

INTEGRATION OF THE PARALLEL RESOURCES TO THE DISTRIBUTED CLOUD INFRASTRUCTURES FOR LARGE SCALE PROJECTS

**S.D. Belov^{1,2 a}, I.S. Kadochnikov^{1,2 b}, V.V. Korenkov^{1,2 c}, N.A. Kutovskiy^{1,g},
I.S. Pelevanyuk^{1,2,e}, R.N. Semenov^{1,2 f}, P.V. Zrelov^{1,2 g}**

¹ *Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia*

² *Plekhanov Russian University of Economics, Stremyanny lane, 36, Moscow, 117997, Russia*

E-mail: ^a belov@jinr.ru, ^b kadivas@jinr.ru, ^c korenkov@jinr.ru, ^d kut@jinr.ru, ^e pelevanyuk@jinr.ru,
^f roman@jinr.ru, ^g zrelov@jinr.ru

Cloud services and data centers use a distributed network architecture for data storage, queries, and search engines. Virtualization is used to improve reliability, reduce services cost, and increase flexibility in deploying services. It is also applied to provide for the ability to migrate from one physical server to another. This paper provides an overview of the Joint Institute's for Nuclear Research (JINR) international distributed cloud infrastructure and the Plekhanov Russian University of Economics (PRUE) cloud infrastructure as a part of the distributed cloud. The paper underlines the unique role of the DIRAC grid interware, which combines the cloud resources as from some JINR Member State organizations as PRUE cloud. A special section is devoted to the integration of the parallel resources of PRUE to the cloud.

Keywords: cloud computing, parallel computing, DIRAC interware

Sergey Belov, Ivan Kadochnikov, Vladimir Korenkov, Nikolay Kutovskiy,
Igor Pelevanyuk, Roman Semenov, Petr Zrelov

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introduction

The experiments at the Large Hadron Collider (LHC) at CERN (Geneva, Switzerland) played a leading role in scientific research not only in High Energy Physics and Nuclear Physics but also in Big Data Analytics. Global distributed system for processing, storage, and analyzing data WLCG (Worldwide LHC Computing GRID) brings together the resources of about 180 computer centers in 50 countries; the total storage capacity is more than 1 Exabytes. Data processing and analysis are carried out using high-performance complexes (Grid), academic, national, and commercial cloud computing resources, supercomputers, and other resources. JINR is actively involved in integrating distributed heterogeneous resources and the development of Big Data technologies to provide modern megaprojects in such high-intensity fields of science as high energy physics, astrophysics, bioinformatics, and others.

The Joint Institute for Nuclear Research (JINR) [1] is an international intergovernmental organization. It is developing as a large multidisciplinary international scientific center incorporating basic research in modern nuclear physics, development and application of high technologies, and university education in the relevant fields of knowledge. Currently, JINR has 18 Member States and 6 countries participating in JINR activities based on bilateral agreements signed on the governmental level.

The research program of the JINR is aimed at conducting ambiguous and large-scale experiments on the Institute's basic facilities and in frames of worldwide cooperation. This program is connected with the implementation of the NICA (Nuclotron-based Ion Collider fAcility) megaproject [2], the construction of new experimental facilities, the JINR neutrino program, the modernization of the Large Hadron Collider (LHC) [3] experimental facilities (CMS, ATLAS, Alice), programs on condensed matter physics and nuclear physics. The recent years' experience shows that the progress in obtaining research results depends directly on computing resources' performance and efficiency. JINR possesses an information-computational complex that has evolved into a set of stand-alone structures with a shared engineering and networking infrastructures. Support of this fully functional infrastructure is the central task of the Laboratory of Information Technologies.

The JINR computing infrastructure combines a broad spectrum of computing components and IT technologies, providing the opportunity to solve various scientific and engineering tasks facing the Institute, from theoretical studies to experimental data processing, storage, and analysis.

