CERN openlab
for DataGrid applications

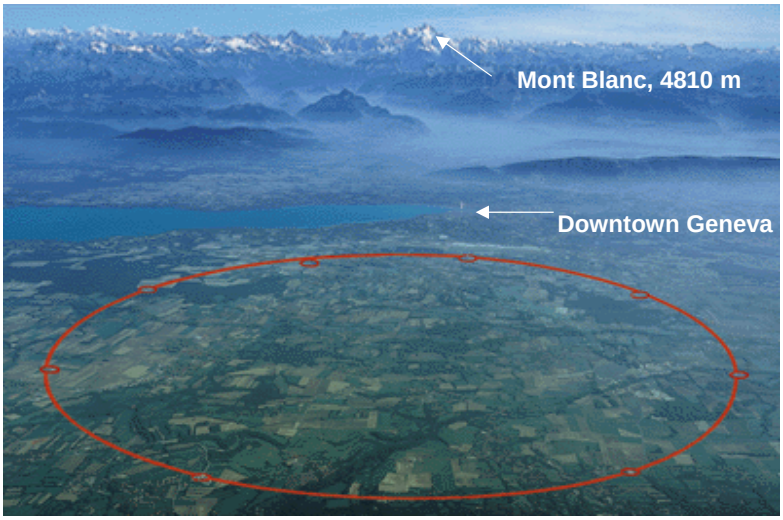
CERN, the LHC and the Grid: an introduction

Francois Grey
CERN IT Department

What is CERN?

- CERN is the world's largest particle physics centre
- 2500 staff (mainly engineers)
- 6500 visiting scientists (1/2 of world's particle physicists)
- Current project: the Large Hadron Collider

CERN site



What is the LHC?

- LHC is a 27km circumference accelerator due to start operation in 2007. It will be the largest scientific instrument on the planet.
- LHC will collide beams of protons at 14 TeV, using the latest super-conducting magnet technologies. It will operate at about -270^{0}C, making it the coldest place in the universe!
- Four major experiments measure results of proton collisions: ALICE, ATLAS, CMS, LHCb. These underground detectors are as big as cathedrals, built by teams of 1000+ scientists.

The LHC Data Challenge

- A particle collision = an event
- Physicist's goal is to count, trace and characterize all the particles produced and fully reconstruct the process.
- Among all tracks, the presence of "special shapes" is the sign for the occurrence of interesting interactions.

- Example the Higgs boson:
  - look for characteristic decay pattern producing 4 muons
  - Record >8 million events per day
  - On average only one event per day will be a Higgs boson

LHC Data

- 40 million collisions per second
- After filtering, 100 collisions of interest per second
- > 1 Megabyte of data per collision
  recording rate > 1 Gigabyte/sec
- 10^{10} collisions recorded each year
  stored data > 10 Petabytes/year

1 Megabyte (1MB)
A digital photo

1 Gigabyte (1GB)
= 1000MB
5GB = A DVD movie

1 Terabyte (1TB)
= 1000GB
World annual book production

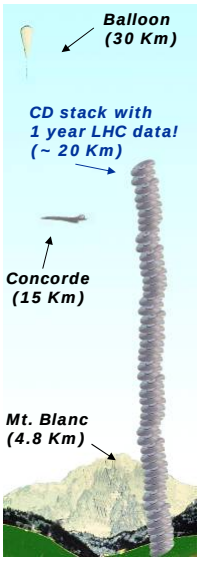
1 Petabyte (1PB)
= 1000TB
Annual production of one LHC experiment

1 Exabyte (1EB)
= 1000 PB
3EB = World annual information production

Challenges: Data Storage

- LHC data correspond to about 20 million CDs each year!

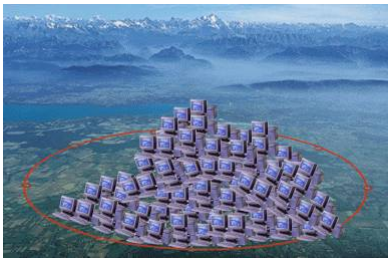
Where will the experiments store all of these data?



Challenges: Data Processing

- LHC data analysis requires a computing power equivalent to ~ 100,000 of today's PC processors!

Where will the experiments find such a computing power?



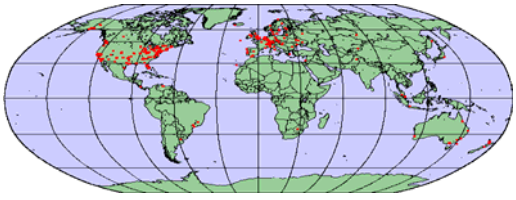
Computing at CERN

- High-throughput computing based on reliable "commodity" technology
- 2000 dual processor PCs
- 3 Petabytes of data on disk and tapes

Nowhere near enough!

Computing for LHC

- Problem: even with Computer Centre upgrade, CERN can provide only a fraction of the necessary resources.
- Solution: CERN has over 250 partner institutes in Europe, over 200 in rest of the world. Most have significant computing resources. Build a Grid that **unites these computing resources.**

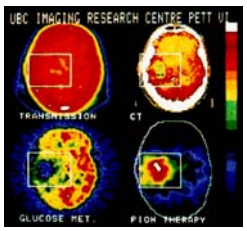


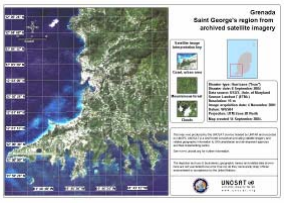
What is the Grid? (I)

- The World Wide Web provides seamless access to information that is stored in many millions of different geographical locations
- In contrast, the Grid is an emerging infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe.

Grid Applications for Science

- Medical/Healthcare *(imaging, diagnosis and treatment)*
- Bioinformatics *(study of the human genome and proteome to understand genetic diseases)*
- Nanotechnology *(design of new materials from the molecular scale)*
- Engineering *(design optimization, simulation, failure analysis and remote Instrument access and control)*
- Natural Resources and the Environment *(weather forecasting, earth observation, modeling and prediction of complex systems)*





Grid @ CERN

- CERN & scientific partners project:
  - LHC Computing Grid (LCG)

- EU-funded project led by CERN:
  - Enabling Grids for E-SciencE (EGEE)

- Industry & CERN funded project:
  - CERN openlab for DataGrid applications