Customer Success Story

University of Cologne

University of Cologne

Panasas ActiveStor

July 2006

Abstract

In 2004, the Center for Applied Informatics at the University of Cologne in Germany, sought to bring high-performance computing (HPC) in the form of Linux cluster computing to one of the oldest and most prestigious institutions of higher education. Over the years as the demand on their HPC resources continued to grow, their existing storage systems could not keep pace. This case study details the challenges faced by the Center at the University of Cologne and how Panasas storage successfully met the Center's requirements for a scalable, ultra high-performance storage infrastructure.

Introduction

The Center for Applied Informatics at the University of Cologne in Germany is a centralized resource for the study of applied computer science and general-purpose computing services for more than 50,000 students and faculty. In addition to serving the University, this testing facility also supports regional academic research projects that require high-performance computing. The facility was established by merging the resources of the University's regional computing center with those of the University's center for parallel processing.

In 2004, the Center sought to bring high-performance computing (HPC) in the form of Linux cluster computing to one of the oldest and most prestigious institutions of higher education. Researchers would be able to use a NPACI Rocks Linux cluster for climate modeling simulations and supporting other applications in chemistry and geophysics. Over time, researchers at other universities and regional technical laboratories would come to rely on the Center to provide the computing capabilities they needed to study complex problems ranging from molecular processes to stellar formation.

The applications being undertaken at the Center required substantial computing power and generated large amounts of data. As the demand on HPC resources continued to grow, the existing storage and associated NFS file systems that supported the HPC environment could not keep pace. A comprehensive evaluation was launched to select a new high-performance storage solution and led to the selection of Panasas storage.

This case study details the challenges faced by the Center for Applied Informatics at the University of Cologne, and how the Panasas storage system has successfully met the Center's requirements for a scalable, ultra high-performance storage infrastructure.

Linux Cluster Architecture

To support HPC users at the University of Cologne, the Center procured two new parallel compute systems including a distributed memory system known as Clio and a shared memory system called Altix1. Clio is a 129-node Linux NPACI Rocks cluster that has been in operation since late 2004. It serves as an entrance computer and it is comprised of the following:

- 2 x 2.2 GHz AMD Opteron processors
- 4GB RAM
- 2 x 36 GB SCSI discs
- 6 x Cisco 3750 switches
- 2GB Ethernet connections
- 1 x 10 GB Infiniband interface
- 1 x Sun Fire Z20z system
- Red Hat Linux Enterprise 3.0

The cluster also supports a shared memory model for parallel programming with an MPI library. For research that is not conducive to efficient parallel processing, the

Center secured an SGI Altix system – an SMP-based computer with 24 Itanium-2, 1.5GHz processors and 48GB of memory. The Center also uses a variety of compute services, including Sun Microsystems SunFire and Sun Ultra Enterprise servers for user applications that require high disk storage and fast I/O to achieve results.

System Characteristics

The NPACI Rocks Linux cluster is used for meteorology research (climate modeling) that requires substantial computing power and generates very large data sets. Additionally, applications in chemistry and geophysics make use of the high-performance cluster to perform complex data analysis and simulations. Research conducted at regional scientific laboratories relies on the Center's HPC resources to test complex theories and conduct compute-intensive research that can only be accomplished with supercomputing capabilities. In each of these projects, the network storage infrastructure must be very fast. It must be capable of approximately one gigabyte per second of throughput with near-linear performance scalability as more storage is added.

Putting the Industry's Best to the Test

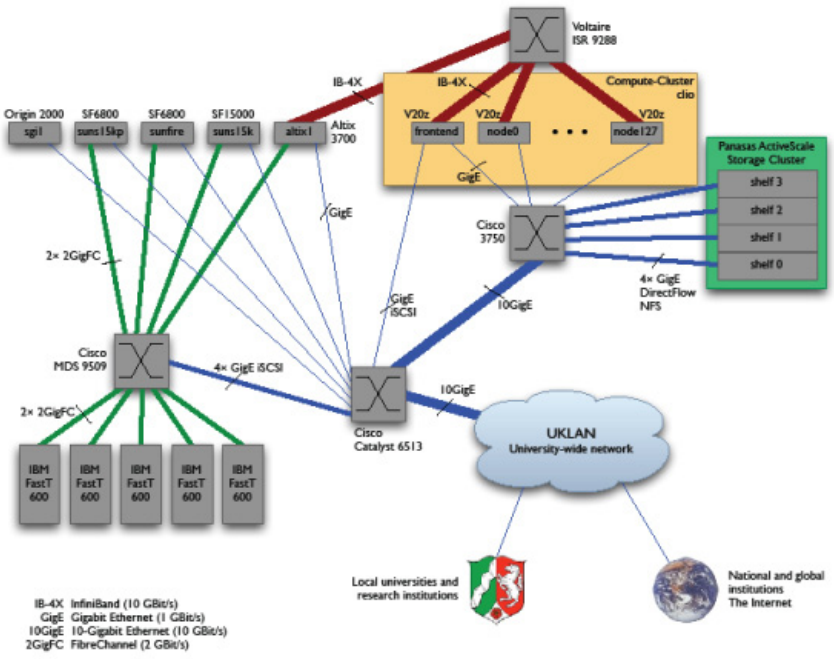
The Center realized its reliance on legacy NFS servers was creating a significant performance bottleneck. The management of the system was also become a burden to the IT staff. The decision was made to upgrade the storage infrastructure.

The Center needed a new storage system that could easily integrate into the existing environment, and scale with ease to support multiple clusters in the future. This would dramatically reduce time to deployment and enable the Center to maximize previous IT investments. Because the Center projected the HPC environment would increase from just over a hundred nodes to thousands of nodes within two years, the new storage solution would have to easily scale to thousands of terabytes without significant reductions in performance. The new storage system also needed to deliver the highest levels of availability as the demand for cluster time was expected to grow. Finally, because the Center had a relatively small IT staff, system reliability, ease-of-use and low management overhead were considered "must-have" features.

Panasas Makes the Grade at University of Cologne

The Center tested solutions from several leading storage vendors using agreed-upon parameters in performance and I/O. The Panasas storage solution was purchased after only a few weeks of evaluation. While benchmarking continues today, the Center has thus far achieved more than a 30X performance improvement over the previous storage system. Benchmarks show a transfer rate of up to 1.3 GB/second when running on 128 dual-processor nodes accessing 20TB of disk space. The Panasas storage solution was deployed in less than four hours while competitive solutions took weeks to implement.

Each of the products offered a different methodology for incorporating NFS, CIFS or GPFS as part of their



overall solution. Panasas' PanFS™, with its fully-parallel DirectFlow® protocol, was the only system to provide direct access between the Linux cluster nodes and storage system to eliminate the performance bottleneck caused by legacy storage systems. Unlike other systems, the Panasas solution required no modifications to the Linux OS kernel.

Summary

The Center for Applied Informatics at the University of Cologne has successfully deployed the Panasas storage system, creating a storage infrastructure with unlimited scalability, higher performance and higher availability. The Center uses Panasas storage with DirectFlow client access software to provide very high bandwidth to all nodes in a NPACI Rocks Linux cluster without the performance bottlenecks associated with legacy file systems. The Panasas storage solution enables the Linux cluster to avoid idle time and continue processing jobs without waiting for I/O processes, improving the speed at which research projects can be completed. As the Linux cluster grows from hundreds to thousands of nodes, Panasas storage is more than capable of keeping pace by scaling performance in near-linear fashion as capacity is increased.