The Challenge

- Share data across a 30 node bioinformatics compute farm

- Solve the performance bottleneck of a single NFS server

- Reduce costs

- Deliver cost-effective file serving performance

- Allow for non-disruptive expansion

The Solution

- 30-node NCBI Blast compute farm

- PolyServe Matrix Server

- Two-node NFS cluster of Intel-based servers with plans for expansion

Results

- Almost doubled NFS throughput by adding an additional server

- Realized better price-performance and absolute performance in NFS throughput than alternatives.

- Saved 60% savings in Total Cost of Ownership (TCO) compared to the alternatives

- Non-disruptive expansion of server and storage capacity

National Yang Ming University and PolyServe Build a Scalable File Serving Cluster for Bioinformatics

When it comes to institutions of advanced research in life sciences and bioinformatics in Taiwan, National Yang Ming University (NYMU) is certainly the first that comes to mind. Established in 1975, NYMU has since grown to become a highly respected institution of education and academic research.

As a university focusing on medical sciences, NYMU is the leader of biomedical research in Taiwan. The collaborative research team of the Veteran General Hospital, Taipei (VGH-TPE) and the National Yang Ming University (NYMU) not only earned international recognition for sequencing human chromosome 4, but also established NYMU as a leading university in genomics research.

NYMU's genome research center, National Yang-Ming University Genome Research Center (http://genome.ym.edu.tw), was established in order to conduct a wide variety of sequencing and genomics projects. The Center has launched several collaborative research projects with other industrial and academic institutions in Taiwan and abroad. To integrate the sequencing, microarray, and proteomic

1

data that these activities generate, NYMU has established the Bioinformatics Research Center (http://ymbc.ym.edu.tw).

THE CHALLENGE

NYMU faced computing performance and scalability obstacles as it continued to push the boundaries of biomedical research. Originally, NYMU used a 6-node compute farm to annotate the human genome in 1999. In addition to using a high performance IBM pSeries 690 computer in its research, NYMU also uses a 30 node compute farm to perform the complex analysis and calculations required in genome and protein sequencing. This compute farm runs a sequence similarity search tool called NCBI Blast. These Intel-based compute nodes constantly query large bioinformatics databases running the popular Blast program. The databases are shared with the compute nodes via NFS.

As the size of the databases and the complexity of the research rapidly grew, NYMU needed to increase the size and performance of the compute farm. However, more compute nodes in the farm exacerbates the I/O performance to the shared bioinformatics data. Originally, NYMU had deployed a single NFS server, but this server acted as a bottleneck to the overall performance of the bioinformatics compute farm.

NYMU needed to solve its performance and scalability challenges without investing large sums in expensive proprietary solutions. Therefore, NYMU was particularly interested in NFS solutions that delivered outstanding price-performance.

THE SOLUTION

NYMU had a pressing performance bottleneck that was only going to get worse as the scale and scope of NYMU's research accelerated. Therefore, after

careful consideration, NYMU decided to implement a file serving cluster with shared SAN storage for the following reasons:

- Scalability
- Performance
- Easier cluster and storage management

NYMU chose to build the cluster using cost-effective Intel-based servers running Linux and PolyServe Matrix Server. Using Intel servers and Linux allowed NYMU to realize higher file serving performance at an attractive price point.

Implemented jointly by PolyServe's distributor, eRexi, and a SAN reseller, NYMU deployed a two-node NFS cluster consisting of two Intel-based SMP Linux servers connected to a Fibre Channel switched SAN with 1TB of shared SAN storage.

"PolyServe and eRexi provided the expertise that made deploying a scalable file serving cluster simple and painless."

Yu-Tai Wang
System Manager,
Bioinformatics Research Center at
National Yang Ming University

PolyServe Matrix Server aggregates the NFS I/O throughput across the servers in the cluster. Matrix Server, a Cluster File System (CFS) with integrated high availability capabilities, enables both NFS servers to concurrently read and write data to the share storage.

With PolyServe, NYMU can horizontally scale its NFS cluster and eliminate the performance bottleneck that has plagued their research. Matrix Server allowed NYMU to overcome the demanding performance requirements of their compute farm of NCBI Blast servers. None of the other clustering solutions that

2

NYMU evaluated provided the scalability and growth path that NYMU required to support its bioinformatics research.
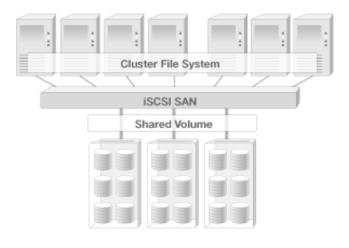
THE RESULTS

The PolyServe solution exceeded the scalability and performance of a single NFS server at a cost effective price point. The PolyServe cluster addressed the scalability and performance bottleneck posed by the NYMU's demanding, data-intensive compute farm. Moreover, NYMU expects the PolyServe solution to continue to pay dividends as its bioinformatics research intensifies.

"We evaluated several alternative solutions and none could match the PolyServe file serving solution in terms of scalability and cost. We were very impressed by the scalability and ease of expansion of the PolyServe solution"

Ueng-Cheng Yang
Director,
Bioinformatics Research Center at
National Yang Ming University

The redundant NFS cluster allows the bioinformatics compute farm to share data without a single point of failure. The pair of clustered NFS servers nearly doubled the aggregate I/O throughput and CPU capacity available to the NFS clients and added high availability capabilities that would not be present had NYMU used a single NAS appliance or NFS file server.

Furthermore, NYMU has a future growth path and investment protection that addresses NYMU's increasing performance requirements. Due to the scalability and online expansion capabilities of PolyServe Matrix Server, NYMU can both expand the size of the storage and add performance to the cluster, while at the same time avoiding the complexity of storage and cluster management.

The PolyServe-powered NFS cluster has successfully been deployed at NYMU since January, 2003. At present, NYMU is considering expanding the NFS cluster by adding a third node. The third node will provide additional performance to the compute farm and provide a backup node from which to perform network-free backups of shared SAN storage.

Cluster File System

iSCSI SAN

Shared Volume