High-availability pairs

Cloud Volumes ONTAP

NetApp
February 13, 2026

Table of Contents

High-availability pairs

Learn about Cloud Volumes ONTAP HA pairs in AWS

A Cloud Volumes ONTAP high-availability (HA) configuration provides nondisruptive operations and fault tolerance. In AWS, data is synchronously mirrored between the two nodes.

HA components

In AWS, Cloud Volumes ONTAP HA configurations include the following components:

- Two Cloud Volumes ONTAP nodes whose data is synchronously mirrored between each other.
- A mediator instance that provides a communication channel between the nodes to assist in storage takeover and giveback processes.

Mediator

Here are some key details about the mediator instance in AWS:

Instance type
t3-micro

Disks
Two st1 disks of 8 GiB and 4 GiB

Operating system
Debian 11

For Cloud Volumes ONTAP 9.10.0 and earlier, Debian 10 was installed on the mediator.

Upgrades
When you upgrade Cloud Volumes ONTAP, the NetApp Console also updates the mediator instance as needed.

Access to the instance
When you create a Cloud Volumes ONTAP HA pair from the Console, you're prompted to provide a key pair for the mediator instance. You can use that key pair for SSH access using the `admin` user.

Third-party agents
Third-party agents or VM extensions are not supported on the mediator instance.

Storage takeover and giveback

If a node goes down, the other node can serve data for its partner to provide continued data service. Clients can access the same data from the partner node because the data was synchronously mirrored to the partner.

After the node reboots, the partner must resync data before it can return the storage. The time that it takes to resync data depends on how much data was changed while the node was down.

Storage takeover, resync, and giveback are all automatic by default. No user action is required.

RPO and RTO

An HA configuration maintains high-availability of your data as follows:

- The recovery point objective (RPO) is 0 seconds.
  Your data is transactionally consistent with no data loss.
- The recovery time objective (RTO) is 120 seconds.
  In the event of an outage, data should be available in 120 seconds or less.

HA deployment models

You can ensure the high availability of your data by deploying an HA configuration across multiple availability zones (AZs) or in a single availability zone (AZ). You should review more details about each configuration to choose which best fits your needs.

Multiple availability zones

Deploying an HA configuration in multiple availability zones (AZs) ensures high availability of your data if a failure occurs with an AZ or an instance that runs a Cloud Volumes ONTAP node. You should understand how NAS IP addresses impact data access and storage failover.

NFS and CIFS data access

When an HA configuration is spread across multiple Availability Zones, *floating IP addresses* enable NAS client access. The floating IP addresses, which must be outside of the CIDR blocks for all VPCs in the region, can migrate between nodes when failures occur. They aren't natively accessible to clients that are outside of the VPC, unless you set up an AWS transit gateway.

If you can't set up a transit gateway, private IP addresses are available for NAS clients that are outside the VPC. However, these IP addresses are static—they can't failover between nodes.

You should review requirements for floating IP addresses and route tables before you deploy an HA configuration across multiple availability zones. You must specify the floating IP addresses when you deploy the configuration. The private IP addresses are automatically created.

For more information, refer to AWS networking requirements for Cloud Volumes ONTAP HA in multiple AZs.

iSCSI data access

Cross-VPC data communication is not an issue since iSCSI does not use floating IP addresses.

