EPJ Web of Conferences **60**, 20054 (2013)
DOI: 10.1051/epjconf/20136020054

The LHC Tier1 at PIC: experience from first LHC run

J. Flix[1,2,a], A. Perez-Calero Yzquierdo[1,2,b], E. Acción[1,3], V. Acin[1,3], C. Acosta[1,3], G. Bernabeu[1,2,c], A. Bria[1,3,d], J. Casals[1,2], M. Caubet[1,2], R. Cruz[1,3], M. Delfino[1,4], X. Espinal[1,3,e], E. Lanciotti[1,3,e], F. López[1,2], F. Martinez[1,2,f], V. Méndez[1,3], G. Merino[1,2,g], A. Pacheco[1,3], E. Planas[1,2], M.C. Porto[1,2], B. Rodríguez[1,3], and A. Sedov[1,3]

[1] *Port d'Informació Científica (PIC), Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain*
[2] *Also at Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, CIEMAT, Madrid, Spain*
[3] *Also at Institut de Física d'Altes Energies, IFAE, Edifici Cn, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain*
[4] *Also at Universitat Autònoma de Barcelona, Department of Physics, Bellaterra (Barcelona), Spain*

Abstract. This paper summarizes the operational experience of the Tier1 computer center at Port d'Informació Científica (PIC) supporting the commissioning and first run (Run1) of the Large Hadron Collider (LHC). The evolution of the experiment computing models resulting from the higher amounts of data expected after the restart of the LHC are also described.

1 Introduction

The LHC, at the European Laboratory for Particle Physics (CERN, Switzerland), started operating in November 2009 and it has generated around 200 Petabytes of raw, simulated and processed data, from all of its detectors, until the stop of the successful first run in February 2013. To analyze the unprecedented rate of PB of data per year, a Grid-based computer network infrastructure was built, the largest scientific distributed computing infrastructure in the world, adding up the computing resources of more than 170 centers in 34 countries: the Worldwide LHC Computing Grid (WLCG [1][2]). In the WLCG, the computing centers are functionally classified in Tiers. Eleven of these centers are the so-called Tier1s, receiving a copy of the raw data in real time from the Tier0 at CERN, and in charge of massive data processing, storage and distribution.

Spain contributes to the WLCG with one Tier1 centre: Port d'Informació Científica (PIC), located in the campus of the Universitat Autònoma de Barcelona, near the city of Barcelona. PIC provides services to three of the LHC experiments, ATLAS, CMS and LHCb, accounting for ~5% of the total Tier1 resources, acting as the reference Tier1 for the Tier2 centers in Spain and Portugal, and sites located in Valparaiso (Chile) and Marseille (France).

PIC has been an active and successful participant in the WLCG project since its start, showing its readiness for the LHC data taking period. In the first phase, contributing to prototyping and testing of the Grid middle-ware and services that were being developed. Later, participating in the Service Challenges carried out by the experiments, testing campaigns aimed to progressively ramp-up the level of load on the infrastructure under conditions as realistic as possible, achieving breakthrough performances.

