Copy and synchronize data

Cloud Manager

NetApp
March 03, 2022

Table of Contents

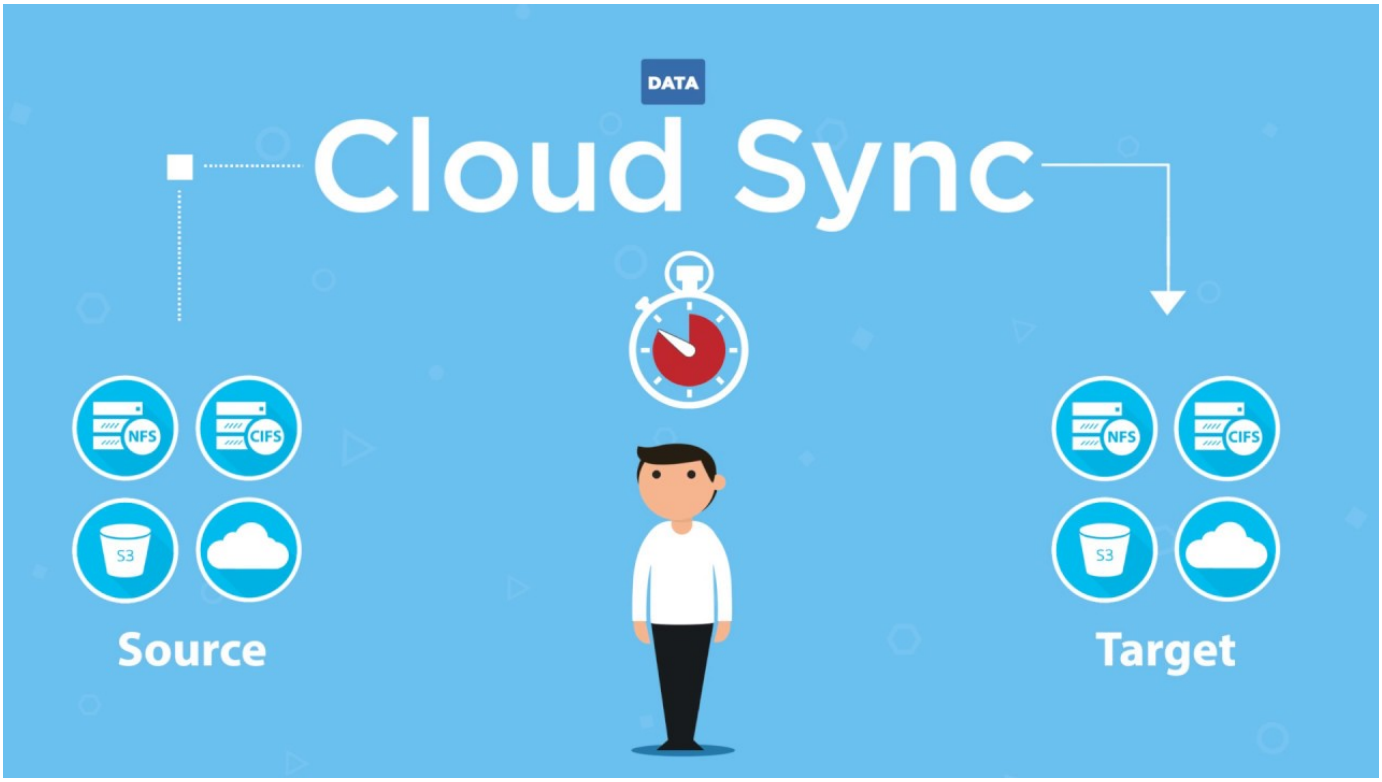
Copy and synchronize data

Cloud Sync overview

The NetApp Cloud Sync service offers a simple, secure, and automated way to migrate your data to any target, in the cloud or on your premises. Whether it's a file-based NAS dataset (NFS or SMB), Amazon Simple Storage Service (S3) object format, a NetApp StorageGRID® appliance, or any other cloud provider object store, Cloud Sync can convert and move it for you.

Features

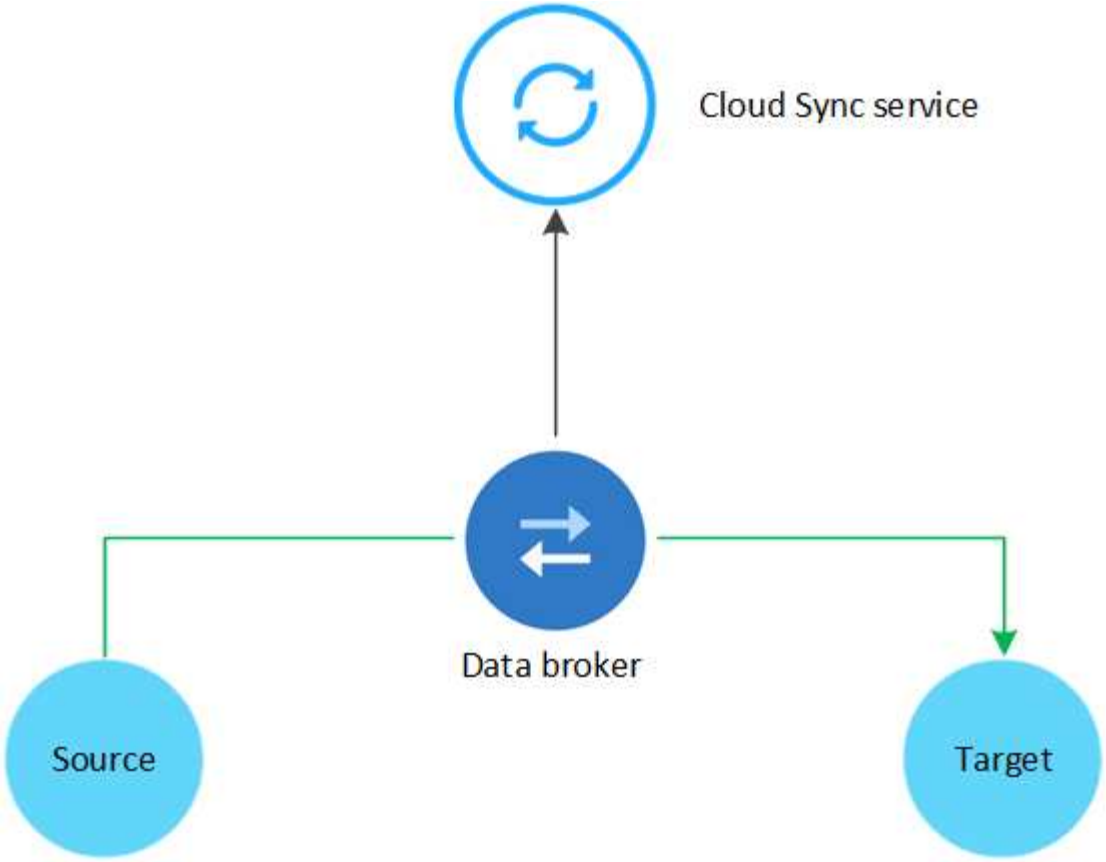
Watch the following video for an overview of Cloud Sync:



How Cloud Sync works

Cloud Sync is a software-as-a-service (SaaS) platform that consists of a data broker group, a cloud-based interface available through Cloud Manager, and a source and target.

The following image shows the relationship between Cloud Sync components:



The NetApp data broker software syncs data from a source to a target (this is called a *sync relationship*). You can run the data broker in AWS, Azure, Google Cloud Platform, or on your premises. A data broker group, which consists of one or more data brokers, needs an outbound internet connection over port 443 so it can communicate with the Cloud Sync service and contact a few other services and repositories. View the list of endpoints.

After the initial copy, the service syncs any changed data based on the schedule that you set.

Supported storage types

Cloud Sync supports the following storage types:

- Any NFS server
- Any SMB server
- Amazon EFS
- Amazon FSx for ONTAP
- Amazon S3
- Azure Blob
- Azure NetApp Files
- Box (available as a preview)
- Cloud Volumes Service
- Cloud Volumes ONTAP
- Google Cloud Storage

- IBM Cloud Object Storage
- On-premises ONTAP cluster
- ONTAP S3 Storage
- SFTP (using API only)
- StorageGRID

Review the supported sync relationships.

Cost

There are two types of costs associated with using Cloud Sync: resource charges and service charges.

Resource charges

Resource charges are related to the compute and storage costs for running one or more data brokers in the cloud.

Service charges

There are two ways to pay for sync relationships after your 14-day free trial ends. The first option is to subscribe from AWS or Azure, which enables you to pay hourly or annually. The second option is to purchase licenses directly from NetApp. Read the following sections for more details.

Marketplace subscription

Subscribing to the Cloud Sync service from AWS or Azure enables you to pay at an hourly rate, or to pay annually. You can subscribe through either AWS or Azure, depending on where you want to be billed.

Hourly subscription

With an hourly pay-as-you-go subscription, you're charged hourly based on the number of sync relationships that you create.

- View pricing in Azure
- View pay-as-you-go pricing in AWS

Annual subscription

An annual subscription provides a license for 20 sync relationships that you pay for up front. If you go above 20 sync relationships and you've subscribed through AWS, you pay for the additional relationships by the hour.

View annual pricing in AWS

Licenses from NetApp

Another way to pay for sync relationships up front is by purchasing licenses directly from NetApp. Each license enables you to create up to 20 sync relationships.

You can use these licenses with an AWS or Azure subscription. For example, if you have 25 sync relationships, you can pay for the first 20 sync relationships using a license and then pay-as-you-go from AWS or Azure with the remaining 5 sync relationships.

Learn how to purchase licenses and add them to Cloud Sync.

License terms

Customers who purchase a Bring Your Own License (BYOL) to the Cloud Sync service should be aware of limitations associated with the license entitlement.

- Customers are entitled to leverage the BYOL license for a term not to exceed one year from the date of delivery.
- Customers are entitled to leverage the BYOL license to establish and not to exceed a total of 20 individual connections between a source and a target (each a “sync relationship”).
- A customer’s entitlement expires at the conclusion of the one-year license term, irrespective as to whether Customer has reached the 20 sync relationship limitation.
- In the event the Customer chooses to renew its license, unused sync relationships associated from the previous license grant DO NOT roll over to the license renewal.

Data privacy

NetApp doesn’t have access to any credentials that you provide while using the Cloud Sync service. The credentials are stored directly on the data broker machine, which resides in your network.

Depending on the configuration that you choose, Cloud Sync might prompt you for credentials when you create a new relationship. For example, when setting up a relationship that includes an SMB server, or when deploying the data broker in AWS.

These credentials are always saved directly to the data broker itself. The data broker resides on a machine in your network, whether it’s on premises or in your cloud account. The credentials are never made available to NetApp.

The credentials are locally encrypted on the data broker machine using HashiCorp Vault.

Limitations

- Cloud Sync is not supported in China.
- In addition to China, the Cloud Sync data broker is not supported in the following regions:
  - Azure US Gov
  - Azure US DoD

Get started

Quick start for Cloud Sync

Getting started with the Cloud Sync service includes a few steps.

Prepare your source and target

Verify that your source and target are supported and set up. The most important requirement is to verify connectivity between the data broker group and the source and target locations. Learn more.

Prepare a location for the NetApp data broker

The NetApp data broker software syncs data from a source to a target (this is called a *sync relationship*). You can run the data broker in AWS, Azure, Google Cloud Platform, or on your premises. A data broker group, which consists of one or more data brokers, needs an outbound internet connection over port 443 so it can communicate with the Cloud Sync service and contact a few other services and repositories. View the list of endpoints.

Cloud Sync guides you through the installation process when you create a sync relationship, at which point you can deploy a data broker in the cloud or download an install script for your own Linux host.

- Review AWS installation
- Review Azure installation
- Review GCP installation
- Review Linux host installation

Create your first sync relationship

Log in to Cloud Manager, click **Sync**, and then drag and drop your selections for the source and target. Follow the prompts to complete the setup. Learn more.

Pay for your sync relationships after your free trial ends

Subscribe from AWS or Azure to pay-as-you-go or to pay annually. Or purchase licenses directly from NetApp. Just go to the License Settings page in Cloud Sync to set it up. Learn more.

Preparing the source and target

Prepare to sync data by verifying that your source and target are supported and setup.

Supported sync relationships

Cloud Sync enables you to sync data from a source to a target (this is called a *sync relationship*). You should understand the supported relationships before you get started.

