BeeGFS®

BEEGFS PARALLEL FILE SYSTEM

A BeeGFS deployment at AbbVie delivered 1.1PB of high performance storage. BeeGFS transparently spreads data across multiple servers, linearly scaling performance and capacity, from small clusters to large HPC systems with thousands of nodes.

ABBVIE CONFIGURATION

- 3x Advanced Clustering BeeGFS Storage Blocks with 1.1 PB of usable space. Additional storage blocks can be added in order to scale both performance and capacity.

- 2x Metadata servers with mirroring for high availability.

- It also includes BeeGFS commercial support from ThinkParQ.

BeeGFS at AbbVie Genomics Research Center

BeeGFS Parallel File System Supports Large and Small Scale BioPharmaceutical Research

AbbVie is the sixth largest research-based biopharmaceutical company in the world, dedicated to addressing some of the world's greatest health needs by delivering leading-edge therapies and medicines. Best known for its blockbuster autoimmune disease drug, Humira, the S&P 100 company reported profits of $5.7 billion in 2018.

One of the company's many focuses lies in the study of the human genome, which contains variations that determine such traits as blood type, eye color and height. Scientists working for AbbVie's Genomics Research Center are examining these genetic markers in search of more personalized approaches to medicines and treatments.

These efforts require significant high performance computing support for such tasks as analyzing hundreds of thousands of complex datasets. The company's researchers now have access to more than 1.1PB of high performance storage with the deployment of Advanced Clustering's Storage Blocks with the BeeGFS parallel file system.

The 1.1PB BeeGFS parallel filesystem consists of two mirrored Metadata Servers and three Advanced Clustering Storage Blocks with BeeGFS. All systems are connected to an InfiniBand network. The system provides 18 GB/s of throughput to the researchers.

Advanced Clustering's Storage Blocks system is designed to provide maximum throughput for storage, which is important because hard disk drive performance has remained stagnate for years while CPU performance has increased steadily.

In other storage systems, performance is tied to a single storage server, which means there is limited capacity, space and I/O operations. The potential for scalability and expansion are limited by the capabilities of each individual storage server.

Advanced Clustering's solution expands the potential for parallel storage by aggregating multiple storage servers into a single filesystem. This means I/O can be processed by multiple servers, increasing storage capacity and enabling scalability. If you need more performance, simply add more storage servers.

"For our genomics workloads in particular, we need scalable and quick storage," said Martin Forde, Systems Administator at AbbVie. "With BeeGFS, we can store multiple petabytes of data but can also access the data quickly because it's shared across multiple servers."

Prior to deploying the BeeGFS storage solution, AbbVie was reliant on data storage solutions that were not scalable.

"Scalability is very important to our users," Forde said. "With the new BeeGFS solution, we've had quicker reads and writes and lower latency. People coming in no longer compete with our bandwidth for the data."

The storage solution has brought faster access to data for the AbbVie team of researchers.

"In terms of transfer speeds, we're experiencing three times faster than what we experienced previously," Forde said. "We support a lot of projects. We need to be able to support that quantity of data. There's a lof of small and mixed data. BeeGFS has made it possible for us to support large files and quickly access the small data – hundreds of millions of smaller files as well."

> **BeeGFS has made it possible for us to support large files and quickly access the small data – hundreds of millions of smaller files as well.**
>
> **– Martin Forde
> Systems Administator
> at AbbVie**