Worldwide LHC Computing Grid

Emil Hedemalm, 2016-06-06
emil-william.hedemalm8@univ-lorraine.fr

Course: Computing clusters, grids, clouds
Lecturer: Andrey Shevel

Overview

The Worldwide LHC Computing Grid (WLCG) is a global infrastructure made to assist the scientific missions surrounding the Large Hadron Collider (LHC), located on the border between France and Switzerland. The grid has multiple clusters on different tiers (officially >170 sites), supporting over 10’000 scientists in mainly 4 experiments (ALICE, ATLAS, CMS and LHCb) surrounding the fundamentals of matter.

Scale of data

The LHC accelerates particles to collide with each other using superconducting magnets. Each collision results in many fragments flying off into all directions. The experiment sites try to capture these fragments, generating a large amount of data for each collision. After some filtering of initial collision data, >99% is removed. The rest, around 30 petabytes annually, is disseminated throughout the WLCG network for storage, reconstruction and analysis.

The real-time speeds of data generation are on average 6 GB/s for the new iteration, with peaks of 10 GB/s. Previously, the same numbers for the first iteration were 1 and 6 GB/s respectively.

History - timeline

1949 - First proposal for joint research facilities
1953 - Location selected
1954 - *European Council for Nuclear Research* was adopted
1957 - First accelerator
1971 - Idea of ring-shaped accelerator conceived
1976 - The *Super Proton Synchotron* is turned on (first ring-shaped accelerator)
1990 - World’s first website and server go live @CERN
2006 - World’s largest superconducting magnet turns on
2007 - LHC Computing grid technical report (plan)
2008 - LHC starts up

The Computing Grid: Tiers

To enable collaboration, the WLCG is divided into several tiers. These tiers ease dissemination of data to all interested parties. Sites belonging to a certain tier may be addressed as *Tier-#* or *T#* sites later on.

Tier-0

Tier-0 consists of the main CERN data centre, with an extension in Hungary - the Wigner Research Center for Physics. Together they keep raw data safe, distribute it, and reconstruct some of the data. Less than 20% of total capacity of the WLCG is located in this tier.

The Wigner Research Centre for Physics is connected to the main CERN data centre using two redundant 100 Gbit/s direct connections, and its activities related to the WLCG are largely managed remotely from CERN.

Tier-1

Tier-1 consists of 13 large computer centers large enough to store LHC data. They provide 24/7 support for the grid and are responsible for storing a proportional share of raw- and reconstructed data. They also perform reprocessing, distribution of data to Tier-2s, and safe-keeping of the results produced by Tier-2 sites.

Each Tier-1 is connected to CERN (Tier-0) with optical-fibre links working at 10 Gbit/s, and the network of Tier-1 and Tier-0 sites is called the LHC Optical Private Network (LHCOPN).

Tier-2

The Tier-2 sites are mainly universities and scientific institutes capable of storing and processing a sufficient amount of data in the project. There amount to around 160 additional sites.

Tier-3

Individual scientists can access the grid as Tier-3 peers, whether it be via individual computers or local clusters in a university department. There is no formal engagement between Tier-3s and the WLCG, but they can still conduct their own research by accessing the Tier-2's data.

Structure

The structure of the WLCG is mainly described as the following parts: Networking, Hardware, Middleware & Physics Analysis.

Networking

As mentioned in *Tiers* part, the CERN-Wigner route features a dual 100Gb/s connection, with an average delay of 25ms, and an optical network (LHCOPN) between T0 and T1 sites operating at 10 Gb/s. Beside them, there is also the LHC Open Network Environment (LHCONE), where T3 peers can connect.

Data is exchanged on the grid via the Grid File Transfer Service (FTS), which solves several problems to require minimal configuration. Multiple backends (Oracle, MySQL), transfer and control protocols and scalability problems are addressed in this system.

Hardware

To enable scalability and reduce manual labor for each new server that is added to the grid, automation software is used. Automated installation, configuration & management is managed via *quattor*, developed at CERN (open source). It installs all necessary libraries (e.g. experiment-specific physics libs) and uploads device information to the Grid scheduling system to make it ready for use with minimal other input.

Tier-1 centres maintain disk and tape storage servers which require upgrades on a regular basis and specialized storage tools to operate.

