DATA AND DISCOVERY: DDN SOLUTIONS FOR SEISMIC PROCESSING

Addison Snell
White paper
September 2012

MARKET DYNAMICS

More Azimuths, More Data
An enduring driver for High Performance Computing (HPC) technologies is a fundamental tenet for science and engineering: Once you solve a problem, it’s no longer interesting. Put another way, you don’t need to design the same bridge twice. The research moves on to the next (usually harder) problem. And until we run out of science – not likely to happen soon – there will always be another minivan to design, another proteome to map, another polymer to simulate. This intrepid progress toward new scientific frontiers drives the need for new technologies.

In seismic processing, this quest is manifest in the desire to create models with ever increasing fidelity. Sensors collect seismic data that is assembled – using multiple HPC techniques – into high-resolution maps of the ancient, long-buried riverbeds that became the final resting places for the dinosaurs that would eventually decompose into petroleum.

There are two ways in which oil companies are pushing their boundaries of scientific discovery. One is in the adoption of higher resolution capture methods, beginning with wide azimuth (WAZ), which uses a multi-sensor array to produce a higher-fidelity image. With WAZ now well-established, oil companies are combining several azimuthal approaches, known as multi-azimuth (MAZ) or in some contexts rich-azimuth (RAZ).

If there is a downside to the transition to these richer formats, it is in the vast increases in data sizes that companies must manage. Single-file sizes in excess of 100TB are not uncommon, and this data must be ingested before it can be processed and analyzed. In some cases a dedicated seismic specialist firm will handle the data capture under contract with an oil company, and in these cases the data still needs to be transferred. One supercomputing class oil company summarized this as follows:

“Before we go out into the field and contract with one of those companies to acquire seismic data, we model the acquisition in our computers. … We’ve got a project that we’ve previously computed and moved the data off to tape, and there’s about 115TB of data that we’re moving back onto spinning disk to do further analysis.”

This assessment is related to the second way in which oil companies continue to reach new levels of innovation: through constant algorithmic evolution. Of all vertical markets in high performance technical computing, oil and gas companies have the highest overall rate of internal software usage – that is, programs and algorithms that are developed in-house.

---

New analytics techniques allow for continued advancement in the interpretation of seismic data, not only for new data coming in, but in many cases, also for historical oil field data. This data is usually saved in the event that new methods are developed to enable the profitable extraction of more oil resources.

**Big Data Driving HPC**

The increasing value of high-performance data management at oil and gas companies is echoed across many other markets that are grappling with the increased demands of Big Data. Big Data is not a specific application, but rather a set of industry trends related to the dramatic rise in volume and availability of data. Precise definitions of Big Data vary by application; what they have in common is an organization’s goal to derive institutional advantage from the information they have access to. Big Data trends reach across many application types, including enterprise, research, and real-time analytics; complex event processing; data mining; and MapReduce analysis of unstructured data. Workflows in seismic processing mirror many of these same trends, particularly with regards to the large volume of data being analyzed.

Another common thread spans all categories of Big Data applications: In a broad-based, end-user survey, both HPC-oriented and non-HPC respondents combined to rate I/O throughput (IOPS) as the top challenge for their Big Data workflows. This comparison included other technical challenges, such as compute cycles and memory bandwidth, as well as datacenter challenges like staffing and facilities costs. Storage capacity ranked second in the technical rating. Users also rated I/O bandwidth and capacity as the two most important storage attributes for Big Data.  

**INTERSECT360 RESEARCH ANALYSIS**

DataDirect Networks (DDN), the largest privately-held provider of high-performance storage, has a large and growing presence in HPC markets. HPC users identify DDN as their storage provider more than any other storage-focused company, with more than twice the mentions of EMC, NetApp, Hitachi Data Systems, or Panasas.  

DDN’s strength in HPC is anchored by its Storage Fusion Architecture (SFA), winner of the HPCwire Editor’s Choice Award for “Best HPC Storage Product or Technology” in each of the past three years. The DDN SFA12000 combines SATA, SAS, and solid-state disks (SSDs) for an environment that can be tailored to a balance of throughput and capacity.

These modular storage systems are built for density and consolidation. With 84 drives in each 4U enclosure, the SFA architecture scales up to 3.6 PB per rack, with a choice of InfiniBand or Fibre Channel connectivity. And leveraging the IOPS capabilities of SSDs, the SFA12000 is capable of 40 GB/sec of throughput for both reads and writes, 1.4 million disk IOPS, and 1.7 million cached IOPS. The combination of modularity and performance features allows users to configure systems that scale up and/or scale out in independent dimensions.

With this combination of I/O and capacity, the SFA architecture is a considerable option for seismic processing applications. For applications that demand

---

performance with minimal data latency, the SFA architecture provides “embedded computing,” a feature that incorporates computational elements within the storage architecture itself. This computation-in-data capability allows the user to perform analytics and queries, or even complete algorithms, without transferring data off the disks to separate clusters of servers. By eliminating the data transfer’s latency, applications that leverage particularly large data sets can see dramatic performance increases.

Another core element of the SFA architecture is its scale, whether it applies to single, large application servers or to multiple application servers with simultaneous access to data. The concept of performance at scale allows organizations to consolidate systems in use by different groups, enabling greater collaboration in developing new approaches. Many organizations are already pursuing IT consolidation for its capital and operational cost efficiencies; DDN presents it as more than a cost savings: it can be a significant competitive advantage.

HPC in seismic processing comes down to a familiar query: How can we maximize the value of our data to our company? And although oil companies traditionally already had some of the biggest data sets anywhere, Big Data trends are at work here too, as higher-fidelity images continue to drive data scalability needs. With embedded computing technologies, the I/O performance of SSDs, and the overall scale of the SFA architecture, DDN can enable oil companies to continue to push the boundary of discovery.

“With embedded computing, applications that leverage particularly large data sets can see dramatic performance increases.”