

The LHC Grid Challenge

F. Malek

LPSC, Université Joseph Fourier Grenoble and CNRS/IN2P3, France.

Abstract

The LHC computing is a worldwide complex machinery relying on several grid technologies such as EGEE and OSG, tens of middlewares, hundred of computing centers and is operated and used by thousands of people, engineers and physicists. This paper describes how this human adventure began and to what extent it is usable today and ready for the start up of the LHC.

1 The Worldwide LHC Computing Grid

The primary reason for the decision to adopt a distributed computing approach to manage LHC data was money. In 1999, when work began on the design of the computing system for LHC data analysis, it rapidly became clear that the required computing capacity was far beyond the funding capacity available at CERN. On the other hand, most of the laboratories and universities collaborating on the LHC had access to national or regional computing facilities. The obvious question was: could these facilities be somehow integrated to provide a single LHC computing service? The rapid evolution of wide area networking, the increase capacity and bandwidth coupled with falling costs made it possible. From there, the path to the Worldwide LHC Computing Grid was set.

During the development of the Worldwide LHC Computing Grid, many additional benefits of a distributed system became apparent. Multiple copies of data can be kept in different sites, ensuring access for all scientists involved, independently of geographical location; It allows optimum use of spare capacity for multiple computer centers, making it more efficient; Having computer centers in multiple time zones eases round-the-clock monitoring and the availability of expert support. The cost of maintenance and upgrades is distributed, since individual institutes fund local computing resources and retain responsibility for these, while still contributing to the global effort; Independently managed resources have encouraged novel approaches to computing and analysis; The so-called "brain drain", where researchers are forced to leave their country to access resources, is reduced when resources are available from their desktop; It provides considerable flexibility in deciding how and where to provide future computing resources. And finally, it allows community to take advantage of new technologies that may appear and that offer improved usability, cost effectiveness or energy efficiency.

The WLCG technical Design Reports [1] were issued in June 2005, describing the grid infrastructure, the common tools to be developed and the computing models for each of the four LHC experiments. A Memorandum of Understanding [2] was agreed in October 2005 between the WLCG collaboration and the participating nations and funding agencies. This MoU guarantees the resources, the quality of services and looks 5-year forward for the resource planning. The quality of services includes a guarantee of the operations 24 hours a day and 7 days a week with intervention to services essential to the running of a center in a time laps of 4 hours. For any site, the target reliability and efficiency to reach is 98% .

The Worldwide LHC Computing Grid combines the computing resources of more than 100,000 processors from 150 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 means to process it.

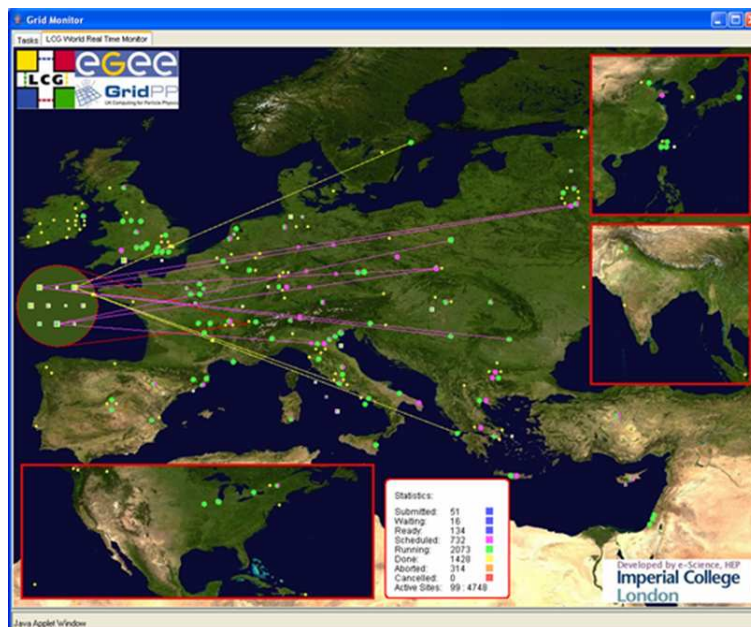


Figure 1: The real time GridPP monitoring showing the WLCG computing facilities processing jobs and transferring data.

