JINR GRID TIER-1@TIER-2

N.S. Astakhov[1], A.S. Baginyan[1], A.I. Balandin[1], S.D. Belov[1], A.G. Dolbilov[1], A.O. Golunov[1], N.I. Gromova[1], I.S. Kadochnikov[1], I.A. Kashunin[1], V.V. Korenkov[1], V.V. Mitsyn[1], I.S. Pelevanyuk[1], S.V. Shmatov[2], T.A. Strizh[1,a], V.V. Trofimov[1], N.N. Voitishin[1], V.E. Zhiltsov[1]

[1] *Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia*

[2] *Laboratory of High Energy Physics, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia*

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

The JINR grid infrastructure is the main component of the JINR Multifunctional Information and Computing Complex (MICC). There are two grid-sites: the Tier-1 for the CMS experiment at LHC and the Tier-2 which provides support to the virtual organizations (VOs) concerning the JINR participation in experiments at LHC (ATLAS, Alice, CMS, LHCb), FAIR (CBM, PANDA), and other VOs (NICA, STAR, COMPASS, NOvA) within large-scale international collaborations with JINR researchers. The grid resources of the MICC JINR are a part of the global Worldwide LHC Computing Grid (WLCG) infrastructure, which was formed to support the LHC experiments. Up to 2015 the main element of the JINR grid infrastructure was the Tier-2 center, one of the best resource centers of the Russian Data Intensive Grid (RDIG) and a part of the global grid infrastructure of WLCG and a member of the European EGI infrastructure. The official inauguration of the JINR Tier-1 for the CMS experiment in March 2015 marked a significant enhancement of the JINR grid computing infrastructure. This was an important contribution to the WLCG infrastructure. During the past two years it has been tuned and upgraded in order to cope with increasing amount of data coming from CMS experiment. The present status of the JINR grid infrastructure and plans for future development are presented.

Keywords: distributed computing, grid, Tier-1, Tier-2, WLCG, LHC

1. Introduction

The development of the LHC experiments data processing system started in parallel with the building of the accelerator and detectors. It was necessary to provide the experiments with the long-term storage petabytes of data and the facility to process and analyze the data. The research of Grid technologies in the Laboratory of Information Technologies (LIT) at JINR started in 2001 as a participation in the EU Data Grid international project. The goal of the project was the creation of the intermediate software to develop and test the functionality of the European Grid infrastructure. Since 2003 LIT JINR, in close collaboration with CERN, is actively participating in large international project WLCG (Worldwide LHC Computing Grid). In addition, JINR took part in EGEE (Enabling Grids for E-sciencE) and EGI (European Grid Infrastructure) projects. Also, JINR is the active member of RDIG (Russian Data Intensive Grid), which is a national federation, established within the EGEE project since September 2003. The most important result of this activity was the deployment of the Grid infrastructure providing the full spectrum of Grid services at JINR.

At present, the largest distributed computing infrastructure for science at LHC - Worldwide LHC Computing Grid [3, 4], is comprised of more than 170 computer centers in 42 countries. In the first period of LHC operation (Run1) this infrastructure became the essential element of the data processing and analysis at the LHC experiments. In Run2 the overall volume of raw, simulated and processed data in all experiments is expected to reach about 300 petabyte [5] which demands more powerful computing resources and data storage systems as well as evolution of the LHC computing model laid by WLCG project.

In the WLCG concept, all computer centers are combined into levels (Tiers) according to their functionality. Tier-0 level produces the raw data and provides initial reconstruction and distribution of the data over Tier-1 level which provides the raw and reconstructed data long-term storage, further processing, and the distribution of the data over Tier-2 level, which provides the data analysis.

Tier-0 includes CERN Data Centre and Wigner Research Centre for Physics in Budapest, Hungary, which are interconnected with three dedicated 100 Gbit/s links.

Tier-1 is comprised of fourteen large computer centers over the world with advanced resources and infrastructures to provide round-the-clock service. Tier-1 sites are interconnected and connected to Tier-0 with 2-3 10 Gbit/s or more links.

