Worldwide LHC Computing Grid (WLCG).

Course: Computing clusters, computing grids, computing clouds.

ITMO University.

June 2017.

**Geraldine Villers.**

Table of content

[1. Introduction 3](#_Toc484504339)

[1.1 LHC Computing Grid 3](#_Toc484504340)

[1.2 WLCG cloud 4](#_Toc484504341)

[2. Component layers of WLCG 4](#_Toc484504342)

[3. Architecture 5](#_Toc484504343)

[4. Services. 6](#_Toc484504344)

[4.1 Storage Elements Services 6](#_Toc484504345)

[4.2 File Transfer Services. 6](#_Toc484504346)

[4.3 Computer Resources Services. 6](#_Toc484504347)

[4.4 Workload Management 7](#_Toc484504348)

[4.5 Database Services 7](#_Toc484504349)

[3.6 Grid Catalogue Services 7](#_Toc484504350)

[5. Cost 7](#_Toc484504351)

[6. Security 7](#_Toc484504352)

[7. Benefits and challenges of computing Grid. 8](#_Toc484504353)

[8. References. 8](#_Toc484504354)

# Introduction

## 1.1 LHC Computing Grid

The LHC Computing Grid project was proposed and approved in order to develop, build and maintain a distributed computing infrastructure for the storage and analysis of data from the four LHC experiments: ALICE, which consist of 33 countries, 116 institutes and over 1000 members. ATLAS, consists of 38 countries, 174 institutes and 3000 members. CMS, with 39 countries, 169 institutions and 3170 members. LHCb, 15 countries, 54 institutes and 754 members.

The LCG project is associated with the EGEE (Enabling Grids for E-Science) project, partially funded by the European Union. It interconnects a large number of sites in 34 countries around the world, integrating several national and regional Grid initiatives in Europe. Moreover, it is associated with the Open Science Grid (OSG) in the US and the Nordic Data Grid Facility (NDGF).



Figure 1. WLHC implementation

The Large Hadron Collider (LHC) located at CERN on the Franco-Swiss border 50 - 175m underground, it consists of 27 km circumference tunnel and the 4 big experiments mentioned before. The LHC produces 15 Petabytes (15 million Gigabytes) approximately of data annually and over 10,000 physicists can access and analyze it. Additionally, all the data needs to be available over 15-year lifetime of the LHC.

Particles collide at high energies inside CERN's detectors, creating new particles that decay in complex ways as they move through layers of subdetectors. The subdetectors register each particle's passage and microprocessors convert the particles' paths and energies into electrical signals, combining the information to create a digital summary of the "collision event".

The data flow from all four experiments is anticipated to be about 25 GB/s (gigabyte per second)

* ALICE: 4 GB/s (Pb-Pb running)
* ATLAS: 800 MB/s – 1 GB/s
* CMS: 600 MB/s
* LHCb: 750 MB/s

In the first stage of the selection, the number of events is filtered from the 600 million or so per second picked up by detectors to 100,000 per second sent for digital reconstruction. In a second stage, more specialized algorithms further process the data, leaving only 100 or 200 events of interest per second. This raw data is recorded onto servers at the CERN Data Centre at a rate around 1.5 CDs per second (approximately 1050 megabytes per second). The particles are accelerated within the LHC ring travelling at 99.9999991% of the speed of light completing 11,245 turns every second. The results are 600 million of collisions every second and only one in a million collision is of interest. By the use of fast electronic preselection passes 1 out of 10,000 events and stores them on computer memory. After storage, the data needs to be computed to be able to find the Higgs boson or reproduce the conditions after the collisions and see what particles have been created, and so on.

In order to analyze the data and compare it with simulations 100,000 CPUs are needed. This amount of resources is outrageous and leads to the need of a distributed model for data storage and analysis of the big data generated (computing Grid). Also, this project uses more than 340,000 cores.

## 1.2 WLCG cloud

As it was mentioned, by using highly-specialized algorithms, the data is transferred to CERN’s data center for processing. The data are saved to tape storage and then distributed to the more than 150 sites worldwide that comprise the Worldwide LHC (Large Hadron Collider) Computing Grid (WLCG) for analysis.

CERN began to investigate various virtualization and cloud platforms. CERN’s IT team decided it would build a private cloud that would need to integrate well with a very heterogeneous environment. CERN’s service consolidation environment is based on the Microsoft Service Center Virtual Machine Manager and the Hyper-V hypervisor. The IT team also built a cloud test bed based on OpenNebula and KVM (Kernel-based Virtual Machine) hypervisor.