2. JINR Cloud infrastructure

A cloud infrastructure at the Joint Institute for Nuclear Research (JINR) was created in 2013. The aim was to manage LIT IT services and servers more efficiently using modern technologies, to combine resources for solving common tasks, increase the efficiency of hardware utilization and service reliability, simplify access to application software and optimize the use of proprietary software as well as provide a modern computing facility for JINR users.

The JINR cloud infrastructure [4] operates on the base of the OpenNebula (release 5.8), enabling relatively easy integration of supporting container virtualization based on OpenVZ, as well as its extensions and additions. It is compatible with the Linux operating system (OS). It has the possibility of running virtual machines with this OS, the required functionality and quality of the software product, the license that permits modification and free use, the availability of clear and accessible documentation, and support from developers when modifying the software. Besides, OpenNebula appears to be an optimal choice in terms of the interrelationship of the hardware functioning in the infrastructure on its basis and the effort required to develop and maintain the cloud.

The JINR cloud resources were increased up to 1,564 CPU cores and 8.54TB of RAM in total. Current hardware resources: 66 servers for VMs, 10 servers for ceph-based software-defined storage (SDS), 3 servers for front-end nodes in high availability setup. The JINR cloud grows not only in the capacity of resources but also in the number of activities. It is used for different system and application tasks, namely, COMPASS production system services [5], a data management system of the UNECE ICP Vegetation, a service for scientific and engineering computations, a service for data visualization based on Grafana, JupyterHub infrastructure for it, gitlab and its runners as well as some others. Along with it, there was a successful attempt to deploy a virtual machine in the JINR cloud with a GPU card transmitted

from the server for developing and running machines and deep learning algorithms for the JUNO experiment.

The approach to cloud integration is based on the OpenNebula cloud platform's built-in mechanism and works well for a small number of joined clouds. However, it sufficiently increases the complexity of such infrastructure maintenance with a growing number of participating clouds.

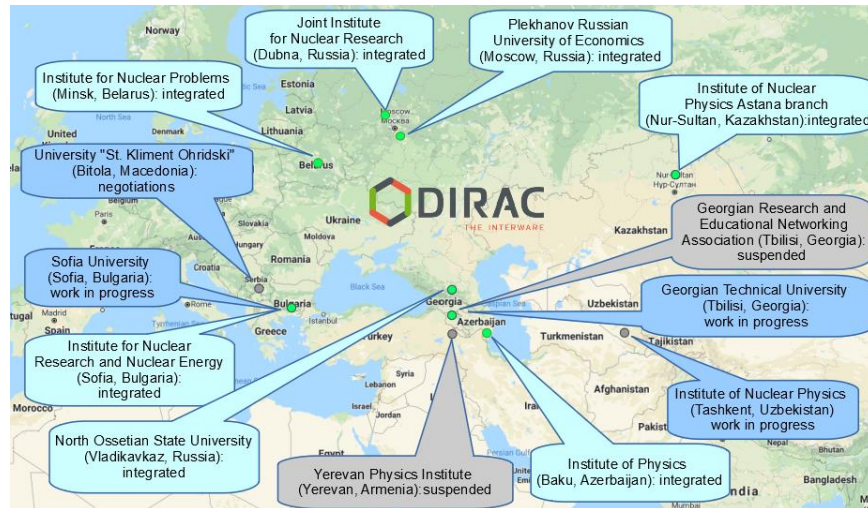


Fig. 1. International distributed cloud infrastructure participants on the map

Another approach uses the possibility to combine clouds by integrating them using a distributed workload management system – DIRAC grid Interware [6]. Different distributed heterogeneous computing and storage resources from clouds of the JINR Member State organizations combine with the help of this approach (Fig. 1).

Fig. 2 shows the contribution of each cloud site to general number of the load test jobs executed by clouds, integrated into a common international JINR cloud infrastructure. Meanwhile, the total number of computational jobs on different clouds does not represent their performance but rather the availability of free resources for the tasks of the distributed cloud infrastructure.

