

PRESENT STATUS AND MAIN DIRECTIONS OF THE JINR CLOUD DEVELOPMENT

**N.A. Balashov¹, A.V. Baranov¹, N.A. Kutovskiy^{1, a}, A.N. Makhalkin¹,
Ye.M. Mazhitova^{1, 2, 3}, I.S. Pelevanyuk^{1, 4}, R.N. Semenov^{1, 4}**

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

² *Institute of Nuclear Physics, I Ibragimova street, Almaty, 050032, Kazakhstan*

³ *Dubna State University, 19 Universitetskaya street, Moscow region, Dubna, 141982, Russia*

⁴ *Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia*

E-mail: ^a kut@jinr.ru

The JINR cloud is growing not only in terms of the amount of resources, but also in the number of activities it is used for, namely, COMPASS production system services, a data management system of the UNECE ICP Vegetation, a service for disease detection of agricultural crops through the use of advanced machine learning approaches, a service for scientific and engineering computations, a service for data visualization based on Grafana, the jupyterhub head and execute nodes for it, gitlab and its runners, as well as some others. Apart from that, there was a successful attempt to deploy a virtual machine in the JINR cloud with a GPU card passed through from the server for developing and running machine and deep learning algorithms for the JUNO experiment. Moreover, the JINR distributed information and computing environment, combining resources from JINR Member State organizations with the help of the DIRAC grid interware, began to be used for running BM@N and MPD experiment jobs. The software distribution on these remote resources was performed using the CernVM File System. All these topics are covered in detail.

Keywords: cloud computing, OpenNebula, clouds integration, ceph, KVM, OpenVZ

Nikita Balashov, Alexandr Baranov, Nikolay Kutovskiy, Alexandr Makhalkin,
Yelena Mazhitova, Igor Pelevanyuk, Roman Semenov

Copyright © 2019 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Hardware resources

One of the key characteristics of the JINR cloud [1] is hardware resources it has. At the time of writing this article, the number of CPU cores is 1564, the total amount of RAM is 8.54 TB, and the total raw capacity of ceph-based storage is about 1.2 PB, about 60% of which are used. Some work was performed to aggregate and visualize information about JINR cloud resources on a single dashboard (see the screenshot from the MICC [1] monitoring system called “LITMon” in Figure 1).

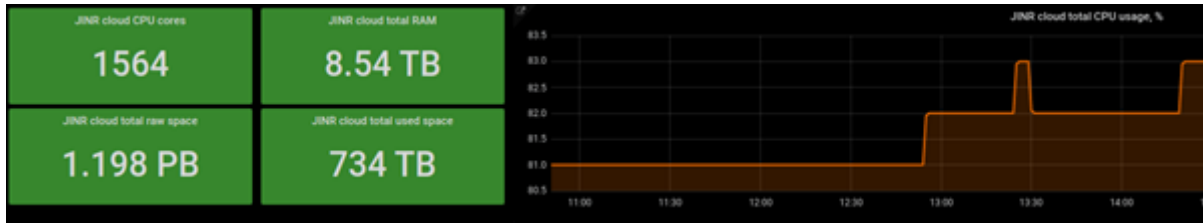


Figure 1. Screenshot from the MICC monitoring system on JINR cloud hardware resources

A sufficient growth of the JINR cloud hardware resources is expected by the end of 2019, which is represented in Table 1.

Table 1. Expected growth of the JINR cloud hardware resources by the end of 2019

Type of resource	Current	By the end of 2019			
	All contributors	NOvA	Baikal-GVD	JUNO	All contributors
CPU cores, items	1564	+120	+300	+2000	3984
RAM, TB	8.54	+0.64	+3	+32	44.18

2. Recent changes

Initially the JINR cloud supported two types of virtualization: full hardware virtualization based on KVM and OS-level virtualization based on OpenVZ. The OpenNebula cloud platform, which the JINR cloud is running on, had no support for OpenVZ, which was used at JINR. The corresponding driver was developed by the JINR cloud team. However, over time, due to lack of manpower, it became hard to support the OpenVZ driver for OpenNebula, i.e. to validate it for each cloud platform release, fix bugs and incompatibilities. Moreover, LXD support (an open sources alternative for OpenVZ containers) appeared in OpenNebula. For all these reasons, it was decided to drop OpenVZ support and migrate OpenVZ user containers running in the JINR cloud to KVM virtual machines (VMs). Another migration process was started in parallel and successfully accomplished, i.e. moving KVM VMs with images on local hard disk drives to ceph-based storage (Fig. 2).

There was a successful attempt to deploy a VM in the JINR cloud with a GPU card passed through from the server for developing and running machine and deep learning algorithms for the JUNO experiment. This approach provides more flexibility for users in terms of the operating system and software environment that users can have inside the VM in comparison with the supercomputer and batch systems.

