IL NUOVO CIMENTO Vol. 32 C, N. 2 Marzo-Aprile 2009
DOI 10.1393/ncc/i2009-10387-6

Colloquia: CSFI 2008

LHCb distributed data analysis

A. Carbone([1]) and S. Perazzini([1])([2])

([1]) *INFN, Sezione di Bologna - viale Berti Pichat 6/2, 40127 Bologna*
([2]) *Dipartimento di Fisica, Università di Bologna - Via Irnerio 46, 40126 Bologna*

(ricevuto il 4 Agosto 2009; pubblicato online il 4 Settembre 2009)

Summary. — This paper describes the LHCb distributed data analysis system. For data analysis, the LHCb Collaboration will exploit the WLCG distributed computing resources spread all over the world. Two main software tools will be used: DIRAC and GANGA. DIRAC is mainly used to easily access distributed computing resources, while GANGA is a user interface that allows users to submit analysis jobs in a data-driven mode.

PACS 29.85.-c – Computer data analysis.
PACS 29.85.Fj – Data analysis.

1. – The LHCb Computing Model

The LHCb [1] experiment has been designed to study the physics of the b quark (CP violation and rare decays) at LHC, exploiting the large value of the expected $b\bar{b}$ cross-section at the 14 TeV p-p collision of LHC, running at a luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, well below the highest achievable luminosity of the machine. The expected event data flux, as shown in fig. 1, from the detector to the storage system, taking data at an event rate of 2 kHz, amounts to about 70 MB/s, with an event size of about 35 kB. The computing resources to be used for the offline data processing are mainly available through the WLCG (Worldwide LHC Computing Grid), a wide-area distributed hierarchical system; CPU resources and disk pools are distributed at WLCG sites across the world. Data processing, reconstruction, selection and user analysis are planned to be performed at the CERN Tier-0 centre and at the 6 Tier-1 National Computing centres. LHCb Tier-2 computing centres are planned to be used mainly for the production of Monte Carlo simulated events. Only larger Tier-2 centres could be used, in case, also for data analysis. More details on the LHCb computing model can be found here [2]. A proof of consistency of the LHCb computing model has been performed during the so-called data challenge activities (latest was DC06), which involved all the steps of the computing model. Large amounts of Monte Carlo data were produced: about 700 million events were simulated over about 120 different computing centres, during a

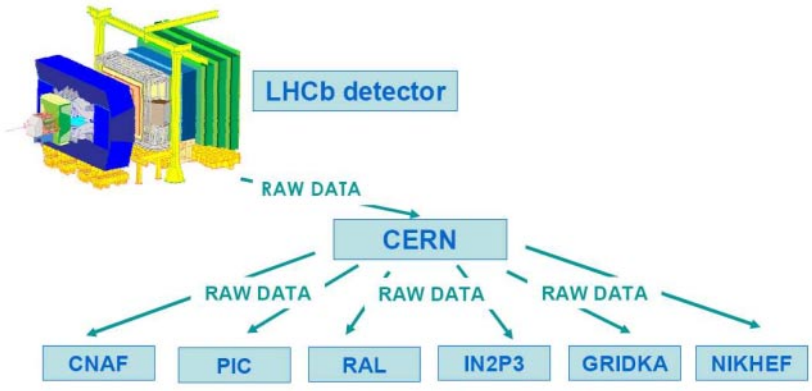

Fig. 1. – Raw data distribution. The expected event data flux (raw data) from the LHCb detector to the CERN Tier-0 is about 2 kHz. The computing model foresees that data processing, reconstruction, selection and user analysis are performed at the CERN Tier-0 centre and at the 6 Tier-1 National Computing centres.

period of 475 days, with a mean value of 1.5 million events produced per day. Up to 10 k jobs were running on the Grid simultaneously.

2. – Distributed data analysis

The distributed analysis model foresees the usage of the LHCb Tier-1 centres for data analysis. Within the LHCb Collaboration, the analysis is realized by means of GANGA (Gaudi/Athena and Grid Alliance) [3], a Grid user interface, specifically designed to provide the user with very easy-to-use means to submit jobs to the Grid through DIRAC [4], *i.e.* the software layer developed in order to provide the collaboration with a tool for an easy access to the distributed computing resources either for data production or analysis.

2˙1. *DIRAC*. – One of the main design goals of DIRAC is the simplicity of the installation, configuration and operation of various services it consists of. Once installed and configured at sites, the system automatically provides most of the management tasks.

The main elements of the DIRAC architecture are:

– *Workload Management System (WMS)*: It consists of a central task queue and a network of agents. The WMS is used as service to submit to the Grid Monte Carlo data production jobs, data processing jobs and user analysis jobs, by means of the "pull" scheduling paradigm. The adopted scheduling mechanism consists in checking the operational environment by means of specialized agents before pulling the real jobs from the central task queue, allowing an optimal exploitation of the available resources.

– *Data Management System*: It includes several tools to handle data management tasks of storage and file catalogs. It provides instruments to perform basic file management operations such as file uploading, replication and registration in a uniform way, masking the diversities of storage systems available at the sites. In addition data distribution and integrity checks of the replicated data are performed.

