

# A Grid-enabled system for 3-D seismic imaging

D. Bevc, I. Musat, A. M. Popovici, and S. E. Zarantonello

3DGeo Development Inc., Suite 401, 4633 Old Ironsides Drive., Santa Clara, CA 95054  
{dimitri, iulian, mihai, sergio}@3dgeo.com

**Summary.** One way to meet the increasing computational demands of the seismic imaging industry is by seamlessly enabling access to compute on demand capability. Grid computing, by leveraging the Globus Toolkit<sup>1</sup> middleware, provides a framework where remote computational resources can be accessed *whenever and wherever needed*. The objective of this paper is to give an overview of a Grid-enabled environment for seismic imaging developed by 3DGeo. This system addresses the demand for advanced seismic imaging applications in the oil and gas industry, and the ensuing need of computational and data resources to run these applications flexibly and efficiently. 3DGeo has developed INSP<sup>2</sup>, a proprietary Java based Internet infrastructure for remote collaborative seismic processing and seismic imaging, and a suite of advanced 3-D imaging applications that can be accessed, executed, and monitored with the INSP system. The conversion of INSP to a Grid-enabled system, providing flexible and secure access to advanced imaging applications and to the resources to run these applications as needed, is a fundamental step in the seismic imaging industry. Advanced compute intensive imaging, enabled by the Grid, will allow the industry to much more effectively meet the imaging challenges presented by the increasingly complicated geological regimes we are exploring today.

**Introduction.** Seismic imaging methods are usually classified as methods based on the Kirchhoff integral equation and methods that operate directly with the wave equation. Wave-equation methods are further classified as shot-receiver (*e.g.* common azimuth and narrow azimuth methods) and shot-profile methods. To put the computational challenge in perspective, in Table 1 we compare the estimated runtimes of hypothetical imaging projects for Deep Gulf of Mexico 3-D marine surveys on a 128-CPU cluster of 2.4 GHz Pentium® 4 processors delivering a sustained performance of 900 Mflops/CPU.

**Table 1. The computational challenge: Gulf of Mexico 3-D marine surveys.**

<i>Size of data</i>		<i>Runtime in days</i>		
<i>Blocks</i>	<i>Gbytes</i>	<i>Kirchoff</i>	<i>Narrow azimuth</i>	<i>Shot profile</i>
10	620	3	31	184
100	6,200	111	1,100	6,640
500	30,700	996	9,960	59,800 (164 yrs !)

<sup>1</sup> I. Foster and C. Kesselman, *Globus: A Metacomputing Infrastructure Toolkit*, in the International Journal of Supercomputing Applications and High Performance Computing, vol. 11, issue 2, pp. 115-128, (1997)

<sup>2</sup> D. Bevc and A. M. Popovici, *Integrated Internet collaboration*, The Leading Edge, vol. 22, pp. 54-57 (2003)

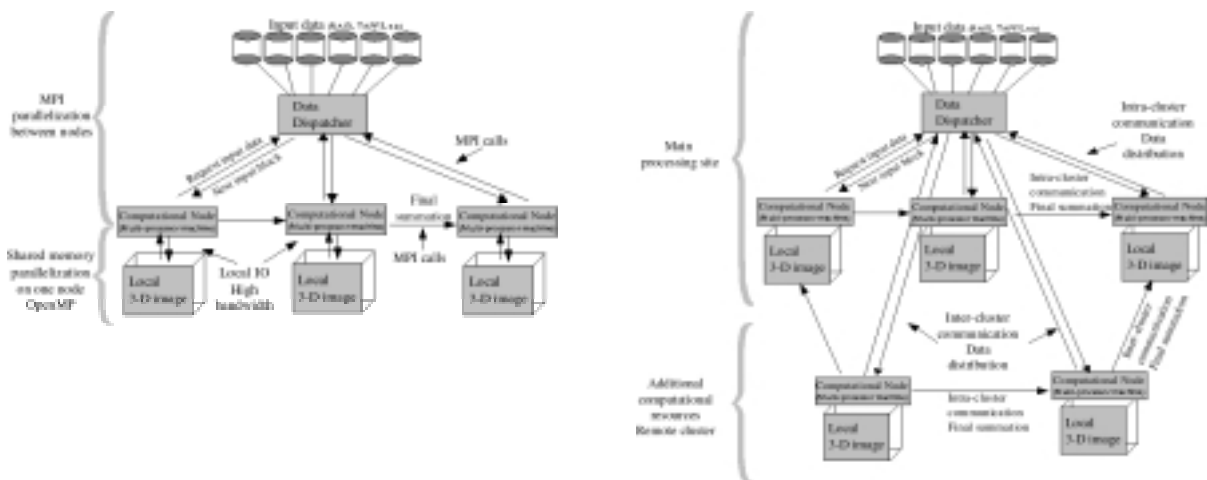
The generally computationally more intensive wave- equation methods give better accuracy than Kirchhoff migration, underscoring the need for more compute resources deliverable through the Grid.

