NEW FEATURES OF THE JINR CLOUD

A.V. Baranov [1], N.A. Balashov [1], A. N. Makhalkin [1], Ye. M. Mazhitova [1,2], N.A. Kutovskiy [1,a], R.N. Semenov [1,3]

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

[2] *Institute of Nuclear Physics, 050032, 1 Ibragimova street, Almaty, Kazakhstan*

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

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

The report covers details on such aspects of the JINR cloud development as migration to high availability setup based on Raft consensus algorithm, Ceph-based storage back-end for VM images and DIRAC-based grid platform for external partner clouds integration into distributed computational cloud environment.

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

1. New high availability setup based on Raft consensus algorithm

Since OpenNebula release 5.4 which the JINR cloud is running on there is a new built-in mechanism for high availability (HA) setup based on so called Raft consensus algorithm [1]. According to OpenNebula documentation [2] a consensus algorithm relies on two concepts:

- System State what in the case of OpenNebula-based clouds means the data stored in the database tables (users, ACLs, or the VMs in the system);
- Log what is a sequence of SQL statements that are consistently applied to the OpenNebula DB in all servers to evolve the system state.

To preserve a consistent view of the system across servers, modifications to system state are performed through a special node called the "leader". The OpenNebula cloud front-end nodes (CFNs) elect a single node to be the leader. The leader periodically sends heartbeats to the other CFNs called followers to keep its leadership. If a leader fails to send the heartbeat followers promote to candidates and start a new election.

Whenever the system is modified (e.g. a new VM is added to the cluster), the leader updates the log and replicates the entry in a majority of followers before actually writing it to the database. It increases the latency of DB operations but enables a safe replication of the system state and the cluster can continue its operation in case of leader failure.

So during a software upgrade on the JINR cloud from OpenNebula 4.12 (see [2] for more details on that architecture) to 5.4 the HA setup based on the Raft consensus algorithm was implemented. Following the OpenNebula documentation recommendations the JINR cloud has odd number of front-end nodes (it equals three in our case). They are represented on the

Figure identically to one marked by the black numeral "2" in the same color square.

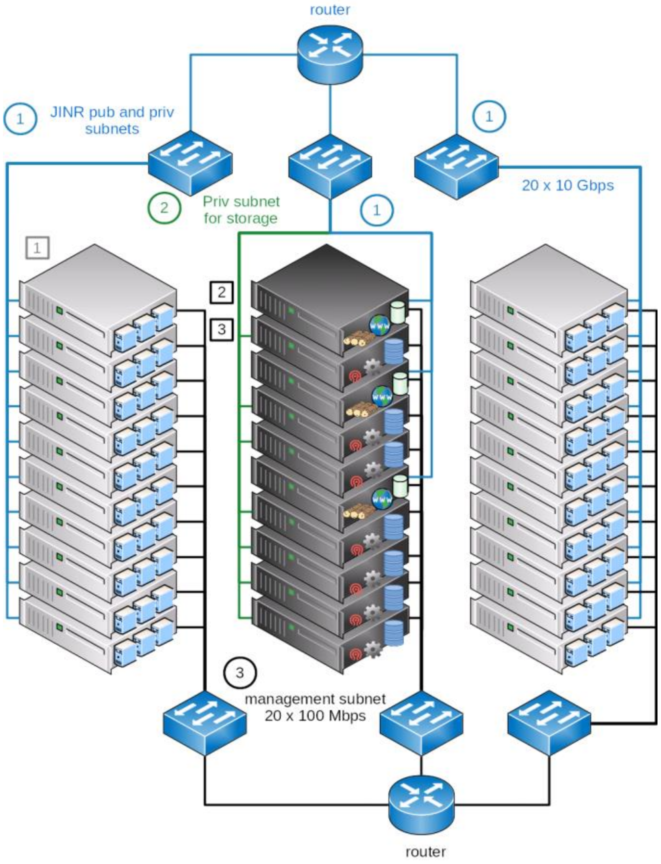

Figure 1. A schema of high availability setup of the JINR cloud based on the Raft consensus algorithm

KVM-based virtual machines (VMs) and OpenVZ-based containers (CTs) are running on cloud worker nodes (CWNs) marked on the

Figure by numeral “1” in a grey square.

All CFNs and CWNs are connected through 10 GbE network interfaces to the corresponding rack switch which in its turn are connected to the router and the JINR network backbone.

At the moment of writing that article the JINR cloud has 1600 CPU cores and 8.1 TB of RAM.

2. Ceph-based software defined storage

VMs and CTs images as well as a user and scientific experiments data are kept in Ceph-based software defined storage (SDS). Its architecture is shown on the Figure 2.

