ISSN 1063-7788, Physics of Atomic Nuclei, 2020, Vol. 83, No. 6, pp. 965–968. © Pleiades Publishing, Ltd., 2020.
Russian Text © The Author(s), 2020, published in Yadernaya Fizika, 2020, Vol. 83, No. 6, pp. 534–538.

Trends and Prospects of the Development of Distributed Computing and Big Data Analytics for Support of Megascience Projects

V. V. Korenkov[*]

Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna, 141980 Russia

Received April 26, 2020; revised April 26, 2020; accepted April 26, 2020

Abstract—The creation and development of computer systems for experimental data processing and for data storage and analysis, of search algorithms, and of means for providing access to data are crucial for megascience projects. Information and computing infrastructures necessary for carrying out research tasks of megascience projects are complex distributed heterogeneous systems including systems of extra-massive parallelism and systems of distributed storage of big data arrays.

DOI: 10.1134/S1063778820050154

Russian research institutes and universities participate actively in international megaprojects, such as the ATLAS, ALICE, LHCb, and CMS experiments at the Large Hadron Collider (LHC) at CERN, the European hard X-ray Free Electron Laser (XFEL) at Deutsches Elektronen-Synchrotron (DESY) in Hamburg (Germany), the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), the CBM and PANDA experiments performed at the Facility for Antiproton and Ion Research (FAIR) at GSI Helmholtzzentrum für Schwerionenfoschung in Darmstadt (Germany), and the International Thermonuclear Experimental Reactor (ITER) in Cadarache (France). The megascience projects at the stage of preparation in Russia include the Nuclotron-based Ion Collider fAcility (NICA) at the Joint Institute for Nuclear Research (JINR, Dubna), the PIK high-flux research nuclear reactor at the Petersburg Nuclear Physics Institute (PNPI, Gatchina), and the Siberian Circular Photon Source (SCPS) at the Budker Institute of Nuclear Physics (BINP, Siberian Branch, Russian Academy of Sciences, Novosibirsk). The implementation of the neutrino research program is currently underway via the Baikal-GVD (Russia), JUNO (China), NOvA and DUNE (USA), and other large-scale projects.

For the purposes of processing, storage, and analysis of data from the experiments at the LHC, a distributed infrastructure based on the grid technology and called the Worldwide LHC Computing Grid (WLCG) was created at CERN [1]. At the present time, the WLCG includes about 1 000 000 CPU cores, about 0.6 exabyte of disk storage, and about 0.8 exabyte of tape robot storage. This provides a long-term storage of data geographically distributed among 170 data processing centers in 42 countries. This system processes more than two million tasks per day and runs hundreds of petabytes of data. The WLCG infrastructure was one the factors that contributed to the success of the first LHC stage, which culminated in the discovery of the Higgs boson.

The experiments at the LHC play a key role in research studies not only in the realms of particle and nuclear physics but also in the realms of big data analytics. Over the past years, the model of computing on the basis of grid technologies for the LHC has undergone some modifications that make it possible to meet the demands of the scientific community to a greater extent. There occurred a transition from the strictly hierarchic processing model [2], where the whole process of data acquisition and processing was distributed among specifically tiered computing centers—Tier0 is the main CERN Data Centre, which collects raw data from all experimental facilities, performs the first step in reconstructing meaningful information on the basis of the raw data, and distributes the raw data and the reconstructed output among 14 computer center of Tier1 for implementing their long-term storage, reprocessing, and analysis and for providing access to these data for Tier2 intended for a simulation and data analysis performed by end users, each of the Tier1 centers being connected with specific Tier2 centers—to a new model where the Tier1 and Tier2 centers are coupled. Moreover, data processing and analysis are being presently performed with resorting to high-performance complexes; academic, national,

[*] E-mail: korenkov@jinr.ru

and commercial resources of cloud computing; supercomputers; and other means [3].

Russian research centers, first of all, National Research Center Kurchatov Institute and JINR, take active part in the integration of distributed heterogeneous resources and in the development of big data technologies for providing support to modern megaprojects in rapidly developing fields of science, such as high energy physics, astrophysics, and bioinformatics. Work on the construction of the unique accelerator complex NICA [4], which requires new approaches to implementing a distributed infrastructure for experimental data processing and analysis is being vigorously performed at JINR.