**Overview of PSDM.** We use PSDM as an example to illustrate design issues that were addressed for Grid deployment. Prestack Depth Migration (PSDM) is 3DGeo's implementation of the three-dimensional Kirchhoff depth migration. PSDM approximately solves the wave equation with a boundary integral method. In a computational setting we express the integral as a sum:

$$\text{Image}(x) = \sum_{x_s} \sum_{x_r} A_s A_r \text{Input}(x_s, x_r, t_s + t_r)$$

where  $A_r$  and  $A_s$  are determined by the transport equation, and  $t_r$  and  $t_s$  are either found by ray-tracing or by solving the eikonal equation. We note that the sums for the different points of the image can be calculated independently from each other. The process of building an exact image is iterative, with successive improvements made to the velocity field and can be extremely demanding in terms of human and computational resources.

**Parallelization of PSDM on a cluster.** PSDM presents different parallelization issues in cluster and Grid environments. 3DGeo's implementation of Kirchhoff migration was designed to achieve maximum efficiency on a cluster of interconnected multiprocessor computers. The input data is distributed among the computational nodes, while the output image is divided into processing blocks that are distributed over parallel processors on each node. At the end, the results from each node are gathered to build the final image. Figure 1 illustrates the PSDM MPI architecture. We chose the OpenMP<sup>3</sup> standard for the implementation of the shared-memory parallelization on each multiprocessor node.



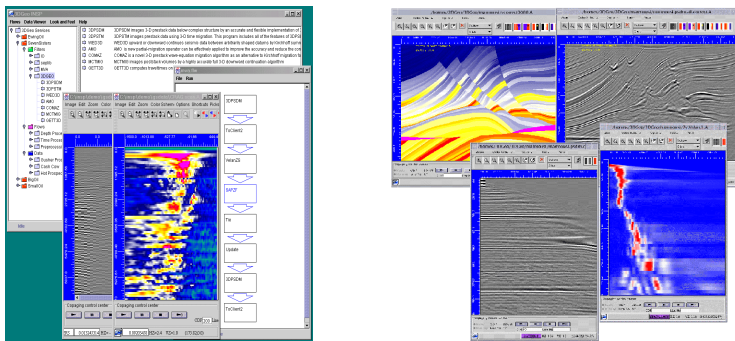
<sup>3</sup> <http://www.openmp.org/>

**Figure 1. Parallelization of PSDM on a cluster (left) and on multiple clusters (right)**

**PSDM on the Grid.** Tests and benchmarking of PSDM on various cluster architectures and configurations shows that in a typical run the I/O operations associated with the input data distribution account for a small fraction of the total processing time. This encouraged us to use, in a first phase, the same architecture for distributing a PSDM job across multiple clusters, interconnected in a computational grid, as shown on the right side of in Figure 1. Using the Globus Toolkit<sup>4</sup> we built such a Grid, interconnecting two of 3DGeo's processing centers (Santa Clara, California, and Houston, Texas) and a Linux cluster at the San Diego Supercomputing Center. The result was a Grid-enabled MPI implementation built on top of MPICH and Globus API.

