Introduction of a Seismology Platform Based on the Armenian National Grid Infrastructure

Mikayel Gyurjyan[1], Hrachya Astsatryan[1], Tigran Shahinyan[1], Lilit Sargsyan[2], Mickayel Enanov[2]

1 - Institute for Informatics and Automation Problems; Yerevan, Armenia
2 - National Survey for Seismic Protection; Yerevan, Armenia
e-mail: mikayel_gyurjyan@ipia.sci.am
e-mail: hrach@sci.am
e-mail: lilit_geo@yahoo.com
e-mail: mikaele@mail.ru

ABSTRACT

To study and analyze very large scale and complex earthquake problems seismologists require to have seismological data sets from different regional areas, experimental results from different earthquake research laboratories, some good data processing, analyzing software and above all high performance computational resources. This can be achieved only by sharing the available resources.

In this paper, the main components of a seismology platform based on the Armenian National Grid Infrastructure (ARMNGI) are introduced. The main core of the platform is the seismic data server that organizes and gives performance access to distributed seismic data.

Keywords
SDS, Seismology, GRID, ARMNGI.

1. Introduction

To study and analyze very large scale and complex earthquake problems seismologists require to have seismological data sets from different regional areas, experimental results from different earthquake research laboratories, some good data processing, analyzing software and above all high performance computational resources. This can be achieved only by sharing the available resources.

In Armenia the seismological data are collected from about thirty stations and stored at the servers of National Survey for Seismic Protection of Armenia (NSSP) [1]. If seismologists want to use this massive data set, they need to download data from a web server and use their local computers to process the data. Therefore seismologists need both computational resources to solve equations that arise in the mathematical modeling of seismic phenomena as well as storage resources in order to store and access historical earthquake information and massive seismic data that are collected either in continuous or discrete data from several geographically distributed sensors.

Grid infrastructures [5] solve these problems by offering a platform where computational storage resources and other miscellaneous resources are available all connected by high speed networks. Grid computing [9] is an advanced technology of the distributed parallel calculations recently intensively developed in Europe, America, and Asia and in other regions of the world. Grid computing technology assumes a collective shared mode of access to network resources and to the services, connected to them using frameworks of globally distributed virtual organizations consisting of the enterprises and the separate experts. Actual Grid-networks consist of large-scale systems of calculations, monitoring, management, complex analysis services and globally distributed sources of the data capable to support structures of scientific, education, government organizations and industrial corporations and forming powerful e-Infrastructure for e-science. The deployment of the seismology platform on the base of Grid infrastructures and porting seismic applications is a natural choice.

At the mean time Armenia actively engaged in different international Grid infrastructures, such as South East European Grid Infrastructure within the framework of EU FP7 Programs [10]. Taking into account the importance of this direction, the interested parties established an Armenian National Grid Initiative that represents an effort to establish a sustainable grid infrastructure in Armenia [11] to expand the high performance computing resources with collaboration of academic and commercial participants and to improve national applications.

2. Seismology Platform

Many earthquake applications, such as deep crucial studies of the earth, propagation of seismic waves in homogeneities of the media need large scale simulations and computational resources. With the maturation of e-Infrastructures, the concept of building distributed systems over the Internet is becoming reality. The suggested seismology platform consists of the following levels

- Seismic data
- Seismic data server
- Programming tools and interfaces
- Applications

2.1. Seismic Data

Seismic data is stored on storage element of the ARMNGI and made available through the File Catalogue Service. The SAC (Seismic Analysis Code) [12] file format is used for seismic data. The data consists of static and dynamic parts. The static data includes information about stations (general information about stations, detailed information about sensors in the stations, response file of stations) and earthquake catalogues including main parameters of earthquakes. The dynamic data mainly includes the seismic waveform data received from various stations of Armenia.

2.2. Seismic Data Server

Seismic data often contain millions of files spread over several storage sites. To find the files of interest, users and applications need an efficient mechanism to discover and query information about their contents. This is provided by associating descriptive attributes (metadata) to files and by exposing this information in catalogues, which can then be queried to locate files based on the value of their attributes. AMGA [2, 3, 4] began as an exploratory project to study the metadata requirements of the LHC [8] experiments, and has since been deployed by several groups from different user communities, including High Energy Physics, Biomed and Earth Observation. More recently, AMGA was incorporated into the gLite [7] software stack as the metadata catalog of the EGEE project [6].

