CERN

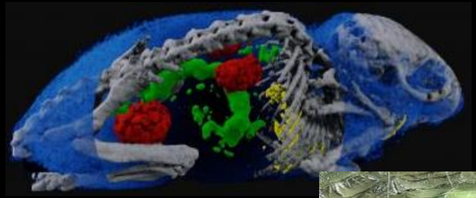
Computación en el CERN: al servicio de la ciencia y su impacto en la sociedad

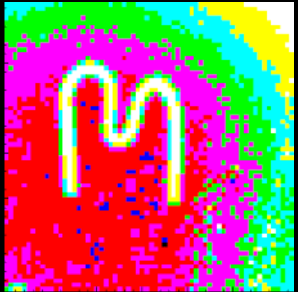
Xavier Espinal
CERN-IT

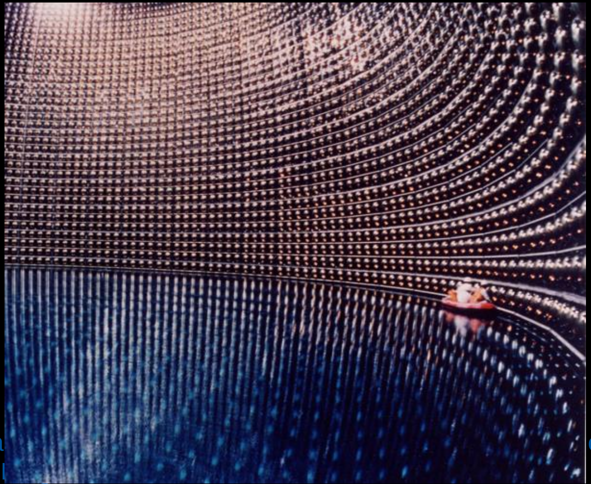
Welcome to CERN-IT!

My name is Xavier Espinal. I obtained the physics degree in 2000 in Barcelona. Afterwards I started Doctoral Studies in Particle Physics finishing in 2006 with a PhD in a Neutrino Oscillations experiment in Japan (K2K). Member of ATLAS computing and the PIC computing center in Barcelona WLCG /Started my career at CERN in 2012 in the Data Storage group.

Currently I am the IT Technical Coordinator for experiments and departments and working on the HL-LHC computing challenges







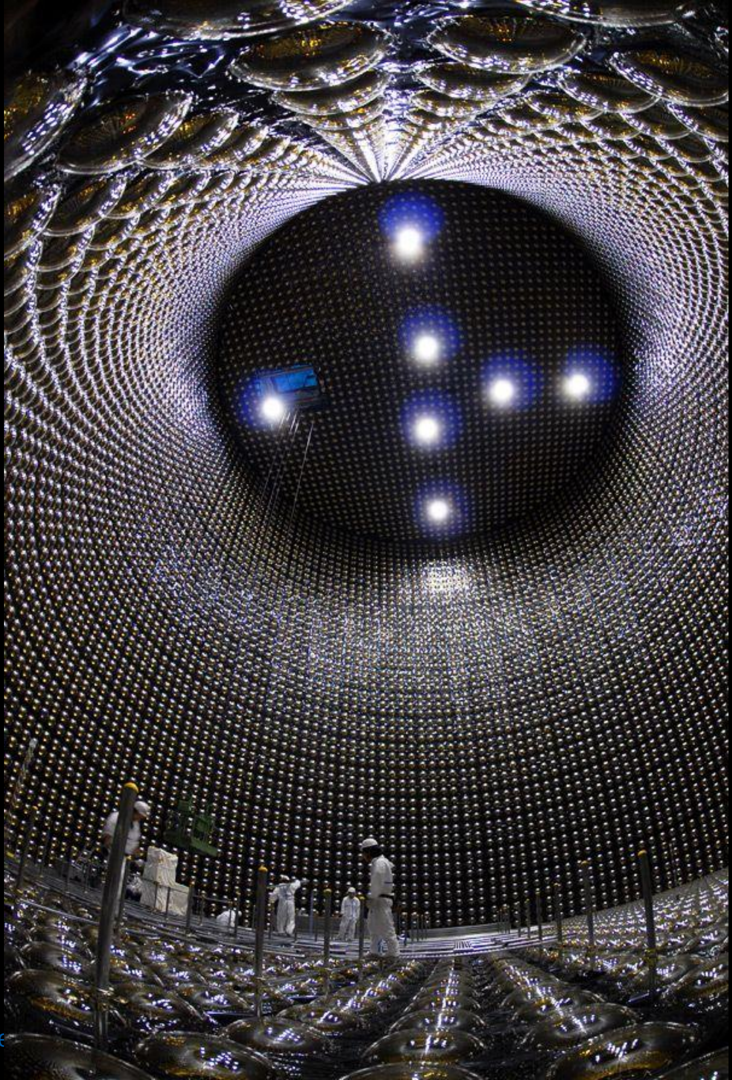


photo @ https://seanmunger.com/

https://home.cern/cern-people/updates/2012/12/remembering-wim-klein

[…]Professor Bakker wrote that Mr Klein had been recommended by the director of the Zeeman laboratory in Amsterdam as a remarkable calculator[…] He needed no desk calculator and performed exceedingly well, exceeding in speed even my own desk calculator[…] I needed tables of combinations of so-called Clebsch-Gordan coefficients […] values of were tabled as decimal numbers, e.g. 0.92308 […] but I needed the explicit form […] he said, 11/13

CERN Data Centre passes the 200-petabyte milestone

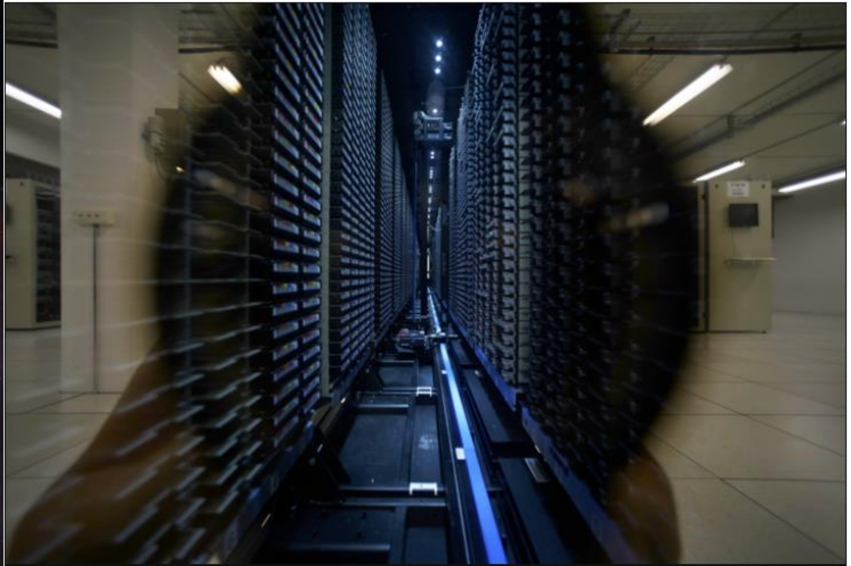
by Mélissa Gaillard


CERN's Data Centre (Image: Robert Hradil, Monika Majer/ProStudio22.ch)

On 29 June 2017, the CERN DC passed the milestone of 200 petabytes of data permanently archived in its tape libraries. Where do these data come from? Particles collide in the Large Hadron Collider (LHC) detectors approximately 1 billion times per second, generating about one petabyte of collision data per second. However, such quantities of data are impossible for current computing systems to record and they are hence filtered by the experiments, keeping only the most "interesting" ones. The filtered LHC data are then aggregated in the CERN Data Centre (DC), where initial data reconstruction is performed, and where a copy is archived to long-term tape storage. Even after the drastic data reduction performed by the experiments, the CERN DC processes on average one petabyte of data per day. This is how the the milestone of 200 petabytes of data permanently archived in its tape libraries was reached on 29 June.

Breaking data records bit by bit

by Harriet Jarlett


Magnetic tapes, retrieved by robotic arms, are used for long-term storage (Image: Julian Ordan/CERN)

This year CERN's data centre broke its own record, when it collected more data than ever before.

During October 2017, the data centre stored the colossal amount of 12.3 petabytes of data. To put this in context, one petabyte is equivalent to the storage capacity of around 15,000 64GB smartphones. Most of this data come from the Large Hadron Collider's experiments, so this record is a direct result of the outstanding LHC performance, the rest is made up of data from other experiments and backups.

