P-208

High Performance Computing in Geosciences: Promises & Challenges

Mukesh Kumar[1]*, D.S.Manral[1], M.K.Banerjee[1], K.Karmakar[1], A.Das[2], B.J.Reddy[1], Dr. R.Dasgupta[1] & S.N.Singh[1], *Oil India Ltd.*

Summary

The demand for hydrocarbon resources as the prime energy driver has been witnessing a continual upside worldwide. The gruelling demands have necessitated geoscientists to explore & discover new hydrocarbon reserves while maximizing production from existing and ageing oilfields. Geoscientist's in the oil & gas industry are tirelessly focusing their knowledge and experience tounravel the mother earth for locating & extracting hydrocarbons from increasingly obscure and challenging geologic locales in the subsurface through integrated mapping, imaging, interpretation and reservoir grade analysis of geoscientific and engineering datasets.

The challenges posed by the invincible mother earth in quest of hydrocarbon resources have necessitated invention, innovation & application of cutting edge technologies, techniques, work processes & knowledge in various facets of the E&P value chain. These advances in varied facets could be successfully translated into practice only through enabling computational technology which has indeed shown an unprecedented growth and complemented the Geo-scientific initiatives head on more often then not. Moreover, if we look back, today computing has evolved to a stage where High performance computing has become affordable, efficient and space convenient and holds a promising future, though challenges will always be part and parcel of it given the appetite of geoscientist to keep on innovating and asking for more.

As a corollary, in the last few decades, the upstream industry has made a journey from seismic trace operations to elastic wave modelling to understand the Earth at higher and higher resolution and on a larger and larger scale. In recent times compute intensive imaging applications such as reverse time migration, iterative and interactive velocity model building and full waveform inversion amongst others have significantly increased the role of High performance computing and made it more vital than ever in the oil & gas industry.

In this article we review and delve on how the computing and E&P industry are complementing each other while emphasizing on the key elements of HPC system, trends, challenges and way forward.

Keywords: *HPC, hydrocarbon, geoscience, computing*

Introduction

The developments in Geosciences and Geophysics in particular have undoubtedly lead to a paradigm shift in all facets of the E&P Value chain right from data acquisition to imaging , interpretation and reservoir management etc.. Several geophysical techniques have evolved over the years covering all domains of the E&P value chain for exploration, exploitation and reservoir characterization viz. 3D seismic, 4D seismic, Multi-component seismic, wave equation based migration, inversion, reservoir simulation methods etc. These techniques are primarily aimed at unraveling the formidable subsurface challenges for identification & delineation of hydrocarbon resources.

Moreover, advanced methods for investigating the Mother Earth involve in use of complex theoretical models requiring extremely compute intensive applications/ algorithms to handle such complexity and the huge data sets generated in the process. These advances have enabled oil

[1.] *Geophysics Department ,Oil India Limited, Duliajan, Assam-786602.*
[2.] *IT Department ,Oil India Limited*
mukesh.gpbhu@gmail.com

& gas industry to take several initiatives for carrying out target oriented exploration & development activities in terms of large scale 3D, 4D & multi- component surveys comprising of high channel counts, finer shot & receiver spacing & fine sampling rate amongst others. This has resulted in an explosion of data volume from several gigabytes to few terabytes. These developments have dramatically increased the volume of data and data imaging requirements, a schematic diagram illustrating computing advances vis a vi advances in geoscientific applications is shown in Figure 1.