A Metadata Catalog stores entries corresponding to the entity being described, typically files. These entries are described by user-definable attributes, which are key/value pairs with type information. Entries are not associated directly with attributes. Instead they are grouped into schemas, with the schemas holding the list of attributes that are shared by all their entries. AMGA structures metadata as a hierarchy, similar to a file-system. Directories play the role of schemas; they may contain both entries and other schemas. This hierarchical model has the advantages of being natural to users as it resembles a file-system, and of providing good scalability as metadata can be organized in sub-trees that can be queried independently.

AMGA is a C++ multi-process server, with an extensible back-end that supports multiple database systems (currently PostgreSQL, Oracle, MySQL and SQLite). It offers two access protocols for clients: Web Services using SOAP and TCP streaming based on a text protocol similar to SMTP or TELNET.

In our model earthquake and station data are kept in AMGA tables whereas waveform data are stored as files. There can be millions of waveform data files. Therefore, an index of these waveform data files is also kept in AMGA tables. The AMGA tables can be queried directly by using various AMGA interfaces (shell, Perl, java, C/C++).

2.3. Programming Tools and Interfaces

Seismic Data Server Application Services (SDSAS) is a set of tools for storage, indexing and providing high level access to massive seismic data. The objective of the tool is to hide the details of where the data files reside and map high level user specification (dates, hours, location etc.) to appropriate pathnames automatically. Available seismic data is massive residing in hundreds of thousands of files. The tool enables users to avoid having to go through directory listings in order to find/use data that is of concern to him. This makes the programs more portable by not requiring the users to hard-code file pathnames.

SDSAS tool consists of two main components. The first component is the data collection component for uploading data using various scripts. The second one is a programming tool in the form of C++ classes and iterators that can be used by programmer for accessing seismic and earthquake data. The aim of the SDS C++ iterators is to provide a higher level interface to seismic data that does not necessitate seismologist users to learn AMGA and which provides additional functionalities. It allows querying AMGA catalogue automatically using high level specifications such date ranges and locations for accessing station data, waveform data files and earthquakes.

2.4. Applications

An earthquake location finding application that is used at IIAP has been gridified. A web interface that displayed the locations of the earthquakes and the stations using the Google Maps interface has been developed.

More seismic applications such as Seismic Risk Assessment, Massive Digital Seismological Signal Processing with the Wavelet Analysis, Numerical Modeling of Mantle Convection are in the deployment process [13].

Except building a seismology platform on ARMNGI and adapting routine analysis software for the Grid seismology, a high level programming web interface for seismic files will be offered to researchers for handling huge amounts of data. This high level programming web interface can be used as a wrapper for seismic data analysis applications. The initial version of the wrapper which provides transparent access to data has already been developed. This web interface has been constructed using the Open Source software resources, Linux, Apache, MySQL and PHP to allow maximum flexibility and customization.

3. Infrastructure

ARMNGI is a collection of clusters interconnected to form a computational Grid based on Academic Scientific Research Computer Network of Armenia (ASNET) [14]. ARMNGI consists of 5 clusters: ArmCluster, a 64 node cluster located in the Institute for Informatics and Automation Problems and used primarily for system software research; and homogeneous 48 processors based cluster systems established within and between the Armenian research organizations and universities (IIAP cluster, YSU cluster, SEUA cluster, IRPHE cluster).

In the seismological stations seismic data is collected from the sensors and sent to NSSP. NSSP analyses the collected data and sends it to the storage element (seismic DB) of ARMNGI. Seismology experts will be able to connect to the SDS using a normal browser and interactively access the necessary data. This infrastructure is shown in figure 1.

4. Conclusion

It is clear that suggested platform is a useful infrastructure for local seismology community. This facility, not only will help to improve the seismic design of the structures, but also, for the seismological studies on the base of available data from the countries in the region on state-of-the-art high performance computational resources spread over the country. On the hand, this will help us to share our ideas, knowledge, resources with the countries of the region by using the Grid infrastructures and hence to save our time. Earthquake research community in the developed countries, have already realized the importance of the Grids, worked upon it and new getting benefited. The usage of the platform will give the following benefits:

- Serve seismic data that is mirrored from national seismology centers using a high level interface that is easy to use/adapt.
- Use the grid platform for its seismology data and applications, which include models for detecting earthquake locations, fault-planes solution and seismic hazard assessment.
- Logical organization, indexing and update of distributed seismic data.
- Performance aspects - (high performance computing, high performance access to massive data).
- High level Web based interface that will provide easy access to seismic data and avoid requiring the users of seismic data to:
    - learn a lot of new tools
    - write data location dependent code
    - modify existing applications drastically
- Seek collaboration with other country seismology related groups/organizations.