"For the last ten years, the data volume stored on tape at CERN has been growing at an almost exponential rate. By the end of June we had already passed a data storage milestone, with a total of 200 petabytes of data permanently archived on tape," explains German Cancio, who leads the tape, archive & backups storage section in CERN's IT department.

Ground-breaking ceremony for the High-Luminosity LHC

by Corinne Pralavorio

Posted by Corinne Pralavorio on 26 Jun 2018. Last updated 26 Jun 2018, 16.21.
Voir en français

The civil engineering work for the High-Luminosity LHC gets under way. Here we see the earthmovers at work on the new 80 metre access shaft at Point 5. (Image: Julien Ordan/CERN)

The earthmovers are at work on the ATLAS site in Meyrin and at CMS in Cessy, digging the new shafts for the High-Luminosity LHC (HL-LHC). The start of the work for this new phase of the project was marked by a ceremony held on 15 June, which was attended by VIP guests including the President of the State Council of the Republic and Canton of Geneva, the Prefect of the Rhône-Alpes-Auvergne region, the Mayor of Meyrin, the Deputy Mayor of Cessy and representatives of CERN's Member and Associate Member States.

"All the chapters of CERN's history have begun with a shovel of earth, and each chapter has begun with the promise of great progress in fundamental knowledge, new technologies that benefit society, and collaboration on a European and now a global scale. This was true of the Large Hadron Collider (LHC) and its experiments and it is true of the project for which we are gathered here today," said Fabiola Gianotti, CERN Director-General.

Time to adapt for big data

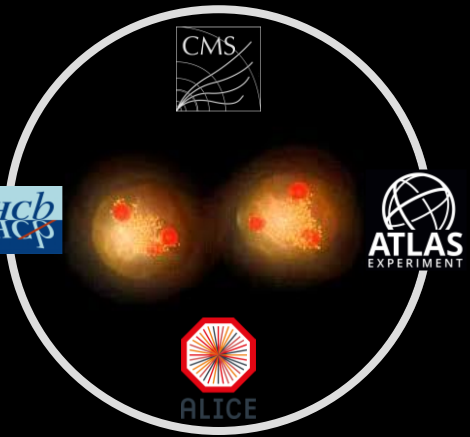
Radical changes in computing and software are required to ensure the success of the LHC and other high-energy physics experiments into the 2020s, argues a new report.

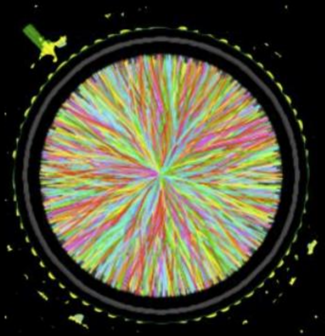
It would be impossible for anyone to conceive of carrying out a particle-physics experiment today without the use of computers and software. Since the 1960s, high-energy physicists have pioneered the use of computers for data acquisition, simulation and analysis. This hasn't just accelerated progress in the field, but driven computing technology generally – from the development of the World Wide Web at CERN to the massive distributed resources of the Worldwide LHC Computing Grid (WLCG) that supports the LHC experiments. For many years these developments and the increasing complexity of data analysis rode a wave of hardware improvements that saw computers get faster every year. However, those blissful days of relying on Moore's law are now well behind us (see panel overleaf), and this has major ramifications for our field.

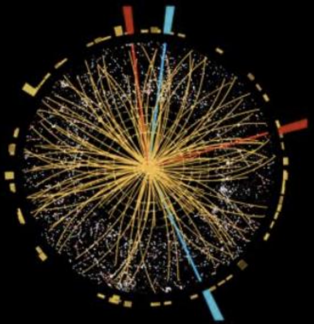
The high-luminosity upgrade of the LHC (HL-LHC), due to enter operation in the mid-2020s, will push the frontiers of accelerator and detector technology, bringing enormous challenges to software and computing (*CERN Courier* October 2017 p5). The scale of the HL-LHC data challenge is staggering: the machine will collect almost 25 times more data than the LHC has produced up to now, and the total LHC dataset (which already stands at almost 1 exabyte) will grow many times larger. If the LHC's ATLAS and CMS experiments project their current computing models to Run 4 of the LHC in 2026, the CPU and disk space required will jump by between a factor of 20 to 40 (figures 1 and 2).

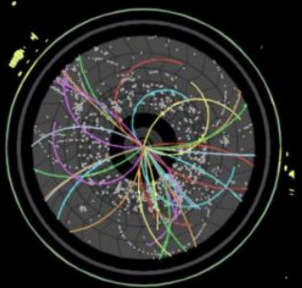
Even with optimistic projections of technological improvements there would be a huge shortfall in computing resources. The WLCG hardware budget is already around 100 million Swiss francs per year and, given the changing nature of computing hardware and slowing technological gains, it is out of the question to simply throw

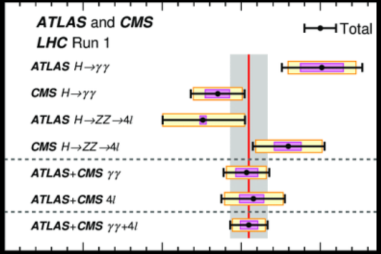
Inside the CERN computer centre in 2017.
(Image credit: J Ordan/CERN.)





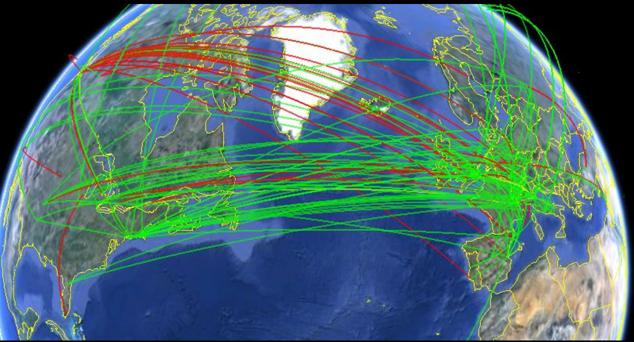






Data Storage - Data Processing - Event generation - Detector simulation - Event reconstruction - Resource accoun

Distributed computing - Middleware - Workload management - Data management - Monitoring