Tier-2 site is typically a university or a research lab computer center with enough power to analyze and simulate the data and communicate with Tier-1 sites at a speed of not less than 1 Gbit/s. There are about 160 such sites over the globe, making the Tier-2 level.

At JINR the Tier-2 site for all four LHC experiments (JINR-LCG2) has been operating at LIT since 2003, and is the best one in RDIG [6].

In March 2011, the Ministry of Science and Education of the Russian Federation proposed to establish the Tier-1 site for four (ALICE, ATLAS, CMS and LHCb) experiments in Russia. CERN supported this proposal. In frames of the Federal Target Program (FTP) of the Ministry of Science and Education of RF "Research and development in priority trends of science and technology in Russia in 2007-2013" the work had been started on the project "Development of the LHC experiments data processing system at Tier-1, and providing the Grid services for distributed processing of the data". In September 28, 2012 the plan of establishing Tier-1 in Russia was approved. The project implied a phased implementation. In December 2012, the Tier-1 prototype was deployed at NRC-KI and JINR. NRC-KI assumed the support of ATLAS, Alice and LHCb experiments, JINR – the support of CMS. In November 2013, the prototype was expanded to the basic Tier-1 level, the new hardware was tested and fine tuned. In March 2015, the full-scale Tier-1 for CMS experiment was officially inaugurated at JINR.

At present, based on the LIT Multifunctional Information and Computing Complex (MICC), Tier-1 for CMS and Tier-2 for all LHC experiments and other RDIG VOs, operate at JINR.

2. Tier-1 for CMS experiment at JINR

The Tier-1 site at LIT JINR (T1_RU_JINR in CMS mnemonics) [7,8] is providing the storage and processing of the CMS data, according to the computing model adopted by CMS collaboration [9]. The T1_RU_JINR is one of seven sites of such a scale in the world, involved in CMS data processing. In the current and the following LHC runs the volume of data is expected to be times as much as before. To be able to manage the increased data volume, keeping the budget under control, the experiment computing model was revised assuming the newest technology achievements [10]. The computing model was modified to provide a flexible and more optimal way of disk space utilization, using automatic removal of unpopular files. This system, however, has to keep the most used datasets locally to reduce the outer network traffic and disk space needs.

The JINR Tier-1 site infrastructure consists of several components: networking, services support system, data processing system, and data storage system. The combined operation of these components provides with the average reliability and availability at the level of 98% 24 hours a day, 7 days a week (the 24x7 mode), which corresponds to the best Tier-1 sites metric.

2.1. Networking

Local networking in Tier-1 is built as distributed data transfer over multichannel communication links using the software-defined networking (SDN) technology. SDN can follow a centralized, hierarchical, or decentralized design. A centralized solution assumes a single control entity having a global view of the network. In hierarchical solutions distributed controllers operate on a partitioned network view with a logically centralized root controller. In distributed approaches controllers operate on their local view or they may exchange synchronization messages to enhance their knowledge.

JINR Tier-1 network is of decentralized distributed design. A configuration file of the Virtual Machine Manager (VMM) fabric is located at every device. In each network segment a value is set which defines the procedure of how this segment gaining control of the network. In case of any loss of connectivity to a device, the overall throughput gets lower but the data processing network is not disrupted [11].

LHC Optical Private Network (LHCOPN) comprises a lot of 10 Gb/s communication channels connecting Tier-0 to Tier-1 sites. LHCOPN provides a very stable and reliable connectivity of high throughput.

In June 2017 IPv6 protocol was put into operation at LIT JINR: IPv6 connectivity was established, outer connectivity routing was deployed (Internet, LHCOne, LHCOPN), the support of IPv6 zones and records in LIT DNS servers was provided. Dual stack configuration allows IPv6 and IPv4 to be used in parallel in the JINR Tier-1 network and communication channels.

2.3. Data Processing System

Data processing system supports 248 64-bit work nodes (WNs) of 12 and 20 cores, which gives 4160 cores (slots) total.

2.4. Data Storage System

The data storage system operates with disks space and long-term tape robot storage. dCache and Xrootd are used to control the storage space. One of the dCache instances is working with the disk servers to provide with fast data access. The second one is dealing with disk servers and tape robot, with the disk space being the temporary buffer zone for tape space, while the tape robot provides with the long-term (permanent) CMS data storage space.

