THE AMERICAN OIL & GAS REPORTER®
JULY 2008
The "Better Business" Publication Serving the Exploration / Drilling / Production Industry

Computers Turning To Parallel Storage

By Larry Jones

FREMONT, CA.–In this era of escalating hydrocarbon prices and more challenging exploration environments, replacement of proven reserves is a critical but increasingly difficult business problem to solve. In addition, companies are faced with deepwater drilling costs that can reach as much as $240 million a well, as in the Tupi Field off the Brazilian coast.

In order to mitigate these challenges and risks, exploration teams are demanding the very best imaging possible for subsalt prospects. Fortunately, a wide array of technical innovations in exploration and well management software is helping oil and gas companies make headway with this problem.

The computing infrastructure also is innovating to support these new software techniques, particularly with the use of new parallel compute clusters. Clusters not only allow the use of much more powerful software tools but deliver superior price and performance. Parallel storage is the key to helping exploration and production organizations find and manage hydrocarbons more effectively.

There are three key trends driving the requirement for a new type of storage known as parallel storage:

- The increasing complexity of processing and imaging algorithms;
- The increasing size and density of the survey data to be stored, processed and imaged; and
- The increasing interrelationship of the phases of the life cycle of exploration and production requiring a tighter workflow.

Advances in the algorithms used to image the geology more accurately have driven advances in hydrocarbon exploration. Algorithmic advances have come in the past 20 years both in wave equation techniques moving from common azimuth to shot point and anticipating interactive PSDM, while ray tracing techniques similarly have evolved from Kirchoff to Gaussian to fast-beam migrations. These more powerful algorithms have been accompanied by much higher computing requirements. Where a Kirchoff migration in the early 1990s may have required a teraflop (1 trillion floating point operations a second) computer on an SMP system, reverse time migrations require much more processing power, on the order of 50 teraflops.

These types of imaging software typically are run on more cost-effective clusters of X86 servers, requiring 1,000 or more processors to achieve this level of computing power. Future algorithms such as elastic modeling or interactive PSDM are expected to require compute resources in the range of a petaflop (1 quadrillion floating point operations a second).

Storage Demands

This level of performance was achieved only this spring by the RoadRunner supercomputer at Los Alamos National Laboratory in a system that used 13,000 conventional X86 processor cores supplemented by 100,000 specialized cell coprocessors. Clearly, the rapid increase and availability of cost-effective compute resources has been a major contributor to the ability of advanced geophysical imaging software to help locate new hydrocarbon deposits.

Demands for storage similarly are increasing to support new land and marine exploration techniques. The amount of surface area in a typical marine acquisition study has increased, steadily driven by the desire to provide a holistic understanding of petroleum systems in basin-wide models. This is complemented by improvements in marine data acquisition technology, including energy source controllers, receivers, positioning and streamer control systems, such that a typical survey has increased by at least a factor of 10–from 300 square kilometers to 3,000 square kilometers, and as much as 8,000 square kilometers, in the past 15 years.

Multi- and wide-azimuth survey techniques have compounded the data explosion by increasing the data density of the surveys from one or two azimuths to as many as six directions, with as many as four vessels a direction for rich azimuth surveys, all of which achieve higher resolution geophysical images.

This higher resolution comes at a cost of much more data to store and process for each survey. A typical 400-square kilometer survey at two azimuths may be three terabytes in size, while a full multiazimuth survey of the same dimensions could require 10 times the storage space. As reservoirs move into production, time-lapse, or 4-D, seismic surveys increasingly are used to determine the changes occurring in the reservoir by comparing the repeated data set surveys of the same reservoir. Those charges typically are the result of hydrocarbon production or injection of water or gas into the reservoir.

The third area of interest is the tighter integration of the exploration and production workflow throughout the life cycle of a hydrocarbon field. This starts with the initial survey, where an imaging project may take six months to complete and require 10 migrations across multiple compute clusters with interspersed quality control and interpretation steps. Once the survey is imaged and turned over to the interpretation team, it may require access to the prestack data to fully assess anomalies in the processed image.

Further downstream, the drilling team may need access to the images to optimize well placement. The production team may complement the reservoir model with additional time-lapse seismic images to optimize reservoir production. Increasingly, the exploration and production environment is becoming data-centric rather than discipline-centric, making the typical practice of copying data across the organization both expensive and inefficient.

Business Challenges

Some businesses are outsourcing companies that process and sell information; others go into actual production of getting oil and gas. Regardless of the type of business, there is a need to get and refine data, and make good decisions on where to look for oil and gas reserves.

The goals of exploration and production organizations have not changed over time: improve exploration success rates, and achieve optimum well production.

To accomplish these goals, IT groups within exploration and production companies face several challenges. The first challenge is how to increase the number of surveys that can be imaged in a given period. This increased throughput allows interpreters to assess more targets and increases the chances that they will find the most valuable ones. This becomes critical when the work is being done as part of a lease auction deadline, where decisions on a wide area must be made with imminent financial ramifications regarding which blocks to choose and how much to bid on them.

For service companies, survey throughput is related directly to revenue, as processed surveys are the ultimate product of these organizations–increased throughput translates to increased profitability. Seismic data is important for drill placement and can become time critical if a problem is encountered in the field and better images are needed to solve drilling bottlenecks.

