



High-Performance File Services for AWS

NetApp Cloud Volumes Service Accelerates I/O-Intensive Workloads

Cloud Volumes Service for AWS: High Performance Use Cases

Cloud Volumes Service delivers great performance and availability for applications that demand high-bandwidth, low-latency shared access including:

- Databases and other high-I/O apps
- Analytics
- Application Dev/Test
- Media Libraries

WHY RUN HIGH-PERFORMANCE FILE WORKLOADS IN AWS?

High performance computing (HPC) users often want to run workloads in the cloud for exactly the same reasons as more mainstream cloud users:

- Availability of resources whenever they are needed
- Access to the latest CPUs, GPUs, and other hardware
- Ability to use services like statistics, analytics, and machine learning
- Budget limitations that put owning the equipment out of reach

Total [cloud spending on high performance computing rose 44% from 2016 to 2017](#), growing much faster than the HPC market as a whole.

For the purposes of this eBook, we're not concerned with the ability to run grand challenges in the cloud. Rather, we'll focus on the wide variety of "high-performance workloads" that require more compute resources and greater shared I/O bandwidth than the typical enterprise cloud application. Examples include genomics, electronic design automation (EDA), seismic data processing, media and entertainment, high-performance databases, and even large software builds.

For these workloads, there are a few challenges that can be a serious impediment to anyone wanting to move them to the cloud.

- **Shared I/O bandwidth.** Typical cloud file sharing services may not meet the performance requirements of these workloads, while parallel file system solutions based on Lustre or GPFS can be complicated and expensive.
- **Management complexity.** Most cloud users just want to run their applications. Time spent architecting or managing a storage solution in the cloud is time that could be better used elsewhere. Solutions that are overly complex to configure and manage can be too much trouble.

- **Data movement.** You may be dealing with significant amounts of data that needs to be moved into the cloud efficiently for processing. Data migration in and out of the cloud can be complicated, time consuming, and expensive.
- **Flexibility.** Your needs for capacity and performance can change from week to week or even hour to hour. Parallel file system solutions may lock you into a configuration.



[NetApp Cloud Volumes Service for AWS](#) addresses these challenges, making it simple and affordable for your organization to run high-performance workloads—and address all your NFS and SMB file sharing needs—using a convenient service in the AWS cloud.

This eBook explores the advantages of Cloud Volumes Service for AWS and showcases how several NetApp customers are using the service to meet the needs of applications with high I/O requirements.

INTRODUCING NETAPP CLOUD VOLUMES SERVICE FOR AWS

If your organization wants to run high-performance workloads in the cloud, NetApp Cloud Volumes Service for AWS is the ideal solution to meet your data storage needs. Cloud Volumes is a fully managed cloud service that delivers extreme performance to meet demanding shared file access requirements.

Cloud Volumes Service for AWS delivers managed file services for NFS and SMB protocols with dual-protocol access to the same shared files when needed. This means that multiple Linux compute instances can access data using NFS—while Windows users can conveniently access data and results files.

The screenshot displays the 'Create volume' configuration interface. It includes the following fields and options:

