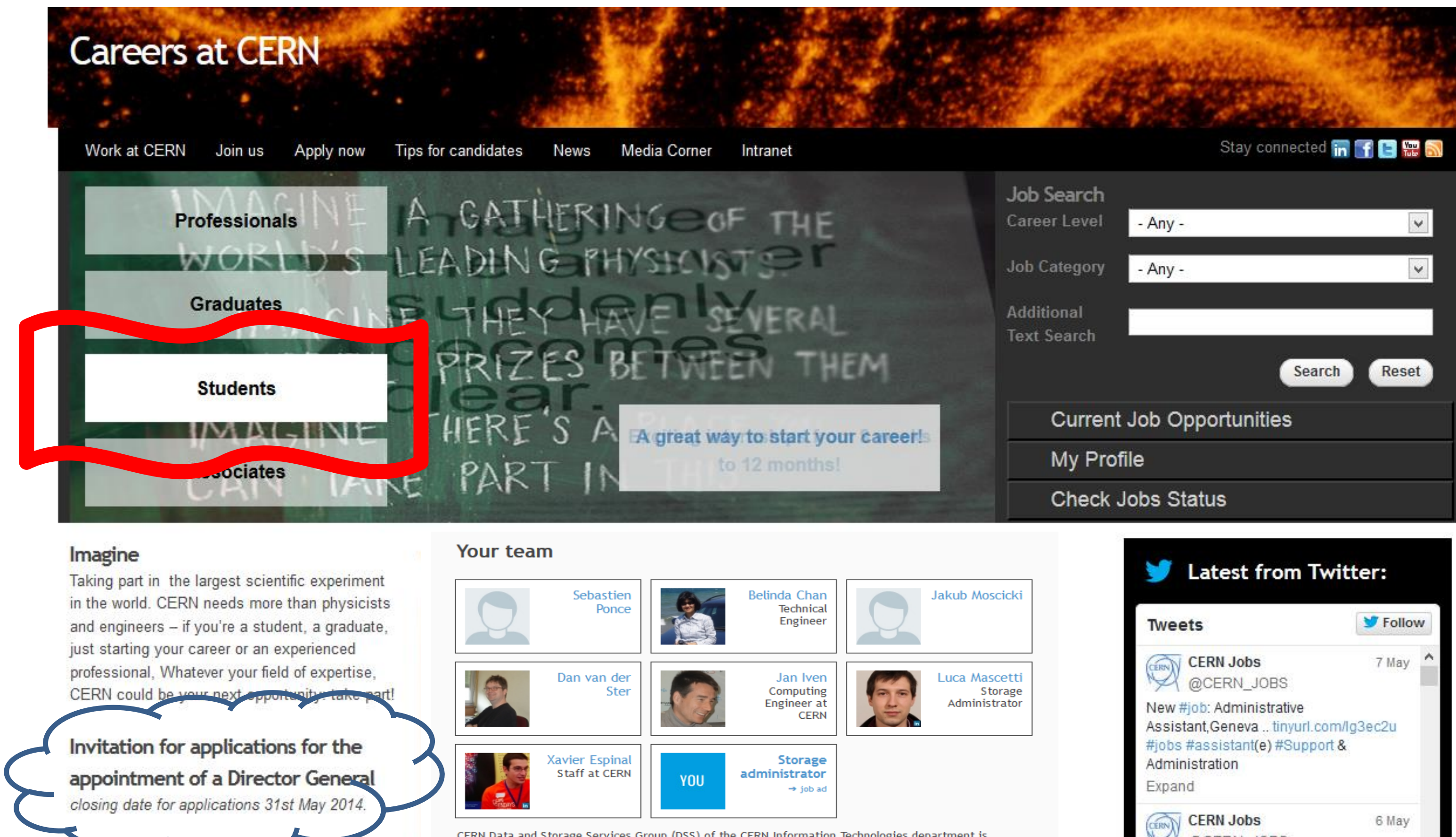
- 2 PB/month over the last 2 years (raw disk space)
- 4 PB/month new data at the end of LHC phase 1
  - Factor of 2 more at the beginning of phase 2?

