JINR Central Information and Computing Complex: Current Status and Perspectives

V.V. Ivanov, V.V. Korenkov, V.V. Mitsyn, N.S. Astakhov, E.A. Tikhonenko, V.V. Trofimov
Laboratory of Information Technologies, JINR

The JINR Central Information and Computing Complex (CICC) at the Laboratory of Information Technologies (LIT) has been substantially modernized and developed during 2006-2007 years. After putting into operation a new supercomputing cluster during this year, the CICC took the 12^{th} rank in the rating of the most high-performance computing systems in Russia and CIS [1,2].

At present, the CICC performance equals 670 kSI2K[1] and the disk storage capacity – 57 TB (47 TB – for user catalogs, software and large data volumes storage). Thus, in the last 2 years the CPU and disk storage resources have been increased 6.7 and 3 times, accordingly. Operating systems are Scientific Linux 4 (SL4) and Scientific Linux 3 (SL3).

CICC logical structure can be considered as a very important achievement. The former CICC structure represented a number of special-purpose clusters allowing two ways of each cluster resources usage: either local or global [3]. Since June 2007, the CICC resources and services have been integrated into a unified information and computing structure (Fig.1). In May 2007, the CICC logical and software structure were substantially restructured. The main changes concerned the security and the batch systems. All the CICC computing and data storage resources can now be used both locally and globally (for distributed computations in the LCG/EGEE grid infrastructure) for all the projects the JINR physicists participate in. The system software has been tuned in an optimal way, providing maximal use of computing resources and the most universal and secure access to the data storage. The Torque batch system and the Maui scheduler are used for computing resources allocation and accounting.

Basically, access to data is provided by the dCache system and partially via NFS. The access to the general-purpose software and user home catalogs is provided by the Andrew File System (AFS). The Kerberos5 system is used for registration and authentication of local users instead of KerberosIV. The main reason is the insufficient security of the KerberosIV protocols with regards to the possibilities of decoding closed data on modern fast CPU. The transition to new protocols asked for essential changes in the turning of all the CICC computers. As the batch system essentially depends on the user authorization system, we had to add to the standard Torque system codes for work with new protocols. Completion has been done at the level of initial codes of the Torque system.

While integrating the support of the Kerberos5 and AFS systems into the Torque batch system, we faced huge difficulties. Only the proper tuning of the Maui scheduler comprising 500 adjustable parameters took us about ten days to test it in a full-scale volume.

A new configuration of the CICC computing resources has been entered also. Instead of the resource separation into a farm for WLCG and a farm for local use, all computing resources became general. This has demanded significant changes in the LCG standard software - gLite, because gLite standard foresees the allocation of resources only under

[1] SI2000 (SI2K) units are usually used for computing performance evaluation (in accordance with the Spec Integer 2000 special test).

