WHITE PAPER - UPDATE

***A Tectonic Shift in Upstream Petroleum*: Breakthrough Advantage with the IBM® System Blue Gene/P® Solution for Exploration and Production**

Sponsored by IBM

Srini Chari, Ph.D., MBA

mailto:chari@cabotpartners.com

Cabot Partners Group, Inc. 59 East 54th St. New York, NY 10022 203-205-0705 www.cabotpartners.com

Cabot Partners
Optimizing Innovation

Executive Summary

Finding oil 40 years ago was mainly hit or miss. Explorers drilled ten "dry holes" for every success. Today, these odds have quintupled largely due to the impressive advances in sensors and remote monitoring and visualization, high-performance computing technologies and algorithms that have enabled multidimensional seismic imaging and reservoir modeling. 3-D seismic imaging is now mainstream practice in the industry for exploration. Furthermore, seismic imaging coupled with very accurate reservoir modeling helps engineers better optimize the number and placement of wells in a reservoir or even on an entire basin composed of multiple reservoirs by "virtual drilling" i.e. without actually drilling. This computational capability is immensely valuable especially since it costs about $60M to drill a new deepwater well.

The IBM Blue Gene/P is the second system in a generation of innovative, ultrascalable architectures that will enable upstream petroleum engineers to make significant improvements in the understanding and the solution of some of the most complex problems in exploration and production. For the first time, it will be possible to combine high fidelity 4-D and even 5-D seismic imaging and very accurate multi-component, multi-scale, reservoir models to further increase the probability of finding and recovering oil. These innovative computational techniques are more imperative in the years ahead as the earth's scarce resources dwindle, and exploration must occur in environmentally hostile regions such as in deep water, wide basins with salt mounds, Siberia, and Alaska. However, this entails a continuing investment by upstream petroleum engineers, geoscientists, mathematicians, and computer scientists to develop new algorithms and applications for ultrascalable parallel computing environments. This investment will be protected as newer and more powerful Blue Gene systems become available. As before, the payoff will far outweigh this investment as energy costs continue to escalate.

IBM has expanded its five year Blue Gene collaboration with the Lawrence Livermore Laboratory to the upstream petroleum community. This collaborative investment has produced impressive early results for some of the most challenging seismic imaging and reservoir modeling problems in industry and research institutions. The scalability and performance obtained from these simulations on the IBM Blue Gene are unsurpassed yet affordable, and easily accessible. Direct hydrocarbon detection and optimized recovery is for the first time – plausible.

A Tectonic Shift in Upstream Petroleum

Unsurpassed Performance for Ultrascale Seismic Imaging and Reservoir Modeling with the IBM® System Blue Gene/P® Solution

Table of Contents

List of Figures

A Tectonic Shift in Upstream Petroleum

Unsurpassed Performance for Ultrascale Seismic Imaging and Reservoir Modeling with the IBM® System Blue Gene/P® Solution

Introduction

Seismic imaging and reservoir modeling have revolutionized upstream petroleum especially the process of exploration and production in the petroleum and energy industries. Seismic imaging and reservoir modeling are routinely used in conjunction with sensor information of fluid and rock properties at given locations to significantly enhance the success of finding oil and to optimize its recovery. These critical interdisciplinary approaches are integral in upstream petroleum and continue to require significant computational capability and benefit considerably from innovations in the information technology industry. The Blue Gene is an example of such a breakthrough innovation.

Seismic Imaging and Reservoir Modeling Combines Geosciences, Fluid Dynamics, Mathematics, and Advanced Computing

In the last four decades, developments in novel algorithms, the increased and affordable access to high performance computers, and the availability of a broad range of commercial, in-house proprietary, and consortia-led academic applications i.e., the Mission-Oriented Seismic Research Program (M-OSRP) led by Dr. Arthur Weglein at the University of Houston and the Integrated Parallel Accurate Reservoir Simulator (IPARS) effort led by Dr. Mary Wheeler from the University of Texas, have together substantially contributed to the widespread acceptance of seismic imaging and reservoir modeling in the petroleum industry.

Effective and accurate sesmic imaging and reservoir modeling require a fusion of interdisciplinary advances in geology and fluid models, mathematical algorithms, information technology architectures, and disciplined software engineering. More accurate imaging and modeling techniques increase the probablility of hydrocarbon discovery and exploitation, extend the life of current reservoirs, improve the risk-reward ratio of locating new hydrocarbon sources, and reduce the overall cost of exploration and production.

The diverse and complex nature of today's large scale upstream petroleum problems (e.g., 3-D/4-D prestack time and depth migration, 4-D and multi-azimuth seismic imaging, dynamic data driven multiphysics reservoir simulation, etc.) requires the balanced use of computing capability for grand challenge simulations coupled with capacity computing capability for production simulations.

Furthermore, a combination of high performance computing systems, massive storage systems, visualization and advanced instrumentation, applications and middleware, all connected by high-speed networks is needed for today's seismic imaging and reservoir modeling infrastructure.

Breakthrough, High Resolution Reservoir Modeling and Seismic Imaging Requires Parallel Ultrascale Computing

Reservoir modeling involves tackling a wide range of complexity in modeling flow physics in or around complex shapes and geometries. The equation describing the flow physics are a coupled system of non-linear, partial differential equations. These problems are often very large, tightly-coupled, and multi-scale. Likewise, as exploration increasingly occurs in environmentally hostile regions with complex geology on wider basins and much deeper in the ground, seismic imaging problems have become very large and information needs to be aggregated from billions of points from the seismic surveys. Also – as the price of oil continues to escalate - revisiting existing reservoir basins and analyzing older seismic data with newer computational algorithms can reveal potential opportunities for production.

These computing needs - to transcend Moore's law - through parallel computing have given rise to clusters, grids, and scalable parallel systems in recent years. Parallel solution approaches based on parallel mesh-decomposition techniques are required to exploit these parallel architectures for large scale reservoir modeling and seismic imaging problems.

While standard clusters and grids can address a sub-class of these problems, an entire range of complex, multi-physics, multi-scale, reservoir and seismic problems require ultrascalable architectures. These systems are large, tightly-coupled computers with high bandwidth and low latency interconnects with an optimized message-passing library, such as MPI (Message Passing Interface). The IBM Blue Gene system is an example of an ultrascale computer that has been very effective in addressing these classes of challenging upstream petroleum problems.

