Письма в ЭЧАЯ. 2016. Т. 13, № 5(203). С. 1103–1107

JINR TIER-1 CENTRE FOR THE CMS EXPERIMENT AT LHC

N. S. Astakhov, A. S. Baginyan, S. D. Belov, A. G. Dolbilov, A. O. Golunov, I. N. Gorbunov, N. I. Gromova, I. S. Kadochnikov, I. A. Kashunin, V. V. Korenkov, V. V. Mitsyn, I. S. Pelevanyuk, S. V. Shmatov, T. A. Strizh[1], *E. A. Tikhonenko, V. V. Trofimov, N. N. Voitishin, V. E. Zhiltsov*

Joint Institute for Nuclear Research, Dubna

An overview of the JINR Tier-1 centre for the CMS experiment at the LHC is given. Special emphasis is placed on the main tasks and services of the CMS Tier-1 at JINR. In February 2015, the JINR CMS Tier-1 resources were increased to the level that was outlined in JINR's rollout plan: CPU 2400 cores (28800 HEP-Spec06), 2.4 PB disks, and 5.0 PB tapes. The first results of the Tier-1 operations received during the LHC Run 2 start are presented.

PACS: 07.05.-t; 07.05.Bx

INTRODUCTION

A computing model for the Large Hadron Collider (LHC) experiments (ATLAS, ALICE, CMS, and LHCb) data processing within the Worldwide LHC Computing Grid (WLCG) [1] was originally built up on the basis of the MONARC model [2]. It assumed a hierarchical, tiered model with a central data store (Tier-0) maintaining an archive copy of all data, and a second copy of the data which is being archived at one of a number of geographically distributed Tier-1s. A number of Tier-2s obtain data from the parent Tier-1 only for processing and analysis. This model has worked well for several years, though not very efficiently. CMS and ATLAS moved away from the MONARC model to a mesh model, where data are accessed directly from on-line copies with no need for Tier-2s to get their datasets from the Tier-1, i.e., now all Tier-1 and Tier-2 centers are connected with each other.

Without implementation of the grid infrastructure on the LHC it would be impossible to process and store enormous amount of data coming from the collider and therefore to make discoveries. The success of grid computing was highlighted by Prof. Rolf-Dieter Heuer, Director-General of CERN, on a festivity dedicated to receiving the Nobel Prize for the discovery of the Higgs boson. He avowed the grid technologies as one of three pillars of success: "It has only been possible because of the extraordinary performance of the accelerators, including the infrastructure, the experiments, and the grid computing".

Nowadays, every large-scale project will fail without using a distributed infrastructure for data processing.

[1]E-mail: strizh@jinr.ru

In November 2011, at the A. A. Fursenko's suggestion at a session of the Committee on Russia–CERN cooperation, it was decided to create a Tier-1 center for the LHC experiments in Russia. A proper infrastructure for the ATLAS, ALICE, and LHCb experiments has been launched on the basis of the NRC "Kurchatov Institute" (NRC KI, Moscow), while the Joint Institute for Nuclear Research (JINR, Dubna) hosted the CMS experiment [3].

A work schedule for Tier-1 in Russia was approved by the Supervisory Council of the WLCG project on September 28, 2012. At the first stage (up to December 2012), the prototype of Tier-1 at NRC KI and at JINR have been built up. The second stage (up to November 2013) included the equipment installation for the base Tier-1 center, its testing and finishing up to the required functional characteristics [4, 5]. Finally, the third stage (up to March 2015) assumed finalization of this complex and commissioning of a full-scale Tier-1 center for the CMS at JINR.

In the sequel, increase of computing facilities and data storage systems is required in accordance with the LHC experiments needs. 100% reliability and availability should be provided and it can be accomplished only by using a present-day engineering infrastructure (in particular, uninterruptible power system and climate control) and a high-speed reliable network infrastructure with a dedicated redundant channel to CERN (LHCOPN).

