CERN

openlab for DataGrid applications

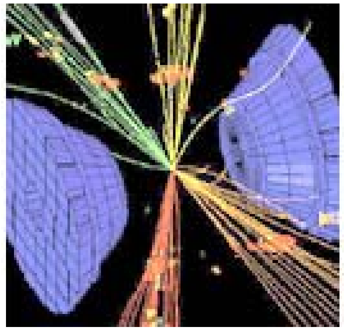
Presentation of the Grid@CERN

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.



LHC data

• 40 million collisions per second

• After filtering, 100 collisions of interest per second

• > 1 Megabyte of data digitised per collision
recording rate > 1 Gigabyte/sec

• 10^{10} collisions recorded each year
stored data > 10 Petabytes/year of data

1 Megabyte (1MB)
A digital photo

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

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

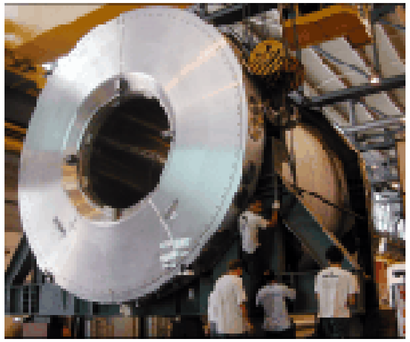
1 Petabyte (1PB)
= 1000TB
Annual production of one LHC experiment