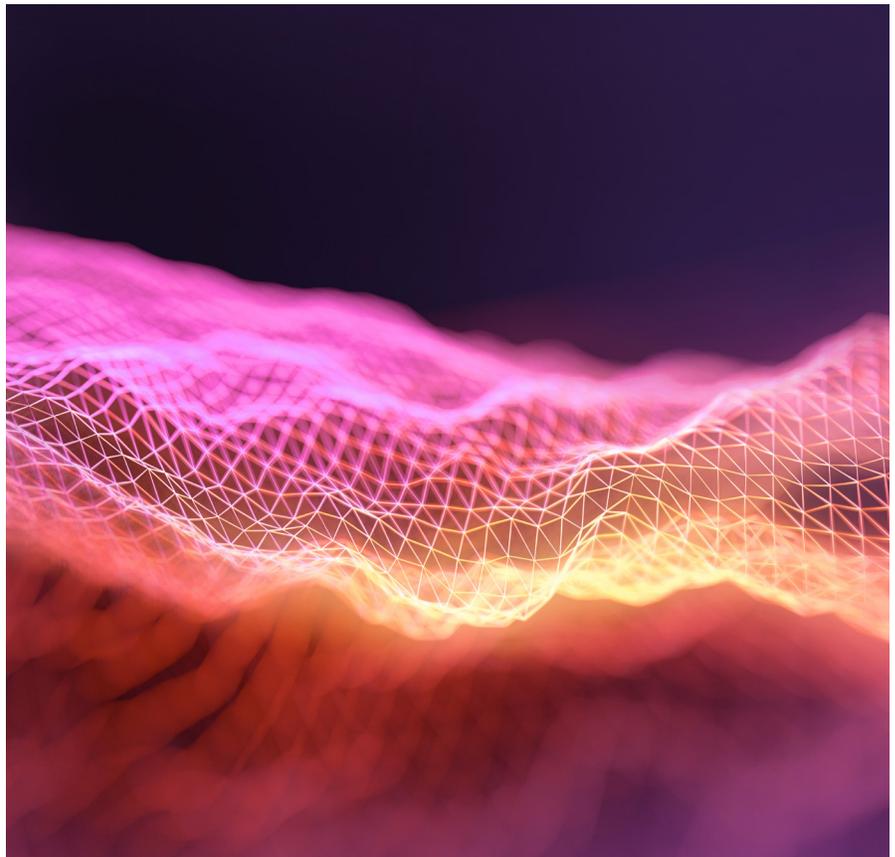
- Name:** HPC
- Region:** us-east-1
- Timezone:** Any
- Volume path:** determined-nauseous-tharp
- Service level:** Extreme
- Allocated capacity:** 10000 GB
- Security style:** UNIX
- Tags:** (empty field)
- Export policy:**
 - Buttons: + Add export policy rule
 - Table:

Rule index	Allowed clients	Access	Protocol
Rule-1	172.0.0.0/8	Read & Write	NFSv3

Cloud Volumes Service for AWS provides a number of advantages over other cloud storage solutions:



- **High performance.** High throughput and low latency file access meets the demands of the largest databases, genomics, and other I/O intensive applications such as software builds, video rendering and transcoding, and financial modeling.
- **Lower cost.** Compared to Lustre and GPFS-based cloud solutions, Cloud Volumes Service for AWS is significantly less expensive to deploy and manage. Because you can adjust performance levels on the fly, it's possible to optimize performance for peak periods without overpaying during off-peak.
- **Fully managed.** NetApp configures and manages your infrastructure, so you don't have to.
- **Easy to deploy.** You can deploy a 100TB cloud volume in just eight seconds and begin working immediately.
- **Enterprise features.** Cloud Volumes Service provides enterprise features like Snapshots, backup, and synchronization that facilitate data management.
- **Guaranteed SLAs.** NetApp service level agreements for performance, availability, and durability allow you to be certain your cloud operations satisfy workload needs.



Reasons to Choose Cloud Volumes Service for AWS

Fully managed. Eliminate the challenge of configuring and managing high-performance cloud storage

Compatibility. Shared file access across Windows and Linux

Integration. Full compatibility with Active Directory

Data Protection. Efficient Snapshots prevent data loss or corruption

Easy Migration. Move data to/ from on-premises or other repositories via Cloud Sync

High availability. Ensure data remains available with guaranteed SLAs

Automation. Schedule tasks easily or take advantage of full REST APIs

WHY CHOOSE NETAPP FOR CLOUD DATA SERVICES?

If you're unfamiliar with NetApp, you may ask why you should trust a company you've never heard of to deliver high-performance cloud services.