3. Resources utilization

The number of users and services, for which the JINR cloud resources are used, is constantly growing. At the moment of writing this article, the total number of registered users is 267 with 336 running VMs, which are used for the following services and purposes: COMPASS production system services, a data management system of the UNECE ICP Vegetation, a service for disease detection of agricultural crops through the use of advanced machine learning approaches, a service for scientific

and engineering computations, a service for data visualization based on Grafana, the jupyterhub head and execute nodes for it, gitlab and its runners, as well as some others. Brief information about some of them is provided below.

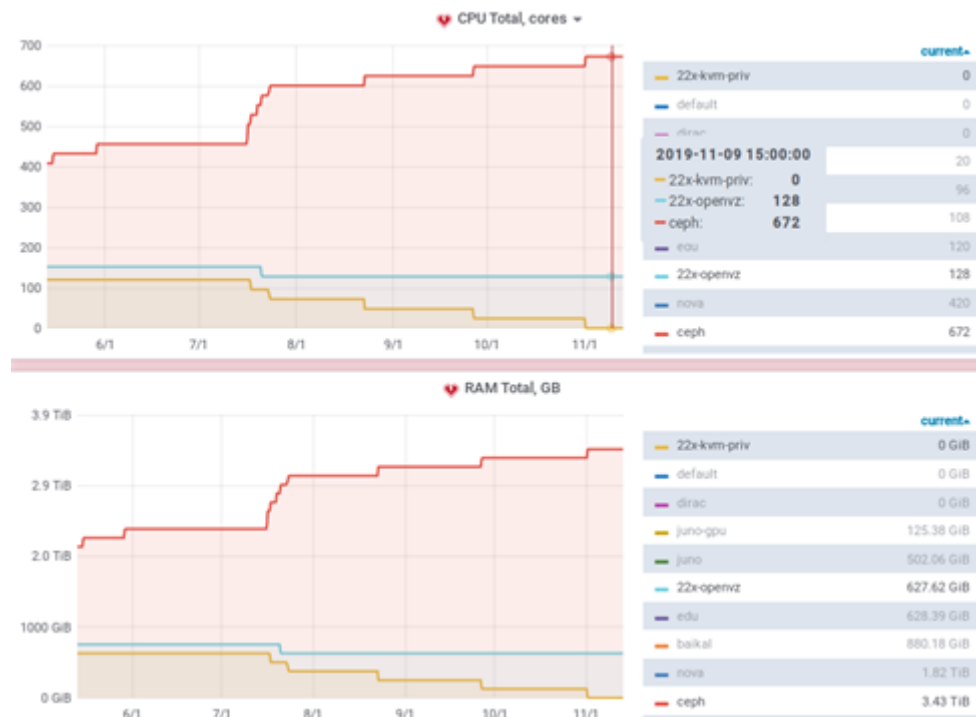


Figure 2. Screenshot of the grafana dashboard showing the number of CPU cores, the amount of RAM in KVM and OpenVZ cloud resources from 2019-06-01 to 2019-11-12

3.1. Service for scientific and engineering computations

To simplify access to MICC resources for small scientific groups from JINR and its Member State organizations, as well as to ease the usage of command line based applications, a cloud service for scientific and engineering computations was developed [3]. Its web interface is available at the following URL: <http://saas.jinr.ru>. At the time of writing this article, two applications are available in this service: stack simulations of long and short Josephson junctions. A user should be authenticated in this service. For unauthorized users there is a demo mode that allows one to navigate the web interface to get an idea of how it looks like. It is planned to add more applications to this service.

3.2. Data management system of the UNECE ICP Vegetation

The data management system (DMS) was deployed on the resources of the JINR cloud to manage UNECE International Cooperative Program (ICP) Vegetation environmental data. DMS [4] provides its participants with a modern unified system of collecting, analyzing and processing biological monitoring data and facilitates IT aspects of all biological monitoring stages starting from the choice of sampling sites and finishing with the generation of pollution maps for a particular area or long-term state-of-environment forecast.

3.3. Multifunctional platform for plant disease detection

The aim of the multifunctional platform for plant disease detection (PDDP) [5] is to provide a new level of service for the farmers' community helping with plant disease detection because crop losses are the major threat to the well-being of rural families, the economy and governments, and food security worldwide. PDDP consists of a set of interconnected services and tools developed, deployed, and hosted with the help of the JINR cloud infrastructure. PDDP was designed using modern approaches and deep learning techniques. The service also has a mobile application that allows users to send photos and text descriptions of sick plants and get the cause of the illness and treatment.

3.4. COMPASS production system

To migrate COMPASS data processing to the grid environment, a set of following services running on the JINR cloud resources were deployed: MySQL database, PanDA, Auto Pilot Factory, Monitoring, CRIC information system and production system (ProdSys) management instance. All these services form the COMPASS production system [6]. It is based on the principles of the service-oriented architecture. Each service of the system is maximally isolated from the others, executed independently, and usually performs only one function, for example, sends jobs, checks their statuses, archives results, and so on.