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

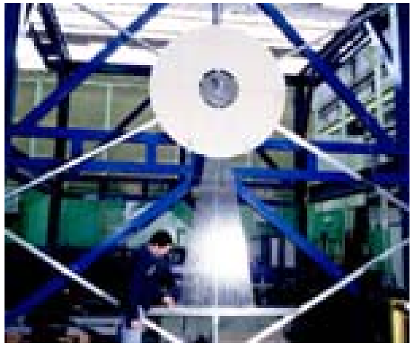
CMS

LHCb

ATLAS



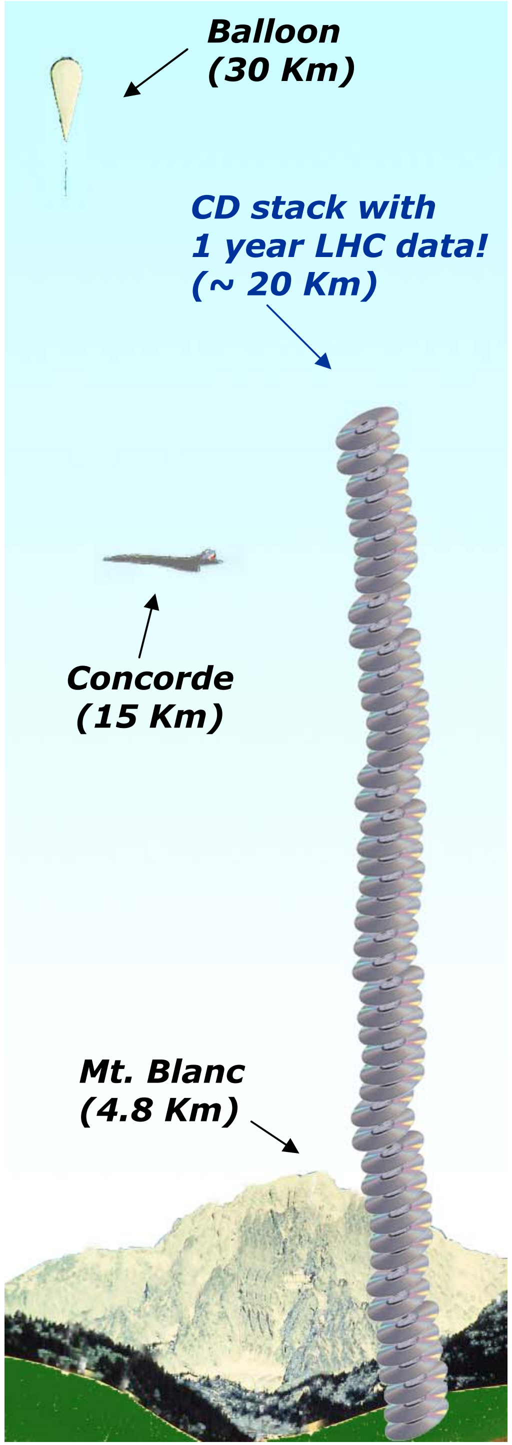
ALICE



LHC data

LHC data correspond to about 20 million CDs each year

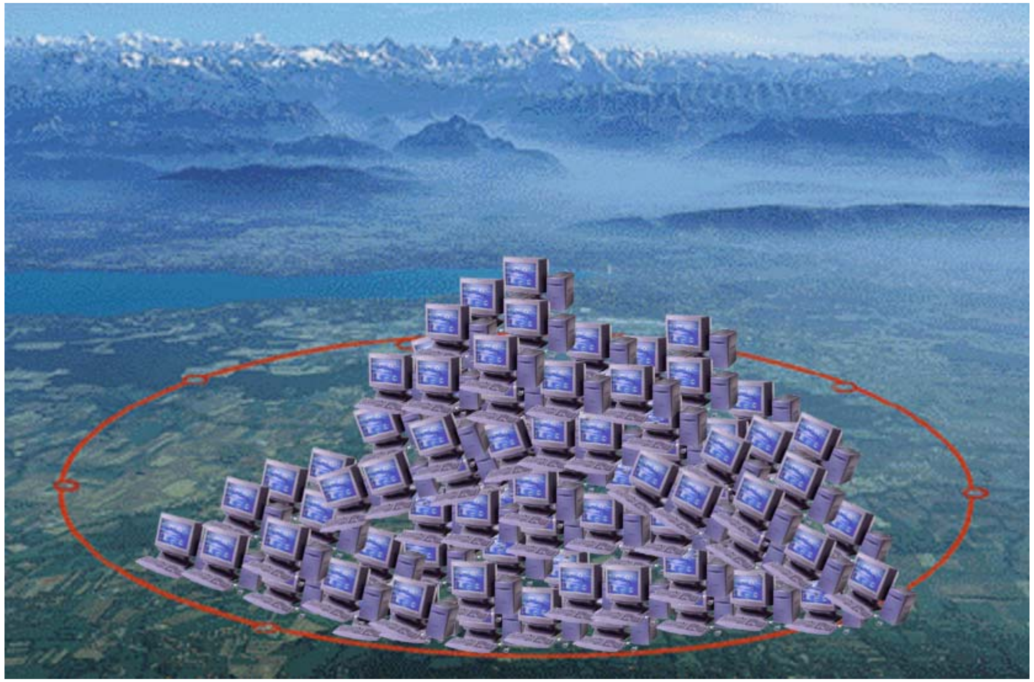
Where will the experiments store all of these data?



LHC processing

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

Where will the experiments find such a computing power?



Computing at CERN

- High-throughput computing based on reliable “commodity” technology
- More than 1000 dual processor PCs
- More than 1 Petabyte of data on disk and tapes

Nowhere near enough!

