



THE GRID AND CLOUD COMPUTING FACILITIES IN LITHUANIA

VAIDAS GIEDRIMAS*, AUDRIUS VARONECKAS† AND ALGIMANTAS JUOZAPAVICIUS‡

Abstract. Nowadays ICT are faced the paradigm shift to the cloud computing. Even already distributed environments such as grids are challenged to change in order to meet new paradigm. This paper presents Lithuanian grid computing achievements, including National grid initiative (LitGrid), and our cloud computing potential. As an example two particular cloud-related applications are presented.

Key words: cloud computing, software services, LitGrid, optimization

1. Introduction. Nowadays ICT are faced the paradigm shift to the cloud computing. Even already distributed environments such as grids are challenged to change in order to meet new paradigm. The solutions for migration into the cloud or to achieve efficient side-by-side usage of the grid and cloud technologies (e.g. StratusLab project [11]) are on the focus on world's scientific research.

This paper presents Lithuanian grid computing achievements in the cloud-computing context. As an example two particular cloud-related applications are presented.

The main aim of this paper is to present current position of Lithuanian National Grid Initiative (NGI) – LitGrid – and its potential in the cloud computing context.

The paper is organized as following. Section 2 outlines main distributed computing projects in Lithuania. Section 3 presents two optimization problems and in Lithuanian's distributed applications for its solving. Finally, conclusions are made and further steps are discussed.

2. Main Distributed Computing Projects in Lithuania.

2.1. LitGrid. There are several private and public distributed computing infrastructures in Lithuania. Each university and some colleges has a cluster. One of the most powerful of them is private cluster constituted of 36 CPUs named Vilkas at Vilnius Gediminas Technical University (<http://vilkas.vgtu.lt>).

In order to connect the clusters and to improve the accessibility to computing resources in academic sector the program “Lithuanian distributed and parallel computing and e-services network (LitGrid)” has been started from 2006's. LitGrid now is not only grid as infrastructure but research community and powerful NGI. Currently it operates over 500 processors, has over 30 TB storage capacity, about 90 users, serves numerous scientific research areas. Lithuanian Grid functionality includes:

- deployment, maintenance and upgrading of the grid infrastructure and e-services;
- analysis, design, test deployment of the cloud computing technology, especially in relation to grids;
- serve for new users, their training tasks;
- creation, maintenance and deployment of the grid certificating procedure;
- presentation of the grid activities and capabilities to the society, to potential users, to public sector and business, developing new grid projects.

LitGrid consist of 13 academic institutions as partners: most active Lithuanian universities, research institutes and colleges. Academic and Research Network in Lithuania (Litnet) provides the network infrastructure with 1Gbps throughput connection between sites. The special communication center to support distributed computing data exchange has been implemented. LitGrid infrastructure is now based on gLite and ARC middleware, however several other operational systems are under consideration to be deployed (Globus, UNICORE, CREAM). Three sites are running central services (WMS, VOMS, LFC) for Lithuanian Grid. The important task for LitGrid is to participate in EGI organization and in the FP7 project EGI-InSPIRE. Such activities grant access to new technologies, developments, innovations, enables to take part in support and development procedures. The essential links for Grid are relations to Scandinavian countries grids, as well as joint efforts with Belarus academic institutions.

*Siauliai University, Visinskio 15, LT-77156 Siauliai, Lithuania, (grid@su.lt)

†Vytautas Magnus University, Vileikos 8, LT-44404 Kaunas, Lithuania, (a.varoneckas@if.vdu.lt)

‡Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania, (algimantas.juozapavicius@mif.vu.lt), <http://www.litgrid.lt/>

2.2. DiSCC. The biggest element of the Lithuanian NGI is Digital Science and Computing Center (DiSCC) of the Vilnius University, Faculty of Mathematics and Informatics (MIF) established together with two business partners - BAIP and TVM. The mission of DiSCC is to provide a translational environment for the transfer of the knowledge and research results to specific areas of the economics and to innovative digital products, maximally shortening the path from generation and exploration of new scientific ideas to computationally intensive market products.

DiSCC pays special attention to cooperation projects between scientific and public or private institutions, their partnership and elaborated feasibilities, in search for effective digitalization areas of Lithuanian economy and public sector, identification of such areas and implementation of suitable cooperation projects.