 In the end, CERN selected OpenStack. As a cloud platform, OpenStack controls and automates pools of compute, storage, and networking resources to turn standard hardware into a powerful cloud computing environment.

# Component layers of WLCG

The four main component layers of WLCG are physics software, middleware, hardware and networking.

On the physics software, the computer centres are made up of multi-petabyte storage systems and computing clusters with thousands of nodes connected by high-speed networks. It includes programs such as ROOT, a set of object-oriented core libraries used by all the LHC experiments; POOL, a framework, that provides storage for event data; and other software for modelling the generation, propagation and interactions of elementary particles.

The middleware is called that way because it sits between the operating systems of the computers and the physics-applications software that can solves a user's particular problem. The most important middleware stacks in the WLCG are the European Middleware Initiative, which combines key middleware providers ARC, gLite, UNICORE and dCache; Globus Toolkit; OMII from the Open Middleware Infrastructure Institute; and Virtual Data Toolkit.

In terms of hardware, each Grid centre manages a large collection of computers and storage systems. It also has the management systems such as Quattor, developed at CERN that automate the installation and upgrading of software. They ensure that the correct software is installed from the operating system all the way to the experiment-specific physics libraries, and make this information available to the overall Grid scheduling system, which decides which centres are available to run a particular job.

For networking, the Grid file-transfer service developed by the Enabling Grids for E-science projects, manages the exchange of information between WLCG centres. The file-transfer service includes authentication and confidentiality features, reliability and fault tolerance, and third-party and partial-file transfer.

Optical-fibre links working at 10 gigabits per second connect CERN to each of the Tier 1 centres around the world. This dedicated high-bandwidth network is called the LHC Optical Private Network (LHCOPN).

# Architecture

The LCG Project implements a Grid to support the computing models of the experiments using a distributed four-tiered model:



Figure 2. LCG Architecture

The original raw data emerging from the data acquisition systems will be recorded at the Tier-0 centre at CERN. The first-pass reconstruction will take place at the Tier-0, where a copy of the reconstructed data will be stored. The Tier-0 will distribute a second copy of the raw data across the Tier-1 centres associated with the experiment. Additional copies of the reconstructed data will also be distributed across the Tier-1 centres according to the policies of each experiment.

On Tier-1 centres varies according to the experiment, but in general they have the prime responsibility for managing the permanent data storage — raw, simulated and processed data — and providing computational capacity for reprocessing and for analysis processes that require access to large amounts of data. Today, 11 Tier-1 centres have been defined, most of them serving several experiments.

 The role of the Tier-2 centres is to provide computational capacity and appropriate storage services for Monte Carlo event simulation and for end-user analysis. The Tier-2 centres will obtain data as required from Tier-1 centres, and the data generated at Tier-2 centres will be sent to Tier-1 centres for permanent storage. More than 100 Tier-2 centres have been identified.

Other computing facilities in universities and laboratories will take part in the processing and analysis of LHC data as Tier-3 facilities. These lie outside the scope of the LCG Project, although they must be provided with access to the data and analysis facilities as decided by the experiments.

# Services.

## Storage Elements Services

A Storage Element (SE) is a logical entity that provides:

* Mass storage system, either disk cache or disk cache front-end backed by a tape system.
* SRM interface to provide a common way to access the MSS. The Storage Resource Manager (SRM) defines a set of functions and services that a storage system provides in an MSS implementation independent way.
* GridFTP service to provide data transfer in and out of the SE to and from the Grid. Normally the GridFTP transfer will be invoked indirectly via the File Transfer Service or through srmcopy.
* Authentication, authorization and audit/accounting facilities. The SE should provide and respect ACLs for files and datasets that it owns, with access control based on the use of extended X509 proxy certificates with a user DN and attributes based on VOMS roles and groups

##  File Transfer Services.

Basic-level data transfer is provided by GridFTP. This can be done directly via the

globus-url-copy command or through the srmcopy command. However, for reliable data transfer it is expected that an additional service above srmcopy or GridFTP will be used. This is generically referred to as a reliable file transfer service (rfts). The service itself is installed at the Tier-0 (for Tier-0– Tier-1 transfers) and at the Tier-1s (for Tier-1– Tier-2 transfers). It can also be used for 3rd-party transfers between sites that provide an SE.

