USING THE "GOVORUN" SUPERCOMPUTER FOR THE NICA MEGAPROJECT

D.V. Belyakov, A.G. Dolbilov, A.N. Moshkin, I.S. Pelevanyuk,
D.V. Podgainy, O.V. Rogachevsky, O.I. Streltsova, M.I. Zuev

Joint Institute for Nuclear Research, 6 Joliot-Curie, 141980, Dubna

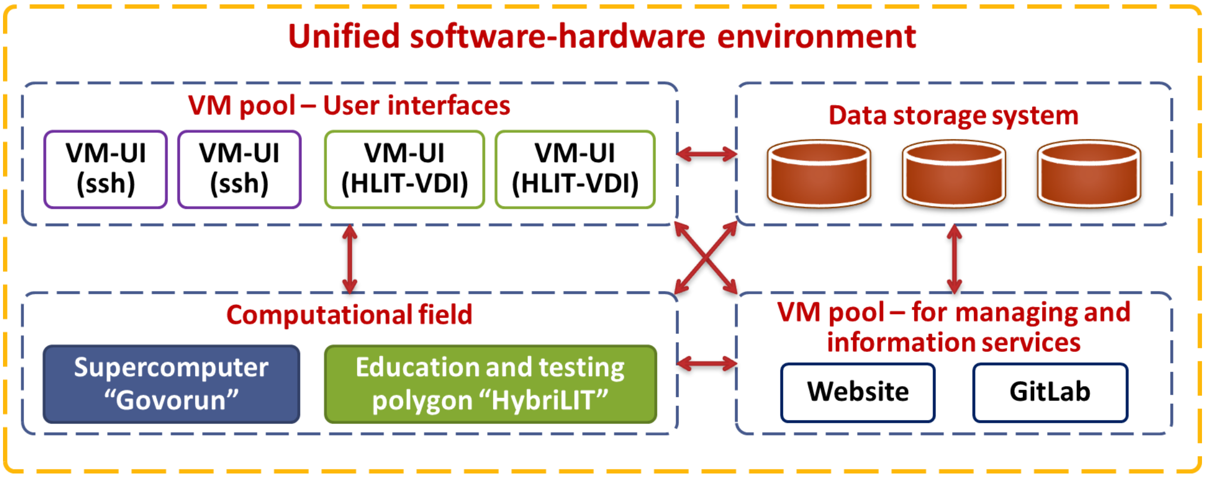
E-mail: podgainy@jinr.ru

At present, the "Govorun" supercomputer is used for both theoretical studies and event simulation for the MPD experiment of the NICA megaproject. To generate simulated data of the MPD experiment, the computing components of the "Govorun" supercomputer, i.e. Skylake (2880 computing cores) and KNL (6048 computing cores), are used; data are stored on the ultrafast data storage system (UDSS) under the management of the Lustre file system with a subsequent transfer to cold storages controlled by the EOS and ZFS file systems. UDSS currently has five storage servers with 12 SSD disks using the NVMe connection technology and a total capacity of 120 TB, which ensures a low time of data access and a data acquisition/output rate of 30 GB per second. Due to the UDSS high performance, by September 2019, over 100 million events for the MPD experiment have already been generated and more than 30 million events have been reconstructed. In the future, other MC generators are expected to be used as well. It is planned to use the DIRAC software for managing jobs and the process of reading out/recording/processing data from various types of storages and file systems. All the enumerated above will allow one to check a basic set of data storage and transmission technologies, simulate data flows, choose optimal distributed file systems and increase the efficiency of event modeling and processing.

Keywords: High-performance computing, hyper-converged systems, data storage and processing systems, middleware for high-energy physics