Dynamic Data Driven Frameworks Coupling Sensor Data and Reservoir Modeling are Evolving in Reservoir Management

Dynamic Data Driven Applications and Simulations (DDDAS) is becoming a new paradigm of computing, one involving symbiotic feedback in which reservoir modeling - simulation and analysis - and sensor field data interact in real time to dramatically improve the fidelity of the analysis, its accuracy, and reliablility. Integrating reservoir modeling with 4-D sensor data enables the ongoing management and optimization of large reservoir basins and the exploration in traditionally difficult terrains. This novel paradigm requires investments in new approaches, massively parallel computing architectures, middleware, and opens up new vistas in high performance analysis in exploration and production thereby yielding a substantial ROI for the petroleum industry especially as the prices of oil and gas continue to go up.

Blue Gene/P: Unsurpassed, Ultrascale, Affordable Computing

The Blue Gene/P is the second in a new generation of novel parallel supercomputers and is the most powerful supercomputer available. Dramatic reductions in power consumption, cost, and space requirements are achieved through the use of innovative technologies in low-power processors; embedded DRAM; system-on-a-chip; advanced power, packaging and cooling; special interconnects delivering very low latency and high bandwidths; and scalable systems management. This performance, scalability, flexibility, and innovative design enables the solution of a wide range of complex reservoir modeling and seismic imaging problems today and in the future.

IBM Collaborates with Upstream Petroleum Application Developers to Foster Breakthrough Innovation and Discovery

Seismic imaging and reservoir engineering applications that are mapped, migrated, and optimized for the Blue Gene architecture will benefit greatly from ultrascalability and extreme performance. This investment in application enablement will permit higher degrees of performance and scaling in newer generation of systems and technologies based on the Blue Gene architecture. Worldwide, IBM is working with upstream petroleum application developers to migrate and optimize their applications on the Blue Gene to solve large challenging optimization problems that are multi-scale and multi-physics, in order to enable breakthrough innovation in industry and academic consortia. Experience to date from the examples described later shows that large-scale reservoir modeling problems (over 10 million to 30 million unknowns), multi-phase flows with intricate physics, and multidimensional seismic imaging problems on complex geology will scale and excel on the Blue Gene.

Computational Challenges and Trends in Upstream Petroleum

Seismic imaging is an inverse reconstruction problem that uses reflected seismic waves to produce images of the earths subsurface by measuring attributes such as time, amplitude, wave-length, frequency, etc. of the reflected wave and record a seismogram. Modern seismic surveys that are vaster and deeper require increasing amounts of data processing, often performed on supercomputers, clusters, or scalable parallel systems. In all but the simplest geological environments, seismic imaging has three fundamental problems: the image starts off being fuzzy, has the wrong shape, and is in the wrong place. Various algorithms – migration types - have been used to solve these wave equations that describe the passage of sound waves through rock.

In the past, poststack migration was very common. Prestack migration takes about 100 times longer than poststack migration but can exploit massively parallel computers more effectively as these algorithms require less inter-processor communication. Prestack migration retains amplitude variation with offset and phase changes which are very useful for subsequent analysis. Prestack time migration is preferred when two or more events occur at the same time but with different stacking velocities. Prestack depth migration is useful when the velocities in the target are complex but requires even more computing power. Production specialists increasingly use several algorithms concurrently to analyze the data driving up the computing requirements even further. But these concurrent analyses are very effective on the Blue Gene as each analysis can be done on separate Blue Gene partitions.

The trend in seismic migration from time to depth, poststack to prestack, and 3-D to 4-D (even 5-D) continues. Some of the challenges that remain are being addressed by a combination of advanced data acquisition systems, high-performance scalable parallel computing systems, and related technologies. The following imaging challenges could be addressed with ultrascalable computing environments with prestack migration:

- Sequence Stratigraphy
- Anisotropy
- 4-D seismic acquisition with offset gathers in the 4th dimension
- 5-D seismic or time-lapse 4-D
- Analysis of multi-azimuth seismic data

Need to Transcend Moore's Law through Parallel Computing

The need to transcend Moore's law through parallel computing has given rise to clusters, grids, and scalable parallel systems in recent years. To process massive amounts of data, which provides valuable engineering insight; applications require massive horsepower, fueling the demand for scalable parallel computing. To meet the demands of the dramatic growth of data and higher accuracy, especially for high fidelity seismic imaging and multi-scale reservoir simulation, ultrascale parallel platforms such as the IBM Blue Gene is desirable.

Novel Partitioning Schemes are Needed for Applications to Scale

In general, many seismic imaging and reservoir simulation applications can be adapted and optimized to scalable parallel architectures for large-scale simulations that require very high degrees of spatial and temporal resolution and accuracy. This requires the use of domain decomposition or mesh partitioning techniques coupled with the solution approaches. Powerful partitioning tools used in the pre-processing phase have enabled many implicit reservoir simulation applications with parallel solvers to scale effectively. Multidimensional seismic data migration problems and 3-D elastic seismic wave propagation problems also benefit significantly from smart partitioning methods.

Large Scale Upstream Petroleum Applications Require Novel Parallel Algorithms and Systems

Very high performance computing is almost always accomplished through parallelism. However, obtaining parallel computing capabilities can be difficult and complex because most practical applications don't multithread beyond a few processes. In order to scale further, parallelism must be at a very high coarse grain level. Novel algorithmic approaches in upstream petroleum engineering based on domain or mesh decomposition strategies and parallel FFTs allow users to obtain the maximum advantage and scalability from parallel machines with large numbers of inter-connected processors. Domain decomposition techniques when coupled with parallel variants of multi-grid, explicit, overrelaxation methods, and parallel FFTs further accelerate complex simulations.

Careful System Specific Tuning is Needed to Further Boost Performance and Scalability on Parallel Platforms

Parallel application development and system performance for large-scale upstream petroleum computing applications also depend on the single processor performance, the communication subsystem performance, I/O performance, and development tools for programming, debugging, and resource management. A significant improvement in performance and scalability can be obtained for applications that are tuned and optimized for the specific parallel architecture. This often requires a combination of deep algorithmic and parallel computing skills and understanding of geosciences and fluid dynamics.

Clusters and Grids are Often Inadequate for Large Scale Upstream Petroleum Computing

An entire class of high fidelity, multi-scale, multiphysics reservoir management problems; e.g., 4-D seismic history matching, require scalable parallel systems with very low latency and high bandwidths between the several thousands of processors for deriving insight.

In order to solve these growing compute requirements for upstream petroleum computing, individual clusters are starting to become insufficient. Standard clusters with thousands of processors are expensive to build and operate. The cost associated with providing support and maintenance grows exponentially. Also, management of such diverse collections (grids) of resources is difficult, and effective software solutions that can scale are only now beginning to appear in the market. Furthermore, the electrical power consumption and the physical facilities required to operate such large clusters are prohibitively expensive. Fortunately, the IBM Blue Gene is architected to address these extreme needs.

