The LHC Computing Grid

Slides mostly by: Dr Ian Bird
LCG Project Leader
18 March 2008
Some precursors...

• Computing for HEP means data handling
  – Fixed-target experiments are also at the forefront

• COMPASS
  – >300 TB per year starting 2001 (still running)
  – Is used to investigating new computing technologies
    • One of the first reconstruction programmes entirely written using C++ and modern techniques
    • Raw data recorded at 35-70 MB/s (CMS expect ~200 MB/s)
    • Test of very-large database technologies
    • First user of CASTOR (transparent access of tape data)
Clusters of Inexpensive Processors

Requirements driven
- We started this phase with a simple architecture that enables sharing of storage across CPU servers, that proved stable and has survived from RISC to Quad-core
- Parallel, high throughput
- Sustained price/perf improvement ~60% /yr

- Apollo DN10.000s
  1989  20 MIPS/proc
- 1990 - SUN, SGI, IBM, HP, DEC, .... each with its own flavour of Unix
- 1996 – the first PC service
- 1998 – COMPASS Computing farm
- 2008 – dual quad core systems
  → 50K MIPS/chip → ~20k cores available == ~20 MSI2K

5 orders of magnitude in 18 years
The LHC Data Challenge

• The accelerator will be completed in 2008 and run for 10-15 years

• Experiments will produce about **15 Million Gigabytes** of data each year (about 20 million CDs!)

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

• Requires many cooperating computer centres, as CERN can only provide ~20% of the capacity

Ian Bird, CERN, IT Department
## Summary of Computing Resource Requirements

**All experiments - 2008**

*From LCG TDR - June 2005*

<table>
<thead>
<tr>
<th>Resource</th>
<th>CERN</th>
<th>All Tier-1s</th>
<th>All Tier-2s</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU (MSPECint2000s)</td>
<td>25</td>
<td>56</td>
<td>61</td>
<td>142</td>
</tr>
<tr>
<td>Disk (PetaBytes)</td>
<td>7</td>
<td>31</td>
<td>19</td>
<td>57</td>
</tr>
<tr>
<td>Tape (PetaBytes)</td>
<td>18</td>
<td>35</td>
<td></td>
<td>53</td>
</tr>
</tbody>
</table>

**Ian Bird, CERN, IT Department**
Solution: the Grid

- Use the Grid to unite computing resources of particle physics institutes around the world

The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations.

The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe.

Ian Bird, CERN, IT Department
How does the Grid work?

- It makes multiple computer centres look like a single system to the end-user.

- Advanced software, called middleware, automatically finds the data the scientist needs, and the computing power to analyse it.

- Middleware balances the load on different resources. It also handles security, accounting, monitoring and much more.

Ian Bird, CERN, IT Department
150 million sensors deliver data ...

... 40 million times per second
LHC Computing Grid project (LCG)

- More than 140 computing centres
- 12 large centres for primary data management: CERN (Tier-0) and eleven Tier-1s
- 38 federations of smaller Tier-2 centres
- 35 countries involved
LCG Service Hierarchy

Tier-0: the accelerator centre
- Data acquisition & initial processing
- Long-term data safekeeping
- Distribution of data → Tier-1 centres

Tier-1: “online” to the data acquisition process → high availability
- Managed Mass Storage – → grid-enabled data service
- Data-heavy analysis
- National, regional support

Tier-2: ~140 centres in ~35 countries
- Simulation
- End-user analysis – batch and interactive

Ian Bird, CERN, IT Department
WLCG Collaboration

- **The Collaboration**
  - 4 LHC experiments
  - ~140 computing centres
  - 12 large centres (Tier-0, Tier-1)
  - 38 federations of smaller “Tier-2” centres
  - ~35 countries

- **Memorandum of Understanding**
  - Agreed in October 2005, now being signed

- **Resources**
  - Focuses on the needs of the four LHC experiments
  - Commits resources
    - each October for the coming year
    - 5-year forward look
  - Agrees on standards and procedures

- **Relies on EGEE and OSG (and other regional efforts)**

Ian Bird, CERN, IT Department
Data Transfer

- Data distribution from CERN to Tier-1 sites
  - The target rate was achieved in 2006 under test conditions
  - Autumn 2007 & CCRC’08 under more realistic experiment testing, reaching & sustaining target rate with ATLAS and CMS active
Grid activity

- WLCG ran ~44 M jobs in 2007 – workload has continued to increase – now at ~165k jobs/day
- Distribution of work across Tier0/Tier1/Tier 2 really illustrates the importance of the grid system
  - Tier 2 contribution is around 50%; >85% is external to CERN
• LCG has been the driving force for the European multi-science Grid EGEE (Enabling Grids for E-sciencE)
• EGEE is now a global effort, and the largest Grid infrastructure worldwide
• Co-funded by the European Commission (Cost: ~130 M€ over 4 years, funded by EU ~70M€)
• EGEE already used for >100 applications, including...

Bio-informatics  Education, Training  Medical Imaging

Ian Bird, CERN, IT Department
The EGEE project

• **EGEE**
  - Started in April 2004, now in second phase with 91 partners in 32 countries

• **Objectives**
  - Large-scale, production-quality grid infrastructure for e-Science
  - Attracting new resources and users from industry as well as science
  - Maintain and further improve “gLite” Grid middleware
Registered Collaborating Projects

25 projects have registered as of September 2007: web page

Infrastructures
geographical or thematic coverage

Applications
improved services for academia, industry and the public

Support Actions
key complementary functions

Ian Bird, CERN, IT Department
Collaborating infrastructures

Countries connected to the EGEE Infrastructure
Countries connected to the Infrastructure via the US Open Science Grid
Countries in the EELA project
Countries in the EUMedGrid project
Countries in the BalticGrid project
Countries in the SEE-GRID project
Countries in the EUIndiaGrid project
Countries in the EUChinaGrid project
Countries in several regional projects
Enabling Grids for E-sciencE

EGEE-II INFSO-RI-031688

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
...

>250 sites
48 countries
>50,000 CPUs
>20 PetaBytes
>10,000 users
>150 VOs
>150,000 jobs/day

21:13:50 UTC
Diseases such as HIV/AIDS, SARS, Bird Flu etc. are a threat to public health due to worldwide exchanges and circulation of persons.

Grids open new perspectives to *in silico* drug discovery:
- Reduced cost and adding an accelerating factor in the search for new drugs.

International collaboration is required for:
- Early detection
- Epidemiological watch
- Prevention
- Search for new drugs
- Search for vaccines

*Avian influenza:*
- Bird casualties
http://wisdom.healthgrid.org/
For more information:

www.cern.ch/lcg  
www.eu-egee.org

www.eu-egi.org/

www.gridcafe.org

Thank you for your kind attention!