Cloud Autonomous Grid Infrastructures for Training, Research, Development and Testing *

N.A. Kutovskiy[1,2]

e-mail: Nikolay.Kutovskiy@jinr.ru, [1]Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna

[2]National Scientific and Educational Centre of Particle and High Energy Physics of the Belarusian State University, Minsk, Belarus

Introduction

Such activities as training, research, development and testing in the grid sphere are less effective on production grid infrastructures from a viewpoint of the speed of the result achievement and the hardware resources utilization than the use of the special complexes for such tasks. Moreover, some of the mentioned tasks can not be solved on the production grid infrastructures at all. An approach to build cloud autonomous grid infrastructures to solve the listed tasks in the cloud and grid spheres was proposed. The implementation of the complex based on the proposed approach and its usage experience are described as well.

Motivation

Grid and cloud technologies are intensively used in different fields nowadays. Their familiarization and development assumes such activities as

- availability of corresponding specialists and/or possibility to train them,
- possibilities to do research work in these fields, implement new functionality and develop new services,
- adaptation of existing apps or development of the new one to run in proper environment.

There is a demand in a special infrastructure that could become a platform for training, research, development, tests and evaluation of modern technologies in distributed computing and data management (these technologies' main components are shown at fig. 1).

Implementation

The infrastructure for the mentioned purposes was set up at LIT JINR (t-infrastructure for short) [1]. Currently it consists of the following components (see fig. 2):

- EMI-based grid site (RU-JINR) as part of the distributed infrastructure;

*The work was partially supported by JINR grant No. 13-602-03

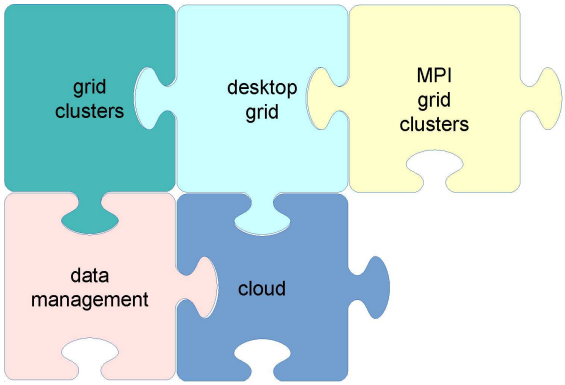

Figure 1: Main components of modern distributed computing and data management technologies

- T3MON testbed to develop monitoring tools for ATLAS Tier-3 sites;
- Russian National Grid Network (RGN) testbed;
- desktopgrid testbed based on BOINC software and 3G-bridge service.

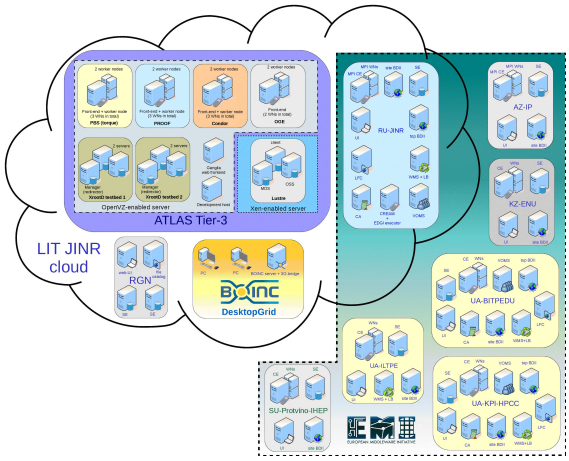

Figure 2: The schema of the distributed cloud grid t-infrastructure

All services of t-infrastructure are running on OpenVZ-based [2] virtual machines deployed in the local private cloud built on OpenNebula software [3].

EMI testbed. Currently EMI testbed consists of the grid sites and services listed in the table 1.

Table 1: The List of the EMI based t-infrastructure services and hosting organizations

site name	hosting organization	services
RU-JINR	Joint Institute for Nuclear Research	User Interface (UI), LCG Computing Element (LCG-CE) with two Worker Nodes (WNs), Disk Pool Manager Storage Element (DPM SE), LCG File Catalogue (LFC), Workload Management System (WMS), Logging&Bookkeeping Service (LB), site BDII (sBDII), top BDII (tBDII), Virtual organizations management service (VOMS)
SU-Protvino-IHEP	Institute of High-Energy Physics, Protvino, Moscow region, Russia	UI, LCG-CE + 2 WNs, dCache SE, WMS, LB, site BDII;
UZ-IMIT	Institute of Mathematics and Information technologies of Academy of Science of Republic of Uzbekistan, Tashkhent, Uzbekistan	UI, MPI enabled LCG-CE + 4 WNs, WMS, LB, site BDII, tBDII;
UA-BITPEDU	Bogolyubov Institute for Theoretical Physics, Kiev, Ukraine	UI, LCG-CE + 8 WNs, DPM SE, LFC, WMS, LB, site BDII, tBDII;
UA-KPI-HPCC	National Technical University of Ukraine "Kyiv Polytechnic Institute", Kiev, Ukraine	UI, LCG-CE + 8 WNs, DPM SE, LFC, WMS, LB, site BDII, tBDII.
KZ-ENU	L.N. Gumilyov Eurasian National University, Astana, Kazakhstan	UI, CREAM + 4 WNs, site BDII, DPM SE
UA-ILTPE	B.Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine, Kharkov,Ukraine	UI, CREAM + 2 WNs, site BDII
AZ-IP	Institute of Physics of Azerbaijan National Academy of Sciences (Baku, Azerbaijan)	UI, CREAM + 2 WNs, site BDII

This testbed is used intensively for different training courses (for users and system administrators from organization of the JINR Member States, semestral educational courses for students of JINR University Centre and Dubna University, for participants of international schools and practices). Apart from that porting applications to run in grid environment, grid services testing and certification as well as functional tests development were done on EMI testbed too [4].