Source location	Supported target locations
Amazon EFS	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB server • StorageGRID
Amazon FSx for ONTAP	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID

Source location	Supported target locations
Amazon S3	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Box [2,3] • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID
Azure Blob	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID

Source location	Supported target locations
Azure NetApp Files	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB server • StorageGRID
Box [2]	• Amazon S3 • IBM Cloud Object Storage • NFS server • SMB Server • StorageGRID
Cloud Volumes ONTAP	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID

Source location	Supported target locations
Cloud Volumes Service	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID
Google Cloud Storage	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • ONTAP S3 Storage • SMB Server • StorageGRID

Source location	Supported target locations
IBM Cloud Object Storage	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Box [2,3] • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID
NFS server	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID

Source location	Supported target locations
On-prem ONTAP cluster	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID
ONTAP S3 Storage	• Google Cloud Storage • SMB server • StorageGRID • ONTAP S3 Storage
SFTP [1]	S3
SMB server	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • ONTAP S3 Storage • SMB Server • StorageGRID

Source location	Supported target locations
StorageGRID	• Amazon EFS • Amazon FSx for ONTAP • Amazon S3 • Azure Blob • Azure NetApp Files • Box [2,3] • Cloud Volumes ONTAP • Cloud Volumes Service • Google Cloud Storage • IBM Cloud Object Storage • NFS server • On-premises ONTAP cluster • ONTAP S3 Storage • SMB Server • StorageGRID

Notes:

1. Sync relationships with this source/target are supported by using the Cloud Sync API only.
2. Box support is available as a preview.
3. You can choose a specific Azure Blob storage tier when a Blob container is the target:
    - Hot storage
    - Cool storage
4. You can choose a specific S3 storage class when Amazon S3 is the target:
    - Standard (this is the default class)
    - Intelligent-Tiering
    - Standard-Infrequent Access
    - One Zone-Infrequent Access
    - Glacier
    - Glacier Deep Archive
5. You can choose a specific storage class when a Google Cloud Storage bucket is the target:
    - Standard
    - Nearline
    - Coldline
    - Archive

Source and target requirements

Verify that your source and targets meet the following requirements.

Networking

- The source and target must have a network connection to the data broker group.

  For example, if an NFS server is in your data center and a data broker is in AWS, then you need a network connection (VPN or Direct Connect) from your network to the VPC.

- NetApp recommends configuring the source, the target, and data brokers to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Target directory

When you create a sync relationship, Cloud Sync enables you to select an existing target directory and then optionally create a new folder inside that directory. So be sure that your preferred target directory already exists.

Permissions to read directories

In order to show every directory or folder in a source or target, Cloud Sync needs read permissions on the directory or folder.

NFS

Permissions must be defined on the source/target with uid/gid on files and directories.

Object storage

- For AWS and Google Cloud, a data broker must have list object permissions (these permissions are provided by default if you follow the data broker installation steps).
- For Azure, StorageGRID, and IBM, the credentials that you enter when setting up a sync relationship must have list object permissions.

SMB

The SMB credentials that you enter when setting up a sync relationship must have list folder permissions.

The data broker ignores the following directories by default: .snapshot, ~snapshot, .copy-offload

Amazon S3 bucket requirements

Make sure that your Amazon S3 bucket meets the following requirements.

Supported data broker locations for Amazon S3

Sync relationships that include S3 storage require a data broker deployed in AWS or on your premises. In either case, Cloud Sync prompts you to associate the data broker with an AWS account during installation.

- Learn how to deploy the AWS data broker
- Learn how to install the data broker on a Linux host

Supported AWS regions

All regions are supported except for the China regions.

Permissions required for S3 buckets in other AWS accounts

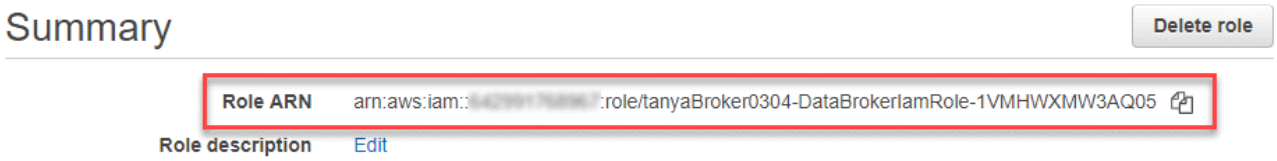
When setting up a sync relationship, you can specify an S3 bucket that resides in an AWS account that isn't associated with a data broker.

The permissions included in this JSON file must be applied to that S3 bucket so a data broker can access it. These permissions enable the data broker to copy data to and from the bucket and to list the objects in the bucket.

Note the following about the permissions included in the JSON file:

1. *<BucketName>* is the name of the bucket that resides in the AWS account that isn't associated with a data broker.
2. *<RoleARN>* should be replaced with one of the following:
   - If a data broker was manually installed on a Linux host, *RoleARN* should be the ARN of the AWS user for which you provided AWS credentials when deploying a data broker.
   - If a data broker was deployed in AWS using the CloudFormation template, *RoleARN* should be the ARN of the IAM role created by the template.

     You can find the Role ARN by going to the EC2 console, selecting the data broker instance, and clicking the IAM role from the Description tab. You should then see the Summary page in the IAM console that contains the Role ARN.



Azure Blob storage requirements

Make sure that your Azure Blob storage meets the following requirements.

Supported data broker locations for Azure Blob

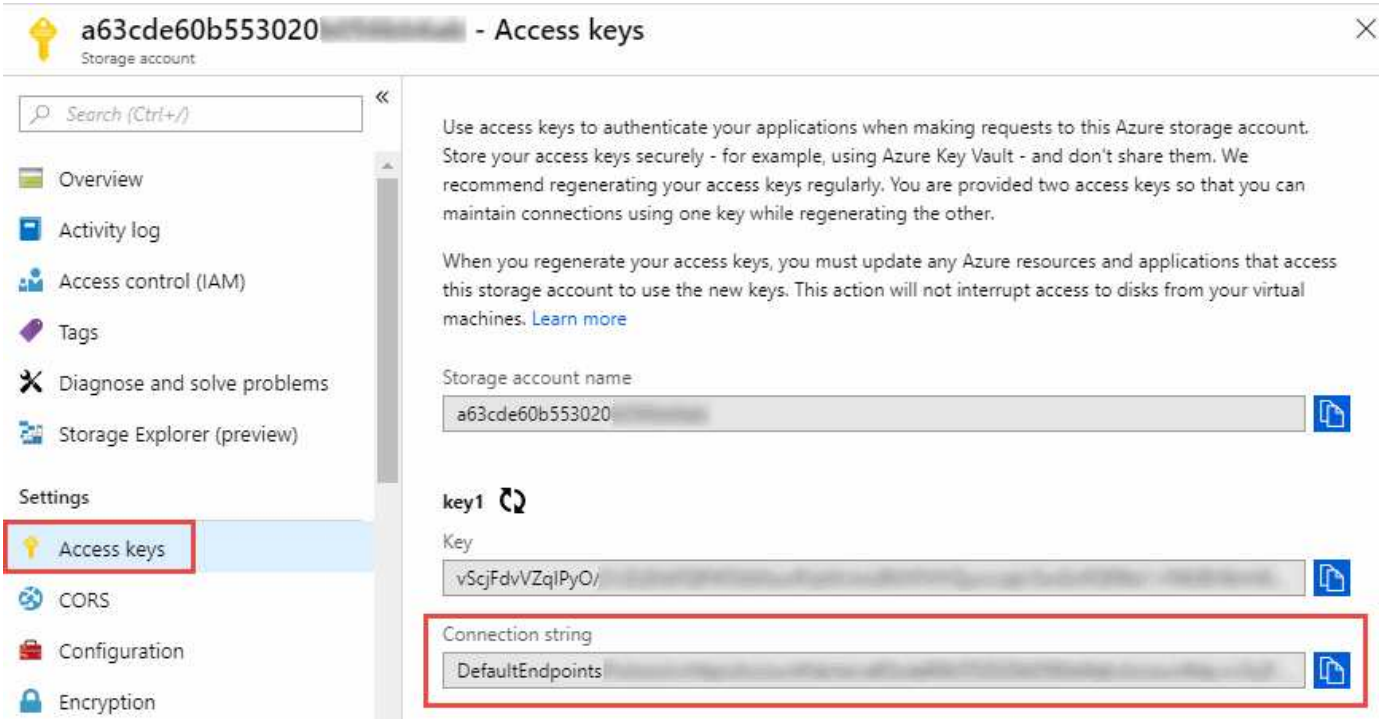
A data broker can reside in any location when a sync relationship includes Azure Blob storage.

Supported Azure regions

All regions are supported except for the China, US Gov, and US DoD regions.

Connection string required for relationships that include Azure Blob and NFS/SMB

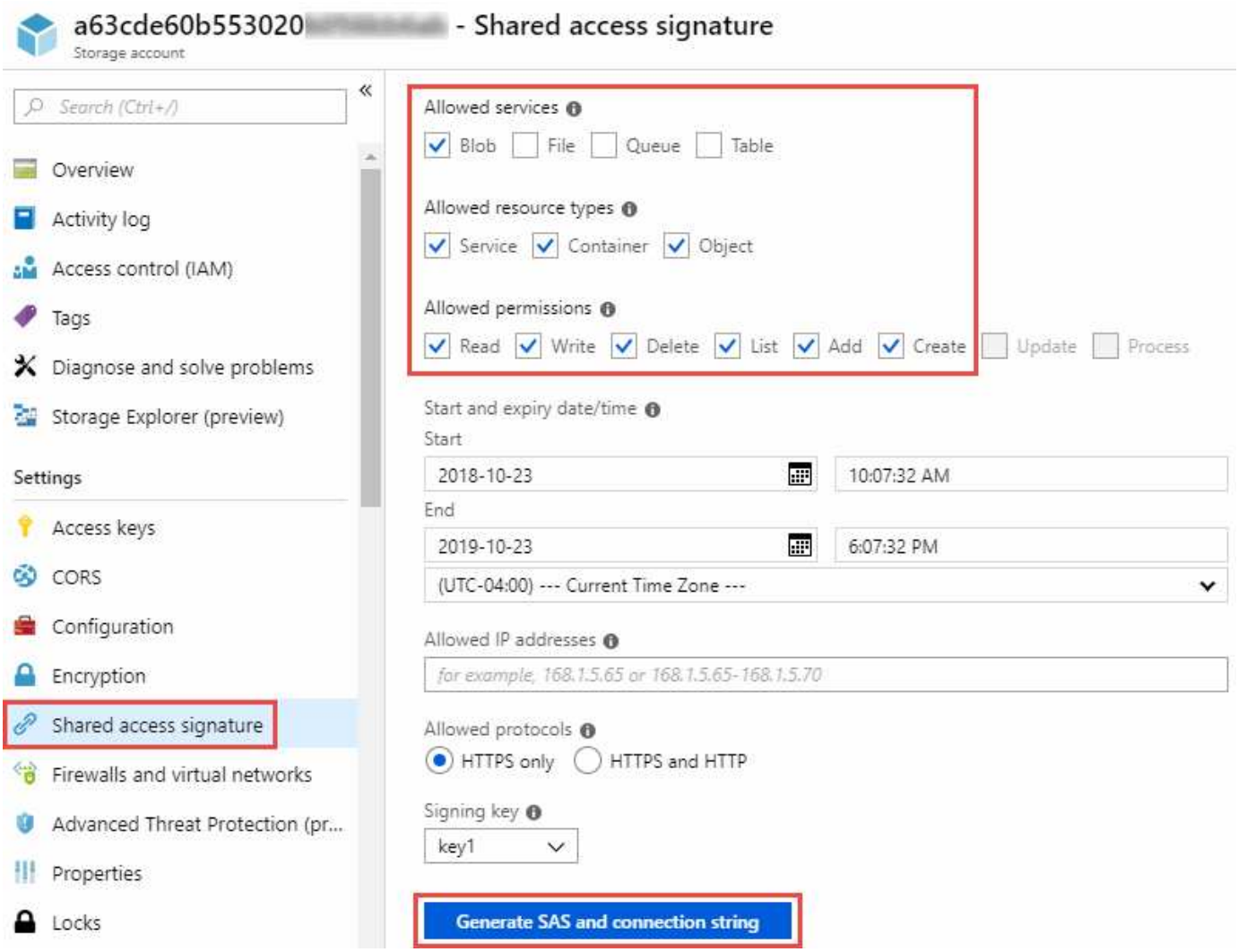
When creating a sync relationship between an Azure Blob container and an NFS or SMB server, you need to provide Cloud Sync with the storage account connection string:



If you want to sync data between two Azure Blob containers, then the connection string must include a shared access signature (SAS). You also have the option to use a SAS when syncing between a Blob container and an NFS or SMB server.