The Russian institutes and JINR as an international organization and its Member State institutes participate in the CMS experiment as part of the Russia and Dubna Member States (RDMS) Collaboration. The CMS Tier-1 computing center at JINR assists to perform the RDMS aims to carry out full-scale researches in the field of particle physics at the nominal values of LHC energies and luminosity.

To implement a complete range of Tier-1 tasks in full, it is also planned to develop and test new technologies for the large-volume data transfer and data processing.

CURRENT STATUS

In accordance with the CMS computing model [6], the CMS Tier-1 centers provide a wide range of high-reliability services for the whole CMS collaboration. They include common WLCG grid interfaces and additional CMS-specific services. It is necessary to assure a high availability of all Tier-1 services and a continuous technical support in a 7×24 operation mode. The JINR CMS Tier-1 performs the following functions:

1) data distribution including CMS RAW (primary) and pre-processed data from the Tier-0 (CERN), data import/export from/to any CMS Tier-1/2 centers;

2) data archiving and disk storage management to store data on tape for long-term safe storage and disks;

3) scheduled data-processing operations (RAW data processing, data reprocessing, Monte Carlo simulation and reprocessing);

4) authorization and authentication to control and prioritize the access to storage and computing resources (data handling and CPU usage);

5) monitoring the software services performance and the hardware components of Tier-1 (CPU farms, disk systems, tape robots, cooling and monitoring system, etc.);

6) alerting services to inform promptly the system administrators about the critical values.

By the end of 2015, the resources comprise: a computing power of 35.88 kHS06, 2400 cores/slots (8 Supermicro Twin blades), disk capacity of 1957.14 TB (30 Disk servers with hardware RAID6) and 521.88 TB used as a buffer for tape storage (8 Disk servers with

hardware RAID6), tape space of 5478.32 TB (IBM TS3500 tape library with $8 \times$ FC8 links to 8 Disk servers). The network is configured as two redundant triangles shared with the NRC KI with 10 Gbit/s LHCOPN connection.

The mass storage system is built on dCache [7] and Enstore [8] as a tape backend for dCache. Torque/Maui is used as a scheduler. The PhEDEx (Physics Experiment Data Export) service is used for the migration of files to mass storage and managing local mass storage stager pools. It also contains a transfer management database and a set of web monitoring tools. The current version installed is PhEDEx-4.1.3. The standard WLCG software stack was used for computing: $2 \times$ CREAM (Computing Resource Execution And Management) for job management operation at the Computing Element (CE) level, $4 \times$ ARGUS (centralized authorization service for distributed services), top-level BDII (Berkeley Database Information Index), site-level BDII, APEL (Accounting Processor for Event Logs) parsers, APEL publisher, EMI-UI (European Middleware Initiative-User Interface), $160 \times$ EMI-WN (Working Node) + gLExec-wn, $4 \times$ FTS3, LFC (LCG File Catalog), WMS (Workload Management System), L&B (Logging&Bookkeeping Service), glite-proxyrenewal.

The JINR CMS Tier-1 monitoring system provides real-time information about work nodes, disk servers, network equipment, uninterruptable power supply elements, and cooling

