

Grid Computing In Seismic: Opportunities and Challenges

R.K. Verma & Ajith C.

GEOPIC, ONGC, Dehradun

ABSTRACT : Current advances in 3D seismic data acquisition have increased the input data volume by many folds. Processing methods have also been changed aiming higher resolution. Further, quest to image seismic data more rapidly while increasing the fidelity of the subsurface structure put forward a paramount increase in the demand for computational power. Most of the seismic algorithms have inherent characteristic of parallelism and enabled the use of parallel processing techniques to meet the growing processing requirements. A large parallel-processing platform is provided by recently evolving grid-computing technology in the form of low cost virtual supercomputer by seamlessly integrating globally distributed heterogeneous computing resources. The opportunity provided by grid computing in the form of enormous computing power is expected to fulfil the need for very high performance computing requirement in seismic processing. However, there are several challenges hindering the exploitation of grid computing in seismic processing. The present paper describes some of these opportunities and challenges in implementing this cutting edge technology in seismic processing industry.

INTRODUCTION

Seismic methods are considered as primary tools in whole gamut of oil and gas exploration. The ultimate aim of the seismic data processing is to transform the large amount of surface collected seismic data into an accurate image of the sub-surface. The main challenge before the geo-scientists today is to reduce the processing cycle time and also to enhance the quality and fidelity of subsurface image. Timely and accurate subsurface imaging not only helps reducing the risk but also provides vital inputs for reservoir management and production optimization. The primary requirement for high-resolution imaging is the high definition input data i.e. seismic data to be acquired at much more closely spaced points along the surface of the earth with finer sampling density. This has led to the increase in input data volume by several folds. Also the 3D imaging, which was traditionally done on a reduced data volume (post stack) to save time and cost, is now routinely emphasized on full volume as pre stack time and depth migration. Thus large data volume and complex mathematical algorithms make seismic data processing an extremely compute and I/O intensive activity, which is beyond the scope of serial architecture machines and therefore high performance computing is the only answer to this challenging application.

A large portion of seismic processing can be parallelised with simple domain decomposition. Thus parallel computing is providing a creative solution to achieve high performance in seismic computing. Recently seismic industry has recognized High Performance Computing Clusters (HPC), which combines multiple Symmetric Multi-Processor (SMP)

together with high speed interconnects, to achieve high performance at much lower cost. Also cost of the computing power is going down day by day as most of the processing companies are opting economic clusters. However, if we consider all the aspects e.g. maintenance, providing and supporting necessary infrastructure etc., it can be realized that owning of computer resources is not at all cheap. Further, it is a common observation that in most of the enterprises approximately 10% of available CPU time is being utilized. This inspiring fact and the dramatic increase in network bandwidth led to the development of highly cost effective computing technology – the grid computing. The basic philosophy of grid computing is to enable the utilization of untapped remotely located computing resources, regardless of their operating characteristics, to be virtually shared, managed and accessed across an organization, an industry or a bigger work group. To a user it gives a single system view and its concept is almost analogous to the electrical power grid, where excess power at one point is utilized at the other. Grid computing will provide seismic industry, a big leap to address highly compute intensive processes like pre-stack depth migration (PSDM) and new upcoming algorithms based on wave equation, elastic modeling, finite difference (phase correction) methods, etc.

GRID COMPUTING: AN OVERVIEW

Grid computing emerged in response to scientific users' need to tackle complex number crunching applications by pooling together large amount computing resources. In a broad sense grid computing technology provides the means of sharing, selection and aggregation of heterogeneous

computing power and storage resources distributed over wide area thereby offering wide area parallel and distributing computing capabilities. There are two aspects of grid computing; compute grids for distributed CPU power and data grids for distributed data. Grid computing infrastructure can be built by coupling globally distributed heterogeneous resources, which may be owned by different organizations, through high-speed network and deploying a number of special services. They include low level services like security, information directory, resource management, scheduling, load balancing and high level toolkit for application development and resource brokerage. Thus a grid can be considered as consisting of several layers (Fig. 1) which can be described broadly as follows:

- **Grid fabric:** This is basically the hardware resources, which could be computers of various makes running on variety of operating software, storage systems and network components.
- **Core grid middle-ware:** This provides services like mediating different computer operating systems, remote process and storage management, resource scheduling, load balancing, prioritizing work from multiple tasks and security, etc.
- **User-level grid middle-ware:** It includes commands and tools to quickly develop grid based applications and provide underlying elements for resource selection, brokerage and portability.
- **Grid applications and portals:** Grid application are normally developed using grid enabled languages and employing message passing libraries like MPI (message passing interface), PVM (parallel virtual machine), OpenMP etc.