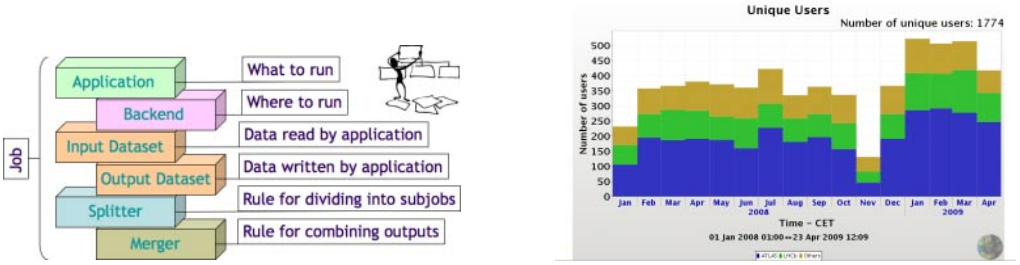

Fig. 2. – Building blocks for constructing a GANGA job (left); number of unique users running analysis jobs from January 2008 until April 2009 for LHCb, Atlas and other collaborations (right).

– *Production Management System*: It provides tools to deal with the execution of production tasks requested by the LHCb Collaboration. These tools allow production managers to handle hundreds of thousands of jobs in a very efficient way during the production shifts of the data taking.

2˙2. *GANGA.* – GANGA deals with configuring the LHCb applications, submitting the jobs and keeping track of the results. An analysis job in GANGA consists of a set of building blocks as shown in fig. 2:

– *Application*, a common interface to set the configuration of programs used, *e.g.*, software version and configuration of program components. On the one hand, this interface has to be adapted to the specific needs of the Atlas and LHCb Collaboration, but must also prove to be flexible enough to interface to different applications in order to be used by other scientific comunities.

– *Backend*, where the job is submitted. User analysis programs can be sent to different local or remote computing sites. One can choose different local batch systems or the GRID. In particular within the LHCb Collaboration is recommended to use DIRAC as the door to the GRID in order to have under the same tool a simplified access to the distributed computing resources, a centralized accounting and a job prioritization system.

– *Input Dataset*, the list of file names to be processed. The user selects the desired dataset from the GANGA-bookkeeping interface specifing only the file name of the data sets as a Logical File Name (LFN) list. Several replicas (at least one) located at Tier-1 sites may correspond to a given LFN. This allows the users not to care about the location of data since jobs are submitted for execution automatically where the data resides (submission is data driven).

– *Output Dataset*, the list of files written by the job. Ganga allows to retrieve logging, error messages and output (*e.g.*, n-tuple files) after a job finishes. All the output files are automatically transferred to the personal computer used for the submission. Only in case the size of the output is over a certain threshold, the system takes care to copy the file to the storage system of the Tier-1 where the job was sent.

– *Splitter*, a tool to split the job in sub-jobs. This allows to submit each sub-job independently according to the geographical location of a given sub-data set. Figure 3 shows a schematic view of the splitter behaviour.

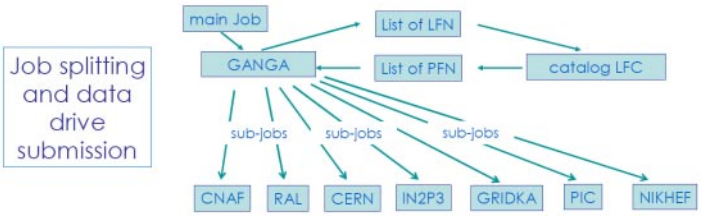

Fig. 3. – A single job can be splitted in sub-jobs. This allows to submit each sub-job independently according to the geographical location of a given data set.

– *Merger*, a tool to merge the sub-jobs output into a single one. After a job finishes it is possible to merge job results if the main job has been executed in parallel sub-jobs. A typical example is the merging of ROOT n-tuples.

In addition to the functionalities described above, Ganga is provided by other two main services. The former is the monitoring system that allows a continuous monitoring of the job status. Information about the remote execution site, queue status and successful termination is collected. Information about possible errors is gathered and categorized. The job scheduler should offer an automatic job resubmission function for failed jobs. The latter consists of a job archive, in order to offer information about terminated user jobs and to provide templates for new user jobs.

Ganga is largely used within the LHCb Collaboration. The number of total unique users from January 2008 until April 2009 running analysis jobs through GANGA were about 350. Figure 2 reports the number of unique users from January 2008 until April 2009, not only for the LHCb Collaboration but also for Atlas Collaboration and for other scientific communities.

REFERENCES

[1] Alves A. A. *et al.* (LHCb Collaboration), *JINST*, **3** (2008) S08005.
[2] LHCb Collaboration, *The LHCb Computing Moldel*, CERN-LHCC-2005-019, LHCb-TDR-11.
[3] Maier A., *J. Phys. Conf. Ser.*, **119** (2008) 072021.
[4] Tsaregorodtsev A. *et al.*, *J. Phys. Conf. Ser.*, **119** (2008) 062048.