

10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)



Contribution ID: 338

Type: not specified

Extension of distributed computing infrastructure and services portfolio for research and educational activities

Friday, 7 July 2023 11:15 (15 minutes)

The importance of the development of computing infrastructure and services for support of research data accumulation, storage, and processing is permanently increasing. e-Infrastructures became more and more demanded and universal tool for support modern science that provide wide set of services for operation with research data, data collection, systematization and archiving, computing resources for complex data processing applications development, porting and execution.

Work on the implementation of distributed computing infrastructure in Moldova started in 2007 when the first Agreement on the creation of the MD-GIRD Joint Research Unit Consortium and accompanying Memorandums of Understanding were signed by seven universities and research institutes of Moldova. Since this time, the works started on the deployment of the national distributed computing infrastructure that included integration of computing clusters and servers deployed in the main national Universities and research institutions. For effective integration of different types of computing resources was into the common distributed infrastructure was used high-capacity communication backbone provided by NREN RENAM [1].

The common computing infrastructure continue to develop due to support of various international and national projects. This distributed infrastructure now unites three main datacenters located in the State University of Moldova (SUM), Vladimir Andrunachievich Institute of Mathematics and Computer Science (VA IMCS) and RENAM Association that are permanently developing and common computing resources now comprises more than 320 CPU cores, 2 NVIDIA T4 Tensor Core GPU units and 54 TB of storage resources [2]. The elaborated concept of the creation of the heterogenous computing infrastructure that includes multi-zone IaaS Cloud infrastructure (fig. 1), pool of virtualized servers that used for permanent resources allocation to execute production services, multiprocessor clusters and bare metal serves that used for running intensive data processing applications. The distributed infrastructure comprises dedicated storage sub-systems for large amount of data archiving and providing resources for the whole distributed infrastructure data backup.

In the first stage, it is planned to deploy a multi-zone IaaS Cloud infrastructure that combines the resources of VA IMCS, SUM and RENAM into distributed computing network for processing scientific data, performing intensive scientific calculations, as well as storing and archiving research data and results of computational experiments. Works on deployment of an updated Scientific multi-zone IaaS Cloud Infrastructure that is based on OpenStack Ussuri have begun in 2021 and are progressing now taking in account continuation of physical computing resources upgrading by installation of new servers' equipment in all three main datacenters. As a result, today in VA IMCS, SUM and RENAM in parallel are available and operating previously deployed resources, based on an of outdated OpenStack versions, updated Cloud platform, based on the OpenStack Ussuri version, and now is deploying Openstack version 2023.1 Antelope, which is currently the most recent stable release and will be actively maintained at least for the upcoming year, offering more features, more processing power, and flexibility of operation [3].

https://cloud.math.md/apps/files_sharing/publicpreview/ZJKysCqoKZ4X6df?x=1914&y=619&a=true&file=Figure%25201.png
Figure 1. Multi-zone Cloud: IMI –RENAM –SUM.

In the current distributed cloud understructure is already implemented useful and important components - block storage and Virtual eXtensible Local Area Network (VXLAN) traffic tagging. These tools will be deployed and used in the developing extended cloud infrastructure too. Block storage allows creation of volumes for organizing persistent storage. In general, in OpenStack, as in other modern Cloud systems, several concepts exist for providing storage resources. When creating a virtual machine, you can choose a predefined

flavor, with a predefined number of CPU, RAM, and HDD space; but previously, when you delete a virtual machine, all data stored on the machine instantly disappears. The Block storage component, used in the created multi-zone IaaS Cloud Infrastructure, is deployed on a separate storage sub-system and allows to create block storage devices and mount them on a virtual machine through special drivers over the network. This is a kind of network flash drive that can be mounted to any virtual machine associated with the project, unmounted and remounted to another, etc., and most importantly, this type of volume is persistent storage that can be reused when the virtual machines are deleted. Thus, you can no longer worry about data safety and easily move data from one virtual machine to another, or quickly scale up VM performance by creating a virtual machine with larger resources and simply mount volumes to it with all scientific data available for further processing.

