# WORLDWIDE LHC COMPUTING GRID

## CLUSTERS, GRIDS, CLOUDS

José Mauricio Perdomo,
PERCCOM Cohort 5
ITMO University
June 7, 2018

## Introduction

Approximately 600 million times per second, particles collide within the Large Hadron Collider (LHC). Each collision generates particles that often decay in complex ways into even more particles. Electronic circuits record the passage of each particle through a detector as a series of electronic signals and send the data to the CERN Data Centre (DC) for digital reconstruction. The digitized summary is recorded as a "collision event". Physicists must sift through the 30 petabytes or so of data produced annually to determine if the collisions have thrown up any interesting physics.

CERN does not have the computing or financial resources to crunch all of the data on site, so in 2002 it turned to grid computing to share the burden with computer centres around the world. The Worldwide LHC Computing Grid (WLCG) – a distributed computing infrastructure arranged in tiers – gives a community of over 8000 physicists near real-time access to LHC data. The Grid builds on the technology of the World Wide Web, which was invented at CERN in 1989.

In 1999, when work began on the design of a computing system for LHC data analysis, it rapidly became clear that the required computing power was far beyond the funding capacity available to CERN. On the other hand, most of the laboratories and universities collaborating on the LHC had access to national or regional computing facilities.

## About LHC and WLCG

These were integrated into a single LHC computing service – the Grid – in 2002. It now links thousands of computers and storage systems in over 170 centres across 41 countries. These computer centres are arranged in "Tiers", and together serve a community of over 8000 physicists with near real-time access to LHC data. The Grid gives users the power to process, analyse and in some cases to store LHC data.

The WLCG is the world's largest computing grid. It is based on two main grids – the European Grid Infrastructure in Europe, and Open Science Grid in the US – but has many associated regional and national grids (such as TWGrid in Taiwan and EU-IndiaGrid, which supports grid infrastructures across Europe and Asia).

This grid-based infrastructure is the most effective solution to the data-analysis challenge of the LHC, offering many advantages over a centralized system. Multiple copies of data can be kept at different sites, ensuring access for all scientists independent of geographical location; there is no single point of failure; computer centres in multiple time zones ease round-the-clock monitoring and the availability of expert support; and resources can be distributed across the world, for funding and sociological reasons.

In short, the Worldwide LHC Computing Grid (WLCG) project is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures.

The mission of the WLCG project is to provide global computing resources to store, distribute and analyse the ~50 Petabytes of data expected in 2018, generated by the Large Hadron Collider (LHC) at CERN on the Franco-Swiss border.

### Using the Grid

With more than 8000 LHC physicists across the four main experiments – ALICE, ATLAS, CMS and LHCb – actively accessing and analysing data in near real-time, the computing system designed to handle the data has to be very flexible.

WLCG provides seamless access to computing resources which include data storage capacity, processing power, sensors, visualization tools and more. Users make job requests from one of the many entry points into the system. A job request can be almost anything – storage, processing capacity, or availability of analysis software, for example. The computing Grid establishes the identity of the user, checks their credentials, and searches for available sites that can provide the resources requested. Users do not have to worry about where the computing resources are coming from – they can tap into the Grid's computing power and access storage on demand.

Tier 0 of the Grid runs around one million jobs per day. Peak data-transfer rates of 10 gigabytes per second – the equivalent of two full DVDs of data per second – are not unusual.

## Hierarchy and Middleware

As mentioned before, WLCG is subdivided in TIERs. WLCG is a widely distributed computing system that spans all the world geographically speaking. Table 1.1 shows the different levels in which the data is processed, stored or transferred.

| Data Journey |
|---|
| **1.** Data is generated by 4 giant experiments of the LHC ring*: LHCb, ATLAS, ALICE,CMS . Conjointly they produce 100 GB/s of data. |
| **2.** Data is aggregated to the Tier 0 CERN Data Centre. Permanent Long-Term Storage ( Tape Storage ) . |
| **3.** Data is distributed to 11 Tier 1 Data Centres ( More permanent Storage, re-processing, and analysis ) |
| **4.** Data is received in 140 Tier 2 Data Centres disseminated around the world (Simulation, and end-user analysis). |
| **5.** Data is sent to TIER 3 universities around the world. |
| 6. Finally, Tier 4 is comprised by end-user PCs**.** |