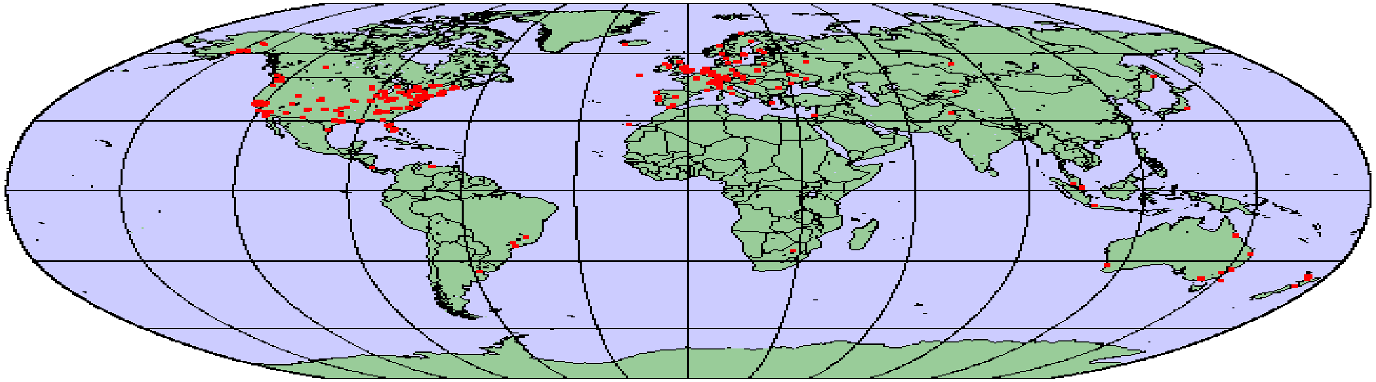
Computing for LHC

Europe:
267 institutes
4603 users

Elsewhere:
208 institutes
1632 users

• Problem: even with Computer Centre upgrade, CERN can provide only a fraction of the necessary resources

• **Solution:** Computing centers, which were isolated in the past, will be connected, **uniting the computing resources of particle physicists worldwide**



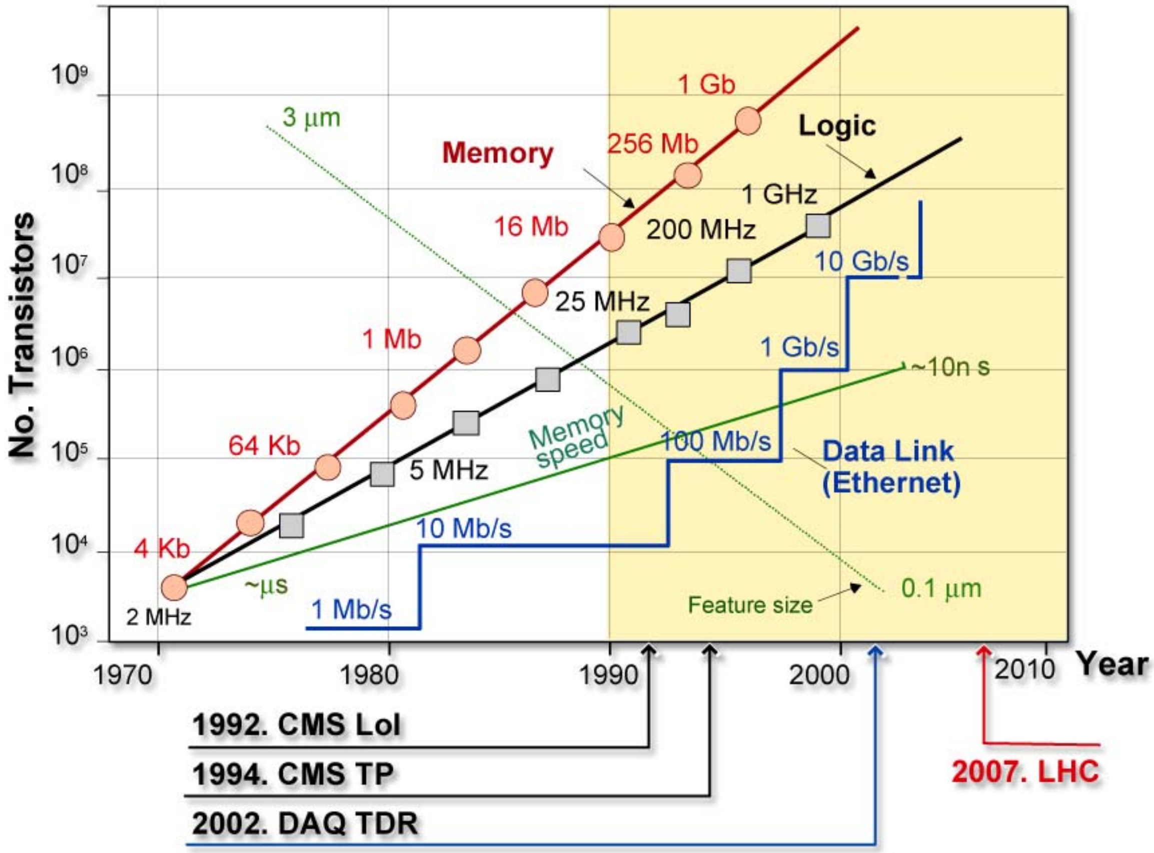
What is the Grid?