The SAS must allow access to the Blob service and all resource types (Service, Container, and Object). The SAS must also include the following permissions:

- For the source Blob container: Read and List
- For the target Blob container: Read, Write, List, Add, and Create



Azure NetApp Files requirement

Use the Premium or Ultra service level when you sync data to or from Azure NetApp Files. You might experience failures and performance issues if the disk service level is Standard.

Consult a solutions architect if you need help determining the right service level. The volume size and volume tier determines the throughput that you can get.

Learn more about Azure NetApp Files service levels and throughput.

Box requirements

- To create a sync relationship that includes Box, you'll need to provide the following credentials:
  - Client ID
  - Client secret
  - Private key
  - Public key ID
  - Passphrase
  - Enterprise ID

- If you create a sync relationship from Amazon S3 to Box, you must use a data broker group that has a unified configuration where the following settings are set to 1:
  - Scanner Concurrency
  - Scanner Processes Limit
  - Transferrer Concurrency
  - Transferrer Processes Limit

  Learn how to define a unified configuration for a data broker group.

Google Cloud Storage bucket requirements

Make sure that your Google Cloud Storage bucket meets the following requirements.

Supported data broker locations for Google Cloud Storage

Sync relationships that include Google Cloud Storage require a data broker deployed in Google Cloud or on your premises. Cloud Sync guides you through the data broker installation process when you create a sync relationship.

- Learn how to deploy the Google Cloud data broker
- Learn how to install the data broker on a Linux host

Supported GCP regions

All regions are supported.

Permissions required for buckets in other Google Cloud projects

When setting up a sync relationship, you can choose from Google Cloud buckets in different projects, if you provide the required permissions to the data broker's service account. Learn how to set up the service account.

ONTAP requirements

If the sync relationship includes Cloud Volumes ONTAP or an on-prem ONTAP cluster and you selected NFSv4 or later, then you'll need to enable NFSv4 ACLs on the ONTAP system. This is required to copy the ACLs.

Permissions for a SnapMirror destination

If the source for a sync relationship is a SnapMirror destination (which is read-only), "read/list" permissions are sufficient to sync data from the source to a target.

NFS server requirements

- The NFS server can be a NetApp system or a non-NetApp system.
- The file server must allow a data broker host to access the exports.
- NFS versions 3, 4.0, 4.1, and 4.2 are supported.

  The desired version must be enabled on the server.

- If you want to sync NFS data from an ONTAP system, ensure that access to the NFS export list for an SVM

is enabled (vserver nfs modify -vserver *svm_name* -showmount enabled).

The default setting for showmount is *enabled* starting with ONTAP 9.2.

ONTAP S3 Storage requirements

When you set up a sync relationship that includes ONTAP S3 Storage, you'll need to provide the following:

- The IP address of the LIF that's connected to ONTAP S3
- The access key and secret key that ONTAP is configured to use

SMB server requirements

- The SMB server can be a NetApp system or a non-NetApp system.
- You need to provide Cloud Sync with credentials that have permissions on the SMB server.
  - For a source SMB server, the following permissions are required: list and read.

    Members of the Backup Operators group are supported with a source SMB server.

  - For a target SMB server, the following permissions are required: list, read, and write.
- The file server must allow a data broker host to access the exports.
- SMB versions 1.0, 2.0, 2.1, 3.0 and 3.11 are supported.
- Grant the "Administrators" group with "Full Control" permissions to the source and target folders.

  If you don't grant this permission, then the data broker might not have sufficient permissions to get the ACLs on a file or directory. If this occurs, you'll receive the following error: "getxattr error 95"

SMB limitation for hidden directories and files

An SMB limitation affects hidden directories and files when syncing data between SMB servers. If any of the directories or files on the source SMB server were hidden through Windows, the hidden attribute isn't copied to the target SMB server.

SMB sync behavior due to case-insensitivity limitation

The SMB protocol is case-insensitive, which means uppercase and lowercase letters are treated as being the same. This behavior can result in overwritten files and directory copy errors, if a sync relationship includes an SMB server and data already exists on the target.

For example, let's say that there's a file named "a" on the source and a file named "A" on the target. When Cloud Sync copies the file named "a" to the target, file "A" is overwritten by file "a" from the source.

In the case of directories, let's say that there's a directory named "b" on the source and a directory named "B" on the target. When Cloud Sync tries to copy the directory named "b" to the target, Cloud Sync receives an error that says the directory already exists. As a result, Cloud Sync always fails to copy the directory named "b."

The best way to avoid this limitation is to ensure that you sync data to an empty directory.

Networking overview for Cloud Sync

Networking for Cloud Sync includes connectivity between the data broker group and the source and target locations, and an outbound internet connection from data brokers over port 443.

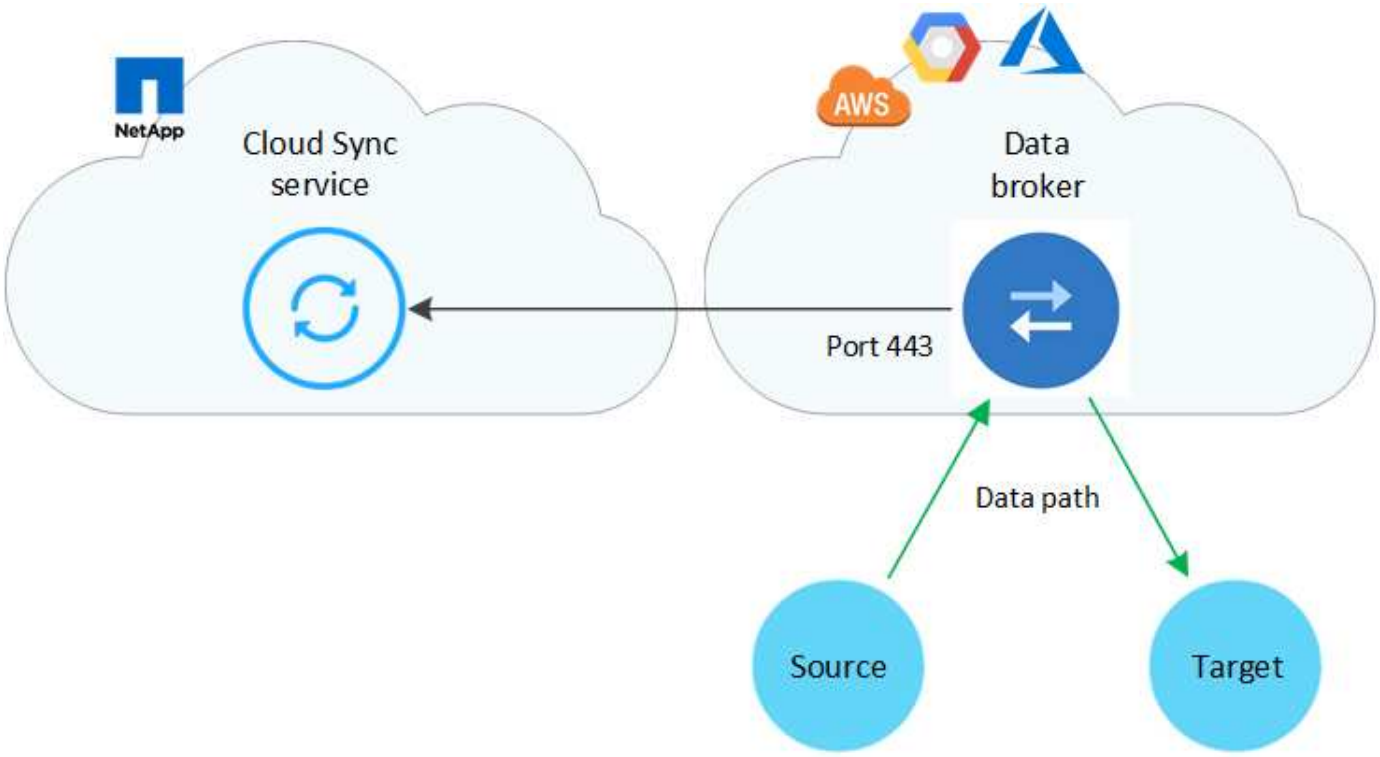
Data broker location

A data broker group consists of one or more data brokers installed in the cloud or on your premises.

Data broker in the cloud

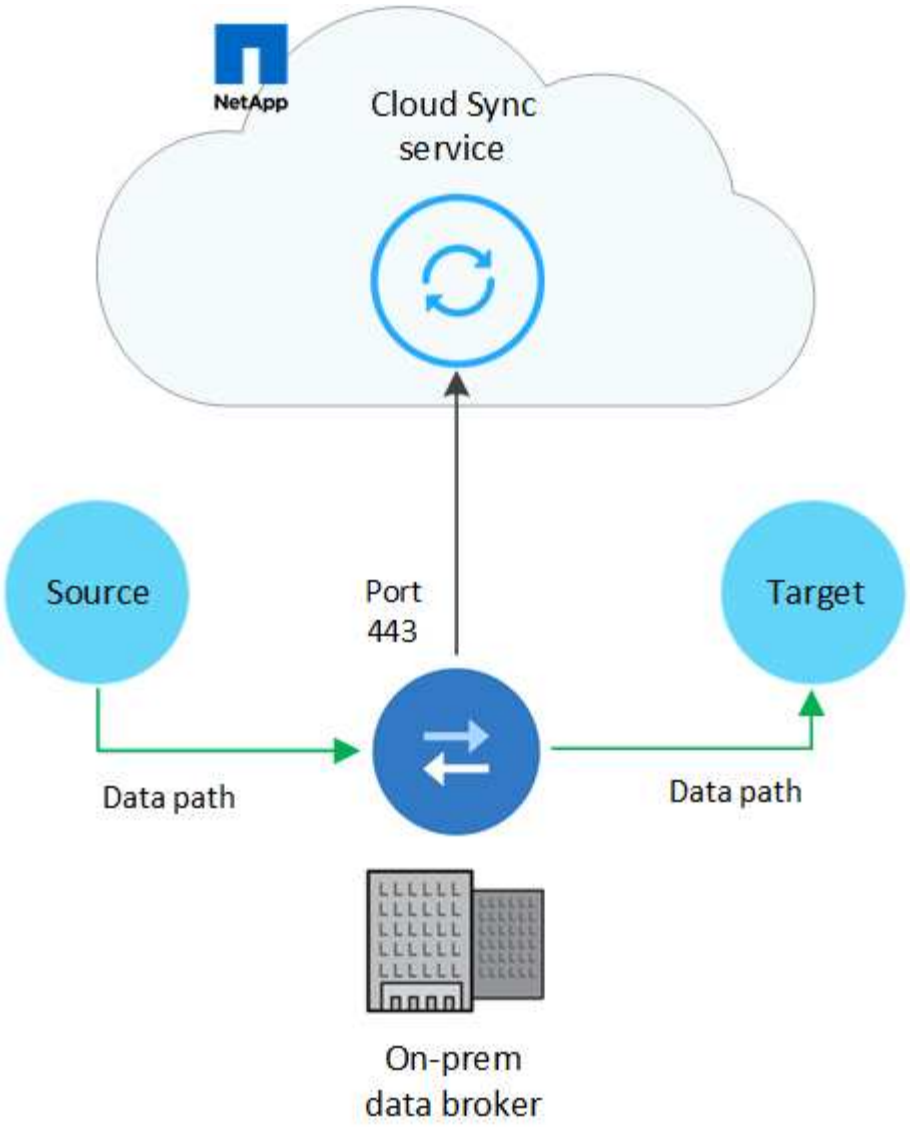
The following image shows a data broker running in the cloud, in either AWS, GCP, or Azure. The source and target can be in any location, as long as there's a connection to the data broker. For example, you might have a VPN connection from your data center to your cloud provider.

When Cloud Sync deploys the data broker in AWS, Azure, or GCP, it creates a security group that enables the required outbound communication.