DiSCC also has a task to manage efficiently, elaborate and use computing resources of the MIF, for the purposes of the students' studies and for research work, for the implementation of newest trends in information technologies. The priority in this field belongs to grid, cloud and HPC computing technologies, to what they can offer to the science and the studies, as well as to public sector or business.

DiSCC is expected to be translated into international center, attractive for computer scientists, industry, foreign partners and for the state as a customer, helping them to digitize areas of the activities by using most suitable scientific and technological initiatives.

DiSCC offers for open access the supercomputer constituted of 1500 CPU cores and 600 TB of storage.

2.3. Main Litgrid applications . LitGrid directly or indirectly serves as a infrastructure for the national and international distributed applications. The main domains of Lithuanian applications are as follows (in the alphabetic order):

- Astronomy (the analysis of star's spectrum);
- Environmental research (the investigation of the Baltic sea ecosystem etc.);
- E-learning;
- Hydrodynamics (the modeling of the floodgates, barrages etc.);
- Health care (epilepsy diagnostics, orthopedics, radiology, fluorescence microscopy, biomarkers);
- Humanities (the analysis of text databases, language processing etc.);
- Materials science (nuclear physics, quantum chemistry, etc.);
- Mathematics (optimization, number theory etc.).

As a resource provider and/or partner LitGrid has been involved in the following projects:

- COST P19 – Multiscale modeling of the materials [10].
- PjezoAdapt - Development and research of mechatronic nanometer resolution multi-dimensional displacements generation / measurement systems
- ITER – the International Thermonuclear Experimental Reactor project [7].
- GridTechno – the Lithuanian national project of the applications of the grid technologies for various research areas. [15].

LitGrid in the cooperation with Sweden, Estonia, Latvia, Poland and Switzerland (CERN) have been participated in the BalticGrid and the BalticGrid-II projects [1].

2.4. The migration towards the cloud. The migration to the cloud computing in Lithuanian academic sector is already started. Small clouds based on various platforms (OpenNebula, Windows Azure, EC2 etc.) has been created in VU, SU and other Lithuanian universities for the test purposes as well as for the scientific research.

Kaunas University of technology (KTU) [8] in cooperation with Vilnius University now is working on the project *Enabling the Researchers with the Gigabit Network Technologies*. One of the tasks of this project is the investigation of three virtualisation platforms:

- virtual machines and resource pools based on VMWare technologies;
- virtual machines based on open-source technologies (Xen,KVM,VirtualBox etc.);
- jail virtual machines (OpenVM,lxc, BSD jail etc.).

The process of KTU's information system migration into VMWare vSphere-based platform is already started.

Vilnius Gediminas Technical University (VGTU) is developing the workplace of the researcher based on cloud technology. Private cloud infrastructure was build based on Eucalyptus open source solution, XEN hypervisor and Rocks Clusters platform. Java Typica framework and JetS3t Toolkit for Amazon's S3 online storage service were used to develop the flexible interface to the user. ANSYS and MATLAB are considered as an interesting and promising cases of the user software [9].

Lithuanian cloud-related research is not limited to local projects only. Lithuania, as a partner of BalticGrid project has been involved in the BG subproject – BalticCloud [5], aimed at developing cloud infrastructure in Baltic states and Belarus. The infrastructure is based on open-source solutions and is available for research and teaching activities within the partner states. Now the BalticCloud is part ECEE initiative [4].

The cloud computing gains the popularity in the commercial sector as well as in academic. There are several companies which provide complete IaaS or PaaS solutions (e.g. [2, 13]). The customers of CC providers are small and medium business companies as well as the government of Lithuania. One of the biggest applications in public sector is the cloud computing solutions for the electronic census of Lithuanian population, performed this year.