Founded in 1992, NetApp pioneered the idea of network attached storage (NAS) and was the first to offer high-performance NAS appliances that greatly simplified the configuration and management of shared file services.

NetApp was instrumental in developing many of the storage efficiency and data protection technologies we take for granted today, including Snapshots, cloning, and much more.

Recognizing early that the public cloud would reshape the way that enterprises approach IT and engage with customers, NetApp created the Data Fabric, an architecture and set of data services that provide consistent capabilities spanning on-premises and cloud environments.

NetApp Cloud Volumes Service for AWS is a key element of this Data Fabric vision. NetApp understands shared file services and data management so you don't have to; you can focus instead on innovation that moves your business forward.

CLOUD VOLUMES SERVICE FOR AWS: THE NETAPP DIFFERENCE

The advanced data management that NetApp Cloud Volumes Service for AWS delivers for high-performance workloads is rocket fuel for innovation. Cloud Volumes Service matches your need for performance and scale with the flexibility to automatically grow or shrink capacity and performance as necessary.

With NetApp Cloud Volumes Service, your teams can get access to robust file services with just a few mouse clicks, unlocking the capabilities they need to be instantly productive.

NetApp Cloud Volumes for AWS Specifications	
Performance Tiers	Standard 16MB / second throughput per allocated TB Premium 64MB / second throughput per allocated TB Extreme 128MB / second throughput per allocated TB
Advanced Data Management	<ul style="list-style-type: none">• Cloud Sync• Snapshot copies• Instant Copy• Backup• Encryption
Protocol Support	<ul style="list-style-type: none">• SMB: 2.1, 3.0, 3.1.1• NFS: NFSv3

NetApp Cloud Volumes Service for AWS is fully managed and optimized to deliver the performance, availability, scalability, security, and data management features you need. Choose the performance level you need today and switch on the fly as requirements change.

"We always had some time outs or some file failures, but when we tested this using NetApp Cloud Volumes, it actually finished in less than an hour. That was a great breakthrough for us."

- WuXi NextCODE Chief Informatics Officer Hakon Gudbjartsson, PhD

CLOUD VOLUMES SERVICE FOR AWS: OPTIMIZED PERFORMANCE AND COST

If your organization is running parallel file systems like Lustre or GPFS on premises, you know that ensuring I/O performance can be a continuing effort. Corresponding services in the cloud are expensive and configurations can be rigid. NetApp Cloud Volumes Service for AWS is less expensive than parallel file services and offers better performance than competing shared file system solutions in the cloud.

With Cloud Volumes Service for AWS, performance concerns are less of an issue. You choose the performance tier you need up front and switch between three different performance levels dynamically without having to move data or reconfigure everything from scratch, enabling you to easily optimize your performance and cost.

Applications that are best suited to Cloud Volumes Service fit the following profile:

- Per instance bandwidth needs are no more than 4.5Gbps full duplex (higher limits are possible with SLES 15 or the Linux 5.3 Kernel).
- Your workload scales horizontally across multiple compute instances
- You need no more than 50M files or directories per volume (if higher inode counts per volume are required, please open a support ticket through the Cloud Volumes Service Portal).

Guaranteed IOPS and Bandwidth

Cloud Volumes Service for AWS provides different performance levels, allowing you to choose the right performance for each volume to meet the needs of your applications. Assuming an I/O size of 4KB, Cloud Volumes Service delivers the performance levels (IOPS and bandwidth) shown below:

NetApp Cloud Volumes for AWS Performance Tiers		
	IOPS per TB	Throughput per TB
Standard	4,000	16MB/sec
Premium	16,000	64MB/sec
Extreme	32,000	128MB/sec

NetApp has carefully benchmarked the performance of Cloud Volumes Service for a variety of workloads. These benchmarks can help you understand whether or not the service will meet your performance needs prior to provisioning.

