

PANASAS® ACTIVESTOR® AND LANDMARK SEISSPACE®/PROMAX®

HALLIBURTON

Parallel Scale-Out Storage Accelerates Discovery

Seismic processing for oil and gas exploration and production has historically been second only to the US government in terms of computer usage. This heavy usage remains a tremendous big data challenge as the size of seismic data sets continues to increase with higher fold and richer azimuths, and as processing algorithms require greater and more complicated access to data. Panasas ActiveStor is optimized for highly demanding HPC storage applications in the energy sector. Landmark SeisSpace/ProMAX software, coupled with JavaSeis and ActiveStor, leads the way in providing solutions for seismic processing for oil and gas exploration.

HIGHLIGHTS

THE SOLUTION

- Panasas ActiveStor – manage your data, not your storage system
- Scale-out NAS performance for big data workloads
- Linear scaling to 12PB and 150GB/s
- Global namespace
- Single point of management

FASTER TIME-TO-RESULTS

- Deployment flexibility
- Seismic processing and imaging
- Visualization and interpretation
- Reservoir simulation and modeling

DISTRIBUTED PARALLEL I/O FOR SEISMIC PROCESSING

Early users of clusters for seismic processing applications used serial I/O on a single node and distributed the data from that one node, creating a major bottleneck. In the mid-2000s, some seismic processing systems began to utilize distributed parallel I/O to eliminate these single node I/O bottlenecks.

Landmark targeted the distributed parallel I/O opportunity by incorporating JavaSeis distributed parallel I/O (see www.javaseis.org) into its SeisSpace®/ProMAX® software, the leading seismic processing system in the oil and gas industry. While the user community applauded this move, it was clear that many of the legacy storage systems simply could not handle the increased load placed on them by distributed parallel I/O. In fact, some users of SeisSpace/ProMAX with JavaSeis saturated their existing storage systems so completely that they froze up and had to be rebooted. Users working on individual jobs sometimes had to coax their storage systems back into use by initializing a few compute nodes at a time.

THE MOVE TO INTERACTIVE SEISMIC PROCESSING

Seismic processing is at its best when it is highly interactive, but interactive seismic processing has been an elusive goal in the oil and gas industry for decades. The challenge is to be able to handle the ever-increasing volumes of data efficiently so that geophysicists can quickly process data through a large number of compute nodes. The value of interactive processing comes from the ability of the analyst to visualize and process data in minutes and make decisions quickly.

It must also be possible for multiple users to execute similar operations simultaneously. Seismic processing is characterized by multiple jobs run by multiple users concurrently on a single cluster, typically accessing a single storage system. Despite the rise in computational intensity of algorithms like reverse time migration, data I/O continues to be a bottleneck in many seismic processing workflows, fueled by the increased need to analyze the data, quality control the data, and refine velocities in interactive and iterative workflows.



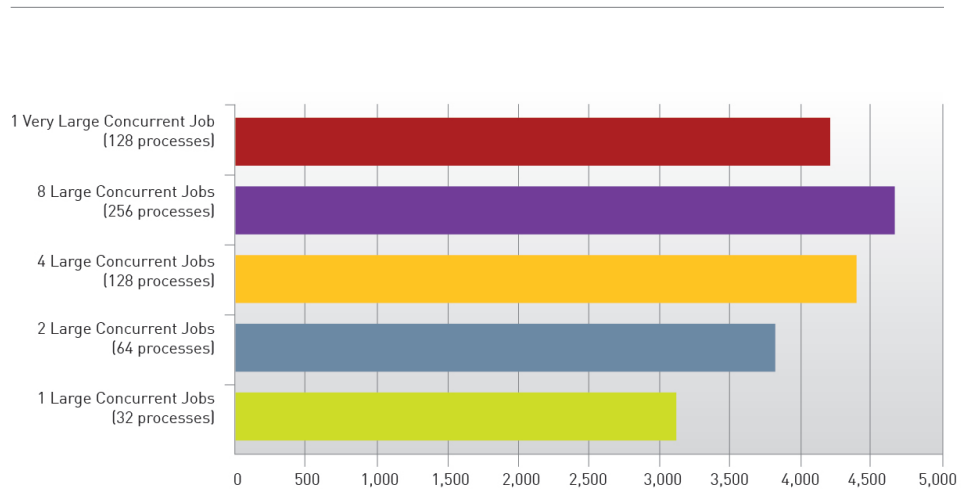
Determination of optimal parameters, and in particular refinement of velocity and statics solutions, requires large amounts of human interaction. While some steps in the workflows are compute intensive, like pre-stack depth imaging, the full workflow includes many other steps that are highly I/O intensive; these include: data preconditioning and quality control of image gathers.