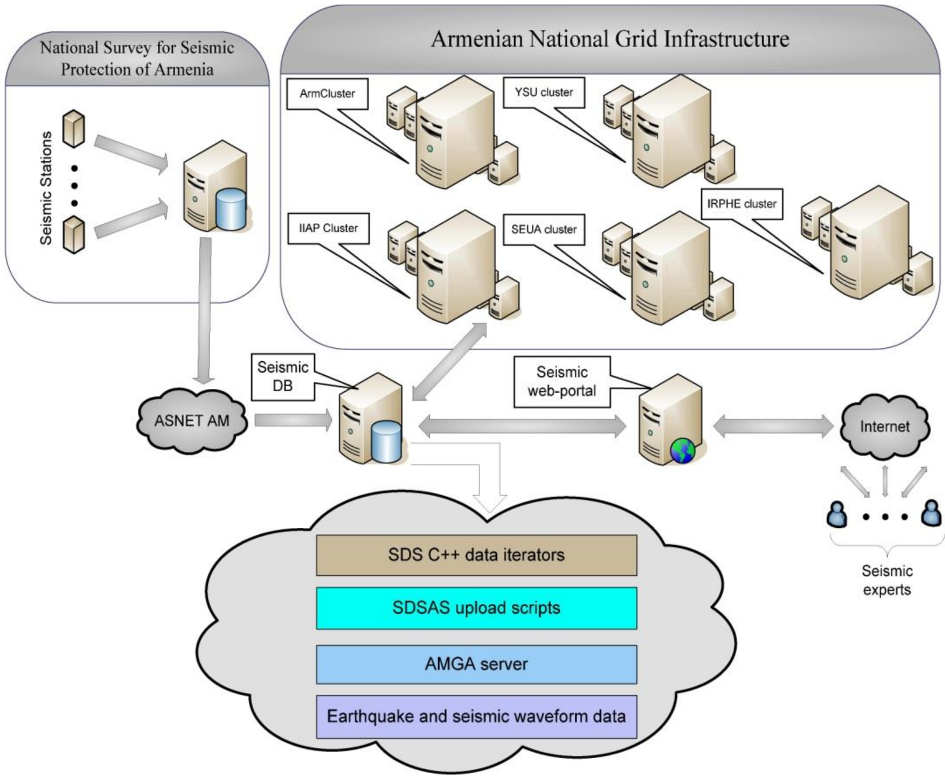

Figure 1. Seismology Platform Infrastructure.

5. Acknowledgement

The joint work has been done within the framework of Armenian State Target Program entitled "Deployment of Armenian National Grid Infrastructure". It has also be supported the EU FP7 South East European Grid e-Infrastructure for e-Science and International Science and Technology Center A-1451 Projects.

REFERENCES

[1] National Survey for Seismic Protection of Armenia: http://www.nssp-gov.am/

[2] N. Santos and B. Koblitz, "Distributed Metadata with the AMGA Metadata Catalog", in Workshop on Next-Generation Distributed Data Management, HPDC-15, Paris, France, June 2006

[3] N. Santos and B. Koblitz, "Metadata services on the grid", in Proc. of Advanced Computing and Analysis Techniques (ACAT'05), Zeuthen, Berlin, May 2005. [4] B. Koblitz, N. Santos, V. Pose, "The AMGA Metadata Service", Journal of Grid Computing, Springer Netherlands, 2008, 61-76.

[5] Grid Infrastructures: http://www.gridipedia.eu/grid-infrastructures.html

[6] Enabling Grids for E-sciencE. http://public.eu-egee.org/

[7] gLite: http://glite.web.cern.ch/glite/

[8] LHC: http://lhc.web.cern.ch/lhc/

[9] Grid Computing Info Centre: http://www.gridcomputing.com/

[10] FP7: http://cordis.europa.eu/fp7/

[11] Armenian National Grid Initiative: www.grid.am

[12] SAC- Seismic Analysis Code: http://www.iris.edu/manuals/sac/SAC_Home_Main.html

[13] SEE-GRID eInfrastructure for regional eScience: http://www.see-grid-sci.eu/

[14] ASNET: www.asnet.am