Data broker on your premises

The following image shows the data broker running on-prem, in a data center. Again, the source and target can be in any location, as long as there's a connection to the data broker.



Networking requirements

- The source and target must have a network connection to the data broker group.

  For example, if an NFS server is in your data center and a data broker is in AWS, then you need a network connection (VPN or Direct Connect) from your network to the VPC.

- A data broker needs an outbound internet connection so it can poll the Cloud Sync service for tasks over port 443.
- NetApp recommends configuring the source, target, and data brokers to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Networking endpoints

The NetApp data broker requires outbound internet access over port 443 to communicate with the Cloud Sync service and to contact a few other services and repositories. Your local web browser also requires access to endpoints for certain actions. If you need to limit outbound connectivity, refer to the following list of endpoints when configuring your firewall for outbound traffic.

Data broker endpoints

A data broker contacts the following endpoints:

Endpoints	Purpose
https://olcentgbl.trafficmanager.net	To contact a repository for updating CentOS packages for the data broker host. This endpoint is contacted only if you manually install the data broker on a CentOS host.
https://rpm.nodesource.com https://registry.npmjs.org https://nodejs.org:	To contact repositories for updating Node.js, npm, and other 3rd party packages used in development.
https://tgz.pm2.io	To access a repository for updating PM2, which is a 3rd party package used to monitor Cloud Sync.
https://sqs.us-east-1.amazonaws.com https://kinesis.us-east-1.amazonaws.com	To contact the AWS services that Cloud Sync uses for operations (queuing files, registering actions, and delivering updates to the data broker).
https://s3.*region*.amazonaws.com For example: s3.us-east-2.amazonaws.com:443 See AWS documentation for a list of S3 endpoints	To contact Amazon S3 when a sync relationship includes an S3 bucket.
https://s3.us-east-1.amazonaws.com	When you download data broker logs from Cloud Sync, the data broker zips its logs directory and uploads the logs to a predefined S3 bucket in the us-east-1 region.
https://cf.cloudsync.netapp.com https://repo.cloudsync.netapp.com	To contact the Cloud Sync service.
https://support.netapp.com	To contact NetApp support when using a BYOL license for sync relationships.
https://fedoraproject.org	To install 7z on the data broker virtual machine during installation and updates. 7z is needed to send AutoSupport messages to NetApp technical support.
https://sts.amazonaws.com	To verify AWS credentials when the data broker is deployed in AWS or when it's deployed on your premises and AWS credentials are provided. The data broker contacts this endpoint during deployment, when it's updated, and when it's restarted.
https://cloudmanager.cloud.netapp.com https://netapp-cloud-account.auth0.com	To contact Cloud Data Sense when you use Data Sense to select the source files for a new sync relationship.

Web browser endpoints

Your web browser needs access to the following endpoint to download logs for troubleshooting purposes:

logs.cloudsync.netapp.com:443

How to install a data broker

Creating a new data broker in AWS

When you create a new data broker group, choose Amazon Web Services to deploy the data broker software on a new EC2 instance in a VPC. Cloud Sync guides you through the installation process, but the requirements and steps are repeated on this page to help you prepare for installation.

You also have the option to install the data broker on an existing Linux host in the cloud or on your premises. Learn more.

Supported AWS regions

All regions are supported except for the China regions.

Networking requirements

- The data broker needs an outbound internet connection so it can poll the Cloud Sync service for tasks over port 443.

  When Cloud Sync deploys the data broker in AWS, it creates a security group that enables the required outbound communication. Note that you can configure the data broker to use a proxy server during the installation process.

  If you need to limit outbound connectivity, see the list of endpoints that the data broker contacts.

- NetApp recommends configuring the source, target, and data broker to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Permissions required to deploy the data broker in AWS

The AWS user account that you use to deploy the data broker must have the permissions included in this NetApp-provided policy.

Requirements to use your own IAM role with the AWS data broker

When Cloud Sync deploys the data broker, it creates an IAM role for the data broker instance. You can deploy the data broker using your own IAM role, if you prefer. You might use this option if your organization has strict security policies.

The IAM role must meet the following requirements:

- The EC2 service must be allowed to assume the IAM role as a trusted entity.
- The permissions defined in this JSON file must be attached to the IAM role so the data broker can function properly.

Follow the steps below to specify the IAM role when deploying the data broker.

Creating the data broker

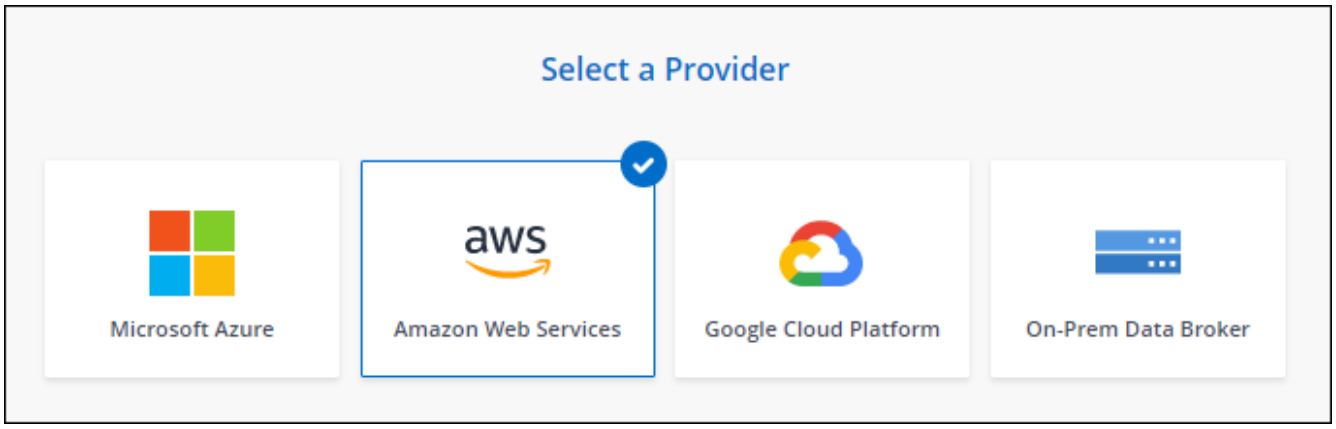
There are a few ways to create a new data broker. These steps describe how to install a data broker in AWS when creating a sync relationship.

Steps

1. Click **Create New Sync**.
2. On the **Define Sync Relationship** page, choose a source and target and click **Continue**.

   Complete the steps until you reach the **Data Broker Group** page.

3. On the **Data Broker Group** page, click **Create Data Broker** and then select **Amazon Web Services**.



4. Enter a name for the data broker and click **Continue**.
5. Enter an AWS access key so Cloud Sync can create the data broker in AWS on your behalf.

   The keys aren't saved or used for any other purposes.

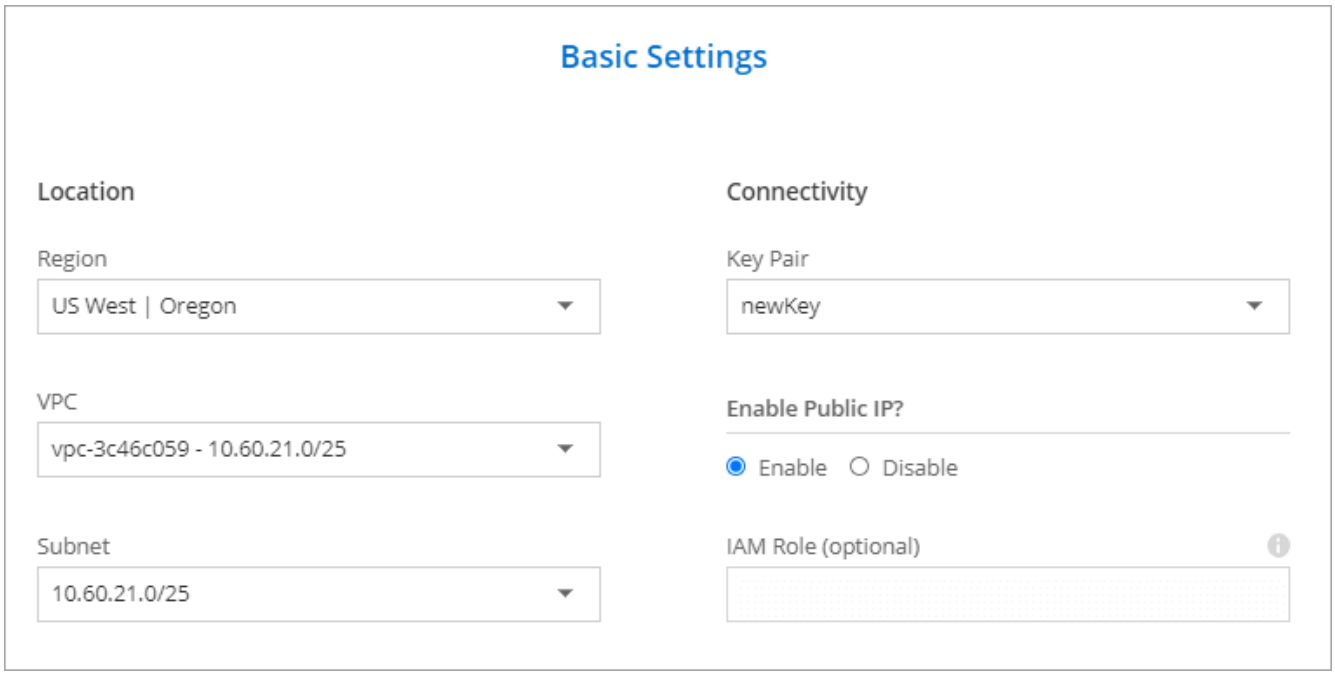
   If you'd rather not provide access keys, click the link at the bottom of the page to use a CloudFormation template instead. When you use this option, you don't need to provide credentials because you are logging in directly to AWS.

   The following video shows how to launch the data broker instance using a CloudFormation template:

   ► https://docs.netapp.com/us-en/occm//media/video_cloud_sync.mp4 *(video)*

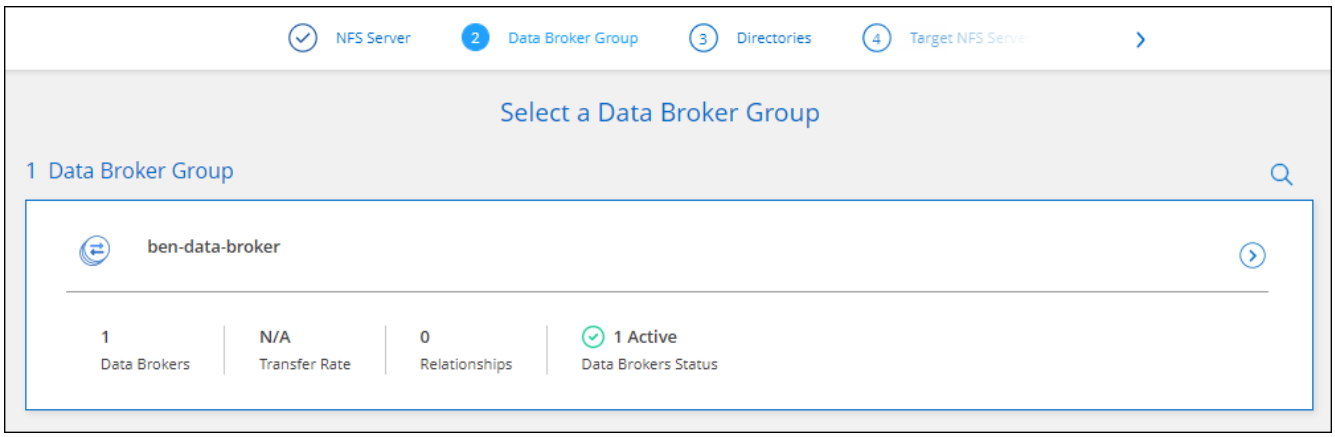
6. If you entered an AWS access key, select a location for the instance, select a key pair, choose whether to enable a public IP address, and then select an existing IAM role, or leave the field blank so Cloud Sync creates the role for you.

   If you choose your own IAM role, you'll need to provide the required permissions.



7. Specify a proxy configuration, if a proxy is required for internet access in the VPC.
8. After the data broker is available, click **Continue** in Cloud Sync.