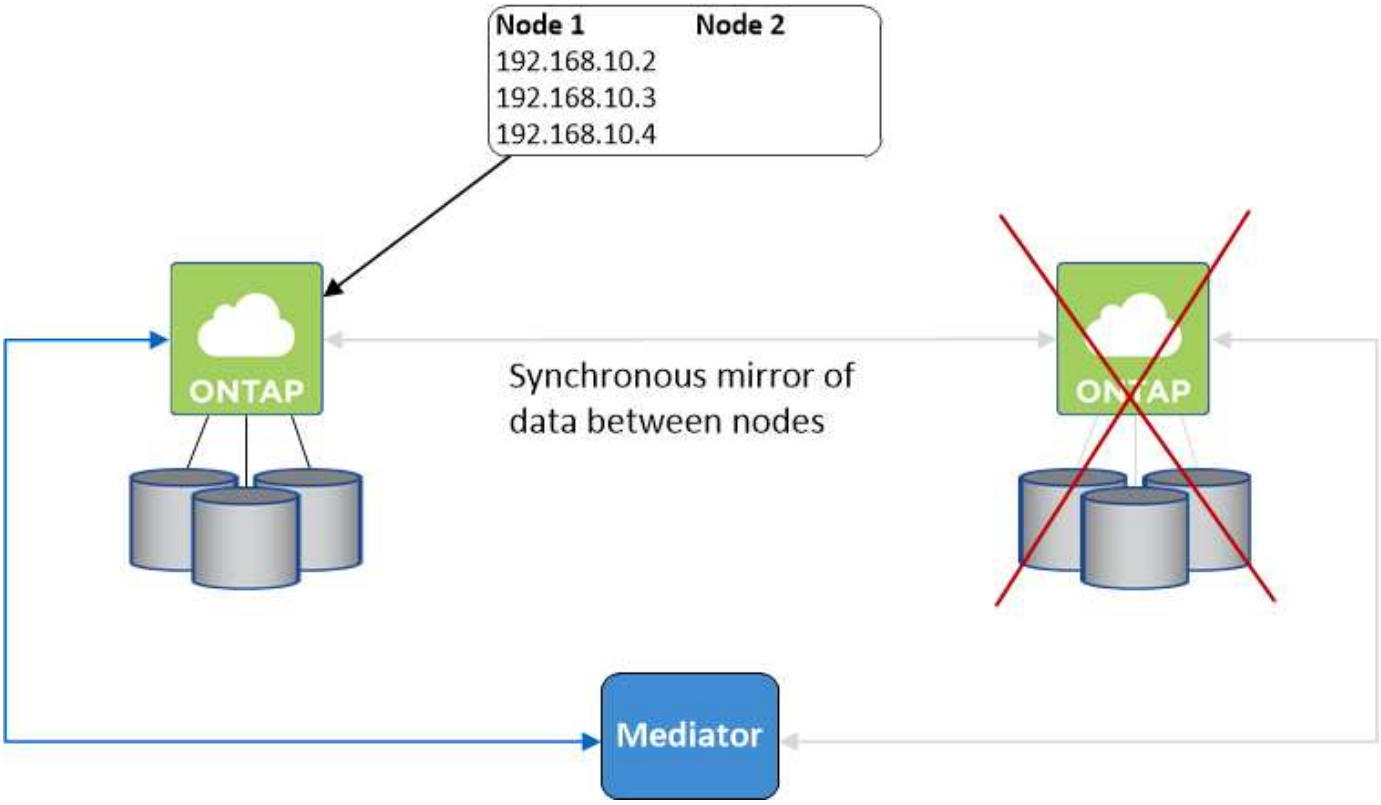
Takeover and giveback for iSCSI

For iSCSI, Cloud Volumes ONTAP uses multipath I/O (MPIO) and Asymmetric Logical Unit Access (ALUA) to manage path failover between the active-optimized and non-optimized paths.

> For information about which specific host configurations support ALUA, refer to the NetApp Interoperability Matrix Tool and the SAN hosts and cloud clients guide for your host operating system.

Takeover and giveback for NAS

When takeover occurs in a NAS configuration using floating IPs, the node's floating IP address that clients use to access data moves to the other node. The following image depicts storage takeover in a NAS configuration using floating IPs. If node 2 goes down, the floating IP address for node 2 moves to node 1.



NAS data IPs used for external VPC access cannot migrate between nodes if failures occur. If a node goes offline, you must manually remount volumes to clients outside the VPC by using the IP address on the other node.

After the failed node comes back online, remount clients to volumes using the original IP address. This step is needed to avoid transferring unnecessary data between two HA nodes, which can cause significant performance and stability impact.

You can locate the correct IP address from the Console by selecting the volume and clicking **Mount Command**.

Single availability zone

Deploying an HA configuration in a single availability zone (AZ) can ensure high availability of your data if an instance that runs a Cloud Volumes ONTAP node fails. All data is natively accessible from outside of the VPC.

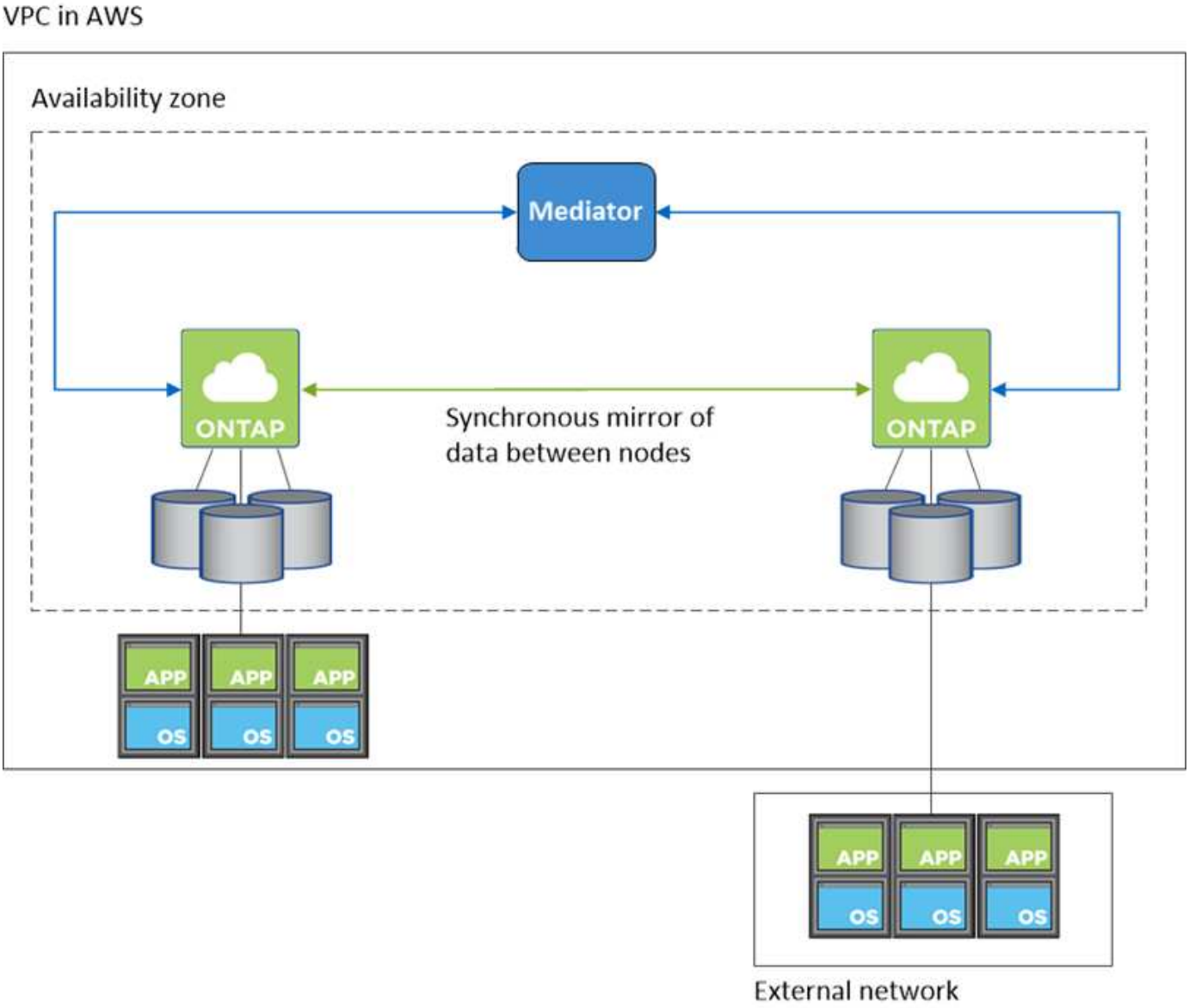
> The Console creates an AWS Documentation: AWS spread placement group and launches the two HA nodes in that placement group. The placement group reduces the risk of simultaneous failures by spreading the instances across distinct underlying hardware. This feature improves redundancy from a compute perspective and not from disk failure perspective.

