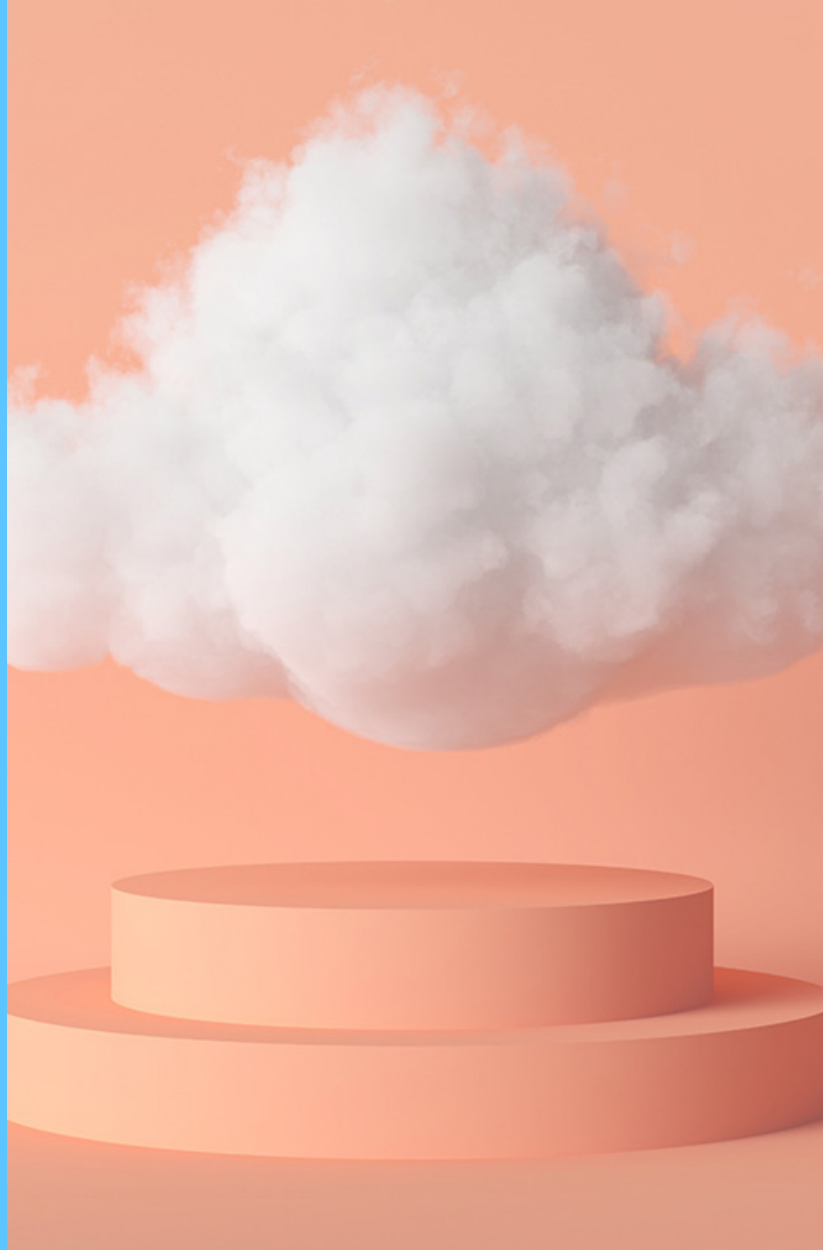
SOLUTION BRIEF

NetApp HPC cloud-connected storage

Connecting service providers and users at the network edge



NetApp

The Challenge

The business of big data is growing fast. IDC says that by 2025, worldwide data will grow 61% to 175 zettabytes, with as much of that data residing in the cloud as in on-premises data centers.

Big data has brought a new paradigm to data architecture. To keep pace with today's compute power, modern high-performance computing (HPC) applications require sophisticated and expensive hardware accelerators. To power their on-demand cloud computing platforms, hyperscalers like Amazon Web Services and Microsoft Azure use different types of accelerators, such as GPUs, FPGAs, and ASICs.

And as supercomputing continues to grow and to evolve, the need for new infrastructures that make data quickly accessible to HPC applications becomes more critical. But with lean teams and limited resources, companies are challenged to find ways to move out of their aging infrastructures and to drive innovation without having to buy massive compute farms.

The Solution

To meet these challenges, NetApp® HPC cloud-connected storage introduces a colocation and hyperscale solution. Your customers get everything that they need to store, to manage, and to deploy their data next to the cloud, including the building, cooling, power, storage, and network elements.

HPC cloud-connected storage is a fully managed parallel file system as a file storage service. This service moves HPC workloads from isolated data centers to colocated data centers by delivering high-throughput and critical file services on demand.