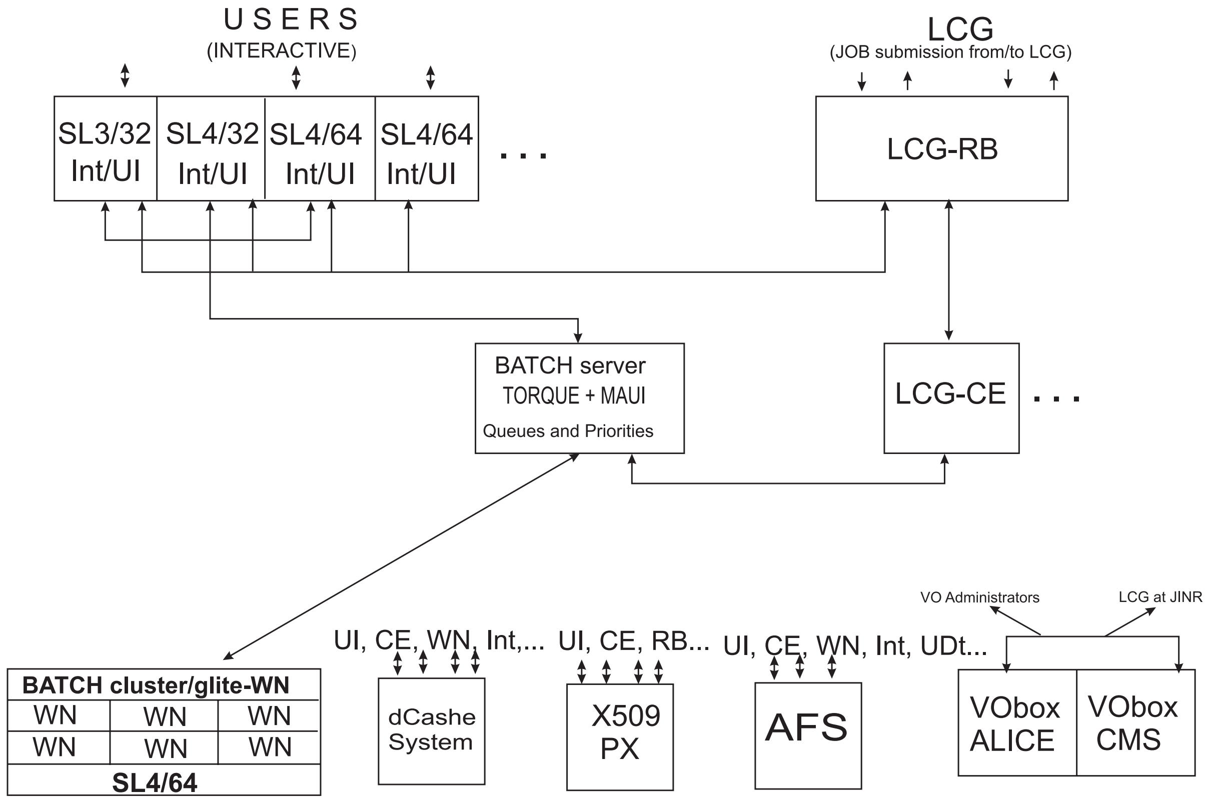

Fig. 1: CICC computing resources, access and maintenance scheme. SL3/32 Int/UI - Interactive nodes/User Interface at 32-bit architecture with SL3, SL4/32 Int/UI - Interactive nodes/User Interface at 32-bit architecture with SL4, SL4/64 Int/UI - Interactive nodes/User Interface at 64-bit architecture with SL4, LCG-RB - LCG Resource Broker, LCG-CE - LCG Computing Elements, WN - Worker Nodes, X509 PX - Proxy, VObox - special node where experiments (ALICE, CMS, etc.) or Virtual organizations (VO) can run specific agents and services to provide a reliable mechanism to accomplish various tasks specific for VO, AFS - AFS servers, dCashe - dCashe servers

LCG/EGEE needs. On the other hand, such association of resources demands special adjustment of numerous parameters of the batch system, that in itself is a long process. Also the integration of the gLite global environment and the CICC local infrastructure required sizeable efforts.

Currently there are the following computing nodes at the CICC:

- 53 special-purpose servers (29 for main the CCIC and JINR information infrastructure support and 24 for the JINR-LCG2 site support in RDIG, WLCG and EGEE);
- 4 interactive nodes;
- 60 computers of 64-bit architecture (2 Xeon 5150 dual-core processors, 8 GB RAM, 160 GB disk, 1 Gb Ethernet);
- 5 computers of 32-bit architecture (2 Athlon 2400+ processors, 2GB RAM, 40 GB disk, 1Gb Ethernet, Myrinet network adapter).

As the dual-core processors are in fact two independent processors, we have 240 CPUs of 64-bit architectures forming the basic part of the CICC supercomputing cluster. Ten CPUs of 32-bit architecture are also in use.

The JINR users can submit their jobs to all these 250 CPUs both locally and globally (in the LCG/EGEE environment). The internal CICC network for the most part make use Gigabit Ethernet. All the basic computing and data storage resources are connected to internal network routers at the rate of 1Gb/sec. The internal CICC network is connected to the JINR backbone network and the border router at the rate of 1 Gb/sec too.

