WHITE PAPER

AWS: NetApp Cloud Volumes and SQL

NetApp®

Executive Summary 3

Introduction 3

Cloud Volumes Service: An Introduction 3

Custom NAS Solutions 3

Building a Database Platform with Amazon EC2 and Cloud Volumes Service 4

Deployment in AWS 4

Conclusion 5

Executive Summary

This white paper describes the benefits of using NetApp® Cloud Volumes Service in an Amazon Web Services (AWS) environment, and demonstrates how a hot standby disaster recovery environment can be created with AWS Auto Scaling Groups and AWS Elastic Load Balancing (ELB). The benefits of using NetApp Cloud Volumes Service for AWS for database storage management are discussed in detail.

Introduction

Relational database systems are an integral component of enterprise architectures and rely on structured, durable storage for an organization's data. Because database systems are primarily concerned with data storage, they are completely dependent on fast, reliable, and scalable storage environments. Moving from on-premises systems to AWS, new database platforms can be deployed using Amazon Relational Database Service (RDS), or built from the ground up using cloud-based compute and storage resources.

Amazon RDS is a fully managed database service that supports a variety of database servers, including MySQL, Oracle, and Microsoft SQL Server. Amazon RDS simplifies the setup and administration of database environments, allowing cloud architects to focus on the delivery of applications and services. However, these benefits come at the cost of relinquishing some control over how the database platform is managed, which may not be suitable in all cases.

Another option is to build a new database platform using Amazon Elastic Cloud Compute (EC2). Although this option involves a certain amount of complexity, database and system administrators retain full control over setting up the operating system of database server nodes, organizing disk storage, managing backup and restore, and much more. Having the power to control all aspects of database server deployment means that database environments can be tailored to the exact specifications of each use case. And once an EC2 instance has been created, it can easily be duplicated to deploy multiple instances.

Cloud Volumes Service is a fully-managed data management service from NetApp that delivers extreme I/O performance with built-in high availability and data protection that can be used seamlessly with AWS deployments. These features make Cloud Volumes Service the ideal storage environment for building new database platforms in the AWS cloud.

This paper describes Cloud Volumes Service in detail, and discusses how to use it along with Amazon EC2 to build enterprise-grade database platforms in AWS.

Cloud Volumes Service: An Introduction

NetApp Cloud Volumes Service is a high-performance platform-as-a-service (PaaS) solution for deploying highly available and scalable shared file systems in AWS. End users can create, manage, and pay for Cloud Volumes Service just as they would for any other service in AWS, such as Amazon S3 or Amazon EFS. Cloud Volumes Service manages all aspects of storage service delivery, automatically taking care of high availability, scalability, and I/O performance. You just create a new file system and start using it.

Cloud Volumes Service is designed to meet the most demanding I/O performance requirements, for both NFS and SMB file systems. The level of IOPS and I/O throughput of each file system can be controlled by configuring its service level, which is selected when the file system is created. Cloud Volumes Service also makes it possible to instantly modify the service level after a file system has been created. Three service levels are available: Standard, Premium, and Extreme.

Service Levels for Cloud Volumes Service

- Standard. Up to 1000 IOPS per TB (16k I/O) and 16MB/s of throughput per TB.
- Premium. Up to 4000 IOPS per TB (16k I/O) and 64MB/s of throughput per TB.
- Extreme. Up to 8000 IOPS per TB (16k I/O) and 128MB/s of throughput per TB.

NFS and SMB shared file systems allow thousands of clients to concurrently read from and write to the same data files. NetApp has capitalized on its industry-leading experience in building enterprise, on-premises NAS solutions to make Cloud Volumes Service a truly cloud-scale platform. The performance, reliability, and scalability of Cloud Volumes Service would be impossible to achieve with custom-built NAS services built on native-cloud compute and storage. Although a custom NAS solution may work initially, it presents several challenges to delivering an enterprise solution.

Custom NAS Solutions

Creating a reliable NAS solution in AWS hinges on the storage architect's ability to ensure high availability and data redundancy across availability zones. Block-level cloud storage services, such as Amazon EBS, store data in more than one location in an availability zone. However, extending the durability of data beyond an availability zone requires some form of data replication. If synchronous replication is not used, there is always the potential for data loss. Cloud Volumes Service solves this issue by using a grid of storage nodes, with built-in data redundancy. These nodes reside in a colocated data center that can be accessed by each availability zone, and are therefore immune to availability zone failures. NetApp also plans to release cross-region data protection in the near future.

In addition to high availability, it's also important to balance storage costs and I/O performance, which requires very careful management of underlying cloud storage resources. Allocating suitable amounts of storage for both hot and cold data can be difficult as requirements change and storage use increases. Moving data between these tiers, and reconfiguring the amount of storage used by each, must be done manually while minimizing downtime, which adds significant administrative

overhead. Cloud Volumes Service allows file systems to be grown instantly and at any time, simply by adjusting the quota. The service level for each file system, which determines the level of I/O performance it will deliver, can also be modified for an active file system without downtime or any adverse effects.

Building a NAS solution that supports advanced data management features such as NetApp Snapshot™ copies, clones, data synchronization, and variable service levels (QoS) requires an organization to have a great deal of in-house storage expertise. Cloud Volumes Service makes it easy to instantly create and manage read-only, point-in-time Cloud Volume Snapshot copies, which can also be used as the basis for creating new, writable Cloud Volume clones. Using the data synchronization features, data can be kept efficiently in sync between Cloud Volumes Service and on-premises systems or other cloud environments, without the considerable manual effort normally required.

