Cloud and Grid Technology Based Educational and Research Computing System

Spodarets D.V.[1], Yegorov S.Y.[1], Dragan G.S.[1]

[1]Odessa National I.I. Mechnikov University

m31@onu.edu.ua

Abstract. The development of computer technology has significantly expanded the possibilities for its use in research and learning process. Modern society requires revitalize the educational process, prepare a professional who not only possesses some knowledge, but still capable for continuous self-improvement, self, self-realization, which requires the current trends of the labor market. Therefore, it was necessary to develop training facilities with appropriate methodological support for the training of high quality, who are able to maintain the equipment.

The aim is to develop teaching and research at the university complex grid cluster pedagogy and methodological support, able to be used in teaching and research.

Keywords

Cloud Computing, GRID, OpenStack.

1 Introduction

In practice, research widely used distributed computing systems consisting of a number of computers connected to a single computer system with a special software. Their disadvantage is the queue when performing calculations by different users, ie, both calculations are performed only one task, while others have to wait for the release system. In other words, scientific computing systems are not designed to isolate because of their use in the classroom is limited.

The solution to this problem is the technique of "cloud computing", which provides a convenient continuous access to computing resources with minimal administrative costs and appeals to the technical staff. Method is to merge all existing physical computers in the so-called "cloud", which is located in the middle of virtual computers that directly serve the users. Feature cloud calculations are able invisible to the end user to transfer payments between physical computers. This, in turn, prevents loss and downtime caused by failures or routine maintenance. At all, the total capacity of the cloud may increased.

In scientific cloud computing environments provide the opportunity to deploy multiple computer systems at one and the same equipment with the required level of isolation, which in turn allows you to simultaneously run multiple performance objectives are incompatible with each other without waiting for their turn.

Basic requirements for complex:

- the ability to create isolated environments for different projects;
- ease of management resources, their distribution among projects;
- scalability complex;
- using only free software;
- the use of complex Nordugrid;
- simplicity to the equipment;

- minimum cost of additional equipment;
- minimum requirements for the qualification of junior technical staff.

2 The software industry.

Analysis of existing systems Eucalyptus, OpenNebula and OpenStack showed that Eucalytpus is fully open and has limited support for hypervisors (KVM and only Xen). OpenNebula and OpenStack have the same free license, but OpenStack has support for more hypervisors, better documented and has a large number of active users and developers, including NASA, SUSE Linux products Gmbh, Canonical Ltd, Dell, RackSpace, HP, AT & T and others. The final argument in favor of OpenStack was the presence of ready modules for Puppet, which made it possible to reduce time needed to add a new node to the cloud from several hours to 15-45 minutes, including hardware mount tasks.

The basic system for building complex was chosen universal free operating system Debian GNU / Linux. The main reason was the largest selection of official repository of software that provides all the necessary set without the need to add a large number of third-party sources of packages that may conflict with each other. For virtual nodes that directly perform the calculations, we used the latest available version of Scientific Linux.

Based on the characteristics of the existing equipment was elected built into the core of the Linux operating system hypervisor KVM. Among its advantages was a great speed, reliability, ability to use as the Guest of any modern operating system without the need of modification. In order to minimize potential errors by operators to automate and manage complex was used Puppet configuration management tool.

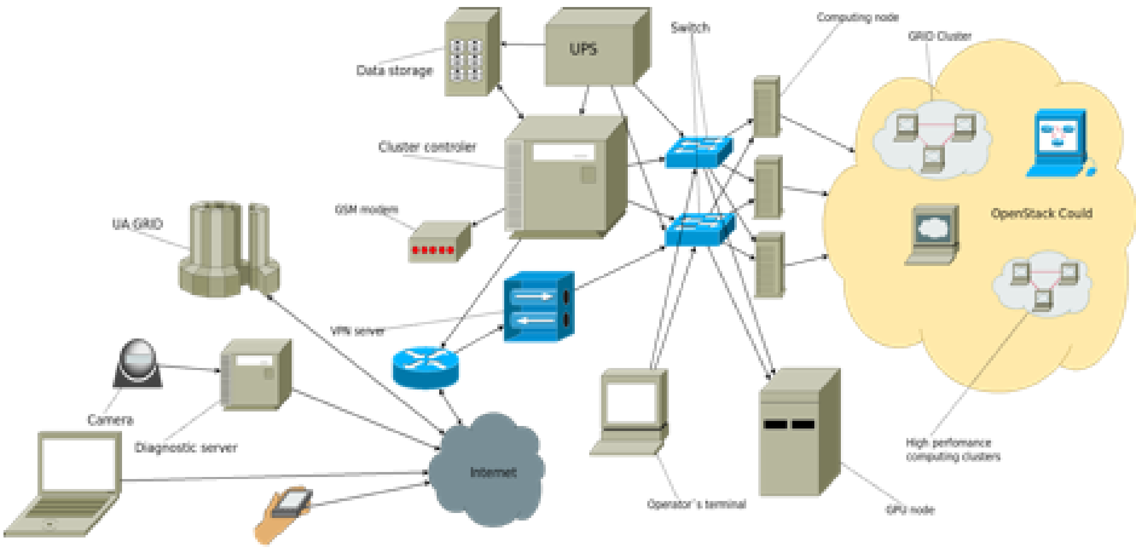

Figure 1. Block diagram of the computer complex.