##  Computer Resources Services.

The Computing Element (CE) is the set of services that provide access to a local batch system

running on a compute farm. Typically, the CE provides access to a set of job queues within the

batch system. A CE provides a mechanism by which work may be submitted to the local batch system and the publication of information through the Grid information system and associated information.

The Computing Element and associated local batch systems must provide authentication and authorization mechanisms based on the VOMS model.

##  Workload Management

The general feature of these services is that they provide a mechanism through which the application can express its resource requirements, and the service will determine a site that fulfils those requirements and submit the work to that site.

##  Database Services

These services provide the database back-end for the Grid file catalogues as either central services

located at CERN or local catalogues at the Tier-1 and Tier-2 sites. Based on scalable and reliable hardware using Oracle at the Tier-0, Tier-1 and large Tier-2 sites, and perhaps using MySQL on smaller sites.

## Grid Catalogue Services

The features for this service are:

• Mapping of Logical file names to GUID and Storage locations (SURL)

• Hierarchical namespace (directory structure)

• Access control

o At directory level in the catalogue

o Directories in the catalogue for all users

o Well-defined set of roles (admin., production, etc.)

• Interfaces are required to:

o POOL

o Workload Management Systems (e.g., Data Location Interface /Storage Index

interfaces)

o POSIX-like I/O service.

# Cost

The planning exercise for the CERN fabric uses the following input parameters to calculate the full cost of the set-up:

1. The base computing resource requirements from the experiments (CPU, disk and tape).

2. Derived resources (tape access speed, networking, system administration) from the combination of the base resources and the computing models.

3. The reference points of the equipment costs.

4. The cost evolution over time of the different resources.

The cost evolution uses a formula for each resource that assumes a smooth evolution over time. In the case of processors and disks this is close to Moore’s Law, a reduction of a factor of two in 18 months.

# Security

The components of the site security approach include firewalls, security monitoring and auditing, intrusion detection, training of system administrators and users, and the speedy patching of systems and applications.

The collaboration of a large number of independent sites into one Grid computing infrastructure potentially amplifies the security problems. Authentication is based on the Grid Security Infrastructure from Globus using a Public Key Infrastructure (PKI) based on X.509 certificates. An essential component of the PKI is the Certification Authority (CA), this being the trusted third-party that digitally signs the certificate to confirm the binding of the individual identity to the name and the public key. The CAs used by LCG are accredited by the three continental Grid Authentication Policy Management Authorities, namely the European, the Americas and the Asia-Pacific, under the auspices of the International Grid Federation.

Authorization to use LCG services and resources is managed via the use of VOMS, the Virtual Organization Membership Service, and local site authorization Grid services, such as LCAS and LCMAPS. The registered users of a VO are assigned roles and membership of groups within the VO by the VO manager.

The production and maintenance of LCG security policies and procedures will continue to be the responsibility of the Joint (LCG/EGEE) Security Policy Group.

# Benefits and challenges of computing Grid.

 The significant costs of maintaining and upgrading the necessary resources for such amount of data are more easily handled in a distributed environment, where individual institutes and national organizations can fund local computing resources and retain responsibility for them.

 In addition, with computer Grid there are no single points of failure. Multiple copies of data and automatic reassigning of computational tasks to available resources ensures load balancing and facilitates access to the data for all the physicists involved, independent of geographical location.

These include ensuring adequate levels of network bandwidth between the contributing resources, maintaining coherence of software versions installed in various locations, coping with heterogeneous hardware, managing and protecting the data so that it is not lost or corrupted over the lifetime of the LHC, and providing accounting mechanisms so that different groups have fair access, based on their needs and contributions to the infrastructure.

Furthermore, there are also some challenges while implementing a computing Grid such as ensuring adequate levels of network bandwidth between the contributing resources, maintaining coherence of software versions installed in various locations, coping with heterogeneous hardware, managing and protecting the data so that it is not lost or corrupted over the lifetime of the LHC, and providing accounting mechanisms so that different groups have fair access, based on their needs and contributions to the infrastructure.

# References.

http://wlcg-public.web.cern.ch

http://wlcg.web.cern.ch/documents-reference

https://indico.cern.ch/event/4896/contributions/2009114/attachments/959710/1362285/wlcg-service-challenge-closeout.pdf

https://home.cern/about/computing/grid-system-tiers