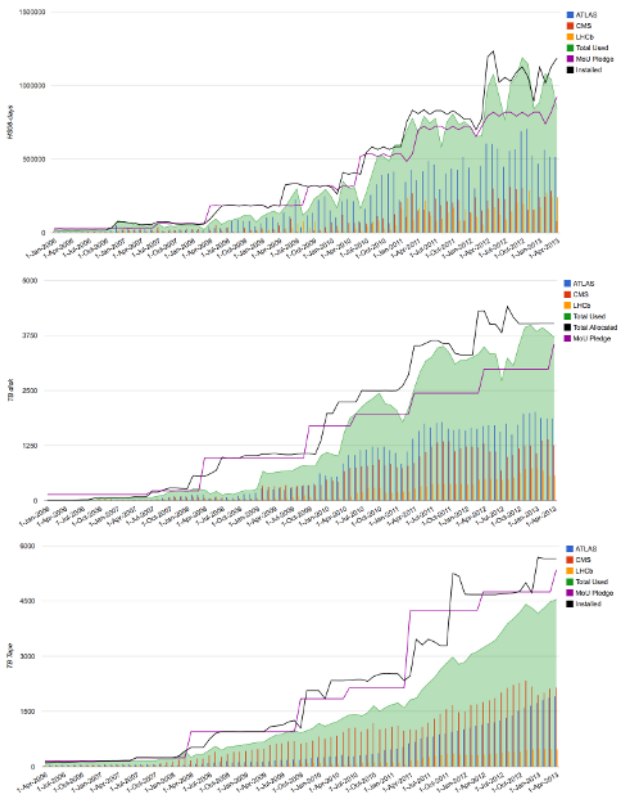

Figure 1: CPU (top), disk (center) and tape (bottom) resources installed and used at PIC, since 2008.

[a] e-mail: jflix@pic.es
[b] e-mail: aperez@pic.es
[c] Currently at FNAL, Chicago, USA
[d] Currently at center de Regulació Genòmica, Barcelona, Spain
[e] Currently at CERN, Switzerland
[f] Currently at Telefónica I+D, Barcelona, Spain
[g] Currently at Wisconsin IceCube Particle Astrophysics Center, USA

As of today, the computing resources installed in PIC are of around 4000 cores (on ~600 CPUs) managed by Torque/Maui. This corresponds to about 35000 HEP-SPEC06 (HS06, see Ref. [3]). The servers are typically two quad-core x86 CPUs, with at least 2GB RAM per core. Each of these nodes has two 10Gbps Ethernet interfaces which are then aggregated in switches and connected to the storage infrastructure. The main servers consist of Blades (HP) and Dual-Twins (Dell).

The storage service at PIC is managed by dCache [4] and Enstore [5] softwares. The dCache software provides uniform and efficient access to the disk space provided by many file servers, and talks to the Enstore software to interface to magnetic tape storage. As of today, 5.5 PB of disk space is installed, by means of around 3000 hard disks of 1, 2 and 3 TBs, distributed on around 70 servers x86, each connected by 4 aggregated 1Gbps Ethernet or one 10Gbps Ethernet, depending on the hardware. The servers brands comprises DataDirect, SGI, and SuperMicro.

The current tape infrastructure at PIC consists of two automated tape libraries, Sun StorageTek 8500SL and IBM TS3500, providing around 8500 tape slots which should be sufficient to cover the Tier1 needs in the coming years. 8 PB of tape storage is managed by Enstore, with access to a total of 4.6 million files. The supported technologies are LTO-3 (read-only), LTO-4, LTO-5 and T10KC, containing 2%, 25%, 36% and 38% of the total data respectively, in around 7000 tape cartridges. A total of 26 tape units are installed to read/write the data (16 LTO-4, 4 LTO-5 and 6 T10KC). Aggregated read/write rate has achieved hourly average rates peaking at 1 GB/s.

2 A reliable, high-capacity service

One of the main characteristics of Tier1 centers, beyond a very large storage and computing capacity, is to provide these resources through services that need to be extremely reliable. Being closely connected to the detectors data acquisition, a maximum time for unintended interruption of the services in a Tier1 is set to 4 hours, and a maximum degradation of Tier0 to Tier1 data acceptance of 6 hours. Critical services in a Tier1 operate in 365x24x7 mode.

Service quality and stability are amongst the cornerstones of the project, therefore they are closely tracked by monitoring two metrics provided by the SAM monitoring framework: site Availability and Reliability. These are built from dozens of sensors which hourly probe all of the site Grid services, which ensures peer pressure and guarantees that the reliability of WLCG service keeps improving.

Figure 2 shows the monthly Reliability results for PIC since May 2006, almost always above the target and the Tier0/Tier1 average, 86% and 88% of the months, respectively. It is worth mentioning that PIC needs to have an expert contact person on site (the liaison), communicating and coordinating priorities with each of the experiments and resolving operational problems. This helps PIC being at top reliability and stability levels.

