IL NUOVO CIMENTO Vol. 32 C, N. 2 Marzo-Aprile 2009
DOI 10.1393/ncc/i2009-10389-4

Colloquia: CSFI 2008

Mass production of event simulations for the BaBar experiment using the Grid

A. Fella(1), D. Andreotti(2), E. Luppi(2) and L. Li Gioi(1)

(1) *INFN CNAF - Bologna, Italy*
(2) *INFN, Sezione di Ferrara - Ferrara, Italy*

(ricevuto il 22 Giugno 2009; pubblicato online il 4 Settembre 2009)

Summary. — The *BaBar* experiment has been taking data since 1999, investigating the violation of charge and parity (*CP*) symmetry in the field of High Energy Physics. Event simulation is an intensive computing task, due to the complexity of the algorithm based on the Monte Carlo method implemented using the GEANT engine. The simulation input data are stored in ROOT format, they are classified into two categories: conditions data for describing the detector status when data are recorded, and background triggers data for including the noise signal necessary to obtain a realistic simulation. In order to satisfy these requirements, in the traditional *BaBar* computing model events are distributed over several sites involved in the collaboration where each site manager centrally manages a private farm dedicated to simulation production. The new grid approach applied to the *BaBar* production framework is discussed along with the schema adopted for data deployment via Xrootd/Scalla servers, including data management using grid middleware on distributed storage facilities spread over the INFN-GRID network. A comparison between the two models is provided, describing also the custom applications developed for performing the whole production task on the grid and showing the results achieved.

PACS `07.05.Bx` – Computer systems: hardware, operating systems, computer languages, and utilities.
PACS `89.20.Ff` – Computer science and technology.
PACS `07.05.-t` – Computers in experimental physics.
PACS `87.10.Rt` – Monte Carlo simulations.

1. – Introduction

The *BaBar* High-Energy Physics (HEP) detector [1] is based at the Stanford Linear Accelerator Center (SLAC), Stanford, USA. The primary goal of the experiment is the study of *CP* violation of simmetries in the decay of the B-meson. The secondary goals are the precision measurement of τ leptons, bottom and charm mesons decays and searches for rare processes accessible because of the high luminosity of PEP-II B-Factory [2], an

$e^+ e^-$ asymmetric collider running at a center of mass of 10.58 GeV. The *BaBar* experiment has accumulated to date about 540 fb^{-1} integrated luminosity. In addition to analyzing the data, a major task is the production of simulated events. At least three times as many simulated events are needed as data events. Jobs are allocated to a specialized distributed computing system based on five computer farms. The experimental and simulated events, the detector conditions, alignment and calibration constants are stored and accessed using the ROOT I/O protocol [3] either directly through network file system or by a load-balanced, fault-tolerant data access service called Xrootd [5] or via General Parallel File System [4] at CNAF Italian Tier-1. The system requires a great deal of local customization with each dedicated farm maintained by a local production manager. Despite these drawbacks, the system run with greater than 98% efficiency. It is natural for such a specialized distributed system to evolve into a simpler and more efficient one by using a general-purpose approach based on the tools and middleware provided by GRID systems such as the LHC Computing GRID Project, LCG [6]. The schema used in this work is based on INFN-GRID [7], the Italian project implemented on the LCG/gLite [8] middleware.

2. – Simulation production using INFN-GRID

At present time the resources accessed by *BaBar* Simulation Production (SP) on INFN-GRID include the CNAF Italian Tier-1 site and eight INFN-GRID sites: Ferrara, Napoli, Padova, Catania, Bari, Perugia, Pisa, Cagliari. The submission and output retrieving operations are hosted on the dedicated User Interface (UI) machine at INFN CNAF site in Bologna and the used Resource Broker (RB) is located at Ferrara.

2·1. *Software scenario*. – The *BaBar* software needed to produce Monte Carlo events has been packaged in a simple compressed tar archive of about 70 MBytes, $\sim$ 450 MBytes uncompressed. The package is then distributed on each INFN-GRID site and run time accessed by jobs. The *BaBar* tools for simulation production are installed on the UI and properly configured. These tools are needed to perform post production operations:

– Merging of several collections of homogeneous events in a single collection.

– Transfer the simulated events to SLAC Mass Storage System.

– Update the central bookkeeping database at SLAC.

In order to submit production runs, a software layer was developed and maintained on the UI. These scripts are involved in dynamically drawing up GRID job setup files that define the job input and output data file set, and specify job brokering requirements and simulation parameters. The latter includes: type of events, run number, background triggers to use, number of events, detector condition files. The *BaBar* software layer permits to manage the jobs submission through the UI to the dedicated RB at Ferrara site that is able to manage Italian and European resources directly involved in *BaBar*. The generated Monte Carlo events are saved as ROOT I/O files and transferred from WNs to a Storage Element (SE) and subsequently transferred to the UI where the merging and export to SLAC stages are performed. A conceptual schema of the work flow described above is shown in fig. 1.