A second set of challenges involve the exploration tools' increasing complexity. The increasingly complex imaging algorithms and massive amounts of data necessary to lower drilling risks make it more difficult to achieve the goal of speeding survey throughput. New processing algorithms–such as fast beam migrations, shot point, reverse time migration, and interactive PSDM–drive more application complexity, require much greater computing resources and often longer computation cycles. Similarly, the new acquisition techniques such as wide-azimuth, multiazimuth, and 4-D are generating higher volumes of data, in some cases 10 times the levels of a decade ago.

The exploration and production process itself is becoming more complex as the parts of the workflow become more interdependent. Imaging a survey often requires five to 10 separate stages. Often, many of the migrations use different computing resources, as do the quality assurance and interpretation steps. Interpreters are using the prestack data, while reservoir managers are using 4-D seismic data to help guide the drilling and production process. All of these factors suggest that centralizing data and eliminating data motion can only help increase operational efficiency.

The third set of challenges is to reduce the total cost of operations while managing a fast-growing IT infrastructure. In order to maximize return on investment of compute resources, the key leverage point is to utilize fully the most expensive assets. A large investment in compute clusters without reliable, balanced input/output makes that investment inefficient and lowers ROI. The storage infrastructure must have sufficient I/O capabilities to meet the needs of the compute requirements.

A standard storage metric often quoted is one gigabyte a second per teraflop of compute capability. Storage performance is necessary but not sufficient; it also must be reliable. Storage is a key system component–uptime is critical, because if the storage system is down the compute cluster also is down.

Many studies have concluded that 80 percent of a data center's cost is the ongoing cost of management resources rather than the initial capital costs. It is critical to minimize the total cost of ownership (TCO) by minimizing the IT resources required to install and maintain it. This is a particularly difficult issue in remote sites, where support must be provided with minimal IT resources and training. In the exploration and production environment, it is a certainty that IT infrastructure budgets will continue to grow dramatically, while IT administrative budgets will grow at a much slower pace, if at all.

Solution: Parallel Storage

Going back to the needs and challenges of the oil and gas industry, the goals are to find oil and gas and do so quickly. It is a competitive market, and the faster a business can locate and access oil and gas resources, the more successful it will be. Parallel storage addresses the challenges that exploration and production companies face today and into the future.

Parallel storage delivers faster throughput of critical processing, interpretation and reservoir modeling projects. With a parallel storage architecture in place, organizations are provided with massively parallel I/O and improved performance capabilities, enabling their Linux compute clusters and workstations to speed the completion of the workflow. As more and more compute nodes and additional storage capacity are brought into the infrastructure, parallel storage clusters can keep pace efficiently by scaling the data throughput rate with near-linear performance gains, unlike the scalability limitations that traditional storage systems such as direct attached storage, storage area network (SAN) and network-attached storage (NAS) systems experience.

In addition, key exploration and production software vendors such as Landmark, Paradigm and Schlumberger are optimizing their applications to take full advantage of parallel storage. An example is the joint effort by Panasas and Paradigm to make a new processing and visualization system at Statoil achieve seven times the throughput of the previous storage system on an important seismic processing application.

Parallel storage is the key to success if the exploration and production industry is going to maintain the trends that have made the industry successful for the past 15 years. Parallel storage clusters have the ability to scale storage performance to support compute clusters that are growing from 10 teraflops to 100 teraflops to one petaflop over the next decade.

These industry benchmarks suggest that storage performance must increase from 10 gigabytes a second now to one terabyte a second over the next decade. Survey data continues to increase in density by 10 times a decade to provide better images for exploration and more images over time for use in reservoir production management. Clearly, traditional SAN and NAS storage architectures cannot meet these challenges, which is why leaders in the industry are turning to parallel storage.

It is also why some companies have suggested a new industry standard, Parallel NFS or NFS 4.1, which has been supported by storage companies, including EMC, NetApp, IBM and Sun. The industry will adopt Parallel NFS in 2008, and it will begin to be deployed by storage and operating system vendors in 2009. It is the standard that will replace NFS for technical computing requirements and provide a firm foundation for the future of the exploration and production industry.

The price that a company pays for storage matters; however, it is the overall total cost of ownership that is of true importance. Without having to purchase additional storage for data copying and replication and without having to manage data versions and synchronization, oil and gas organizations are impacted at the bottom line. The ability to reduce TCOs through leveraging the parallel storage infrastructures is achieved primarily by being able to "compute in place" rather than have to copy data from one island of storage to another.

The bottom line is that companies will be able to shorten business cycles because of the superior benefits that parallel storage provides compared to traditional storage systems. In fact, the consistent power and performance that parallel storage clusters provide already are allowing many oil and gas organizations to more easily achieve business objectives, further leverage existing capital investments, and achieve a competitive edge in a highly volatile marketplace. ❒

LARRY JONES, *Panasas' vice president of marketing, is responsible for all product marketing at the company. He previously served as vice president of product management at Driveway, a personalized online storage company. Prior to that, Jones was the director of marketing of E*Trade's Business Solutions Group, where he increased customer levels by 200 percent and revenue by 130 percent for the company's business-to-business services. He also has held senior marketing management roles at Lotus Development/IBM for Lotus Notes and cc:Mail, and product management positions at Farallon Computing and Northern Telecom. Jones received a B.A. in history from the University of Michigan.*