The Blue Gene/P Solution

The IBM® System Blue Gene® Solution is the result of an IBM super-computing project which began in collaboration with Lawrence Livermore National Laboratory over five years ago. It was dedicated to building a new and innovative family of supercomputers optimized for bandwidth, scalability and the ability to handle large amounts of data while consuming a fraction of the power and floor space required by today's high performance systems. The IBM® Blue Gene®/P (BG/P) Solution is the second generation machine in IBM's Blue Gene program. It adheres to the key design strategies providing petaflop scale performance that is efficient in terms of power, cooling and floor space, thereby reducing the total cost of ownership. Compared to Blue Gene/L, its predecessor, Blue Gene/P extends performance through a doubling of processor cores and a frequency increase, and adds 4-way SMP functionality, hardware DMA, 10 Gb Ethernet, and employs aggressive power management. Blue Gene/P also provides a standard programming environment and supports a wide range of IBM and open source software libraries and middleware.

An Ultrascalable, Environmentally Efficient Architecture

The Blue Gene system is built with a very large number of compute nodes, each of which has a relatively modest clock rate contributing to both low power consumption and low cost. It utilizes IBM PowerPC® embedded processors, embedded DRAM and system-on-a-chip techniques that allow for integration of all system functions including compute processor, communications processor, three cache levels, and multiple high speed interconnection networks with sophisticated routing onto a single chip. Because of a relatively modest processor cycle time, the memory is close, in terms of cycles, to the processor. This is also advantageous for power consumption and enables construction of dense packages in which 1024 dual-processor compute nodes can be placed within a single rack.

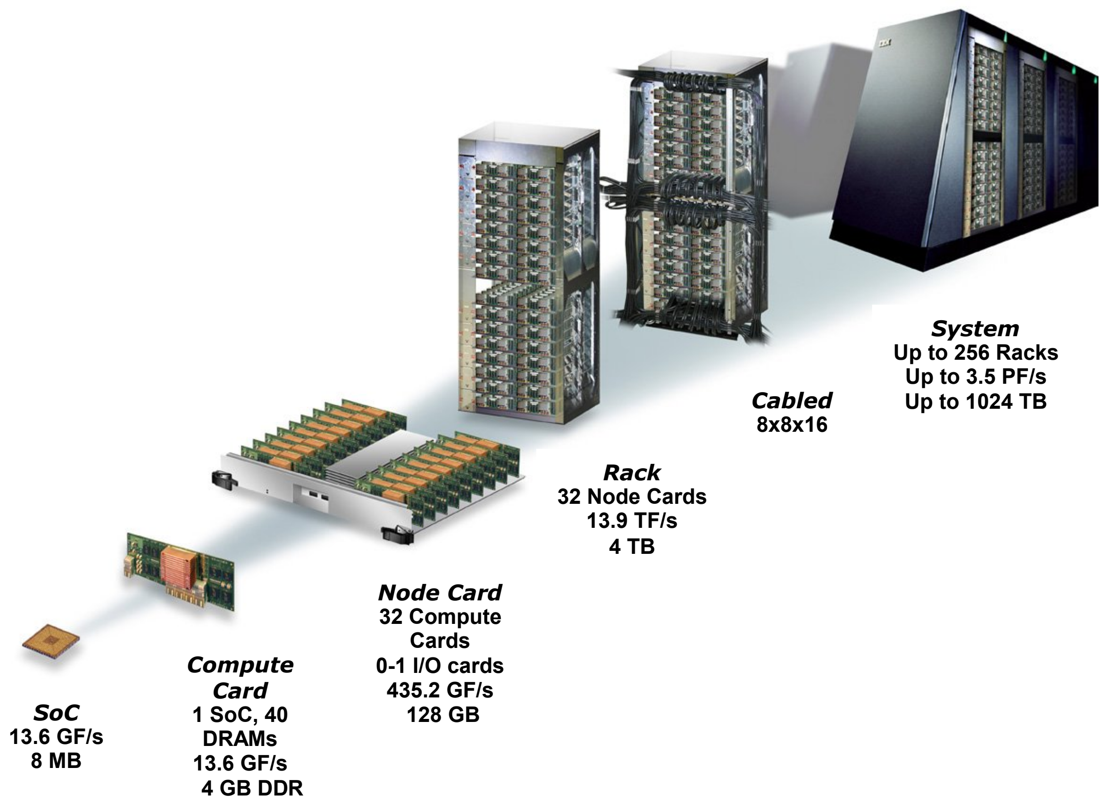


Figure 1: The Blue Gene/P Architecture

The nodes are interconnected through five networks: a 3-dimensional torus network for point-to-point messaging between compute nodes, a global collective network for collective operations over the entire application, a global barrier and interrupt network, and two gigabit Ethernet networks for machine control and for connection to other systems. The torus network is particularly effective for applications with locality of communication. And the global collective network is useful for speeding up MPI collective communications constructs.

Large-scale reservoir modeling applications and seismic migration algorithms based on finite-differences that use domain decomposition approaches benefit significantly from this local-global capability of the torus network. The computations done in each sub-domain are kept local from one fine grain iteration to the next and can be done in parallel across all of the sub-domains. After several local fine-grain iterations, the computations are synchronized across the entire domain using the collective communications constructs before proceeding further with more fine-grain, local iterations. This local-global approach accelerates the solution convergence for a large class of non-linear reservoir simulation applications, permits the aggregation of multidimensional seismic analysis along each dimension, and scales very well on the IBM Blue Gene/P.

A Familiar, Optimized Software Environment

Three fundamental principles were followed when the system software was designed for the Blue Gene system: simplicity, performance and familiarity. Driving toward simplicity in the software design has allowed development of software that takes advantage of hardware features to deliver high performance without compromising stability and security. And by creating a programming and administration environment based on familiar programming languages, libraries, job management tools and parallel file systems, Reservoir modeling and seismic application developers benefit from the innovative design elements of the Blue Gene system without facing a steep learning curve. The flexible and user friendly IBM software stack contains the following items

- XL (FORTRAN, C, and C++) compilers
- LoadLeveler scheduler
- GPFS parallel file system
- ESSL/MASSV libraries
- MPI library

Other Software Support for IBM Blue Gene/P includes

- Parallel File Systems
- Job Schedulers
- Parallel Debugger
- Libraries

- Performance Tools

Integrated with High Performance Visualization, File Systems, and Middleware

As mentioned earlier, in addition to high performance computing systems, massive file/storage systems, visualization and advanced instrumentation, middleware, all connected by high-speed networks are also needed for today's upstream petroleum infrastructure.