1. Introduction. The HybriLIT platform.

The Heterogeneous platform "HybriLIT" [1] is the part of the JINR Multifunctional Information and Computing Complex [2] for high-performance computing. The HybriLIT platform consists of two elements, i.e. the education and testing polygon and the "Govorun" supercomputer, combined by a unified software and information environment. It is noteworthy that information technology is one of the most rapidly developed fields in terms of the hardware development (development of computing architectures, data storage systems, network solutions), as well as the development of methods, algorithms, software for calculations on novel computing architectures and the software development using novel frameworks and libraries. All the enumerated above creates serious requirements to a software and information environment, namely, flexibility, the ability to quickly form an IT environment for solving specific applied tasks; scalability, for a fast expansion/reduction of the computing field; user-friendly, a prompt response to user requests from the platform (community) solving problems in various fields developed at the Institute. The indicated requirements are satisfied through different mechanisms. Flexibility is achieved due to the formed dynamic pool of virtual machines designed for debugging and launching jobs (the VM user interface supporting only the SSH-protocol) and VMs for solving graphical tasks (the HLIT-VDI service based on the Citrix technology). Scalability is attained due to the possibilities of the SLURM job scheduler allowing, if necessary, to reallocate computing resources in queues. The user-friendly requirement of the software and information environment is realized by the rapid deployment of necessary IT environments and work with heterogeneous computing group users through services, which are actively developed and supported by the team. The structure of the created and supported software and hardware environment is presented in Figure 1.


Unified software-hardware environment

VM pool – User interfaces

VM-UI (ssh)

VM-UI (ssh)

VM-UI (HLIT-VDI)

VM-UI (HLIT-VDI)

Data storage system

Computational field

Supercomputer "Govorun"

Education and testing polygon "HybriLIT"

VM pool – for managing and information services

Website

GitLab

Figure 1. Structure of the software and hardware environment

Platform users are able to develop and debug their applications on the education and testing polygon, furthermore, carry out calculations on the supercomputer, which allows them to effectively use the supercomputer resources. The given possibility is provided by the unified software and information environment, including the unified system level (the operation system, the job scheduler, file systems and software), as well as a set of services allowing users to quickly get answers to their questions, jointly develop parallel applications, receive information about conferences, seminars and meetings dedicated to parallel programming technologies [3].

The "Govorun" supercomputer [4] commissioned in 2018, is aimed to cardinally accelerate complex theoretical and experimental studies in all projects underway at JINR especially for NICA megaproject [5]. The supercomputer is a heterogeneous computing platform containing the GPU component based on the NVIDIA graphics accelerators and the CPU component based on two Intel computing architectures. The GPU component consists of 5 NVIDIA DGX-1 servers with 8 GPU NVIDIA Tesla V100 in each and InfiniBand interconnect between them is 100 Gbits/s. The supercomputer CPU component is implemented on the high-density architecture "RSC Tornado" with direct liquid cooling developed by specialists of the Russian company "RSC Group" [6]. CPU computing nodes have a two types Intel server products, namely, 21 nodes contain 72-core server processors Intel® Xeon Phi™ 7290 and 40 nodes have processors Intel® Xeon® Scalable (models Intel® Xeon® Gold 6154). The last one nodes contain a high-speed solid-state disks Intel® SSD DC P4511 with the NVMe interface and a capacity of 1 TB. For high-speed data transfer between CPU computing nodes, the supercomputer uses a switching technology Intel® Omni-Path, which ensures the speed of non-blocking switching up to 100 Gbit/s based on 48-port switches Intel® Omni-PathEdgeSwitch 100 Series with 100% liquid cooling.

To speed up work with data, an ultrafast data storage system (UDSS) is implemented in the "Govorun" supercomputer under the Lustre file system. At present, UDSS has five storage servers with 12 SSD drives with the NVMe connection technology, which reduces the time of access to data. The total capacity of UDSS is currently 120 TB and the data acquisition/output rate is 30 TB per second. It is noteworthy that UDSS has the ability to linearly increase productivity (speed of working with data) and the storage volume without changing the principles of the architectural design of the system. It should also be highlighted that a new approach for the industry of high-performance systems, based on the principle of hyper-convergence of the system, the essence of which is to combine resources for computing and storage on each node of the system, is implemented in the "Govorun" supercomputer. Each node of the system is both part of the computing subsystem and part of the distributed system of user data storage, performing two types of load at once (compute/store). It allows linearly scaling the system resources with an increase in the number of nodes. Unlike the classical approach in HPC, when the computing system and the storage system are separate and scale separately, in the hyper-converged system, as the number of nodes increases, both the computing power and the volume/speed of the distributed data storage system grow in parallel [7].