**Demonstration on the Virtual Computer.** The SDSC cluster was configured to support the seismic imaging software and was connected to 3DGeo's distributed monitoring Grid. The connection to the monitoring grid was performed by the Ganglia Resource Monitoring tool. Ganglia<sup>5</sup> is an Open Source application, originally developed by the Berkeley Millenium. Ganglia gives a view of resource utilization, and includes graphs showing the evolution in time of machine load, memory usage, number of processes, *etc.* It was thus possible to inspect from a single site the load of the machines geographically distributed in Santa Clara CA, San Diego CA and Houston TX. The final step was to incorporate the Grid-enabled applications such as PSDM within the INSP framework.

**Tying Grid resources together with INSP.** INSP<sup>6</sup> is a collaborative environment, developed by 3DGeo for building and launching workflows of computationally intensive parallel and distributed jobs, visualizing data on client workstations, and monitoring jobs. Figure 2 demonstrates some of the INSP functionality. It shows a screen shot of the INSP Explorer interface. Through this interface, for example, tree structures for four servers in a 3DGeo operational Grid can be shown: Santa Clara, CA. (3DGeo Services), San Diego, CA (SDSC cluster), Houston, TX (3DGeo Houston Processing Center), and Terabyte<sup>7</sup> (a Houston-based provider of computational resources to the oil and gas industry).



<sup>4</sup> I. Foster and C. Kesselman, *Globus: A Metacomputing Infrastructure Toolkit*, in the International Journal of Supercomputing Applications and High Performance Computing, vol. 11, issue 2, pp. 115-128, (1997)

<sup>5</sup> <http://ganglia.sourceforge.net/>

<sup>6</sup> D. Bevc, M. Popovici, and B. Biondi, *Will Internet processing be the new paradigm for depth migration interpretation and visualization ?*, First Break, vol. 20, no. 3, March 2002

<sup>7</sup> <http://www.terabyte.com/>

**Figure 2. INSP Explorer is the Internet-based GUI for remote processing services.**

**Conclusions.** Large 3-D seismic imaging projects typically involve input data sets of 10 to 15 terabytes in size. With current technologies, it can take six to nine months of computer time to generate an image. By leveraging Grid resources effectively through our technology we expect this turn-around time to decrease by at least a factor of 10 and also make it practical to apply more accurate - but computationally more expensive imaging algorithms on multiterabyte data sets. In summary, A Grid environment for seismic imaging signifies a quantum leap over current methods in terms of efficiency and capability. Landmark and Schlumberger Information Systems, through their association with IBM in the Deep Computing initiative, and GXT technologies, are all presently deploying Grid-enabled products for the oil and gas industry. We believe these Grid environments are not as optimally designed and deployed for 3-D seismic imaging as they could be, and that their functionalities overlap at most minimally with the Grid technologies presented in this paper. Ultimately, the goal of Grid Computing is to provide compute resources on demand, and enable the purchase of compute power over the internet as seamlessly as electricity is purchased over electrical grids today.



**Figure 3. In today's operational scenario data are physically transported between the steps of acquisition, data processing, data storage, and the end user (oil company). The fully-enabled Grid scenario allows greater access to resources (data, computers, personnel), reduces turn around time, and ultimately shortens the time to making a drilling decision. In the fully-enabled Grid scenario, all components of the process become Grid nodes.**

Seismic imaging is an application area where Grid computing holds great promise. Today's operational environment, shown in the left panel of Figure 3, involves many inefficiencies that are seamlessly resolved in a Grid environment. A large imaging project today can easily require one to two years to complete. In a fully-optimized Grid operational environment, shown on the panel on the right, the time to complete the same project can be reduced to one half-year. These advantages are compelling, thus motivating 3DGeo's efforts to be at the forefront of bringing the Grid to the energy exploration industry.

**Acknowledgement.** This work was supported in part by NASA. We take this opportunity to thank Tom Hinke, of NASA and Kevin Walsh of SDSC for their insight and support.