RESEARCH CLOUD COMPUTING ECOSYSTEM IN ARMENIA

H. Astsatryan[a], W. Narsisian, A. Mirzoyan, V. Sahakyan

Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia, 1, Paruyr Sevak str., Yerevan 0014, Armenia

E-mail: [a] hrach@sci.am

Growing needs for computational resources, data storage within higher-educational institutions and the requirement for a lot of investment and financial resources the idea or the concept of “National Research Cloud Platform (NRCP)” is crucial to provide necessary IT support for educational, research and development activities, which allow access to advanced IT infrastructure, data centers, and applications and protect sensitive information. In this article we will illustrate the concept of NRCP, background, deployment stages and architecture and finally some use cases.

Keywords: IaaS, NRCP, Openstack, ArmCloud, ArmCluster, ArmGrid, Earth science, Life Science, VM

Hrachya Astsatryan, Wahi Narsisian, Artashes Mirzoyan, Vladimir Sahakyan

1. Introduction

Virtualization transforms the IT industry landscape providing capabilities to run various virtual machines (VM) in the same hardware capacities, enhancing resource sharing and improving performance [1]. Low overhead costs for implementing this technology, high and constantly growing demand for computing resources and the need to provide more flexible services have led to the transition from the use of bare-metal servers and towards providing of virtualized resources (virtual machines, storage and even network infrastructures) that are easier to scale and provide a sufficient level of reliability. On the other hand, the Cloud computing environment has proven to be the base of these changes, which increased the Cloud services and computing resources requirement throughout scientific institutions and universities [2]. The term started use by Amazon Company in 2008. Later, this novel technology was developed and provided as a service by GAFA (Google, Apple, Facebook, and Amazon) and other public cloud providers [3]. The main approach of Public cloud providers is to deliver on-demand services through the Internet to anyone who registers and pays for the services. Instead of public clouds, the private cloud infrastructures built-in for a couple of institutions or companies to host the facilities on their side [4]. For instance, national research cloud platforms provide cloud services to the academic and research community on the top of the research and education networks. It is possible to combine public and private cloud deployment models to create a synergy, calling hybrid cloud. Usually, the resources of Public clouds are supported the elasticity of computational resources in case of the need.

Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) are the leading cloud computing service layers [5]. IaaS provides the infrastructure such as VMs and other resources like VM disk image library, block and file-based storage, firewalls, load balancers, IP addresses, or virtual local area networks. PaaS or platform as a service model delivers computing platforms typically including an OS, programming language execution environment, database, and web server. SaaS provides a new service delivery model to access application services over the web without worrying about installing, maintaining, or coding the software. The SaaS provider manages the software maintenance or setup. Therefore, the software is available to access and operate without downloading or installing any piece of design or OS.

In addition to the leading cloud computing service layers, a wide range of services can be provided by Cloud, with an extra layer of flexibility and scalability, such as provisioning high-demand virtual HPC resources [6]. The critical challenge of deploying such cloud services is the complexity and the cost to purchase and maintain the computing resources needing a lot of human efforts to keep all the services up to date and reliable. The cost is a significant limitation for developing countries, like in the case of Armenia. In 2018, the Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia (IIAP) started to realize the “National Research Cloud Platform (NRCP)” initiative. NRCP aims to deliver on-demand cost-effective cloud computing resources and services to the local institutions and research communities. The market analyzes with scientific communities and stakeholders aimed to find out the demands and complexity of scientific problems facing to solve, and gather the information related to the communities’ tools and packages. As a result, IIAP deployed user-oriented Cloud services to fulfill almost all types of demands in Armenia, ranging from general to domain-specific services. Section 2 represents the architecture and design of NRCP, while the conclusions and lessons learned follow in Section 3.