With IBM's Deep Computing Visualization, high-end graphical images are created in two visualization modes:

- Scalable Visual Networking (SVN)
- Remote Visual Networking (RVN)

SVN supports multiple high-resolution monitors or projectors for immersive, stereo visualization. RVN distributes graphical images to remote (collaborative) client stations. These features help derive more accurate engineering and geological insights from the higher resolution of the complex data obtained from reservoir simulations and seismic data on the Blue Gene/P.

The following High Performance Computing cluster software is now available on Blue Gene/P: the Engineering and Scientific Subroutine Library (ESSL) for Linux on POWER, General Parallel File System (GPFS) for Linux on POWER, and LoadLeveler for Linux on POWER. ESSL provides over 150 math subroutines that have been specifically tuned for performance on the Blue Gene/P. GPFS is the top performing cluster-wide file system for the Blue Gene/P, providing superior scalability and high reliability. LoadLeveler is a job scheduler designed to maximize resource utilization and job throughput to get the most out of the available resources.

This combination of middleware enables the optimization and scaling of the Blue Gene/P resources (processors and storage) for performing several, concurrent seismic data migrations and reservoir simulations typical in dynamic data driven application studies in reservoir management and optimization. The GPFS uses dedicated I/O nodes on each rack of the Blue Gene/P to dramatically increase the I/O performance for large multidimensional seismic imaging problems. The optimized FFT subroutines in ESSL can be used to speed-up many seismic migration applications.

Both High Performance and High Throughput Computing Modes are Possible with the Blue Gene/P

Floating point operations per second achieved on a single application have long been the yardstick to rank HPC systems. These metrics are now being supplemented by high throughput computing metrics which quantify the total amount of computational work done over a period of time on a mix of applications typically found in many industrial environments. With the Blue Gene/P both High Performance and High Throughput Computing modes are possible on a single platform. The features of HPC and HTC are tabulated in Figure 2.

High Performance Computing (HPC) Mode	High Throughput Computing (HTC) Mode
▪ Parallel, tightly coupled applications • Single Instruction, Multiple Data (SIMD) architecture ▪ Programming model: typically MPI ▪ Applications need tremendous amount of computational power over short time period	▪ Large number of independent tasks • Multiple Instruction, Multiple Data (MIMD) architecture ▪ Programming model: non-MPI ▪ Applications need large amount of computational power over long time period ▪ Traditionally run on large clusters

Figure 2: Blue Gene/P HPC and HTC Modes

Blue Gene/P System Details at a Glance

The Blue Gene/P system is accompanied by a product roadmap that will deliver petaflop performance. Available from 1-112 racks, the salient system details and benefits for Blue Gene/P are summarized in the following table.

ATTRIBUTE	DETAILS	BENEFITS
Processor	IBM® PowerPC® 450 850 MHz; four per node	Low power allows dense packaging; better processor memory balance
Memory per node	4 GB SDRAM-DDR (Model 0206-850)	Wider reach applications
Networks	1. 3D Torus - 5.1 GB/s; 3.5 µsec latency 2. Collective Network – 1.7 GB/s; 2.5 µsec latency 3. Global Barrier/ Interrupt 4. Optical 10 GbE (machine control and outside connectivity) 5. 1 GB Control Network (system boot, debug, monitoring)	Special networks speed up internode communications; designed for MPI programming constructs; improve systems management
Compute Nodes I/O Nodes (10 GbE)	Quad SMP processor; 1024 per rack Quad SMP processor; Configurable from 8 to 64 per rack. 8 is the default configuration	Double FPU improves performance Increases relative I/O performance

Operating Systems Performance	Compute node – Lightweight proprietary kernel I/O Node- Linux® operating system Front End and Service Nodes- SUSE LINUX SLES 10 Peak Performance per rack – 13.9TFlops	Kernel tailored to processor design; industry standard distribution preserves familiarity to end users and administrators Highest available performance benefits capability customers
Power	40 kW power consumption per rack (maximum) 200-240 VAC 3-phase; 175 amp service per rack	Low power draw enables dense packaging
Cooling	Air conditioning ~13 tons/rack (minimum)	Low cooling requirements enable extreme scale up
Dimensions (includes air duct)	Height 1956mm, Width 1220mm Depth 966mm; Weight 782kg Service clearances 914mm all sides Raised floor height - 40.64 cm minimum	Design allows "brickwall" layout for better floor space utilization

Figure 3: Blue Gene/P System Details at a Glance

Blue Gene/P comparison with Cray XT4 and IBM cluster 1350

With 1024 processors/rack Blue Gene/P distinctly proves to be superior when compared to Cray XT4 which has 96 processors/rack. The Blue Gene/P occupies less space and requires significantly less energy and cooling compared with a Cray XT4. A comparison between Cray XT4 and Blue Gene/P is shown in Figure 4.

	IBM BG/P	**Cray XT4**
Processor	IBM PowerPC®	AMD Opteron®
Frequency	850MHz	2.6GHz
Cores / Processor	4	1, 2, 4
Scalable	1M+ Cores	120K+ Cores
Processors / Rack	1,024	96
Cores / Rack	4,096	192(with Dual Core)
Memory / Rack	4,096 GB	768GB (maximum)
Peak Performance / Rack	13.9 TF	1/2, 1 or 4 TF
Integrated Networks	Torus, Collective, Barrier	Torus (SeaStar)
Rack Dimensions (HxWxD)	77 x 27.6 x 38	80.5 x 22.5 x 56.75
Rack Weight	1700 lbs	1529 lbs
Cooling Required / Compute Node	12.7 kg	62.5 kg

~100 TF Comparison		
Racks	8	96
Peak Performance (TF)	111.2	94.6 (Dual Core)
Floor Space (sq ft)	313	1344
Power (kVA)	346	2089
Cooling (tons)	104	576

Figure 4: Comparison of Blue Gene/P and Cray XT4

The Total Cost of Ownership (TCO) comparisons between the Blue Gene/P and the IBM Cluster 1350 are shown in Figure 5. The power consumption for Blue Gene is significantly less compared to a cluster. Other cost such as high speed interconnects which are critical for clusters are not an issue for the fully integrated Blue Gene/P. The floor space occupied by the Blue Gene/P is also comparatively less. As can be seen, the Total Cost of Ownership for a Blue Gene/P is about three-fifth that of a cluster.