At the moment, based on the experience of many completed tasks (within the framework of the Folding@Home project as well), the most reliable and effective are the JINR and PRUE clouds.

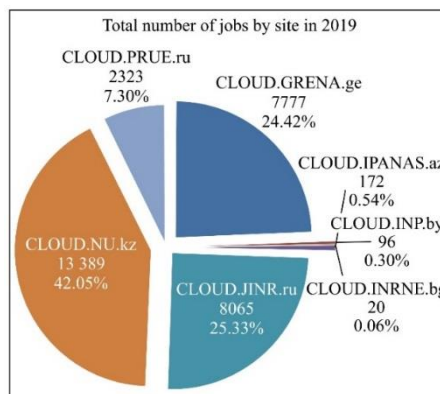


Fig.2. Distribution of jobs by clouds of the Member States' organization

3. PRUE Cloud infrastructure

Cloud infrastructure of the Scientific laboratory of cloud computing and Big Data analytics [7] of Plekhanov Russian Economic University (after this — the cloud, cloud service) operates based on software OpenNebula 5. As a storage system, it is used software-defined storage (SDS) based on Ceph version 12. All servers are running under Linux Centos 7 operating system.

Currently, the cloud service is deployed on eight servers. A single server takes the lead role and hosts the following main components:

- cloud infrastructure core (OpenNebula core),
- OpenNebula scheduler,
- MySQL database server,
- interfaces for accessing the cloud (user-defined web interface and command-line interface and the application programming interface).

Four servers operate as cloud worker nodes (CWNs) that directly host virtual machines (VMs). Three servers act as storage nodes and, at the same time, as cloud-based worker nodes.

The network part of the cloud consists of two subnets (Fig. 3). One subnet is intended for virtual machines and has a connection to the network switch at 1 Gbit. Another subnet is dedicated to SDS traffic and is connected to the network switch at a speed of 10 Gbit. Internet access is provided with the network equipment of the Plekhanov Russian University of Economics.

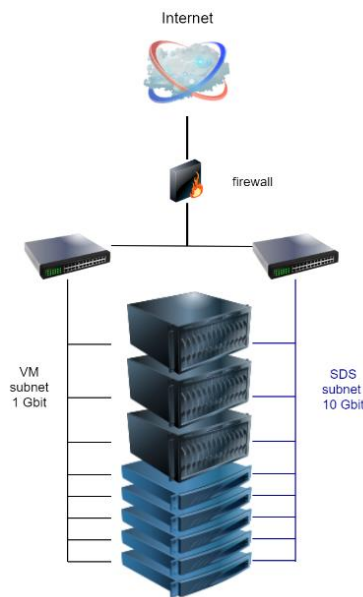


Fig. 3. The network structure of the PRUE cloud

The total resources of the cloud: Processors: 264 cores, RAM: 544 GB, Disks: 200 TB.

Currently, PREU cloud resources are used in several ways:

- training, research and test tasks, as well as development in various projects;
- hosting services with high availability and reliability;
- computing resources, including as an extension of computing capabilities of grid infrastructures.

In addition to processing requests from PRUE users, the cloud service is integrated with other clouds that are part of the computing resources of the JINR member organizations' clouds.

The integration of cloud infrastructures is carried out using the DIRAC grid platform (distributed infrastructure with remote Agency management). The reasons to chose this interware platform:

- provides all necessary functionality, including operation and data management;
- cloud as a computing base support;
- simplify the deployment and maintenance of services compared to other platforms with similar functionality (for example, EMI).

This approach also allows you to share the resources of each cloud between external network users and local non-network users.

Currently, the integration of JINR member states' clouds into a distributed DIRAC based platform is at various stages (the stages and locations of participants in the distributed cloud infrastructure are shown on the map in Fig. 1).

4. Integration of parallel resources to the distributed cloud

The DIRAC Interware is a software framework for distributed computing, providing a complete solution to one or several user communities requiring access to distributed resources. The DIRAC software offers a common interface to a number of heterogeneous computing and storage resources. From the user's perspective, DIRAC is a system that accepts their computing jobs and uploads results to storage.

