

Smart Cloud Scheduler

**N.A. Balashov^{1,a}, A. V. Baranov¹, I.S. Kadochnikov¹, V.V. Korenkov^{1,2},
N.A. Kutovskiy^{1,2}, A.V. Nechaevskiy¹, I.S. Pelevanyuk¹**

¹ Joint Institute for Nuclear Research, 6 Joliot-Curie street, Dubna, Moscow region, 141980, Russia

² Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia

E-mail: ^abalashov@jinr.ru

The rapid development of cloud technologies has led to the partial shift of scientific computations to the Cloud computing model. At present, a lot of scientific centers establish their own private virtualized datacenters. A cloud infrastructure deployed at the Laboratory of Information Technologies of the Joint Institute for Nuclear Research (JINR) and based on Infrastructure as a Service (IaaS) model has been extensively used by the JINR engineers and researchers for the last few years. The authors of this paper carried out the analysis of the experience gained in usage and management of the JINR cloud service and revealed some weak spots in its functioning. We show that one of the key problems is an inefficient utilization of the cloud resources that may also be inherent to other scientific clouds. One possible way to cope with the problem is through dynamic reallocation and consolidation of virtual machines, and we also give an overview of the smart cloud scheduling method proposed by the authors earlier. This method allows one to release underloaded cloud resources and to put them back to the pool of free resources. This work presents a review of several possible strategies to reuse the freed cloud resources and propose a new strategy based on integration of the cloud service with a batch system (e.g. HTCCondor). This novel strategy provides a way for making the reuse of free but allocated resources more stable by means of the dynamically deployed additional virtual batch system worker-nodes which can be safely destroyed on demand releasing reused cloud resources.

Keywords: cloud computing, virtualization, self-organization, intelligent control, datacenters, VM consolidation

The work was supported by the RFBR grant 15-29-07027

© 2016 Nikita N. Balashov, Alexander V. Baranov, Ivan S. Kadochnikov, Vladimir V. Korenkov, Nikolay A. Kutovskiy, Andrey V. Nechaevskiy, Igor S. Pelevanyuk

1. Introduction

Modern cloud platforms are characterized by high elasticity and scalability and in some cases by its universality that led as a result to wide use in commercial area. In science application of cloud technologies continue gaining popularity and they are used in development of complex distributed computing systems for processing of big volumes of data [Timm et al., 2015] as well as in supporting information systems (IS) functioning and providing a universal computing resource to the end users, which gives a more simple access to computing resources needed for solving of their everyday tasks.

An example of one of such scientific cloud computing environments is a cloud service of the Joint Institute of Nuclear Research (JINR) [Baranov, Balashov, Kutovskiy, Semenov, 2016]. Accumulated for the last several years experience in working with the cloud computing environment was analyzed by the JINR experts and allowed to make a conclusion that computing resources of the JINR cloud service are used not efficiently enough. As a result a task was set to analyze the reasons of the low utilization efficiency and to develop methods and strategies to increasing the efficiency.

2. Problem description and its cause (analysis of cloud usage)

Complexity of modern IS and applications, ranging from long-running scientific simulations to transactional applications, makes the problem of assessment of the sufficient resources almost unsolvable and for this reason most of the users tend to stick to the same strategy when requesting computing resources: request as much resources as possible. Following this strategy inevitably leads to the inefficient use of computing resources from the point of view of those who provides these resources due to underutilization of the provided computing capacities.

JINR cloud service is a universal computing resource whose users can be conditionally split into 4 groups:

- x Software developers;
- x System administrators;
- x Physicists;
- x Automated systems.

Profile of computational load of virtual machines (VM) of software developers is characterized by high enough resources utilization during a short period of time. Generally it is a result of setting up system environment, installation of additional software and libraries that are required for the further work. This period generally lasts from one to several days and directly depends on the complexity of the project to be developed. In the following period VM is characterized by low resources utilization and periodic short-term load increases, sometimes reaching the peak load. Increased load during this period arise mostly during developed software builds, compilation and testing. Frequency of these loads directly depends on the intensity of the particular project development. By the end of the main software development cycle generally starts a period of decreasing load frequency proportionally to decreasing intensity of the project development process. Such virtual machines often live during the whole life-cycle of the project and are deleted only after development is finished or when an owner of the VM leaves the development team of the particular project.

