QUANTITATIVE AND QUALITATIVE CHANGES IN THE JINR CLOUD INFRASTRUCTURE

N.A. Balashov[1], I.S. Kuprikov[2], N.A. Kutovskiy[1,a], A.N. Makhalkin[1], Ye. Mazhitova[1,3], R.N. Semenov[1,4]

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

[2] *Dubna State University, 19 Universitetskaya str., Dubna, Moscow region, 141980, Russia*

[3] *Institute of Nuclear Physics, 050032, 1 Ibragimova str., Almaty, Kazakhstan*

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

E-mail: [a] kut@jinr.ru

The high demand for JINR cloud resources facilitated its sufficient growth. It triggered changes that needed to be made to overcome encountered problems and keep QoS for users: the main part of computational resources was reorganized as pre-deployed worker nodes of the HTCondor-based computing element to decrease the load on OpenNebula services during mass jobs submission; new SSD-based ceph pool for RBD VM disks with strong disk I/O requirements; dedicated ceph-based storage for the NOvA experiment; reorganized from scratch "Infrastructure-as-a-Code" approach based on a role and profile model implemented with the help of foreman and puppet; migration to a prometheus-based software stack for resource monitoring and accounting, and some other changes.

Keywords: cloud computing, OpenNebula, clouds integration, cloud bursting, DIRAC, ceph

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

1. Introduction

The JINR cloud is one of the components of the Multifunctional Information and Computing Complex (MICC) [1] hosted at the Meshcheryakov Laboratory of Information Technologies. The JINR cloud has been actively developed during last few years. The amount of its resources grew substantially. In this regard, some changes in the service architecture were made. A set of services hosted in the JINR cloud was expanded.

2. Changes in the virtualization part

Over the past few years (since the Grid'2018 conference), the amount of JINR cloud [2] resources has sufficiently increased. There are now 176 servers for virtual machines (VMs) in total (+96 servers since Grid'2018 contributed by the JUNO and NOvA experiments – 90 and 6 servers respectively), 5,044 CPU non-hyperthreaded cores (+3400) and 60 TB of RAM (+52 TB).The amount of RAM per CPU core varies from 5.3 GB up to 16 GB.

Thousands of jobs simultaneously running on the increased amount of cloud resources led to the saturation of the 10 Gbps network link, which connected the cloud to the JINR backbone (Fig. 13). It resulted in the misbehavior of services deployed in the JINR cloud. Some measures were taken, including switching to faster network equipment with a higher bandwidth.
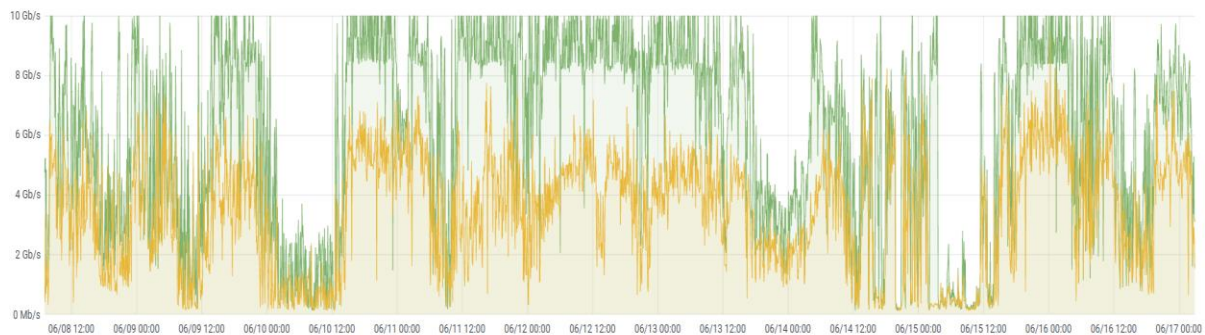

Figure 13. 10 Gbps network link load from 6 to 17 June 2021

Due to the lack of manpower to support the OpenVZ driver for the OpenNebula platform, on which the JINR cloud is based, it was decided to drop support for the driver in the JINR cloud. All users' OpenVZ containers were migrated to KVM-based VMs. Most of the KVM VMs have disks as block devices in the ceph-based software-defined storage.

3. Changes in the storage part

In addition to the general-purpose ceph-based storage with a total raw capacity of 1.1 PiB, two new storage elements (SEs) were deployed: the first is for NOvA experiment needs only, and the second is the pure SSD-based ceph storage for a set of production services and users with high demands in terms of disk I/O. The main parameters of all these cloud SEs are listed in Table 1.