GAUDI-LHCb

ATHENA-ATLAS

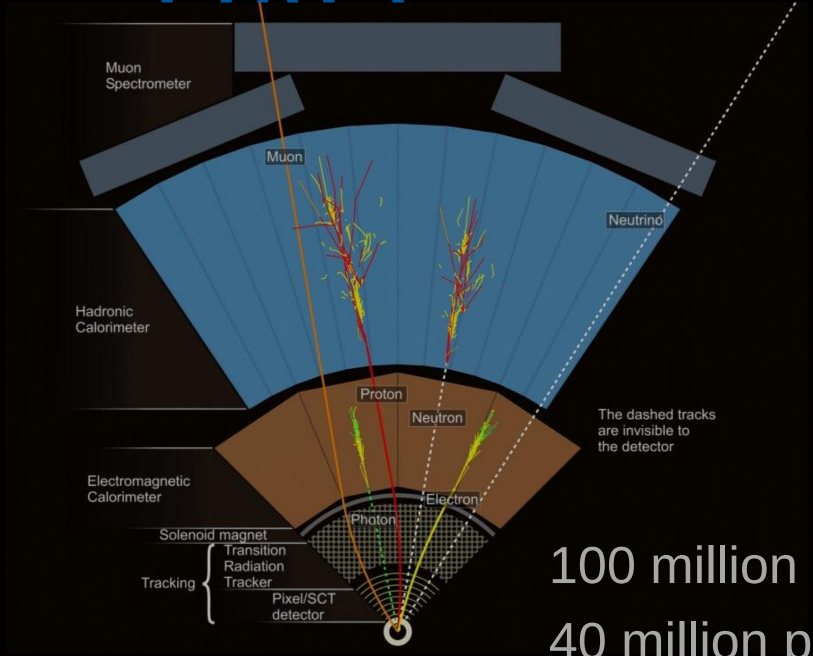
CMSSW-CMS

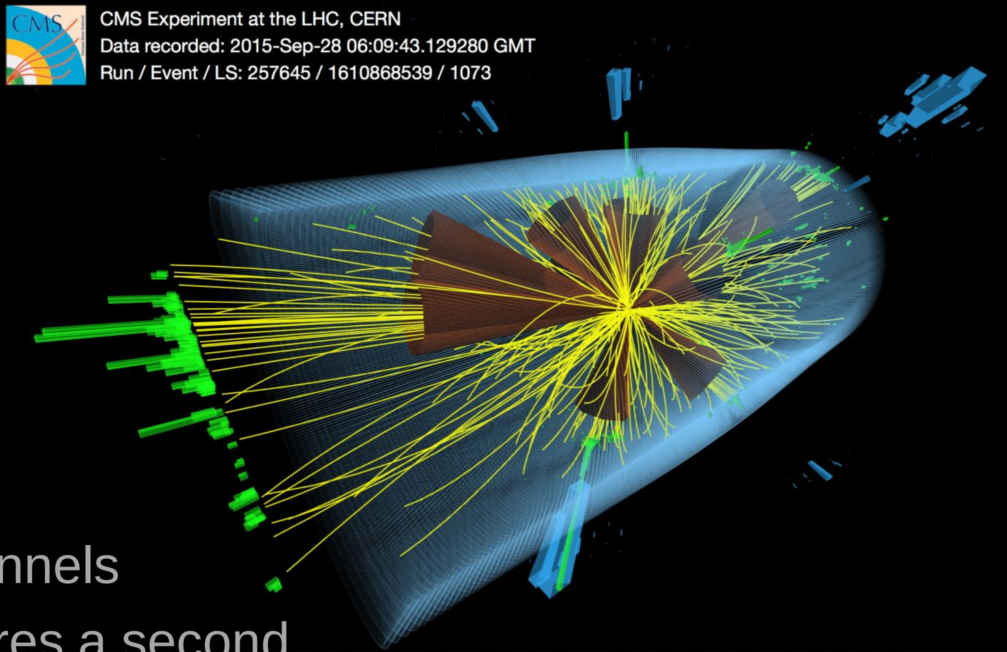
GEANT4 A SIMULATION TOOLKIT

SHERPA

PYTHIA

From the Hit to the Bit: DAQ





100 million channels

40 million pictures a second

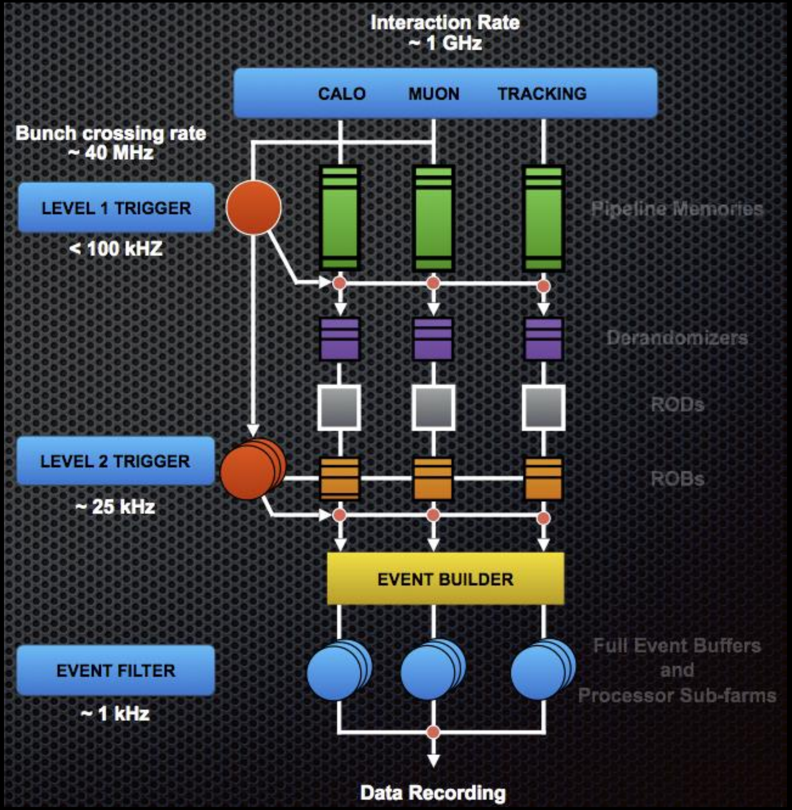
Synchronised signals from all detector parts

From the Hit to the Bit: event filtering (1/2)

- L1: 40 million events per second
  - Fast, simple information
  - Hardware trigger in a few micro seconds

- L2: 100 thousand events per second
  - Fast algorithms in local computer farm
  - Software trigger in <1 second

- EF: Few 100 per second recorded for study

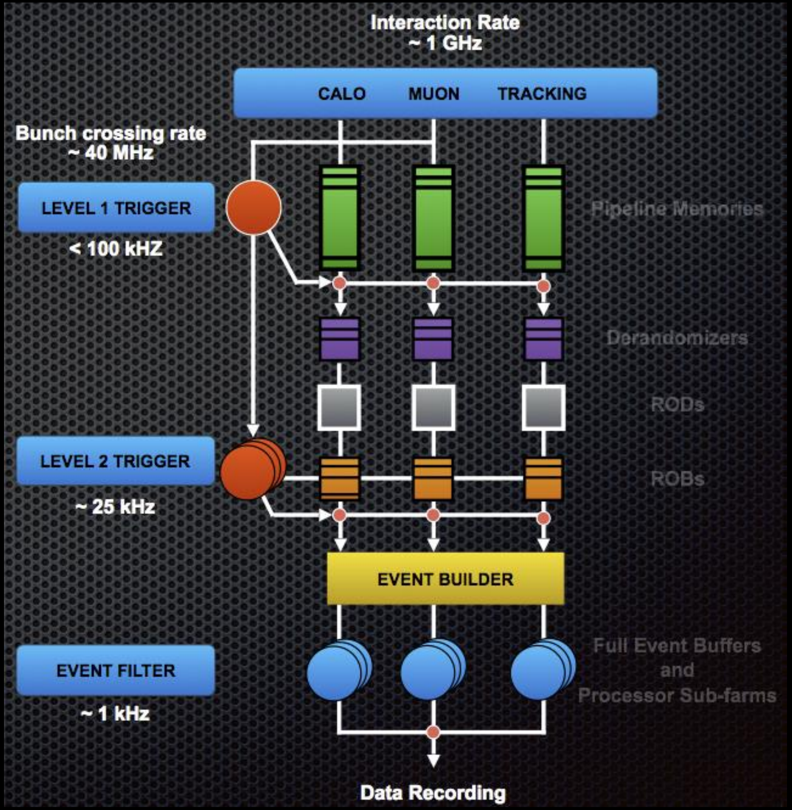


From the Hit to the Bit: event filtering (2/2)

- L1: this is ~1 Petabyte per second!
  - Cannot afford to store it
  - 1 year's worth of LHC data at 1 PB/s would cost few hundred trillion euros

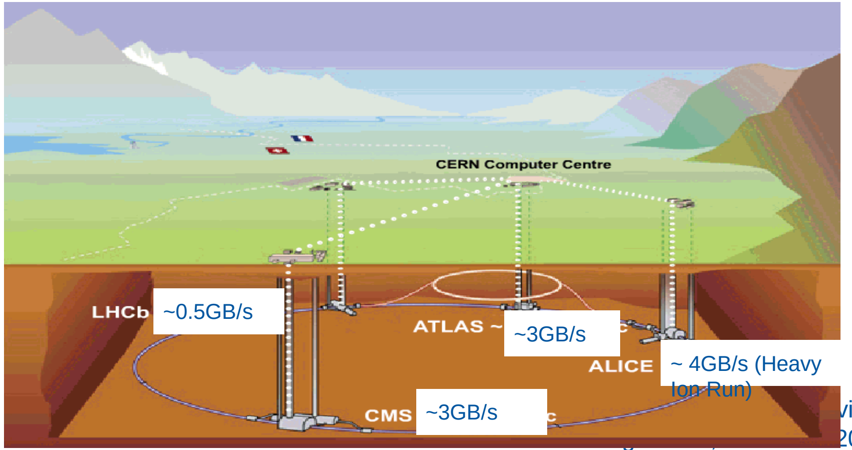
- Have to filter in real time to keep only "interesting" data
  - We keep ~1 event in a million
  - Yes, 99.9999% is thrown away

