CLOUDS OF JINR, UNIVERSITY OF SOFIA AND INRNE JOIN TOGETHER

V.V. Korenkov[1], N.A. Kutovskiy[1], N.A. Balashov[1], V.T. Dimitrov[2,a], R.D. Hristova[2], K.T. Kouzmov[2], S.T. Hristov[3]

[1] *Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia*

[2] *Faculty of Mathematics and Informatics, University of Sofia, 5 James Bourchier Blvd., Sofia, 1164, Bulgaria*

[3] *Institute for Nuclear Research and Nuclear Energy, 72 Tsarigradsko shose, Sofia, 1784, Bulgaria*

E-mail: [a] cht@fmi.uni-sofia.bg

JINR develops a cloud based on OpenNebula launched for integration with the clouds from the member states. The paper presents a state of the 3-year project that aims to create a backbone of the cloud in Bulgaria. The University of Sofia and INRNE participate in this initiative. This is a target project funded by JINR within the plan for scientific research and cooperation of the Institute.

Keywords: cloud computing, OpenNebula, clouds integration

1. The project

The project "Creation of distributed information and computing infrastructure using cloud and grid technologies based on resources of the Laboratory of Information Technologies of the Joint Institute for Nuclear Research, Institute for Nuclear Research and Nuclear Energy of the Bulgarian Academy of Science and Sofia University 'St. Kliment Ohridski'" is a 3-year project with a priority funding in the period of 2017-2019.

2. Targets

This Project's aim is to create a common distributed information and computational environment using Cloud and Grid technologies on the capabilities available at LIT JINR, INRNE BAS and FMI SU. This environment will be used for further scientific research, training and education in common concerns of the participants as well as third parties.

The establishment of such an environment is important to join the participants' efforts and resources in solving fundamental scientific and application problems that can be successfully solved only with the help of new methods and approaches to distributed and parallel computations and with systems for a distributed storage of large amounts of data.

3. Project problems

The target research in frames of the project is to find out how to integrate LIT JINR, INRNE and FMI SU resources in an optimal way. The resulted infrastructure must help to fasten the scientific research underway at JINR and to establish a base for training and education in Bulgaria of high-skilled specialists on advanced computer technologies for big scientific data storage, processing and analyses. The experience and skills acquired in this environment could be used for solving problems in many other areas.

4. Project foundations

There are many scientific research centers in Bulgaria; among them are INRNE and University of Sofia, which have a long-time traditional, high level cooperation with JINR within the fundamental scientific research and investigations.

New emerging technologies from the last years on networking, computing and data management impose the JINR resources to be in common share with those of its member states and in particular with Bulgarian scientific research centers, targeting particular scientific and application problems. A lot of problems could not be solved without new modern approaches in the area of information and computational environments for storage, processing and analysis of big scientific data. The establishment of such an environment will intensify the cooperation between JINR and Bulgarian institutes and reduce time required for obtaining new scientific results as well as will serve as a powerful impetus on the development and usage of new emerging technologies in Bulgaria.

5. Current research

Numerous present-day scientific studies foresee using systems for collecting, storing, processing and analyzing information. Moreover, often the data volumes are collected and must be

analyzed within the capabilities of one, even largest, scientific organization. In that context, the task for resources integration of research centers in one common infrastructure looks an urgent one.

In the biggest worldwide projects, this task is solved using Cloud and Grid technologies. Such a global infrastructure is the Worldwide LHC Computing Grid (WLCG) established for the Large Hadron Collider (LHC) computing. This infrastructure includes hundreds of data processing and computing facilities worldwide. Starting from 2003, JINR, INRNE and SU have been the WLCG participants.

The project goals are in synchrony to the world trends in the development of new information technologies and computing. The establishment of the common Cloud and Grid infrastructure of JINR - Bulgaria, on the one hand, will meet the scientific researchers' needs for more resources, will improve the resources utilization, and from a strategic viewpoint, will strength the Bulgaria's position among the countries that use and implement advanced information technologies.

6. Expected results

The expected results from the project are as follows:

1. The establishment of the distributed information and computing infrastructure segments in INRNE and SU shared with a corresponding infrastructure at JINR.
2. The researchers from JINR, INRNE and SU will use together all shared resources to solve concrete problems within the JINR scientific program in collaboration or separately with participating Bulgarian scientific organizations.
3. A software stack will be developed for information and computing resources integration and corresponding knowledge and training for its application will be delivered in Bulgaria.
4. Training events will be organized and conducted for Bulgarian professionals on the environment resource usage.
5. Virtual computers will be instantiated for INRNE, SU and other BAS institutes, for cooperative or separate research with JINR.