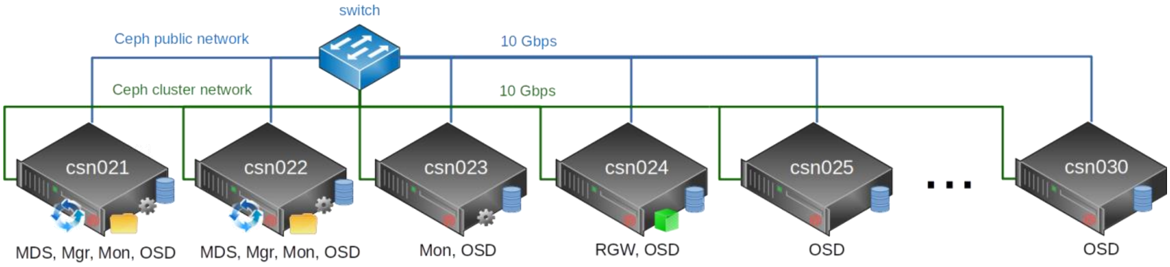

Figure 2 A schema of the Ceph-based software-defined storage deployed at JINR

Total amount of raw disk space in that SDS is about 1 PB. Due to triple replication an effective disk space available for users is about 330 TB. More details on the JINR Ceph-based SDS can be found in [3].

3. Clouds integration

Apart from increasing the JINR cloud resources by buying new servers and maintain them locally at JINR there is another activity on resources expansion: integration of part of computing resources of the partner organizations’ clouds.

Initially such integration was done with help of cloud bursting driver [5] developed by the JINR cloud team. But a growing number of participants of such distributed cloud-based infrastructure increases a complexity of its maintenance sufficiently (every new cloud integration requires changes in configuration files of every integrated cloud as well as appropriate services restart). That’s why a research work was started to evaluate possible alternatives. Among existing software platforms for distributed computing and data management a DIRAC (Distributed Infrastructure with Remote Agent Control) [7] one was chosen because of the following reasons:

- it provides the whole needed functionality including both job and data management;
- cloud as a computational back-end support (although an appropriate plugin required some development);
- easier services deployment and maintenance in comparison with other platforms with similar functionality (e.g. EMI).

A schema of clouds integration using DIRAC grid middleware is shown on the Figure 2.

Such approach also allows to share resources of each cloud between external grid users and local non-grid users.

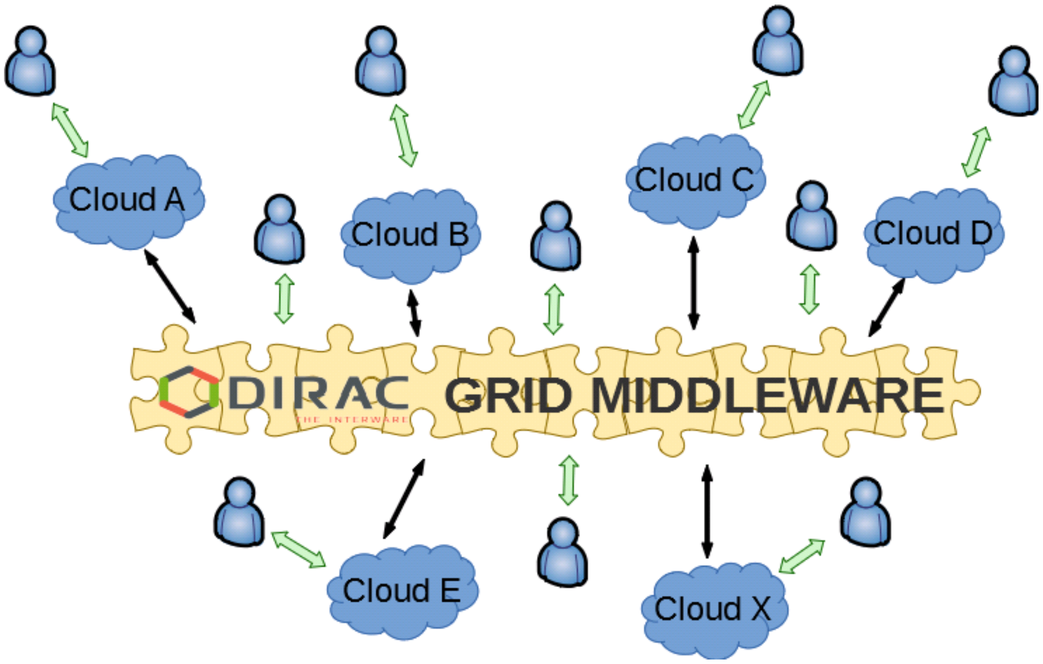

Figure 2. A scheme of clouds integration using DIRAC grid middleware

At the moment of writing that article the integration process of the clouds of the JINR Member State organizations into DIRAC-based distributed platform is at different stages, in particular (locations of such distributed cloud infrastructure participants are shown on the map on the Figure):

- Astana branch of the Institute of Nuclear Physics - Private establishment "Nazarbayev University Library and IT services" (Astana, Kazakhstan, integrated);
- Scientific Research Institute of Nuclear Problems of the Belarusian State University (Minsk, Belarus, integrated);
- Institute of Physics of the National Academy of Sciences of Azerbaijan (Baku, Azerbaijan, integrated);
- Yerevan Physical Institute (Yerevan, Armenia, integrated);
- Plekhanov Russian Economic University (Moscow, Russia, integrated);
- Institute for Nuclear Research and Nuclear Energy (Sofia, Bulgaria, negotiations);
- Georgian Technical University (Tbilisi, Georgia, in the process of integration);
- St. Sophia University "St. Kliment Ohridski" (Sofia, Bulgaria, negotiations);
- Institute of Nuclear Physics (Tashkent, Uzbekistan, negotiations);
- University "St. Kliment Ohridski" (Bitola, Macedonia, negotiations).

