IL NUOVO CIMENTO Vol. 32 C, N. 2 Marzo-Aprile 2009
DOI 10.1393/ncc/i2009-10438-0

Colloquia: CSFI 2008

The computing of the LHC experiments

M. Paganoni
Università di Milano-Bicocca and INFN, Sezione di Milano - Milano, Italy

(ricevuto il 4 Agosto 2009; pubblicato online il 23 Ottobre 2009)

Summary. — The LHC experiments have thousands of collaborators distributed worldwide who expect to run their physics analyses on the collected data. Each of the four experiments will run hundreds of thousands of jobs per day, including event reconstruction from raw data, analysis on skimmed data, and production of simulated events. At the same time tens of petabytes of data will have to be easily available on a complex distributed computing fabric for a period of at least ten years. These challenging goals have prompted the development and deployment of reliable Grid services, which have been thouroughly tested and put at the needed scale over the last years. This paper concentrates on CMS computing needs for data taking at LHC and highlights the INFN-Grid contribution to the effort.

PACS 89.90.+n – Other topics in areas of applied and interdisciplinary physics.
PACS 07.05.Hd – Data acquisition: hardware and software.

1. – The CMS Distributed Computing Model

The CMS experiment has developed a distributed computing system capable of serving, processing and archiving the large number of events, amounting to several petabytes per year, which will be generated during data taking. The CMS Computing Model [1, 2] foresees about 20% of the resources at CERN (Tier-0 and CAF), where computing activities with significant latency constraints, like prompt reconstruction, are performed. Another 40% of the resources are concentrated in large-size computing centers (Tier-1), responsible for serving a portion of the data and simulated events, for reprocessing and skimming activities and for the long-term custody of the data. Presently seven Tier-1 are active for CMS: ASGC in Taiwan, FNAL in US, GridKA in Germany, IN2P3-Lyon in France, INFN-CNAF in Italy, PIC in Spain, RAL in UK. The remaining computing resources are located at about 40 medium-size centers (Tier-2) which are the primary locations for physics analysis and generation of simulated events. In the CMS model the location of the data drives the activities on the sites as the jobs are steered where the data have been pre-located. In table I the computing resources presently available for CMS are summarized.

Table I. – *Computing resources presently available for CMS.*

	CPU (kHS06)	Disk (PB)	Tape (PB)
Tier-0 and CMS CAF	44	2.2	10
Tier-1	58.9	9.7	16.5
Tier-2	141.8	9.2	-

2. – The main CMS workflow components

The architecture of the CMS computing relies on a distributed system of services and resources that are part of the Worldwide LHC Computing Grid (WLCG) [3]. The CMS Workflow Management system (WM) manages the large-scale data processing that is optimized to let the final users access efficiently the data streams and perform their physics analysis. It supports all CMS necessary workflows (data re-reconstruction, calibration activities, Monte Carlo production, AOD production, skimming and physics analysis) and shields users from the full complexity of the underlying architecture. The CMS WM makes use in a coherent way of:

– the CMS Grid Workload Management System (Grid WMS) which schedules jobs onto the distributed resources according to the CMS policy and priorities and monitors the jobs status;

– the CMS Data Management system (DM) which allows to discover, transfer and access data sets of different kinds across the Grid.

In order to manage and track the datasets, CMS has developed its own data management services:

– DBS, the dataset bookkeeping service, knows how a group of files forms a dataset and maps the files into logical quantities called data blocks, the smallest quantity we expect to track in the data transfer system and the global data catalog.

– PhEDEx [4], a reliable and scalable dataset replication tool, guarantees managed and structured data flow, by monitoring the data transfer and data integrity at the level of the file blocks.

3. – The commissioning of the sites

In the last years a remarkable work was devoted, through Data Challenges, to demonstrate that Grid services and resources scale well beyond what is needed at the LHC startup. Furthermore a continuous monitoring and reporting process improved substantially the reliability of the Grid infrastructure. A detailed procedure was defined in order to guarantee that data processing can be performed in an efficient and reliable way. Hence all relevant aspects of a Grid site were extensively tested in a continuous mode before integrating it into the daily operations. Automatic functionality and reliability tests (SAM tests) are used in order to constantly monitor the quality of the site. At the same time the Debugging Data Transfer (DDT) task force uses the PhEDEx LoadTest

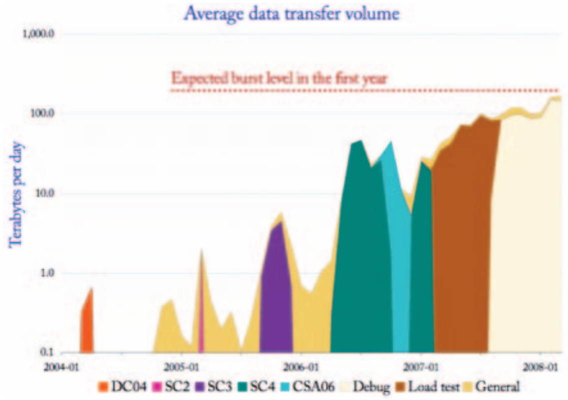

Fig. 1. – The overall CMS data transfer over the years of the computing commissioning.

tool to generate data traffic transfer among sites and commission links between all CMS Tier centers, including crosslinks between Tier-1 sites and upload/download links between Tier-1 and Tier-2 sites. As can be seen in fig. 1 the overall CMS data transfer has steadily increased over the years to reach the target at the LHC startup.

4. – The physics analysis at the Tier-2 sites

CMS expects to process approximately 100k jobs per day at the startup of the data taking. One of the most complex issues of the Computing Model is to ensure that data can be analyzed in an efficient way at the Tier-2 sites by the community of about 3000 physicists from all over the world. INFN has designed and developed CRAB (CMS Remote Analysis Builder) [5], a specific tool that allows end-users an easy and transparent access to the distributed data. CRAB has adopted a client-server architecture in order to automate the analysis workflow, leaving to the user just minor actions. In this way CRAB fully integrates into WLCG Grid infrastructure and into the CMS data and workload management system. CRAB is in production and has been extensively used since Spring 2004. Presently an increasing number of Tier-2 centers are successfully hosting data samples and accepting analysis jobs. Peaks of daily submissions equal to 70000 jobs/day, for a total of 10 million jobs in 2008, on simulated data and real cosmic data, have been reached. The number of distinct CRAB users is about 1500, which is approaching the 50% of the whole CMS community and more than 100 users submit jobs each day. The CMS goal for grid submission success this year is 90%, but much higher efficiency is desired in the final system.

5. – The use of the INFN Grid infrastructure

The INFN Grid infrastructure is used continuously with good performance by final users. In particular in the Tier-2 sites both production of simulated events and end-user's analysis are taking place. It is now standard to see more than 100k concurrent CMS jobs running worldwide. More than 10k jobs per day using only the Italian resources can be achieved. This proves the stability and the maturity of the INFN-GRID infrastructure.

REFERENCES

[1] The CMS Collaboration, *The CMS Computing Model*, CMS NOTE 2004-031.
[2] The CMS Collaboration, *The CMS Computing Project Technical Design Report*, CERN/LHCC 2005-023.
[3] Jacobs D. *et al.*, *WLCG Memorandum of understanding*, CERN-C-RRB-2005-01/Rev.
[4] Tuura L. *et al.*, *J. Phys. Conf. Ser.*, **119** (2008) 072030.
[5] Spiga D. *et al.*, *Nucl. Phys. B, Proc. Suppl.*, **177-178** (2008) 267.