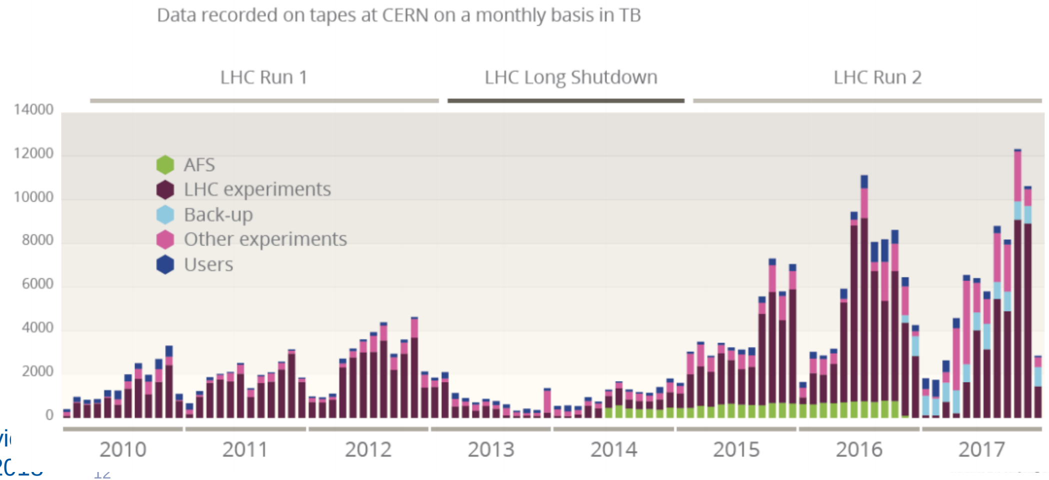
- Final rate is O(Gigabyte per second)*



Data Processing

- Experiments sent 70 Petabytes of data in 2017 year
  - 40 Petabytes from the four LHC experiments
- The LHC data is aggregated at the CERN data centre to be stored and processed





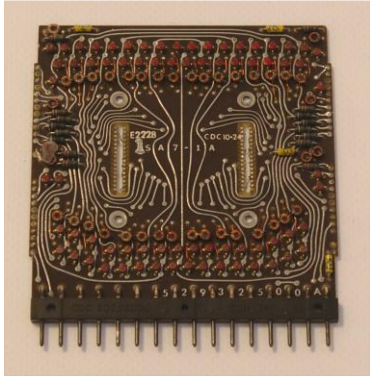
CERN Data Center

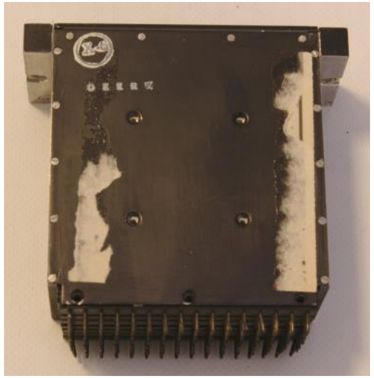
- Built in the 70s on the CERN site (Meyrin-Geneva)
  - 3.5 MW for equipment
- Extension located at Wigner (Budapest)
  - 2.7 MW for equipment
  - Connected to the Geneva CC with 3x100Gb links (21 and 24 ms RTT)
- Hardware generally based on commodity
- **15,000** servers, providing **230,000** processor cores
- **90,000** disk drives providing **280,000 TB** disk space
- **30,000** tapes drives, providing **0.4EB** capacity (1EB=1000PB)

CDC7600 SUPERCOMPUTER (1972)

60-bit word size and 36MHz processor

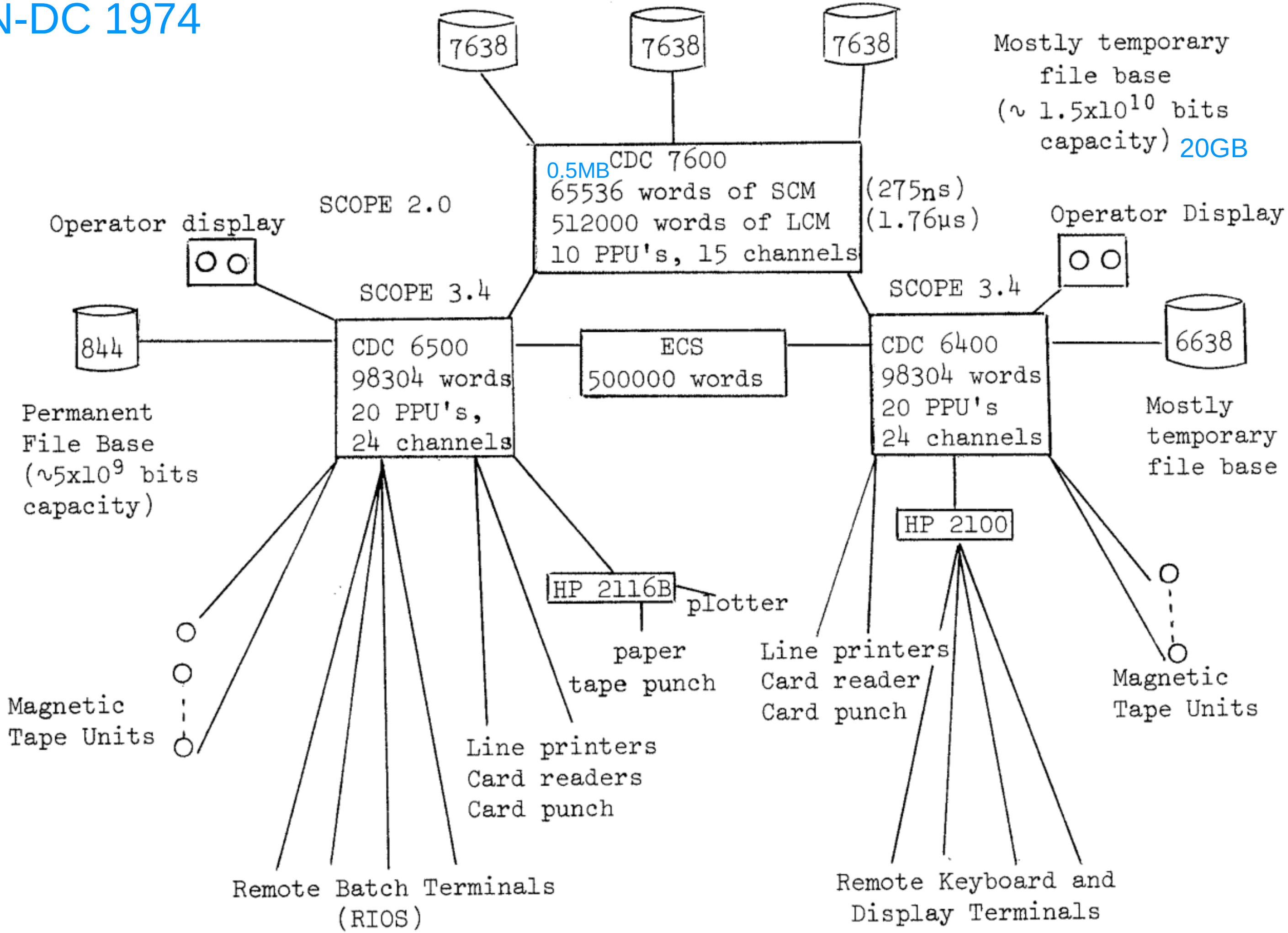
The computing centre was enlarged with the installation of a CDC 7600, which is shown here during its assembly in February 1972.





http://cerncourier.com/cws/article/cnl/24597
https://en.wikipedia.org/wiki/CDC_7600
https://videos.cern.ch/record/43172
https://videos.cern.ch/record/43113