Fig. 1 shows a snapshot of the WLCG real time monitoring where the worldwide computing centers are processing data and exchanging information. The computing centers providing resources for the Worldwide LHC Computing

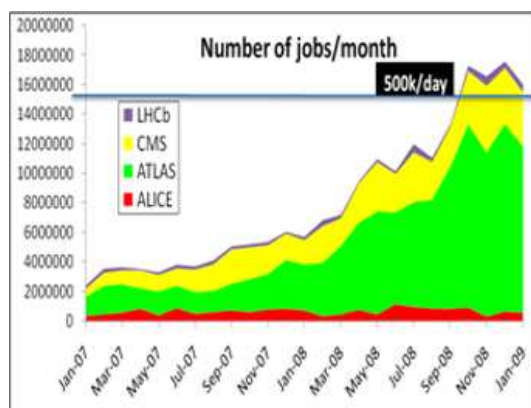


Figure 2: LHC experiments physics simulation jobs on WLCG infrastructure in 2009.

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.

WLCG ran around 44 million "jobs" or tasks in 2007 and more than 65 million jobs in 2008. In preparation for LHC start-up in 2009, WLCG has been running up to 500,000 jobs per day, simulating the running conditions of the LHC, see Fig. 2.

2 The WLCG distributed computing model

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.

The WLCG infrastructure is based on three "tiers" and 33 countries are formally involved:

- The Tier0 is a single site: the CERN Computing Center. All data passes through this central hub but it provides less than 20% of the total computing capacity.
- The Tier1s consist of eleven sites, located in Canada, France, Germany, Italy, the Netherlands, the Nordic countries, Spain, Taiwan, the UK, and two sites in the USA.
- The Tier2s consist of 140 sites, grouped into 60 federations covering countries from Australia to U.S., passing by Asia and Europe. Together, these sites will provide around 50% of the capacity needed to process the LHC data.
- The Tier2 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.

The CERN facility is linked to other major grid centers using 10 Gigabit per second optical wide area links, as well as to the general global education and research network infrastructure. The four LHC experiments computing models rely on the distributed computing facilities sketched on Fig. 3. More details can be found in the computing models described in the technical design reports for each of the experiments [3].

3 Computing within the LHC experiments

The Tier0 and the CAF (Cern Analysis Facility) have very much the same functionality for all experiments. Each LHC experiment is a Virtual Organization (VO) with a set of policies (certification, security, grid middleware, specific software, etc.). The Tier0 is the central hub where all the data passes through before they are distributed to the Tier1's for reprocessing. The CAF is mainly used for the calibrations, first streams analysis and local users analyses. The Tier1's have the principal role of reprocessing the data and the Tier2's deal with the simulation and the physics analysis. As an example, the CMS computing model is sketched on Fig. 4.