It is noteworthy that, initially, grid technologies relied on the HTC (high-throughput computing), but the evolution of the computing model resulted in combining various technologies, including HTC, HPC (high-performance computing), volunteer computing, and commercial and noncommercial cloud computing resources. This approach is necessary for meeting the requirements of megascience experiments both in the data processing rate and in the data storage volume. Moreover, it is necessary to introduce further modifications in the computing model for each experiment in order to optimize the use of resources. Today, a great deal of effort is going into the development of software aimed at improving a general performance in employing the state-of-the-art architectures (the use of many cores, graphical processors, and so on). It is also necessary to optimize processing processes, storage systems, and amount of the data copies stored.

An intermediate connectivity software (platform) that permits performing joint work in information and computing systems is a key point in organizing such infrastructures—in particular, WLCG. In the ATLAS experiment at the LHC, for example, the PanDA (Production and Distributed Analysis) Workload Management System (WMS) platform was developed for running computing resources [5]. It is an automated and tunable system for controlling jobs and optimizes access of users to distributed resources. By means of PanDA, the users see a unified computing resource intended for experimental data processing, even though the resource centers are scattered worldwide. PanDA isolates physicists from the hardware, system software and middleware, and other technological difficulties associated with network and hardware setup. Computing problems are automatically traced and fulfilled. At the present time, the PanDA controls hundreds of computing centers in 50 countries worldwide, hundreds of thousands of computing nodes, hundreds of millions of jobs per year, and thousands of users.

The DIRAC (Distributed Infrastructure with Remote Agent Control) Interware [6] is yet another version of the connectivity middleware; this is software for the integration of heterogenous computing resources and systems for data storage into a unified infrastructure—the so-called interware platform. The integration of the resources is based on the use of standard data access protocols (such as xRootD and GridFTP) and pilot jobs. This provides the users with an abstract interface where they can launch jobs, control data, align processes, and monitor their fulfillment. Batch-processing systems, grid-sites, clouds, supercomputers, and even individual computing nodes may become computing resources within the DIRAC platform. Pilot jobs are an important concept in the DIRAC Interware. It is owing to them that almost any computing resource becomes integrable. In dealing with data, the DIRAC provides the whole set of necessary instructions. For all of the instructions to function correctly, the storage system should support grid-protocols of data transfer.

At the present time, much attention is being given to new promising directions in the creation of distributed data storage systems (DataLake) [7]. This permits combining a substantially improved efficiency of big data storage with a high rate of access to the data.

The Laboratory of Information Technologies at JINR plays an important role in the development of computing for megascience projects, and its main task is to develop networking information and computing infrastructure of JINR for research and production activities of this institute [8]. The Multifunctional Information and Computing Complex (MICC) of JINR is being actively developed [9]. It meets the requirements for a modern highly efficient scientific and computing complex: multifunctionality, high performance, multiply tiered data storage system, high reliability and accessibility, data security, scalability, individual software environment for various groups of users, high-performance telecommunications, and state-of-the-art local computing network. Currently, MICC of JINR has the following basic component:

(i) the Central Information and Computing Complex (CICC) of JINR with in-house-built computer and mass storage elements,

(ii) the Tier2 cluster for all experiments at the LHC and other virtual organizations (VOs) in the grid environment [10] (4128 cores; the total useful volume of disk servers is 2.929 petabytes),

(iii) the Tier1 cluster for the CMS experiment [11] (10 688 cores; the useful volume is 10.4 petabytes, and that of tape robots is 51 petabytes),

(iv) the HybriLIT heterogeneous platform for high-performance computing (HPC) on the basis of the GOVORUN supercomputer [12] (the total peak supercomputer capacity is 860 teraflops for double-precision operations),

(v) the cloud infrastructure [13] (1564 cores),

(vi) The data storage system of the basis of EOS file system (3.740 petabyte of disk space).

In recent years, attention in the realms of mega-science projects has been given primarily to developing the telecommunication and network infrastructure, including the upgrade of local computing networks with the aim of providing resources for data storage and processing.

Owing to the exploitation features and data storage systems of the CMS Tier1 site of JINR, which is a basic grid component of MICC, it is currently ranked second in the world among the other CMS Tier1 sites in the number of processed events.

The Tier2 site of JINR has also been actively developed. It allows data processing for all four LHC experiments (ALICE, ATLAS, CMS, and LHCb) and lends support to a number of VOs beyond the LHC (BESIII, BIOMED, COMPASS, MPD, NOvA, STAR, and ILC). The MICC also provides computational resources for off-grid-environment calculations. This is of special importance for such experiments as NOvA, PANDA, BESIII, and NICA/MPD/BM@N and for local users from all JINR laboratories.

The cloud infrastructure of JINR is yet another MICC component. The cloud infrastructures of the JINR member states were integrated within this infrastructure.