4. JINR distributed information and computing environment

To combine resources for solving common tasks, as well as distribute a peak load across facilities of partner organizations from JINR Member States, a JINR distributed information and computing environment (DICE) was created.

The list of its participating organizations and their geographical locations are shown in Figure 3.



Figure 3. List of participating organizations and their geographical locations

The JINR DICE resources are configured to support the following virtual organizations (VOs): BM@N, MPD, Baikal-GVD. More than 15 000 Monte-Carlo simulation jobs for the BM@N VO were successfully completed. Apart from that, trainings for students and fellows from JINR Member State organizations are carried out on its resources.

A pie chart with the distribution of work done among the cloud resources of JINR DICE participants is shown in Figure 4.

4. Future plans

A certain number of services, for which the JINR cloud resources are used, would gain a sufficient performance boost in case of enabling the SSD-based caching pool in ceph storage, on which images of JINR cloud VMs are running. Therefore, to make this possible, it is planned to add several so-called “all flash” caching servers.

As for JINR DICE, in addition to extending a number of its participating organizations, it is planned to adopt a set of experiment workflows to run on their resources, as well as attract individual scientists and a small research group from JINR Member States.

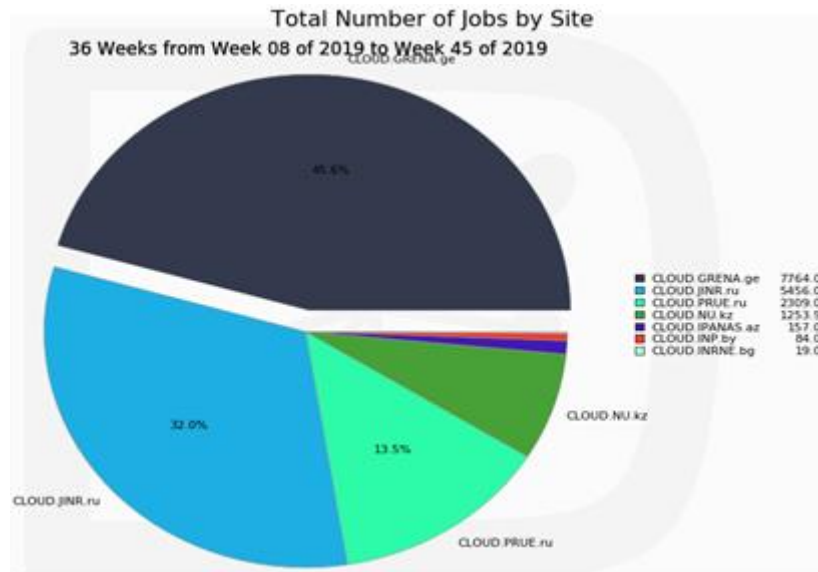


Figure 4. Pie chart with the distribution of work done among the cloud resources of JINR DICE participants

5. Acknowledgments

The work related to the development and maintenance of the JINR SaaS service is supported by the Russian Science Foundation grant #18-71-10095.

References

- [1] A.V. Baranov, N.A. Balashov, A. N. Makhalkin, Ye. M. Mazhitova, N.A. Kutovskiy, R.N. Semenov New features of the JINR cloud, Proceedings of the 8th International Conference «Distributed Computing and Grid-technologies in Science and Education (GRID'2018)», CEUR Workshop Proceedings, ISSN:1613-0073, vol. 2267 (2018), pp. 257-261.
- [2] V. Korenkov, A. Dolbilov et al, The JINR distributed computing environment, EPJ Web of Conferences 214, 03009 (2019), <https://doi.org/10.1051/epjconf/201921403009>
- [3] N. Balashov et al, Service for parallel applications based on JINR cloud and HybriLIT resources, EPJ Web of Conferences 214, 07012 (2019), <https://doi.org/10.1051/epjconf/201921407012>
- [4] A. Uzhinskiy, G. Ososkov, M. Frontasyeva Data Management System of the UNECE ICP Vegetation. Advances in Ecological and Environmental Research (ISSN 2517-9454). Vol.4. Issue 6, (2019), pp. 168-176.
- [5] P. Goncharov, G. Ososkov, A. Nechaevskiy, A. Uzhinskiy, Architecture and basic principles of the multifunctional platform for plant disease detection, Proceedings of the 8th International Conference «Distributed Computing and Grid technologies in Science and Education (GRID'2018)», CEUR Workshop Proceedings, , ISSN:1613-0073, vol. 2267(2018), pp. 200-206.
- [6] A. Petrosyan COMPASS Production system overview, EPJ Web of Conferences 214, 03039 (2019), <https://doi.org/10.1051/epjconf/201921403039>