CERN-DC 1974

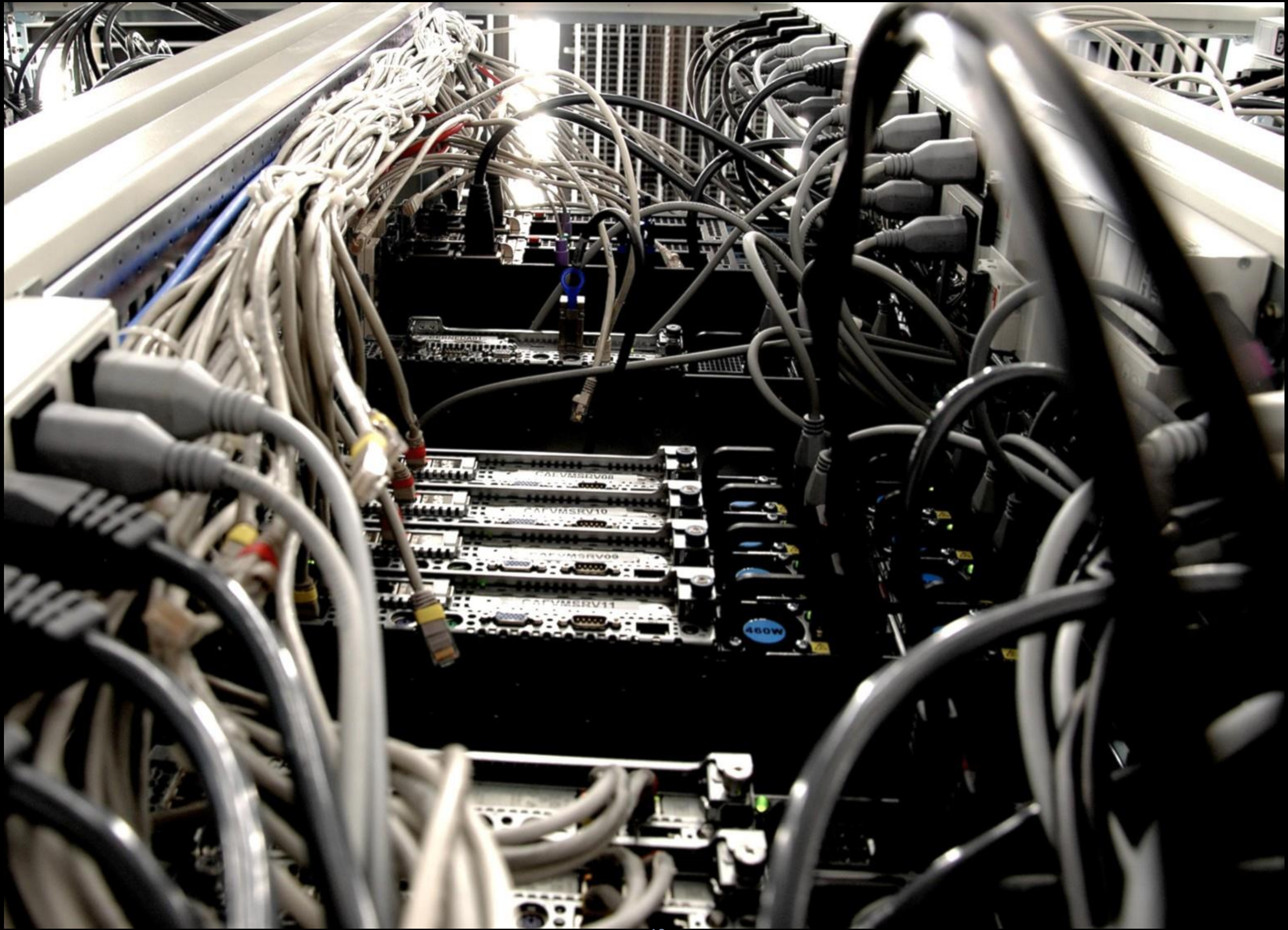


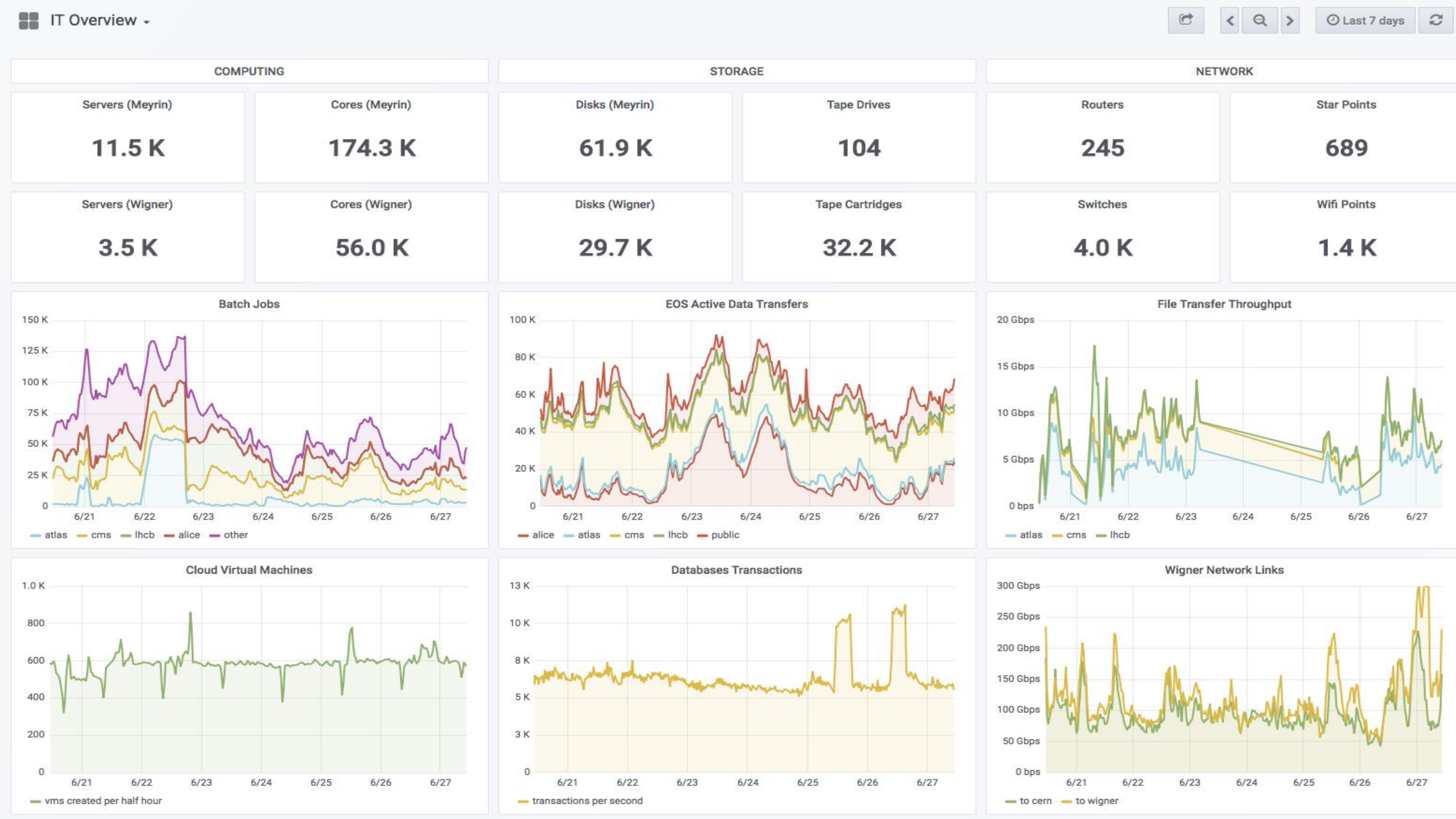
7638
7638
7638
Mostly temporary file base (∼ 1.5x10^10 bits capacity) 20GB
0.5MB CDC 7600
65536 words of SCM
512000 words of LCM
10 PPU's, 15 channels
(275ns)
(1.76μs)
SCOPE 2.0
Operator display
Operator Display
SCOPE 3.4
SCOPE 3.4
844
CDC 6500
98304 words
20 PPU's,
24 channels
ECS
500000 words
CDC 6400
98304 words
20 PPU's
24 channels
6638
Permanent File Base (∼5x10^9 bits capacity)
Mostly temporary file base
HP 2100
HP 2116B
plotter
paper tape punch
Line printers
Card reader
Card punch
Magnetic Tape Units
Magnetic Tape Units
Line printers
Card readers
Card punch
Remote Batch Terminals (RIOS)
Remote Keyboard and Display Terminals