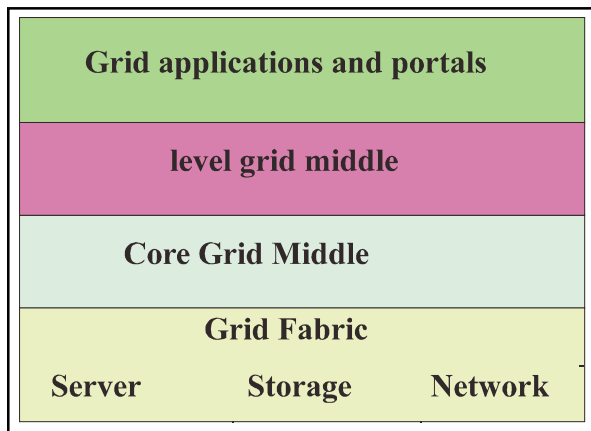


Figure 1 : The basic Grid Architecture

Grid portal is the web-enabled interface through which users submit their jobs and receive the results.

Various toolkits and grid middle-ware are now available, even in open source domain, to implement grid infrastructure by connecting remote computing resources through high-speed network. Important ones among them are Globus toolkit, Legion, AppLes, Condor, NetSolve, etc. Projects are also underway to enrich and modify them further.

OPPORTUNITIES

There are many opportunities which Grid computing has in its offing for seismic environment. A few of them are described below:

A source for vast computing power

Seismic applications, which are characterized by complex number crunching operation on a very large volume of input data, have always been computer resources hungry and pioneer in adopting new developments in computing technology. Large-scale exploitation of some of the seismic algorithms is still limited by the availability of processing power even on a supercomputer/ large parallel cluster. Computational grids equip geo-scientists with enormous computing power in the form of co-ordinated mix of workstations, supercomputers and clusters to feed all such algorithms and the processes like 3D pre-stack depth migration, wavelet processing, full wave form inversion, etc. may be routinely exercised. Reduction in turn around time may even bring to reality the dream of geo-scientists that complete surveys to be shot, analyzed and ready for interpretation by the time all data is captured.

Economic and productivity benefits

Cost of computing resources are going down day by day but owning them is not really cheap. If resources are over-provisioned for estimated future peak load then there is danger of unutilized capacity. Provisioning of resources for average base load may lead to higher cycle time and loss of productivity. Grid computing provides a resilient/ highly available infrastructure, which can be scaled up or down by a dynamic cost component thereby turn high performance computing from fixed cost to variable cost. Corporates no longer need to estimate and predict their computing needs and also save themselves from resource obsolescence. Even a small-scale, organization wide corporate grid set up optimizes resource utilization by eliminating under utilization. Therefore, grid computing offers substantial reduction in infrastructure

and resource management costs on one hand while increasing throughput and performance on the other.

LICENSE SHARING

Seismic processing and interpretation software licenses are very expensive and they should be most optimally utilized. Corporate grid computing set up automatically provides a means of sharing unused application software licenses among geographically separated work centres of an upstream/ seismic service company. A tremendous advantage can be obtained if these work centres fall under opposite time zone.

VISUALIZATION AND DISTANT COLLABORATIVE WORKING

The physical grid infrastructure, in the form of already in-place high-speed broadband network, can provide a new dimension to multidimensional visualization. Grid computing will provide the needed resolution, speed and realism to visualize the enormous geo-scientific data in a variety of manners to find the key subtle traps and stratigraphic features. The real time visualization also provide an added opportunity to increase multi-disciplinary collaboration among experts within an organization and across partners through on-line remote conferencing, display forum, and virtual reality (VR) centres.

DATA GRID; MORE DATA ONLINE

The data-grid offshoot of grid computing provides an opportunity to store massive seismic datasets online by combining the capacity of multiple data centres through secured broadband networks. Data can be processed then and there without being physically moved. It also provides a platform for setting up federation of data, which enable increased access to data and collaboration in industry. Data bank and data mining tools are embedded in most of the grid middle-wares.

CHALLENGES

There are many challenges on the way to exploit this potent technology of grid computing in seismic environment. Some of them are addressed below:

COMMUNICATION BANDWIDTH

The basic communication infrastructure for grid computing as generally thought of is the Internet. However, in a developing country like India, barring some privileged locations, achieving a speed of a few MBPS on a common Internet platform is a distant dream. Therefore grid computing

can not be built on Internet, especially for seismic environment where a large amount of data is to be transferred almost in real time across grid nodes to achieve the real benefit. This challenge can only be addressed by employing corporate or leased, highly available very fast network; not only for seismic but for other high performance computing requirements also.

Batch oriented seismic jobs do not demand real-time system response. So available high-speed network back plane can be shared among other corporate applications to increase productivity. Recently researchers at Stanford Linear Accelerator Center could achieve a transcontinental data transfer speed record of 923 Mbps over a distance of 11,000 Km using standard TCP/IP protocol stack on Inetnet-2 platform. It will be some time before this speed feat gets translated into real life use. Nevertheless, coupled with ever increasing high speed storage and I/O capabilities on today's computer systems, network speed of this magnitude will boost the confidence on grid capabilities. Seismic processing centers can be connected using fiber optic links, satellite links, microwave etc. to create the communication infrastructure.