7. The networking and computing infrastructure of JINR

The Networking and Computing Infrastructure of JINR is a distributed software and hardware complex based on specialized software and multifunctional equipment. The core of this infrastructure is the Multifunctional Information - Computational Complex of JINR. This infrastructure uses JINR local area network that connects all JINR resources into single information and computing environment accessible to all JINR users. The complex implements Grid technologies. It supports the data exchange among the scientific laboratories and the administration.

The complex has telecommunication channels that provide a remote access to Russian and foreign scientific networks and support the remote access to JINR resources.

The purpose of this infrastructure requires being supported high secure, reliability and scalability levels in combination with ease system maintenance.

7.1. Telecommunication channels of JINR

Currently, JINR has an external telecommunication channel for data transmission with the speed of 100 Gbps. The JINR local area network connects 7947 computers and nodes.

The central communication node has been upgraded to be the fault tolerant core of this LAN communication structure. It supports high network security and reliability levels, has good data transmission parameters and tools for monitoring of, management of, and remote access to the nodes.

The backbone networks operates at a data transfer rate of 10 Gbps.

7.2. The computing infrastructure of JINR

The multifunctional information and computing complex of JINR consists of:

1. Tier 1 level JINR center for CMS experiment at LHC with configuration and parameters:
    - 4160 computing nodes;
    - 5.3 PB disk storage;
    - 1.1.PB disk buffer space;
    - 9 PB tape library.
2. Tier 2 level JINR center operating also as the computer cluster for the non-Grid computing including parallel computing for experiments BM@N, MPD, CMS, ATLAS, ALICE, LHCb, COMPASS, STAR, NOvA, BESIII, DIRAC, OPERA, NEMO, Mu2e, NUCLON, FUSION, BIOMED, etc., for physical processes modeling and experimental data analyses, with configuration and parameters:
    - 3640 computing nodes;
    - 2 PB disk storage.
3. The JINR LIT Cloud infrastructure consists of 600 cores, RAM 2700 GB, distributed file system of 16 TB, the local disk space on the servers with virtual machines is 50 TB.
4. Heterogeneous computing cluster "HybriLIT" is intended for carrying out calculations using parallel programming technologies. It can be characterized as a modular heterogeneous High Performance Computing (HPC) complex with configuration and parameters
    - Peak performance – 142 Tflops,
    - CPU cores – 252,
    - GPU cores – 77184,
    - MIC cores – 182.

8. INRNE resources

Currently, there is no production Grid farm at the INRNE. The planned Cloud infrastructure will consists of available 6-10 servers with parameters: Intel Xeon 2.80 GHz, RAM 2 GB, and HDD 80 GB. Additional 10-12 TB disk storage must be added to theses servers.

For the reliable and stable operations of this local Cloud, suitable air conditioning and UPS must be acquired.

New servers and components will be acquired for this Cloud infrastructure after its initial establishment.

9. Resources of Sofia University

The Grid farm at the University of Sofia is managed and supported by the Faculty of Mathematics and Informatics and the Faculty of Physics. It is located at the Faculty of Physics. This production farm is part of the European Grid infrastructure. Its configuration is given in Figure 1.

It consists of:

- 40 WORKER NODES:
  - 2 x CPU per node
  - Processor Information:
    - Version: Intel(R) Xeon(R) CPU E5345
    - Max Speed: 5000 MHz
    - Current Speed: 2333 MHz
    - Core Count: 2
    - Thread Count: 4
  - 8 x Ram per node
    - Memory Device
    - Size: 2048 MB
    - Type: DDR2
    - Speed: 667 MHz
- 8 STORAGE Servers:
  - 2 x CPU per node
    - Processor Information
    - Version: Intel(R) Xeon(R) CPU E5345
    - Max Speed: 5000 MHz
    - Current Speed: 2000 MHz