There are 4 computers at the CICC for interactive user's work: 2 computers of 64 bit-architecture (Core 2 Duo, 4 GB RAM, 250 GB HDD, 1 Gb Ethernet) and 2 computers of 32-bit archirecture (Dual Athlon 2400+, 2 GB RAM, 80 GB HDD, 1 Gb Ethernet).

The special-purpose servers are used to support the work of users and JINR services: batch, WWW, MySQL and Oracle DBs, e-mail, DNS and some others. These servers work at different hardware platforms – from 32-bit Pentium-3 up to 64-bit Core 2 Duo.

The AFS distributed file system with a highly secure data access is used at JINR for user home catalogs and general-purpose software. The total AFS space is about 1.45 TB (1.40 TB – at CICC servers). There are 8 AFS servers installed and supported at the CICC.

Information stored at the CICC disk servers can be subdivided into 2 categories:

- user home catalogs and software products in different forms;
- "physical" information, i.e. data obtained at physical experiments and simulated data samples stored in the dCache system.

The dCashe system has finally replaced the Castor system. Castor servers are put out of operation, so that access to data in that format is not supported at JINR. We have chosen the dCache system instead the Castor, because dCashe responds the best to the requirements which are demanded of the JINR as part of distributed Ru-Tier2 center in the WLCG (Worldwide LHC Computing Grid) hierarchy. dCache hardware and software complex serves to store large data volumes. It includes one server for hardware control, 5 front-end servers for system monitoring and a work with data and 14 data storage systems (back-end or pool). Athlon 2000+, Pentium 4, Pentium D and Xeon hardware platforms are in use. All storage systems are constructed with the usage of RAID5 hardware mechanism. Internal sub-connections are implemented at the 1Gb Ethernet technology. All the front-end systems are also connected to the 1 Gb Ethernet. Now the volume of dCache data storage pools is about 43 TB, and it is planned to increase it up to 90 TB in the nearest future. The dCache system has been continuously tested during 2006-2007 years, and we continue the testing to check the JINR dCache system readiness to receive physical data from the LHC experiments after the LHC start. The tests showed that single files (1-2 GB size) can be transferred via grid utilities to the JINR dCache system from CERN or other scientific center with a rate of 2-4 MBps. The rates for single files do not essentially decrease, if several files are transferred in parallel, but a total transfer rate up to 30 MBps can be achieved for all files, transferred in parallel. During the last six months, in between 1 to 1.5 TB test data were successfully transferred to the JINR dCache daily. As an example, the averaged CERN-JINR connection throughput during CMS test transfers from 29.09.07 to 12.10.07 is shown at Fig.2.

Just as the AFS system, dCache provides local and global access to files which are logically combined into one tree, but physically are located on different disks. In the local Linux environment at JINR it is possible to mount these trees and transparently manipulate the files, but without access to the file's content. In contrast to the AFS