Table 6. Cloud storage elements characteristics

Name	Disk type	Consumers	Ceph version	Raw capacity, PiB	Replication	Connectivity
Regular cloud storage	HDD	all	14.2.21	1.1	3x	2x10GBase-T
NOvA storage	HDD	NOvA exp.	15.2.11	1.5	3x	2x10GBase-T
Fast cloud storage	SSD	High disk I/O users	15.2.13	0.419	3x	4x10GBase-T (bonding) + 2x100Gbps – to be connected

4. Monitoring, alerting and accounting

JINR cloud servers and some of its services are monitored by the Nagios software [3]. Apart from that, JINR cloud metrics are gathered via a custom collector developed several years ago at JINR. It stores collected data into the InfluxDB time series database (TSDB). Moreover, metrics from all cloud ceph servers are aggregated using ceph prometheus plugins and the Prometheus TSDB [4]. As one can see, the set of software components used for JINR cloud metrics collection is quite wide. Keeping it consistent and up to date takes some effort. To reduce it, it was decided to switch to a prometheus-centric software stack: node_exporters were deployed on all cloud servers to provide Prometheus scrapers with servers state data. The OpenNebula collector was modified to be capable to provide metrics for the Prometheus TSDB as well (except InfluxDB). Alerting is implemented at the Prometheus level. Grafana is used for data visualization.

JINR cloud accounting is based on data gathered by the OpenNebula collector. Its visualization is performed using grafana dashboards (an example is given in Fig. 14).
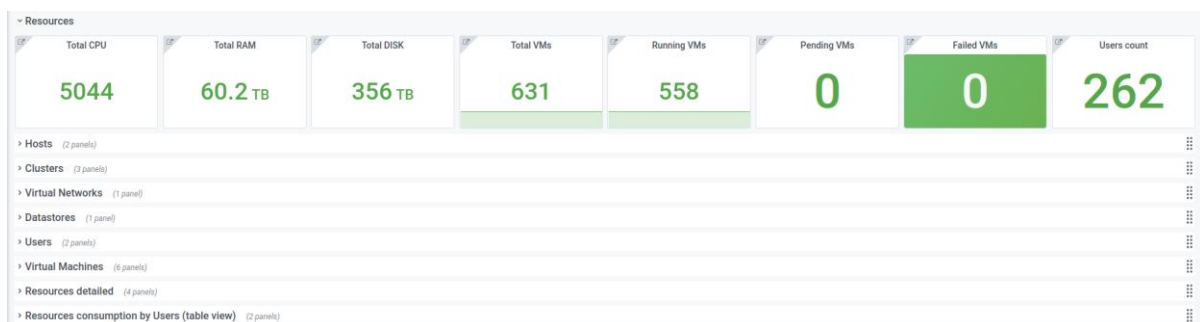

Figure 14. Grafana dashboard with JINR cloud metrics

Another source of information about the JINR cloud is OpenNebula log files. They are collected with the help of filebeat [5] and sent to ElasticSearch via logstash. Data visualization is carried out in Kibana [6].

5. Infrastructure management and hardware inventory

The JINR cloud is managed following the "Infrastructure-as-a-Code" approach when host provisioning and management are done through machine-readable definition files. For this, the Foreman [7] and Puppet [8] software are used. All host definition files (called "manifests" in terms of the puppet software) implement the role and profile method. Manifests are stored in the JINR git control versioning system [9]. Sensitive information (RSA/DSA keys, passwords, etc.) is kept in HashiCorp Vault [10].

The hardware inventory of the JINR cloud is performed on the basis of the iTop software [11], which is Information Technology Service Management (ITSM) and IT Infrastructure Library (ITIL). All data on the JINR cloud hardware (server vendors, models, locations, statuses, RAID controllers,

installed network cards, spare parts, used and free IP addresses, and much more) is kept in the iTop-based service.

6. Resource utilization

JINR cloud resources are used to solve a wide range of tasks: various services for the NOvA/DUNE, Baikal-GVD, JUNO, Daya Bay experiments; COMPASS production system [12] components; UNECE ICP Vegetation data management system; service for the disease detection of agricultural crops using advanced machine learning approaches; service for scientific and engineering computations; Grafana-based service for data visualization; jupyterhub head and execute nodes for it; gitlab and its runners, as well as a few others. Brief information about some of them is provided below.