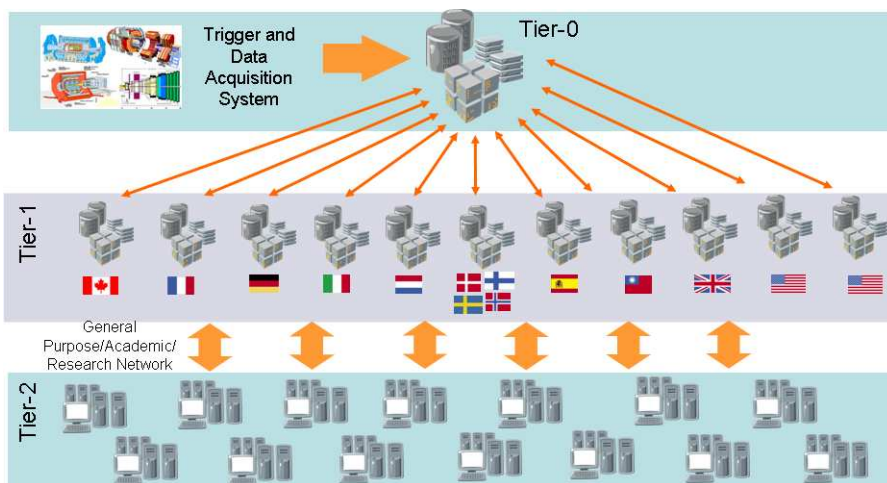


Figure 3: The LHC distributed computing model

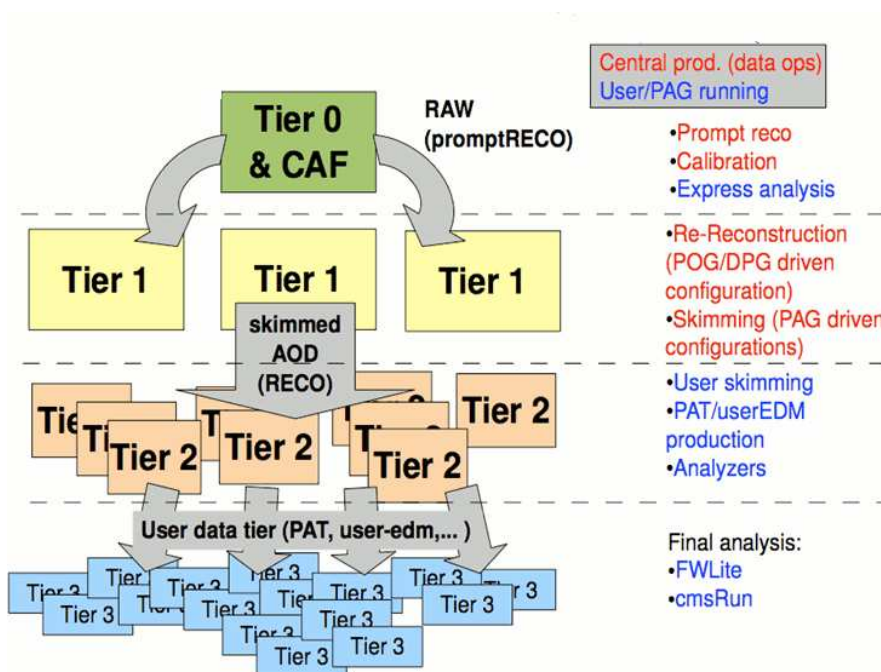


Figure 4: CMS Data flow structure.

However, each of the four LHC experiment rely on its own computing model which differ somehow from the others, especially in the services (data management, data distribution, Tiers relations, etc.). For example, CMS analysis jobs in Tier2s can get data from any Tier1 while for ATLAS, the analysis jobs in Tier2s can get data only from the Tier1 within the same cloud. The clouds in ATLAS are defined as federation of Tiers within the same country or a defined region. Fig. 5 shows the flowchart of ATLAS clouds where eight of them are operated under EGEE, the European grid infrastructure. The IN2P3 cloud, for example, is a region which comprises the French sites and the associated ones: Japan, China and Romania. Also, CMS analysis coordination is per Tier2 while in ATLAS the coordination is done per physics group and/or cloud.

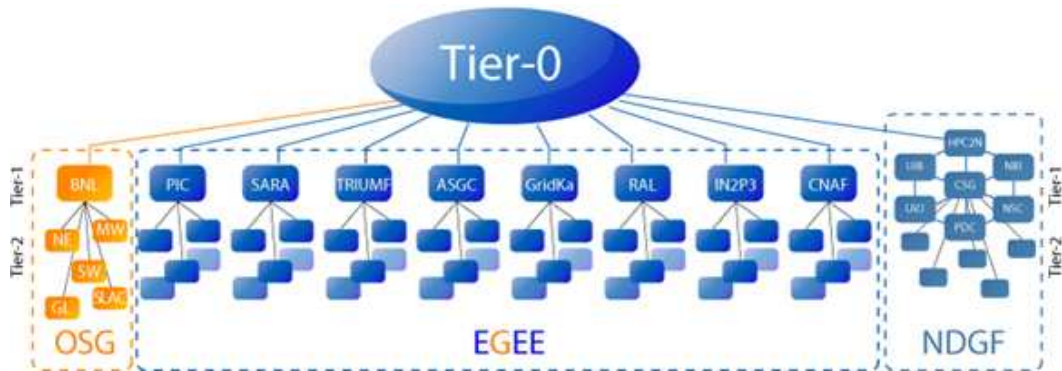


Figure 5: The ATLAS experiment Clouds infrastructure.

The four LHC experiment rely on some common software like physics generation (Genser, HepMC, ...) or detector simulation (Geant4, Fluka, Garfield, ...). They also have some common core libraries and services, LCG applications packages like ROOT (Physics analysis), COOL, POOL (condition data bases) etc. But, each of the four LHC experiments has its own software infrastructure: Aliroot, Gaudi, CMSSW, Athena. The analysis tools and packages (PANDA, GANGA, ALIEN, CRAB, ...) are also quite different from one experiment to the other. All these tools should run on some 20 different platforms and operating systems. Last but not least, since the precision of calculation is an asset, the whole infrastructure must run on both 32 bits and 64 bits machines.