The association of Lithuanian ICT companies *InfoBalt* [6] have performed the survey of its members (N=78), mainly SMEs (71% of the respondents). This survey shows the following prospects in Lithuanian cloud computing market:

- 33% of the respondents already provide CC services, 6% have plans to start such services this (2011) year and 15% ave plans to start next two years;
- Most popular CC services are related with CRM (33%) and ERP(30%) but almost 52% of the respondents have named other small services, e.g. e-mail, data managment etc.;

3. Optimization problems and grid-oriented applications in Lithuania. The researchers of LitGrid community develops and uses a large set of the grid-oriented applications. Part of these applications can be adopted (or already are under adopting) to the cloud computing, e.g. materials science applications [12]. It is not possible to describe all applications in one short paper. In this section examples of developed applications in Vilnius University Institute of Mathematics and Informatics (VU IMI), Vytautas Magnus University (VMU) and Siauliai University (SU) were taken.

3.1. VU IMI and VMU approach. Visualization of the multidimensional data is a large scale numerical problem which is time consuming and difficult to solve on a single personal computer. The Grid is an excellent candidate for providing the infrastructure needed for solving such problems.

Real world objects usually are described by multidimensional data sets. Visual representation can be very useful to grasp the structure of such data sets. There are a lot of the visualization techniques, which can be used for multidimensional data visualization. One of the powerful multidimensional data visualization techniques is multidimensional scaling (MDS). Multidimensional scaling is an exploratory technique for data analysis, widely usable in different applications, e.g. psychometrics, market analysis, data mining, visualization of general multidimensional data, visualization of the observation points in interactive global optimization.

Let a set of n multidimensional vectors (representing the considered objects) have to be visualized in a $p = 2$ dimensional space. Pairwise dissimilarities measured between all pairs of the objects are denoted by $\delta_{ij}, i, j = 1, \dots, n$. It is supposed that dissimilarities are symmetric $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$. The points x_1, x_2, \dots, x_n that constitute a set of n objects in p dimensional space should be found fitting pairwise distances of points to given pairwise dissimilarities $\delta_{ij}, i, j = 1, \dots, n$. The fitness criterion, called STRESS function should be minimized:

$$\min S(X) = \sum_{i=1}^n \sum_{j=i+1}^n w_{ij} (d(x_i, x_j) - \delta_{ij})^2$$

where $X = (x_1, x_2, \dots, x_n)$, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip}, d(x_i, x_j))$ denotes the distance between the points x_i and x_j , and w_{ij} denotes weights.

Quality of the projection of the multidimensional data into low-dimensional space depends on $d(x_i, x_j)$, and therefore choice of the metrics of the embedding space is very important. Pairwise dissimilarities between pairs of the object may be considered as a distance in multidimensional original space, and may be estimated using different norms in \mathcal{R}^p . The most widely used distance measure is a Minkowski distance [3]:

$$d(x_i, x_j) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^r \right)^{1/r}.$$

Parameter r influences the quality of the projection into low-dimensional space. Usually two well known special cases of Minkowski distance are used in MDS: Euclidean distance when $r = 2$, and city-block distance when

$r = 1$. The points $X = (x_1, x_2, \dots, x_n)$ found by means of the minimization of STRESS function using different distance metrics are different nonlinear projections of the set of the objects in original multidimensional space to the lower dimensional embedding space.

MDS is a difficult global optimization problem. Although STRESS is defined by an analytical formula, which seems rather simple, its minimization is difficult. The function normally has many local minima. When city-block distances are used, STRESS can be non differentiable even at the minimum point [16]. The minimization problem is high dimensional (number of variables is $N = n \times m$) global optimization problem. When computing power of usual computers is not sufficient to solve a problem, the high performance parallel computers, clusters of the computers, computational grids and cloud computing may be helpful. An algorithm is more applicable in case its parallel implementation is available, because larger practical problems may be solved by means of parallel computation. Parallel version of genetic algorithm with multiple populations for MDS with Euclidean distances has been implemented in grid [14].

3.2. SU approach. There are three main distributed computing-related research areas of the Siauliai University:

- Interoperability of the e-learning and grid and cloud computing systems.
- Numerical modeling and investigation of the systems. Modeling of the temperature, stability and optimal design of the constructions (shells, plates, sticks) is investigated. New technology of numerical modeling of the constructions of composite materials using grid and cloud computing is being developed.
- Software synthesis for grid, cloud computing and other distributed systems. The synthesis distributed software to solve scientific problems of the numerical modeling and optimization of the systems.

On the umbrella of third research area new project at Siauliai University is started from the October 1 of 2010s. The main goal of this project is to develop the environment for scientific software synthesis using grid, cloud and wiki-oriented technologies.

