1

BeeGFS – THE FILE SYSTEM FOR BIG DATA AND AI

The success of current AI technologies such as neural networks is based on the increased power of today's processors – mostly GPU´s – but, above all, on the availability of very large amounts of data. For example, new medical devices, autonomous vehicles, and genome analyses supply ever more fine resolution data in quick succession forming the basis for future AI solutions. Developed at ITWM and distributed by ThinkparQ, the parallel file system BeeGFS (also known as Fraunhofer Parallel File System – FhGFS) helps in mastering the large data volumes with a very flexible software solution.

BeeGFS is a parallel file system where storage capacity as well as read and write speeds grow linearly with the number of linked storage units. As a pure software solution, it can be flexibly installed both on existing hardware and on the latest, superfast flash-memory systems. In addition to very good scalability, our system development team attaches great importance on easy handling and a high degree of flexibility for a variety of potential use cases.

BeeGFS on NVMe

Training deep neural networks (Deep Learning) demands that existing data be provided several times very quickly to the computing units. Most external storage systems are hardly suitable for this task, so the data is cached directly to the computer servers on fast local systems (NVMe). Since these have relatively small capacities, the need arises for data to be distributed on several units in parallel.

The BeeGFS software system is specially optimized for high speed requirements even with a large number of files and this ability is its biggest strength. BeeGFS can be installed directly on the computer servers and is scalable to high I/O rate of 1 TByte/sec and more. Japanese AI researchers were convinced: BeeGFS is now successfully deployed on the two major Japanese AI systems TSUBAME 3.0 (HPE) and AI Bridging Cloud Infrastructure (ABCI, Fujitsu).

Open-source license

The software is distributed with an open-source license and source files are provided on the BeeGFS website. A spin-off of Fraunhofer ITWM, ThinkparQ, supplies worldwide commercial support for BeeGFS and manages further development from a customer perspective. The joint development team also successfully applies its extensive knowledge in several EU funded projects that focus on the use of BeeGFS on future Exascale computing systems.

1   *BeeGFS architekture*