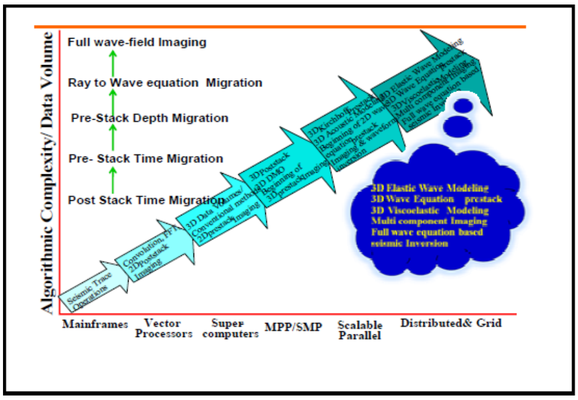
Algorithmic Complexity/ Data Volume
Full wave-field Imaging
Ray to Wave equation Migration
Pre-Stack Depth Migration
Pre- Stack Time Migration
Post Stack Time Migration
Seismic Trace Operations
Convolution, FFT 2DPoststack Imaging
3D Data Volumes/ Conventional methods 2Dprestack Imaging
3DPoststack 3D DMO Beginning of 3Dprestack Imaging
3DKirchhoff prestack 3D Acoustic Modeling Beginning of 2D wave equation prestack Imaging & waveform inversion
3D Elastic Wave Modeling 3D Wave Equation prestack 3DViscoelastic Modeling Multi component Imaging Full wave equation based seismic Inversion
3D Elastic Wave Modeling
3D Wave Equation prestack
3D Viscoelastic Modeling
Multi component Imaging
Full wave equation based
seismic Inversion
Mainframes
Vector Processors
Super computers
MPP/SMP
Scalable Parallel
Distributed& Grid

Figure 1: Computing complementing developments

Key Drivers of HPC in Geosciences

The oil and gas industry has always been at the forefront of adopting and implementing technologies from various disciplines be it physical, biological, material, or computational sciences amongst others in quest of hydrocarbon resources. Therefore, it was not surprising when it used the unique capabilities provided by supercomputing for over decades, wherein it took buildings to house the hardware and months to completely iterate through a project's dataset. Infact the past generation of geoscientist were often so far ahead of the curve that they had to wait patiently for computing technology to evolve to their expectations for testing/implementing innovative algorithms for solving complex 3D problems thereby indicating the presence of a symbiotic relationship between Geoscientist and compute technology.

Geoscientist's in the oil & gas industry still retain the same appetite and keenness as their past generation to innovate in their own domain and continually place demand on the computational technologies, while being in a ready state to tap the compute power of leading high-performance computing technologies available as on date. Moreover, the multidisciplinary activities of Geoscientists engaged them in adoption and application of leading edge technology and spurred the development of new computer technology and enhanced the discovery of energy resources.

The developments in geoscience over the last few decades and the ever increasing challenges to explore and develop hydrocarbon resources have led to the usage of several cutting edge techniques and technology. The increased data volume thereof, usage of compute and data intensive imaging algorithms in pre stack time & depth domain for unravelling geologically complex multifarious subsurface, reservoir modelling, simulation studies, high resolution visualization & real time data rendering have necessitated the usage of high performance computing, scalable storage, high performance interconnect fabrics, high bandwidth, visualization facilities and efficient data management solutions amongst others. The Recent advances in seismic interferometry and accurate depth migration algorithms have enabled geoscientists to image complex geology with a higher degree of confidence. In fact the rapidly evolving field of exploration seismology is marked by continual algorithmic advancements and the processing and analysis of large datasets that are indeed pushing the limits of HPC.

However, it is pertinent to note that traditional approaches cannot possibly manage the explosion of dataset and their compute intensive nature for making informed & confident decisions that are relevant to data in huge volumes (3D/4D/3C) seismic, derived seismic attributes, empirical relationship mapping, rock physics, wells, interpretation objects, time-variant geologic processes etc. Moreover, the continual focus on adoption and implementation of emerging technologies, burgeoning data volumes generated thereof and advanced analytical/model based approaches/ processes have placed exceptional demands on computing landscape& decision makers in the E&P industry worldwide to execute their oil & gas workflows/processes efficiently. A typical oil & gas work flow/process employing cutting edge technology is illustrated with the help of a schematic diagram in Figure 2.

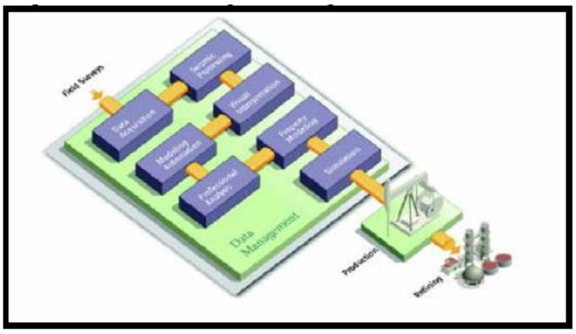

Figure 2: Typical Oil & Gas work flow.

