



Overview of BeeGFS Parallel File Systems

BeeGFS Parallel File Systems

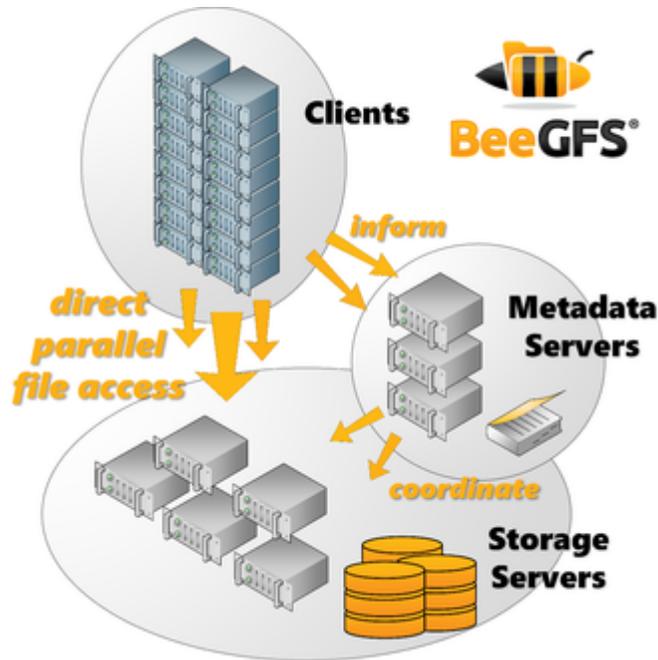
When it comes time to adding high performance storage to your HPC cluster, file systems can become a limiting factor. The cost of the storage expansion can sometimes be as expensive as the HPC cluster itself. This is where BeeGFS, and other parallel file systems come in as a cost-effective, highly scalable, high performance alternative to traditional storage platforms.

BeeGFS is a leading parallel cluster file system. As a parallel file systems, BeeGFS spreads user data across multiple servers in an effort to simplify I/O intensive workloads. By increasing the amount of servers and disks in a system, users can scale the performance and capacity of their storage to the appropriate scale necessary. With BeeGFS it is possible to start with a small storage cluster and grow into an enterprise wide, multi-Petabyte system (comprising hundreds or thousands of nodes) in the most seamless way possible.

BeeGFS General Architecture

BeeGFS utilizes metadata servers to coordinate file placement and striping among storage servers, informing the clients about file details when necessary. When users access their file contents, they directly contact the storage servers to run workloads and communicate with multiple servers simultaneously, giving applications truly parallel access to file data. Metadata is distributed across multiple servers to keep access latency at a minimum. Because BeeGFS is a software-defined storage solution, applications don't need to be rewritten or customized to be run on BeeGFS. It's structure is truly built for ultimate user convenience and flexibility. Users experience increased productivity by the fast data retrieval that BeeGFS allows for.

Depending on how users are accessing the BeeGFS file system, a gateway node may server as an intermediary between clients and the file system. The gateway node will allow both Windows and Linux users simple access to the BeeGFS file system.



Architectural Diagram of BeeGFS Cluster from Wikipedia

Comparing BeeGFS to other Parallel File Systems

When compared to other popular parallel file systems like Gluster and Lustre, the performance of BeeGFS in terms of simplicity and functionality is unparalleled. From our experience, BeeGFS offers the best overall mix of features, performance, scalability, and ease of deployment and management. Although alternatives each offer other of their own benefits in terms of simplicity (Gluster) or performance/scalability (Lustre), it's BeeGFS that allows us to deploy turnkey storage solutions for our customers without sacrificing either of the aforementioned. Note: Though we tend to favorite BeeGFS, we are able to utilize other parallel file systems if desired.

BeeGFS Benchmarks

To really see the power of BeeGFS in action, we ran tests using FIO (flexible IO) benchmark software. We configured a small 10 node HPC Cluster writing to a 9 node flash storage node based BeeGFS cluster. The results are very impressive. With over 50 GB/sec sustained reads & writes and over 10Mil IOPS, BeeGFS can certainly deliver the type of data transfer rates necessary for the most demanding HPC environments. As expected, the first machine run slightly slower than others, an unavoidable event due to the additional overhead incurred as the management node. Overall, performance benchmarks were above and beyond.