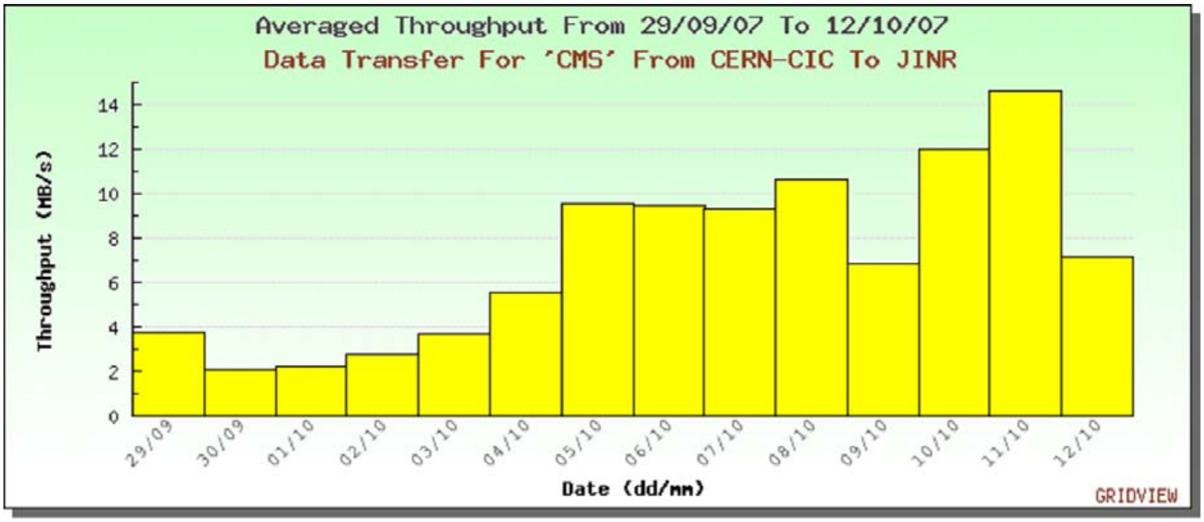

Fig. 2: Averaged JINR-CERN network connection throughput during CMS test transfers

system, dCache uses its own data access protocol named DCAP or gridFTP-protocol for the access to files.

The higher level tools – FTS, SRM and Xrootd – are used in a global grid environment. dCache is oriented to the work with WORM (Write Once, Read Many) data. To work with a file, usually it is necessary to copy it to a local file space. Also it is possible to work with files from applications both statically or dynamically with the usage of libdcap library. Unfortunately, the dCache 1.7.0.39 version does not support the mechanism of quotas, but it is expected that a new version supporting quotas to be released in the nearest time. The guide on the dCache usage at JINR and complete current information for JINR users can be found at the CICC web-site [4]. The dCache system is considered to be the only system at the JINR for storage of large data volumes. CICC resources

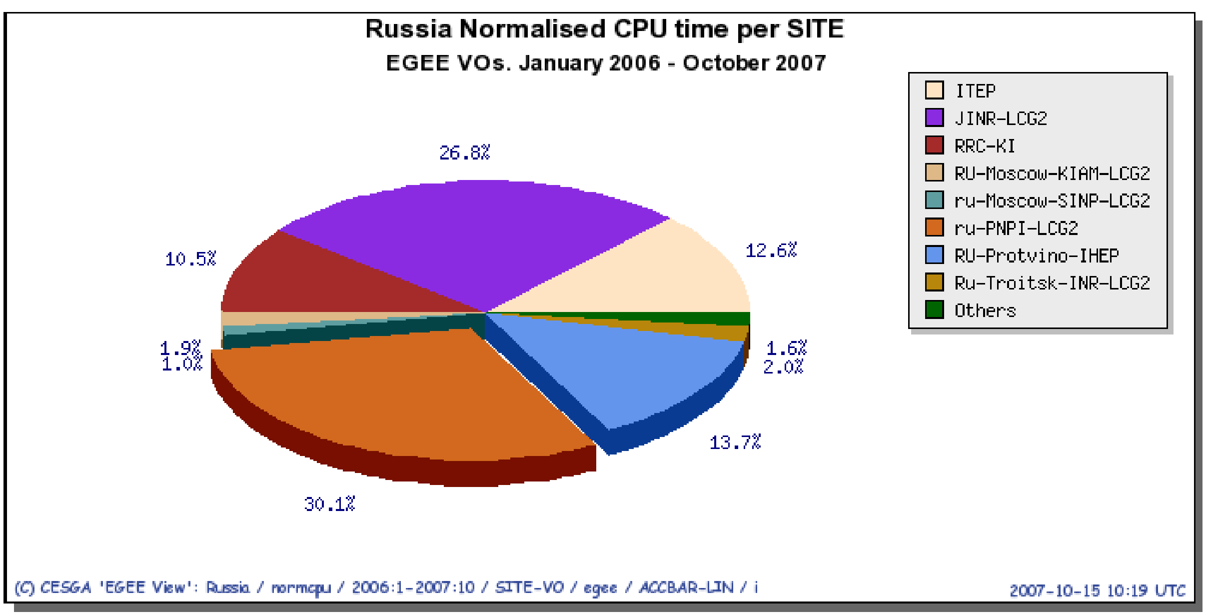

Fig. 3: Normalized CPU time per site in the RDIG infrastructure (Jan.2006-Oct.2007)

included into the EGEE/LCG and RDIG (Russian Data Intensive Grid) infrastructures