In view of above backdrop it emerges that all aspects of computing, including data management, processor arithmetic-unit speed, memory bandwidth and latency, interconnect performance, IO bandwidth, Visualization, and power consumption play a crucial role in exploration & development activities right from data acquisition to reservoir simulation and monitoring. Hence for imaging & integrated interpretive geoscientific applications one requires fit for purpose hardware architectures/landscape to optimally utilize the various applications throughout the E&P value chain.

In corollary, HPC landscape is continually evolving to meet the pressing demands and hence has emerged as a mainstream technical tool in the oil and gas industry worldwide. Fuelled by new and more economic server-based cluster computing technologies, and by modeling/simulation software developed specifically for exploration and production, supercomputing has today reached to desktop. Today HPC has become a potent and critical tool spanning diverse aspects including everything from timely modeling and simulation of large-scale physical systems, to processing, analysis and interpretation and visualization of huge seismic data volumes. Few applications, such as waveform inversion and true wave equation MVA which were previously thought to be computationally impossible or impractical are now making inroads into routine commercial application.

Brief Computing Chronicle: Geophysical Perspective

In the late 1950s, the introduction of computers to the oil & gas sector transformed analog seismic processing into digital, and fostered the growth of new downhole logging tools. Earlier in late 1970s the seismic data was processed via punch cards shown in Figure 3. Data were saved on 12 inch reel tapes and were uploaded by slow tape drives at 1600 BPI.These punch cards were replaced by mainframe computers and 32-bit floating point array processor (Phoenix, Tempus32 & Megaseis) in course of time. Theses processors enhanced the computational aspect of algorithms based on Fast Fourier Transform. Datas were still stored on reel tapes and storage & Maintenance of these reel tapes was an issue.

Continuous development of geophysical technology along with computer industry eventually brought seismic Interactive interpretation workstations (IIW) by the mid-1980s.These workstations allowed the integration of various geophysical and geological data sets for improved analysis in reduced project cycle time and the concept of 'multidisciplinary geo-scientific team' came in to existence. The success of the IIW led to design enhancements and introduction of RISC (reduced instruction set computer) workstations.

Figure 3: Punched card and card reader, Computer tape drive

After the revolution in geophysical industry in early 1990s through development of data acquisition technologies like 3D seismic survey (multi-azimuth and wide azimuth) and pre-stack time & depth migration of 2D seismic data, Geophysicists ardently sought faster computational technology/system to shorten processing project time. From The beginning, 3-D seismic was enabled by computers. Acquisition required computing technology in the field just To record, store and manage the digital data; processing required powerful systems in the computing center; and interpretation and visualization of large data volumes required a new breed of interactive workstations in the office. The fastest processors of that time, supercomputers, did not compete with the need of exploration industry effectively and economically. Also, it's designing, accommodation, safe operation and high cost was an issue. Supercomputers were replaced by massive parallel processor (MPP) systems. Traditional computers used a single CPU to solve computational problems while with MPP many interconnected CPUs could be used simultaneously to complete a job in shorter time. Towards the end of decade, the geophysical industry concentrated its resources in improving the multitasking capabilities (enhanced graphics & memory) of the interactive workstations, menu-driven user interface; window based interactive processing techniques etc. Geophysicists no longer limited themselves to 2D & 3D seismic data API but adopted and implemented other technologies for carrying out 3D reservoir characterization, modelling and other simulations studies. This all demanded more data storage, memory and standardization and networking capabilities. However, the shared memory computer initially inhibited the performance of MPP systems because the increasing number of CPUs gaining access to the main memory became a bottleneck and performance was impaired. The problem was corrected by distributing memory between the processors and interconnecting them via a communication network to form a distributed system. This enhanced the computational speed manifold. The concept of distributed memory and more number of computational chips increased the computation in the range of Teraflops. But unfortunately the price remains the problem still. Using the same concept of parallel computing, the geophysical industry began to adopt the scalable clustering architecture and operating system.

High Performance Computing: Industry's Choice

In recent years, tremendous advancement in computer capabilities in terms of ease of programming (object oriented), flexible storage, computational speeds volumetric visualization have led to a paradigm shift in the way large volume of datasets and compute-intensive applications are applied in varied facets of the E&P value chain. The computing systems and their efficiency have changed with time and will continue to evolve. The advances and fundamental premises of the cluster based high performance computing systems have made them the choice of industry instead of mainframe computers, stand alone supercomputers and massive parallel processors.