Data access

Because this configuration is in a single AZ, it does not require floating IP addresses. You can use the same IP

address for data access from within the VPC and from outside the VPC.

The following image shows an HA configuration in a single AZ. Data is accessible from within the VPC and from outside the VPC.



Takeover and giveback

For iSCSI, Cloud Volumes ONTAP uses multipath I/O (MPIO) and Asymmetric Logical Unit Access (ALUA) to manage path failover between the active-optimized and non-optimized paths.

> For information about which specific host configurations support ALUA, refer to the NetApp Interoperability Matrix Tool and the SAN hosts and cloud clients guide for your host operating system.

For NAS configurations, the data IP addresses can migrate between HA nodes if failures occur. This ensures client access to storage.

AWS Local Zones

AWS Local Zones are an infrastructure deployment where storage, compute, database, and other select AWS services are located close to large cities and industry areas. With AWS Local Zones, you can bring AWS services closer to you which improves latency for your workloads and maintain databases locally. On Cloud Volumes ONTAP,

You can deploy a single AZ or multiple AZ configuration in AWS Local Zones.

AWS Local Zones are supported when using the Console in standard and private modes. At this time, AWS Local Zones are not supported in restricted mode.

Example AWS Local Zone configurations

Cloud Volumes ONTAP in AWS supports only high availability (HA) mode in a single availability zone. Single node deployments are not supported.

Cloud Volumes ONTAP does not support data tiering, cloud tiering, and unqualified instances in AWS Local Zones.

The following are example configurations:

- Single availability zone: Both cluster nodes and the mediator are in the same Local Zone.
- Multiple availability zones
  In multiple availability zone configurations, there are three instances, two nodes and one mediator. One instance out of the three instances must be in a separate zone. You can choose how you set this up.

  Here are three example configurations:

  - Each cluster node is in a different Local Zone and the mediator in a public availability zone.
  - One cluster node in a Local Zone, the mediator in a Local Zone, and the second cluster node is in an availability zone.
  - Each cluster node and the mediator are in separate Local Zones.

Supported disk and instance types

The only supported disk type is GP2. The following EC2 instance type families with sizes xlarge to 4xlarge are currently supported:

- M5
- C5
- C5d
- R5
- R5d

Cloud Volumes ONTAP supports only these configurations. Selecting unsupported disk types or unqualified instances in AWS Local Zone configuration might result in deployment failure. Data tiering to Amazon Simple Storage Service (Amazon S3) is not supported if your Cloud Volumes ONTAP system is in an AWS Local Zone, because accessing the Amazon S3 buckets outside of the Local Zone involves higher latency and impacts Cloud Volumes ONTAP activities.

AWS Documentation: EC2 instance types in Local Zones.

How storage works in an HA pair

Unlike an ONTAP cluster, storage in a Cloud Volumes ONTAP HA pair is not shared between nodes. Instead, data is synchronously mirrored between the nodes so that the data is available in the event of failure.

Storage allocation

When you create a new volume and additional disks are required, the Console allocates the same number of disks to both nodes, creates a mirrored aggregate, and then creates the new volume. For example, if two disks are required for the volume, the Console allocates two disks per node for a total of four disks.

Storage configurations

You can use an HA pair as an active-active configuration, in which both nodes serve data to clients, or as an active-passive configuration, in which the passive node responds to data requests only if it has taken over storage for the active node.

You can set up an active-active configuration only when using the Console in the Storage System View.

Performance expectations

A Cloud Volumes ONTAP HA configuration synchronously replicates data between nodes, which consumes network bandwidth. As a result, you can expect the following performance in comparison to a single-node Cloud Volumes ONTAP configuration:

- For HA configurations that serve data from only one node, read performance is comparable to the read performance of a single-node configuration, whereas write performance is lower.
- For HA configurations that serve data from both nodes, read performance is higher than the read performance of a single-node configuration, and write performance is the same or higher.

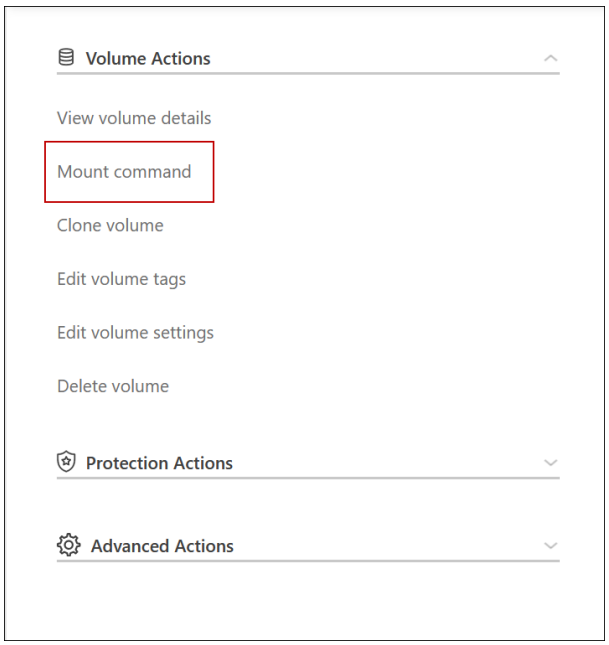
For more details about Cloud Volumes ONTAP performance, refer to Performance.

