The Worldwide LHC Computing Grid: at a glance

What does the Worldwide LHC Computing Grid do?

- The Worldwide LHC Computing Grid combines the computing resources of more than 100,000 processors from 140 institutions in 33 countries, producing a massive distributed supercomputer that will provide more than 7000 physicists around the world with near real-time access to LHC data, and the power to process it.

How much data are we talking?

- The LHC will produce around 15 petabytes (15 million gigabytes) of data every year for ten to fifteen years. This is enough to fill 3,000,000 DVDs, every year. Viewing 3,000,000 DVDs would take around 500 years.
- If LHC data were to be burned to CD, a tower of CDs around 20 kilometers high would be created within a year. This is more than twice as high as Mt Everest.

What's been happening so far?

- The Worldwide LHC Computing Grid ran around 44 million "jobs" or programs in 2007 and more than 65 million jobs so far in 2008.
- In preparation for LHC start-up, the Worldwide LHC Computing Grid has been running between 250,000 and 500,000 jobs per day, simulating the running conditions of the LHC.

How much has it cost and who's paying?

- Individual institutions contributing to the Worldwide LHC Computing Grid are responsible for their own expenses, and computers that are part of the Worldwide LHC Computing Grid also contribute to other scientific grids.
- CERN has spent 100 million Euros on personnel and materials for the Worldwide LHC Computing Grid, up to and including 2008. Funding for CERN's contribution has come from national governments and the European Union. CERN's contribution is estimated to be around 20% of the total cost of the Worldwide LHC Computing Grid.
- CERN's budget for equipment for the Worldwide LHC Computing Grid in 2009 and following years in the medium term is 22.5 million Swiss Francs (14 million Euros) annually.

What benefits have we already seen?

- Lessons learned from the Worldwide LHC Computing Grid have driven further innovation in grids all over the world, changing the way science is done. Grids are being used in the fight against disease, climate change, air pollution and more. Any science that requires intensive simulation or calculation can benefit from grid computing.
- The computing centres providing resources for the Worldwide LHC Computing Grid are also active in other grids, in particular Enabling Grids for E-sciencE

(EGEE) in Europe and Open Science Grid (OSG) in the United States, but also several national and regional grid structures such as GridPP in the UK, INFN Grid in Italy and NorduGrid in the Nordic countries.

Who is involved in the Worldwide LHC Computing Grid?

- The Worldwide LHC Computing Grid comprises three “tiers” and 33 countries are formally involved:
    - Tier-0 is one site: the CERN Computing Centre. All data passes through this central hub but it provides less than 20% of the total compute capacity.
    - Tier-1 comprises eleven sites, located in Canada, France, Germany, Italy, the Netherlands, the Nordic countries, Spain, Taipei, the UK, and two sites in the USA.
    - Tier-2 comprises over 140 sites, grouped into 60 federations covering Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, India, Israel, Japan, Republic of Korea, the Netherlands, Norway, Pakistan, Poland, Portugal, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Taipei, Turkey, the U.K, Ukraine, and the U.S. Tier-2 sites. Together, these sites will provide around 50% of the capacity needed to process the LHC data.
    - Tier-2 sites then feed their data to PC clusters in physics institutes around the world, such that groups of scientists and individuals can analyze LHC data from their own desks.

What were the challenges of developing the Worldwide LHC Computing Grid?

- Managing the sheer volume of data that has to be moved reliably around the grid.
- Administering the storage space at each of the sites.
- Keeping track of the tens of millions of files generated by more than 7000 physicists as they analyse the data.
- Ensuring adequate network bandwidth: optical links between the major sites, but also good reliable links to the most remote locations.
- Guaranteeing security across a large number of independent sites while minimizing red-tape and ensuring easy access by authenticated users.
- Maintaining coherence of software versions installed in various locations
- Coping with heterogeneous hardware.
- Providing accounting mechanisms so that different groups have fair access, based on their needs and contributions to the infrastructure.

Open Science Grid and Enabling Grids for E-scienceE

- The Worldwide LHC Computing Grid has been the driving force behind the European multi-science grid Enabling Grids for E-SciencE (EGEE), which began in April 2004 and continues to grow in size and diversity of usage.
    - EGEE currently involves more than 240 institutions in 45 countries, supporting science in more than 20 disciplines, including bioinformatics, medical imaging, education, climate change, energy, agriculture and more.

- The Worldwide LHC Computing Grid has also been the driving force behind Open Science Grid (OSG), a grid computing infrastructure for large scale science, built and operated by a consortium of U.S. universities and national laboratories. OSG also continues to increase its size and breadth of reach.
    - The OSG Consortium was formed in 2004 and enables diverse communities of scientists to access a common grid infrastructure and shared resources.

Future challenges for the Worldwide LHC Computing Grid

As more data is gathered from collisions inside the LHC, there will be increasing demand upon the storage and processing capacity of the Worldwide LHC Computing Grid. These challenges will include:

- Increase of available resources in response to planned upgrades to the LHC accelerator, as well as to the increasing data requirements of the four LHC experiments.
- Smooth uptake of and conversion to new and evolving technologies: even during the planning and design of the Worldwide LHC Computing Grid, there were significant changes to available technologies, so we need to remain flexible and adaptable.
- Continuing to optimize the use of multi-core processors, where cores increase beyond two or four or eight processors to 16- or 32-core processors.
- Working with limitations in terms of the cooling and power requirements of large data centres, which is an ongoing issue shared by large data centres all over the world.
- Adaptation to longer term sustainable grid infrastructures.

More information

Head of WLCG: Ian Bird (ian.bird@cern.ch) +41 22 767 5888
EGEE: Bob Jones (bob.jones@cern.ch) +41 22 767 1482
Open Science Grid: Elizabeth Clements (lizzie@fnal.gov) +1 630 840 2326
GridTalk: Cristy Burne (cristy.burne@gridtalk-project.eu) +41 76 487 0486

Adapted from LCG Dissemination (http://LCG.web.cern.ch/LCG/dissemination.html)