- 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?

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



What is the Grid?

From an industry point of view:

- The Grid is, for the time being, many Grids
- The Grid is not just cycle scavenging (SETI@home)
- Grids are a framework for Resource Virtualisation
- Grids are a label for Wide-Area Distributed Computing
- Grids will rely on new models of Inter-Organisational Security
- Grids will be used by Virtual Organisations
- Grids are a platform for Resource Discovery

Grid Services will be based on Web Services

Grid development has been driven by the academic community

Industrial variants are on-demand computing, Utility computing

What are the challenges for CERN?

Must share data between >7000 scientists with multiple interests

Must link major computer centres, not just PCs

Must ensure all data accessible anywhere, anytime

Must grow rapidly, yet remain reliable for more than a decade

Must cope with different management policies of different centres

Must ensure data security: more is at stake than just money!

Must be up and running by 2007

Grid projects in the world

- NASA Information Power Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF TeraGrid
- DOE ASCI Grid
- DOE Earth Systems Grid
- DARPA CoABS Grid
- NEESGrid
- DOH BIRN
- NSF iVDGL

- UK e-Science Grid
- Netherlands – VLAM, PolderGrid
- Germany – UNICORE, Grid proposal
- France – Grid funding approved
- Italy – INFN Grid
- Eire – Grid proposals
- Switzerland - Network/Grid proposal
- Hungary – DemoGrid, Grid proposal
- Norway, Sweden - NorduGrid

- DataGrid (CERN, ...)
- EuroGrid (Unicore)
- DataTag (CERN,…)
- Astrophysical Virtual Observatory
- GRIP (Globus/Unicore)
- GRIA (Industrial applications)
- GridLab (Cactus Toolkit)
- CrossGrid (Infrastructure Components)
- EGSO (Solar Physics)

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)

- EC funded projects led by CERN:
  Enabling Grids for E-Science in Europe (EGEE)
  +others

- Industry funded projects:
  CERN openlab for DataGrid applications

LCG

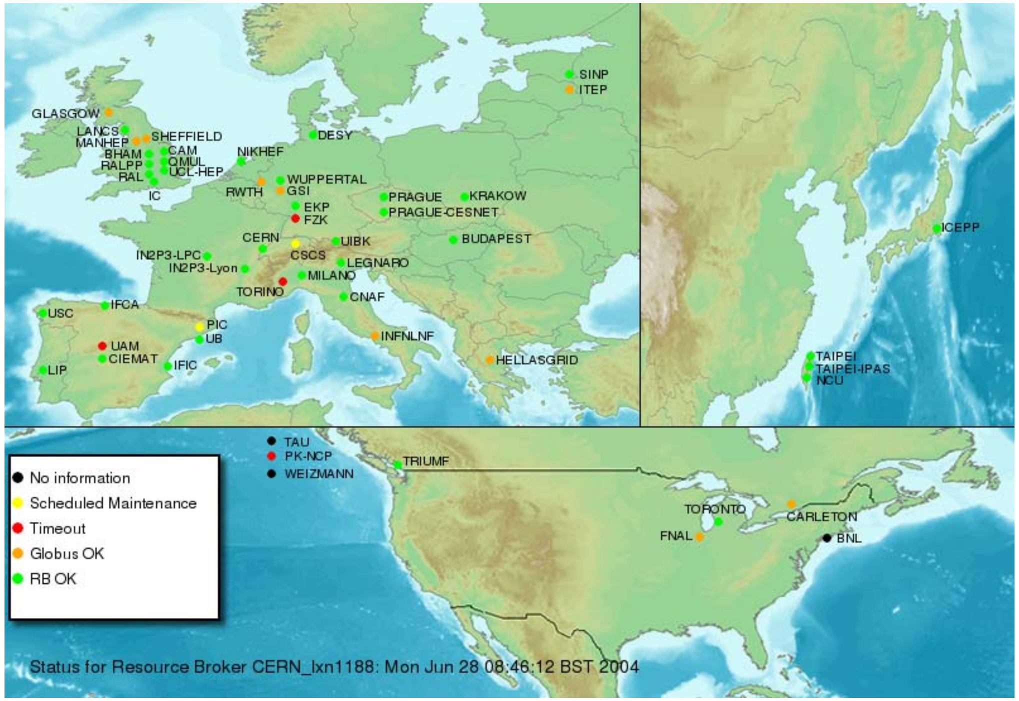
egee
Enabling Grids for
E-science in Europe