2.5. Tier-1 Resources and Usage

At present, Tier-1 includes 4160 core/slots for batch (typically SuperMicro Blade), disk only space 5.3PB (typically SuperMicro and DELL), total disk buffer space - 1.1 PB, and tape robot IBM TS3500 - 9PB.

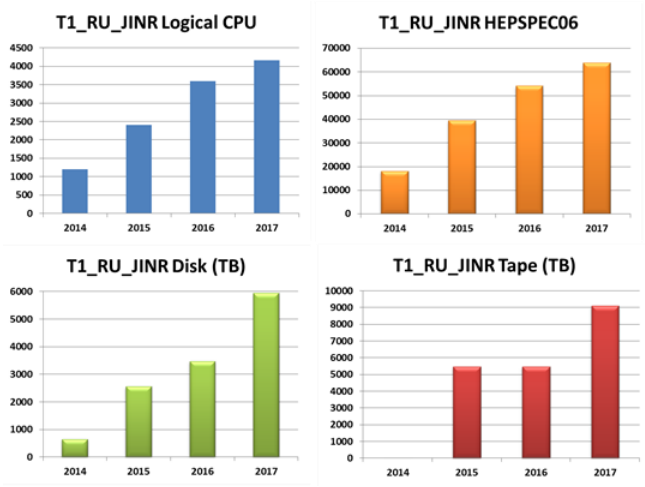

Figure 1. The growth of JINR Tier-1 resources in 2014-2017

Figure 1 shows the growth of the JINR Tier-1 resources in 2014-2017.

The contribution of the Tier–1 site at JINR to the processing of the CMS experiment data from March 2015 to September 2017 was 10.5%, while the greatest contribution to the data processing was provided by the FNAL (47.5%) site (Figure 2). Summing up, since its launch in 2015, the JINR site has executed 20 036 812 jobs [14]. Figure 3 shows the distribution of the events processed in between March 2015 and September 2017 over the Tier-1 sites.

One of the primary tasks of a Tier-1 site is the storing the non-processed and simulated experimental data. In Figure 3 and Figure 4 the weekly distribution of written and read files and their volume in TB in 2017 is presented. These plots show the intensity of the tape robot usage.

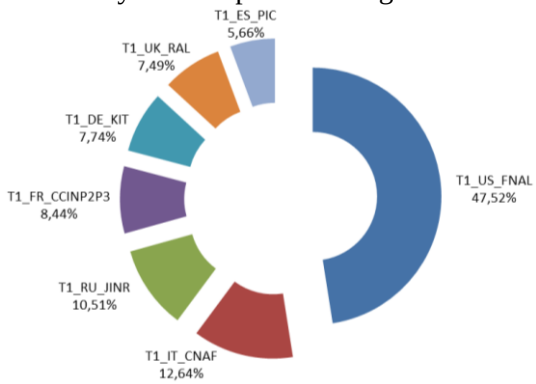

Figure 2. The distribution of the jobs processed between March 2015 and September 2017 over the Tier-1 sites. Total: 20 036 812 jobs. JINR part is 10,5%

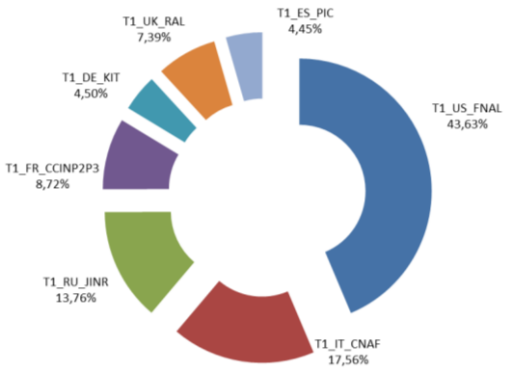

Figure 3. The distribution of the events processed between March 2015 and September 2017 over the Tier-1 sites. Total: 13 77 733 millions events. JINR part is 14%