2. National research cloud platform

In a first stage, a federated cloud infrastructure in the Black Sea Region has been deployed, enabling user communities from participant countries (Armenia, Georgia, Moldova, and Romania) to join the local virtualized resources providing them with VMs, networks, and storages [7]. The federated infrastructure offers user communities to use local or remote resources and makes user communities’ regional collaboration easier. The federated cloud platform based on OpenNebula middleware address the regional problems that require large amounts of computational resources, even when the actual simulations happen not in the zone where the data is stored. In the next stage, the virtualization is widely implemented for the core services of the Armenian National Grid (ArmGrid) infrastructure providing on-demand access to a sustainable computing platform [8]. The ArmGrid infrastructure consists of seven Grid sites located in the leading research centers and universities of Armenia. The total number of processors at the Grid sites was approximately 450 CPU cores. Instead of a single system ArmCluster (Armenian Cluster), ArmGrid is an autonomous decentralized system with distributed job management and scheduling opportunities [9]. Finally, a hybrid research computing platform has been deployed combing HPC with Grid and Cloud Computing based on ArmCluster HPC cluster, resource sharing ArmGrid Grid, and on-demand service provisioning federated cloud infrastructures. Each infrastructure identifies rules for making up the resources and executing applications, like resource ownership and sizing, application portability, or resource allocation policy. Based, on these experimental infrastructures, a NRCP is suggested in Armenia aims to have better hardware utilization, increase the storage systems reliability and services management, and offer higher services with the virtualization support. The infrastructure provides VMs and networking services, consists of a cloud core service and a scheduler, APIs, databases, and nodes where VMs are running. The full virtualization on the KVM hypervisor for each computational node has been implemented. NRCP consists of three different Zones of Cloud resources, GPU resources, and a data lake [fig. 1]. Combining these three solutions under a single umbrella provides domains specific services with high-availability and scalable services. The NRCP is a critical element of the Armenian e-infrastructure [10], a complex national IT infrastructure consisting of both communication and distributed computing infrastructures. Most importantly, all the input and output data reside on the NRCP side to reduce the time to process the data and possibly is shared data between different scientific groups.

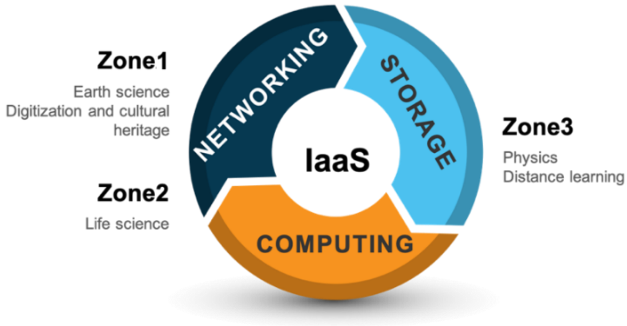

Figure 1. National research cloud platform

NRCP architecture design is mainly built on multiple Cloud controllers dedicated to different scientific communities to split the Cloud resources and the Cloud storage based on several scientific domains. The technical information of computational resources is generalized in Table 1.

Table 1. NRCP technical specification

Server type	**Q-ty**	**CPU/GPU model**	**Server parameters**			**Total cores**
			CPU/GPUs	**Cores**	**RAM (GB)**	
Thin	4	Intel Xeon E5-2630 v4	2	20	256	80
Fat	2	Intel Xeon Gold 6138	4	80	512	160
Accelerated	2	Intel Core i9-10900KF	1	10	128	20
	2	Intel Xeon E5-2680 v3	2	24	128	48
		Intel Xeon Phi 7120P	2	122		244
	2	Intel Xeon Gold 5218	2	32	192	64
		Nvidia V100 32GB	2	10240		20480
Total (cores)						**21096**

Therefore, NRCP provide compute services consist of 616 physical and 20480 GPU cores, about 3 terabytes of memory and 1620 terabytes for data storage. For instance, all Earth science production groups are consolidated under a single Zone with the same storage node, enabling them to share data if needed very easily and opens the door to a better collaboration possibility.

The OpenStack has been deployed and customized for NRCP providing orchestration needed to virtualize servers, storage, and networking. The Controller, compute, network, and storage components have been used for the deployment. Various flavors of VMs ranging from small instances such as 2 CPU and 2 GB RAM with 40 GB HDD, to a very huge instance with 128 CPU cores, 256 GB. The Data Lake provides a scalable and secure platform that allows all users to upload and download their data with a high-speed, process the data in real-time, use the data for different simulations, share the data between groups. For instance, in the domain of astrophysics, the infrastructure's core is the Armenian VO repository providing an advanced experimental platform for data archiving, extraction, acquisition, correlation, reduction, and use [11]. Another example is the Armenian Data Cube [12], a complete and up-to-date EO (Earth Observation) archive data (e.g., Landsat, Sentinel). EO, using precise and reliable data, is a critical element to address different environmental challenges, like water, soil, or plants.

For the last three years NRCP serves multiple scientific projects and communities. For instance, several domain specific services have been developed for earth science user community to address critical societal challenges, such as weather prediction, air quality monitoring and prediction, water quality and quantity monitoring, or earth observation [13-14].