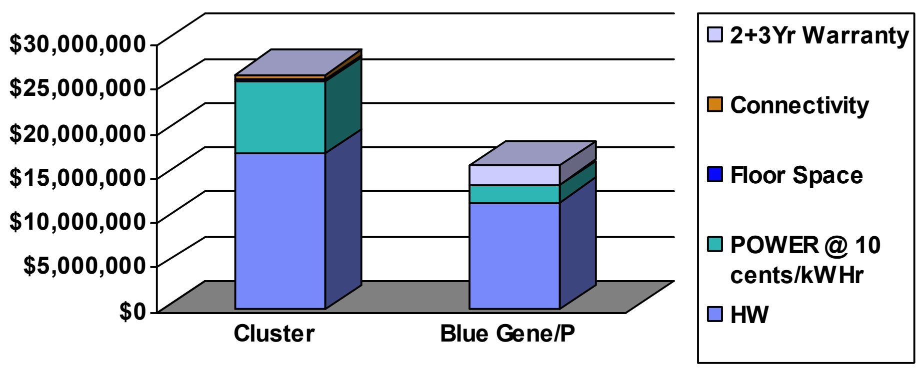


Figure 5: Comparison of TCO of Cluster and Blue Gene/P

The IBM Blue Gene/P Excels at Parallel Seismic Imaging and Reservoir Simulation

Recent advances in domain decomposition techniques coupled with parallel solvers make the Blue Gene solution very suitable for a large class of large-scale upstream petroleum problems. These crucial geosciences and engineering problems in the industry will require the computing performance and scalability that is practical only with the Blue Gene architecture. Reservoir modeling community could look for the following specific features which IBM Blue Gene/P offers to them.

- Reservoir models can be run in minutes instead of hours.
- Proper multi-physics and multi-scale modeling of the reservoir can be achieved to actually solve the behavior of all fluids in the ground.
- Dynamic reservoir modeling will find the 'actual' minima or maxima for the whole reservoir, and not just local min/max solution.
- Oil reserves can be more accurately estimated and then extracted optimally. This would increase reserves while minimizing errors.

Prestack and Higher Dimension Seismic Migration Scales Well on the Blue Gene

As mentioned earlier, trend in seismic migration is from time to depth, poststack to prestack, and 3-D to 4-D. 2-D prestack and poststack migration problems and 3-D poststack migration find solutions in IBM's Cluster servings. While Blue Gene/P performs well for 2-D and 3-D migration problems, the superior performance of Blue Gene/P is ideal for challenging problems like 3-D prestack migration and 4-D problems. In addition, it excels for newer applications in sequence stratigraphy, 3-D elastic wave propagation, and multi-azimuth seismic. The timely solution of these newer challenging problems is only possible with architectures such as the Blue Gene that can scale gracefully to thousands of processors. Moreover, because of its low power consumption and dense packaging, the Blue Gene can be easily housed and operated on deep sea exploration vessels for near real-time seismic processing.

Blue Gene Performs	Blue Gene Excels
▪ 2-D Poststack Migration ▪ 2-D Prestack Migration ▪ 3-D Poststack Migration	▪ 3-D Prestack Migration ▪ 4-D Seismic Imaging ▪ Multi-Azimuth Seismic ▪ Sequence Stratigraphy ▪ Anisitropy ▪ Forward Modeling ▪ Reverse Time Modeling

Figure 6: Affinity of the Blue Gene for Seismic Imaging

Basin Wide Reservoir Optimization Problems with Multiphysics are Suitable for the Blue Gene

The range of reservoir simulation problems that can benefit from the Blue Gene architecture is sketched in Figure 7. Large scale reservoir simulation problems with over 10 million unknowns scale very well especially when domain decomposition techniques that are effective in balancing the computing load are coupled with parallel solvers. Multi-component problems with widely varying thermal and velocity fields often require a very fine mesh that results in very large numbers of unknowns. Also, interdisciplinary problems with multiple physics and/or fronts increase the problem scale making the Blue Gene even more attractive.

Complex, coupled, computationally challenging problems in reservoir optimization over an entire basin and parametric studies iterating over thousands of sensor input parameters, and chemically reacting flows with multiple species will surely benefit significantly from the Blue Gene architecture. These problems have additional unknown variables per grid point beyond the normal fluid flow variables (saturation and concentration) thus increasing the computational problem size. Also, hundreds of parametric simulations can be done concurrently on multiple Blue Gene partitions. All this will significantly reduce simulation time for large-scale reservoir optimization.