Building a Database Platform with Amazon EC2 and Cloud Volumes Service

Database systems have traditionally used block-level devices for the storage of database data and log files. However, all major vendors now also support shared file services, such as NFS and SMB, including Oracle, MySQL, PostgreSQL, and Microsoft SQL Server. Shared file systems offer many benefits in terms of storage management and the ability to provision hot standby servers without the need for data replication.

Cloud Volumes Service allows users to increase or decrease their storage usage instantly and at any time. This is a major benefit for storage administrators, because databases are always growing in size, and this growth normally requires provisioning new disks for existing database servers. With Cloud Volumes Service, administrators just set a new quota size and additional storage capacity is available immediately.

One primary concern of all database administrators is to ensure that recoverable backups exist for all production database systems. Data errors are a frequent occurrence that can be rectified only through easy access to previous versions of the database. However, database backups for large databases can take a long time to create and require significant additional storage space. Cloud Volumes Service offers an ideal solution for point-in-time database recovery with its Snapshot feature, which can be used to instantly create a space-efficient copy of a database of any size.

Cloud Volumes Benefits for Databases

- Recoverable, space-efficient backups
- Increase and decrease storage rapidly and on the fly
- Mountable Snapshot copies
- Rapid clone volume creation

To create a consistent backup of a database system, NetApp recommends temporarily suspending database I/O operations and flushing buffers in memory to disk. This procedure ensures that all in-flight database data is captured in the Snapshot copy and prevents data corruption. Each database system uses its own specific commands and routines to quiesce database I/O. After the Snapshot copy is complete (instantly, with Cloud Volumes Service), normal database operations can be resumed.

Cloud Volumes Service provides immediate access to the data in each Snapshot copy as a read-only view of the database files at the point in time that the copy was created. These database files can be mounted by a database server to obtain read-only access to the information in the copy, and a selective restore can be performed to the production system if necessary. Cloud Volumes Service also supports performing a full, instant restore of a Snapshot copy back to the source cloud volume from which it was created.

Snapshot copies can also be used to quickly create new, writable clones of the original cloud volume from which the copy was taken. This capability offers huge benefits to DevOps engineers who need to provision test environments with up-to-date copies of production database systems. Cloud Volumes Service can be used to create multiple clones from the same Snapshot copy to facilitate parallel testing by multiple teams, as well as to support continuous integration and continuous delivery environments.

Deployment in AWS

Even though NFS is a shared file protocol, only one instance of a database server is normally connected to a set of database files at one time. However, because the file storage is abstracted from the compute resource, if a database server node fails, AWS can be used to simply spin up a new node to connect to the storage in Cloud Volumes Service and continue serving out database requests. This type of Amazon EC2 setup can be achieved by using Amazon EC2 Auto Scaling, which automatically maintains a fleet of Amazon EC2 instances. Using a fleet size of 1 and AWS Elastic Load Balancing ensures that at least and at most 1 database server instance is running and is always accessible over the network.

Multiple auto scaling groups can be used to scale database compute independently of the database storage layer by splitting the hosted databases across multiple database server instances. High availability is achieved in the same way. Scaling can be done while centrally managing all database storage in Cloud Volumes Service, which easily scales to accommodate hundreds or thousands of concurrent client connections.

Oracle Database can take full advantage of the scalability of Cloud Volumes Service without creating multiple database server instances. Using its Direct NFS feature, Oracle DB can create multiple parallel client connections to an NFS file service, which massively increases scalability and I/O throughput.

Deploying a Database on AWS

- Allocate database storage in Cloud Volumes Service
- Deploy database server to Amazon EC2
- Set up Amazon EC2 Auto Scaling and AWS Elastic Load Balancing

For horizontally scalable database platforms, such as those using NoSQL, Amazon EC2 Auto Scaling can be used to create a cluster of database compute nodes with all storage managed through Cloud Volumes Service. In case of a node failure, a new node can be created to reconnect to the previously failed node's database storage, significantly reducing the amount of time needed to rebalance the data on the new node from the rest of the cluster. If a single cloud volume is used for all data storage across the cluster, a single Snapshot copy can be used to restore a database to a point in time for all nodes.

To initialize a new cloud volume with data from an existing database system, Cloud Volumes Service can be used with NetApp Cloud Sync data synchronization features to migrate database files from on-premises systems or other cloud environments. This feature is very useful for getting a new database platform up and running quickly. After that, database replication mechanisms can be used to keep data synchronized.

Conclusion

Cloud Volumes Service is a fully managed NFS file service with a feature set that makes it highly suited to hosting and managing database data in the cloud. Used together with Amazon EC2, it's easy to create custom, highly available database environments, with the advantage of being able to manage and scale compute and storage independently.

One of the most compelling benefits of using Cloud Volumes Service for database storage is the ability to instantly create Snapshot copies of a database of any size, and instantly restore that database to the point in time when the copy was taken. An additional benefit is having very easy access to the data in the copies without needing to restore them, and being able to rapidly create writable clones of a database as it existed at the point in time when the Snapshot copy was created.

To find out more, or to try Cloud Volumes Service in AWS for yourself, visit us at NetApp Cloud Central.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations. For more information, visit www.netapp.com. #DataDriven

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

WP-7277-1018

NetApp®