   The following image shows a successfully deployed instance in AWS:



9. Complete the pages in the wizard to create the new sync relationship.

Result

You have deployed a data broker in AWS and created a new sync relationship. You can use this data broker group with additional sync relationships.

Details about the data broker instance

Cloud Sync creates a data broker in AWS using the following configuration.

Instance type

m5n.xlarge when available in the region, otherwise m5.xlarge

vCPUs

4

RAM

16 GB

Operating system

Amazon Linux 2

Disk size and type

10 GB GP2 SSD

Creating a new data broker in Azure

When you create a new data broker group, choose the Microsoft Azure to deploy the data broker software on a new virtual machine in a VNet. Cloud Sync guides you through the installation process, but the requirements and steps are repeated on this page to help you prepare for installation.

You also have the option to install the data broker on an existing Linux host in the cloud or on your premises. Learn more.

Supported Azure regions

All regions are supported except for the China, US Gov, and US DoD regions.

Networking requirements

- The data broker needs an outbound internet connection so it can poll the Cloud Sync service for tasks over port 443.

  When Cloud Sync deploys the data broker in Azure, it creates a security group that enables the required outbound communication.

  If you need to limit outbound connectivity, see the list of endpoints that the data broker contacts.

- NetApp recommends configuring the source, target, and data broker to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Authentication method

When you deploy the data broker, you'll need to choose an authentication method: a password or an SSH public-private key pair.

For help with creating a key pair, refer to Azure Documentation: Create and use an SSH public-private key pair for Linux VMs in Azure.

Creating the data broker

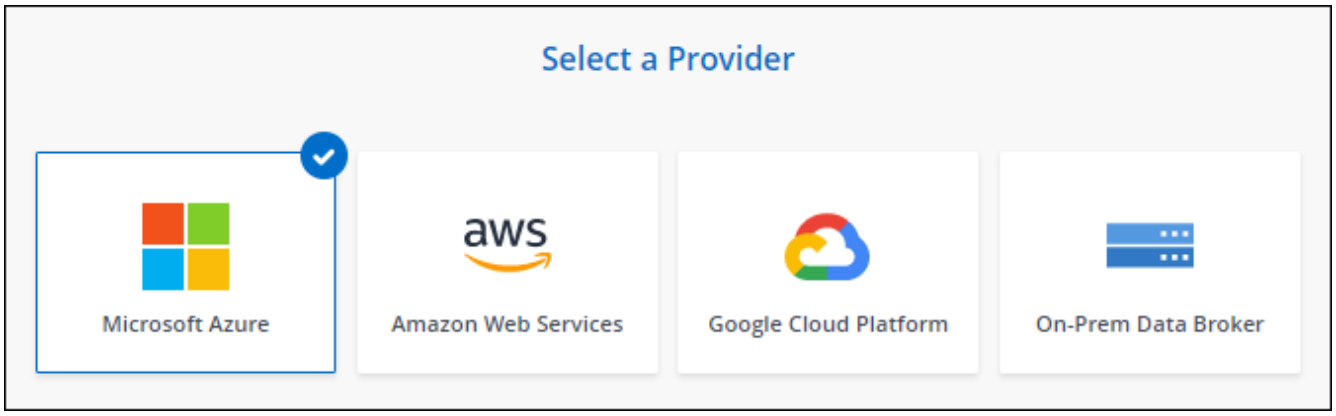
There are a few ways to create a new data broker. These steps describe how to install a data broker in Azure when you create a sync relationship.

Steps

1. Click **Create New Sync**.
2. On the **Define Sync Relationship** page, choose a source and target and click **Continue**.

   Complete the steps until you reach the **Data Broker Group** page.

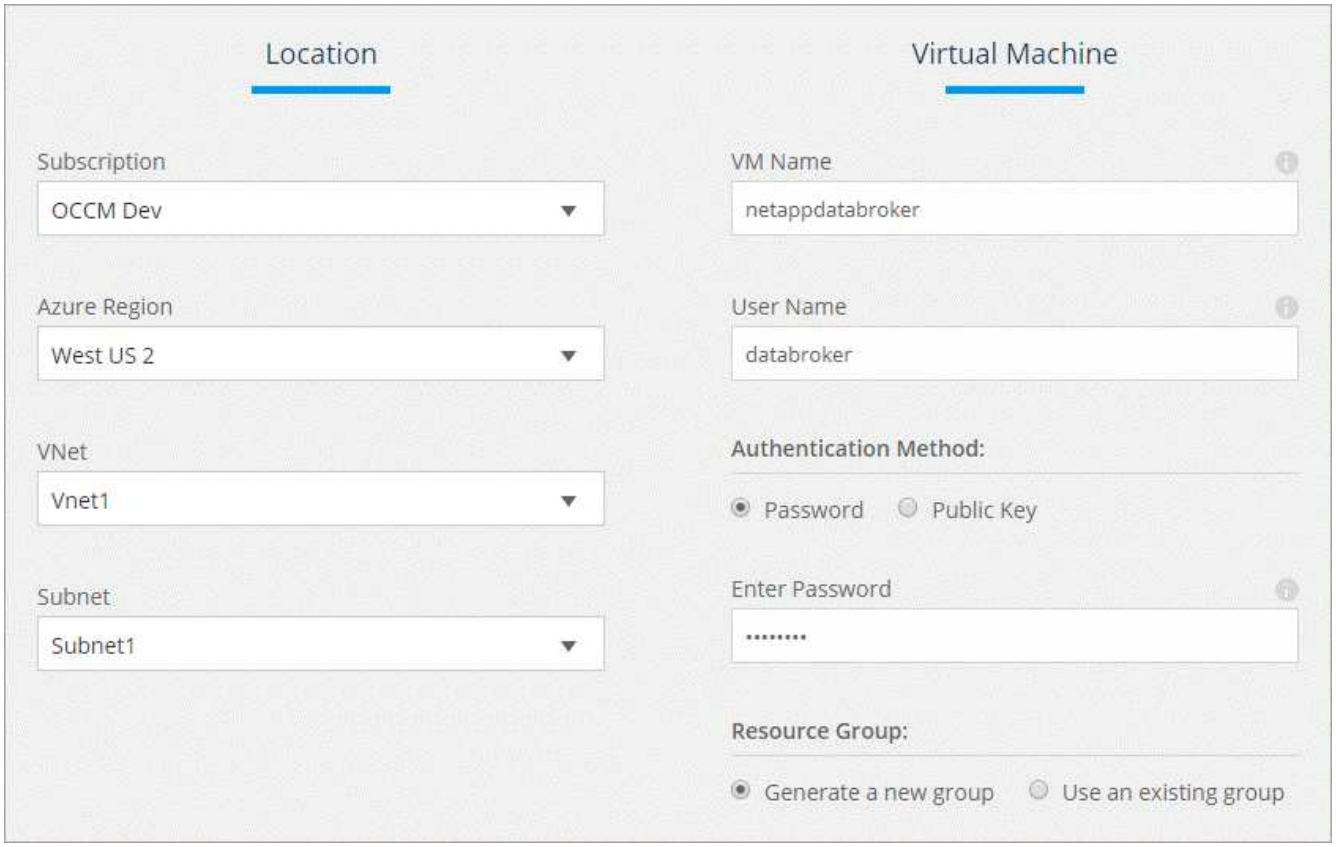
3. On the **Data Broker Group** page, click **Create Data Broker** and then select **Microsoft Azure**.



4. Enter a name for the data broker and click **Continue**.
5. If you're prompted, log in to your Microsoft account. If you're not prompted, click **Log in to Azure**.

   The form is owned and hosted by Microsoft. Your credentials are not provided to NetApp.

6. Choose a location for the data broker and enter basic details about the virtual machine.



7. Specify a proxy configuration, if a proxy is required for internet access in the VNet.

8. Click **Continue** and keep the page open until the deployment is complete.

   The process can take up to 7 minutes.

9. In Cloud Sync, click **Continue** once the data broker is available.
10. Complete the pages in the wizard to create the new sync relationship.

Result

You have deployed a data broker in Azure and created a new sync relationship. You can use this data broker with additional sync relationships.

Getting a message about needing admin consent?

If Microsoft notifies you that admin approval is required because Cloud Sync needs permission to access resources in your organization on your behalf, then you have two options:

1. Ask your AD admin to provide you with the following permission:

   In Azure, go to **Admin Centers > Azure AD > Users and Groups > User Settings** and enable **Users can consent to apps accessing company data on their behalf**.

2. Ask your AD admin to consent on your behalf to **CloudSync-AzureDataBrokerCreator** using the following URL (this is the admin consent endpoint):

   https://login.microsoftonline.com/{FILL HERE YOUR TENANT ID}/v2.0/adminconsent?client_id=8ee4ca3a-bafa-4831-97cc-5a38923cab85&redirect_uri=https://cloudsync.netapp.com&scope=https://management.azure.com/user_impersonationhttps://graph.microsoft.com/User.Read

   As shown in the URL, our app URL is https://cloudsync.netapp.com and the application client ID is 8ee4ca3a-bafa-4831-97cc-5a38923cab85.

Details about the data broker VM

Cloud Sync creates a data broker in Azure using the following configuration.

VM type
Standard DS4 v2

vCPUs
8

RAM
28 GB

Operating system
CentOS 7.7

Disk size and type
64 GB Premium SSD

Creating a new data broker in Google Cloud

When you create a new data broker group, choose Google Cloud Platform to deploy the data broker software on a new virtual machine instance in a Google Cloud VPC. Cloud Sync guides you through the installation process, but the requirements and steps are repeated on this page to help you prepare for installation.

You also have the option to install the data broker on an existing Linux host in the cloud or on your premises. Learn more.

Supported GCP regions

All regions are supported.

Networking requirements

- The data broker needs an outbound internet connection so it can poll the Cloud Sync service for tasks over port 443.

  When Cloud Sync deploys the data broker in GCP, it creates a security group that enables the required outbound communication.

  If you need to limit outbound connectivity, see the list of endpoints that the data broker contacts.

- NetApp recommends configuring the source, target, and data broker to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Permissions required to deploy the data broker in GCP

Ensure that the GCP user who deploys the data broker has the following permissions:

```
- compute.networks.list
- compute.regions.list
- deploymentmanager.deployments.create
- deploymentmanager.deployments.delete
- deploymentmanager.operations.get
- iam.serviceAccounts.list
```

Permissions required for the service account

When you deploy the data broker, you need to select a service account that has the following permissions:

```
- logging.logEntries.create
- resourcemanager.projects.get
- storage.buckets.get
- storage.buckets.list
- storage.objects.*
- iam.serviceAccounts.signJwt
```

The "iam.serviceAccounts.signJwt" permission is required only if you're planning to set up the data broker to use an external HashiCorp vault.

Creating the data broker

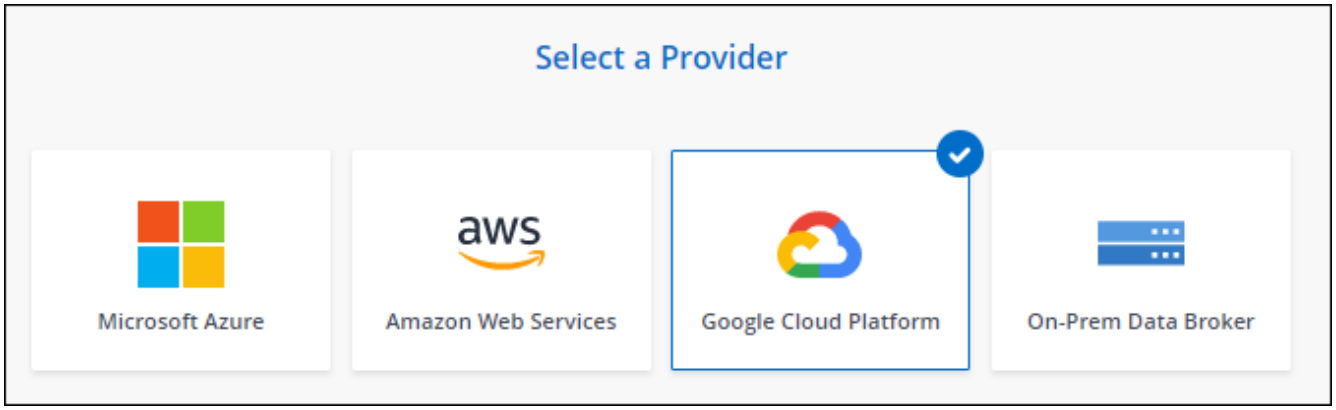
There are a few ways to create a new data broker. These steps describe how to install a data broker in Google Cloud when you create a sync relationship.

