INP ALMATY CLOUD AND ITS INTEGRATION INTO THE JINR DICE

N. Balashov[a], N. Burtebayev[b,c], V. Korenkov[a], N. Kutovskiy[a], A. Makhalkin[a], Ye. Mazhitova[a,b,] *, I. Pelevanyuk[a], I. Satyshev[a,b], R. Semenov[a,]

[a]Joint Institute for Nuclear Research, Dubna, Russia

e-mail: emazhitova@jinr.ru

[b]Institute of Nuclear Physics, Almaty, Kazakhstan

[c]Al-Farabi Kazakh National University, Almaty, Kazakhstan

Abstract – Nowadays, the successful implementation of a significant part of scientific projects involves the use of a distributed information and computing environment (DICE) for storing, processing and analyzing data. The DICE initiative of the Joint Institute for Nuclear Research (JINR) is dedicated to the creation, support and development of such an environment by combining the resources of educational and research organizations of the JINR Member States. One of such organizations is the Institute of Nuclear Physics (INP) in Almaty (Kazakhstan). Based on INP resources, a cloud infrastructure was installed, the necessary services were deployed; the INP cloud was integrated into the JINR DICE. A group of pioneer scientists started to use INP cloud resources to conduct research in their areas. The work is presently focused on user support.

INTRODUCTION

Currently, the volumes of information received are increasing significantly, new methods of data processing, storage and transmission are emerging. Cloud technologies and distributed systems, together with the latest software developments and modern methods of data analysis, allow combining data of different types, processing and classifying huge volumes of data, as well as reaching a new level of processing and analysis, which is basically impossible without using these technologies. Today, a lot of scientific projects would not be possible without using a distributed data processing infrastructure. Based on this, it was planned to create an INP cloud and integrate it into the distributed information and computing environment of the JINR Member States using the DIRAC platform [1, 2].

INP CLOUD INFRASTRUCTURE

From a variety of cloud service models, such as software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS), the last one enables the on-demand deployment of computing power, as well as transparent control over the allocation and use of the provided resources. To create the INP cloud infrastructure, the same open source solution, OpenNebula, was chosen, it is already used in other JINR DICE clouds. The utilization of the same software stack sufficiently simplifies infrastructure maintenance. The OpenNebula product is a cloud-computing platform for managing heterogeneous distributed data center infrastructures. This platform manages the virtual infrastructure of the data center to create private, public and hybrid infrastructure-as-a-service implementations. OpenNebula enables cloud deployments by implementing an IaaS model that provides the ability to use a cloud infrastructure to independently manage computational, storage, networking and other types of resources, as well as to provide a hybrid scheme by combining the resources of a local data center and external cloud providers [3, 4].

It should be noted that in 2020-2022, a cloud infrastructure was already deployed in Kazakhstan at the Astana branch of the Institute of Nuclear Physics, however, due to a number of difficulties, it was decided to stop the development of this cloud, to transfer the existing hardware resources to the parent organization in Almaty and start deploying the cloud infrastructure anew. The INP cloud infrastructure at the Astana branch was implemented in the form of two main components: cloud front-end node, i.e., one virtual host (virtual machine), which contained the cloud core and interfaces for interacting with users and services; four physical servers as working nodes running users' virtual machines and having the following hardware characteristics: 4 x CPU E5-2660 v4 2.00GHz; RAM256GB; 2 x HDD SAS 600GB in RAID1 mode (mirroring); 2 x 10G + 4 x 1G + 2 x FC.

To expand the INP cloud infrastructure, additional equipment was purchased at the end of 2022 (four physical Dell R640 servers, D-Link configurable L2+ switch with 24 10GBase-T ports and four 25GBase-X SFP28 ports, MikroTik industrial router, UPS HP Enterprise R/T3000 3000VA, Rack/Tower 2U, Gen5).

It was also decided to add one of the necessary elements of infrastructure maintenance, namely, a monitoring system. Several years ago, a custom collector for aggregating various OpenNebula-based cloud metrics was developed at JINR [5]. It can store the collected data in the Prometheus time series database via the node_exporter component. Using the Thanos sidecar, the collected metrics are sent to the ceph object storage. Grafana is used for data visualization. It requests data from the object storage via Thanos Querier and Thanos Store [6-9]. Such a collector was installed in the INP cloud infrastructure.

At the time of writing this article, the INP cloud infrastructure has the following hardware resources: the number of central processor cores is 90, the total amount of random access memory is 944 GB, the total ceph-based storage capacity is 6.96 TB. Pioneer users were defined. Their hardware and software needs were identified. A set of training events for

OpenNebula-based cloud usage were conducted. Since then, INP cloud resources are used to solve various scientific tasks. For now, 12 users are registered in the INP cloud. 11 virtual machines occupying a total of 55 CPU cores, 227.5 GB of RAM and 1.9 TB of disk storage are deployed.

The representatives of the following organizations use the INP cloud infrastructure:

- Institute of Nuclear Physics (Almaty, Kazakhstan),
- Joint Institute for Nuclear Research (Dubna, Russia),
- Al-Farabi Kazakh National University (Almaty, Kazakhstan),
- L.N. Gumilyov Eurasian National University (Astana, Kazakhstan),
- Institute of Physics and Technology (Almaty, Kazakhstan).