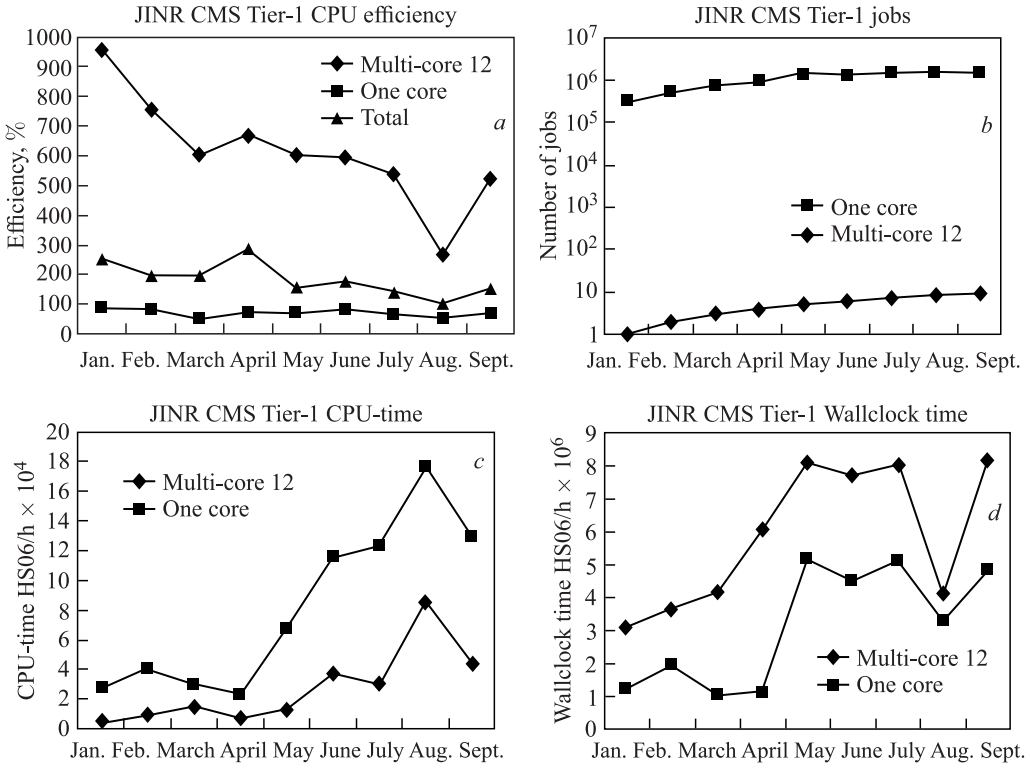

Fig. 1. Multi-core vs. one-core job processing during the last month: *a*) CPU efficiency for one-core jobs, 12-core jobs and total; *b*) number of 12-core and one-core jobs; *c*) CPU-time for two types of jobs; *d*) wallclock time for two types of jobs

system. It also can be used for creating network maps and network equipment load maps, for drawing state tables and different plots.

To improve the efficiency and power of the CMS computing infrastructure for operations during the LHC Run 2, there is a long-felt need to make several modifications and consolidations in the computing services. The most considerable improvements expected concern a transition to multi-core workflows. CMS expects to take advantage of the pilot system to facilitate the transition to a multi-core scheduling. If the pilot system consists of groups of processors or entire nodes, the system can schedule the resources as if they are batch nodes. It will create a mix of multi-core and single-core jobs to make the most efficient use of the available resources. A special queue for 12-core jobs for CMS was organized at the JINR CMS Tier-1. The last month statistics for multi-core and one-core jobs is shown in Fig. 1.

The Experiment Dashboard system [9, 10] is used for the JINR CMS Tier-1 site status and functionality control. It provides common solutions for monitoring job processing, data transfers, and site/service usability in the LHC experiments. For example, the Site Status Board is a monitoring tool that describes the status of sites based on metrics. The metrics are collected regularly and represent the number of jobs, number of successful transfers, status of