Note that AWS architectural limits exist as such:

- Writes per instance to 5Gibps (600MiB/s)
- Reads per network flow to 5Gibs (600MiB/s)
 - o A network flow is defined as the combination of (src ip and port : destination ip and port)

Understanding Storage Performance

Different applications generate different I/O workloads. It's useful to think about the I/O characteristics of the workload or workloads you need to support in terms of three metrics.

IOPS

Transaction-oriented applications such as databases generate small, random reads and writes. This type of storage performance is measured in I/O Operations Per Second or IOPS.

Throughput

Applications such as data warehouses and video rely on sequential access to data in large blocks, resulting in a workload that's dramatically different from transaction-oriented applications. Throughput is a measure of the amount of data that can be moved in or out of storage and is usually reported in MB/sec or GB/sec.

Latency

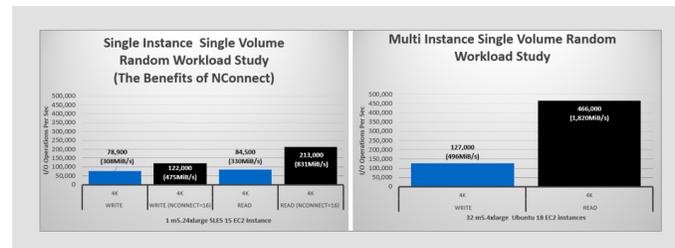
Latency is a measure of how long it takes to satisfy an I/O. It is an important metric for time-critical applications such as real-time trading and online transaction processing (OLTP), which are extremely latency sensitive. Interactive users notice and react to differences in latency.

Higher performance for your workloads is achieved by splitting a workload across multiple instances and or shifting to SLES 15 (or the Linux 5.3 kernel). Though currently available only in the SUSE Linux Enterprise Server (SLES15), support for the nconnect NFS mount option has entered General Availability in the Linux 5.3 kernel. We anticipate that this support will be backported into forthcoming RHEL distributions as well. As documented by suse.com, the nconnect mount option requests multiple transport connections between the client and the storage IP. With this development in place, the I/O from even a single mount point is parallelizable across up to 16 network flows compared to a single flow for a single storage endpoint without this option.

Please be aware: although [the documentation](#) (see [item 4.6.2](#) in [suse-sles15 release notes](#)) calls out nconnect as a workaround for NFSv4 performance regressions, the new capability in Cloud Volumes Service works very well with NFSv3.

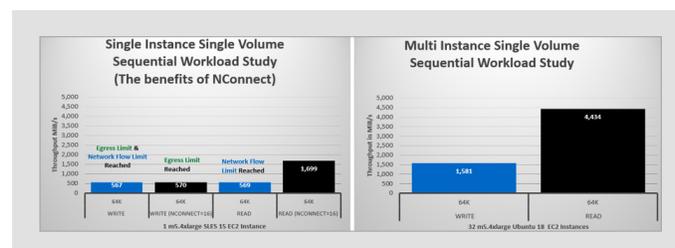
IOPS

NetApp tested random read and write performance of single and multiple clients against a single instance of Cloud Volumes Service for AWS. Depending upon client operating system and mount option (nconnect to be specific), maximum read performance for a single client is between ~85K and ~215K IOPS. Maximum observed performance for multiple clients against a single volume is over 460K IOPS.



Throughput

NetApp tested the sequential read and write performance of single and multiple clients against a single instance of Cloud Volumes Service for AWS. As above, maximum read performance for a single client is predicated upon client operating system/Linux Kernel and the use of the nconnect mount option. Without nconnect, clients are restricted to approximately 600MiB/s of NFS read bandwidth, with nconnect that AWS imposed restriction is bypassed and up to 1,700MiB of single client reads have been observed. The AWS imposed per instance egress limit of 5Gib/s (@600MiB/s) is still in effect. Maximum read performance for multiple clients against a single volume exceeded 4GB/s.



Sizing A Cloud Volumes Deployment

Accurate sizing to deliver the necessary bandwidth for your workload can be critical when it comes to delivering the fastest results and optimizing costs.