**Table 1.1** Data Journey throughout the WLCG

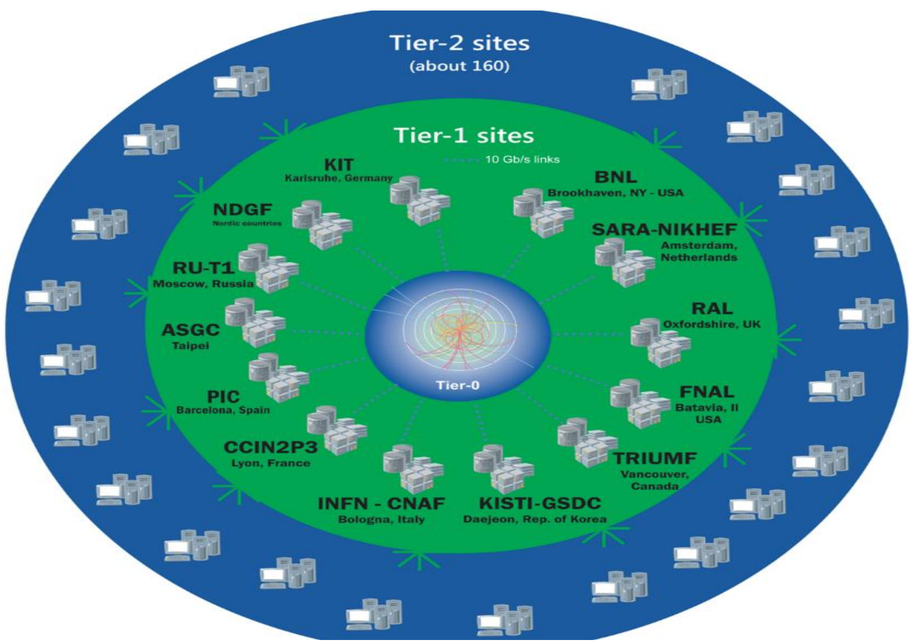



**Figure 1.1** Partial Hierachy of WLCG

## Middleware used in the Grids that comprise WLCG

Computing resources are allocated to each experiment, so that scientists have the right amount of computing power they need for conducting their research.

ARC

The Advanced Resource Connector (ARC) middleware integrates computing resources (usually, computing clusters managed by a batch system or standalone workstations) and storage facilities, making them available via a secure common Grid layer.

Virtual Data Toolkit

The Virtual Data Toolkit (VDT) is an ensemble of distributed computing software that can be easily installed and configured. The goal is to make it as easy as possible for users to deploy, maintain and use distributed computing software. The VDT is a product of the Open Science Grid (OSG), which uses the VDT as its software distribution.

Globus Toolkit

The Globus Toolkit is a fundamental enabling technology for building grids that allow distributed computing power, storage resources, scientific instruments, and other tools to be shared securely across corporate, institutional, and geographic boundaries.about Globus Toolkit

EMI - European Middleware Initiative

EMI is a collaboration of the three major middleware providers in Europe, ARC, gLite and UNICORE, and other consortia. It aims to deliver a consolidated set of middleware components for deployment in EGI, PRACE and other DCIs; extend the interoperability between grids and other computing infrastructures; strengthen the reliability of the services; and establish a sustainable model to maintain and evolve the middleware, fulfilling the requirements of the user communities.

**LHC Home**

LHC@Home, as show in Figure 1.2 is a volunteer computing platform where you donate idle time on your computer to help physicists compare theory with experiment, in the search for new fundamental particles and answers to questions about the Universe**.**

**Figure 1.2** Large Hadron Collider Home

All of the LHC home projects run using BOINC (Berkeley Open Infrastructure for Network Computing) - a long-established platform which is used by the vast majority of volunteer-computing projects around the world.

The main projects you can participate on are:

| | |
|---|---|
| ATLAS@HOME | Known" physics and "new" phenomena - want to create alternative models of the universe? |
| BEAUTY | "b" is for "Beauty" - the gorgeous little particle in antimatter physics. |
| CMS@HOME | CMS is on the lookout for completely new, unpredicted phenomena. |
| SixTrack | Help CERN accelerator engineers to run intensive simulations to check the stability of the twin proton beams circulating in the LHC machine. |
| Test4Theory | Simulate high-energy particle collisions and help tune the theory to the experimental results. |

An important remark is that end-users that connect to the grid are involved in simulation tasks rather than computing with direct data generated from LHC. The WLCG as such is interfaced by the aforementioned projects.

Moreover, these projects require various different types of middleware like Virtualbox, CERNVM or Vboxwrapper to compute more complex tasks. However, all of the LHC home projects run using BOINC ( Berkeley Open Infrastructure for Network Computing ) - a long-established platform which is used by the vast majority of volunteer-computing projects around the world.

In essence, BOINC is a software that can use the unused CPU and GPU cycles on a computer to do scientific computing. BOINC's architecture may be found in Figure 1.3. As shown, the figure depicts the way different hosts are leveraged to perform the different jobs to achieve a common goal.