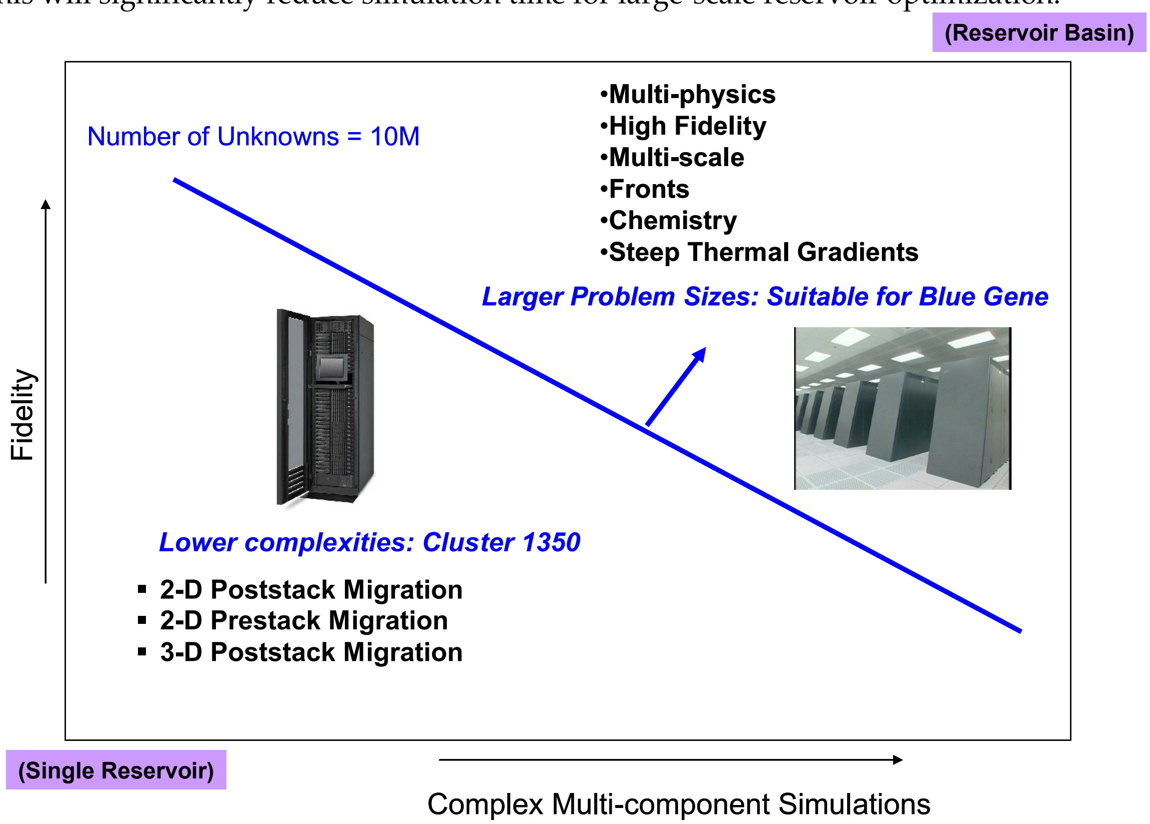


Figure 7: Affinity for Parallel Reservoir Simulation on the IBM Blue Gene

Tuning Further Enhances Scalability and Performance

To take advantage of the Blue Gene architecture, substantial performance tuning can be achieved by careful load balancing, maximizing single processor performance, maintaining data locality, minimizing cache misses, and maximizing the computation-to-communication ratio. Application performance can be further enhanced by the use of mathematical libraries that are optimized for the Blue Gene.

This requires a very structured approach. It consists of first porting the application, then validating it on Blue Gene, and finally optimizing the application on Blue Gene. Careful performance tuning can enhance the performance of the application significantly but requires a deep understanding of the application and the Blue Gene architecture. The investment made in developing these optimized versions can enable the solution of larger-scale problems with a substantial payoff in terms of obtaining insight and value. Furthermore, this investment is protected as newer generations of Blue Gene systems become available. IBM provides access to the Blue Gene system through the Enablement center. Seismic and reservoir applications developers and prospective users can test out and get in-depth experience on the Blue Gene in a cost-effective manner.

Examples

IBM is working with upstream petroleum application developers to migrate and optimize their applications on the Blue Gene to solve challenging problems that are multidimensional, multi-scale, and multi-physics. These representative examples range from 3-D depth migration, to dynamic data-driven chemically reacting, fine-scale, integrated reservoir optimization studies. Performance and scalability results for a few representative examples are presented here.

3-D Prestack Migration

As mentioned earlier, prestack time and depth migration is a very compute-intensive application that has almost perfect parallelism. Figure 8 depicts a typical prestack time and depth migration. The circled migration steps are major computational bottlenecks that can be accelerated significantly using the Blue Gene.

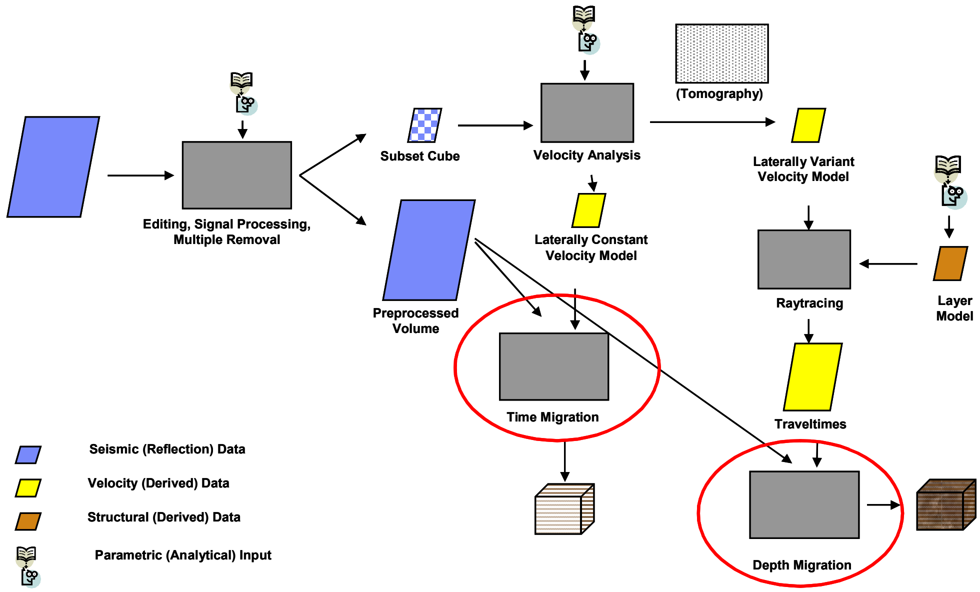


Figure 8: Prestack Migration Process and Blue Gene/P Affinity

A Kirchoff-based 3-D prestack migration problem with 120 offsets was run on multiple racks of the Blue Gene. The results show almost perfect speedup and performance for 2048 processors on the Blue Gene. The time to migrate this data was reduced from 60 days on one Blue Gene rack to 4 days on 16 racks. This dramatic reduction and possibly more reduction with more Blue Gene racks opens up the possibility of near real-time 3-D prestack migration.

Blue Gene/P **Nodes / Racks** (*Dual-processor nodes*)	**Time to Solution** (3D Survey, 120 Offsets)
1024 / 1 rack	25 days
2048 / 2 racks	12.5 days
4096 / 4 racks	6.25 days
8192 / 8 racks	3.3 days
16384 / 16 racks	1.6 days

Figure 9: Estimated Performance of a Kirchoff-based Prestack Migration on Blue Gene/P

Mission - Oriented Seismic Research Program at the University of Houston

The Mission-Oriented Seismic Research Program (M-OSRP) is a research program and a petroleum industry consortium at the University of Houston to address seismic exploration and production problems whose solutions would optimize the location and production of hydrocarbons. M-OSRP is pioneering and developing approaches to challenges in petroleum seismology to locate and define hydrocarbon targets beneath complex media, e.g., salt and basalt. IBM is a collaborator in this effort.

Preliminary results for optimizing internal multiple attenuation algorithms on the Blue Gene demonstrate near-linear scalability.