Client access to storage

Clients should access NFS and CIFS volumes by using the data IP address of the node on which the volume resides. If NAS clients access a volume by using the IP address of the partner node, traffic goes between both nodes, which reduces performance.

If you move a volume between nodes in an HA pair, you should remount the volume by using the IP address of the other node. Otherwise, you can experience reduced performance. If clients support NFSv4 referrals or folder redirection for CIFS, you can enable those features on the Cloud Volumes ONTAP systems to avoid remounting the volume. For details, refer to the ONTAP documentation.

You can easily identify the correct IP address through the *Mount Command* option under the manage volumes panel in.



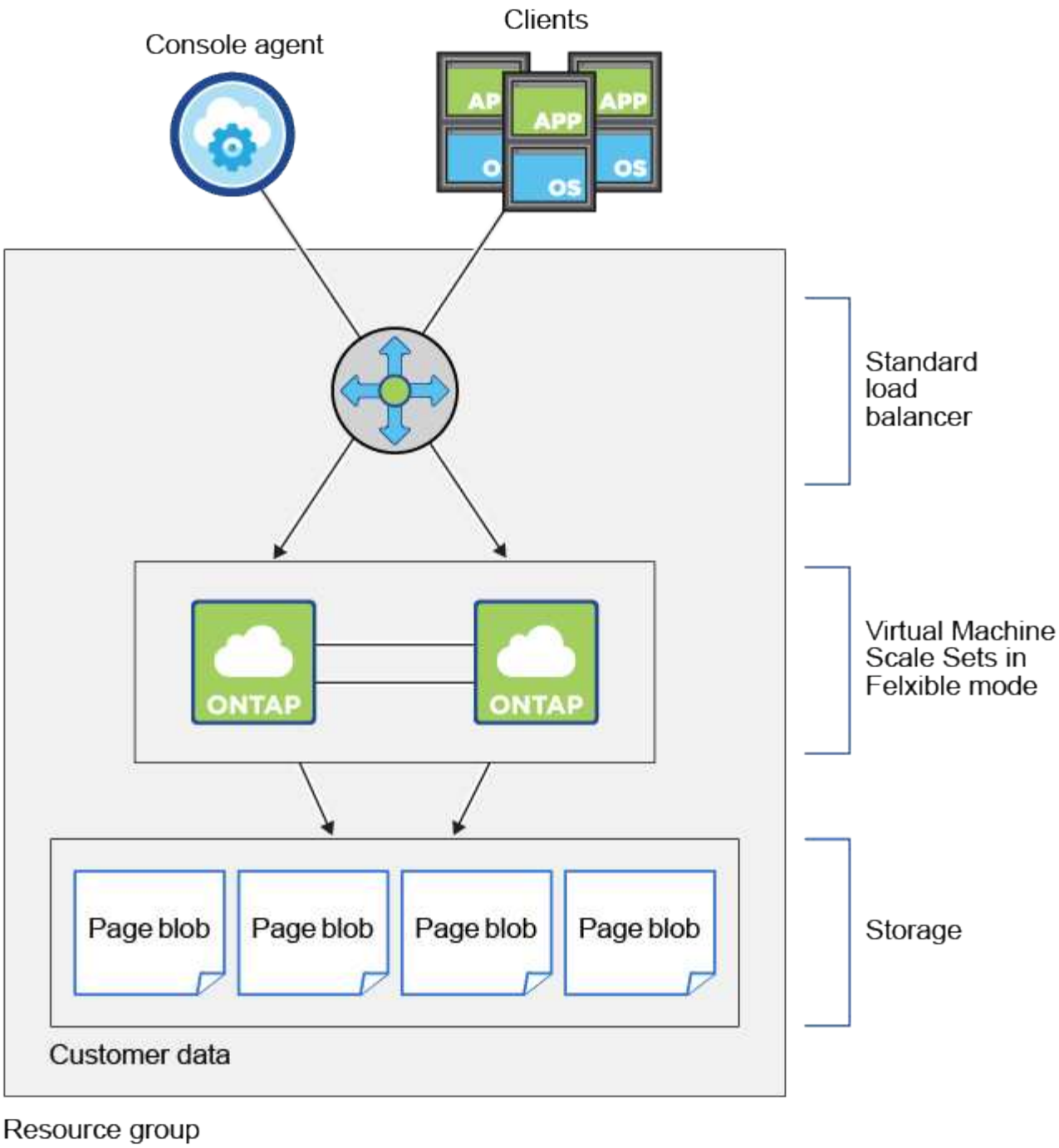
Learn about Cloud Volumes ONTAP HA pairs in Azure

A Cloud Volumes ONTAP high-availability (HA) pair provides enterprise reliability and continuous operations in case of failures in your cloud environment. In Azure, storage is shared between the two nodes.

HA components

HA single availability zone configuration with page blobs

A Cloud Volumes ONTAP HA page blob configuration in Azure includes the following components:



Note the following about the Azure components that the NetApp Console deploys for you:

Azure Standard Load Balancer

The load balancer manages incoming traffic to the Cloud Volumes ONTAP HA pair.

VMs in single availability zones

Beginning with Cloud Volumes ONTAP 9.15.1, you can create and manage heterogeneous virtual machines (VMs) in a single availability zone (AZ). You can deploy high-availability (HA) nodes in separate fault domains within the same AZ, guaranteeing optimal availability. To learn more about the flexible orchestration mode that enables this capability, refer to the Microsoft Azure documentation: Virtual Machine Scale Sets.

Disks

Customer data resides on Premium Storage page blobs. Each node has access to the other node's storage. Additional storage is also required for boot, root, and core data.

Storage accounts

- One storage account is required for managed disks.
- One or more storage accounts are required for the Premium Storage page blobs, as the disk capacity limit per storage account is reached.

