A Grid Computing Infrastructure for Research in Geophysics: EGEODE

Dominique Thomas*

Compagnie Générale de Géophysique

Summary

The paper explains the main benefits of this Grid Computing emerging technology for geosciences and particularly for data sharing and processing in the geophysical industry.

Then we introduce EGEODE, the "Expanding Geosciences On Demand" Open Virtual Organization, with a practical description of how to join the community and become an actor in this exciting project started few months ago.

Introduction

Modern seismic data processing and geophysical simulations require more and more huge amount of computing power. The research community hardly keeps pace with this evolution resulting in difficulties for small or medium research centers to deliver their innovative algorithms.

Grid Computing is an opportunity to foster sharing of computer resources and to give access to large computing power for a limited period of time at an affordable cost. Capability to solve new scientific complex problems and to validate innovative algorithms on real size problem is also a way to attract and keep the brightest researchers for the benefit of both the academic and industrial R&D geosciences communities.

EGEODE is an initiative to create a Virtual Organization dedicated to research in Geosciences for both industrial R&D (public and private) and Academic Laboratories. It was launched by CGG end 2004 in the context of EGEE (Enabling Grids for E-sciencE), the largest worldwide grid infrastructure project to date.

EGEODE stands for "**Expanding Geosciences On Demand**". It raises already interest across Europe, Russia and Asia and delivers a framework for e-collaboration in Industry and Research.

Grid computing for sciences

Grid computing, a form of distributed computing, has been developed by scientific communities since its inception in the early 90 years, to solve very large problems requiring huge computation resources (CPUs and data storage). Grid computing was born with the simplest idea to exploit under-used computation resources to solve large problems and by sharing existing resources to create virtual huge computer center, which can be accessed by anybody from anywhere as Internet is today (I. Foster and C. Kesselman, 1998).

The capacity of grids for enabling access to geographically distributed computing power and storage facilities has been seriously demonstrated and new international projects are now focused on operational grids used for day to day production, like EGEE.

The general benefits of grid computing are:

- Access to computing resources without investing in large IT infrastructure,
- Lower the total cost of IT by sharing available resources with other members of the community

And the specific benefits for Research community:

- Free the researcher from the additional burden of managing IT hardware and software complexity and limitations,
- Have a framework to share data and project resources with other teams across Europe and the rest of the world,
- Share best practices, support and training more easily,
- Enable cross-organizational teamwork and partnership.

EGEE is an international project that aims to integrate current national, regional and thematic Grid efforts, in order to create a seamless Grid infrastructure for the support of scientific research. This infrastructure is built on the EU Research Network GEANT.

EGEE provides researchers in academia and industry with round-the-clock access to major computing resources, independent of geographic location. The infrastructure supports distributed research communities, which share common Grid computing needs and are prepared to integrate their own computing infrastructures and agree on common access policies. Mostly funded by EU funding agencies, this project has a worldwide mission and receives important contributions from the US, Russia and other non-EU partners. Early 2005, EGEE integrates more than 100 sites, 10000 cpus and several Petabytes of data storage.

The rapid advances in computer and information technologies are driving the community to change its mode of operation by organizing its efforts through interdisciplinary initiatives.

The proposal to port Earth Sciences applications to EGEE was warmly supported. "EGEE committee recognizes the very important integration effort made by this community and the high scientific value attached to its objectives". These applications cover several domains:

In ***Earth Observation***, all the applications concern an intensive use of data from space borne instruments. Auxiliary data are needed to validate or complement the results obtained. Several scientific topics are approached, especially atmospheric chemistry.

In ***Climate***, the applications concern an ensemble of multi-model earth system simulations, performed on supercomputer. The goal is a collaborative evaluation, processing, exchange and comparison of geographically distributed climate data.

In ***Hydrology***, the application concerns the management of coastal water in the Mediterranean region.

In ***Solid Earth physics*** the applications concern high precision geodetic measurements with GPS data, determination of earthquakes mechanisms and synthetic seismograms to share with a large community.

In ***Geosciences***, the applications are devoted to R&D applications like a seismic processing generic platform and intensive computation for modelling and imaging.

We choose a phased approach for implementation, with focus in two out of the several proposed applications. The first one is Earth Observation that consists of the production of ozone profiles from the satellite experiment GOME and their validation by using lidar data. The second one concerns Geocluster, a seismic processing generic platform, used by the Compagnie Générale de Géophysique and a large community of industrial R&D and academic users. Today both applications are delivering initial results on the Grid. The activity is supported by two virtual organisations (VO): Earth Sciences Research (ESR) and EGEODE.

Mid 2005, the deployment on the GRID on the initial applications is well advanced. The community building has been spectacular. Further deployment of two new applications, Solid Earth Physics and Hydrology is now on going.