PLATFORM INDEPENDENT CODE FOR SEISMIC PACKAGES

The concept of grid computing is based on employing heterogeneous computing resources running on a wide variety of operating systems like Linux, Unix or Windows. Therefore, the very nature of computational grid demands platform independent binary code i.e. the executable code, which can run on any hardware or software platform seamlessly. As of now in seismic industry, perhaps none of the commercially available, reputed seismic software qualify this requirement. However, most of the compute intensive commercial seismic software, at source code level is platform independent and mostly developed using open parallel processing tools. Reputed seismic software venders provide executables on multiple platforms on a single licensing agreement. This scenario enables the processing of data using the same application on multiple platforms, provided executables of the same software are installed on these platforms (or installed centrally on one server shared by appropriate resource at desired time). Therefore, to some extent, this challenge can be addressed and the benefit of grid computing still can be had by selecting the desired platforms from among the vast grid. It is also expected, as the grid computing further develops, seismic software venders may come with grid enabled seismic software.

Application software licensing

There is also a challenge posed by licensing issues which is very important and critical in the present scenario. As a normal industry practice, for seismic data processing

software, licensing fee is charged based on the throughput obtained with the software. Thus licensing fee not only depends on number of licenses but also on the number of processors and power (MIPS equivalent) of each processor. For example, licensing fee of a processing software on a desk top work station is very small but same software, with same code may cost several times more on a multi processor server. That means, if we increase the number of CPUs of the server or even replace existing server nodes with faster nodes, then the existing licenses will not be valid unless the necessary licensing fee is paid to the software vendor. In grid computing, as the concept goes, large but variable number of processors shall be employed with large variations in their speeds. Thus, rationalizing the software licensing costs is yet to be made clear by the software providers. They are to come out with a convenient methodology for licensing fee in grid computing environment. However, in case of interpretation software, the issue is not that critical as the licensing fee normally depends on the number of seats (licenses) and does not depend on the power and capacity of workstation/ server.

Security

Data security is vital in current E&P scenario. None of the issue is more challenging than the security of data itself which may either come from the security of the nation as data may belong to some sensitive location or due to any other reason. Most of the seismic processing organizations, therefore keep their processing network isolated from the Internet or build a strong firewall around it. This is also a challenge before the processing service providers, if grid computing based on world wide resources needs to be used, as the vendors have to give security undertaking to their clients. Therefore, grid computing in seismic, in general currently limited to connecting processing centers of a service provider/ upstream corporate employing their own networks.

Standardization

There are many other challenges in general in the concept of grid computing. To achieve real success in a grid-computing environment, distributed systems components need not only talk to one another but also need to retrieve data from remote resources. Therefore, standards must be set for combination of remote data access, communication protocols, networking and possibly encryption also. In the dimension of standardization of grid protocols, major efforts are underway within Global Grid Forum (GGF) to define the Open Grid Service Architecture (OGSA). Similar efforts are also effectively taken-up by Interoperable Informatics Infrastructure Consortium (I3C). As of today, on practical front, some de-facto standards are produced by Globus toolkit and other similar open source products. These are to be modernized

and extended to meet the emerging new requirements with the support of global computer industry. Open standards help accelerate the availability of the grid computing technology. Further, the resources employed in grid computing will be owned by different organizations with their management policies, usage and cost model for different users at different time zones. Thus the concept of computing commodity unit also needs to be developed and standardized. Prospectwise accounting in seismic scenario is also an issue to be looked into.

CONCLUSION

The concept of grid computing in providing low cost computational power, by harnessing unused computing resources is a noble one. It serves as a scalable computing platform for executing large scale computational and data intensive problems. Such grids already exist for providing rapid solution to large-scale problems in science, industry and commerce. Grid computing is committed to deliver significant value to resource owners, service providers and consumers by providing a means in improving their capital investment and operating cost structure. Seismic processing industry is also looking for many opportunities in this technology. However, when the practical aspects are considered, on ground reality, there are challenges and concerns which are to be addressed before leveraging this emerging technology in seismic processing environment.

ACKNOWLEDGEMENT

Authors wish to thank ONGC for providing necessary support in writing this paper. Authors are grateful to Dr. C. H. Mehta, GGM - Head GEOPIC, ONGC for his encouragement and keen interest in the concept. The authors also express their gratitude to Shri S. M. Bandopadhyay, DGM (Programming), GEOPIC, ONGC for his continuous guidance. The support extended by all the colleagues are also deeply acknowledged.

REFERENCES

- Buyya R. April 2002, Thesis entitled "Economic-based Distributed Resource Management and Scheduling for Grid Computing" submitted in fulfillment of the requirement for the degree of Doctor of Philosophy in the School of Computer Science and Software Engineering, Monash University, Melbourne, Australia.
- Karbarz F. Grid Computing for Seismic Processing, The Leading Edge, January 2003, p 58-60
- Premont R. Grid Computing: will it fly? Hart's E&P, June 2002, p 77-78
- Telecom Asia, April 2003