DIRAC uses a pilot mechanism to run jobs on heterogeneous resources. The general idea of the pilot is the following: jobs do not run on computing resources directly but through a special program called «Pilot».

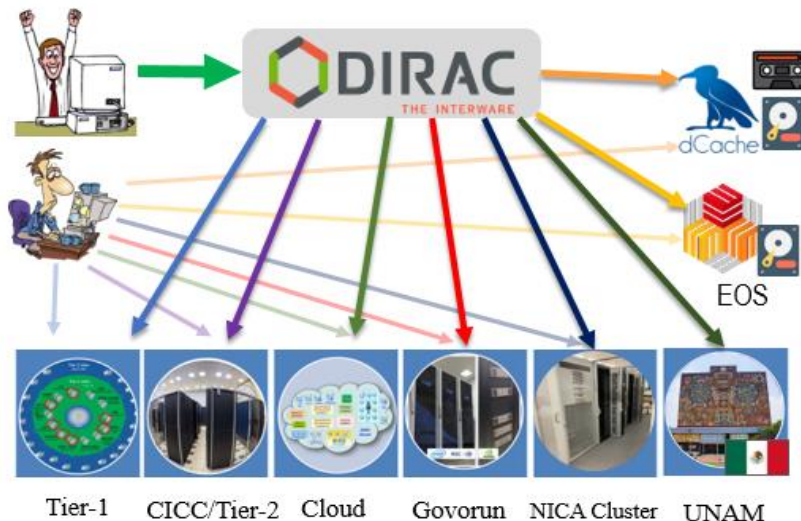


Fig. 4. Scheme of integration of geographically distributed heterogeneous resources based on the DIRAC Interware

The JINR DIRAC Installation has been installed and gradually improved since 2017. At present, the following computing resources of JINR are integrated into the JINR DIRAC Installation (Fig. 4) [8]: Tier1/Tier2, «Govorun» supercomputer, JINR cloud, NICA cluster, and dCache and EOS storage resources, JINR Member States clouds. A cluster of the National Autonomous University of Mexico (UNAM) has recently joined the system, the DIRAC-based unified environment, which includes both computing resources and data storage systems, is used to generate and reconstruct events of the MPD experiment, to study the SARS-CoV-2 virus within the Folding@Home project on available cloud resources and to integrate clouds of the JINR Member States' organizations into a distributed platform.

The load of the parallel cluster is not always 100%. That is natural that the load is not constant on this type of system. To improve resource utilization it is possible to use it for scientific projects which do not require dedicated resources but could benefit from opportunistic resources. Another possible use is introducing a parallel cluster to the studying process.

The parallel cluster was integrated into the DIRAC instance in the Joint Institute for Nuclear Research. This instance was supported and gradually improved since 2016. It supports educational and scientific groups of users: Multi-Purpose Detector, Baryonic Matter at Nuclotron, and Baikal-GVD collaboration groups. That allows getting scientific jobs by REA resources in case of scientific collaboration. To integrate REA parallel cluster the pilot jobs should be able to run on the resource. The pilot job works like a wrapper job for user workload. It checks the environment, operating system, software, RAM, and CPU performance. This information is sent to DIRAC to match the appropriate job for the pilot. Since SLURM is used as a batch system on the parallel cluster, a special DIRAC module has been used. An additional system user named "dirac" was created to represent jobs submitted by DIRAC users.

The first issue was related to the default Python version. On the parallel cluster, it is Python 3.6.5, but right now, DIRAC works only on Python 2.7 version. Special settings for DIRAC users allowed

changing the default Python version to the correct one. The second issue with integration was related to time settings on the cluster. Since DIRAC is being oriented to distributed systems, it requires all resources to keep time precisely. After fixing time DIRAC was able to submit user jobs to pilots.