The following tasks are solved:

- SPD (Spin Physics Detector), MPD (Multi-Purpose Detector), BM@N (Baryonic Matter at Nuclotron) experiments of the NICA megascience project,
- OLVE-HERO, the High Energy Cosmic Ray Observatory studies cosmic radiation in the "knee" region of 1012 - 1016 eV,
- Tasks in the field of cosmic ray physics and high and ultra-high energy physics, in particular CORSIKA (COsmicRaySImulationsforKAscade, a Monte Carlo tool designed to analyze the properties and evolution of widespread air showers).

Moreover, virtual machines are deployed in the cloud infrastructure to test the functioning of various software. In addition to the utilization of INP cloud resources by local users, the integration of the cloud into the JINR DICE infrastructure was performed. The JINR DICE was created to combine computational power for solving common scientific tasks, to distribute peak loads across the participants, as well as to disseminate knowledge and practical skills in cloud and grid technologies among users and administrators of the JINR Member States' research and educational organizations. INP is one of these organizations. All the listed above pioneers' studies are so far performed on INP cloud resources only. The web interface of the INP cloud is available at the following URL: https://cloud.inp.kz/. Further work assumes an evaluation of user workflows and software to be run on the other resources of the JINR DICE infrastructure.

The cloud infrastructures of JINR and its Member States' organizations are also based on the OpenNebula open source solution. The JINR cloud is the core of this infrastructure. It hosts DIRAC services that manage computational tasks and data over the resources of JINR and its Member States' organizations. DIRAC plays the role of a connecting layer that transforms computing resources (computational grids, clouds or clusters) distributed around the world into a unified environment, providing users with a single entry point for data management, processing and analysis [10]. The integration of the INP cloud infrastructure into the JINR DICE allows using INP cloud resources for high-throughput computing workload.

The software is distributed to remote resources using the CernVM file system (CVMFS). It is a web-oriented global file system with version control optimized for software distribution [11]. The content of the file system is installed on a central web server, from where they can be mirrored and cached by other web servers and web proxies. File system clients load data and metadata on request and cache it locally. Cryptographic hashes and digital signatures ensure data integrity and authenticity.

For the more stable operation of CVMFS, the Frontier-squid software package, which is a corrected version of the standard squid HTTP-proxy caching software, is installed [12]. The frontier-squid package contains default settings and bug fixes that work well with applications used over the network.

CONCLUSION

Work to transfer the cloud infrastructure (including equipment) from the Astana branch of the Institute of Nuclear Physics to the parent organization in Almaty, as well as to increase the capacity of cloud resources, was carried out. The INP cloud infrastructure running on the OpenNebula cloud platform was deployed, tested and put into operation. It became a part of the JINR DICE infrastructure. The CernVM file system was chosen as a tool for distributing the software across JINR DICE resources. A group of pioneer scientists started using INP cloud resources to conduct research in their areas. The work is presently focused on software maintenance, service launches, failure cloud recovery, as well as user support, which includes conducting training events on working with the OpenNebula web interface, assistance in installing the necessary software products (including graphical interfaces) to solve user problems on virtual machines. Another direction of further work is to evaluate user workflows and software to be run on the other resources of the JINR DICE infrastructure.

ACKNOWLEDGEMENT

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant # BR10965191 “Complex research in nuclear and radiation physics, high-energy physics and cosmology for the development of competitive technologies”).

REFERENCES

1. DIRAC web portal. URL: https://dirac.readthedocs.io/en/latest/ (accessed November 13, 2024).
2. Balashov N., Burtebayev N., Korenkov V., Kutovskiy N., Mazhitova Ye., Satyshev I., Semenov R. CERN-JINR-INP-KazNU Data Center: Current Status and Plans // Physics of Particles and Nuclei Letters. – 2022. – Vol. 19, No. 5. – P. 547–549.
3. OpenNebula web portal. URL: https://opennebula.io/ (accessed November 13, 2024).
4. Milojicic D., Llorente I.M., Montero R.S. “OpenNebula: A cloud management tool”. IEEE Internet Computing, Vol.15, March 2011.
5. Balashov N.A., Kuprikov I.S., Kutovskiy N.A., Makhalkin A.N., Mazhitova Ye., Pelevanyuk I.S., Semenov R.N. JINR distributed information and computing environment: participants, features and challenges // CEUR Workshop Proceedings, ISSN:1613-0073, 2021. – Vol. 3041. – P. 280-284.
6. Prometheus web portal. URL: https://prometheus.io (accessed November 13, 2024).
7. Prometheus node_exporter github page. URL: https://github.com/prometheus/node_exporter (accessed November 13, 2024).
8. Thanos sidecar github page. URL: https://github.com/thanos-io/thanos (accessed November 13, 2024).
9. Grafana web portal. URL: https://grafana.com/grafana/ (accessed November 13, 2024).
10. Balashov N.A., Kutovskiy N.A., Makhalkin A.N., Mazhitova Ye., Pelevanyuk I.S., Semenov R.N. Distributed information and computing infrastructure of JINR member states’ organizations // AIP Conference Proceedings, ISBN: 978-0-7354-4132-3, 2021. – Vol. 2377. – 040001.
11. CernVM File System web portal. URL: https://cernvm.cern.ch/fs/ (accessed November 13, 2024).
12. Frontier-squid cache server-installing page. URL: https://twiki.cern.ch/twiki/bin/view/Frontier/InstallSquid (accessed November 13, 2024).