VMs of the first two groups of users all have one common similarity - the out of hours VMs load drops significantly.

Profiles of VMs of the third group of users, physicists, is characterized by the long enough (sometimes up to month) time intervals having peak loads alternating with close to minimal ones (CPU utilization may drop down to 0%).

VMs of the fourth group of the JINR cloud users are the VMs automatically created by different IS and workload management systems (WMS) of various physical experiments. Currently these are the WMS of BES-III, ALICE and NOvA experiments.

Also we should mention virtual machines which host commissioned information systems, users of which potentially include all the employees of JINR - the workload in such VMs directly depends on popularity of the particular IS among its users.

It is clear that such usage of cloud resources leads to highly inefficient distribution of cloud resources and Figures 1 and 2 demonstrate that the problem remains with the growth of the JINR cloud service.

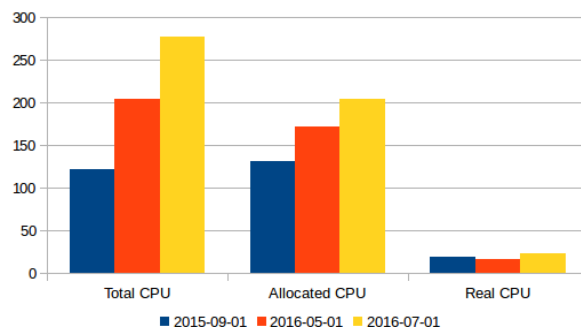


Fig. 1. JINR cloud CPU load

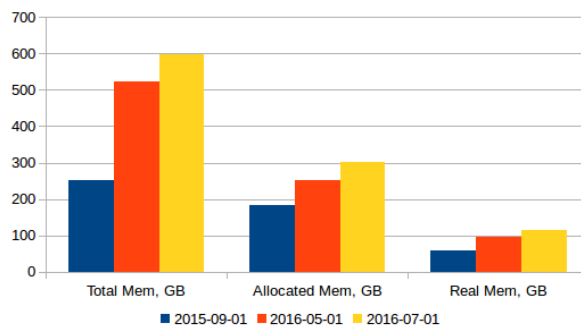


Fig. 2. JINR cloud memory load

3. Method of dynamic cloud resources reallocation

To solve the above-mentioned problem of ineffective utilization of JINR cloud resources the authors of the article proposed a method of dynamic virtual machines reallocation in cloud environment based on the analysis of historical records of the virtual machines resources consumption [Balashov, Baranov, Korenkov, 2016]. The method allows to release some part of the allocated but underutilized computing resources by consolidating virtual machines using overcommitment technologies. Released using this method resources are supposed to be brought back to the pool of free resources, the workload on which would be automatically reallocated by the built-in cloud scheduler according to its built-in algorithms. This approach makes it unnecessary to disable or modify in any way the built-in cloud scheduler of the deployed cloud platform, the only possible requirement on some platforms (eg. OpenStack) that should be met is disabled built-in overcommitment management.

Following this approach to the re-use of the released resources decreases reliability of the virtual machines functioning because of the deliberate allocation of a larger amount of virtual resources than the physical resources the cloud really has. To have control over reliability of virtual machines the method uses a system of ranks of real and virtual resources that is basically a management tool for controlling overcommitment rates of particular servers and virtual machines.

4. Re-use of computing resources

As a result of application of the suggested method arises a new problem of proper and accurate re-use of the released resources. One way is to return them back to the pool of free resources as it was suggested before, but it is not the only possible strategy. As an alternative solution it is possible to put servers that don't have any virtual machines running into low-power mode [Beloglazov, Buyya, 2015]. By doing so it is possible to significantly decrease power consumption. One more solution that can be used by scientific clouds (similar to the JINR cloud service) is by integrating the cloud with the batch processing systems eventually allowing to safely re-use freed cloud resources and to minimize the drop of reliability of the virtual machines functioning in the conditions of their highly consolidated distribution over servers. Some of the batch processing systems already have the ability to launch computing jobs to cloud infrastructures based on IaaS model, e.g. workload management system DIRAC [McNab, Stagni, Luzzi, 2015]. At the same time there are developments of complementary systems to enable suchlike functionality for some other batch systems, e.g. HTCCondor, that don't have it by default (examples of such developments include VCondor [IHEP-CC, 2016] and Vcycle [McNab, Love, MacMahon, 2015]).