Steps

1. Click **Create New Sync**.
2. On the **Define Sync Relationship** page, choose a source and target and click **Continue**.

   Complete the steps until you reach the **Data Broker Group** page.

3. On the **Data Broker Group** page, click **Create Data Broker** and then select **Microsoft Azure**.

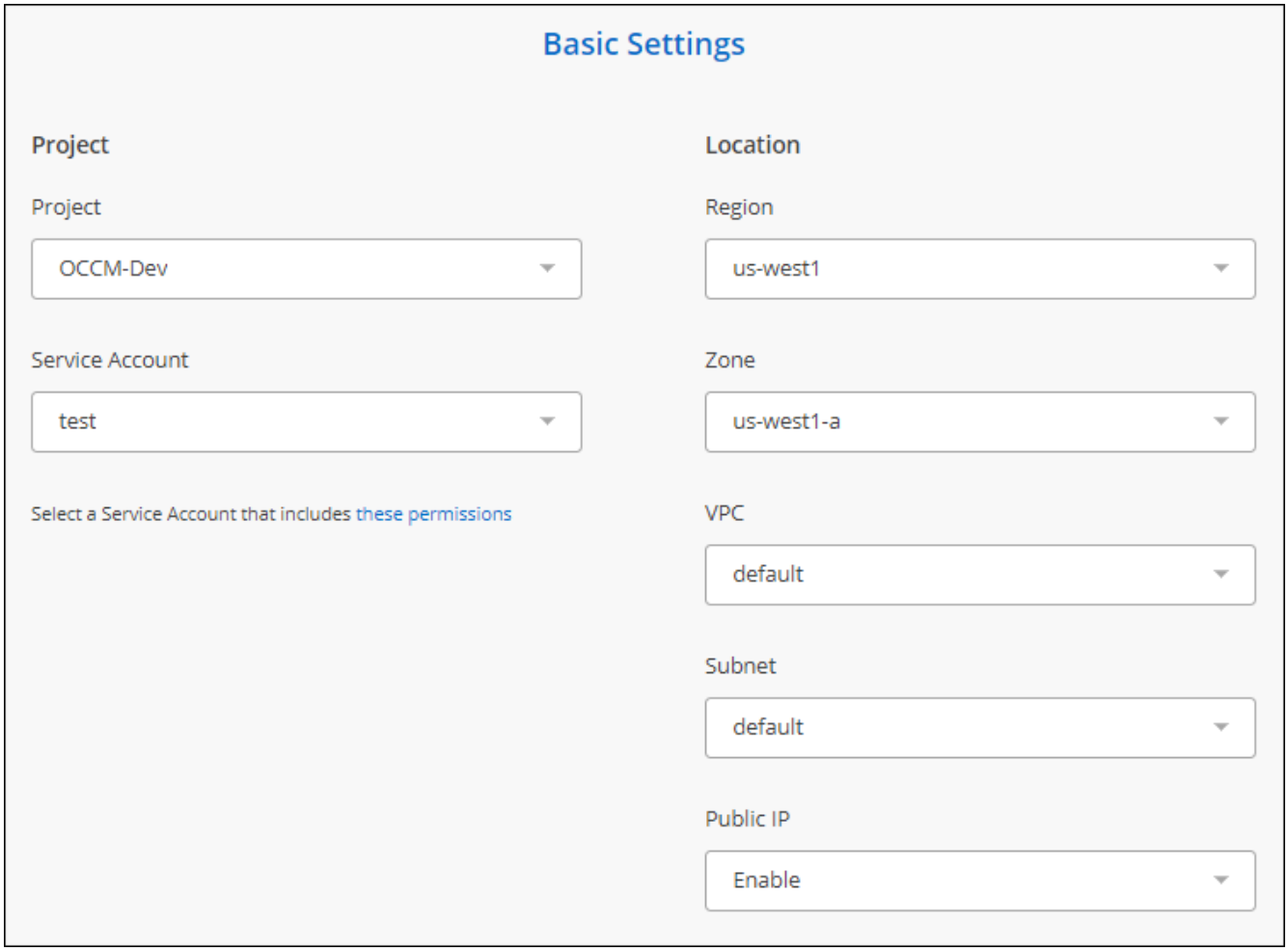


4. Enter a name for the data broker and click **Continue**.
5. If you're prompted, log in with your Google account.

   The form is owned and hosted by Google. Your credentials are not provided to NetApp.

6. Select a project and service account and then choose a location for the data broker, including whether you want to enable or disable a public IP address.

   If you don't enable a public IP address, then you'll need to define a proxy server in the next step.



7. Specify a proxy configuration, if a proxy is required for internet access in the VPC.

   If a proxy is required for internet access, then the proxy must be in Google Cloud and use the same service account as the data broker.

8. Once the data broker is available, click **Continue** in Cloud Sync.

   The instance takes approximately 5 to 10 minutes to deploy. You can monitor the progress from the Cloud Sync service, which automatically refreshes when the instance is available.

9. Complete the pages in the wizard to create the new sync relationship.

Result

You've deployed a data broker in Google Cloud and created a new sync relationship. You can use this data broker with additional sync relationships.

Providing permissions to use buckets in other Google Cloud projects

When you create a sync relationship and choose Google Cloud Storage as the source or target, Cloud Sync enables you to choose from the buckets that the data broker's service account has permissions to use. By default, this includes the buckets that are in the *same* project as the data broker service account. But you can choose buckets from *other* projects if you provide the required permissions.

Steps

1. Open the Google Cloud Platform console and load the Cloud Storage service.

2. Click the name of the bucket that you'd like to use as a source or target in a sync relationship.
3. Click **Permissions**.
4. Click **Add**.
5. Enter the name of the data broker's service account.
6. Select a role that provides the same permissions as shown above.
7. Click **Save**.

Result

When you set up a sync relationship, you can now choose that bucket as the source or target in the sync relationship.

Details about the data broker VM instance

Cloud Sync creates a data broker in Google Cloud using the following configuration.

Machine type
n1-standard-4

vCPUs
4

RAM
15 GB

Operating system
Red Hat Enterprise Linux 7.7

Disk size and type
20 GB HDD pd-standard

Installing the data broker on a Linux host

When you create a new data broker group, choose the On-Prem Data Broker option to install the data broker software on an on-premises Linux host, or on an existing Linux host in the cloud. Cloud Sync guides you through the installation process, but the requirements and steps are repeated on this page to help you prepare for installation.

Linux host requirements

- **Operating system**:
  - CentOS 7.0, 7.7, and 8.0

    CentOS Stream is not supported.
  - Red Hat Enterprise Linux 7.7 and 8.0
  - Ubuntu Server 20.04 LTS
  - SUSE Linux Enterprise Server 15 SP1

    The command `yum update all` must be run on the host before you install the data broker.

A Red Hat Enterprise Linux system must be registered with Red Hat Subscription Management. If it is not registered, the system cannot access repositories to update required 3rd party software during installation.

- **RAM**: 16 GB
- **CPU**: 4 cores
- **Free disk space**: 10 GB
- **SELinux**: We recommend that you disable SELinux on the host.

  SELinux enforces a policy that blocks data broker software updates and can block the data broker from contacting endpoints required for normal operation.

Networking requirements

- The Linux host must have a connection to the source and target.
- The file server must allow the Linux host to access the exports.
- Port 443 must be open on the Linux host for outbound traffic to AWS (the data broker constantly communicates with the Amazon SQS service).
- NetApp recommends configuring the source, target, and data broker to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

Enabling access to AWS

If you plan to use the data broker with a sync relationship that includes an S3 bucket, then you should prepare the Linux host for AWS access. When you install the data broker, you'll need to provide AWS keys for an AWS user that has programmatic access and specific permissions.

Steps

1. Create an IAM policy using this NetApp-provided policy. View AWS instructions.
2. Create an IAM user that has programmatic access. View AWS instructions.

   Be sure to copy the AWS keys because you need to specify them when you install the data broker software.

Enabling access to Google Cloud

If you plan to use the data broker with a sync relationship that includes a Google Cloud Storage bucket, then you should prepare the Linux host for GCP access. When you install the data broker, you'll need to provide a key for a service account that has specific permissions.

Steps

1. Create a GCP service account that has Storage Admin permissions, if you don't already have one.
2. Create a service account key saved in JSON format. View GCP instructions.

   The file should contain at least the following properties: "project_id", "private_key", and "client_email"

   When you create a key, the file gets generated and downloaded to your machine.

3. Save the JSON file to the Linux host.

Enabling access to Microsoft Azure

Access to Azure is defined per relationship by providing a storage account and a connection string in the Sync Relationship wizard.

Installing the data broker

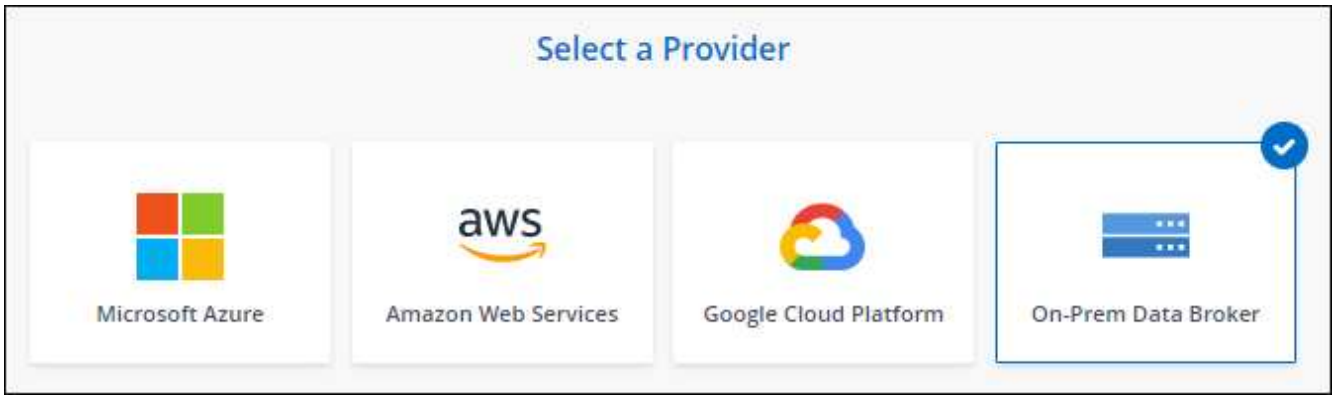
You can install a data broker on a Linux host when you create a sync relationship.

Steps

1. Click **Create New Sync**.
2. On the **Define Sync Relationship** page, choose a source and target and click **Continue**.

   Complete the steps until you reach the **Data Broker Group** page.

3. On the **Data Broker Group** page, click **Create Data Broker** and then select **On-Prem Data Broker**.



> Even though the option is labeled ***On-Prem*** **Data Broker**, it applies to a Linux host on your premises or in the cloud.

4. Enter a name for the data broker and click **Continue**.

   The instructions page loads shortly. You'll need to follow these instructions—they include a unique link to download the installer.

5. On the instructions page:
   a. Select whether to enable access to **AWS**, **Google Cloud**, or both.
   b. Select an installation option: **No proxy**, **Use proxy server**, or **Use proxy server with authentication**.
   c. Use the commands to download and install the data broker.

      The following steps provide details about each possible installation option. Follow the instructions page to get the exact command based on your installation option.

   d. Download the installer:
      - No proxy:

        ```
        curl <URI> -o data_broker_installer.sh
        ```

      - Use proxy server:

```
curl <URI> -o data_broker_installer.sh -x <proxy_host>:<proxy_port>
```

- Use proxy server with authentication:

```
curl <URI> -o data_broker_installer.sh -x
<proxy_username>:<proxy_password>@<proxy_host>:<proxy_port>
```

URI

Cloud Sync displays the URI of the installation file on the instructions page, which loads when you follow the prompts to deploy the On-Prem Data Broker. That URI isn't repeated here because the link is generated dynamically and can be used only once. Follow these steps to obtain the URI from Cloud Sync.

e. Switch to superuser, make the installer executable and install the software:

Each command listed below includes parameters for AWS access and GCP access. Follow the instructions page to get the exact command based on your installation option.