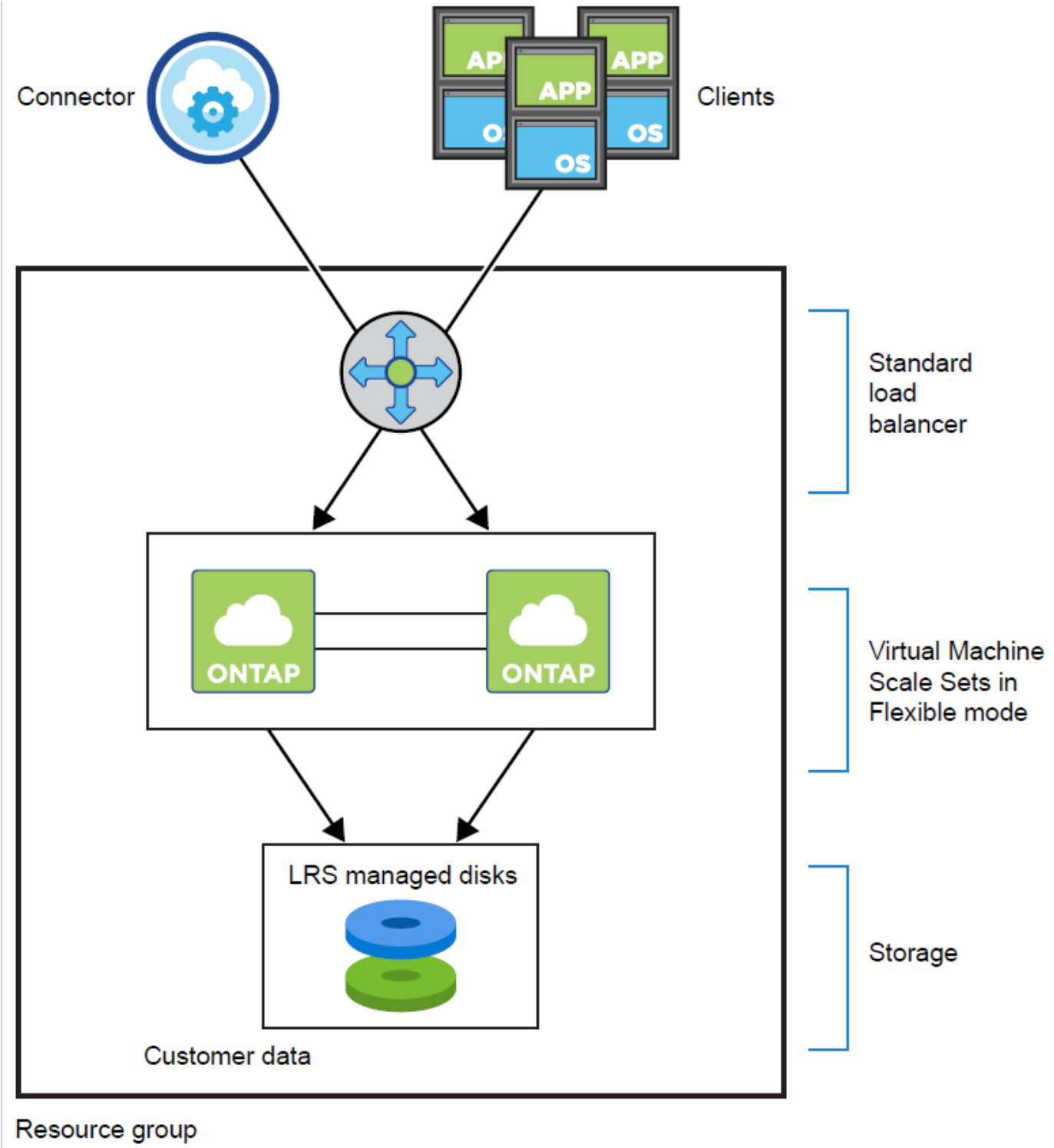
  Microsoft Azure documentation: Azure Storage scalability and performance targets for storage accounts.

- One storage account is required for data tiering to Azure Blob storage.
- Starting with Cloud Volumes ONTAP 9.7, the storage accounts that the Console creates for HA pairs are general-purpose v2 storage accounts.
- You can enable an HTTPS connection from a Cloud Volumes ONTAP 9.7 HA pair to Azure storage accounts when adding a Cloud Volumes ONTAP system. Note that enabling this option can impact write performance. You can't change the setting after you create the system.

> Starting with Cloud Volumes ONTAP 9.15.0P1, Azure page blobs are no longer supported for new high-availability pair deployments. If you currently use Azure page blobs in existing high-availability pair deployments, you can migrate to newer VM instance types in the Edsv4-series VMs and Edsv5-series VMs.
> Learn more about supported configurations in Azure.

HA single availability zone configuration with shared managed disks

A Cloud Volumes ONTAP HA single availability zone configuration running on top of shared managed disk includes the following components:



Note the following about the Azure components that the Console deploys for you:

Azure Standard Load Balancer

The load balancer manages incoming traffic to the Cloud Volumes ONTAP HA pair.

VMs in single availability zones

Beginning with Cloud Volumes ONTAP 9.15.1, you can create and manage heterogeneous virtual machines (VMs) in a single availability zone (AZ). You can deploy high-availability (HA) nodes in separate fault domains within the same AZ, guaranteeing optimal availability. To learn more about the flexible orchestration mode that enables this capability, refer to the Microsoft Azure documentation: Virtual Machine

Scale Sets.

The zonal deployment uses Premium SSD v2 Managed Disks when the following conditions are fulfilled:

- The version of Cloud Volumes ONTAP is 9.15.1 or later.
- The selected region and zone support Premium SSD v2 Managed Disks. For information about the supported regions, refer to Microsoft Azure website: Products available by region.
- The subscription is registered for the Microsoft Microsoft.Compute/VMOrchestratorZonalMultiFD feature.