This video series will help you size a workload for Cloud Volumes Service to get the most from the service.

- **Part 1. Architecture**
- **Part 2. Performance**
- **Part 3. Oracle on AWS**



Hadoop Spark

NetApp also benchmarked the performance of Hadoop Spark. Spark is able to achieve an average throughput of 3.1GB/s against a single Cloud Volumes Service volume when run on 15 C5.9xlarge Amazon EC2 instances. Additional tests are planned wherein the higher limits seen above are anticipated.

The figure below illustrates that the total execution cost of a simulated Spark job with a 71TB working set size is actually more favorable using the CVS premium service than using the standard service. Increased cost of the premium service is more than offset by reduced runtime and reduced compute costs.



To see all the latest performance benchmarks for Cloud Volumes Service for AWS, visit: cloud.netapp.com/cloud-volumes-service/aws-benchmarks

"We love how NetApp has invested, leaning forward into the cloud, so that we can enable our customer's success"

–Michael Denari Sr.
Manager, Procore



Cloud Volumes Service for AWS: Scalability

Scaling is another significant challenge that comes with high-performance file systems. In addition to providing extreme levels of I/O performance, NetApp Cloud Volumes Service for AWS is also highly scalable and the only file system with guaranteed SLAs1.

As the number of active sessions increases, I/O performance increases to accommodate demand. Single applications can scale performance by parallelizing data access across multiple client sessions. The ability to scale up automatically without needing to tune infrastructure is a major advantage of Cloud Volumes Service for AWS.

If you exceed the purchased capacity in a Cloud Volume, you can continue to write data to the volume beyond the allocated capacity, however the performance of the volume will not increase until the allocated capacity is increased.

Cloud Volumes Service for AWS: Availability and Data Protection

Availability and data protection are also critical for high-performance workloads. Cloud Volumes Service is designed to deliver 99.99% data availability using built-in HA features including network failover and advanced data protection. Cloud Volumes Service for AWS is immune to Availability Zone failures and provides consistent performance in each region. Redundancy is implemented at every level. NetApp technologies protect against a variety of disk errors, improving data durability and integrity versus native AWS services.

Cloud Volumes Services for AWS provides instant protection for file shares of any size through the use of space-efficient Snapshots. A Snapshot copy takes only moments to create, and Snapshots remain completely stable and incur no performance overhead. You can store up to 255 Snapshot copies per volume. Snapshots can be run manually or on a set schedule.

[NetApp Cloud Backup Service](#) complements Cloud Volumes Service for AWS with backup to Amazon S3 object storage for long-term archive and recovery, reducing backup costs versus traditional methods.

Cloud Volumes Service for AWS: Migrating and Managing Data

Data management is a challenge in any environment. Cloud Volumes Service for AWS capabilities take the pain out of moving, organizing, protecting, and accessing data.

[Cloud Sync](#) is NetApp's service for rapid and secure data synchronization. Whether you need to transfer files between on-premises repositories and Cloud Volumes Service or between volumes in different regions, Cloud Sync does the job quickly and securely. Cloud Sync is integrated and included for free with Cloud Volumes Service for AWS.

A variety of use cases from development to analytics to AI require copies of existing datasets. Cloud Volumes Service Instant Copies solve this challenge. You can create new copies of datasets from Snapshots of a volume.

"Public archives for raw sequencing data have been doubling in size every 18 months. Leveraging these data requires researchers to use large-scale computational resources. Cloud computing, a model whereby users rent computers and storage from large data centres, is a solution that is gaining traction in genomics research."

-Nature



Deploying Cloud Volumes Service for Genomics

The total amount of genomics data is large and growing rapidly. A 2015 journal article estimates that by 2025, genomics data will require more annual data storage than YouTube.