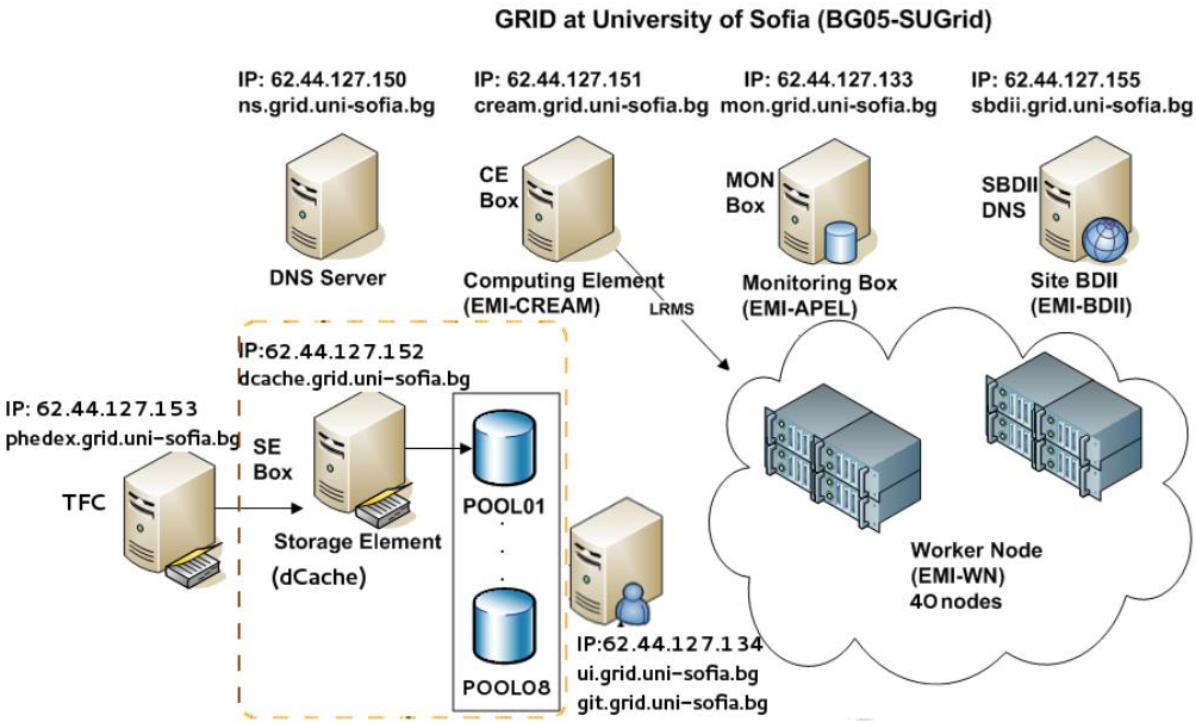

Figure 1. University of Sofia Grid farm

  - 4 x RAM per node
    - Memory Device
    - Size: 2048 MB
    - Type: DDR2
    - Speed: 667 MHz
  - 1 x RAID 6 Storage per node

        - Disk Size 9.1 TB
- TOTAL RESOURCES:
    - COMPUTING NODES:
        - CPUs: 80
        - Cores: 160
        - Threads: 320
        - RAM: 655360 (640 GB)
        - Lan: 1 GigB
    - STORAGE:
        - 8 x RAID arrays:
        - Total: 72.8 TB
        - Configured Arrays: RAID6
        - Distributed File system: dCache

10. Working plan

The following activities are planned:

1. Investigation of available solutions for building common distributed information and computing environment based on the resources of LIT JINR, INRNE, and SU.
2. Acquisition of necessary equipment and establishment of test sites at INRNE and SU, based on the selected solution.
3. Organization of training for Bulgarian specialists.
4. Integration of JINR LIT and Bulgarian sites into common distributed information and computing environment.
5. Testing the distributed information and computing environment.
6. Exploitation of the distributed information and computing environment for concrete scientific and application problems at JINR and Bulgarian centers.
7. Acquisition and installation of new resources for increasing the power this distributed information and computing environment for solving scientific and application problems.

11. Acknowledgement

This work is mainly supported by the JINR priority funding project "Creation of distributed information and computing infrastructure using cloud and grid technologies based on resources of the Laboratory of Information Technologies of the Joint Institute for Nuclear Research, Institute for Nuclear Research and Nuclear Energy of the Bulgarian Academy of Science and Sofia University 'St. Kliment Ohridski'" on the problem plan topic 05-6-1118-2014/2019 "Information-computing infrastructure of JINR", proposed for funding by the Bulgarian Plenipotentiary at JINR.

The work also is partially supported by the Bulgarian National Science Fund project ДН 02/9 - 17.12.2016, "GloBIG: A Model of Integration of Cloud Framework for Hybrid Massive Parallelism and its Application for Analysis and Automated Semantic Enhancement of Big Heterogeneous Data Collections".

12. Conclusion

Currently, the project is in its initial phase.

The selected solution for integration is based on the OpenNebula [1-2].

The training of the Bulgarian specialists has been conducted this year.

A test site was partially installed on the currently available equipment at the University of Sofia. By the end of the year, this site will be moved in the newly acquired equipment for the final establishment of the site infrastructure.

References

[1] Baranov A.V. et al. JINR cloud infrastructure evolution // Physics of Particles and Nuclei Letters, ISSN 1547-4771, eISSN: 1531-8567, 2016, vol. 13, No. 5, pp. 672–675. DOI: 10.1134/S1547477116050071.

[2] Baranov A.V. et al. Approaches to cloud infrastructures integration // Computer Research and Modeling, 2016, vol. 8, No. 3, pp. 583–590 (in Russian).