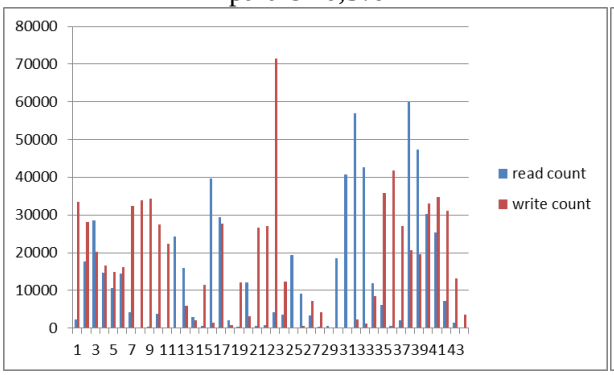

Figure 4. The number of files written and read by tape robot in 2017

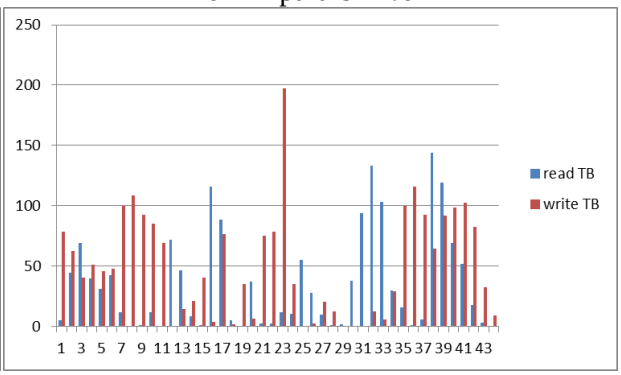

Figure 5. The volume of data (TB) written and read by tape robot in 2017

3. Status of Tier–2 (JINR-LCG2) site

The JINR Tier-2 site is operating as part of the cluster which provides the computing power and data storage to the JINR users and Virtual Organizations representing the four LHC collaborations and several collaborations of other physics experiments in Russia and abroad. Totally, for 12 years of JINR-LCG2 operation, 37 920 945 LHC jobs are executed (1 135 101 164 CPU Work HS06 Hours) which makes the site to be number 20 among 147 WLCG Tier-2 sites.

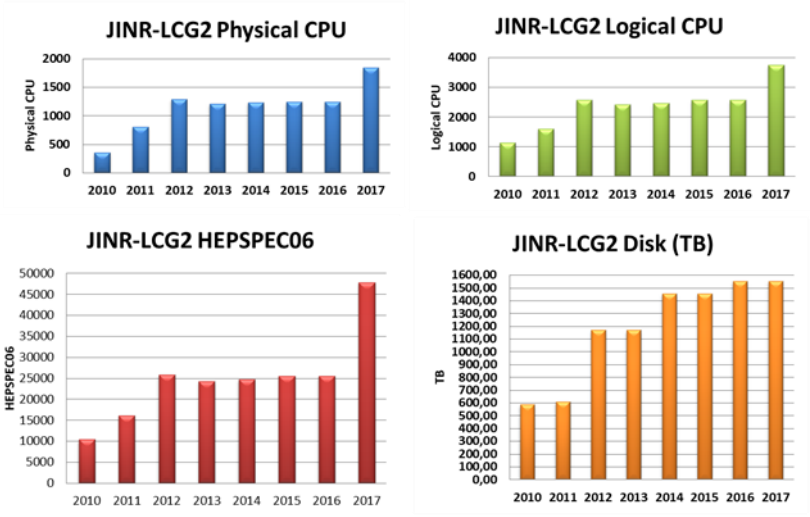

Figure 6. JINR Tier-2 resources growth in 2014-2017

Figure 6. shows the growth of the JINR Tier-2 in 2014-2017.

Currently, Tier-2 at JINR includes 279 computers (3640 cores) of 46866.52 HEP-SPEC06 performance. The primary users of the JINR grid-resources are Virtual Organizations of all LHC experiments. Computing resources are distributed over VOs as the following: LHC: Alice – 15%, ATLAS - 20%, CMS - 20%, LHCb 15%. The data storage system of 1909.8 TB is distributed over VOs this way: LHC: Alice @ xrootd&EOS – 31% (587.45 TB), ATLAS @ dcache – 34,6% (660.33 TB), CMS @ dcache – 34,7% (662.02 TB).