CERN
openlab for DataGrid applications

LHC Computing Grid (LCG)

Timeline:

- 2002: start project
- 2003: service opened (LCG-1 started in September with 12 sites)
- 2004 LCG-2 released
- 2002 - 2005: deploy the LCG environment
- 2006 – 2008: build and operate the LCG service



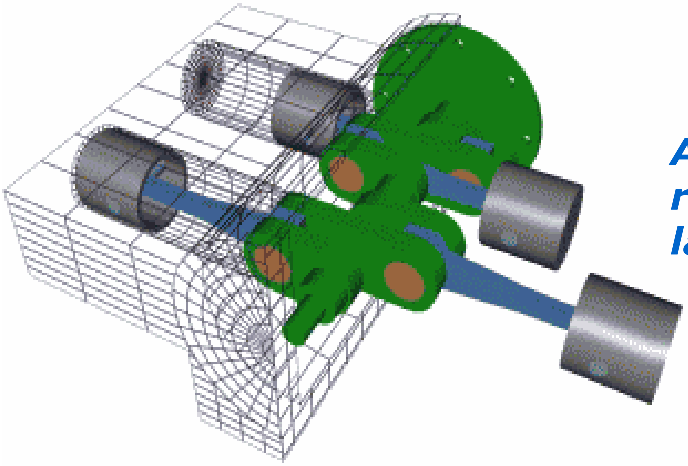
http://goc.grid-support.ac.uk/gppmonWorld/gppmon_maps/lcg2.html

- **As of July: 71 sites in 22 countries, 6400 CPU**
  - 49 Europe, 2 US, 5 Canada, 6 Asia, 2 HP
  - Coming: New Zealand, China, Korea
  - Industry: HP (Brazil, Singapore)

The EGEE Vision

Access to a production quality GRID will change the way science and much else is done in Europe

An international network of scientists will be able to model a new flood of the Danube in real time, using meteorological and geological data from several centers across Europe.



A team of engineering students will be able to run the latest 3D rendering programs from their laptops using the Grid.

A geneticist at a conference, inspired by a talk she hears, will be able to launch a complex biomolecular simulation from her mobile phone.

CERN

openlab for DataGrid applications

sponsored by

IBM

intel

hp invent

ORACLE

ENTERASYS NETWORKS

CERN openlab student program 2004

- **13 Students** from CH, DK, FI, NO, PO, UK, US (BSc-MSc-PhD levels)
- **3 Projects** involving Grids (openlab, Grid Cafe, openlogbook)
- **Co-funding** from participating institutions, HP and Nordic Grid Facility
- **2 months at CERN** + company visits + pre- and post- visit projects

Friday 6th 14:00-16:00

Presentation of CERN openlab and student programme activities, tour of Computer Centre

Please email Francois.Grey@cern.ch (limited to 40 persons)

If you want to know more, visit...

GridCafé

www.gridcafe.org