Scientists need the cloud to access the resources necessary to store and analyze genomics datasets of increasing scale. The cloud not only offers the elasticity to access CPU cores and storage on an as-needed basis, but it can also facilitate collaboration between dispersed researchers.

But the cloud is not a panacea. It can be difficult for scientists to create and manage the data repositories they need for their work. This was the exact challenge faced by WuXi NextCODE.

Leading Genomics Company Relies on NetApp Cloud Volumes Service

With the world's largest collection of human genome sequences, the team at WuXi NextCODE is at the forefront of data-driven precision medicine, enabling healthcare organizations on four continents to unlock the power of the genome to transform medical care.

This is no small task. According to Rob Brainin, the CEO of WuXi NextCODE, "It was only 15 years ago that the first whole human genome was sequenced. It took two decades and cost almost \$3 billion to do. The cost has come down radically to under a \$1,000 and a few days, but that's generated a tsunami of data." And even though all humans share the same DNA, the sequencing of it is 100% unique to the individual. If you and a friend were to compare your DNA, you would find about five million differences."

"The challenge is to figure out the differences or mutations that are important—which ones are the causes of rare diseases, chronic conditions, and cancers, and how to individualize patient treatment," Dr. Hakon Gudbjartsson, CIO of WuXi NextCODE explained.

At the heart of the WuXi NextCODE platform is the genomic relational database, the only data architecture designed to optimize the use of massive genomic data. By leveraging [NetApp Cloud Data Services](#), the genome platform makes it possible to integrate data on the fly to deliver unprecedented computational efficiency.

"The benchmark for analyzing genomic data is accessing data from 100,000 individuals," Dr. Gudbjartsson said. "We always had timeouts or file failures. But when we tested this using the [NetApp Cloud Volumes Service](#), it actually finished in less than an hour. That was a great breakthrough for us."

WuXi NextCODE is a crucial part of the data-driven precision medicine equation. Its architecture underpins the preeminent genomics efforts on four continents and is the emerging global standard for organizing, mining, and sharing large-sequence datasets that are transforming medical care today.

And it's powered by NetApp.



Deploying Cloud Volumes Service for Media & Entertainment

In order to stay competitive, content creators in media and entertainment are under pressure to produce more original content and distribute it faster. Without the cloud it would be almost impossible to create, store, manage, and deliver huge quantities of digital content in a dynamic and cost-effective manner. Hybrid cloud deployments provide greater flexibility, allowing critical workloads like video editing and rendering to shift seamlessly in response to unforeseen resource needs.

Cloud computing enables media and entertainment companies to contain technology and storage costs to balance rising costs in other areas. By shifting from a CapEx model to an OpEx model, companies reduce their sunk costs. Using different tiers of cloud storage for hot vs. cold assets can reduce storage costs considerably.

Workloads such as rendering and transcoding, in particular, may require high-speed, shared access to the same files from large numbers of compute instances, making these workloads an ideal choice for NetApp Cloud Volumes Service for AWS.

Getting Started with Cloud Volumes Service

Cloud Volumes Service for AWS is an enterprise-grade solution for deploying high-performance file sharing to support a wide range of cloud uses cases, including genomics, electronic design automation (EDA), seismic data processing, media and entertainment, high-performance databases, and large software builds.

This fully managed solution lets you easily deploy scalable file shares, migrate data as needed, and get to work without delays. Flexible service levels allow you to optimize performance and cost to meet your application needs without breaking your budget.

Visit the [Cloud Volumes Service page on NetApp Cloud Central](#) for the latest information. Request a demo with a NetApp Cloud expert, sign up for the Cloud Volumes Service there, or go directly to the [AWS Marketplace](#) to get started.

© 2019 NetApp, Inc. All Rights Reserved. NETAPP, the NETAPP logo, and the marks listed at netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.

¹ Guaranteed SLAs so in the event any of the "Service" does not meet the Service Levels for performance and/or availability, you will be eligible to receive a "Service Credit" from NetApp for the future use of the Service.