From Data to Knowledge

CERN Open Days
14 & 15 September 2019
Between LHC Particles Colliding …

Lately inside the LHC: 2 protons 0.0000000000000001 sec before the collision
… and the Presentation of Physics Results

Press coverage of the CERN Seminar in July 2012 on the Higgs boson search results
Computing is Involved in All Stages
Definition: Event

“Observation of the collision of two particles and their subsequent production of new particles”
A Particle’s Life

Detector

Selection

Data Preparation & Reconstruction

Publication

Data Analysis

30 Million events per second

1.000 events per second

Month(s) - Year(s)

Year(s)

Week(s) - Month(s)

# events

Background

Signal

Relevant quantity

30 Million events per second

1.000 events per second

Year(s)
Selection of events (Trigger)

- Filtering stages installed in the experimental areas around the LHC
- Level 1 trigger implemented in hardware
- High Level Trigger (HLT) running on farms of several 10 thousand computing cores
- Final filtered collisions then exported to CERN computer center and further worldwide for next processing stages
Software Trigger Selection
(every dot represents 1,000 events)
Google searches 98 PB

LHC data ~200 PB

Facebook uploads 180 PB

SKA Phase 1 – 2023
~300 PB/year science data

HL-LHC – 2016
50 PB raw data

LHC – 2016
50 PB raw data

Google Internet archive
~15 EB

Yearly data volumes

SKA Phase 2 – mid-2020’s
~1 EB science data

HL-LHC – 2026
~1 EB Physics data

HL-LHC – 2026
~600 PB Raw data

Facebook
Phase 1
180 PB

~200 PB

~15 EB

~300 PB/year

~1 EB science data

~1 EB Physics data
The Worldwide LHC Computing Grid (WLCG)

- 170 computing sites in 42 countries providing
  - Up to 900,000 computer cores
  - ~1 exa-byte of disk and tape storage

→ Used for all later remaining data processing
Data Reconstruction

Convert measured data points by the detector and energy into “tracks” and particles

→ Physics data ready for analysis
Simulation

- Particle collision and detector response simulated with computer programs
- Simulation is essential for data analysis and design of new detectors
- Majority of WLCG resources needed for simulation
Data Analysis

Physicists analyze the reconstructed data with statistical methods to verify or falsify theoretical predictions.
You can help to process LHC data at home …

http://lhcathome.web.cern.ch/
Demonstrator in the restaurant on the other side of the road

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

• Many thanks to Anna Sfyrla (U. Geneva) for sharing her summer student lecture slides !!!
• Icons from https://www.flaticon.com/
• Many thanks for comments and suggestions from Concezio Bozzi, Alessandro Di Girolamo, Maarten Litmaath, Servesh Muralidharan, Markus Schulz, Andrea Sciabà
• Translation of slides by Julien Collet, Herve Rousseau, Nicole Cremel, Catherine Delamare, Melissa Gaillard
Backup
Selection of events (Trigger)

- Filtering stages installed in the experimental areas around the LHC
  - Level 1 trigger implemented in hardware
  - High Level Trigger (HLT) running on farms of several 10 thousand computing cores
- Final filtered collisions then exported to CERN computer center and further worldwide for next processing stages