T3MON testbed. The current ATLAS Tier-3 infrastructure consists of a variety of sites of different sizes and with a mix of local resource management systems (LRMS) and mass storage system (MSS) implementations. The Tier-3 monitoring suite, having been developed in order to satisfy the needs of Tier-3 site administrators and to aggregate Tier-3 monitoring information on the global VO level, needs to be validated for various combinations of LRMS and MSS solutions along with the corresponding Ganglia plugins. Since a performance of specific components is not a critical issue for development and validation, whereas easy management and deployment are crucial, all services were installed on virtual machines in the LIT JINR private cloud. This testbed [5] allows simulation of various computational cluster and storage solutions. It provides the ability to run testbeds with various LRMS and MSS implementations, and with the capability to quickly redeploy particular cluster or their components.

As a result a set of Ganglia-based monitoring tools for such LRMS and MSS used on ATLAS Tier-3 sites as PBS, Condor, XRootD, Lustre, PROOF, OGE were created and tested on the testbed's basis. These tools allow having information on Tier-3 sites operation both on local and global levels. Monitoring metrics sent from sites via active messages queue are collected at CERN and then are presented at Dashboard [6] which is a single entry point to the monitoring data collected from the distributed computing systems of the LHC virtual organizations.

RGN testbed. To meet JINR's obligations in Russian National Grid Network project the following instances were deployed on t-infrastructure: graphical user web-interface for problem-oriented interfaces (POIs) development, two GridFTP storages and the virtual machine for data management system development [7].

POIs are intended to simplify user's work with particular application in RGN infrastructure. The POI itself is a plug-in for RGN graphical user web-interface service. LIT JINR team developed POIs for the 5 applications: DL_POLY (parallel molecular dynamics simulation package), Elmer (open source multiphysical simulation software), GEANT4-DNA (Geant4 Monte Carlo simulation toolkit extended with processes for the modeling of early biological damages induced by ionizing radiation at the DNA scale), Fire Dynamics Simulator (low-speed flows simulation with an emphasis on smoke and heat transport from fires) and ZondGeoStat (geophysical and geometrical sounding data processing) . These POIs let users specify a unique job name, choose a particular application version, define an application specific parameters, specify an archive name with input files as well as a

name for archive with output files. Apart from that there is a field at the bottom of the web-page where system replies are printed. For the rest operations like e.g. user authentication and authorization, job management, input and output files uploading and downloading, a RGN graphical user web-interface service is used.

The main task of the data management service is to provide a convenient and reliable mechanism for users to manage their data. It is one of the core services of RGN. The service is based on X.509 and GSI and using GridFTP as basic data transfer protocol. The data management is performed via web-interface and RESTful API. That service takes care about access control, interactions with different grid storages, operations on datasets, replication, backup, etc.

DesktopGrid testbed. To use idle CPUs resources of the desktop PCs at JINR and organizations from its member states for computational tasks one needs to build desktop grid (DG) infrastructure and adopt applications for it. The last item requires a testbed which can be deployed on t-infrastructure and integrated within its EMI part.

The DG testbed consists of the following components: BOINC server with 3G-bridge, a few BOINC clients installed on the PCs of the University Centre of JINR, EMI CREAM computing element with EDGI executor.

Some test applications were run successfully on that testbed including job submission using EMI User Interface. For the time being few organizational aspects related to production DG infrastructure building and its usage are in-depth study.

Conclusion

The t-infrastructure is based on such modern information technologies as cloud and grid ones. Their synthesis allows to increase an efficiency of hardware utilization and extend the spectrum of t-infrastructure applications. It includes different type of resources: clusters, supercomputers, cloud, heterogeneous desktopgrid PCs, distributed storage and data management systems and is intensively used for training, research, development and testing tasks.

References

[1] N. Kutovskiy Cloud autonomous grid infrastructures for training, research, development and testing tasks // Informatization of education and science, 2013, vol. 4 (Russian).

[2] http://openvz.org

[3] http://opennebula.org

[4] Kutovskiy N.A. Distributed training and testing grid infrastructure evolution // The 5th International Conference "Distributed Computing and Grid-technologies in Science and Education" (Dubna, 16-21 July 3, 2012): proceedings. Dubna, 2012. P. 180-185.

[5] S. Belov et al VM-based infrastructure for simulating different cluster and storage solutions used on ATLAS Tier-3 sites // Journal of Physics: Conference Series. 2012. Vol. 396. Part 4. 5 pp. Doi:10.1088/1742-6596/396/4/042036

[6] http://dashboard.cern.ch

[7] S. Belov et al JINR participation in the Russian National Grid Network project, ibid, pp.????

[8] V. V. Korenkov, N. A. Kutovskiy, R. N. Semenov An experience of the application software packages adaptation for running in grid environments // Computer Research and Modeling, 2012, vol. 4, no. 2, pp. 339–344 (Russian).

[9] Kutovskiy N.A., Lensky I.I., Semenov R.N. Problem-oriented web-interfaces for Russian grid network // The 5th International Conference "Distributed Computing and Grid-technologies in Science and Education" (Dubna, 16-21 July 3, 2012): proceedings. Dubna, 2012. P. 186-188.