To handle the performance requirements of the most demanding workflows, HPC cloud-connected storage uses BeeGFS as its parallel file system. BeeGFS enables companies to streamline I/O patterns for small and large files, providing direct access to data faster.

Key benefits

- Easy deployment and administration of a customer's data center
- Designed for scalable I/O-intensive data storage
- Building-block approach to growth
- Connects to the compute resources of any hyperscaler of choice
- Option to lease compute resources in a colocation data center

The combination of HPC cloud-connected storage with BeeGFS provides a powerful platform for easy deployment and administration of a customer's data center infrastructure. NetApp and its partners set up and manage the underlying infrastructure, so customers can focus on application development and other priorities.

The result is a solid and superior storage infrastructure that connects to the compute resources of any hyperscaler—without sacrificing control of a customer's data. Customers are not locked in to any one vendor, so they can switch cloud providers at any time and avoid costly data migrations.

This infrastructure-ready solution is colocated and multitenant to deliver scalability and agility in an integrated, optimized, and cost-effective package—allowing you to scale customer environments to their specific needs.

Feature Overview

HPC cloud-connected storage uses the BeeGFS parallel file system in combination with NetApp storage hardware, which is racked behind network infrastructure in a colocated data center. This setup provides on-demand raw data access to applications running in the cloud or on the premises in the colocated data center facility.

Colocation lease option. For businesses that must meet compliance regulations such as HIPAA or PCI or that need submillisecond latency for compute, NetApp offers the option to lease compute resources in its colocation data center.

Fully managed service. Together with our partners, we manage capacity growth, perform software upgrades, and handle any failed components so that customers can focus on their business and application needs instead of storage maintenance tasks.

Nonstop reliability. Our extensive partner ecosystem helps us validate configurability, stability, interoperability, and reliability. Customers can be certain that the HPC cloud-connected storage service will deliver the 24/7 availability that their operations require.

Elastic storage capacity. Our modular solution design offers a granular, building-block approach to growth. You can scale your customer systems seamlessly by using an on-site buffer to scale up on demand, with customers paying only for the capacity that they consume.

Integration with BeeGFS. HPC cloud-connected storage integrates with BeeGFS to provide a powerful set of tools to easily add compute and storage to a meet your customers' performance requirements. This integration offers:

- Highly available infrastructure that implements industry-leading data protection techniques and best practices
- Open source and free to use (basic) features with very lightweight metadata and management services
- An easy-to-use wizard to automate the process of setting up and configuring BeeGFS in the HPC cloud-connected storage environment
- Metrics, health checks, and graphical administration and monitoring to confirm that BeeGFS is working properly
- File handling of both large and small files simultaneously with little impact on performance

Under the Hood

HPC cloud-connected storage configured with BeeGFS is designed for scalable I/O-intensive data storage.

BeeGFS is optimized to enable you to:

- Distribute data across any number of storage servers.
- Get high streaming throughput and high IOPS.
- Gain parallel data access from all compute nodes.

Simplicity. Simplicity is among the most attractive features of the BeeGFS parallel file system. Because BeeGFS requires no kernel changes and runs in user space, you can set up the BeeGFS system in minutes.

Designed to scale. BeeGFS supports distributed file contents with flexible striping across storage servers on a per-file or per-directory basis and supports distributed metadata. By increasing the number of servers and disks in your system, you can scale performance and capacity to the level that a customer needs.

For example, if you have the necessary GPUs and want to process up to 90,000 images per second, you might find that acquiring just one HPC cloud-connected storage stack is enough to support that task. A parallel file system such as BeeGFS also enables a user or a group of users to run more than one job simultaneously by separating BeeGFS clients to run different jobs.

The Right Fit

HPC cloud-connected storage provides shared persistent storage with high throughput and low latency to meet the demands of large databases and HPC applications.

It is optimal for workloads such as machine learning, genome sequencing, video rendering and transcoding, electronic design automation, financial simulation models, oil and gas 3D modeling and fluid dynamics.

Summary

The handling of big data is very complex. It requires new, innovative, and scalable technology to access, to host, and to analytically process the vast amount of data that is stored in the cloud and in data centers.

NetApp HPC cloud-connected storage combined with BeeGFS delivers a cost-effective and easy-to-scale solution to meet the demands of HPC applications and workloads. With HPC cloud-connected storage, the right data is quickly accessible. And when data is brought together quickly, customers can make better, smarter, real-time, data-driven decisions that can improve the way that they handle operations and compete in the marketplace.

About NetApp

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services and applications to the right people—anytime, anywhere. www.netapp.com

NetApp