



PANASAS PARALLEL STORAGE CLUSTERS AT A MAJOR INTERNATIONAL OIL AND GAS COMPANY

CASE STUDY | MARCH 2007

CHALLENGE

This customer's existing SAN-based storage was unable to deliver the necessary bandwidth critical to process large seismic data sets used in the search for oil. This limitation seriously impacted their time to market.

SOLUTION

500 Terabytes of Panasas high performance parallel clustered storage. The fully integrated software/hardware solution included the Panasas ActiveScale® Operating Environment and the PanFS™ Parallel file system with Panasas DirectFLOW® protocol.

RESULTS

- 300% performance increase over previous seismic processing storage infrastructure
- 3.5 GB/sec sustained bandwidth (expected to scale to 17GB/sec as racks are added)

ABSTRACT

One of the most recognized and respected brands for Oil & Gas production has turned to parallel clustered storage technology to improve its seismic processing operations and its research & development efforts. Panasas storage plays an important role in their goal to accelerate delivery of oil to energy markets worldwide, and provides a competitive edge in a highly volatile and demanding marketplace. This case study provides details of the unique challenges of this Oil & Gas conglomerate and how the Panasas storage architecture met their requirements for bandwidth, scalability, manageability, and reliability.



INTRODUCTION

As the worldwide thirst for oil continues to grow with soaring demands from emerging economies such as China and India, the ability to locate and extract oil as quickly as possible has become not only a key differentiator, but a survival imperative for Oil and Gas exploration and production companies. One of the most respected of these companies turned to Panasas to upgrade their computing infrastructure and gain a real competitive advantage.

This industry giant operates in over 100 countries around the globe. Most of their sophisticated 3D seismic processing applications are developed in house and deployed on high performance Linux clusters. Their seismic processing applications perform complex simulations that guide geologists to where and how deep to drill for oil. Time is of the essence, however, these data intensive computations can take days or weeks to complete.

As the number of client and data streams used in seismic imaging operations continued to increase, the company soon realized that its existing SAN-based storage and NFS servers would be taxed to the limit. It was evident that their IT infrastructure would be unable to deliver the ever-growing aggregate bandwidth that is so critical to processing the large seismic datasets used in the search for oil.

THE PROBLEM

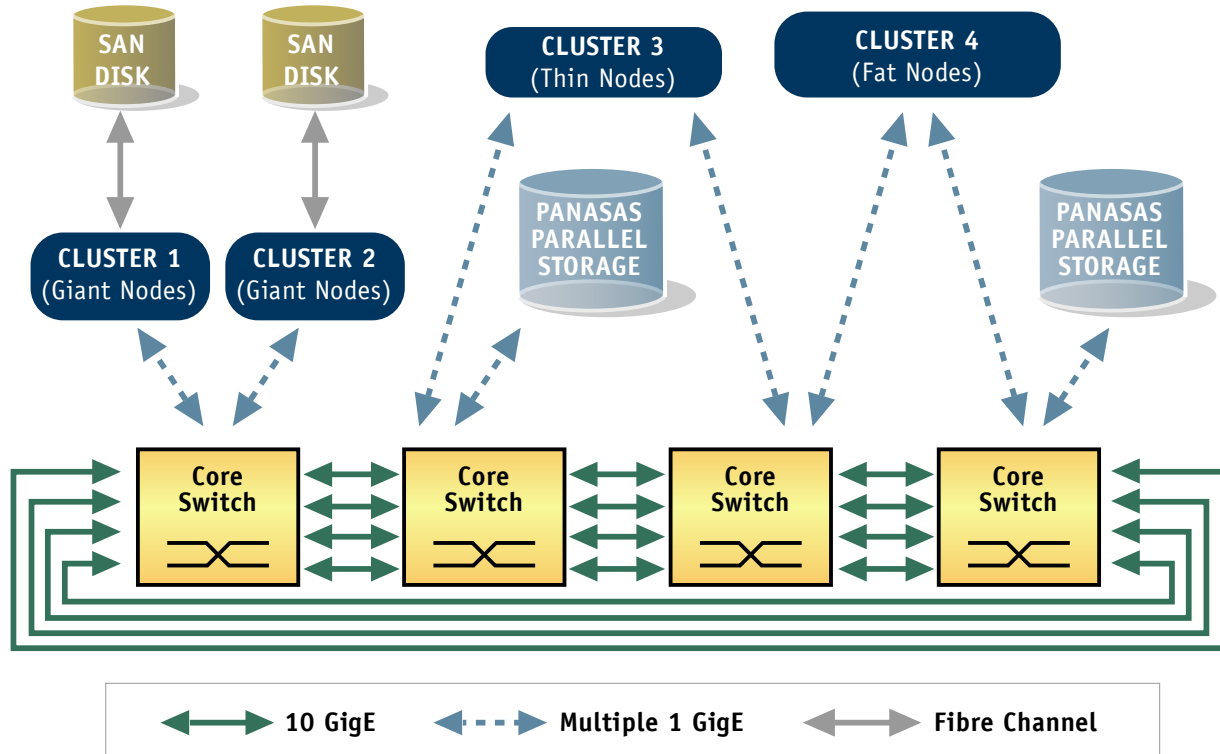
Historically, most of the storage for this Oil & Gas customer was SAN-based and was managed by a global file system. Although every user had access to the global file system, the SAN-based storage created a bottleneck which limited the number of compute systems that could be attached to the storage. In addition, the existing NFS servers were creating performance limitations. A new, higher performance approach to storage for their Linux clusters, such as that provided by the Panasas ActiveStor Parallel Storage Clusters, was required.