High performance computing (HPC) has been revolutionized by a fairly new scalable clustering technology approach which is based on open architecture and open standards. HPC is basically cluster-based computing technologies that harness the networked power of multiple smaller and more affordable commodity hardware pieces. Today the HPC clusters are capable of providing the computational power equal to that of supercomputers thereby making them an appropriate candidate to address the varied challenges of oil and gas exploration.

The growth of cluster computing has been fueled by a number of key technology developments in recent years. First is the rapid advancement of CPU technology, in accordance with Moore's Law. Second is the commoditization of high performance networking technology, which provides the interconnection network, required for cluster computers to communicate with one another. Last is the maturation of the software infrastructure required to orchestrate the activities of hundreds or thousands of cluster computers, making the task of effectively harnessing these hardware advancements accessible to a larger group of programmers. The key factors which made high performance computing systems more demanding in E&P sector are following:

- Improved seismic data processing efficiency
- Able to perform data intensive computations and tackle complex algorithmic challenges.
- Lowered the risk and project cycle time.
- Eliminated single point of failure.
- Eliminated the performance bottleneck
- Better visualization solutions
- Flexible storage infrastructure
- Vale addition at reasonable cost
- Open architecture and open standards
- Energy & Space efficiency

Over the past few decades, geoscientists have developed new technologies and techniques for unraveling the Earth employing complex theoretical models for gaining more subtle insight of the subsurface while making it applicable to huge data sets. Many of the advances in geoscience have changed the way we perceived, imaged and characterized the subsurface. Notable advances among them include 3D anisotropic prestack depth migration, Beam Migration, Reverse Time Migration for imaging the Earth's reflectivity, 3D wavefield tomography for deriving complex velocity models, 3D visualization techniques for effective interpretation & prospect identification, more accurate time lapse and simulation studies, sophisticated fluid-flow simulators for understanding oil and gas production, and digital rock simulators for understanding the subtle relationships between the measured fields and the underlying geology.

The continual evolution/development of high performance computing system have today enabled geoscientist's to acquire, store, manage, process, visualise and interpret very large-scale data volumes quickly and with confidence. Infact High performance computing has become exciting not because it is fascinating rather, it is now a tool that addresses the technical issues and limitations involved in analyzing and displaying large amounts of data. Moreover, HPC has exploded because of the convergence of inexpensive cluster computing, parallel applications that harness clusters, and high performance storage. HPC has therefore become pervasive part our economy and will continue to drive the competitive edge in businesses, as well as to improve the quality of life by solving varied problems spanning several disciplines that are relevant to today's world.

Typical HPC Architecture

HPC clusters built with off-the-shelf technologies offer scalable and powerful computing to meet high-end computational demands. HPC clusters typically consist of networked, high-performance servers, each running their own operating system. Clusters are usually built around dual-processor or multi-processor platforms that are based on off-the-shelf components. The low cost of these components have brought Linux clusters to the forefront of the HPC community. Clusters are extremely good for applications that need very high computational power or increased memory. However, many applications that rely on clusters still require a large amount of I/O throughput. These applications generate huge amounts of data while running, and need to efficiently and effectively write data to a storage medium, usually disk drives.

The general structure and design of cluster architectures that may enable high performance parallel computing are the ones where master node(s), compute node(s), storage & parallel file system are interconnected by a high performance network system are illustrated in Figure 4. The master node is the architecture's gateway to external resources. In order to make the master node highly available to users, High Availability (HA) clustering might be employed.

The compute nodes on the other hand are the cluster workhorses and are used to execute parallel jobs. Typically, access and management of compute nodes are provided via remote interfaces, such as network and/or serial port connections through the master node. Since compute nodes do not need to access machines outside the cluster, nor do machines outside the cluster need to access the compute nodes directly —compute nodes commonly use private IP addresses.