The HybriLIT heterogeneous computing platform is an important MICC component. It consists of an education and testing polygon and the GOVORUN supercomputer, which jointly employ unified software and information environment. The GOVORUN supercomputer is aimed at performing resource-intensive and massive parallel calcul ations required in tackling a wide range of problems that JINR scientific groups have to solve. This becomes possible owing to the heterogeneity (presence of computing accelerators of various types) hardware architecture of the supercomputer.

In order to extend the potential for developing mathematical models and algorithms and for performing resource-intensive calculations, including those that involve graphical computing accelerators, which reduce substantially machine time, an ecosystem for the problems of computer-assisted/deep learning and data analysis was created for HybriLIT users and is being actively developed.

The MICC projects turned out to be a successful integration of all computing and infrastructure resources. It provides a reliable and duly organized computing environment for performing research studies by physicists from JINR and its member states. The availability of cutting-edge computing resources, such as the GOVORUN supercomputer and the CMS Tier1 site contributes to a substantially broader recognizability of JINR all over the world.

The developed multilevel monitoring system [14] of MICC makes it possible to obtain information from different components of this computing complex: engineering infrastructure, network, computing nodes, task management systems, data storage elements, and grid services. This guarantees a high level of reliability of MICC.

A heterogeneous distributed information and computing cluster is being created for the NICA mega-science project. This will make it possible to meet to the highest possible degree the requirements of the participants of the project both in the field of theoretical investigations and in the field of processing, storage, and analysis of experimental data from the BM@N, MPD, and SPD detectors. The basic configuration of the distributed information and computing cluster of the NICA complex is expected to ensure the processing and storage of up to 10 petabyte of data per year. The complex consists of territorially distributed on-line and off-line clusters connected with one another by a high-speed computing network boasting a data-transfer rate of up to 4×100 Gb/s.

The developed computing models should take into account trends in evolving network solutions, computing architectures, and IT solutions making it possible to integrate supercomputer (heterogeneous), grid, and cloud technologies and to create, on this basis, distributed software-configurable HTC and HPC platforms. A distributed scalable hybrid cluster was created in order to ensure support of experiments at the NICA accelerator complex. It can be readily reconfigured in order to meet the needs of various classes of problems and various users. A distributed two-level (disk—tape) storage system is an important component of this cluster.

The GOVORUN supercomputer is used within the distributed NICA cluster to solve problems requiring massive parallel calculations in lattice QCD for studying the properties of hadron matter at a high energy density; to perform a mathematical simulation of antiproton interactions with protons and nuclei by means of the DPM, FTF, and UrQMD + SMM generators developed at JINR, which are of interest for the NICA-MPD experiment; and to simulate the dynamics of relativistic heavy ion collisions.

The ultrafast data storage system (UDSS) implemented within the GOVORUN supercomputer under the management of the Lustre file system is yet another component of the NICA cluster. At the present time, the UDSS is based on SSD (solid-state drive) disks employing the NVMe (Non-Volatile Memory express) connection technology. This reduces the time of access to data and provides a data acquision/output rate in excess of 300 Gigabyte per second.

The use of the DIRAC platform made it possible to combine the computing resources of MICC JINR, including Tier1/Tier2; the GOVORUN supercomputer; the JINR cloud; and the UDSS Lustre, dCache, and EOS storage resources. These results make a significant contribution to the development of a digital platform for megascience projects.