BENCHMARKING LANDMARK SEISSPACE® SOFTWARE ON PANASAS ACTIVESTOR

GeoUsers.org was created to promote collaboration between the users of Landmark's geophysical seismic processing software, SeisSpace, and ProMAX (see www.geo-users.org/seisspace). GeoUsers.org maintains a set of standard benchmarks that test the known seismic processing stress points with distributed parallel I/O. The centerpiece of these tests is the "Heavy Load Effect" benchmark which compares storage device I/O performance under a range of heavy loads. It is the acid test for distributed parallel I/O for seismic processing.

GeoUsers.org's suite of I/O benchmarks, including the Heavy Load Effect, was conducted on 32 cluster nodes accessing four Panasas ActiveStor 12 shelves via a 10 GbE switch. The Panasas storage system showed routine performance across the test suite of more than 3 gigabytes per second, and more than 4.5 gigabytes per second during the Heavy Load Effect benchmark. The results showed that Panasas ActiveStor provides the best parallel I/O performance for SeisSpace®/ProMAX® software, delivering superior top-line performance under heavy loads in all of the tests.

The single, very large concurrent job at 128 processes provides capability computing performance at more than 4GB/s on over 11TB of trace data. Demonstrating the Heavy Load Effect on 1-8 large concurrent jobs, each with 32 processes generating multi-TBs of traces, the capacity performance scales as additional I/O load from the compute cluster is required.



GeoUsers.org Heavy Load Effect Benchmark with Panasas ActiveStor 12

ACTIVESTOR IN A LARGE-SCALE PRODUCTION ENVIRONMENT

The Society of Exploration Geophysicists Advanced Modeling ("SEAM") Phase I Project was conducted, as a consortium of 23 top oil and gas companies (see <http://www.seg.org/resources/research/seam/seamphasei>). The project was one of the largest seismic simulations ever attempted. It included over 7 million CPU-hours of computation and approximately 200 terabytes of data. The project was carried out on a tight budget and with strict deadlines. After initial struggles with slower than expected I/O, Panasas ActiveStor was deployed to handle the heavy workload, both to meet the high I/O demands of the multiple quality control processes applied to the data, and as a cost-effective bulk storage system, meeting all key requirements of the SEAM project. The project was completed with Panasas storage earlier than expected and within budget. This type of large-scale deployment shows what can be accomplished with Panasas ActiveStor and lends further credence to the stellar results demonstrated in the GeoUsers.org benchmark testing.

THE ACTIVESTOR DIFFERENCE

Panasas is the premier provider of high performance parallel storage solutions for technical computing applications and big data workloads. All Panasas ActiveStor scale-out storage appliances leverage Panasas® PanFS®, the Panasas parallel file system, to deliver superior performance, data protection, scalability, and manageability.

The Panasas DirectFlow® protocol eliminates I/O bottlenecks found in traditional NAS systems by allowing cluster nodes to access storage directly and in parallel. Files are written directly to storage without having to first funnel I/O through a file server. Similarly, clients read files by fetching data directly from the storage. As a result of this parallel access, ActiveStor, coupled with SeisSpace/ProMAX software, provides unmatched performance for Linux clusters of all sizes.

Users can simply add individual blade chassis or entire racks to non-disruptively scale capacity and performance as storage requirements grow. This makes it easy to linearly scale capacity to 12 petabytes and performance to a staggering 150GB/s. For these reasons, ActiveStor is perfectly suited for SeisSpace/ProMAX software with JavaSeis to meet the demanding I/O requirements of seismic processing applications.