- No proxy configuration:

```
sudo -s
chmod +x data_broker_installer.sh
./data_broker_installer.sh -a <aws_access_key> -s <aws_secret_key> -g
<absolute_path_to_the_json_file>
```

- Proxy configuration:

```
sudo -s
chmod +x data_broker_installer.sh
./data_broker_installer.sh -a <aws_access_key> -s <aws_secret_key> -g
<absolute_path_to_the_json_file> -h <proxy_host> -p <proxy_port>
```

- Proxy configuration with authentication:

```
sudo -s
chmod +x data_broker_installer.sh
./data_broker_installer.sh -a <aws_access_key> -s <aws_secret_key> -g
<absolute_path_to_the_json_file> -h <proxy_host> -p <proxy_port> -u
<proxy_username> -w <proxy_password>
```

AWS keys

These are the keys for the user that you should have prepared following these steps. The AWS keys are stored on the data broker, which runs in your on-premises or cloud network. NetApp doesn't use the keys outside of the data broker.

JSON file

This is the JSON file that contains a service account key that you should have prepared following these steps.

6. Once the data broker is available, click **Continue** in Cloud Sync.
7. Complete the pages in the wizard to create the new sync relationship.

Create sync relationships

When you create a sync relationship, the Cloud Sync service copies files from the source to the target. After the initial copy, the service syncs any changed data every 24 hours.

Before you can create some types of sync relationships, you'll first need to create a working environment in Cloud Manager.

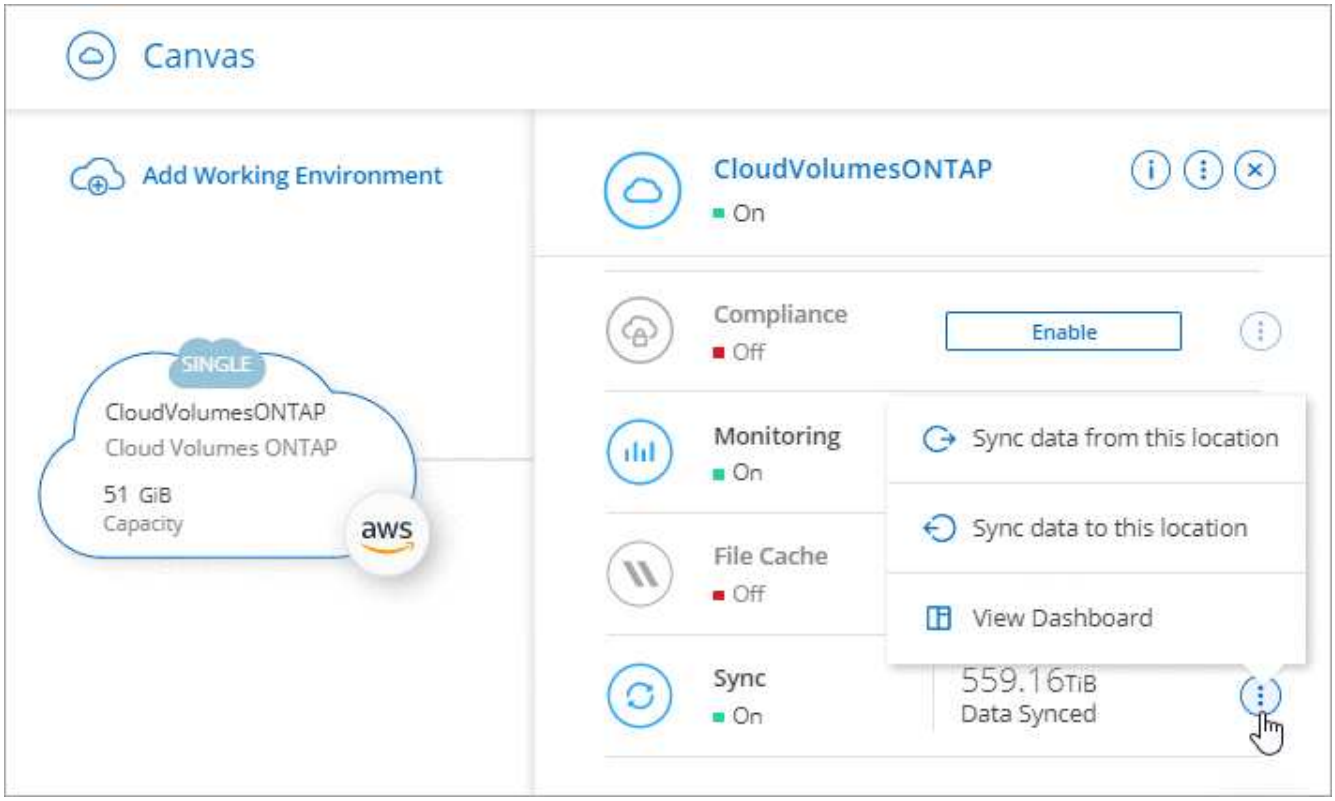
Create sync relationships for specific types of working environments

If you want to create sync relationships for any of the following, then you first need to create or discover the working environment:

- Amazon FSx for ONTAP
- Azure NetApp Files
- Cloud Volumes ONTAP
- On-prem ONTAP clusters

Steps

1. Create or discover the working environment.
    - Create an Amazon FSx for ONTAP working environment
    - Setting up and discovering Azure NetApp Files
    - Launching Cloud Volumes ONTAP in AWS
    - Launching Cloud Volumes ONTAP in Azure
    - Launching Cloud Volumes ONTAP in GCP
    - Adding existing Cloud Volumes ONTAP systems
    - Discovering ONTAP clusters
2. Click **Canvas**.
3. Select a working environment that matches any of the types listed above.
4. Select the action menu next to Sync.



5. Select **Sync data from this location** or **Sync data to this location** and follow the prompts to set up the sync relationship.

Create other types of sync relationships

Use these steps to sync data to or from a supported storage type other than Amazon FSx for ONTAP, Azure NetApp Files, Cloud Volumes ONTAP, or on-prem ONTAP clusters. The steps below provide an example that shows how to set up a sync relationship from an NFS server to an S3 bucket.

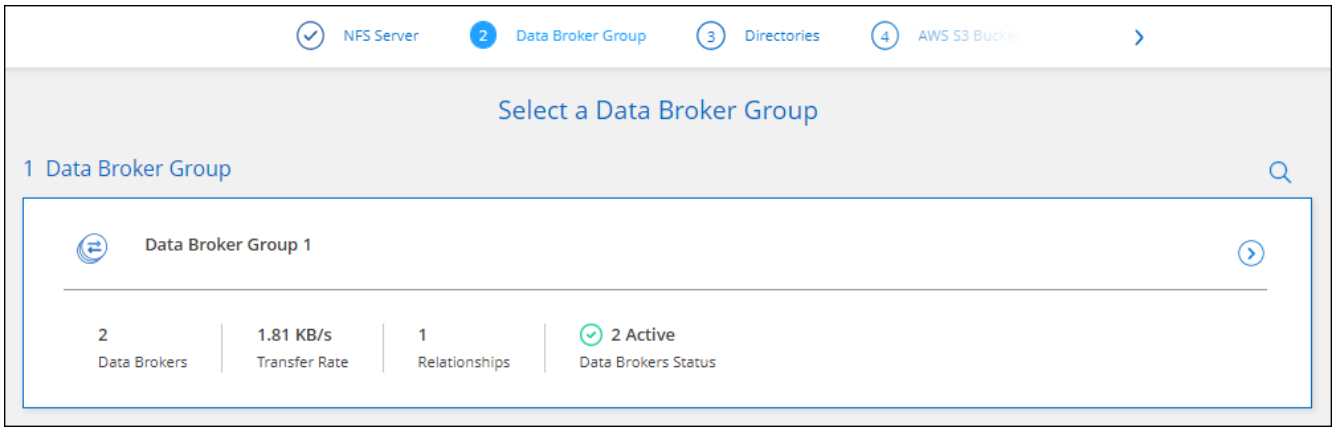
1. In Cloud Manager, click **Sync**.
2. On the **Define Sync Relationship** page, choose a source and target.

   The following steps provide an example of how to create a sync relationship from an NFS server to an S3 bucket.

3. On the **NFS Server** page, enter the IP address or fully qualified domain name of the NFS server that you want to sync to AWS.
4. On the **Data Broker Group** page, follow the prompts to create a data broker virtual machine in AWS, Azure, or Google Cloud Platform, or to install the data broker software an existing Linux host.

   For more details, refer to the following pages:

   - Create a data broker in AWS
   - Create a data broker in Azure
   - Create a data broker in Google Cloud
   - Installing the data broker on a Linux host
5. After you install the data broker, click **Continue**.

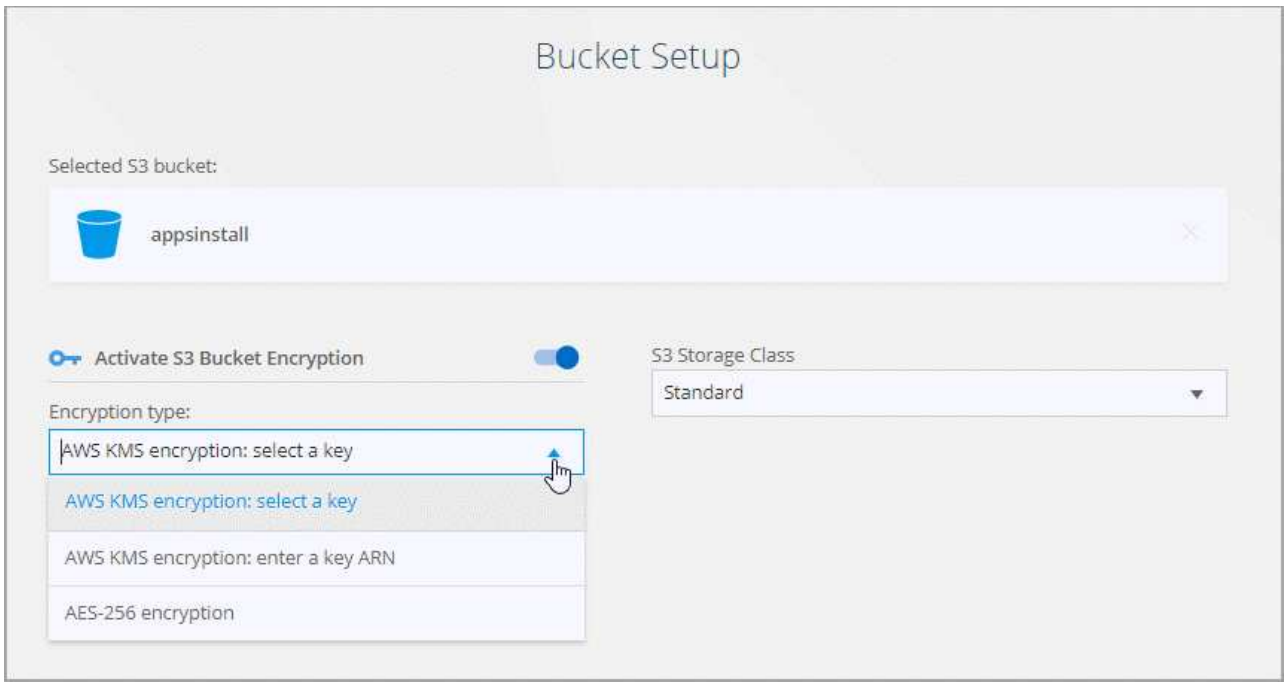


6. On the **Directories** page, select a top-level directory or subdirectory.

   If Cloud Sync is unable to retrieve the exports, click **Add Export Manually** and enter the name of an NFS export.

   If you want to sync more than one directory on the NFS server, then you must create additional sync relationships after you are done.

7. On the **AWS S3 Bucket** page, select a bucket:
   - Drill down to select an existing folder within the bucket or to select a new folder that you create inside the bucket.
   - Click **Add to the list** to select an S3 bucket that is not associated with your AWS account. Specific permissions must be applied to the S3 bucket.
8. On the **Bucket Setup** page, set up the bucket:
   - Choose whether to enable S3 bucket encryption and then select an AWS KMS key, enter the ARN of a KMS key, or select AES-256 encryption.
   - Select an S3 storage class. View the supported storage classes.