4 Conclusion and Future challenges

The WLCG applications support and service meet with satisfaction the requirements and the baseline services are in production. There is, as expected, a continuously increasing capacity and workload and the general site reliability is improving. It has already reached a quality level of 95% for most of the Tier1s, the target value being 98% as fixed by the WLCG MoU.

The data and storage services remain the weak points of WLCG. Nevertheless, sites and experiments are working well together to tackle the problems. Many data transfers, LCG services, experiment challenges are now involving most sites and the most realistic scenarii are taken into account.

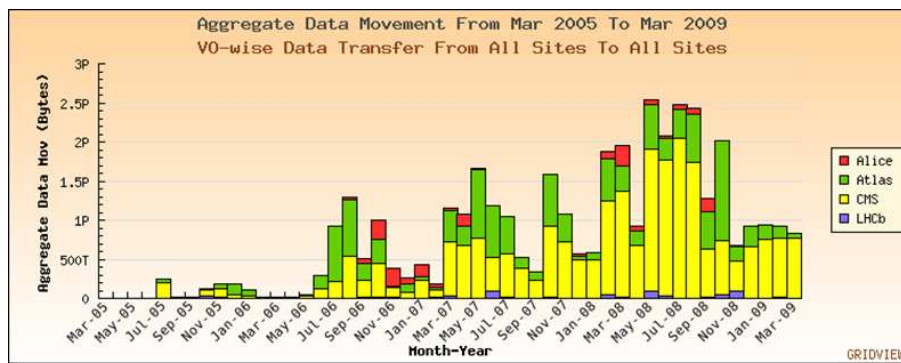


Figure 6: VO-wise data transfer from all sites to all sites from march 2005 to March 2009.

Combined computing readiness challenges in Feb-May 2008 (called CCRC08) and in May-June 2009 (called STEP 09) were essential to provide experience for site operations and storage systems. The infrastructure as a whole was stressed simultaneously by all four experiments. The plot on Fig. 6 show the results of the data transfer exercises done by the four LHC experiments since 2005. A data movement of around 2.5 petaBytes was reached in 2008.

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 the 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.

Besides this, conversion to new and evolving technologies is important. 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 is essential. And finally, working with limitations in terms of the cooling and power requirements of large data centers, is an ongoing issue shared by large data centers all over the world.

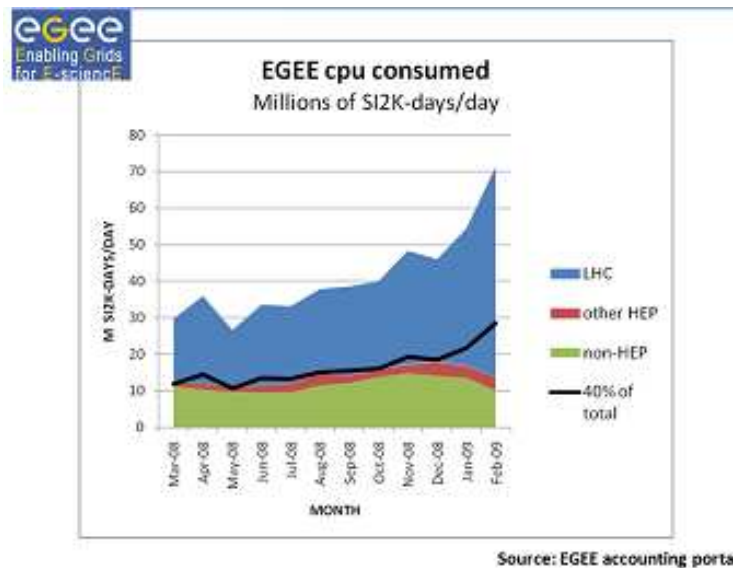


Figure 7: Usage of the EGEE infrastructure by all scientific communities from 2008 to early 2009.

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. Fig. 7 shows the usage of the worldwide grid infrastructure EGEE. One can observe that since a year, 40% of the CPU resources are used by non HEP communities which indicates that grid infrastructures are becoming a useful tool for all scientific fields.

References

- [1] **CERN-LHCC-2005-024**, see also: <http://lcg.web.cern.ch/LCG/tdr/>.
- [2] **CERN-C-RRB-2005-01**.
- [3] **ALICE-TDR-012: CERN-LHCC-2005-018;**
ATLAS-TDR-017: CERN-LHCC-2005-022;
CMS-TDR-007: CERN-LHCC-2005-023;
LHCb-TDR-11: CERN-LHCC-2005-019.