"The system is blazingly fast. We've been able to eliminate our I/O bottleneck and thus carry out data analysis more quickly."

Terry Gaasterland
Associate Professor of
The Rockefeller University

Gaasterland Laboratory

The Gaasterland Laboratory of Computational Genomics at The Rockefeller University aims to create and use new software tools to explore the unfolding world of genomes. The tools are designed to integrate, analyze, and visualize the output of high-throughput molecular biology experiments in the context of complete genome sequence data. In an effort to apply new tools to specific biological questions in the most efficient way possible, a prerequisite is a high performance storage solution that is easy to access and manage.

SUMMARY

Industry: Life Sciences

THE CHALLENGE

Explosive data storage had caused I/O bottlenecks, increasing the time required to develop new software tools used in molecular biology experiments. Managing the storage system had become a huge administrative burden and it was disrupting productivity, increasing analysis time and posing data availability problems.

THE SOLUTION

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

THE RESULT

- Improved System Reliability
- Increased I/O Performance
- Exceptional Ease of management

The Challenge

When the Gaasterland Laboratory started in 1998, Dr. Terry Gaasterland, Associate Professor of The Rockefeller University, initially speculated that she would only need 500 GB of data capacity. However, as the files to be analyzed grew in number and size, she quickly realized that she needed far more storage. "Our data needs were huge," said Dr. Gaasterland. "The data sets were growing and changing all the time." Not only were the number of data sets to test increasing, but as new tools were developed, each tool would result in a new data set that needed to be stored and accessed for comparison to the original. Data growth was explosive and managing these ever-changing data sets was an administration burden.

To attack this problem, Dr. Gaasterland set out to find a storage system to achieve three key objectives. First, the system had to scale seamlessly without disruption to current users or mount points. Second, it needed to provide fast data throughput in order to reduce the time necessary to analyze data. Genomic data analysis is I/O bound and the new storage solution should eliminate the I/O bottleneck traditionally experienced with network storage solutions. Finally, the system needed to be extremely reliable and available to an extended community outside the Lab.

The Solution

During the evaluation process, Dr. Gaasterland and her team entertained solutions from IBM, Sun and Apple, but the Panasas® Storage system was selected for its performance, reliability and ease of management. Having experience with distributed file systems, Dr. Gaasterland knew that the Panasas Operating Environment architecturally had the capability to solve all their problems. The multi-TB solution was deployed to interact with Apple and Sun desktop workstations.

"Panasas met all the requirements on my checklist," said Dr. Gaasterland. "Even though Panasas was relatively new, we felt confident that they could provide not only the product but the support necessary to meet all our needs."

The Result

In the six months after moving into production, the Panasas Storage system was everything the team expected and more. "The system is blazingly fast," said Gaasterland. "We've been able

to eliminate our I/O bottleneck and thus carry out data analysis more quickly." The Panasas system is designed from the ground up to deliver exceptional random I/O and data access throughput.

Further, the single seamless namespace and dynamic growth capabilities had a significant impact on the organization. "The product is 'plug-and-play' at all levels," commented Gaasterland. "Our team doesn't have to think about it at all. We can focus our attention on biology research questions instead of systems administration issues." With the Panasas Storage system, there is no reconfiguration required to expand data sets. The manageability features enable the system to grow without having to grow the overhead to support it.

Finally, during a recent two-day power outage the system's reliability was tested and proved to be more than capable. While other vendors' systems failed to provide a graceful shutdown to guarantee data integrity, the Panasas Storage system was the only storage system in the lab that maintained complete data integrity. "At the end of the day, data availability and integrity is essential to do our job," Gaasterland said. "Panasas proved itself to be more than reliable enough for our environment."

> **"Our team doesn't have to think about it at all. We can focus our attention on biology research questions instead of systems administration issues."**
>
> Terry Gaasterland
> Associate Professor of The Rockefeller University

About Panasas

Panasas, Inc., the leader in high-performance scale-out NAS storage solutions, enables enterprise customers to rapidly solve complex computing problems, speed innovation and bring new products to market faster. All Panasas solutions leverage the patented PanFS™ storage operating system to deliver exceptional performance, scalability and manageability.

panasas | Phone: 1-888-PANASAS | www.panasas.com