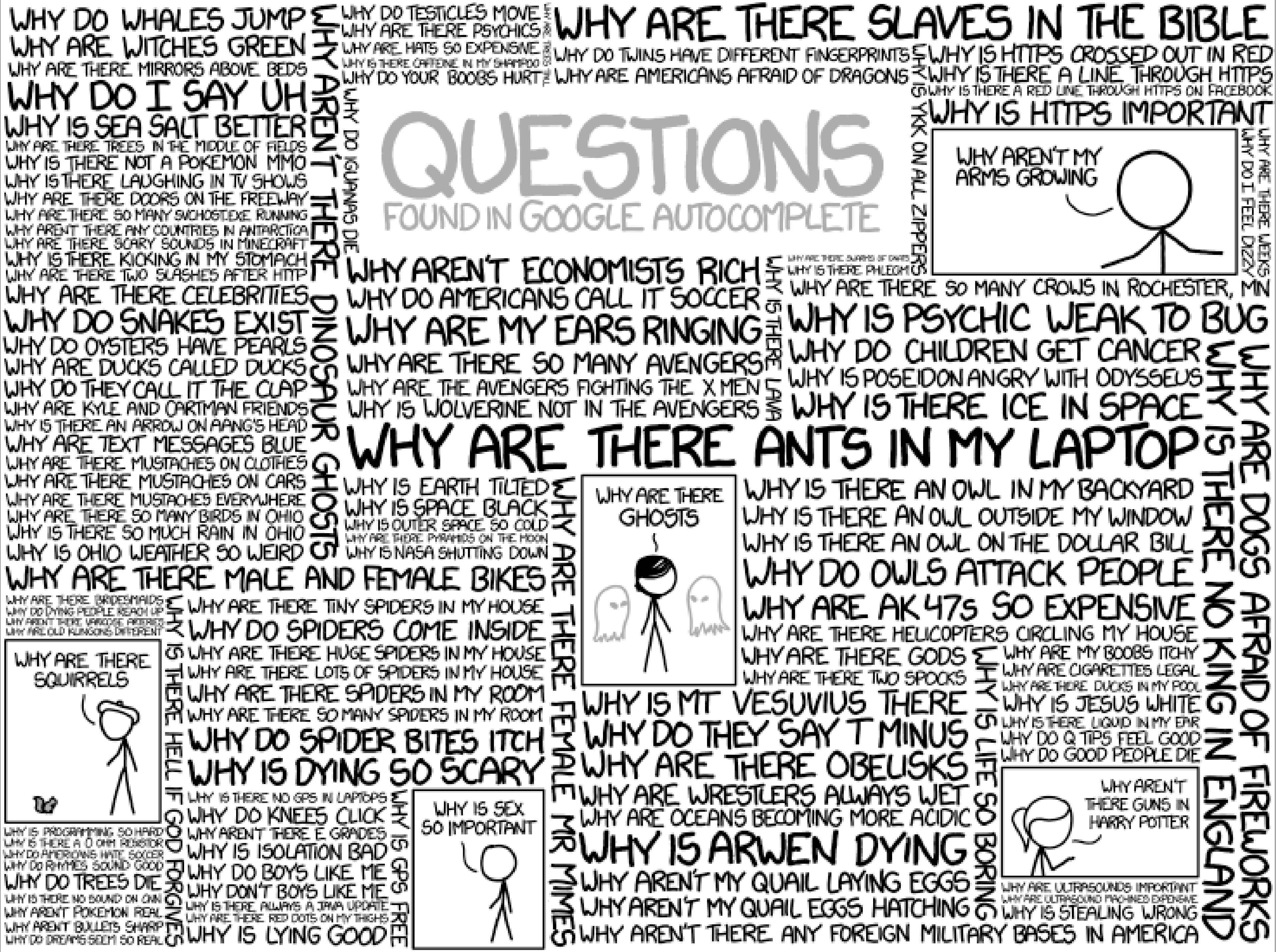
Staff, fellows and students!

- Staff, fellow and students
  - **cern.ch/jobs**
  - Many opportunities for students!
    - Physicists & Engineers (incl. Computer science)
  - Challenging but exciting!
    - Summer Student (Jan/Feb → 8-12 weeks in summer)
    - OpenLab Summer Students (Jan/Feb → 8-12 weeks in summer)
    - Technical Students (2 times per year → 6-12 months)
    - …maybe too early for applying for DG but check out the student pages



Xavier Cortada (with the participation of physicist Pete Markowitz), "In search of the Higgs boson: H -> ZZ", digital art, 2013.





QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY AREN'T THERE DINOSAUR GHOSTS
WHY DO TESTICLES MOVE
WHY ARE THERE PSYCHICS
WHY ARE HATS SO EXPENSIVE
WHY IS THERE CAFFEINE IN MY SHAMPOO
WHY DO YOUR BOOBS HURT
WHY DO IGUANAS DIE
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS YKK ON ALL ZIPPERS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS THERE LAVA
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY ARE THERE FEMALE MR MIMES
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY ARE MY BOOBS ITCHY
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD RUNSONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS THERE HELL IF GOD FORGIVES
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS GPS FREE
WHY IS SEX SO IMPORTANT
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY IS LIFE SO BORING
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