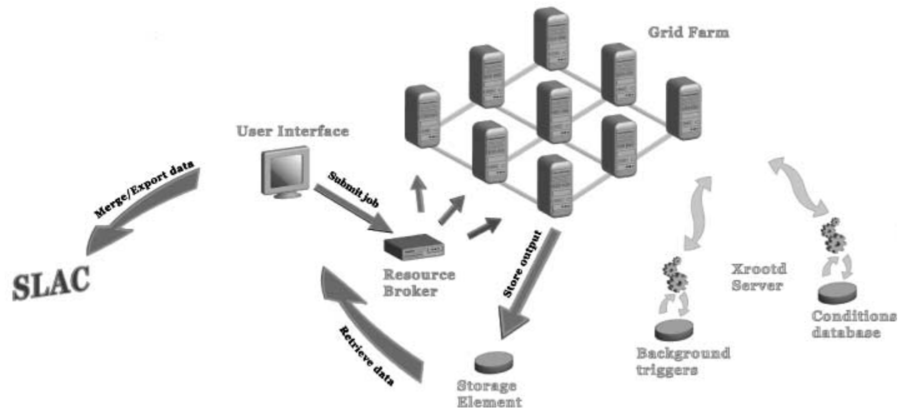

Fig. 1. – BaBar simulation production workflow on the INFN-GRID system.

3. – Input and output data management

As already mentioned, the *BaBar* simulation job involves the mixing of genuine random trigger data and Monte Carlo generated events in order to have a realistic simulation of the detector background. The production of simulated events needs as input background collections, the detector conditions and calibration constants. Background collections are produced on a monthly basis, and consist of about 100 k random triggers, each typical production job reads 2000 background collections. Moreover the detector conditions and calibration constants consist of about 30 GBytes data, frequently updated. The typical amount of condition data read by a single simulation job is about 1/50 of the total.

Background collections, detector conditions and calibration constants are stored as ROOT I/O files in three selected INFN-GRID sites involved in simulation production, and accessed by the Xrootd server facility locally or through the Wide Area Network (WAN). At CNAF Tier-1 site the metadata access is performed by GPFS.

3˙1. *Job output data handling.* – During 2007 the code involved in data transfer procedures was substantially improved in terms of reliability. In this section we are going to discuss the implementation of the job output data transfer from the WN where the job runs, to the SE, and from the SE to the UI at job completion time. The SP jobs transfer the output to the site Storage Element (Closest SE) by LCG utilities. The LCG data management system refers to a distributed file replica catalogue called LCG File Catalogue (LFC) [9] in which all the logical and physical file addresses information is maintained. The analysis of failure services scenario brought us to develop a multi-level transfer recovery structure:

– All the systems work properly: the system default behaviour is to use the LCG utilities integrated with LFC service. The job registers the simulated events file to the LFC service and transfers it to the Closest SE.

– The Closest SE is unreachable: at job completion the output file is registered to the LFC service and transferred to the CNAF SE by LCG utilities.

– The LFC service is down: the job output file is transferred directly to the Tier-1 SE using low-level Globus Toolkit [10] utilities.

4. – Monitor system

The process at the base of the monitor system relies on the LCG Logging and Bookkeeping information and on *BaBar* SP job logs collected via the GRID Job Output Sandbox. The process runs as a daemon on the UI machine and queries (via LCG utilities) the job status for all the jobs previously prepared by *BaBar* tools on the file system. At job completion the daemon applies a retry policy in case of job failure or takes care of retrieving the output logs, performing the produced simulated events transfer (via LCG data management tools) from the SE to the UI, updating the job status. A web-based monitor system accesses the data collected on the UI, performs SLAC Data Base queries and publishes several graph views and statistics [11]. A wider treatment of monitor system structure is available in refs. [12, 13].

5. – Conclusion

The simulation of HEP events is now feasible using the GRID architecture; so far the *BaBar* Simulation Production in the period 2005-2007 on LCG-based GRID counts 700 million events generated in the UK and Italy. The peak production rate in Italy was 25 million events per week with $\sim$ 1000 jobs running simultaneously on INFN-GRID sites. The *BaBar* SP on INFN-GRID is doubled in 2007 with respect to the 2006 results. Reliability approaches that of private *BaBar* computing farms. An optimal data availability in the WAN scenario is garanteed by the use of the LCG utility and Xrootd data access system addressing, respectively, the job output transfer to UI and the job access to the background trigger files.

REFERENCES

[1] Aubert B. *et al.*, *Nucl. Instrum. Methods A*, **479** (2002) 1.
[2] PEP-II An Asymmetric B Factory CDR SLAC, June 1993.
[3] Brun R. and Rademakers F., *Nucl. Instrum. Methods A*, **389** (1997) 81.
[4] General Parallel File System (Online). Available: `http://publib.boulder.ibm.com/infocenter/clresctr/vxrx/topic/com.ibm.cluster.gpfs.doc/gpfsbooks.html`.
[5] Hanushevsky A., Dorigo A. and Furano F., in *Interlaken 2004, Computing in high energy physics and nuclear physics*, pp. 680-683.
[6] LHC Computing Grid Project (online). Available: `http://lcg.web.cern.ch/LCG`.
[7] GridIt, INFN Production GRID (online). Available: `http://www.grid.it`.
[8] The gLite Project (online). Available: `http://glite.web.cern.ch/glite`.
[9] Baud J., Lemaitre S., Nicholson C., Smith D. and Stewart G., UK e-Science All Hands Conference, Nottingham, September 2005 (EPSRC) 2005, ISBN 1-904425-53-4.
[10] The Globus Alliance (online). Available: `http://www.globus.org`.
[11] BaBarGRID Web Monitor (online). Available: `http://webserver.infn.it/babargrid`.
[12] Fella A., Andreotti D. and Luppi E., *IFAE08, Incontri di Fisica delle Alte Energie, Bologna, Italy, 26-28 marzo 2008. Nuovo Cimento B*, **123** (2008) 944. DOI 10.1393/ncb/i2008-10570-2.
[13] Fella A., Andreotti D. and Luppi E., *EELA07, e-infrastructure shared between Europe and Latin America, December 2007*. ISBN 978-84-7834-565-6.