VXLAN is a more advanced and flexible model of interaction with the network. In the upgraded cloud infrastructure, in addition to the usual “provider network” model, which allocates one real IP address from the pool of provider network addresses to each virtual machine, a self-service network is also available. A self-service network allows each project to create its own local network with Internet access via NAT (Network Address Translation). For a Self-service network, the user can create a virtual router for the project with its own address space for the local network. VXLAN traffic tagging is used to create such overlay networks that prevent the occurrence of address conflicts between projects in case several projects will use network addresses from the same range. To ensure the functioning of NAT, one IP address from the provider network is allocated to the external interface of the virtual router, which serves as a gateway for virtual machines within the project. Also, when using the self-service model, the floating IP technology becomes available, which allows you to temporarily bind the IP address from the provider network to any of the virtual machines in the project, and at any time detach it and reassign it to any other virtual machine of the project. Moreover, the replacement occurs seamlessly, that is, the address does not change inside the machine, but remains the same - the address is from the internal network of the project, but the changes occur at the level of the virtual router. Incoming to the external address packets are forwarded by the virtual router to the internal interface of the selected virtual machine. This allows you efficiently to use IP addresses and not allocate an external address to each virtual machine. The external IP address remains assigned to the project and can be reused by other machines within the project.

For the deployment of new computing infrastructure, the process of transition to a 10G network has started according to the elaborated plan. The New Juniper switch already has been installed and all storage servers with 10G cards on board have been connected to this switch. Now is realizing procedure of switching connection of all remaining servers to N x 10G interfaces [4].

Distributed computing infrastructure providing for uses of national research and educational community the following production services, software platform and tools:

- Jupyter Notebook - is a web-based interactive computing platform. The notebook combines live code, equations, narrative text, visualizations. Jupyter Notebook allows users to compile all aspects of a data project in one place making it easier to show the entire process of a project to your intended audience. Through the web-based application, users can create data visualizations and other components of a project to share with others via the platform.
- BigBlueButton is a purpose-built virtual classroom that empowers teachers to teach and learners to learn.
- TensorFlow 2 - an end-to-end open source machine learning platform for everyone
- Keras: Deep Learning for humans. Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.
- Anaconda Distribution: equips individuals to easily search and install thousands of Python/R packages and access a vast library of community content and support.
- The Apache Tomcat® software is an open-source implementation of the Java Servlet, JavaServer Pages, Java Expression Language and Java WebSocket technologies.
- Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Nextcloud is a self-hosted, open-source file sharing and collaboration platform that allows users to store, access, and share their data from any device or location.

Acknowledgment

This work was supported by “EU4Digital: Connecting research and education communities (EaPConect2)” project funded by the EU (grant contract ENI/2019/407-452) and Grants from the National Agency for Research and Development of Moldova (grant No. 20.80009.5007.22 and grant No. 21.70105.9SD).

References

1. P. Bogatencov, G. Secieru, N. Degteariov, N. Iliuha. Scientific Computing Infrastructure and Services in Moldova. In Journal “Physics of Particles and Nuclei Letters”, SpringerLink, 04 September 2016, DOI: 10.1134/S1547477116050125), <http://link.springer.com/journal/11497/13/5/page/2>
2. Petru Bogatencov, Grigore Secieru, Boris Hîncu, Nichita Degteariov. Development of computing infrastructure for support of Open Science in Moldova. In Workshop on Intelligent Information Systems (WIIS2021) Proceedings, Chisinau, IMI, 2021, pp. 34-45, ISBN 978-9975-68-415-6.

3. Petru Bogatencov, Grigore Secrieru, Radu Buzatu, Nichita Degteariov. Distributed computing infrastructure for complex applications development. In the Proceedings of Workshop on Intelligent Information Systems (WIIS2022), Chisinau, VA IMCS, 2022, pp. 55-65, ISBN 978-9975-68-461-3.
4. Grigore Secrieru, Peter Bogatencov, Nichita Degteariov. DEVELOPMENT OF EFFECTIVE ACCESS TO THE DISTRIBUTED SCIENTIFIC AND EDUCATIONAL E-INFRASTRUCTURE. Proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2021). Dubna, Russia, 5-9 July 2021; Vol-3041, urn:nbn:de:0074-3041-3, pp. 503-507; DOI:10.54546/MLIT.2021.21.73.001, ISSN 1613-0073

Summary

Primary author: SECRIERU, Grigore (Vasile)

Co-authors: DEGTEARIOV, Nichita (RENAM); BOGATENCOV, Peter (RENAM, Moldova)

Presenter: SECRIERU, Grigore (Vasile)

Session Classification: Cloud Technologies

Track Classification: Cloud Technologies