At present virtual computing cluster based on HTCCondor system is being deployed at JINR (needed to process computing job of the NOvA neutrino experiment), that would be able to re-use spare resources of the JINR cloud service using the described above approach.

HTCCondor is a batch processing system in which computing jobs are distributed among separate computing nodes. A virtual machine can be a computing node in this system and so in advance configured in particular way virtual machine can at any moment join HTCCondor pool as a computing node. By default HTCCondor doesn't have the ability to request additional nodes from a cloud service, but configured in a proper way virtual machine is able by itself contact and join the HTCCondor computing infrastructure, so it is sufficient for the cloud scheduler (in case there are spare resources) to launch this virtual machine and it will automatically become a computing node of the HTCCondor infrastructure. To release occupied resources the scheduler at any time can stop the required number of virtual HTCCondor computing nodes.

The described above scheme is the simplest scheme not taking into account the possible absence of jobs in HTCCondor queues. In this case HTCCondor virtual computing nodes are still get launched leading to an inexpedient consumption of computing power. At the same time implementation of the queue tracking mechanism by the cloud scheduler would lead to the loss of the scheduler universality. As a solution of this problem the authors suggest to introduce the virtual machine labeling system to the cloud scheduling algorithm for the VMs that are temporary HTCCondor computing nodes, so that the cloud scheduler could stop them when it's needed to release the resources, and to delegate handling of virtual nodes and tracking the batch queue to a third-party system (e.g. VCondor).

5. Conclusion

Modern cloud platforms are widespread in commercial as well as in scientific areas. Given the limited amount of resources comprising cloud infrastructures the problem of its efficient usage inevitably arises. As a possible solution of the task to increase the efficiency of the cloud resources utilization the authors of the paper proposed a new method of smart scheduling of cloud resources utilizing the ability of dynamic resources reallocation. The method allows to free some amount of the allocated but underutilized resources that could be reused by one of the suggested ways in accordance with the chosen strategy and that may significantly increase the efficiency of cloud resources utilization.

Acknowledgements

The work was supported by the RFBR grant 15-29-07027.

References

- Timm, S. et al.* Cloud Services for the Fermilab Scientific Stakeholders // J.Phys.Conf.Ser. — Dec. 2015. — Vol. 664, No. 2.
- A.V. Baranov, N.A. Balashov, N.A. Kutovskiy, R.N. Semenov* JINR cloud infrastructure evolution // Physics of Particles and Nuclei Letters. — 2016. — Vol. 13, No. 5. — P. 672–675.
- N. Balashov, A. Baranov, V. Korenkov* Optimization of over-provisioned clouds // Physics of Particles and Nuclei Letters. — 2016. — Vol. 13, No. 5. — P. 609–612.
- A. Beloglazov, R. Buyya* OpenStack Neat: A Framework for Dynamic and Energy-Efficient Consolidation of Virtual Machines in OpenStack Clouds // Concurrency and Computation: Practice and Experience (CCPE). - 2015. - Vol. 27, No. 5 — P. 1310-1333.
- A. McNab, F. Stagni, and C. Luzzi* LHCb experience with running jobs in virtual machines // J. Phys.: Conf. Ser. 664 - 2015. - Vol. 664.
- Computing Center of the Institute of High Energy Physics (IHEP-CC) Vcondor - virtual computing resource pool manager based on HTCondor [online] URL: <https://github.com/hep-gnu/VCondor> (accessed on: 08.11.2016)
- A. McNab, P. Love, and E. MacMahon* Managing virtual machines with Vac and Vcycle 2015 // J. Phys.: Conf. Ser. 664 - 2015. - Vol. 664.