If you choose Premium SSD Managed Disks for an environment that fulfils the above criteria, the Console automatically deploys Premium SSD v2 Managed Disks. You cannot switch to Premium SSD v1 Managed Disks.

Disks

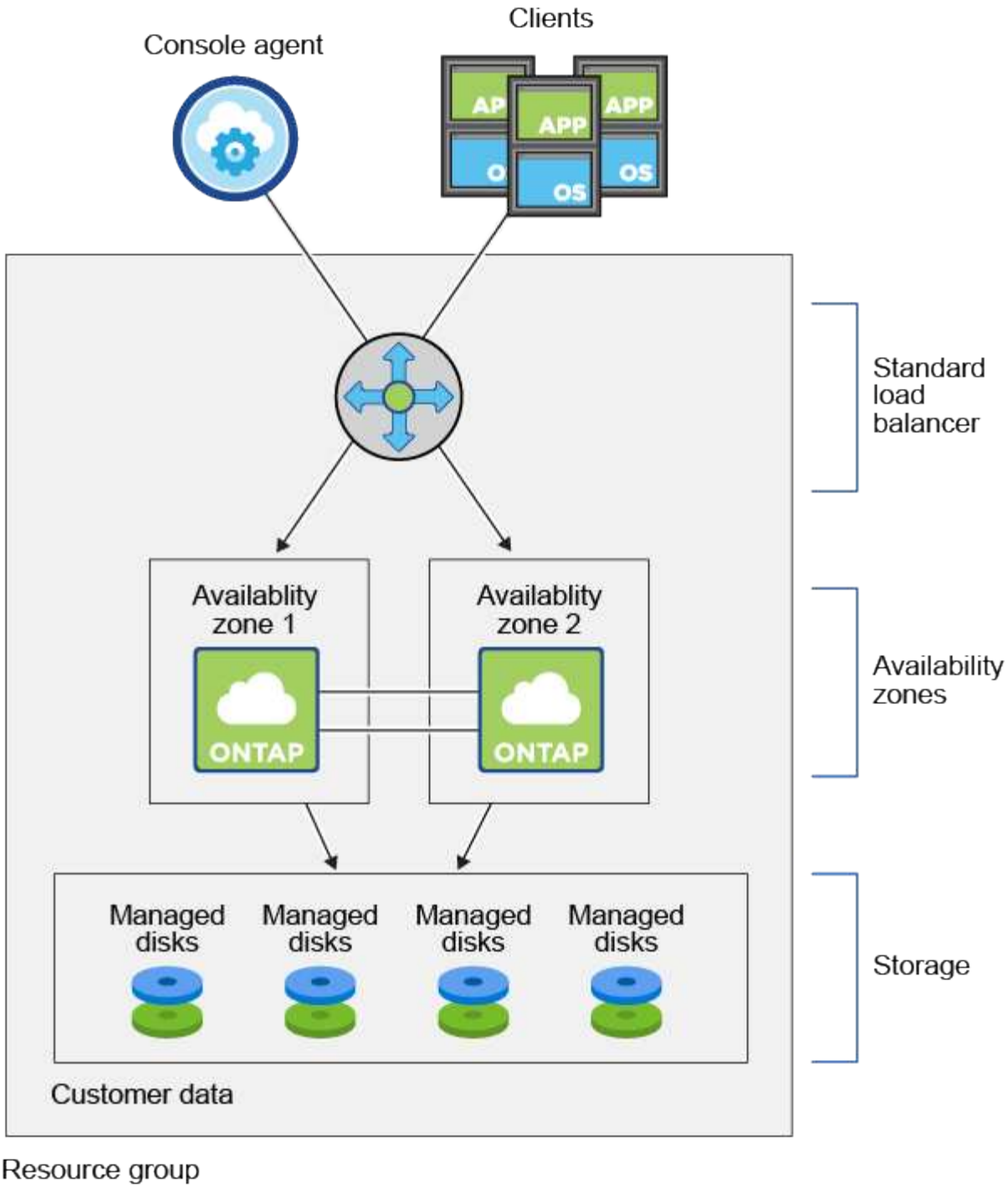
Customer data resides on locally-redundant storage (LRS) managed disks. Each node has access to the other node's storage. Additional storage is also required for boot, root, partner root, core, and NVRAM data.

Storage accounts

Storage accounts are used for managed disk based deployments to handle diagnostic logs and tiering to blob storage.

HA multiple availability zone configuration

A Cloud Volumes ONTAP HA multiple availability zone configuration in Azure includes the following components:



Note the following about the Azure components that the Console deploys for you:

Azure Standard Load Balancer

The load balancer manages incoming traffic to the Cloud Volumes ONTAP HA pair.

Availability Zones

HA multiple availability zone configuration utilizes a deployment model where two Cloud Volumes ONTAP nodes are deployed into different availability zones, ensuring that the nodes are in different fault domains to provide redundancy and availability. To learn how Virtual Machine Scale Sets in Flexible orchestration mode can use availability zones in Azure, refer to the Microsoft Azure documentation: Create a Virtual Machine Scale Set that uses Availability Zones.

Disks

Customer data resides on zone-redundant storage (ZRS) managed disks. Each node has access to the other node's storage. Additional storage is also required for boot, root, partner root, and core data.

Storage accounts

Storage accounts are used for managed disk based deployments to handle diagnostic logs and tiering to blob storage.

RPO and RTO

An HA configuration maintains high availability of your data as follows:

- The recovery point objective (RPO) is 0 seconds.
  Your data is transactionally consistent with no data loss.
- The recovery time objective (RTO) is 120 seconds.
  In the event of an outage, data should be available in 120 seconds or less.

Storage takeover and giveback

Similar to a physical ONTAP cluster, storage in an Azure HA pair is shared between nodes. Connections to the partner's storage allows each node to access the other's storage in the event of a *takeover*. Network path failover mechanisms ensure that clients and hosts continue to communicate with the surviving node. The partner *gives back* storage when the node is brought back on line.