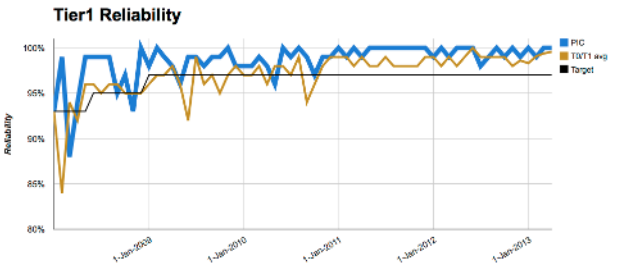

Figure 2: Site Reliability for Tier0 and Tier1s since May 2006. PIC Tier1 (blue) is compared to the average (yellow) and to the target set by the project (black).

3 PIC performance during LHC Run1

PIC responsibilities as a Tier1 include running data processing jobs. Monte-Carlo (MC) generation and processing are also among its tasks, along with a small proportion of analysis jobs, normally run at Tier2s. Millions of jobs run annually at PIC, main customers being ATLAS-Tier1, CMS-Tier1 and LHCb-Tier1, but also covering other experiments needs (ATLAS Tier2, ATLAS Tier3, Cosmology, ...). The CPU efficiency of LHC jobs has increased along the years, being around 90% since the LHC start.

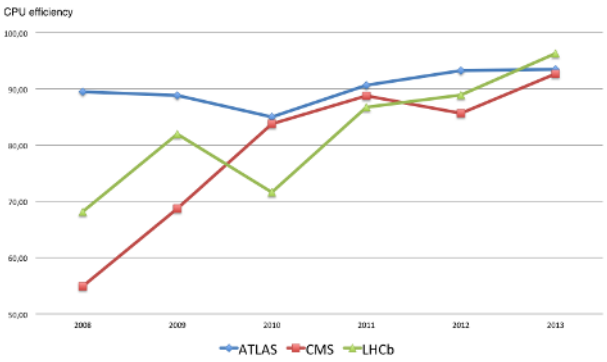

Figure 3: Yearly average of CPU efficiency for jobs run at PIC since 2008.

In order to ensure the needed quality and performance in the critical data transfers from the Tier0 to the Tier1 centers, the WLCG project deployed an Optical Private Network (OPN, see Ref. [6]) using GÉANT2 and the National Research and Educational Networks (NRENs). The OPN connects CERN and the Tier1 centers through point-to-point dedicated links of 10Gbps. The OPN link connecting PIC and CERN is operative since 2007 and uses infrastructure from the Catalan NREN (Anella Científica). Many extended periods of high load and a notably high average load for a link of this capacity are observed in PIC, with a quite stable behaviour. A back-up line is provided as well, in case the primary goes down incidentally.

WLCG involves massive data transfers between Grid sites. Good performance links and reliable data transfer systems are a must. During Run1 the monthly averaged

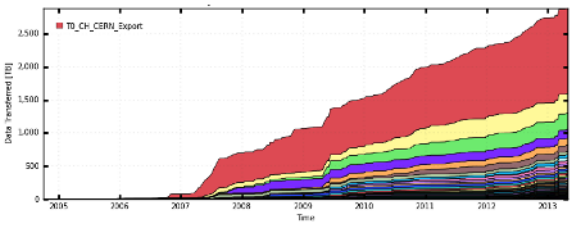

Figure 4: CMS data transfers to PIC from Tier0.

value for incoming (outgoing) transfers to PIC has been of about 250 (400) MB/s, with hourly peaks exceeding 2 GB/s. Recently, a PerfSONAR-PS installation has been deployed to monitor the health of the network. Tier1s are currently connected to Tier2s via NRENs. All sites will soon be connected by a cutting edge dedicated network (LHCONE). PIC coordinates the PerfSONAR-PS and LHCONE deployments for all of the Iberian computing sites.