3. Conclusion and lessons learned

The article summarizes the experiences gained so far, and highlighted a few scientific use cases, where the community is intensively using NRCP resources. Throughout the deployment and the implementation phases of NRCP deployment for diverse scientific communities with specific domain-oriented approaches, a list of recommendation has been collected:

- To consider the complete infrastructure and its capabilities before the deployment, boosting to choose the best possible options and tools satisfying the needs;
- To conduct benchmarks and experiments before putting the solution into production; to confirm the systems' reliability by handling different scenarios, even if the deployment of some packages needs to be done several times;
- To prepare a well-documented tutorial with exact details of all services and solutions, considering that not every user has an IT background when using the system;
- To conduct a training campaign with potential communities and explain the opportunities and challenges. It will help to understand the benefits of such solutions by boosting scientific experiments and simulations;
- To minimize the manual deployment as much as possible. For instance, it is planned to implement multiple bash scripts and Puppet automation;
- To maximize the overall resource usage in computing, networking, and storage resources considering the energy consumption minimization;
- To use federated identity authentication based on SAML 2.0 (Security Assertion Markup Language) to make easy user access.

It is planned to develop and provide user-specific high-level services, like SaaS solutions for all those communities, conducting the experiments without accessing the computing resources. Instead, the communities may use the browser to access any domain-specific service and run the experiment from it, further simplifying cloud resource usage. The OpenStack Ironic will be implemented for economic and most efficient use of computing resources focusing on HPC Cloud solutions provisioning based on completely virtual, bar-metal, and hybrid architectures. The future ultimate goal is the establishment of a National Open Science Cloud Initiative and its further integration with the European Open Science Cloud and European Research Infrastructures, like ELIXIR or Openaire.

References

[1] M.F. Mergen, V. Uhlig, O. Krieger, J. Xenidis. Virtualization for high-performance computing // ACM SIGOPS Operating Systems Review, 2006 Apr 1, vol. 40, issue 2, pp. 8-11

[2] M. Armbrust, et al. A view of cloud computing // Communications of the ACM, 2010 Apr 1, vol. 53, issue 4, pp. 50-58

[3] V. Chang, G. Wills, D. De Roure. A review of cloud business models and sustainability // In2010 IEEE 3rd International Conference on Cloud Computing 2010 Jul 5, pp. 43-50

[4] Y. Jadeja, K. Modi Cloud computing-concepts, architecture and challenges // In2012 International Conference on Computing, Electronics and Electrical Technologies, 2012 Mar 21, pp. 877-880

[5] S.K. Sowmya, P. Deepika, J. Naren. Layers of Cloud–IaaS, PaaS and SaaS: A Survey // International Journal of Computer Science and Information Technologies. 2014, vol. 5, issue 3, pp. 4477-4480

[6] R.R. Expósito, et al. Performance analysis of HPC applications in the cloud // Future Generation Computer Systems, 2013 Jan 1, vol. 29, issue 1, pp 218-229

[7] H. Astsatryan, , et al. Environmental science federated cloud platform in the bsec region // International Journal of Scientific & Engineering Research, 2014, vol. 1, issue 1, pp. 1130–1133

[8] H. Astsatryan, Yu. Shoukouryan, V. Sahakyan. Grid activities in Armenia // In Proceedings of the International Conference Parallel Computing Technologies. Novgorod, Russia, March, 2009

[9] H. Astsatryan, Yu Shoukourian, V. Sahakyan. Creation of High-Performance Computation Cluster and DataBases in Armenia // In Proceedings of the Second International Conference on Parallel Computations and Control Problems (PACO '2004), pp. 466–470, 2004

[10] H. Astsatryan, V. Sahakyan, Yu. Shoukourian, P.H. Cros, M. Dayde, J. Dongarra, P. Oster. Strengthening Compute and Data intensive Capacities of Armenia // IEEE Proceedings of 14th RoEduNet International Conference - Networking in Education and Research, Craiova, Romania, pp. 28-33, September 24-26 2015

[11] A.M. Mickaelian, et al. Ten Years of the Armenian Virtual Observatory // ASPC, vol. 505, issue 16, 2016

[12] Sh. Asmaryan, et al. Paving the way towards an armenian data cube // Data, vol. 4, issue. 3, pp. 1–10, 2019

[13] H. Astsatryan, et al. An interoperable cloud-based scientific GATEWAY for NDVI time series analysis // Elsevier Computer Standards & Interfaces, vol. 41, pages 79-84, 2015

[14] H. Astsatryan, W. Narsisian and Sh. Asmaryan. SWAT hydrological model as a DaaS cloud service // Springer Earth Science Informatics, vol. 9, issue 3, pages 401–407, 2016