We are working on the integration of OSG type Computing Element - HT-CONDOR in our Tier-2 infrastructure. At the moment it works for STAR VO mainly, but can be extended to support other VOs in the future.

Table 1. The usage of the Tier-2 resources by LHC VOs

	2016	**2017**	**total**	**%**
alice	53416446	47385875	100802322	25.77%
atlas	65241643	60981847	126223490	32.27%
cms	64017857	37078294	101096151	25.84%
lhcb	28753595	34300972	63054567	16.12%
Total	**211429541**	**179746988**	**391176530**	

Table 1 shows the usage of the Tier-2 resources by LHC VOs during 2016 and 9 months of 2017 in Normalized Elapsed time (HEPSPEC06) * Number of Processors (hours) by LHC VO.

Besides operating as a Tier-2 site the JINR computer cluster supports the non-Grid computing power including parallel computing, which is required by such experiments as NOvA, PANDA, BESIII, NICA/MPD and others, as well as local users. Figure 7 shows such cluster resources usage during 9 months of 2017.

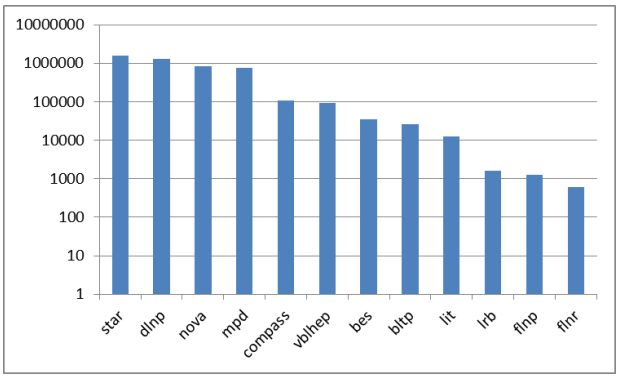

Figure 7. The cluster resources usage by local (non-Grid) users during 9 months of 2017 in CPUclock Normalized to 1000 Specint2000

4. Monitoring system

To effectively keep track of the whole infrastructure functionality, the monitoring system based on Nagios has been developed [12].

The system allows one to trace and control power supply and air conditioning systems, computing and data storage systems, including information on processors load, number of active and waiting jobs, network load. The operation center of the JINR Multifunctional Information and Computing Complex (MICC) is deployed and currently exploited [13].

In addition to the global WLCG monitoring in CMS [14], a local Tier-1 services monitoring system is needed to have a detailed picture of the current local status of the services at any time [15].

5. Conclusion

Since 2001, the LIT JINR are actively involved in study, utilization, and development of the Grid technologies. An important result of this activity is the establishing and reliable operation of the Grid sites at JINR, which are fully integrated into the WLCG infrastructure of the LHC data processing, analysis and storage. Two Grid sites are successfully operating at JINR: since 2004 - full-scale Tier-2 site (JINR-LCG2) for ATLAS, Alice, CMS, LHCb, and since 2015 — full-scale Tier-1 site (JINR-T1/T1_RU_JINR) for CMS.

Importance of the Tier-1 center at JINR is:

- to ensure full-scale participation in CMS data processing and analysis for the JINR physicists, JINR Member States, and whole RDMS CMS Collaboration;
- the invaluable experience of launching the Tier-1 center will be used to create a system of data storage and processing for the NICA megaproject and other large scale projects of the JINR member countries.

Successive upgrades of the LHC accelerator and the experiments facilities are resulting in the data volume increasing times as much. It is just impossible for any single computer center to store and process such volumes locally. All four LHC collaborations require still more and more data processing and storage power (both disk and tape) at all WLCG sites. According to these requirements, the development of the JINR Grid infrastructure is being planned within the RDMS CMS Collaboration framework. Table 2 shows a planned growth of the Tier-1 resources in 2018-2019 in the absolute values with a percentage increase compared to the previous year.