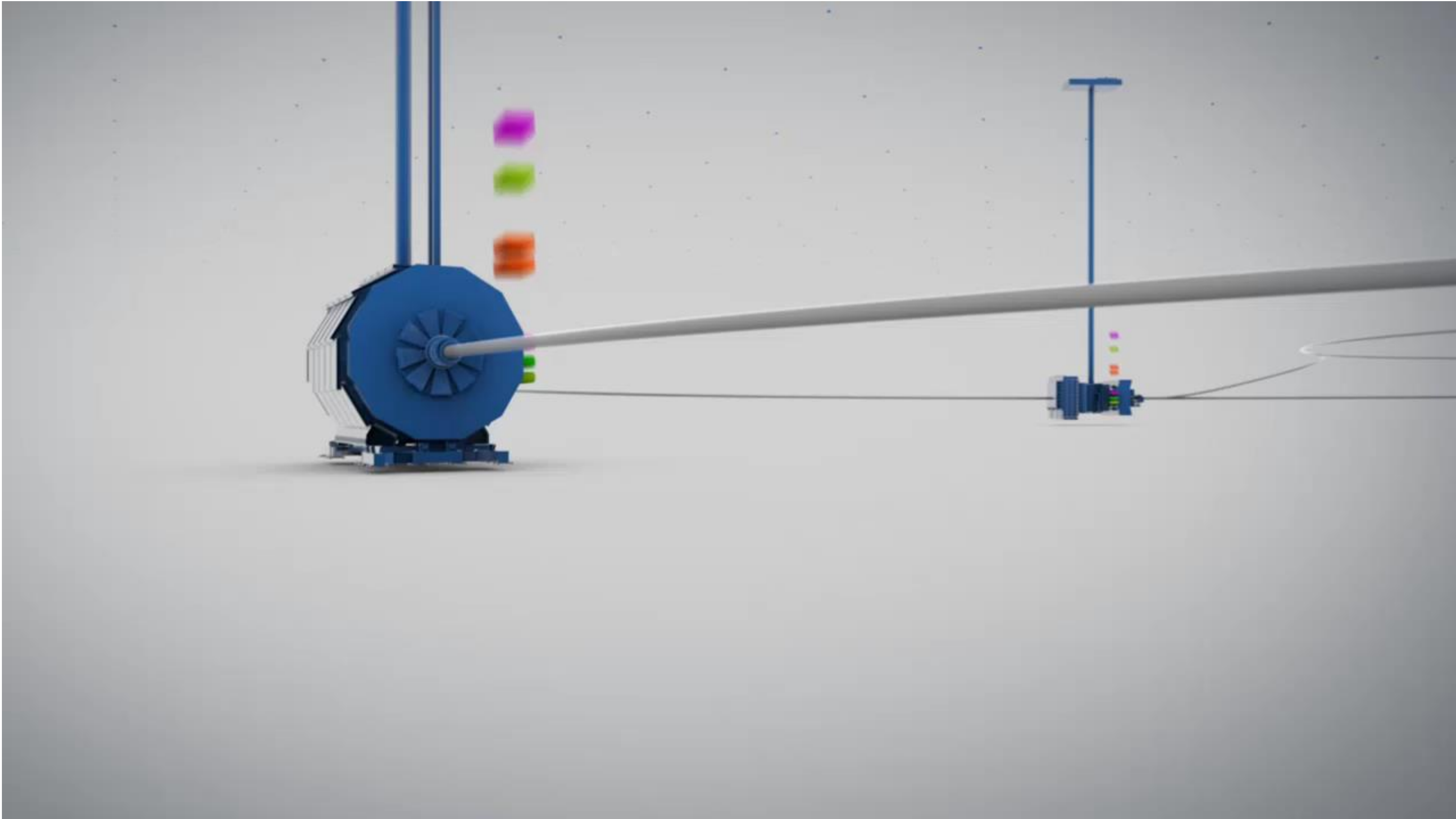
CERN Data Center

CERN Data Center

CERN Data Center

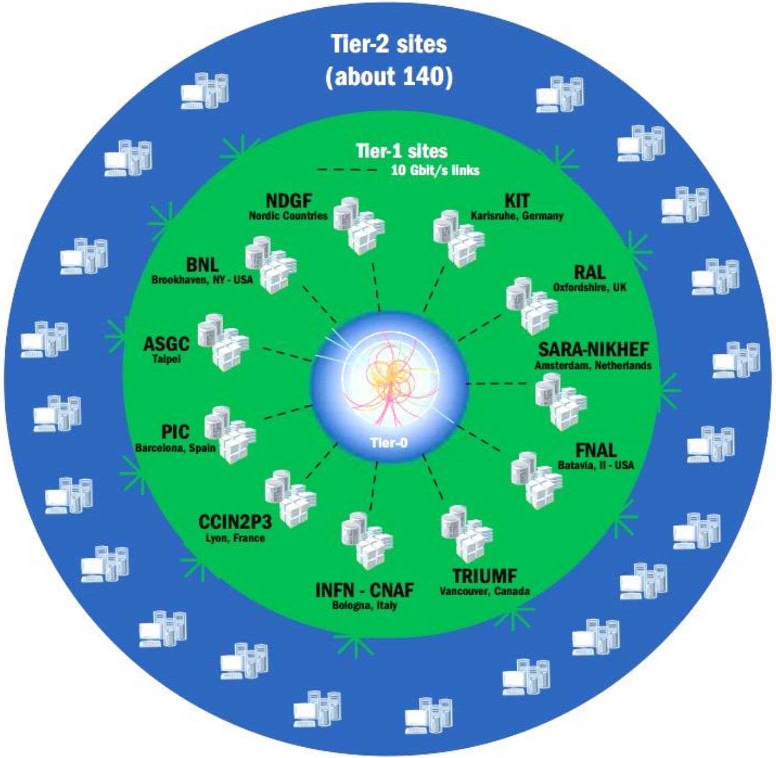



IT Overview
Last 7 days
COMPUTING
STORAGE
NETWORK
Servers (Meyrin)
11.5 K
Cores (Meyrin)
174.3 K
Disks (Meyrin)
61.9 K
Tape Drives
104
Routers
245
Star Points
689
Servers (Wigner)
3.5 K
Cores (Wigner)
56.0 K
Disks (Wigner)
29.7 K
Tape Cartridges
32.2 K
Switches
4.0 K
Wifi Points
1.4 K
Batch Jobs
atlas cms lhcb alice other
EOS Active Data Transfers
alice atlas cms lhcb public
File Transfer Throughput
atlas cms lhcb
Cloud Virtual Machines
vms created per half hour
Databases Transactions
transactions per second
Wigner Network Links
to cern to wigner



Worldwide LHC Computing Grid (WLCG)

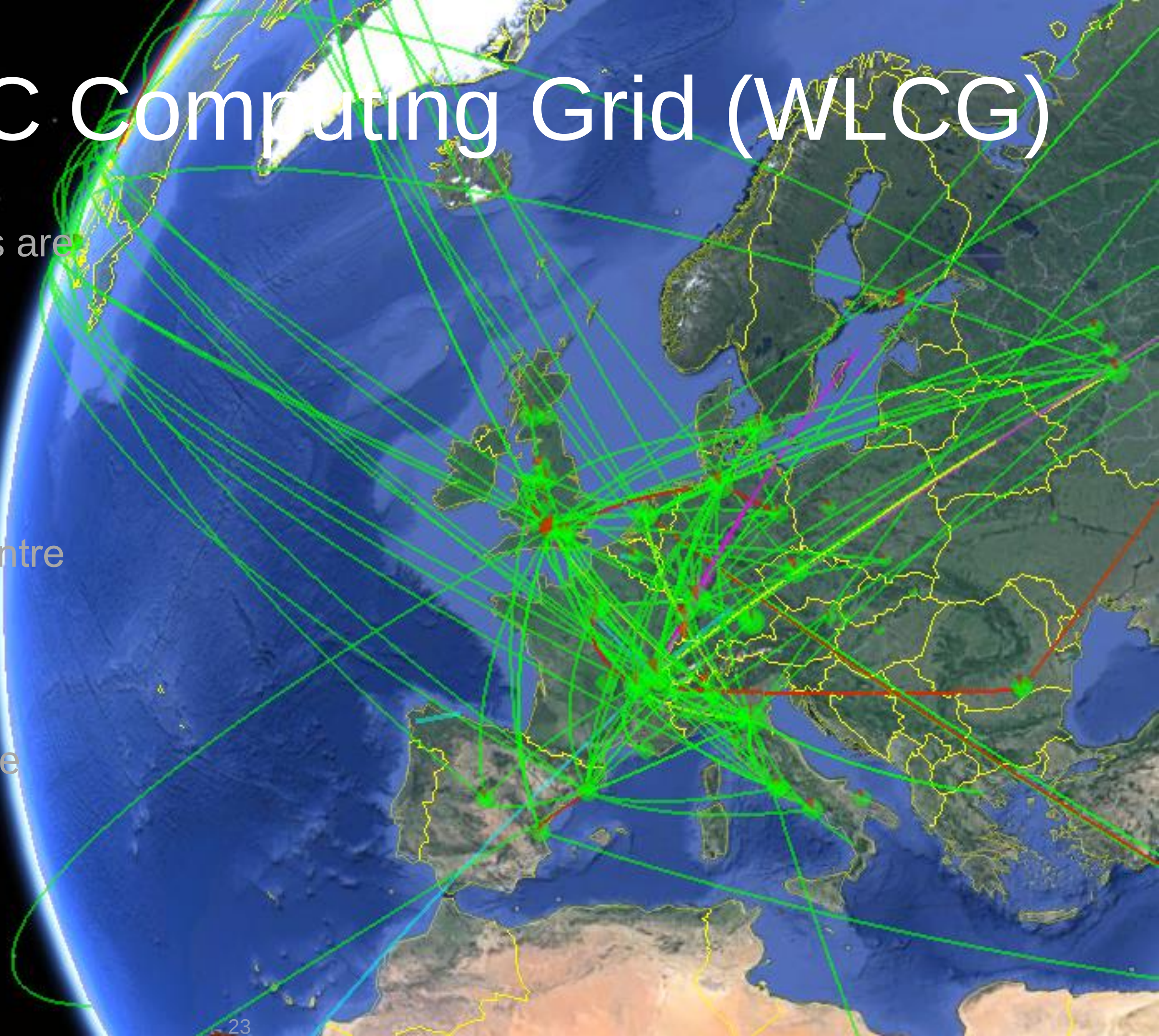
- The Worldwide LHC Computing Grid (WLCG) is a global collaboration of more than 170 data centres around the world, in 42 countries
- The CERN data centre (Tier-0) distributes the LHC data worldwide to the other WLCG sites (Tier-1 and Tier-2)
- WLCG provides global computing resources to store, distribute and analyse the LHC data
- The resources are distributed – for funding and sociological reasons



