The LHC Computing Grid

Indian Grid Computing Initiative

Dr Ian Bird
LCG Project Leader
28th November 2008
CERN’s Large Hadron Collider

Large Hadron Collider
27 km circumference

Lake Geneva
The LHC Computing Challenge

• The scale and complexity of the data
  – Much more than in previous experiments
  – Large number of sensors * event complexity
    * high event rate
  → 15 Million GigaBytes (15 PetaBytes) of
    new data each year

• 7,000 researchers all actively analysing the
  data
  – Event complexity * large number of events *
    thousands of users
  → 100 k processors
  → 45 PetaBytes of disk storage to be
    added and managed each year

• The physics that is discovered will govern
  the
  way in which the data will be accessed
  → Flexible computing service able to
    respond rapidly to changing demands
What do we mean by a Computing Grid?

Interconnected Computing Centres

Virtual Organisations – map people to subsets of the resources of the grid

Software makes it all look like a single computing service for each Virtual Organisation

ATLAS Virtual Organisation

CMS Virtual Organisation

Les Robertson
LHC Computing Service Hierarchy

**Tier 0**
initial processing
long-term data archive

**Tier 1s**
data curation
data-intensive analysis
national, regional support

**Tier 2s**
end-user analysis
simulation
~130 centres
in 33 countries

The Tier-1 Centres
- Canada - Triumf (Vancouver)
- France - IN2P3 (Lyon)
- Germany - Forschungszentrum Karlsruhe
- Italy - CNAF (Bologna)
- Netherlands - NIKHEF/SARA (Amsterdam)

Nordic countries - distributed Tier-1
- Spain - PIC (Barcelona)
- Taipei - Academia Sinica
- UK - Rutherford Lab (Oxford)
- US - FermiLab (Illinois), Brookhaven (NY)
Building and Operating the Worldwide LHC Computing Grid

A collaboration between:

- The physicists and computing specialists from the LHC experiments
- The projects in Europe and the US that have been developing Grid middleware
- The regional and national computing centres that provide resources for LHC
- The research networks
Tier 0 at CERN: Acquisition, First pass processing

Storage & Distribution

1.25 GB/sec (ions)
Tier 0 – Tier 1 – Tier 2

Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (11 centres):
- Permanent storage
- Re-processing
- Analysis

Tier-2 (~130 centres):
- Simulation
- End-user analysis
Network Infrastructure status

Showing end-end links made by “stitching” circuits together from different contributors and backup links for resiliency.
Data Management

- Full experiment rate needed is 650 MB/s
- Desire capability to sustain twice that to allow for Tier 1 sites to shutdown and recover
- Have demonstrated far in excess of that
- All experiments exceeded required rates for extended periods, & simultaneously
- All Tier 1s achieved (or exceeded) their target acceptance rates
Usage Patterns

- The grid concept really works – all contributions – large & small contribute to the overall effort!
- Tier 2s consistently deliver ~50% of total
Recent grid activity

- In readiness testing, WLCG ran more than 10 million jobs/month.
- (1 job is ~ 8 hours use of a single processor)

These workloads are at the level anticipated for 2009 data.
WLCG depends on two major science grid infrastructures ....

**EGEE** - Enabling Grids for E-Science

**OSG** - US Open Science Grid

... as well as many national grid projects

Interoperability & interoperation is vital significant effort in building the procedures to support it
Collaborating e-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 EU-MedGrid project
Countries in the BalticGrid project
Countries in the SEE-GRID project
Countries in the EUIndiaGrid project
Countries in the EUCinaGrid project
Countries in several regional projects
Partnerships with India

- As part of WLCG there are 2 Tier 2 sites in India:
  - TIFR in Mumbai, and VECC in Kolkata
- 1 Gb/s network connection from CERN to Mumbai/TIFR permits data transfers at adequate rates
  - Acts as gateway for LHC partners in India
  - In October 2009 this will be replaced with a Dante (GEANT) link
- BARC team provide software development effort to LCG in grid monitoring
- India had close relationship with EGEE via EUIndiaGrid which had a budget of 1.2 million EUR for a period of 2 years starting from Oct 2006.
  - EUIndiaGrid composed of 5 European and 9 Indian partners
- Would hope that GARUDA continue to maintain close ties to EGEE and to WLCG, and in future to EGI (European Grid Infrastructure)
  - Work on interoperation is key to success
Long term infrastructures

- WLCG has built up the service to a significant level:
  - Resources – CPU and storage
  - Workloads
  - Data transfer rates
- It now relies on EGEE and OSG (and other national infrastructures) to provide operational, middleware, and user support
  - Today many European Tier 1 and Tier 2 sites use the EGEE operational and support infrastructure to provide services for WLCG
  - The EGEE procedures and mechanisms are also the basis of interoperability and interoperation with the US Open Science Grid
- Vital for WLCG now to understand the long term evolution of these infrastructures
  - Can they be relied on over the coming years?
European Grid Initiative

Goal:
• Long-term sustainability of grid infrastructures in Europe

Approach:
• Establishment of a new federated model bringing together NGIs to build the EGI Organisation

EGI Organisation:
• Coordination and operation of a common multi-national, multi-disciplinary Grid infrastructure
  – To enable and support international Grid-based collaboration
  – To provide support and added value to NGIs
  – To liaise with corresponding infrastructures outside Europe
European Grid Initiative timeline

Testbed phase | Initial production phase | Routine usage phase

EDG | EGEE | EGEE-II | EGEE-III


Sustainable e-Infrastructure

EGI - DS | EGI

09/07 | 11/09

Must be no gap in the support of the production grid
• EGI Design Study proposal approved by the European Commission (started 1st September’07)
• Supported by 35+ National Grid Initiatives (NGIs)
  http://web.eu-egi.eu/partners/ngi/
• 2 year project to prepare the setup and operation of a new organizational model for a sustainable pan-European grid infrastructure
• Draft EGI Blueprint produced:
  Functions Description http://www.eu-egi.eu/functions.pdf
Conclusions

- The Worldwide LHC Computing Grid has successfully demonstrated the implementation of a grid infrastructure
  - This is in daily use by the LHC experiments and is supporting significant workloads

- The need for a long term evolution of this infrastructure in Europe and world wide is clear for WLCG

- The EGI_DS project is outlining a proposal for such a long term European infrastructure

- Will this be in place to take over from EGEE and other grid infrastructure projects in mid-2010?
  - This is an extremely optimistic timescale and of real concern to WLCG