

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

THE CUSTOMER

The Andrej Sali Laboratory at the University of California, San Francisco uses computation grounded in the laws of physics and evolution to study the structure and function of proteins. The Sali Laboratory strives to improve and apply methods for predicting the structures of proteins, determining the structures of proteins and macromolecular assemblies and annotating the functions of proteins using their structures. By contributing to structure-based functional annotation of proteins, this research enhances the impact of genome sequencing, structural genomics, and functional genomics on biology and medicine. To meet their scientific goals, the Sali Laboratory leverages advanced computing solutions for high-throughput structural and functional studies of proteins.

THE CHALLENGE

In order to find the structural similarities of protein models, the team at Sali Laboratory knew they needed a computing solution that could handle a massive I/O workload. The environment is responsible for analyzing 2,000,000 protein sequences against 30,000 experimental structures in an effort to predict 3D structures. “Our system has extreme I/O requirements,” said Dr. Ursula Pieper, Assistant Researcher at Sali Laboratory, “We constantly process millions of small files and need an IT solution to return results to our researchers as quickly as possible.”

For the compute side of the solution, the Sali Laboratory grew a 370 dual processor Linux cluster. While this solution met their computational needs, storage system performance and administrator management were major concerns. The Sali Laboratory was looking for a solution that could deliver exceptional I/O and had the ability to scale capacity. Key to capacity scaling was finding a single, global namespace to manage all of the data sets in a single system image. Using direct attached storage in the past had delivered inconsistent performance results and the Sali Laboratory knew that management problems were on the horizon as the system scaled in size. Finally, realizing that they were



Industry: Life Sciences

The Challenge:

This customer needed a computing solution that could process millions of small files as quickly as possible for researchers identifying the structural similarities of protein models. They needed to eliminate poor storage system performance and significantly decrease administrator management time and within a limited budget.

Panasas Solution:

The fully integrated software/hardware solution included the Panasas® ActiveScale® Operating Environment and the PanFS™ parallel file system with the Panasas DirectFLOW® protocol.

Key Results:

- Up to 5X Performance Improvement
- A single namespace for simplified cluster management
- Maximized ROI from clustered computing environment

working with a constricted academic budget, price/performance and overall system value were key requirements. “Certain financial constraints are common in academia, so we had to find a solution with the best price/performance,” said Dr. Pieper.

“After evaluating several solutions, it was clear that Panasas had the most comprehensive offering.”

- Dr. Ursula Pieper
Assistant Researcher at Sali Laboratory



THE SOLUTION

An extensive evaluation process included detailed reviews of several high performance and next-generation network attached storage solutions. Ultimately, the Panasas® Parallel Storage Cluster was selected for its random I/O performance capabilities, ease of management through a seamless global namespace and extreme value. The multi-TB solution was connected to the Linux cluster for storing and retrieving computational data sets. “After evaluating several solutions,” said Dr. Pieper. “It was clear that Panasas had the most comprehensive offering.”

The Panasas Storage Cluster leverages a distributed file system to provide direct disk to client access via the Panasas DirectFLOW® protocol. Further, the system uses finely tuned hardware components – Panasas StorageBlade® module and Panasas DirectorBlade® module – to support record-setting random I/O. The Panasas ActiveScale Operating Environment is built on an object-based architecture, which allows seamless scaling in capacity with a unified, global namespace. Finally, by leveraging industry standard components, Panasas is able to offer this solution at extremely competitive price points.

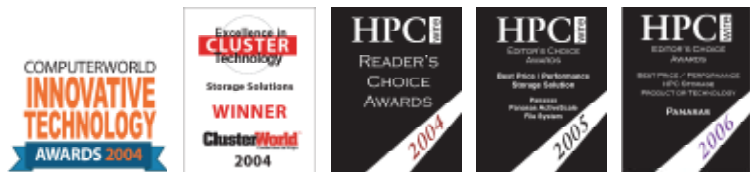
THE VALUE

Upon moving the Panasas Storage Cluster into production, the Sali Laboratory immediately experienced a performance improvement. “The Panasas I/O performance is extremely high,” said Dr. Pieper. “We’ve eliminated our I/O bottleneck to complete our computations more quickly. In fact, the Panasas system has cut the job times for the computation of protein interactions, one of our most I/O intensive projects, from 40 hours to less than 8 hours.” Even more impressive is the fact that as the number of clients accessing the system increases, performance remains the same.

Further, the integrated software/hardware solution and single unified namespace streamlined system installation and management. “The system is easy to manage,” said Dr. Pieper. “We don’t want to worry about IT issues, and with the Panasas solution we don’t have to.” The single namespace offered by the Panasas Storage Cluster allows significant capacity growth with no worries about re-partitioning or managing discrete data islands. Finally, Panasas has provided exceptional service and support. “The entire team at Panasas has been great in helping us every step of the way,” said Dr. Pieper.

“We’ve eliminated our I/O bottleneck to complete our computations more quickly. In fact, the Panasas system has cut the job times for the computation of protein interactions, one of our most I/O intensive projects, from 40 hours to less than 8 hours.”

- Dr. Ursula Pieper
Assistant Researcher at Sali Laboratory



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 Incorporated. All rights reserved. Panasas, the Panasas logo, Accelerating Time to Results, DirectFLOW, DirectorBlade, StorageBlade and ActiveScale are trademarks or registered trademarks of Panasas, Inc. in the United States and other countries. All other trademarks are the property of their respective owners.