Unlocking High-Performance Storage for Next-Gen Artificial Intelligence Workloads with BeeGFS

AI Solution Brief

Artificial Intelligence (AI) workloads demand storage systems that can keep pace with rapid data growth, complex compute requirements, and real-time analytics. BeeGFS is a SuperPOD* certified solution, validating its capabilities to deliver the highest performance, scalable, and flexible solutions tailored for AI environments.

Key Benefits

BeeGFS delivers exceptional performance and scalability by distributing both data and metadata across multiple servers, enabling massive parallelism that accelerates training and inference for large-scale AI models. Its support for cutting-edge technologies such as GPUDirect and CSI drives ensuring maximum performance across heterogeneous AI environments. The system's modular design and intuitive management tools simplify deployment, scaling, and maintenance, allowing data scientists and engineers to focus on innovation instead of infrastructure.

BeeGFS also integrates seamlessly with leading AI frameworks and container orchestration platforms, streamlining workflows and speeding time-to-insight. Built with a future-ready architecture, it continues to evolve with advanced data management capabilities and multi-site collaboration features, empowering organizations to meet the growing demands of AI with confidence.

For enterprises and research organizations aiming to accelerate AI innovation, BeeGFS offers a proven, high-performance storage foundation that drives faster model development, efficient resource utilization, and scalable growth making BeeGFS the ideal data platform for your success.

*NVIDIA DGX SuperPOD with NetApp BeeGFS building blocks

High Performance

- Parallel I/O for massive throughput
- Low latency access to small and large files
- Optimized for HPC workloads

Robust

- High availability with BeeGFS on HA
- Fault-tolerant services using Pacemaker/Corosync
- Resilient under heavy metadata and I/O load

Scalable Architecture

- Scale metadata and storage nodes independently
- Seamless scale-out as data grows
- No single point of bottle neck

Ease of Use

- Fast deployment & management tools
- Monitoring via Grafana/ Prometheus
- Integration via Slurm, Kubernetes, Singularity

Data Management

- BeeOND for burst buffer on compute nodes
- Temporary, job-specific scratch space
- Fine-grained control with QoS and striping

Architecture Agnostic

- All kinds of Linux distributions
- Multiple network technologies
- Multitude of CPU/GPU architectures
- All types of storage technologies

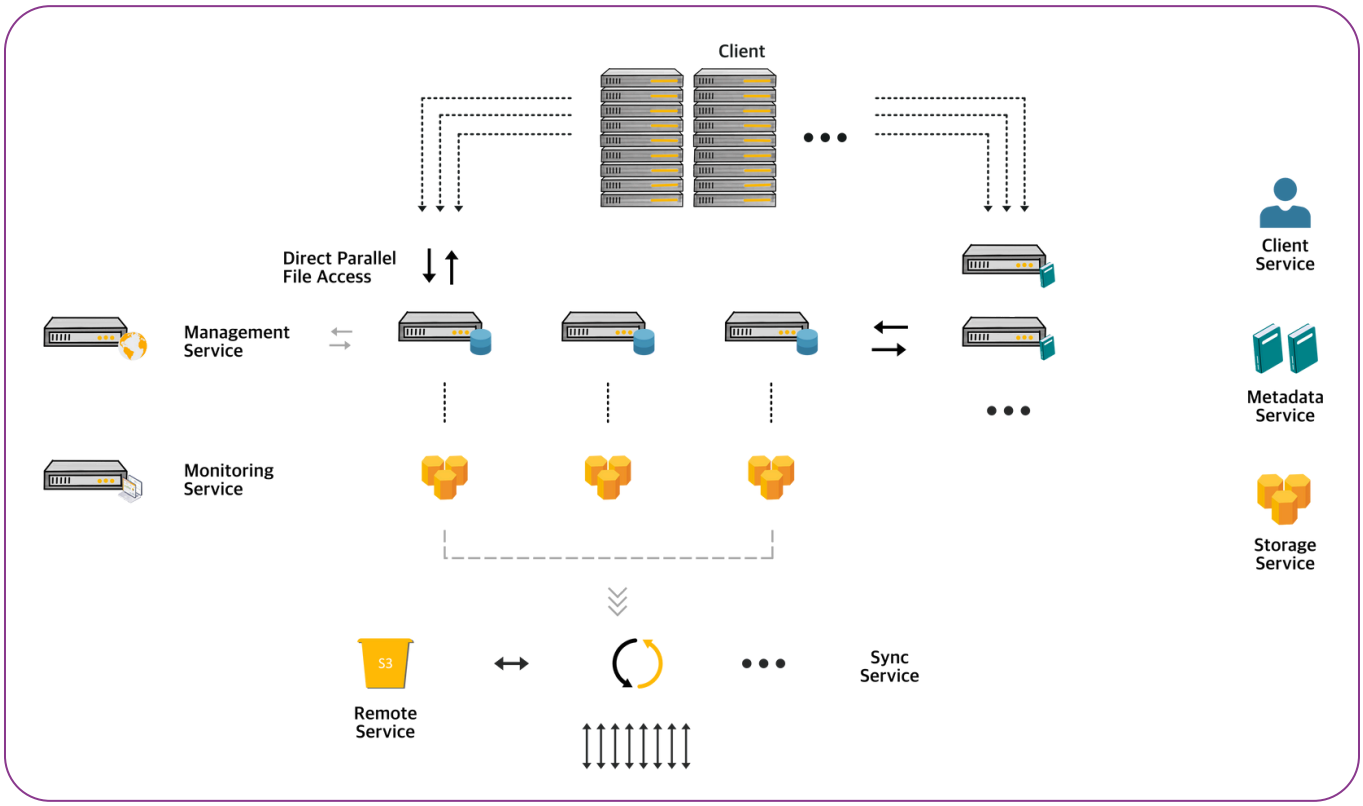
Eliminating I/O Bottlenecks in HPC/AI Workloads

I/O Throughput & Bandwidth	Stripes data across multiple storage servers to boost throughput. Enables low-latency, high-speed transfers using RDMA and user-space I/O, minimizing CPU load.
Metadata Performance	Eliminates metadata bottlenecks by distributing operations across nodes—ideal for small file workloads.
Scalability	Add servers or disks to scale linearly. No downtime, No disruption.
Concurrency & Contention	Handles many clients accessing the same data simultaneously, with no single point of contention.
Latency	Bypasses kernel and enables direct memory transfers, reducing latency and CPU overhead.
Cost-Effectiveness & Management	Runs on standard Linux servers with HDDs, SSDs, or NVMe, flexible and cost-effective. BeeOND: Creates fast, temporary file systems on local storage for burst-buffering HPC jobs.

Choose BeeGFS if you value:

- Fast, scalable storage that is simple to deploy and manage
- A trusted technology used in leading HPC sites worldwide
- High-performance for both large and small I/O workloads
- The ability to scale flexibly without vendor lock-in

The BeeGFS Architecture



Explore the BeeGFS Hive of Possibilities

ThinkParQ offers BeeGFS as a self-supported Community Edition and a fully supported Enterprise Edition with additional features and functionalities.

Features	BeeGFS Community	BeeGFS Hive Enterprise
Scratch File System	✓	✓
Access Control Lists	✓	✓
Monitoring Tool	✓	✓
Community Mailing List	✓	✓
Commercial License Key		✓
High Availability		✓
Mirroring		✓
Quota Enforcement		✓
Storage Pools		✓
Remote Storage Targets		✓
Copy		✓
Watch		✓
Index		✓

www.beegfs.io