6.1. Neutrino computing platform (NCP)

Within the cooperation of the Dzhelepov Laboratory of Nuclear Problems (DNLP) and the Meshcheryakov Laboratory of Information Technologies (MLIT), a computing platform for neutrino experiments (NCP) was created. It consists of a set of HTCondor-based services (submit nodes, cluster manager, worker nodes, computing elements) for the NOvA, DUNE and JUNO experiments, several general-purpose interactive virtual machines, a forum for the Baikal-GVD experiment. At the time of writing this article, 2,000 CPU cores for JUNO users and 1,020 CPU cores for NOvA and DUNE users are exposed via HTCondor-CE.

To optimize the utilization of NCP resources, it was proposed to share them among the most resource-consuming DLNP neutrino experiments (NOvA, DUNE, JUNO and Baikal-GVD). The Cloud Meta-Scheduler (CMSched) [13] is intended to implement such sharing by the dynamic scaling of the HTCondor cluster on demand. The CMSched prototype was deployed and is now under testing.

6.2. Service for scientific and engineering computations

A service for scientific and engineering computations [14] was developed to simplify the usage of JINR MICC resources by providing scientists with an intuitive web interface to run computational jobs. The list of supported applications was extended sufficiently and now includes the following:

1) Long Josephson junction (JJ) stack simulation,

2) Superconductor-Ferromagnetic-Superconductor Josephson junction simulation,

3) Annular Array of JJs average,

4) Long Josephson junction coupled with a ferromagnetic thin film,

5) Stack of short JJs,

6) Stack of short JJs with LC shunting.

7. Conclusion

JINR cloud resources are growing, as is the number of its users. Most of the hardware contribution to the JINR cloud is made by neutrino experiments. Quantitative changes entail changes in the architecture: splitting ceph storage into several instances, ceph with SSD disks for VMs sensitive to disk I/O, network upgrade. The migration from nagios-based monitoring to prometheus-based one is in progress. There is ongoing work to enhance the degree of automation of the provisioning and management of JINR cloud servers, as well as of deployed services, by adding more profiles and roles in the foreman and puppet systems.

8. Acknowledgment

References

[1] A.G. Dolbilov et al., Multifunctional Information and Computing Complex of JINR: Status and Perspectives // Proc. of the 27th International Symposium NEC'2019, Budva, Montenegro, Vol. 2507, 2019, pp. 16-22

[2] Balashov N.A. et al., Present Status and Main Directions of the JINR Cloud Development // Proceedings of the 27th International Symposium Nuclear Electronics and Computing (NEC'2019), CEUR Workshop Proceedings, ISSN:1613-0073, vol. 2507 (2019), pp. 185-189

[3] Nagios monitoring software official web portal. Available at: https://www.nagios.org (accessed 03.09.2021)

[4] Prometheus web portal. Available at: https://prometheus.io (accessed 03.09.2021)

[5] Home page of the filebeat component of the ElasticSearch software stack. Available at: https://www.elastic.co/beats/filebeat (accessed 06.09.2021)

[6] Balashov N.A. et al., Using ELK Stack for Event Log Acquisition and Analysis // Modern Information Technologies and IT Education, Vol. 17, #1, 2021, ISSN: 2411-1473, pp. 125-134. DOI 10.25559/SITITO.17.202101.731

[7] Foreman software web portal. Available at: https://theforeman.org (accessed 06.09.2021)

[8] Puppet software web portal. Available at: https://puppet.com (accessed 06.09.2021)

[9] JINR git portal. Available at: https://git.jinr.ru (accessed 06.09.2021)

[10] HashiCorp Vault web portal. Available at: https://www.vaultproject.io (accessed 06.09.2021)

[11] iTop software web portal. Available at: https://www.combodo.com/itop-193 (accessed 06.09.2021)

[12] Petrosyan A., COMPASS Production System Overview // EPJ Web of Conferences, Vol. 214, 2019. DOI: 10.1051/epjconf/201921403039

[13] N. Balashov, N. Kutovskiy, N. Tsegelnik, Resource Management in Private Multi-Service Cloud Environments // to be appeared in the proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2021), Dubna, Russia, July 5-9, 2021

[14] N. Balashov et al., JINR Cloud Service for Scientific and Engineering Computations // Modern Information Technologies and IT Education, Vol. 14 (1), 2018, pp. 57-68. DOI: 10.25559/SITITO.14.201801.061-072