CERN openlab

for DataGrid applications

The Grid and CERN: an overview

Francois Grey
CERN IT Department

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

CERN openlab
for DataGrid applications

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

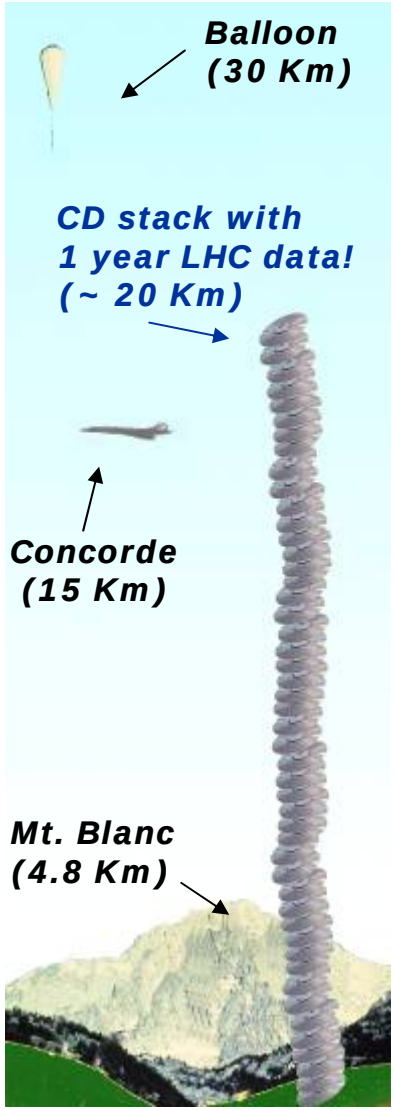
1 Petabyte (1PB)
= 1000TB
Annual production of one LHC experiment

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

Challenges: Data Storage

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

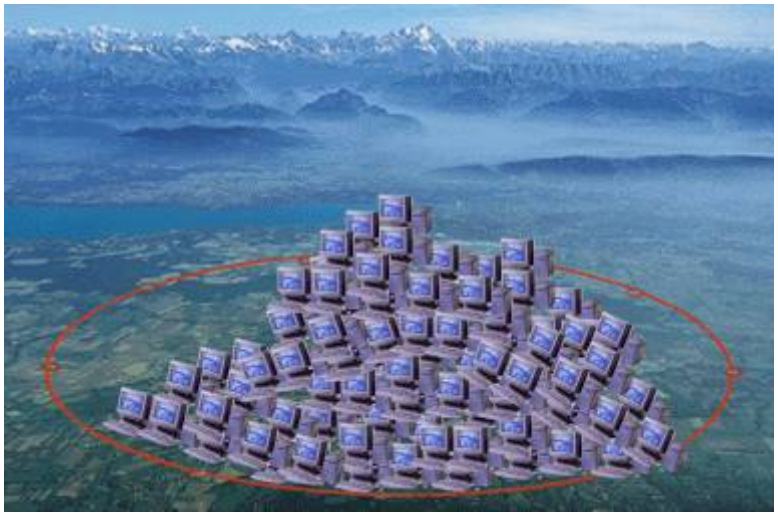
Where will the experiments store all of these data?



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?



CERN openlab
for DataGrid applications

Computing at CERN

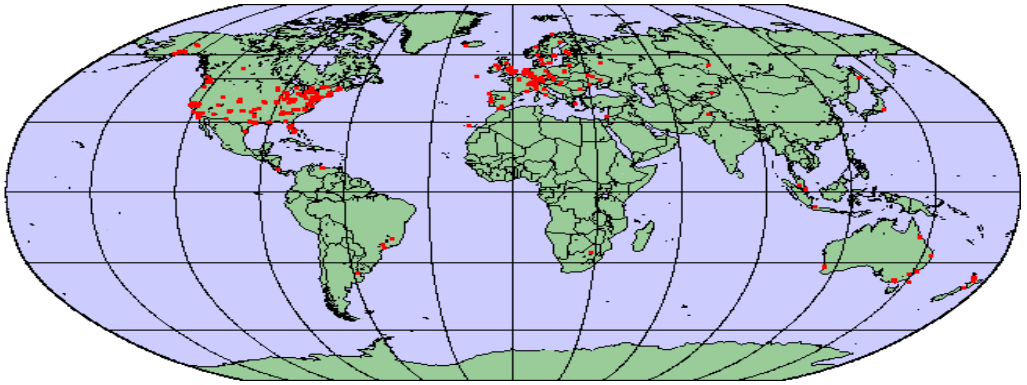
- High-throughput computing based on reliable “commodity” technology
- 2000 dual processor PCs
- 3 Petabytes of data on disk and tapes