While the company had traditionally operated on large-scale symmetric multiprocessing-based systems, its longer-term goal was to upgrade its clustered Linux environment to one that allowed more compute-intensive processing at a more cost effective price point. A complementary high-performance storage solution was also required to enable the processing of complex, seismic calculations into usable data that could help the oil company achieve its business objectives and carve out that highly-prized competitive business advantage.

THE SOLUTION: PARALLEL STORAGE OPTIMIZED FOR CLUSTERS

The computing environment of the Oil & Gas company is a flat network architecture with a core set of Ethernet switches. An expert team of geophysicists use the company's Linux clusters to develop complex imaging algorithms and oil exploration simulations to pinpoint potential deposits deep beneath the ocean floor. It is a time-consuming, expensive and high-risk operation in which mistakes can lead to drilling in the wrong location or lower than expected yields resulting in disastrous financial consequences. On the production side, the Linux cluster administrators control the multiple data streams to ensure that a consistent data rate of information is available where and when it is needed. While most of the cluster computing is done by the teams at the oil company's Houston R&D center, users in Cairo, London, and other centers around the world also access the storage in supporting their leading-edge seismic imaging applications.

HPC Network Configuration



Due to the crippling limitations of their existing storage architecture, an evaluation unit of the Panasas Storage Cluster was first deployed in 2002 to provide sustained I/O from these many data streams into a common, global file system. Today, nearly half a petabyte of Panasas storage has been deployed across five racks to support multiple clusters. The primary Linux cluster has 800 nodes and another secondary Xeon-based Linux cluster has 720 nodes. The Panasas storage solution allows the company to utilize a unified storage infrastructure for both clusters. Performance has increased 300%, a dramatic improvement over their prior seismic processing storage infrastructure.

A typical Panasas storage cluster configuration consists of eight to ten shelves. With their data organized in this way, the company has achieved a sustained bandwidth rate of 3.5 GB/sec. from as many as 400 to 500 client data streams. And as additional 8-to 10-shelf configurations are added to the Panasas Storage Cluster, the data throughput will scale even higher with linear performance to 17 GB/sec. for this five-rack solution.

Transitioning to a more productive and efficient Linux compute clustered coupled with Panasas Parallel Storage Clusters has reduced the time it takes to run simulations and significantly sped up critical business decisions.

PANASAS OUTPACES THE REST IN STRINGENT TESTS

During their search for a solution that would improve upon the existing NFS servers, extensive competitive testing of advanced file systems were conducted. The Panasas Parallel Storage Cluster outpaced other storage products by delivering a consistent aggregate bandwidth when large numbers of data streams were running against a single file system. Panasas' integrated software/hardware design, in conjunction with the parallel Panasas DirectFLOW® Protocol, offered a solution with the critical capabilities that the oil company sought:

- No bandwidth limitations for an ever-growing number of compute nodes and data flows
- Ease in administering systems and integrating them into an existing computing infrastructure
- A unified storage solution with a single global namespace eliminating the need for islands of storage
- A smooth transition path to next-generation supercomputers so that the backend storage remains constant
- No kernel changes to the Linux OS

The Panasas solution provided all of these capabilities. With new levels of compute performance their teams of scientists can achieve results more quickly and make decisions more accurately in their race against time to discover and analyze where and how to drill for oil.

SUMMARY

One of the world's largest Oil & Gas producers has successfully deployed Panasas Storage Clusters to not only provide unprecedented performance, scalability, manageability and availability for its ever-growing seismic imaging needs, but to also create a storage infrastructure that is poised for next-generation Petascale computing as it becomes available. Using Panasas storage, the company has been able to eliminate major problems slowing down production and affecting their bottom line; the time-consuming and costly bottlenecks resulting from existing NFS servers, and bandwidth and scalability limitations of a prior SAN-based file system that restrict the number of compute nodes that can be attached to the storage. As more and more compute nodes and additional storage capacity are brought into the infrastructure, the Panasas Parallel Storage Cluster can efficiently keep pace by scaling the data throughput rate with near-linear performance gains. With consistently having the power they need, the Oil & Gas company can more easily achieve business objectives, further leverage existing capital investments, and achieve a competitive edge in a highly volatile marketplace.



Accelerating Time to Results™

6520 Kaiser Drive Fremont, California 94555 Phone: 1-888-PANASAS Fax: 510-608-4798 www.panasas.com
 1-888-PANASAS (US & Canada) 00 (800) PANASAS2 (UK & France) 00 (800) 787-702 (Italy) +001 (510) 608-7790 (All Other Countries)

©2007 Panasas, Inc. All rights reserved. Panasas, the Panasas logo, and DirectFlow are trademarks or registered trademarks of Panasas, Inc. in the United State and other countries. All other trademarks are the property of their respective owners. Information supplied by Panasas, Inc. is believed to be accurate and reliable at the time of publication, but Panasas, Inc. assumes no responsibility for any errors that may appear in this document. Panasas, Inc. reserves the right, without notice, to make changes in product design, specifications and prices. Information is subject to change without notice.