Run	Number of Blue Gene Compute Nodes	Run Time in Seconds (Estimated[1])
1	256	3766.3
2	512	1888.4
3	1024	957.9

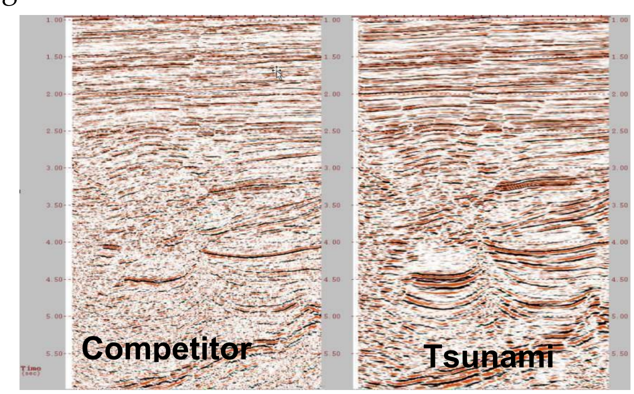
Figure 10: MO-SRP: Optimizing Internal Multiple Attenuation Algorithms

Tsunami

Tsunami is a Houston based company that provides advanced seismic imaging software for the petroleum industry. Current applications include Kirchoff based prestack time and depth migration. IBM is sponsoring an effort to port and optimize Tsunami applications on the Blue Gene in collaboration with CERT – the largest private education provider in the middle-east. A 2048 CPU has been installed at CERT's park in the UAE. One goal is to provide a Blue Gene version of Tsunami for users who need to achieve performance possible from thousands of nodes.

Tsunami Development (ISV) had the following salient features:

- Kirchoff-based PSTM and PSDM seismic imaging apps
- Port complete, production tests to 4096 processors
- High Resolution Seismic Imaging
- Process 300 SqKm/day/rack PSTM, 41Million traces.
- Tsunami states that one Blue Gene rack equals approximately a 1500 processor Opteron cluster

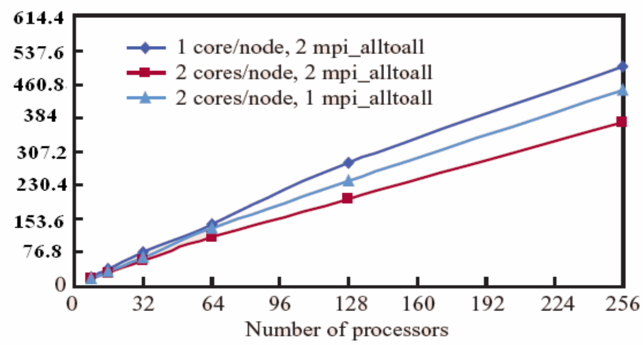


Multi-Azimuth Seismic Simulation

Seismic imaging is an application that benefits massively from parallel computer systems. Three-dimensional (3D) prestack time migration (PSTM) and pre-stack depth migration (PSDM) are key components of seismic imaging and require very large computing resources. The PSDM technique is the most preferred method used for 3D seismic imaging. The common azimuth migration (CAM) technique makes use of a PSDM algorithm that is specifically suited for the management of marine seismic data due to its speed and the quality of its results. The computing and communication capabilities of Blue Gene/P make it possible to modify the CAM algorithm to obtain a

[1] Estimated by extrapolating measured Blue Gene/L data

massively parallel version of this algorithm. Volumetric FFT as part of the CAM algorithm is one of the most computationally intensive components of the common azimuth method. Several studies have demonstrated that the most scalable and natural implementation of the 3D FFT on a Blue Gene system is based on a volume domain decomposition. This method achieves scalability to a larger number of nodes (up to 16,368 nodes) because it allows the distribution of work for an N x N x N FFT over the natural torus topology of the Blue Gene system.



Estimated three-dimensional FFT speedup with a size of 2,304 * 1,680 * 64 on the Blue Gene/P system, depending on the number of cores per compute node, with homogeneous (red plot) and non homogeneous (light blue plot) output partitioning.

Figure 11: Three-dimensional FFT speed up on Blue Gene system[2]

The speedup on the BG/L system occurs with up to 256 processors, with a steady efficiency of greater than 0.8, with one core per node for the homogeneous version and two cores per node for the non-homogeneous version. In order to integrate the fastest 3D FFT into the overall algorithm, a phase shift correction routine was modified to manage data decomposition in an x or y direction. The communication part, in the sequence of three steps (FFT, then phase shift correction, followed by inverse FFT), can be divided by two, with a 30% or greater improvement in the overall 3D-FFT time. Figure 11 gives the three dimensional FFT speed up on Blue Gene system incorporated using the PSDM and PSTM techniques.

<u>Acoustic 3D</u>

This is a key application used by a multinational energy company engaging in upstream operations such as oil and gas exploration, development and production, LNG, etc. and

[2] "A massively parallel implementation of the common azimuth pre-stack depth migration", H.Calandra, et.al , IBM J. RES. & DEV. VOL. 52 NO. 1/2 JANUARY/MARCH 2008

also downstream operations such as refining, marketing and the trading and shipping of crude oil and petroleum products. Analyzing the acoustic velocities of sub-surface formations is critical to the petroleum industry. Post drill analysis is important and valuable no doubt but there is a substantial benefit to be able to perform these calculations while drilling operations are ongoing. Real time seismic analysis, pore pressure estimation, and formation evaluation enable better planning and execution of expensive drilling and evaluation operations. IBM performed a number of tests to determine the time taken to accomplish certain acoustic 3D calculations. The number of processors necessary to carry out the tests Bench1 and Bench2 is dictated by the size of the total memory. The case "bench1" requires a minimum of 256 processors and "bench2" needs 1024 processors. These tests were performed using the MPI communication protocol and the comparison between the theoretical values and the measured values are shown in Figure 12.

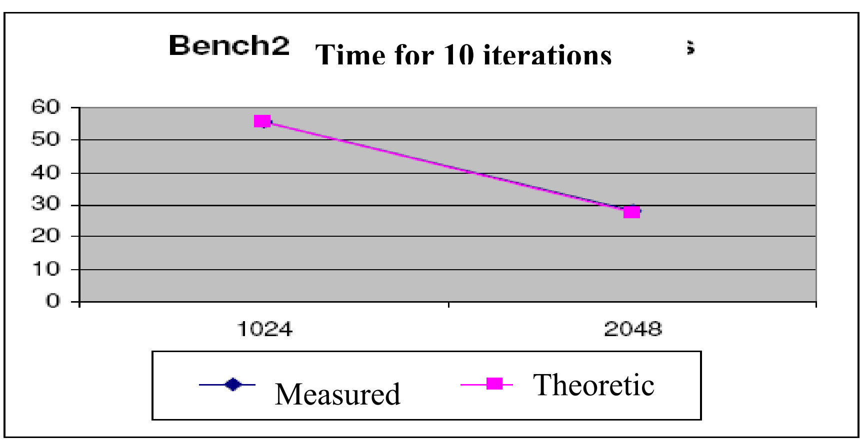


Figure 12: Time Taken for 10 Iterations on the Blue Gene/P

IPARS from the University of Texas

The University of Texas in Austin has developed a simulator framework IPARS (Integrated Parallel Accurate Reservoir Simulator) to serve as a test-bed for multiphase flow models, numerical discretizations, solvers, and upscaling. IBM is one of the sponsors of this effort. IPARS is a comprehensive dynamic data-driven framework based on a multi-block approach for performing scalable simulations of multiphysics, multi-scale, and multi-algorithm reservoir optimization problems.