Over 50GB/s

WRITE: bw=3248MiB/s (3406MB/s), 3248MiB/s-3248MiB/s (3406MB/s-3406MB/s), io=3169GiB (3402GB), run=999002-999002msec
WRITE: bw=6837MiB/s (7169MB/s), 6837MiB/s-6837MiB/s (7169MB/s-7169MB/s), io=6670GiB (7162GB), run=999001-999001msec
WRITE: bw=6334MiB/s (6642MB/s), 6334MiB/s-6334MiB/s (6642MB/s-6642MB/s), io=6180GiB (6635GB), run=999001-999001msec
WRITE: bw=6411MiB/s (6723MB/s), 6411MiB/s-6411MiB/s (6723MB/s-6723MB/s), io=6255GiB (6716GB), run=999003-999003msec
WRITE: bw=6146MiB/s (6445MB/s), 6146MiB/s-6146MiB/s (6445MB/s-6445MB/s), io=5996GiB (6438GB), run=999002-999002msec
WRITE: bw=6631MiB/s (6953MB/s), 6631MiB/s-6631MiB/s (6953MB/s-6953MB/s), io=6469GiB (6946GB), run=999001-999001msec
WRITE: bw=7093MiB/s (7438MB/s), 7093MiB/s-7093MiB/s (7438MB/s-7438MB/s), io=6920GiB (7430GB), run=999002-999002msec
WRITE: bw=5547MiB/s (5816MB/s), 5547MiB/s-5547MiB/s (5816MB/s-5816MB/s), io=5412GiB (5811GB), run=999002-999002msec
WRITE: bw=6844MiB/s (7176MB/s), 6844MiB/s-6844MiB/s (7176MB/s-7176MB/s), io=6676GiB (7169GB), run=999001-999001msec
WRITE: bw=5662MiB/s (5937MB/s), 5662MiB/s-5662MiB/s (5937MB/s-5937MB/s), io=5524GiB (5931GB), run=999002-999002msec

Over 10M IOPs

Write: IOPS=831k, BW=3246MiB/s (3404MB/s)(3167GiB/999002msec)
Write: IOPS=1226k, BW=4789MiB/s (5021MB/s)(4672GiB/999002msec)
Write: IOPS=1170k, BW=4569MiB/s (4791MB/s)(4457GiB/999001msec)
Write: IOPS=1201k, BW=4691MiB/s (4918MB/s)(4576GiB/999001msec)
Write: IOPS=1198k, BW=4682MiB/s (4909MB/s)(4567GiB/999001msec)
Write: IOPS=1200k, BW=4688MiB/s (4915MB/s)(4573GiB/999001msec)
Write: IOPS=1250k, BW=4883MiB/s (5120MB/s)(4764GiB/999001msec)
Write: IOPS=1169k, BW=4567MiB/s (4789MB/s)(4455GiB/999001msec)
Write: IOPS=1231k, BW=4810MiB/s (5044MB/s)(4693GiB/999001msec)
Write: IOPS=1133k, BW=4426MiB/s (4641MB/s)(4318GiB/999001msec)

Creating Our Parallax Storage Clusters with BeeGFS

PSSC Labs Parallax Storage Clusters come application-optimized with popular parallel file systems, including BeeGFS. Built for maximum IOPS performance with the ability to scale out with your unique capacity needs, no AI and HPC workload is a match. Our Parallax Storage Cluster systems have been deployed at universities, government agencies, and commercial organizations around the country, including for Villanova, Cal Poly San Luis Obispo, the U.S. Navy, U.S. Army and many more. Like our entire product line, our Parallax Storage Cluster is also delivered production-ready, so you experience no down time from the time of delivery. In addition, Parallax Storage Clusters are typically a fraction of the cost of proprietary offerings from traditional storage vendors and do not require expensive licensing and support contracts.



The Parallax Storage Cluster will always be built exactly to your specifications, particularly regarding metadata nodes, data nodes and gateway nodes. Our metadata nodes are small, lightweight boxes comprised of SSDs or NVMEs, designed for fast data seek times. Data nodes are larger storage spaces that we typically configure with RAID 6 to ensure the protection of user data. Additionally, we can also add gateway nodes, perfect for users that will be accessing their system with Linux or Windows operating systems. Our solutions are truly customizable from the ground up, so each component even beyond those mentioned above can be designed exactly for your needs. To learn more about the Parallax Storage Cluster, visit us here: <https://pssclabs.com/products/storage-cluster/>