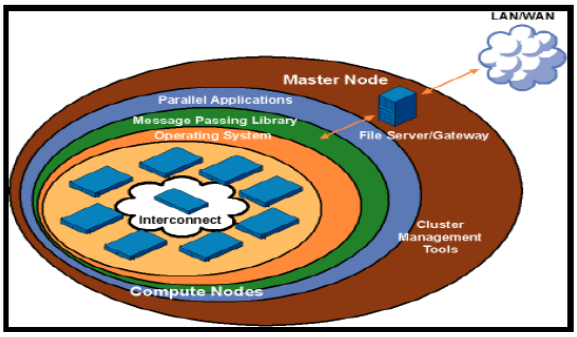

Figure 4: Generic Architecture of HPC system.

The integral components for complete High Performance computing system are:

- Operating system
- Compute nodes & Cluster Management tools
- Network and node file system
- Message passing libraries
- Application workload manager
- High performance interconnect & Storage
- Performance benchmarking & Development tools

Trends in High Performance Computing

The term High Performance Computing (HPC) was originally used to describe powerful, number-crunching supercomputers. As the range of applications for HPC has grown, however, the definition has evolved to include systems with any combination of accelerated computing capacity, superior data throughput, and the ability to aggregate substantial distributed computing power. The architectures of HPC systems have also evolved over time, and the same is illustrated with the help of schematic diagram in Figure 5.

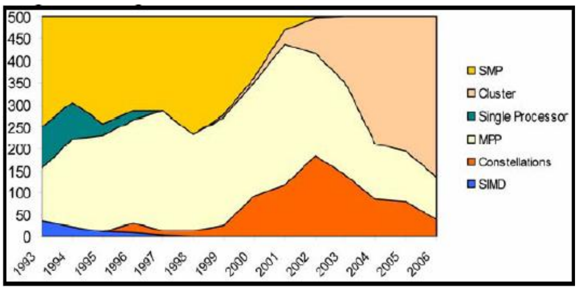

Figure 5: The history of HPC architectures shows a shift toward cluster computing.

Ten years ago symmetric multiprocessing (SMP) and massively parallel processing (MPP) systems were the most common architectures for high performance computing. More recently, however, the popularity of these architectures has decreased with the emergence of a more cost effective approach: **cluster computing.**

Cluster computing has achieved this prominence because it is extremely cost-effective. Rather than relying on the custom processor elements and proprietary data pathways of SMP and MPP architectures, cluster computing employs commodity standard processors and uses industry-standard interconnects. Applications that are a good fit for the cluster architecture tend to be those that can be “parallelized,” or broken into sections that can be independently handled by one or more program threads running in parallel on multiple processors. Such applications are widespread, appearing in industries such as finance, engineering, bioinformatics & oil and gas exploration. Application software for these industries is now routinely being delivered in “cluster aware” forms that can take advantage of HPC architectures. Clusters are becoming the preferred HPC architecture because of their cost effectiveness; however, these systems are starting to face challenges. The single-core processors used in these systems are becoming denser and faster, but they are running into memory bottlenecks and dissipating ever-increasing amounts of power. The presence of memory bottlenecks has two components: a limited number of I/O pins on the processor package that can be dedicated for memory access and an increasing latency due to the use of multi-layered memory caches. Higher power dissipation is a direct result of increasing clock speeds, and is forcing a need for extensive cooling of the processors.

The increasing power demand of processors is of particular concern in data centers and many other applications where calculations per watt are increasingly important. System cooling requirements are limiting cluster sizes, and therefore limit performance levels achievable. Their existing cooling systems are simply running out of capacity and increasing cooling capacity carries a high price.

The industry’s current solution to these growing problems is to move towards multi-core processors. Increasing the number of processor cores in a single package offers increased node performance at somewhat lower power dissipation than that of an equivalent number of single-core processors. But the multi-core approach does not address the memory bottlenecks inherent in the packaging. An even more significant HPC trend with impact on geoscience applications is the move toward distributed computing on semi-independent processing nodes.

HPC Benchmarking & Performance Evaluation

Benchmarking and Performance Evaluation allow us to understand how the HPC platforms perform in solving complex problems and the characteristics of computational applications. Performance evaluation efforts with real applications begun to emerge in 1988 with the Perfect Benchmarks. This set of scientific and engineering applications was created with the intent to replace kernel-type benchmarks and the commonly used peak-performance numbers. The Perfect benchmarks were a significant step in the direction of application-level benchmarking. There is general agreement that overall performance must be evaluated using wallclock time measurements. One open and often controversial issue is how to combine such measurements for multiple benchmarks into one number. Kernel benchmarks evaluate a wide variety of system components. Accordingly, the metrics vary widely. This is also true for metrics that characterize computational applications. Examples are working set sizes, hardware counter values (cache hit rates, instruction counts), and software metrics (code statistics, compiler results).