For NAS configurations, data IP addresses automatically migrate between HA nodes if failures occur.

For iSCSI, Cloud Volumes ONTAP uses multipath I/O (MPIO) and Asymmetric Logical Unit Access (ALUA) to manage path failover between the active-optimized and non-optimized paths.

For information about which specific host configurations support ALUA, refer to the NetApp Interoperability Matrix Tool and the SAN hosts and cloud clients guide for your host operating system.

Storage takeover, resync, and giveback are all automatic by default. No user action is required.

Storage configurations

You can use an HA pair as an active-active configuration, in which both nodes serve data to clients, or as an active-passive configuration, in which the passive node responds to data requests only if it has taken over the storage for the active node.

Learn about Cloud Volumes ONTAP HA pairs in Google Cloud

A Cloud Volumes ONTAP high-availability (HA) configuration provides nondisruptive operations and fault tolerance. In Google Cloud, data is synchronously mirrored between the two nodes.

HA components

Cloud Volumes ONTAP HA configurations in Google Cloud include the following components:

- Two Cloud Volumes ONTAP nodes whose data is synchronously mirrored between each other.
- A mediator instance that provides a communication channel between the nodes to assist in storage takeover and giveback processes.
- One zone or three zones (recommended).

  If you choose three zones, the two nodes and mediator are in separate Google Cloud zones.

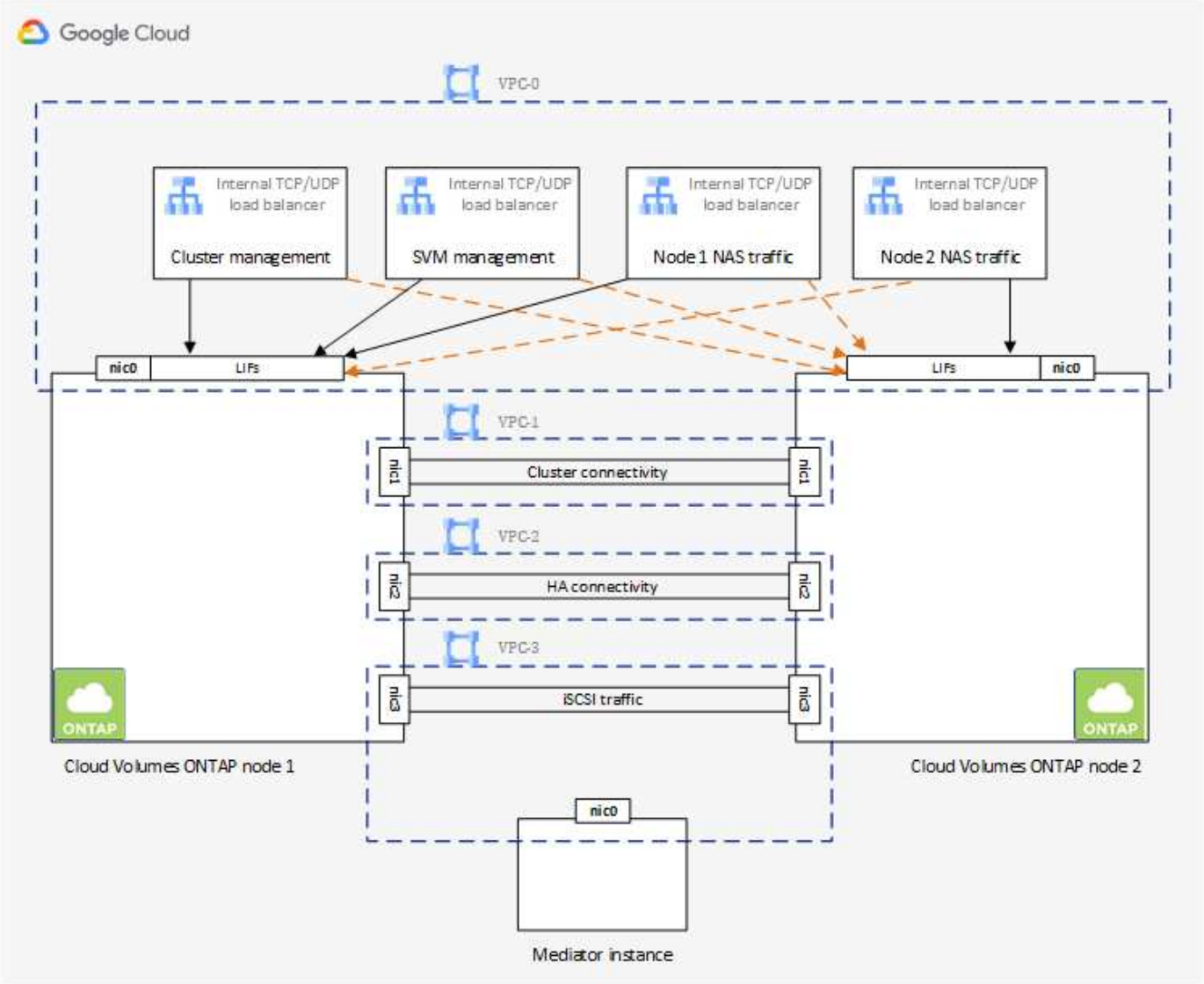
- Four Virtual Private Clouds (VPCs).

  The configuration uses four VPCs because GCP requires that each network interface resides in a separate VPC network.

- Four Google Cloud internal load balancers (TCP/UDP) that manage incoming traffic to the Cloud Volumes ONTAP HA pair.

Learn about networking requirements, including more details about load balancers, VPCs, internal IP addresses, subnets, and more.