Thus, the uniqueness and advantage of the CPU component of the given supercomputer are 100% liquid cooling, which ensures a high density of computing nodes, i.e. 150 per cabinet, and high energy efficiency, about 10 GFlop/W; hyper-convergence and the ultrafast data storage system enabling to significantly accelerate data processing and being unique in the field of high-performance computing; heterogeneity, which allows solving tasks that require different computing architectures.

The total peak supercomputer performance is 1 PFlops for single-precision operations and 500 TFlops for double-precision operations.

2. "Govorun" supercomputer for NICA

At present, the computing resources of the "Govorun" supercomputer are actively used for both theoretical studies in lattice quantum chromodynamics and event generation and reconstruction for the NICA experiments. For theoretical research, the GPU and CPU components are used, while for event generation and reconstruction, only the CPU components, namely, Skylake and KNL, are used; data are stored on UDSS with a subsequent transfer to cold storages controlled by the EOS and ZFS file systems. The DIRAC software [8] is used to manage jobs and the process of reading out/recording data from different types of storages and file systems (Fig.2). The DIRAC Interware is a product for integrating heterogeneous computing and data storage resources into a unified platform [9]. The integration of resources is based on using standard data access protocols (xRootD, GridFTP, etc.) and pilot jobs. Owing to it, a user is provided with a unified environment, in which it is possible to run jobs, manage data, build processes and monitor their implementation. Batch processing systems, grid computing elements, clouds, supercomputers and even separate computing nodes can act as computing resources in the framework of DIRAC. When working with data, DIRAC provides all necessary commands. For correct operation of all commands, the storage system should support standard grid file transfer protocols. This allows the pilot to acquire data on any of the resources and upload them back. However, a user can always send his job to a specific computing resource and then work with the local file system on the resource. In this case, additional effort may be required to make data accessible from anywhere.

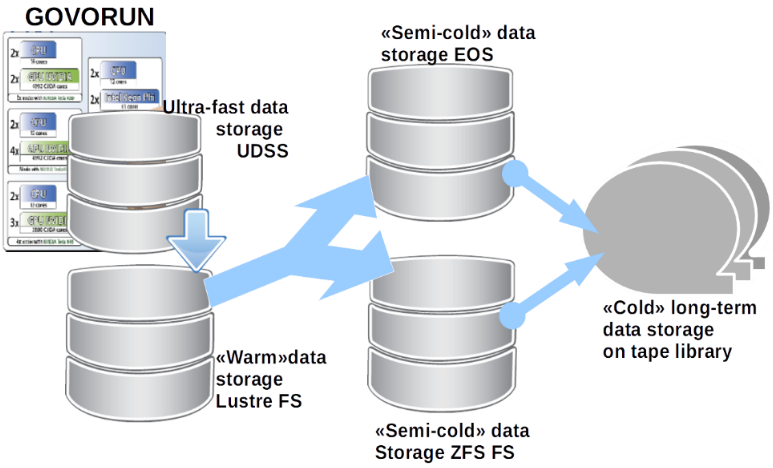

GOVORUN
Ultra-fast data storage UDSS
«Semi-cold» data storage EOS
«Cold» long-term data storage on tape library
«Warm»data storage Lustre FS
«Semi-cold» data Storage ZFS FS

Figure 2. Scheme of data transfer on the "Govorun" supercomputer for computing modeling for the NICA megaproject and event simulation for the MPD experiment implemented with the help of the DIRAC Interware

Using DIRAC, the computing resources of the JINR MICC, i.e. Tier1/Tier2, the "Govorun" supercomputer, the JINR cloud and storage resources such as Lustre UDSS, dCache and EOS, were combined. Within Monte-Carlo data generation for the MPD experiment, 70,000 jobs were performed on the MICC Tier1/Tier2 components using the DIRAC platform, and 15,000 jobs were carried out on the computing resources of the "Govorun" supercomputer using UDSS. As a result, 4.5 TB of data were generated and sent to dCache.

In total, by September 2019, over 167000 tasks on all computing components were completed by all groups performing calculations on the supercomputer. It is noteworthy that about 31% of them refer to computing for the NICA megaproject. Moreover, more than half are the theoretical calculations carried out on all computing components of the supercomputer. More than 40% directly relate to event generation and reconstruction for the MPD experiment, namely, 75 million events have been generated using UrQMD, LAQGSM, PHSD and other models and 30 million events have been reconstructed (Fig.3).