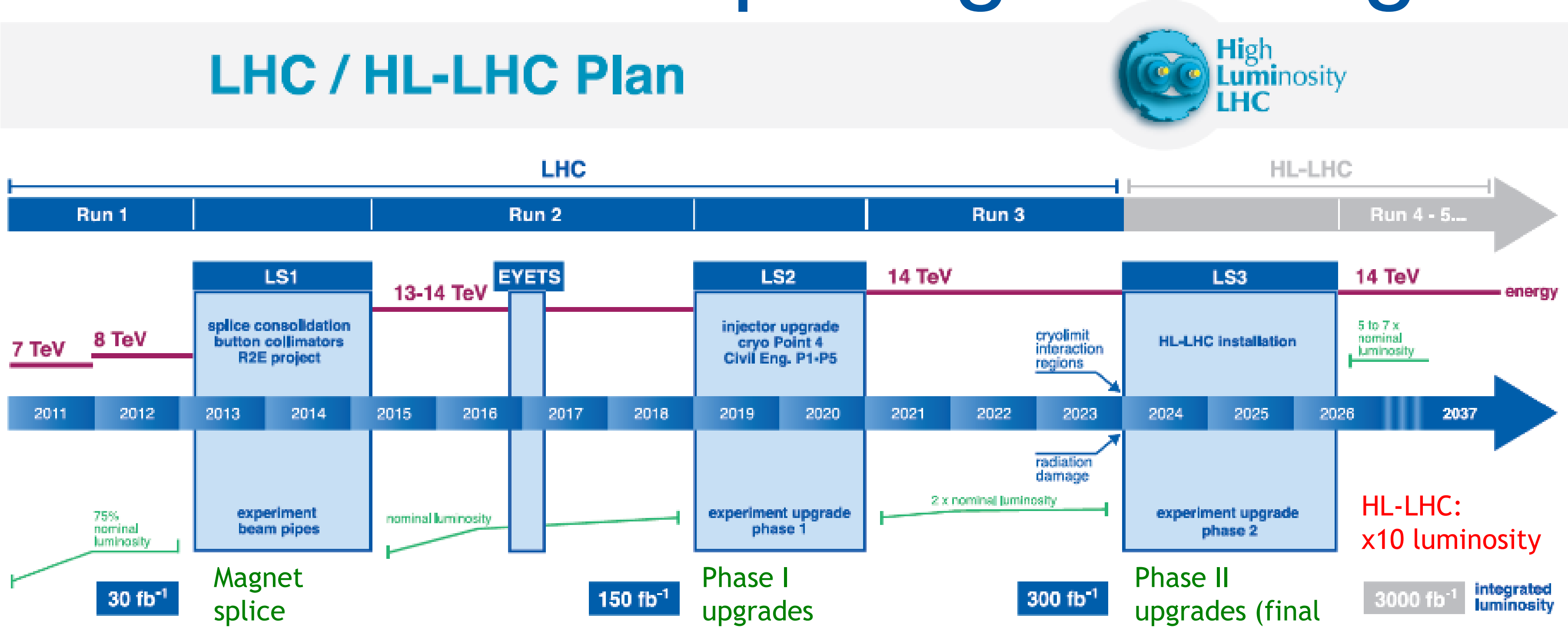

Tape Storage
Tape Storage
Tape Storage

Worldwide LHC Computing Grid (WLCG)

- A community of 10,000 physicists are the WLCG users
- On average around 250,000 jobs running concurrently
- 600,000 processing cores
- 15% of the WLCG computing resources are at CERN's data centre
- 500 petabytes storage available worldwide
- 20-40-80-100 Gbit/s optical-fiber links connect CERN to each of the 13 Tier-1 institutes