The following conceptual image shows a Cloud Volumes ONTAP HA pair and its components:



Mediator

Here are some key details about the mediator instance in Google Cloud:

Instance type

e2-micro (an f1-micro instance was previously used)

Disks

Two standard persistent disks that are 10 GiB each

Operating system

Debian 11

For Cloud Volumes ONTAP 9.10.0 and earlier, Debian 10 was installed on the mediator.

Upgrades

When you upgrade Cloud Volumes ONTAP, the NetApp Console also updates the mediator instance as needed.

Access to the instance

For Debian, the default cloud user is `admin`. Google Cloud creates and adds a certificates for the `admin` user when SSH access is requested through the Google Cloud Console or gcloud command line. You can specify `sudo` to gain root privileges.

Third-party agents

Third-party agents or VM extensions are not supported on the mediator instance.

Storage takeover and giveback

If a node goes down, the other node can serve data for its partner to provide continued data service. Clients can access the same data from the partner node because the data was synchronously mirrored to the partner.

After the node reboots, the partner must resync data before it can return the storage. The time that it takes to resync data depends on how much data was changed while the node was down.

Storage takeover, resync, and giveback are all automatic by default. No user action is required.

RPO and RTO

An HA configuration maintains high availability of your data as follows:

- The recovery point objective (RPO) is 0 seconds.

  Your data is transactionally consistent with no data loss.

- The recovery time objective (RTO) is 120 seconds.

  In the event of an outage, data should be available in 120 seconds or less.

HA deployment models

You can ensure the high availability of your data by deploying an HA configuration in multiple zones or in a single zone.

Multiple zones (recommended)

Deploying an HA configuration across three zones ensures continuous data availability if a failure occurs within a zone. Note that write performance is slightly lower compared to using a single zone, but it's minimal.

Single zone

When deployed in a single zone, a Cloud Volumes ONTAP HA configuration uses a spread placement policy. This policy ensures that an HA configuration is protected from a single point of failure within the zone, without having to use separate zones to achieve fault isolation.

This deployment model does lower your costs because there are no data egress charges between zones.

How storage works in an HA pair

Unlike an ONTAP cluster, storage in a Cloud Volumes ONTAP HA pair in GCP is not shared between nodes. Instead, data is synchronously mirrored between the nodes so that the data is available in the event of failure.

Storage allocation

When you create a new volume and additional disks are required, the Console allocates the same number of disks to both nodes, creates a mirrored aggregate, and then creates the new volume. For example, if two disks are required for the volume, the Console allocates two disks per node for a total of four disks.

Storage configurations

You can use an HA pair as an active-active configuration, in which both nodes serve data to clients, or as an active-passive configuration, in which the passive node responds to data requests only if it has taken over storage for the active node.

Performance expectations for an HA configuration

A Cloud Volumes ONTAP HA configuration synchronously replicates data between nodes, which consumes network bandwidth. As a result, you can expect the following performance in comparison to a single-node Cloud Volumes ONTAP configuration:

- For HA configurations that serve data from only one node, read performance is comparable to the read performance of a single-node configuration, whereas write performance is lower.
- For HA configurations that serve data from both nodes, read performance is higher than the read performance of a single-node configuration, and write performance is the same or higher.

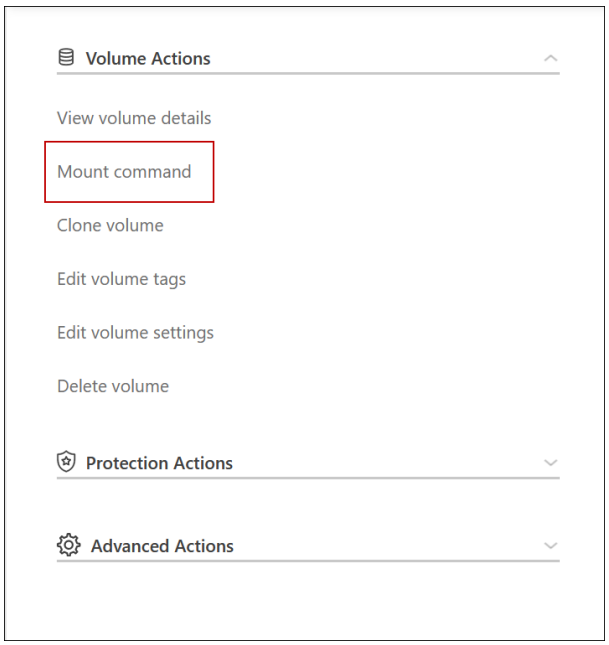
For more details about Cloud Volumes ONTAP performance, refer to Performance.

Client access to storage

Clients should access NFS and CIFS volumes by using the data IP address of the node on which the volume resides. If NAS clients access a volume by using the IP address of the partner node, traffic goes between both nodes, which reduces performance.

If you move a volume between nodes in an HA pair, you should remount the volume by using the IP address of the other node. Otherwise, you can experience reduced performance. If clients support NFSv4 referrals or folder redirection for CIFS, you can enable those features on the Cloud Volumes ONTAP systems to avoid remounting the volume. For details, refer to ONTAP documentation.

You can locate the correct IP address from the Console by selecting the volume and clicking **Mount Command**.



Related links

- Learn about networking requirements
- Learn how to get started in GCP

Operations unavailable when a node in Cloud Volumes ONTAP HA pair is offline

When a node in an HA pair isn't available, the other node serves data for its partner to provide continued data service. This is called *storage takeover*. Several actions are unavailable until in storage giveback is complete.

When a node in an HA pair is unavailable, the state of the system in the NetApp Console is *Degraded*.

The following actions are unavailable from storage takeover:

- Support registration
- License changes
- Instance or VM type changes
- Write speed changes
- CIFS setup
- Changing the location of configuration backups
- Setting the cluster password
- Managing disks and aggregates (advanced allocation)

These actions are available again after storage giveback completes and the state of the system changes back to normal.

Copyright information

Trademark information