

# UNIVERSITY OF CALIFORNIA, SANTA CRUZ

## THE CUSTOMER

The Center for Biomolecular Sciences and Engineering at University of California, Santa Cruz (UCSC) launches interdisciplinary research and academic programs that address the scientific questions of the post-genomic era. The Center uses computational, mathematical, and statistical approaches to probe and analyze biological data, from DNA to biological processes to healthcare systems. One of the Center's major projects is the UCSC Genome Browser, a web-based tool that allows researchers to view all 23 chromosomes of the human genome from as large as a full chromosome down to an individual nucleotide. The UCSC Genome Browser integrates the work of numerous scientists in laboratories worldwide and includes work generated at UCSC in an interactive, graphical display.

## THE CHALLENGE

The UCSC Genome Browser leverages extremely fast search software that runs on the KiloKluster, a second-generation 1000+ node bioinformatics Linux cluster. It enables researchers to match any DNA sequence to the human genome in seconds and maps experimental data to the reference sequence. In order to process the Browser's huge quantity of data, the Center searched for a storage solution that delivered high performance random I/O to a large number of cluster nodes. "Our KiloKluster really taxes the capabilities of a storage system," said Robert Baertsch, Research Assistant at UCSC. "To be fully effective, a storage solution in our environment needs to be able to deliver exceptional performance with a large number of cluster nodes."

UCSC's system needed to scale in performance as well as capacity. As a result, the Center searched for a solution that had the potential to scale as a large single pool of data. Similar to many universities, the UCSC researchers work on several complex projects at any one time. The ability to focus solely on their tests, instead of the systems that were running them, was critically important. "Many of our programs would take years to run on a single CPU. Having a cluster with many nodes shortens run time to days or

# UC SANTA CRUZ



**Industry:** Life Sciences

### The Challenge:

Slow I/O and downtime impacted the run times of their Genome Browser search tool used by scientists in their work to solve questions of the post-genomic era. They were searching for a storage solution that delivered high performance random I/O to an exceptionally large number of cluster nodes and one that would allow them to focus solely on their tests instead of the systems running them.

### 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:

- Exceptional I/O Performance
- A single namespace for simplified cluster management
- Maximized ROI from their clustered computing environment

even hours making our research possible," said Baertsch. "Slow I/O and downtime can really impact overall run times and it is critical to have a storage system that can scale to thousands of nodes with a single system image." Finally, price is always a major consideration for the Center. A fundamental requirement is to deliver a compelling price point.

“*The Panasas Storage Cluster has reduced run times by over 40 hours.*”

- Robert Baertsch  
Research Assistant at UCSC



## THE SOLUTION

The Center conducted an extended evaluation process including detailed testing with many high performance network-attached and direct-attached storage solutions. After thorough testing, the Panasas® Parallel Storage Cluster was selected for its ability to deliver high-speed random I/O performance in a large cluster environment, simplified management through a scalable, shared pool of storage and exceptional value. The Panasas Storage Cluster is now connected to the KiloKluster to store and retrieve reference sequences. “A typical NFS server is brought to its knees when 1000 cluster nodes are pulling data from it,” said Baertsch. “With the Panasas Storage Cluster and its object-based architecture, we are able to simultaneously read from and write to all cluster nodes.”

The Panasas Storage Cluster helps organizations like the Center for Biomolecular Sciences and Engineering accelerate the speed and accuracy of its decisions and ultimately, lead to real world breakthroughs that improve people’s lives. The Panasas Storage Cluster enables the Center to maximize the benefits of Linux cluster computing by breaking down the storage bottleneck created with legacy network storage technologies. The solution is powered by the Panasas ActiveScale Operating Environment and the company’s unique object-based storage architecture. In addition to exceptional performance benefits, the system enables seamless growth of a single namespace, greatly improving system manageability. Finally, by leveraging industry standard components, Panasas is able to offer this solution at an extremely competitive price.

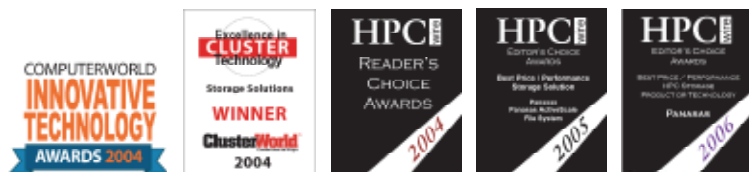
## THE VALUE

The Center for Biomolecular Sciences and Engineering was able to see a significant performance boost once the Panasas Storage Cluster was moved into production. “The consistently high I/O performance delivered by the Panasas solution enables our researchers to get their results more quickly,” said Baertsch. “By using the Panasas DirectFLOW® protocol we’ve been able to eliminate our I/O bottleneck.” In fact, for specific batches of jobs the Panasas Storage Cluster has reduced run times by over 40 hours. Perhaps even more important, by leveraging the object-based architecture, Panasas has been able to offer the Center new ways to look at increasing overall performance. “Panasas has given us many ideas on how to scale I/O and we look forward to further experiments,” said Baertsch.

The Panasas Storage Cluster’s ease of management also added tremendous value to the Center’s solution. The single unified namespace ensures administrator management will be streamlined today and in the future. As the system capacity requirements increase, the Center is confident that the Panasas Storage Cluster can increase in size with no impact to administrator management.

“By using the Panasas DirectFLOW® protocol we’ve been able to eliminate our I/O bottleneck”

- Robert Baertsch  
Research Assistant at UCSC



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 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.