Table 2. Planned growth of the Tier-1 resources in 2018-2019

	2018	**2019**
Processor power of the core /kHS06	5200/83,2 (23%)	10000/160 (52%)
Disc storages (TB)	6100 (20%)	8000 (80%)
Tape storages (TB)	10000 (0%)	20000 (100%)

Upgrade of the outdated compute element (CE) and of the data storage element (SE) as well as new additions according to the existing schedule of the JINR Tier-2 site is foreseen. This site supports LHC VOs (ATLAS, Alice, CMS, LHCb), and also FAIR (CBM, PANDA), and other VOs within large-scale international collaborations, as well as non-Grid computing tasks of local JINR users. Table 3 shows a planned growth of the resources of CICC/Tier-2 per years in the absolute values and in a percentage growth compared to the previous year.

Table 3. Planned growth of the CICC/Tier-2 resources in 2018-2019

	2018	**2019**
Processor power of computing core / kHS06	4700/75,2 (27%)	6000/96,0 (74%)
Disc storages (TB)	3400 (14%)	5000 (26%)

References

[1] Belov S.D. et al.: Grid in JINR and participation in the WLCG project, in Proc. of GRID'2012 conference, Dubna, 2012, p.23.

[2] Klimentov A., Korenkov V.: The distributed computing systems and their role in discovering a new particle, Supercomputers, No.3 (11), 2012 p. 54 (in Russian).

[3] LHC Computing Grid Technical Design Report, CERN-LHCC-2005-024, 20 June 2005.

[4] Bird I.: Computing for the Large Hadron Collider, Annual Review of Nuclear and Particle Science, Vol. 61: 99-118, November 2011.

[5] Update of the Computing Models of the WLCG and the LHC Experiments, LCG-TDR-002: https://cds.cern.ch/record/1695401/files/LCG-TDR-002.pdf

[6] Gavrilov V., Golutvin I., Kodolova O. et al.: Status of RDMS CMS computing Particles and Nuclei Letters, v.13,no 5, pp.1108-1111, 2016.

[7] Astakhov N.S., Baginyan A.S., Belov S.D. et al.: JINR Tier-1 centre for the CMS Experiment at LHC. Particles and Nuclei, Letters, v.13,no 5, pp.1103-1107, 2016.

[8] Astakhov N.S., Baginyan A.S., Belov S.D. et al.: Tier-1 CMS at JINR: Status and Perspective CEUR Workshop Proceedings (CEUR-WS.org) http://ceur-ws.org/Vol-1787/1-14-paper-1.pdf

[9] Grandi C., Stickland D., Taylor L. et al.: The CMS Computing Model Preprint CERN-LHCC-2004-035/G-083.

[10] Grandi C., Bonacorsi D., Colling D., Fisk I., Girone M.: CMS computing model evolution, Journal of Physics: Conference Series 513 (2014) 0320399.

[11] Baginyan A.S., Dolbilov A.G., Korenkov V.V., Network for data-center Tier 1 at JINR for experiment CMS (LHC), T-Comm. 2016. Vol. 10. No.1, pp. 25-29 (in Russian).

[12] Kashunin I., Dolbilov A., Golunov A., Korenkov V., Mitsyn V., Strizh T.: The monitoring system of Multifunctional Information and Computing Complex CEUR Workshop Proceedings (CEUR-WS.org), http://ceur-ws.org/Vol-1787/256-263-paper-43.pdf (in Russian).

[13] Golunov A., Dolbilov A., Kadochnikov I., Kashunin I., Korenkov V., Mitsyn V., Pelevanyk I., Strizh T.: Operation center of the JINR Multifunctional information and computing complex) CEUR Workshop Proceedings (CEUR-WS.org), http://ceur-ws.org/Vol-1787/235-240-paper-39.pdf (in Russian).

[14] CMS Dashboard: http://dashboard.cern.ch/cms/

[15] Kadochnikov I., Pelevanyuk I.: JINR Tier-1 service monitoring system: Ideas and Design CEUR Workshop Proceedings (CEUR-WS.org), http://ceur-ws.org/Vol-1787/275-278-paper-46.pdf