



## BEEGFS – DAS DATEISYSTEM FÜR BIG DATA UND KI

Der Erfolg aktueller KI-Technologien wie neuronaler Netze basiert auf der gestiegenen Rechenleistung heutiger Prozessoren – meist GPUs – aber vor allem auf der Verfügbarkeit sehr großer Datenmengen. Neue medizinische Geräte, autonome Fahrzeuge und Genomanalysen liefern immer größere und feiner aufgelöste Daten in schneller Folge und damit die Basis für zukünftige KI-Lösungen. Das von uns entwickelte und von ThinkparQ vermarktete parallele Dateisystem BeeGFS hilft dabei, mit einer sehr flexibel einsetzbaren Software-Lösung der Datenmengen Herr zu werden.

1 BeeGFS-Architektur

BeeGFS ist ein paralleles Dateisystem, bei dem sowohl die Speicherkapazität als auch die Schreib- und Lesegeschwindigkeit mit der Anzahl der angeschlossenen Storage-Einheiten linear anwächst. Es ist eine reine Software-Lösung, die flexibel sowohl auf existierender Hardware als auch auf den neuesten superschnellen Flash-Speichersystemen installiert werden kann. Neben der sehr guten Skalierbarkeit des Systems legt unser Entwicklerteam großen Wert auf eine unkomplizierte Handhabung und ein hohes Maß an Flexibilität bei den potentiellen Einsatzszenarien.

### Zwischenspeicherung entfällt

Beim Training tiefer neuronaler Netze (Deep Learning) müssen darüber hinaus die vorhandenen Daten mehrfach sehr schnell den Compute-Einheiten zur Verfügung gestellt werden. Dafür sind die meisten externen Speichersysteme kaum geeignet und die Daten werden direkt in den Computerservern auf schnellen lokalen Systemen (NVMe) zwischengespeichert. Da deren Kapazitäten recht klein sind, ergibt sich hier auch die Notwendigkeit, die Daten parallel auf mehrere Einheiten zu verteilen. Das BeeGFS-Softwaresystem ist speziell für hohe Geschwindigkeitsanforderungen auch bei sehr vielen Dateien optimiert und kann hier seine Stärken ausspielen. So kann BeeGFS direkt auf den Computerservern installiert werden und skaliert bis zu hohen I/O-Geschwindigkeiten von 1 TByte/sec und mehr. Dies hat auch japanische KI-Forscher überzeugt: BeeGFS wird jetzt auf den beiden großen japanischen KI-Systemen TSUBAME 3,0 (HPE) und AI Bridging Cloud Infrastructure (ABCI, Fujitsu) erfolgreich eingesetzt.

### Open-Source-Lizenz

Die Software wird unter einer Open-Source-Lizenz vertrieben; die Quelldateien können auch von der BeeGFS Website bezogen werden. Unser Spin-off ThinkparQ bietet weltweit kommerziellen Support für BeeGFS an und steuert die weitere Entwicklung aus Kundensicht. Das gemeinsame Entwicklerteam bringt sein umfangreiches Wissen auch in mehrere EU-Projekte ein, in denen es um die Nutzung von BeeGFS auf zukünftigen Exascale-Rechnersystemen geht.