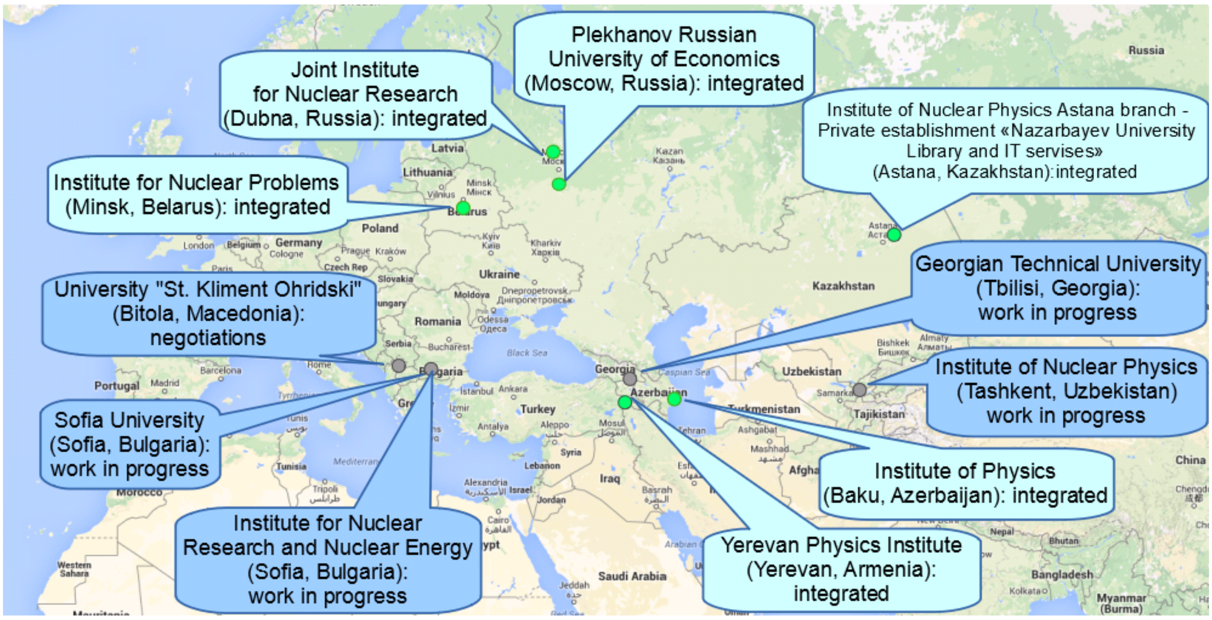
Plekhanov Russian University of Economics (Moscow, Russia): integrated
Joint Institute for Nuclear Research (Dubna, Russia): integrated
Institute of Nuclear Physics Astana branch - Private establishment «Nazarbayev University Library and IT servises» (Astana, Kazakhstan):integrated
Institute for Nuclear Problems (Minsk, Belarus): integrated
Georgian Technical University (Tbilisi, Georgia): work in progress
University "St. Kliment Ohridski" (Bitola, Macedonia): negotiations
Institute of Nuclear Physics (Tashkent, Uzbekistan) work in progress
Sofia University (Sofia, Bulgaria): work in progress
Institute of Physics (Baku, Azerbaijan): integrated
Institute for Nuclear Research and Nuclear Energy (Sofia, Bulgaria): work in progress
Yerevan Physics Institute (Yerevan, Armenia): integrated

Figure 4. Map with the distributed cloud infrastructure participants

6. Conclusion

The JINR cloud is rapidly developing. A demand in its resources as well as a spectrum of tasks it is used for is growing permanently. Such scientific experiments as JUNO, Daya Bay, Baikal-GVD started to use its resources.

The JINR cloud front-end node configuration was migrated from shared across two physical servers DRDB partition to front-end nodes HA architecture based on the Raft consensus algorithm.

The ceph-based software defined storage was put into production and now it is used for virtual appliances images as well as for keeping user and scientific experiments data.

References

[1] The Raft consensus algorithm web-site [Online]. Available: https://raft.github.io/. [Accessed on 06.11.2018]

[2] A. V. Baranov et al. JINR cloud infrastructure development // The 7th International Conference «Distributed Computing and Grid-technologies in Science and Education (Grid'2016)», CEUR Workshop Proceedings, ISSN: 1613-0073, vol. 1787, 2016, pp. 15–19.

[3] N.A. Balashov et al. JINR cloud service for scientific and engineering computations // Modern Information Technologies and IT-Education, ISSN 2411-1473, Vol. 14, No. 1, 2018, pp. 61-72.

[4] A. V. Baranov et al. Approaches to cloud infrastructures integration // Computer Research and Modeling, vol. 8, no. 3, 2016, pp. 583–590.

[5] DIRAC web-portal [Online]. Available: http://diracgrid.org [Accessed on 06.11.2018]