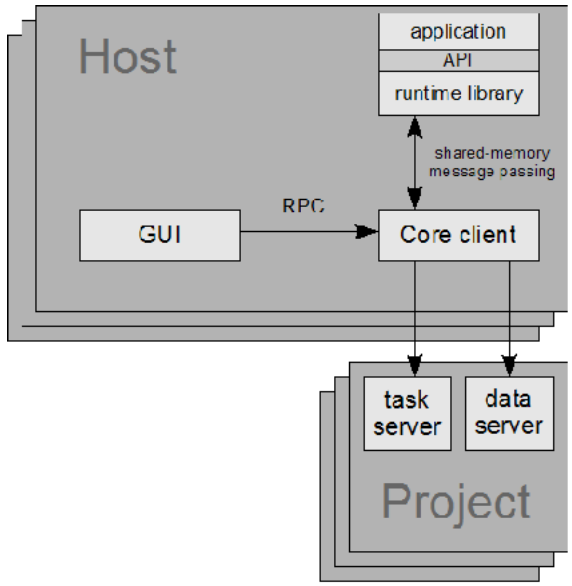
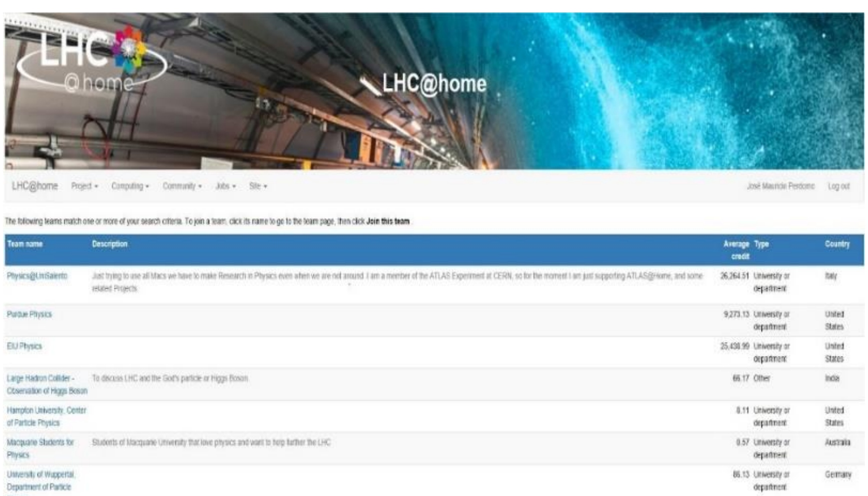


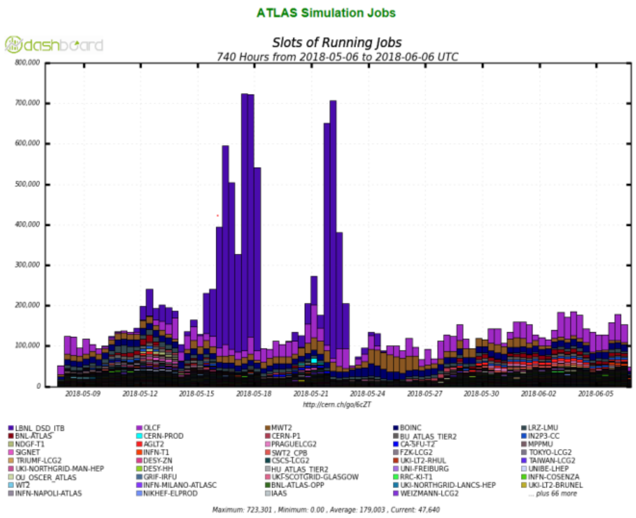

**Figure 1.3** BOINC Architecture

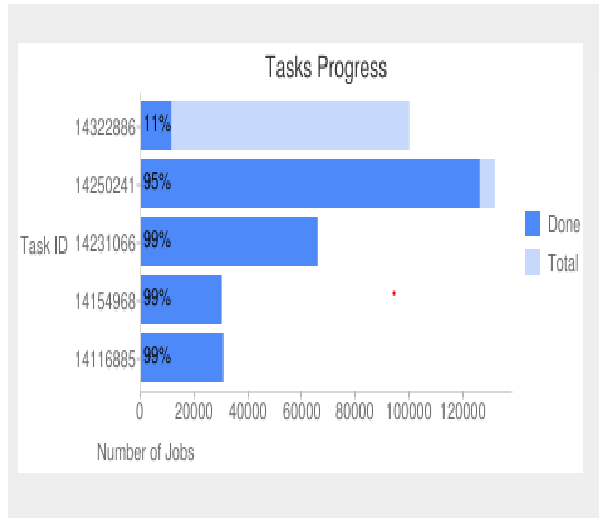
In addition, LHC Home permits you join different LHC computing teams throughout the world. It has a gamification system embedded in the page that allows you to accumulate points for you and your team. Generally, these teams are comprised by university students, or science enthusiasts that want to provide benefit, although miniscule, to scientific research.

More information on, https://bit.ly/2JlzDUQ



Furthermore, the webpage displays real-time graphs of the computed jobs of a particular project and those that are remaining to compute as well. Figure 1.5 and Figure 1.6 show the ATLAS Simulation computed jobs divided in Grids and remaining jobs that need to be computed respectively.




## Conclusion

In summary, WLCG is a good example of the immense benefits a computing grid may achieve. Grids facilitate collaboration between members of supported distributed communities. Grids allow distributed resources to be shared uniformly and securely for common goals. Grids may have complex infrastructures. Grids are useful for many scientific disciplines and projects. The Worldwide LHC Computing Grid is vital for the success of the LHC experiments!