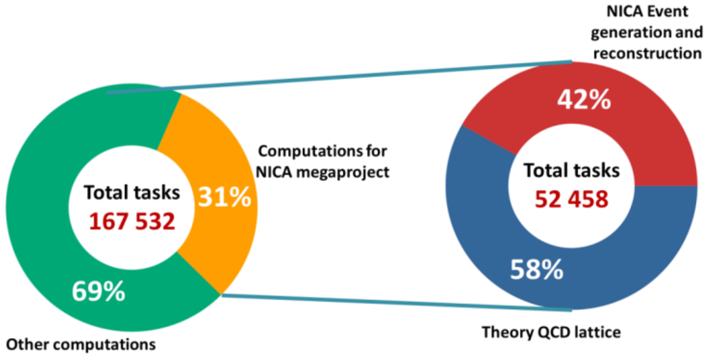


Figure 3. Using the resources of the "Govorun" supercomputer for the NICA megaproject

3. Conclusion and future plans

One of the major tasks, for which the resources of the supercomputer will be used, is the development of computing for the NICA megaproject. The created computing for the NICA megaproject should provide data acquisition from detectors and data transmission for processing and analysis. To perform the given tasks, computing has certain requirements, including the requirements to the network infrastructure, computing architectures, storage systems, as well as the appropriate software of the system and for data processing and analysis. Developed computing models should take into account the trends in the development of network solutions, computing architectures and IT solutions, which allow combining supercomputer (heterogeneous), grid and cloud technologies and creating distributed, software-configured HPC platforms on its basis. The use of such solutions for data processing and analysis requires the creation of software environments, which provide the necessary code abstraction enabling to implement the required functionality for a wide range of computing tools.

The implementation of various computing models for the NICA megaproject entails the confirmation of the model performance, i.e. meeting the requirements for temporal characteristics of data acquisition from detectors with their subsequent transfer to processing, analysis and storage, as well as the requirements for the efficiency of modeling and processing events in the experiment. For these purposes, it is necessary to carry out tests in a real software and hardware environment, which should contain all the required components. Supercomputer "Govorun" is the hyperconverged computational system and contains all newest computational architecture and ultrafast data storage system and that why it is one of the major elements for performing a full cycle of testing a computing model for NICA megaproject.

Acknowledgments

The studies in the given direction were supported by the RFBR grant ("Megascience – NICA"), №18-02-40101.

References

[1] Heterogeneous platform "HybriLIT", http://hlit.jinr.ru/en/

[2] A.G. Dolbilov, I.A. Kashunin, V.V. Korenkov, Multifunctional information and computing complex of jinr: status and perspectives, ibid; http://micc.jinr.ru

[3] V.V. Korenkov, D.V. Podgainy, O.I. Streltsova, Educational program on HPC technologies on the basic of the HybriLIT heterogeneous cluster (LIT JINR). Modern Information Technology and IT-education V13, №4, 2017, P141-146 (in Russian)

[4] Supercomputer "Govorun", http://hlit.jinr.ru/en/about_govorun_eng/

[5] NICA (Nuclotron-based Ion Collider fAcility), http://nica.jinr.ru

[6] RSC Group. http://www.rscgroup.ru/en/company

[7] P. Lavrenko, Disaggregated composable environment for high performance problems, "Storage News" № 2 (74), 2019.

[8] Federico Stagni, Andrei Tsaregorodtsev, ubeda, Philippe Charpentier, Krzysztof Daniel Ciba, Zoltan Mathe, ... Luisa Arrabito. (2018, October 8). DIRACGrid/DIRAC: v6r20p15 (Version v6r20p15). Zenodo. http://doi.org/10.5281/zenodo.1451647

[9] Gergel V., Korenkov V., Pelevanyuk I., Sapunov M., Tsaregorodtsev A., Zrelov P. (2017) Hybrid Distributed Computing Service Based on the DIRAC Interware // DAMDID/RCDL 2016. Communications in Computer and Information Science, vol 706. Springer, Cham. DOI: 10.1007/978-3-319-57135-5_8