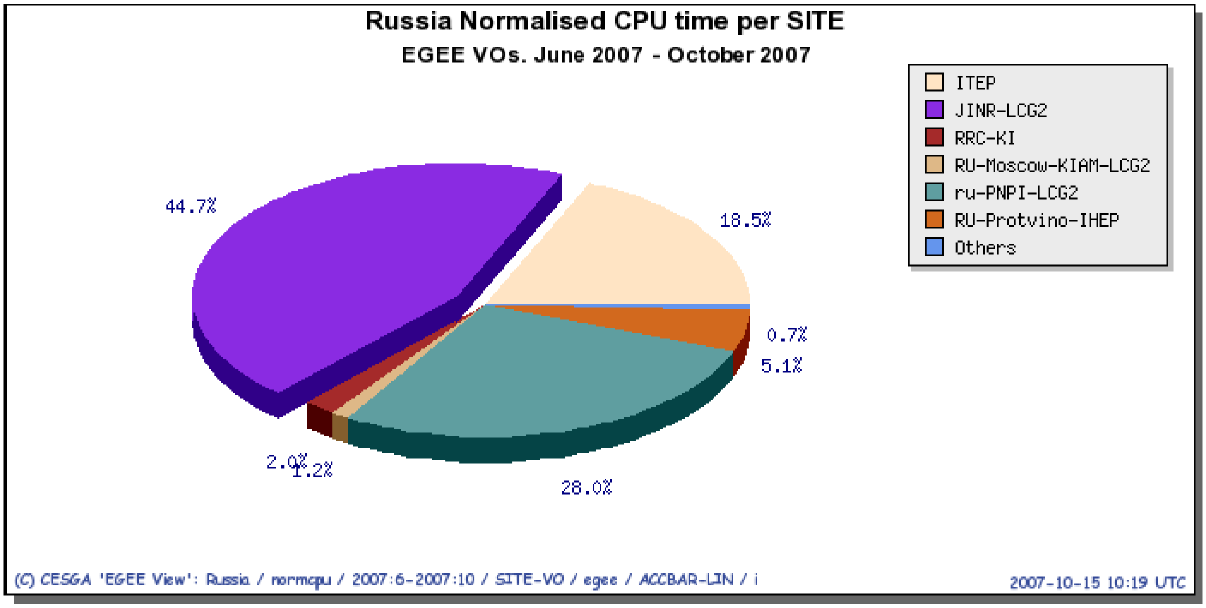

Fig. 4: Normalized CPU time per site in the RDIG infrastructure (June 2007-Oct.2007)

have been reliably and successfully used during 2006-2007 years, and its percentage of usage in the RDIG infrastructure is about 27% (January 2006-October 2007), and after putting into operation of the new supercomputing cluster (from June, 2007) - about 45% (see Fig.3 and Fig.4).

Table 1:

Year	2007	2008	2009	2010
CPU (kSI2K)	1000	1250	1750	2500
Disk (TB)	150-200	400	800	1200
Tapes (TB)	-	-	100	200

By the end of 2007 year, we plan to increase the JINR CICC CPU performance up to 1000 kSI2K and disk storage – up to 100 TB. Further extension, founded in accordance with the JINR scientific program, is presented in Table 1.

To estimate the performance of the JINR supercomputing cluster, testing on the High Performance LINPACK (HPL) benchmark has been accomplished [2].

References

[1] http://www.supercomputers.ru/?page=rating

[2] V.V.Korenkov, V.V.Mitsyn, E.B.Dushanov, A.S.Airyan, A.I.Lutsenko, Testing of the new supercomputing cluster of the JINR CICC, in this volume, pp. 43-46.

[3] N.S.Astakhov, V.V.Ivanov, V.V.Korenkov, V.V.Mitsyn, A.P.Sapozhnikov, E.A.Tikhonenko, The JINR CICC resources and services: current status and perspectives, LIT JINR Scientific Report 2004-2005, Dubna, 2005, pp.33-36 (in Russian).

[4] http://lit.jinr.ru (Computing&Information Resources →CICC→ News)