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.**



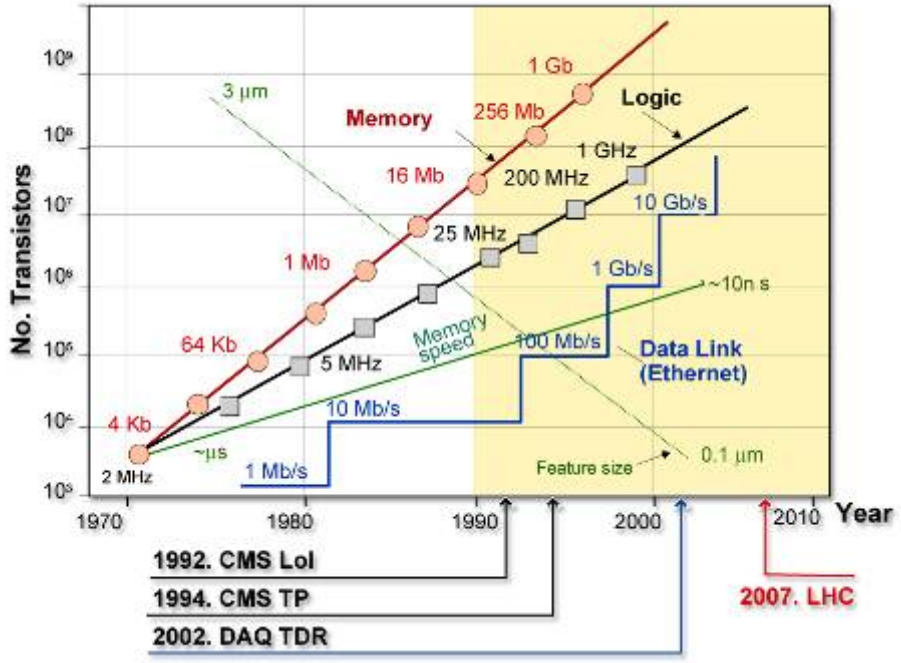
What is the Grid? (I)

- 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.

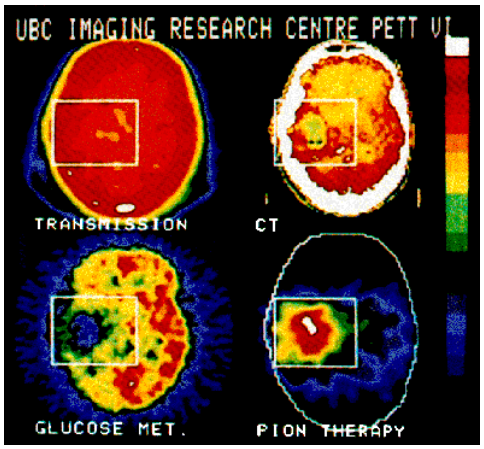
What is the Grid? (II)

- The Grid has ancestors in distributed computing (e.g. metacomputing). Difference now is global scale, due to data transfer rates evolving more rapidly than Moore's law for processors and memory.



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 projects:
  - LHC Computing Grid (LCG)
- EU-funded projects led by CERN:
  - Enabling Grids for E-SciencE (EGEE)
- Industry funded projects:
  - CERN openlab for DataGrid applications

CERN openlab student programme

- 15 Students from Europe, US, Asia (BSc-MSc-PhD levels)
- Projects involving Grid technology, team-based
- Co-funding from participating institutions, industry partners
- 2 months at CERN + company visits + pre- and post- visit projects

Related events

- Friday 12th August 15:00-17:00 at IT Auditorium (B31-3-004)
  - Presentation of CERN openlab and some student projects
  - Tour PC farm, opencluster, tape storage silos…
  - Interested persons please email Laura (first-come-first-served)
- Wednesday 24th August: afternoon Grid tutorial (exact times/place tbd)
  - 4hrs, including hands-on experience
  - Interested persons please email Laura (first-come-first-served)

Related events

- CERN School of Computing
  - Helsinki 2006, open for applications in February

  NB: opening for 1 student to support CSC 2005 in St Malo 2-17 September.

  If interested please contact **Jackie.Franco-Turner@cern.ch**
- International Summer School on Grid Computing
  - Vico Equense 2007