REFERENCES

1. The Worldwide LHC Computing Grid (WLCG). http://wlcg.web.cern.ch/LCG.
2. Technical Design Report, Document LCG-TDR-001, CERN-LHCC-2005-024 (CERN, 2005).
3. Ph. Charpentier, in *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), Sofia, 2018,* Ed. by A. Forti, L. Betev, M. Litmaath, O. Smirnova, and P. Hristov, EPJ Web Conf. **214**, 09009 (2019). https://doi.org/10.1051/epjconf/201921409009
4. NICA Megaproject. https://nica.jinr.ru/ru/.
5. T. Maeno, J. Phys.: Conf. Ser. **119**, 062036 (2008).
6. F. Stagni, A. Tsaregorodtsev, Ch. Haen, Ph. Charpentier, Z. Mathe, W. J. Krzemien, and V. Romanovskiy, in *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), Sofia, 2018,* Ed. by A. Forti, L. Betev, M. Litmaath, O. Smirnova, and P. Hristov, EPJ Web Conf. **214**, 03012 (2019). https://doi.org/10.1051/epjconf/201921403012
7. I. Bird, S. Campana, M. Girone, X. Espinal, G. McCance, and J. Schovancová, in *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), Sofia, 2018,* Ed. by A. Forti, L. Betev, M. Litmaath, O. Smirnova, and P. Hristov, EPJ Web Conf. **214**, 04024 (2019). https://doi.org/10.1051/epjconf/201921404024
8. V. Korenkov, A. Dolbilov, V. Mitsyn, I. Kashunin, N. Kutovskiy, D. Podgainy, O. Streltsova, T. Strizh, V. Trofimov, and P. Zrelov, in *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), Sofia, 2018,* Ed. by A. Forti, L. Betev, M. Litmaath, O. Smirnova, and P. Hristov, EPJ Web Conf. **214**, 03009 (2019). https://doi.org/10.1051/epjconf/201921403009
9. A. Dolbilov, I. Kashunin, V. Korenkov, N. Kutovskiy, V. Mitsyn, D. Podgainy, O. Stretsova, T. Strzh, V. Trofimov, and A. Vorontsov, in *Proceedings of the 27th Symposium on Nuclear Electronics and Computing, Montenegro, Budva, 2019,* Ed. by V. Korenkov, T. Strizh, A. Nechaevskiy, and T. Zaikina, CEUR Workshop Proc. **2507**, 16 (2019).
10. A. Baginyan, A. Balandin, A. Dolbilov, A. Golunov, N. Gromova, I. Kadochnikov, I. Kashunin, V. Korenkov, V. Mitsyn, D. Oleynik, I. Pelevanyuk, A. Petrosyan, S. Shmatov, T. Strizh, A. Vorontsov, V. Trofimov, et al., in *Proceedings of the 27th Symposium on Nuclear Electronics and Computing, Montenegro, Budva, 2019*, Ed. by V. Korenkov, T. Strizh, A. Nechaevskiy, and T. Zaikina, CEUR Workshop Proc. **2507**, 321 (2019).
11. N. Astakhov, A. Baginyan, S. Belov, A. Dolbilov, A. Golunov, I. Gorbunov, N. Gromova, I. Kadochnikov, I. Kashunin, V. Korenkov, V. Mitsyn, I. Pelevanyuk, S. Shmatov, T. Strizh, E. Tikhonenko, V. Trofimov, et al., Phys. Part. Nucl. Lett. **13**, 714 (2016); A. Baginyan, A. Balandin, S. Belov, A. Dolbilov, A. Golunov, N. Gromova, I. Kadochnikov, I. Kashunin, V. Korenkov, V. Mitsyn, I. Pelevanyuk, S. Shmatov, T. Strizh, V. Trofimov, N. Voytishin, and V. Zhiltsov, in *Proceedings of the 8th International Conference on Distributed Computing and Grid-technologies in Science and Education, Dubna, 2018,* Ed. by V. Korenkov, A. Nechaevskiy, T. Zaikina, and E. Mazhitova, CEUR Workshop Proc. **2267**, 1 (2018).
12. Gh. Adam, M. Bashashin, D. Belyakov, M. Kirakosyan, M. Matveev, D. Podgainy, T. Sapozhnikova, O. Streltsova, Sh. Torosyan, M. Vala, L. Valova, A. Vorontsov, T. Zaikina, E. Zemlyanaya, and M. Zuev, in *Proceedings of the 8th International Conference on Distributed Computing and Grid-technologies in Science and Education, Dubna, 2018,* Ed. by V. Korenkov, A. Nechaevskiy, T. Zaikina, and E. Mazhitova, CEUR Workshop Proc. **2267**, 638 (2018).
13. N. Balashov, A. Baranov, N. Kutovskiy, A. Makhalkin, Y. Mazhitova, I. Pelevanyuk, and R. Semenov, in *Proceedings of the 27th Symposium on Nuclear Electronics and Computing, Montenegro, Budva, 2019,* Ed. by V. Korenkov, T. Strizh, A. Nechaevskiy, and T. Zaikina, CEUR Workshop Proc. **2507**, 185 (2019).
14. A. Baginyan, N. Balashov, A. Baranov, S. Belov, D. Belyakov, Y. Butenko, A. Dolbilov, A. Golunov, I. Kadochnikov, I. Kashunin, V. Korenkov, N. Kutovskiy, A. Mayorov, V. Mitsyn, I. Pelevanyuk, R. Semenov, et al., in *Proceedings of the 26th International Symposium on Nuclear Electronics and Computing (NEC 2017), Budva, 2017*, Ed. by V. Korenkov and A. Nechaevskiy, CEUR Workshop Proc. **2023**, 226 (2017).