HL-LHC: a computing challenge

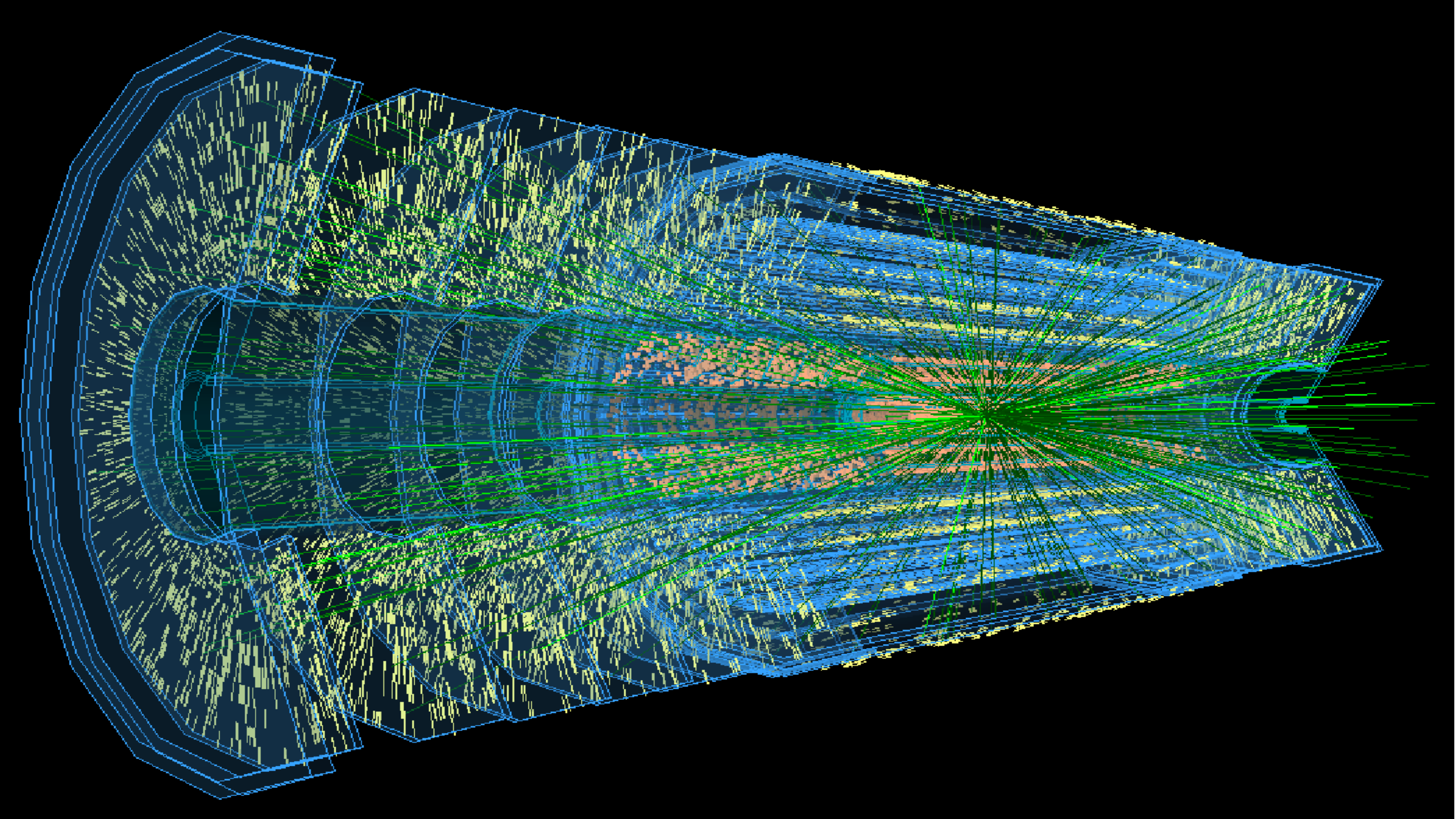


Magnet splice update

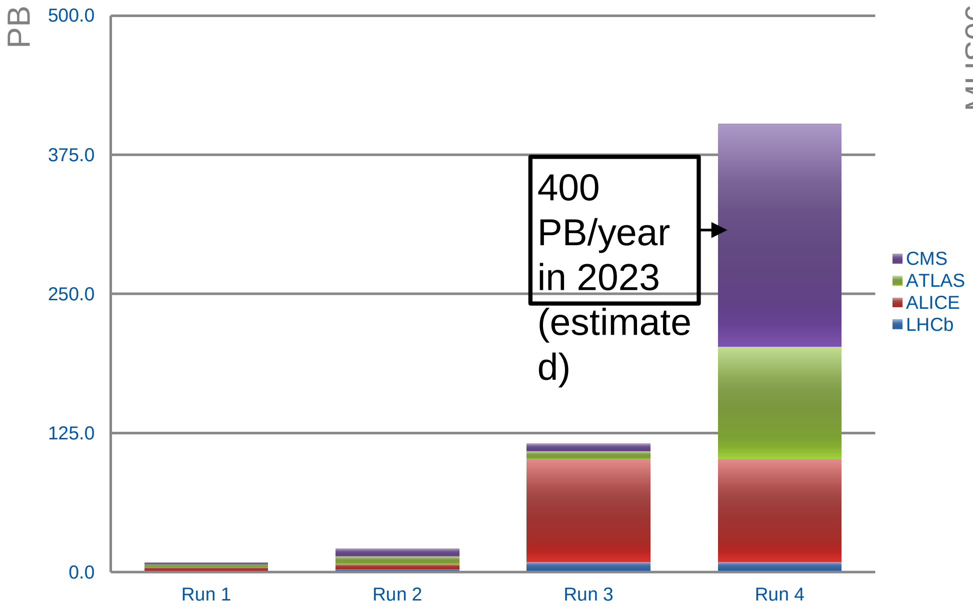
Phase I upgrades (injectors)

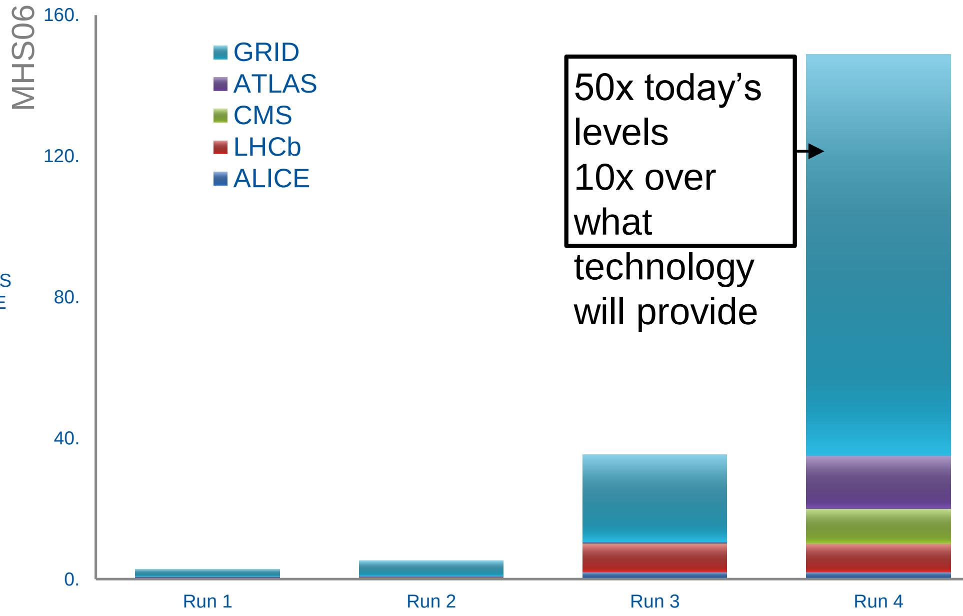
Phase II upgrades (final focus)

HL-LHC: x10 luminosity



HL-LHC: a computing challenge





Future in Computing

- Manage storage and computing continuous growth (budget and infrastructures)
  - Storage technologies evolution: HDD, SSD, Tapes
  - CPU speed and multicore/vector exploitation
  - Data Center engineering: optimize energy consumption (PUE and green IT)
- Provide to the experiments and the users the computing requirements while optimizing resources
  - Worldwide LHC Computing Grid (WLCG)
- Improve software performance
  - HEP (High Energy Physics) Software Foundation, HSF
- Data preservation:
  - “We are nonchalantly throwing all of our data into what could become an information black hole without realising it.” Vint Cerf, vice president of Google and an early internet pioneer, February 2015
  - How to ensure that all the data collected and published is still readable by the next generations … and how to make sense out of it
- CERN is leading a global effort for HEP, that others will inevitably face soon or later

CERN-IT: pushing boundaries

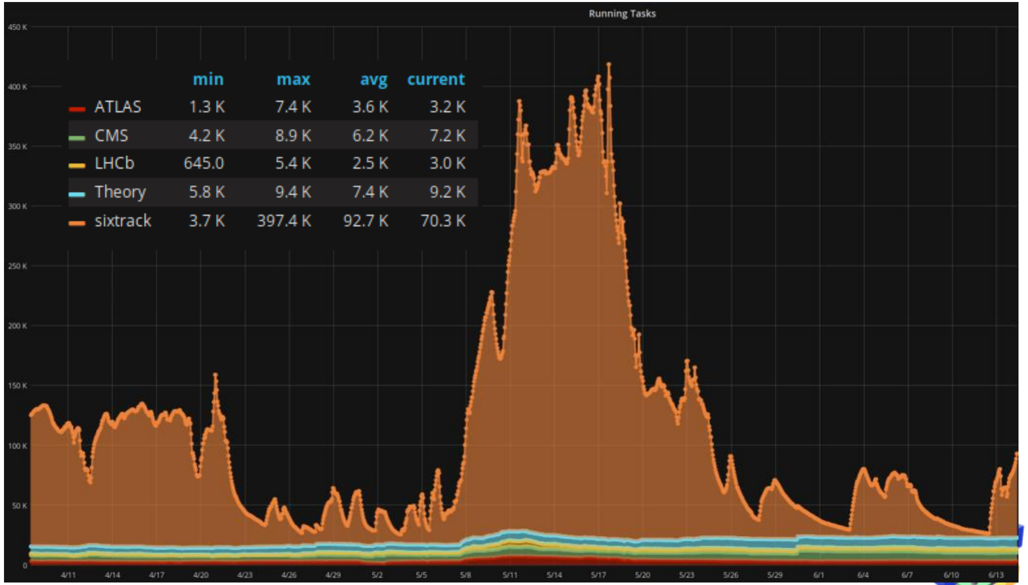
- CERN-IT impact on society through computing:
  - Need for collaboration tools for Global Science led to invent the **World Wide Web**
  - Need for collaboration of computing resources for the Global LHC led to adopt **Grid Computing** and first concept of **Computing Clouds**
  - Need for sharing the results had led CERN to pave to way to open access to documents and now data: **LHC@home** and **CERN Opendata Portal**

- Could these technologues have been originated somewhere else?
  - Probably. But often we are faced to challenges 5 to10 years before others, pushed by the needs of the detectors and accelerators. Pushed by fundamental science.

LHC@Home

- Simulations from ATLAS, CMS, LHCb and Theory running under CernVM and VirtualBox
- You can contribute by running BOINC on your computer outside working hours
- The BOINC client can be configured to run 17:30-8:30 or when your computer is idle

@
Volunteer computing for the LHC



http://lhcathome.web.cern.ch/

opendata
CERN

About

CERN OPENDATA

Explore more than **1 petabyte**
of open data from particle physics!

Start typing...

Search

search examples: collision datasets, keywords:education, energy:7TeV

Explore

datasets
software
environments
documentation

Focus on

ATLAS
ALICE
CMS
LHCb
OPERA

Muchas gracias!