Tier1s provide the experiments with mass storage on tape for custodial replicas of raw and processed data, as well as MC samples. Read and write average rates have increased along the years as the amount of data increases and new tape technologies become available. Total rate at PIC has achieved hourly average rates peaking at 1 GB/s.

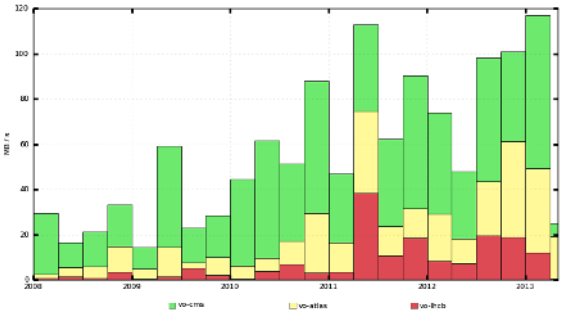

Figure 5: PIC tape service read/write rate, since 2008.

4 Getting prepared for the LHC restart

Though data taking is stopped at the moment, the experiments continue reprocessing the full sample of collected data, adding parked data in Tier0 (CMS case), and producing MC files with the expected Run2 conditions. Additionally, new functionalities and tools are being tested and integrated in the system.

In 2015, LHC experiments will restart data taking at increased collision energy and trigger rates. LHC computing needs to prepare for, at least, twice the amount of data. A number of improvements, currently under development, have been proposed. Firstly, advance towards generic tools used by the LHC VOs (job submission, network and storage monitoring, etc). Data transfers and storage resources will undergo an integration into storage federations (e.g. CMS AAA project), benefiting from increased network rates (LHCONE) and including new transfer protocols, e.g. xrootd/http. This will effectively decouple where data is and where jobs run. Parallel Computing and multi-core job might be a must, as increased pileup require processing more complex events with improved memory management in multi-thread applications running on multi-core CPUs. Cloud Computing and opportunistic resources are key points as well. PIC is involved in many of these projects, either participating or coordinating them.

The successful operation of PIC during LHC run1 helped in the discovery of the Higgs Boson and hunting for exciting new physics. Despite hardware retirements, PIC resources will continue growing to cope with experiments requirements, being full ready to enter into LHC Run2. Among many ongoing tasks, it is worth mentioning that an extension of a private network to Tier2 and Tier3 sites is being built, the so called LHCONE, bringing those sites into a more robust framework. PIC coordinating the Iberian deployment of this new setup. The current network is based in IPv4, and PIC is extensively testing all of the services in IPv6, to migrate to the new schema by 2015.

5 Acknowledgements

The Port d'Informació Científica (PIC) is maintained through a collaboration between the Generalitat de Catalunya, CIEMAT, IFAE and the Universitat Autònoma de Barcelona. This work was supported in part by grants FPA2007-66152-C02-00 and FPA2010-21816-C02-00 from the Ministerio de Ciencia e Innovación, Spain. Additional support was provided by the EU 7th Framework Programme INFRA-2007-1.2.3: e-Science Grid infrastructures Grant Agreement Number 222667, Enabling Grids for e-Science (EGEE) project and INFRA-2010-1.2.1: Distributed computing infrastructure Contract Number RI-261323 (EGI-InSPIRE).

References

[1] "LHC Computing Grid Technical Design Report", CERN-LHCC-2005-024, 20 June 2005.
[2] "Computing for the Large Hadron Collider", Ian Bird, "Annual Review of Nuclear and Particle Science", Vol. 61: 99-118, November 2011.
[3] https://hepix.caspur.it/benchmarks/doku.php.
[4] http://www.dcache.org.
[5] J. Bakken et al., "Enstore Technical Design Document", http://www-ccf.fnal.gov/enstore/design.html.
[6] LHC Optical Private Network, http://lhcopn.cern.ch.