After these corrections, DIRAC was able to submit user jobs and get results. However, it took much time for pilots to start. The starting of DIRAC pilots means downloading the TAR archive with all python code and extracting it. This is IO intensive operation, and shared file systems usually do not perform it well. The shared file system on the parallel cluster is based on CEPH. To fix that issue DIRAC was instructed to set the working directory for pilots not on CEPH but disk. With that fixes DIRAC started to work effectively.

The second part of the integration was checking parallel cluster performance. The cluster has 32 slots for jobs and equipped with Intel Xeon Gold 6130 CPU. It is important to know the performance of an individual slot. DIRAC uses a special benchmark called DB12 for analysis of CPU performance. The benchmark showed results of around 25 HEP-SPEC06 for one slot. That is a good result which is better or comparable with most processors used in JINR distributed infrastructure up to now. Then the network was tested with a standard upload/download test. The result showed 100MB/s transfer speed. It is worth notice that jobs running through DIRAC may use GPU resources on the cluster. That makes it useful not only for CPU load but also for GPU load.

Provided numbers show that high computing power of the REA parallel cluster may be accessed now through DIRAC. This allows using it by students in the education process. Another possible use is the participation of the REA parallel cluster in international scientific experiments by providing computing power. And if some resources are idle they may be utilized in the Folding@HOME project or other projects.

5. Conclusion

In the paper, we have discussed several perspectives on creating and maintaining the international distributed cloud infrastructure and underlying architecture, including network aspects. There were presented ways of gluing clouds together to provide system services for distributed computing, resources, and interfaces for application tasks. There is an overview of the international distributed cloud infrastructure of the Joint Institute for Nuclear Research (JINR) and the cloud infrastructure of the Plekhanov Russian University of Economics (PRUE) as a part of the distributed cloud. The paper emphasizes the unique role of the DIRAC grid interware in integrating cloud resources from some JINR Member State organizations as PRUE cloud. The particular part is about the integration of the high-performance parallel resources to the cloud.s

Acknowledgement

The study was carried out at the expense of the Russian Science Foundation grant (project No. 19-71-30008).

References

- [1] Joint Institute for Nuclear Research. Web: <http://www.jinr.ru/>
- [2] NICA (Nuclotron-based Ion Collider fAcility). Web: <http://nica.jinr.ru/>
- [3] LHC (Large Hadron Collider). Web: <https://home.cern/science/accelerators/large-hadron-collider>
- [4] N.A. Balashov, A.V. Baranov, N.A. Kutovskiy, A.N. Makhalkin, Ye.M. Mazhitova, I.S. Pelevanyuk, R.N.Semenov, Proc. of the 27th International Symposium on Nuclear Electronics & Computing (NEC'2019), <http://ceur-ws.org/Vol-2507/185-189-paper-32.pdf>, (2019).
- [5] A.Sh. Petrosyan, EPJ Web of Conf., Vol. 214, 03039, 2019, <https://doi.org/10.1051/epjconf/201921403039>.

- [6] NA. Balashov, R.I. Kuchumov, N.A. Kutovskiy, I.S. Pelevanyuk, V.N. Petrunin, A.Yu. Tsaregorodtsev, Proc. of the 27th International Symposium on Nuclear Electronics & Computing (NEC'2019), <http://ceur-ws.org/Vol-2507/256-260-paper-45.pdf>, (2019).
- [7] Scientific laboratory of cloud technologies and Big Data Analytics of Plekhanov Russian University of Economics. <https://www.rea.ru/ru/org/managements/unitscires/Laboratorija-Oblachnykh-tekhnologijj-i-analitiki-Bolshikh-dannykh/Pages/lotiabd.aspx>.
- [8] Korenkov V., Pelevanyuk I., Tsaregorodtsev A. (2020) Integration of the JINR Hybrid Computing Resources with the DIRAC Interware for Data Intensive Applications. In: Elizarov A., Novikov B., Stupnikov S. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2019. Communications in Computer and Information Science, vol 1223. Springer, Cham. https://doi.org/10.1007/978-3-030-51913-1_3.