The hardware of the complex:

- server-router with an AMD Athlon ™ 64 3500 + and 2 hybrid RAM, as it used to perform some infrastructure management tasks, and can't be replaced by special hardware;
- server to manage complex OpenStack with additional function of the central data storage based on a single processor Intel ® Xeon ® E31220 with 16 hybrid RAM and disk arrays 6 TIB-based disk controller 3ware 9650SE-4LPML, data storage is planed to be separated, but funding limitations enforced us to join this tasks on one server;
- two computing servers based on Intel ® Xeon ® CPU E5-2609 with 16 hybrid RAM with 32 hybrid solid state drives to install the base operating system;
- computer server equipment for CUDA-computing (not included in the structure of clouds, but interacts with it);
- three diskless computing servers based on 32-bit processors Intel Xeon

- for use as an educational complex distributed computing;
- UPS APC Smart-UPS SC1500;
- network switch D-Link WebSmart 16 network ports with Jumbo Frame support and VLAN.

Major requirement for computing servers is hardware virtualization technologies, like Intel VT-x and a lot of RAM. Also, we used non-branded hardware, assembled in our Center, as result of funding limits. Network hardware was chosen as cheapest one, with 802.1Q VLAN support. Deploying complex was spit in 7 basic steps:

- mounting hardware;
- the installation of the base operating system for infrastructure management servers;
- installation and configuration of management software;
- development of automated management policies;
- automatic deployment of private clouds;
- creation of virtual systems for use in GRID-cluster;
- deployment of cluster and join to GRID infrastructure.

To install the operating system:

- installation Server option in the router ≪DNS and DHCP server≫;
- installation management server in a private cloud version of ”base system”;
- add minimum version of one computing node and then cloning it to others.

Among the supporting software is crucial Centralized configuration management Puppet. Managing a complex template is for a private cloud based on Debian and OpenStack, which was elaborated to suit your needs. The time required to add new hardware to the cloud node is approximately 30 minutes after the hardware is turned on and has a linear dependence on network speed and disk subsystem of used hardware.

3 Technical problems and troubleshooting.

The first problem was the lack of support for Network Address Translation (hereinafter - NAT) in complex Nordugrid. Given the fact that technology is the only method of NAT destination addresses in public OpenStack, it threatened to reduce all work on the development of the complex. But in modern versions of the Linux has improved support for NAT, which allowed to ≪bypass≫ Nordugrid limitation, like it works without NAT. Also, public IP address was specified as network firewall for Nordugrid.

Another problem is the very low computing power complex that was identified test Linpack. The difference between the physical server and virtual machine to a value approaching 90%. The reason for the loss of function of the program performance proved Virtualization QEMU-KVM. QEMU computer emulator allows you to change the CPU instruction set, available for use by the virtual system. In the standard setting Gests operating system had very limited access to the capabilities of the CPU, which actually turned processor Xeon E5 at a very fast class processor Pentium MMX. In setting the cloud option was added, which clearly indicated processor type for the virtual environment as host-passthru. That Gests operating system was able to use the additional features of the CPU, such as a set of instructions SSE4, SSSE3 and the like. This problem is QEMU-specific, but other hypervisors, like Xen, may hide some CPU capabilities for guests and it may cause the same issue.

4 Pedagogical approaches use complex.

Application and implementation grid and cloud technologies in various areas of educational activities, contributing significantly to the modernization of education, the transition to a new level. Thus, a cluster grid can create a private cloud environment to simultaneously engage in the work of the entire group of students. Then grid cluster is a teaching and research platform for experimental and solving scientific problems, calculations can't be performed on individual computational facilities. Properly configured cloud complexes involve calculating all the calculating power, the cloud will provide user interaction both on a local and global level. Of course, this provides a new kind of educational process, but new information technologies provide for the introduction of modern teaching approaches, methods and techniques.

In methodological and research work cloud technology can be implemented on a grid cluster, which will be implemented own teaching facilities, research and development for the educational process. Specifically, the program provides quality training highly qualified specialists in computer science that can use grid technology in cloud environments.

References

[1] Gillam, Lee. Cloud Computing: Principles, Systems and Applications / Nick Antonopoulos, Lee Gillam. - L.: Springer, 2010.

[2] SoCC '10: Proceedings of the 1st ACM symposium on Cloud computing / Hellerstein, Joseph M. - N. Y.: ACM, 2010.

[3] The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology. NIST (20 October 2011).

[4] Official documentation OpenStack http://docs.openstack.org/

[5] Official Document of Debian http://wiki.debian.org/