Middleware

To help access and use the other computers distributed in the grid network, middleware softwares are used. These help with job submission, user authentication and authorization, etc. At the WLCG the following middlewares are used:

- The European Middleware Initiative (EMI), which combines other middleware providers (ARC, gLite, UNICORE and dCache) to satisfy user requirements concerning security, compute, data and information systems.
- The Globus Toolkit
- OMII, from the Open Middleware Infrastructure Institute, and
- Virtual Data Toolkit

Physics Analysis

Due to the changing and large demands which commercially available software cannot satisfy, specialized software is used for physics analysis. The physics software used on the grid includes:

- ROOT, a set of object-oriented core physics/mathematics libraries used by all LHC experiments, providing functionalities to deal with big data, statistical analysis, visualization and storage. Written in C++, but includes bindings for Python and R.
- POOL persistency framework, a framework for event data storage made specifically for the LHC project (full name Pool Of persistent Objects for LHC),
- Also other software which is used for modelling the various interactions of the elementary particles.

Storage and data systems

Most of the data in CERN is stored in magnetic tapes, using the CERN Advanced Storage system (CASTOR), and the rest is stored on a disk pool system called EOS allowing for multiple concurrent users for fast analysis access.

For the second iteration of the LHC experiments 140 petabytes of raw disk space is available, split between the CERN Data Centre and the Wigner Data Centre. Using chunking and segmentation options to keep some data readily available, this translates into about 60 petabytes of storage including back-up files. The chunking and scattering techniques also reduce data loss and reconstruction algorithms can reproduce content even if multiple disks fail.

Security

To maintain high uptime, the WLCG operates in accordance to several policies. For example, there is a 9-page document concerning the general Grid Security Policy[1], detailing how users are given authorities, the description of Virtual Organizations (VOs) which users have to belong to in order to use the network, mentioning of the Grid Acceptable Use Policy, etc.

There are two teams working on security operations surrounding the WLCG: the EGI CSIRT, which handles operational security and incident reports including forensics, and the OSG Security Team, which has security awareness guides, certification services etc.

[1] https://edms.cern.ch/ui/file/428008/5/Security_Policy_V5.7a.pdf , accessed 2016-06-06

Tools

Due to the sheer scale and complexity of the WLCG, it is not viable to describe all tools used here. But you can read more about the various live dashboards, middlewares (mentioned earlier), etc on the WLCG website: http://wlcg.web.cern.ch/tools

Related research

There is various research on the topics surrounding big grid operations at the WLCG (e.g. some 126+ papers published in 2016 with the keywords “WLCG & LHC”). Some recent examples are the following:

- The paper by Dubenskaya et al on “*New security infrastructure model for distributed computing systems*”
    - This paper tries an attempt to use a Hash table and values to replace the traditional public-key storage technique for authentication, due to the time-constraints of the unpredictability required for processing some calculation jobs.

- The paper by Forti et al on “*Multicore job scheduling in the Worldwide LHC Computing Grid”*
    - This paper describes the WLCG Multi-core task force’s work. It mainly concerns problems with scheduling jobs when not all jobs can be parallelized, and their attempts to minimize inefficiency due to the nature of the scheduling mechanisms.

References

- http://press.cern/press-releases/2013/06/cern-and-wigner-research-centre-physics-inaugurate-cern-data-centres , Accessed 2016-06-03
- http://home.cern/about/computing/grid-software-middleware-hardware , Accessed 2016-06-04
- https://wiki.egi.eu/wiki/EGI_CSIRT:Main_Page Accessed 2016-06-06
- https://twiki.opensciencegrid.org/bin/view/Security/WebHome Accessed 2016-06-06
- Dubenskaya J., Kryukov A., Demichev A., & Prikhodko N. (2016). *New security infrastructure model for distributed computing systems*. In Journal of Physics: Conference Series (Vol. 681, No. 1, p. 012051). IOP Publishing.
- Forti A., Yzquierdo A. P. C., Hartmann T., Alef M., Lahiff A., Templon J., ... & Filipcic A. (2015). *Multicore job scheduling in the Worldwide LHC Computing Grid*. In Journal of Physics: Conference Series (Vol. 664, No. 6, p. 062016). IOP Publishing.