Challenges & Way Forward

The theoretical peak speed of high performance clusters is impressive. However, the architecture of clusters need to be optimized for numerical computations. The real HPC challenges today are to integrate all the compute units to transfer large data volumes to and from the processing units (efficient data access), and to process data in small chunks distributed over many processing units (efficient algorithms). The most limiting factor in the optimization or configuration of the clusters is internal memory bus; data throughput and computer interconnect fabric. The growing number of computing components within the hardware architecture means large efforts must be made for the parallelisation of application programs. It is a fact that parallelisation tools are far behind the possibilities offered by HPC hardware.

As a user one need to make a number of choices before assembling a cluster system. This may include a number of key perimeters to be considered in view of the expected performance viz. What hardware will the nodes run on? Which processors will you use? Which operating system? Which interconnect? Which programming environment? What application software will run on the cluster for optimal and efficient usage? Each decision will affect the others, and some will probably be dictated by the intended use of the cluster. Therefore computational tools are often not easy to use and require considerable judgment and expertise.

The payoff for developing/implementing the high performance computational tools and the computer comes when the production user employs the computational tools to solve real problems. A large, massively parallel computation may produce terabytes of data. Extracting information from such datasets is a massive challenge. Thus, while computational science and engineering has great potential, there are significant challenges to realize that promise.

On the other side substantial amount of energy consumption & cooling requirements need immediate attention. It is a well-known fact that the energy consumption of HPC data centers will double in the next four to five years, if the current trend continues. HPC manufacturers and data centers have to concentrate on energy efficiency. The development of HPC systems that reduce the energy consumption (50 to 70 per cent of the power is normally used to cool the system) is absolutely necessary.

A serious competitor catching up the multi-core CPU is represented by graphical processing units (GPUs), which are graphic cards used for scientific computing. They are fast and will get a lot faster. GPU's are cheap & use less power than CPUs when compared on a performance-per-watt basis. But the limitations of the GPUs are only good for tasks that perform some type of number crunching. The GPU's were designed specifically to process graphics, and that means processing streams of data. The fastest graphics chips are already in the Teraflops range whereas normal Multi-Core chips are slowly touching this border.

The real problem with GPUs is that they may not be programmed as it is for normal common CPUs. That's the reason that GPUs offer the support of the CUDA (compute unified device architecture) library that provides a set of user-level subroutines and allows the GPU to be programmed with standard C or Fortran without the need to use a graphics specific API. For the nearest future scenario of HPC systems the hardware architecture will be a combination of specialized CPU and GPU type cores.

Furthermore, since exploration data analysis makes use of a broad mix of applications. Some algorithms are CPU-intensive, others are memory-intensive, others still are I/O-intensive, and some are all three. This again places varying IOPS and throughput demands on a storage system. What's needed is a storage solution that has the flexibility to handle the application mix found in most energy exploration organizations.

Moreover, given the enormous amounts of data in geosciences that need to be manipulated, analyzed, integrated, moved, and visualized, the key to success is to have highly honed & improved computational workflows in place to handle the work. Additionally, exploration organizations worldwide need to continuously try accelerating their workflows. This can be achieved throught optimization of increasingly sophisticated analysis algorithms to take advantage of a hardware-assisted speedup by running them on GPUs and FPGAs. Use of these technologies can significantly change the IOPS and throughput demands on a storage system.

Also in order to perform seismic analysis on the scale required today, geoscientists must be more than experts in signal processing & geological formations. There is a need to give more impetus to computational geophysics so that industry professionals learn low-level software programming to quickly analyze large datasets on high-performance computing (HPC) resources. Going from terascale to petascale HPC systems means that the number of elements (cores, interconn ect, storage) within such a system will grow enormously.

Acknowledgement

Authors are thankful to the management of Oil India Limited for granting permission to publish this paper.

References

Lawerence M.Gochioco, Computer technology: From punch cards to clustered supercomputers,TLE, November 2002

Alexei kuzmin, cluster approach to high performance computing, computer modelling & new technologies, (2003) 7-15.