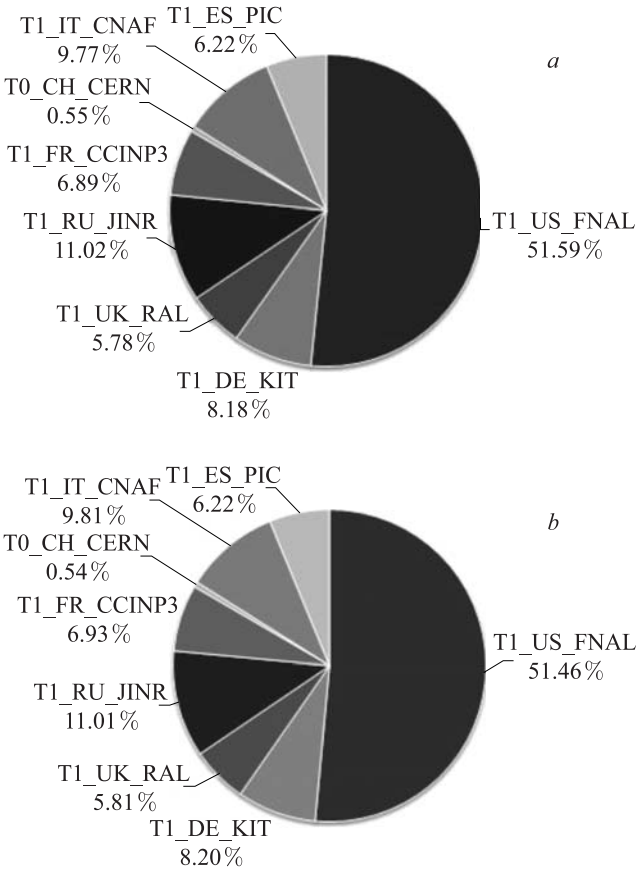

Fig. 2. Last month jobs over all CMS Tier-1 centres: *a*) completed jobs; *b*) submitted jobs

service, etc. (see daily metrics for the JINR CMS Tier-1 in [11]). The JINR CMS Tier-1 site is named as T1_RU_JINR in Dashboard systems and JINR-T1 in EGI accounting and WLCG monitoring systems.

The Historical Views application of the CMS Dashboard [12] offers the job statistics distributed over time and is used to check-up the site status at any customized period of time As example, diagrams presenting the number of jobs submitted and completed on CMS Tier-1 sites for last month are shown in Fig. 2.

CONCLUSIONS

The creation of the CMS Tier-1 center at JINR provides facilities for the RDMS CMS physicists as well as for the CMS Collaboration to realize a full-scale participation in processing and analysis of data of the CMS experiment at the Large Hadron Collider. The JINR CMS Tier-1 has shown its stable state for the whole period after putting it into the full-operation mode.

The invaluable experience of launching the Tier-1 center might be used for creating a system of storage and data processing for other projects in Russia and all over the world that require operations with big data flows.

REFERENCES

1. The Worldwide LHC Computing Grid (WLCG). http://wlcg.web.cern.ch/LCG.
2. MONARC — Models of Networked Analysis at Regional Centres for LHC Experiments. http://monarc.web.cern.ch/MONARC/.
3. *Berezhnaya A. et al.* LHC Grid Computing in Russia: Present and Future // J. Phys.: Conf. Ser. 2014. V. 513. P. 062041.
4. *Korenkov V. V. et al.* Creation at JINR of the Data Processing Automated System of the TIER-1 Level of the Experiment CMS LHC // Proc. of the 5th Intern. Conf. "Distributed Computing and Grid Technologies in Science and Education". Dubna, 2012. P. 254–265.
5. *Astakhov N. S. et al.* Tier-1-Level Computing System of Data Processing for the CMS Experiment at the Large Hardon Collider // Inform. Technol. Comp. Syst. 2013. No. 4. P. 27–36.
6. *Bonacorsia D. et al. (CMS Collab.).* The CMS Computing Model // Nucl. Phys. B. Proc. Suppl. 2007. V. 172. P. 53–56.
7. dCache. https://www.dcache.org.
8. Enstore. http://www-ccf.fnal.gov/enstore.
9. *Andreeva J. et al.* Experiment Dashboard — a Generic, Scalable Solution for Monitoring of the LHC Computing Activities, Distributed Sites and Services // J. Phys.: Conf. Ser. 2012. V. 396. P. 032093.
10. CMS Dashboard. http://dashboard.cern.ch/cms/
11. Site Status Board. http://dashb-ssb.cern.ch/dashboard/request.py/sitehistory?site=T1_RU_JINR#currentView=default.
12. Historical Views. http://dashb-cms-jobsmry.cern.ch/dashboard/request.py/dailysummary.