IPARS was easily ported to the Blue Gene in a few days. A small reservoir simulation problem - 1 million unknowns - scaled well on the Blue Gene.

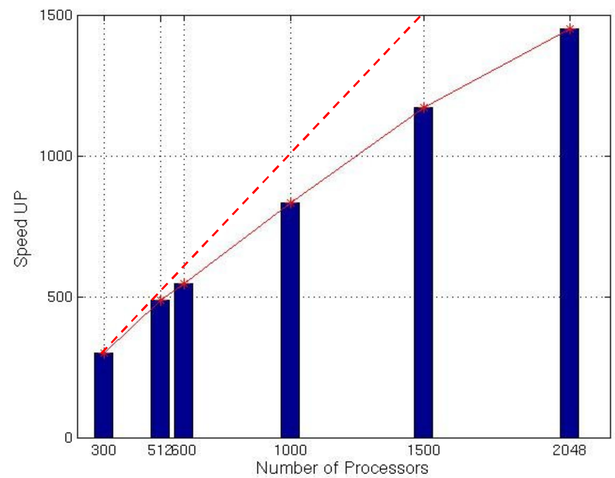


Figure 13: IPARS on the Blue Gene/P - 1 million unknowns completed in 3 minutes

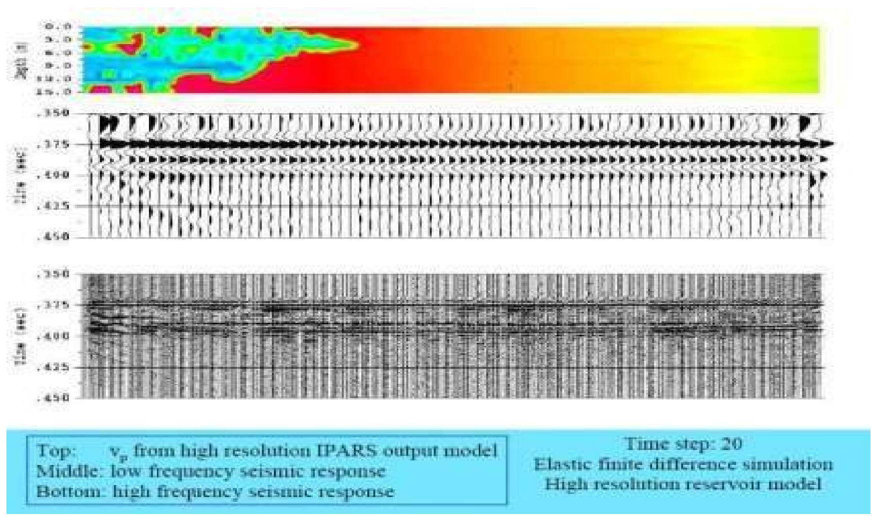
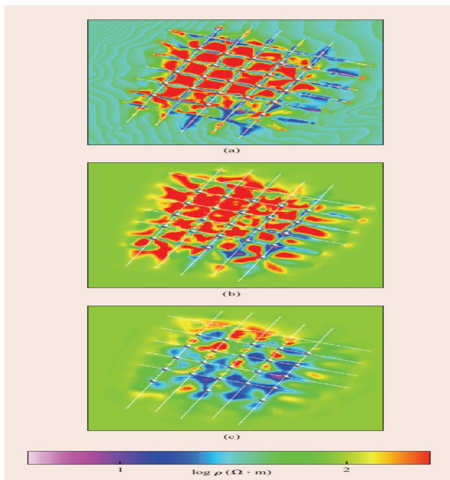


Figure 14: Reservoir Seismic Model Using Blue Gene/P

Controlled Source Electro-Magnetic (CSEM) 3D geophysical imaging

Large scale 3D CSEM geophysical imaging is highlighted here for petroleum exploration. The old seismic imaging method has a disadvantage of not being able to discriminate different types of fluids contained in the rock pore space like the brines, water, oil and gas. The CSEM technique uses low frequency electromagnetic signals to map variations in the sub-surface electrical conductivity or resistivity of offshore oil and gas prospects. Electromagnetic field measurements have shown to be highly sensitive to changes in the pore fluid types and locations of hydrocarbons, given a sufficient resistivity contrast between hydrocarbons and fluids such as brine or water. Figure 15 shows an average resistivity computed over three depth ranges using the CSEM method.



Average resistivity computed over three depth ranges for solution 72: (a) water bottom to 500 m below mud line (BML); (b) interval 500 to 1,500 m BML; (c) interval 1,500 to 2,500 m BML. Resistivity is rendered on a base 10 log scale.

Figure 15: Average Resistivity Computed Over Three Depth Ranges[3]

Conclusions: The Blue Gene/P Enables a Tectonic Shift in Upstream Petroleum

As shown in the examples, large-scale upstream petroleum simulations are now viable with the IBM Blue Gene/P. There has been a lot of progress over the last several years to overcome challenges in solution approaches, algorithms, and domain decomposition techniques. Multidimensional prestack seismic migration and large-scale reservoir simulation problems on complex geometries such as an entire reservoir basin will scale well on the IBM Blue Gene/P. Upstream petroleum applications are once again poised to benefit immensely from Blue Gene's unsurpassed scalability and performance. This will continue to advance the field of exploration and production further.

---

[3] "Massively parallel electrical-conductivity imaging of hydrocarbons using the IBM Blue Gene/L supercomputer", M.Commer et.al , IBM J. RES. & DEV. VOL. 52 NO. 1/2 JANUARY/MARCH 2008.

A tectonic shift is underway in upstream petroleum computing. Dynamic Data Driven Applications and Simulations (DDDAS) is becoming a new paradigm for upstream petroleum computing, one involving symbiotic feedback in which reservoir modeling - simulation and analysis - and sensor field data interact in near real time to dramatically improve the fidelity of the analysis, its accuracy, and reliablility. Energy companies who invest in these breakthrough approaches will obtain big payoffs as they continue to solve the grand challenge problems in reservoir optimization using a combination of innovative parallel solution approaches and more powerful systems based on the Blue Gene/P architecture.

More Information

To learn more about the IBM System Blue Gene Solution, please contact your IBM marketing representative or visit the following Web site: http://www.ibm.com/servers/deepcomputing/bluegene.html.