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

Teacher Programme, 28th June 2018

• Brain, stone, papyrus/jeroglifics, books

photo @ https://seanmunger.com/
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.
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 milestone of 200 petabytes of data permanently archived in its tape libraries was reached on 29 June.

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 Canelo, who leads the tape, archive & backups storage section in CERN’s IT department.
Ground-breaking ceremony for the High-Luminosity LHC

by Corinne Pralavorio

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.
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 microseconds

- **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 Data Center
CERN Data Center
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.
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

LHC / HL-LHC Plan

<table>
<thead>
<tr>
<th>LHC</th>
<th>HL-LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>Run 4 - 5...</td>
</tr>
<tr>
<td>Run 2</td>
<td></td>
</tr>
<tr>
<td>Run 3</td>
<td></td>
</tr>
</tbody>
</table>

**Magnet splice update**

**Phase I upgrades (injectors)**

**Phase II upgrades (final focus)**

**HL-LHC: x10 luminosity**
Computación en el CERN: al servicio de la ciencia y su impacto en la sociedad

Teacher Programme, 28th June 2018
HL-LHC: a computing challenge

- 400 PB/year in 2023 (estimated)
- 50x today’s levels
- 10x over what technology will provide
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 to 10 years before others, pushed by the needs of the detectors and accelerators. Pushed by fundamental science.
LHC@Home

- Simulaciones de ATLAS, CMS, LHCb y Theory ejecutándose bajo CernVM y VirtualBox

- Puedes contribuir corriendo BOINC en tu computadora fuera de las horas de trabajo

- El cliente BOINC se puede configurar para correr de 17:30 a 8:30 o cuando tu computadora está inactiva

http://lhcathome.web.cern.ch/
Muchas gracias!