The applications ported so far within these VO are already providing several demonstration and operational tools for wide range of Earth Sciences application domains and are fostering more collaborative work with other communities. The applications gathered around 40 academic and industrial partners in 9 countries. EGEE permits to achieve some important goals:

- In atmospheric chemistry by satellite observations, i.e. 7 years of data retrieval with 3 different algorithms. It has been relatively easy to perform a very large data production and (especially) validation in a relatively short time. The results will be submitted soon in a scientific journal.
- In seismology an innovative seismic software application, allows researchers at Institut de Physique du Globe de Paris to rapidly determine the mechanism and the central coordinates of strong-to-major earthquakes (more than 5.5 on the Richter scale) around the world. Thanks to the Grid, the best of 30,000 solutions could be found within 30 hours after the earthquake took place. This best solution was later confirmed by comparing it with results from other analysis methods.
- CGG achieved the implementation of its generic seismic platform software, based on Geocluster commercial software, which includes 400 geophysical modules. EGEE will permit to users both from academia and private companies to use it from anywhere in the world at anytime.

EGEODE membership

EGEODE is the Virtual Organization, which includes this seismic processing application. It is a Virtual Organization opened to both Academic *and Industry* research

laboratories. Research centers, which share common interest, can join as new members. Seismic data processing platform is already available. New applications on Reservoir Simulation are expected.

Membership has two folds

1 - *Authentication and authorization*. To use grid resources, each individual user must have a secure and recognized identity. This is achieved with a digital personal certificate, delivered by a specific certification authority. With this certificate, the user can then register to the VO he wants to join, subject to the agreement from the VO manager, and get a login to the grid resources. Actual certification authorities are country dependent. The EGEE web site references most of them. The EGEODE VO Manager is the central point of contact.

2 - *Sharing resources*. Each participant (or organization he belongs to) brings a small part of the global resources, according to his possibility. The current economic model assumes that, *in average*, each participant brings what he will use in term of computing and storage resources and should provide the community with a physical grid node, consisting in

- A network connection (either Internet or a direct powerful link to the research network GEANT in Europe)
- Front-end Linux PC on which EGEE middleware is installed
- One or more computing or storage PC

The current hardware, operating system and grid middleware are fully based and will evolve with EGEE infrastructure and will likely become the de-facto standard for science. The middleware is Open Source.

Conclusions

The paper shortly described Grid Computing benefits, then EGEODE, the "Expanding Geosciences On Demand" open virtual organization, with a practical description of how to join the community and become an actor in this exciting project.

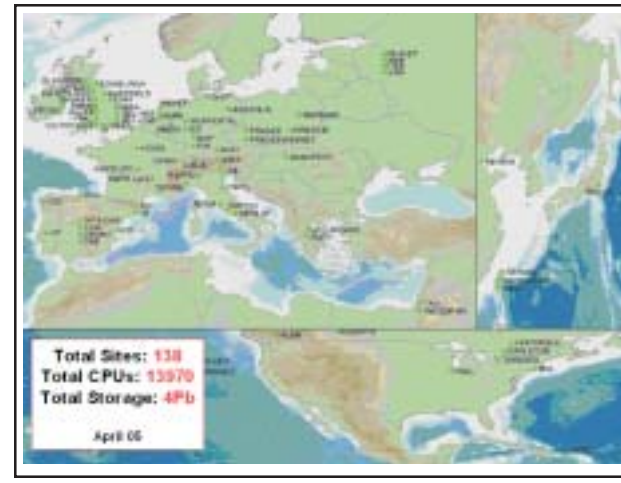

Fig. 1: Workdwide EGEE deployement

Several issues in computer sciences are still open, like MPI usage, software license management and database access to fully get all benefits of grid computing in geosciences. A major step is also expected from new economic and usage models to facilitate new membership.

The most convincing argument for collaborative, grid-based production is the achievement of scientific results which are difficult to obtain without using the Grid and which may also represent new, and original results. This is our common target.

References

Foster and C. Kesselman, The Grid, Blueprint for a new computing infrastructure, Morghan and Kaufmann Publishers Inc., San Francisci, CA (USA), 1998.

Site EGEE:
http://egee-intranet.web.cern.ch/egee/intranet/gateway.html

Site Earth Sciences Research:
http://datagrid.nadc.nl/twiki/bin/view/ESR/WebHome

EGEODE current VO Manager: vo-egeode-managers@cgg.com

Earth Sciences: Applications and Requirements, for the EGEE Generic Application Advisory Panel, EGEE Internal 2004. M. Petitdidier, J. Biercamp, C. Boonne, L. Fusco, J. Linford, O. Marti, G. Moguilny, P. Renard, W. Som de Cerff, D. Thomas and J.P. Vilotte.