We are stating the hypothesis that using together wiki-based technologies, software synthesis methods and the power of the grid/cloud infrastructure scientific software can be developed more rapidly and the quality of the software will be better. The project consist of three main stages:

1. The development of the portal for the wiki-based mass-collaboration. This portal will be used as the UI enabling scientists to specify the problems for software development, to rewrite/refine the specifications and software artifacts given by other researchers, to contribute all software developing for particular domain process. As the target domain for soft-ware development we have chosen the set of the statistical simulation and optimization problems. In the future the created environment can be applied to other domains.
2. The development of the model of the interoperability bridge between wiki-based portal and the LitGrid or other grid/cloud-based infrastructure. For this purpose currently the private OpenNebula-based cloud is created at SU.
3. To refine existing methods for software synthesis using the power of distributed computing infrastructures.

The results of the project will have direct positive impact in the scientific software development, because of the bridging two technologies, each of them promise good performance. The power of the wiki-technologies will ensure the ability of the interactive collaboration on software developing using the terms of particular domain. On the other hand the bridge of new environment and the grid/cloud infrastructure will give the possibility to use all power of distributed computing infrastructures.

Conclusions and Future Work. Lithuanian infrastructure for the distributed and parallel computing, including grid and cloud computing, is modern, advanced and continuous growing. The scientific LitGrid community is growing and the variety of applications is increasing. The applications from very different domains are observed, starting with the nuclear physics, ending with the public sector.

The migration to the cloud computing in Lithuanian academic sector is already started and first very promising results are already achieved. As a test cases for this migration scientific problems of optimization methods are selected. Our future work includes the continuous migration to cloud computing platform and investigation of the interoperability of grid and cloud technologies as well as the interoperability of different clouds. Lithuanian National Grid Initiative is open for collaboration for other NGI's, universities and business

Acknowledgements. The authors of this paper would like to address special thanks to all indirect contributors of this paper, especially Dr. D.Mazeika (VGTU), Dr. K. Paulikas (KTU), Prof. L.Sakalauskas (SU)

and Mr. A.Pleckaitis (InfoBalt association).

REFERENCES

- [1] BalticGrid: Balticgrid-ii web site. <http://www.balticgrid.org/>
- [2] BlueBridge: Bridge to cloud. <http://www.bridge2cloud.com/>
- [3] Cox, T., Cox, M.: Multidimensional Scaling. Chapman and Hall/CRC (2001)
- [4] ECEE: Enabling clouds for escience. <http://www.scientific-cloud.org/>
- [5] Edlund, A.: Balticcloudproject. <http://cloud.balticgrid.eu/>
- [6] INFOBALT: Infobalt association. http://www.infobalt.lt/sl/index_en.php
- [7] ITER: The way to the energy. <http://www.iter.org/>
- [8] KTU: Kaunas university of technology. <http://en.ktu.lt/>
- [9] Mazeika, D., Kaceniauskas, A., Pacevic, R., Katkevicius, T.: Engineering application on university private cloud. EGI User Forum 2011 abstract session.
- [10] P19, A.: Multiscale modelling of materials. <http://www.ipm.cz/costp19/>
- [11] StratusLab: Enhancing grid infrastructures with virtualization and cloud technologies. <http://stratuslab.eu/>
- [12] Tamuliene, J., Vaisnoras, R., Badenes, G., Balevicius, L.: Point of view on magnetic properties of con (n=6,8,10,12) based on quantum chemistry investigations
- [13] TEO: The cloud computing-based voice telephony services. <http://www.teo.lt/en/press-en/2733>
- [14] Varoneckas, A., Zilinskas, A., Zilinskas, J.: Parallel multidimensional scaling using grid computing: assessment of performance. Information technology and control 37, 52–56 (2008)
- [15] VGTULPC: Gridtechno project. <http://1s1.vgtu.lt/en>
- [16] Zilinskas, A., Zilinskas, J.: Two level minimization in multidimensional scaling. Journal of Global Optimization 38, 581–596 (2007)

Edited by: Dana Petcu and Jose Luis Vazquez-Poletti

Received: November 1, 2011

Accepted: November 30, 2011