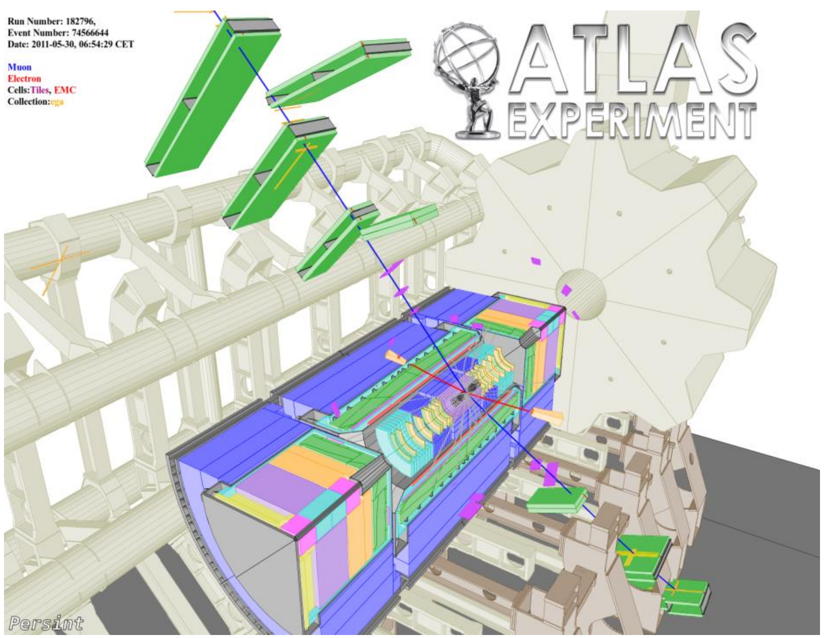
But finally why so much data?

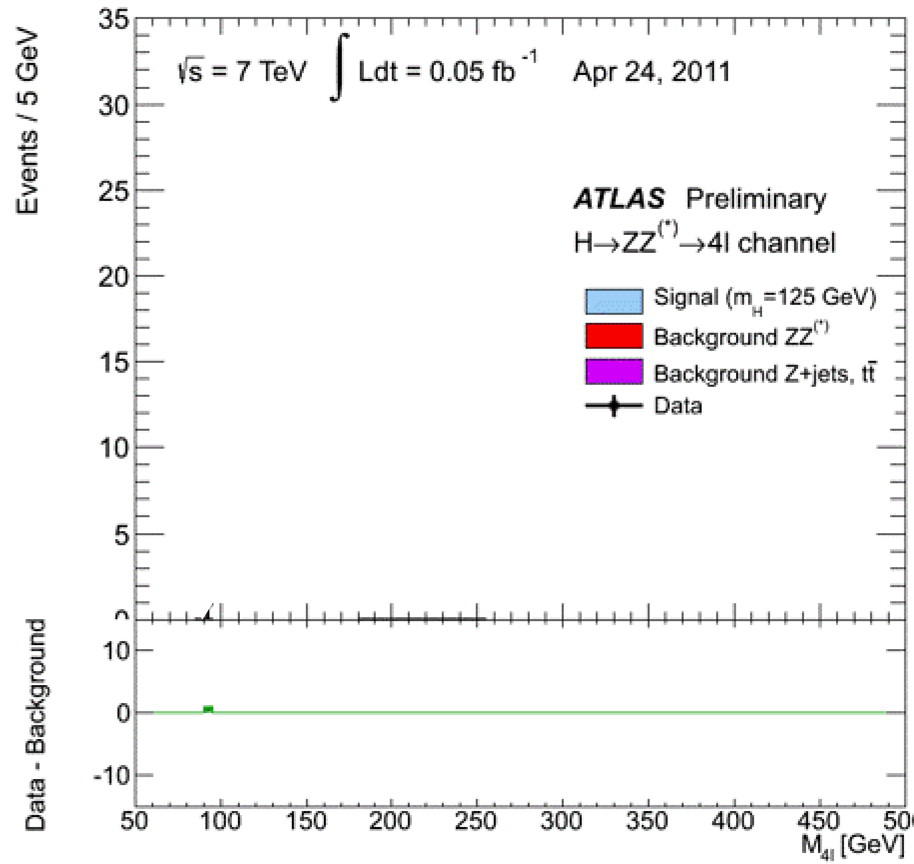
$\int$ Ldt → Luminosity

It is a measurement of the potential statistics accumulated in the collisions over time (here it accounts for the 2011 and 2012 data taking)

Think to the exposure time of a camera when taking pictures

Events with 2 Zs decaying in 2-lepton pairs (here e^+e^- and $\mu+\mu-$)





But finally why so much data?

$\int Ldt$ → Luminosity

It is a measurement of the potential statistics accumulated in the collisions over time (here it accounts for the 2011 and 2012 data taking)

Think to the exposure time of a camera when taking pictures

Events with 2 Zs decaying in 2-lepton pairs (here e^+e^- and $\mu^+\mu^-$)