9. On the **Settings** page, define how source files and folders are synced and maintained in the target location:

Schedule

Choose a recurring schedule for future syncs or turn off the sync schedule. You can schedule a relationship to sync data as often as every 1 minute.

Retries

Define the number of times that Cloud Sync should retry to sync a file before skipping it.

Compare By

Choose whether Cloud Sync should compare certain attributes when determining whether a file or directory has changed and should be synced again.

Even if you uncheck these attributes, Cloud Sync still compares the source to the target by checking the paths, file sizes, and file names. If there are any changes, then it syncs those files and directories.

You can choose to enable or disable Cloud Sync from comparing the following attributes:

- **mtime**: The last modified time for a file. This attribute isn't valid for directories.
- **uid**, **gid**, and **mode**: Permission flags for Linux.

Copy for Objects

Enable this option to copy object storage metadata and tags. If a user changes the metadata on the source, Cloud Sync copies this object in the next sync, but if a user changes the tags on the source (and not the data itself), Cloud Sync doesn't copy the object in the next sync.

You can't edit this option after you create the relationship.

Copying tags is supported with sync relationships that include an S3-compatible endpoint (S3, StorageGRID, or IBM Cloud Object Storage).

Copying metadata is supported with "cloud-to-cloud" relationships between any of the following endpoints:

- AWS S3
- Azure Blob
- Google Cloud Storage
- IBM Cloud Object Storage
- StorageGRID

Recently Modified Files

Choose to exclude files that were recently modified prior to the scheduled sync.

Delete Files on Source

Choose to delete files from the source location after Cloud Sync copies the files to the target location. This option includes the risk of data loss because the source files are deleted after they're copied.

If you enable this option, you also need to change a parameter in the local.json file on the data broker. Open the file and update it as follows:

```
{
"workers":{
"transferrer":{
"delete-on-source": true
}
}
}
```

Delete Files on Target

Choose to delete files from the target location, if they were deleted from the source. The default is to never delete files from the target location.

File Types

Define the file types to include in each sync: files, directories, and symbolic links.

Exclude File Extensions

Specify file extensions to exclude from the sync by typing the file extension and pressing **Enter**. For example, type *log* or *.log* to exclude *.log files. A separator isn't required for multiple extensions. The following video provides a short demo:

► https://docs.netapp.com/us-en/occm//media/video_file_extensions.mp4 *(video)*

File Size

Choose to sync all files regardless of their size or just files that are in a specific size range.

Date Modified

Choose all files regardless of their last modified date, files modified after a specific date, before a specific date, or between a time range.

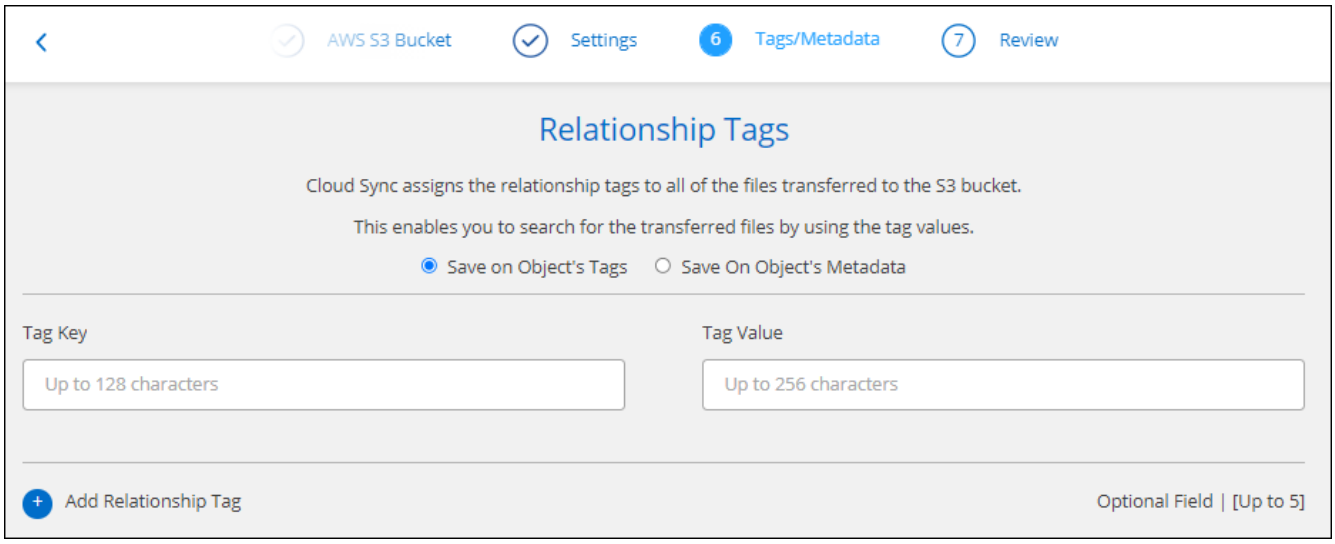
Date Created

When an SMB server is the source, this setting enables you to sync files that were created after a specific date, before a specific date, or between a specific time range.

ACL - Access Control List

Copy ACLs from an SMB server by enabling a setting when you create a relationship or after you create a relationship.

10. On the **Tags/Metadata** page, choose whether to save a key-value pair as a tag on all files transferred to the S3 bucket or to assign a metadata key-value pair on all files.



This same feature is available when syncing data to StorageGRID and IBM Cloud Object Storage. For Azure and Google Cloud Storage, only the metadata option is available.

11. Review the details of the sync relationship and then click **Create Relationship**.

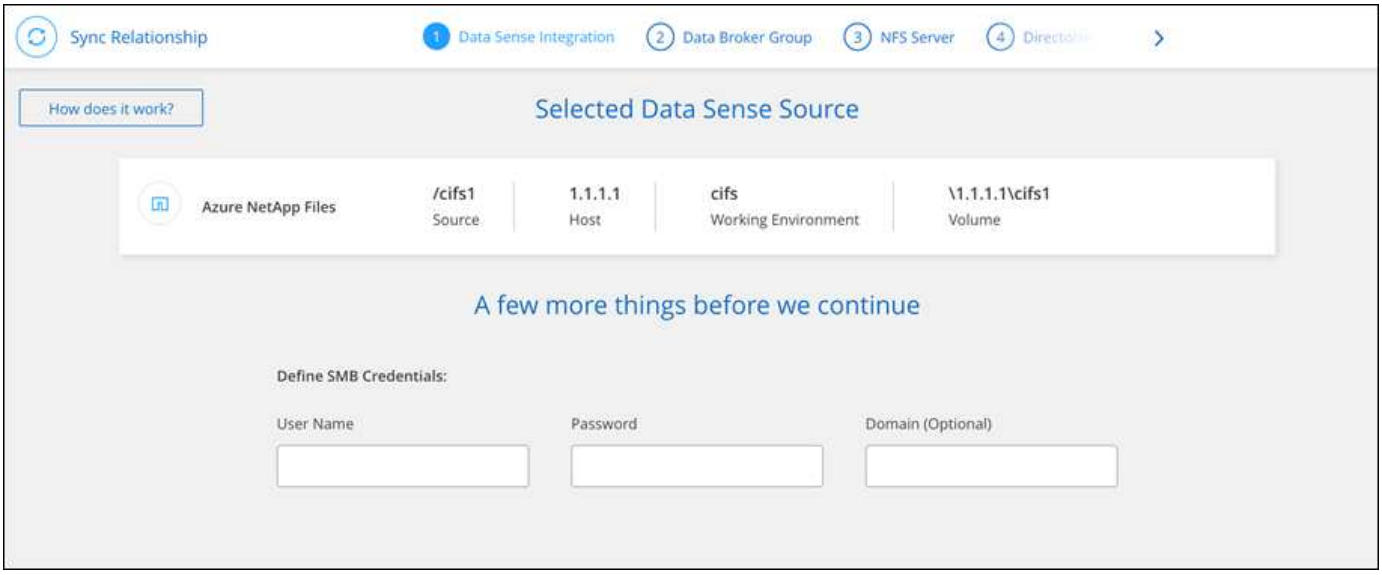
Result

Cloud Sync starts syncing data between the source and target.

Create sync relationships from Cloud Data Sense

Cloud Sync is integrated with Cloud Data Sense. From within Data Sense, you can select the source files that you'd like to sync to a target location using Cloud Sync.

After you initiate a data sync from Cloud Data Sense, all of the source information is contained in a single step and only requires you to enter a few key details. You then choose the target location for the new sync relationship.



Learn how to start a sync relationship from Cloud Data Sense.

Paying for sync relationships after your free trial ends

There are two ways to pay for sync relationships after your 14-day free trial ends. The first option is to subscribe from AWS or Azure to pay-as-you-go or to pay annually. The second option is to purchase licenses directly from NetApp.

You can subscribe from either the AWS Marketplace or the Azure Marketplace. You can't subscribe from both.

You have the option to use licenses from NetApp with a Marketplace subscription. For example, if you have 25 sync relationships, you can pay for the first 20 sync relationships using a license and then pay-as-you-go from AWS or Azure with the remaining 5 sync relationships.

Learn more about how licenses work.

What if I don't immediately pay after my free trial ends?

You won't be able to create any additional relationships. Existing relationships are not deleted, but you cannot make any changes to them until you subscribe or enter a license.

Subscribing from AWS

AWS enables you to pay-as-you-go or to pay annually.

Steps to pay-as-you-go

1. Click **Sync > Licensing**.
2. Select **AWS**
3. Click **Subscribe** and then click **Continue**.
4. Subscribe from the AWS Marketplace, and then log back in to the Cloud Sync service to complete the registration.

   The following video shows the process:

   ► https://docs.netapp.com/us-en/occm//media/video_cloud_sync_registering.mp4 *(video)*

Steps to pay annually

1. Go to the AWS Marketplace page.
2. Click **Continue to Subscribe**.
3. Select your contract options and click **Create contract**.

Subscribing from Azure

Azure enables you to pay-as-you-go or to pay annually.

What you'll need

An Azure user account that has Contributor or Owner permissions in the relevant subscription.

Steps

1. Click **Sync > Licensing**.
2. Select **Azure**.

3. Click **Subscribe** and then click **Continue**.
4. In the Azure portal, click **Create**, select your options, and click **Subscribe**.

   Select **Monthly** to pay by the hour, or **Yearly** to pay for a year up front.

5. When deployment is complete, click the name of the SaaS resource in the notification pop-up.
6. Click **Configure Account** to return to Cloud Sync.

   The following video shows the process:

   ► https://docs.netapp.com/us-en/occm//media/video_cloud_sync_registering_azure.mp4 *(video)*

Purchasing licenses from NetApp and adding them to Cloud Sync

To pay for your sync relationships up front, you must purchase one or more licenses and add them to the Cloud Sync service.

What you'll need

You'll need the serial number for your license and the user name and password for the NetApp Support Site account that the license is associated with.

Steps

1. Purchase a license by contacting NetApp.
2. In Cloud Manager, click **Sync > Licensing**.
3. Click **Add License** and add the required information:
   a. Enter the serial number.
   b. Select the NetApp Support Site account that is associated with the license that you're adding:
      - If your account was already added to Cloud Manager, select it from the drop-down list.
      - If your account wasn't added yet, click **Add NSS Credentials**, enter the user name and password, click **Register**, and then select it from the drop-down list.
   c. Click **Add**.

Updating a license

If you extended a Cloud Sync license that you purchased from NetApp, the new expiration date won't update automatically in Cloud Sync. You need to add the license again to refresh the expiration date.

Steps

1. In Cloud Manager, click **Sync > Licensing**.
2. Click **Add License** and add the required information:
   a. Enter the serial number.
   b. Select the NetApp Support Site account that is associated with the license that you're adding.
   c. Click **Add**.

Result

Cloud Sync updates the existing license with the new expiration date.

Tutorials

Copying ACLs from SMB shares

Cloud Sync can copy access control lists (ACLs) between a source SMB share and a target SMB share, or from a source SMB share to object storage (except for ONTAP S3). If needed, you also have the option to manually preserve ACLs between SMB shares by using robocopy.

Cloud Sync doesn't support copying ACLs back from object storage to SMB shares.

Choices

- Set up Cloud Sync to automatically copy ACLs
- Manually copy the ACLs between SMB shares

Setting up Cloud Sync to copy ACLs from an SMB server

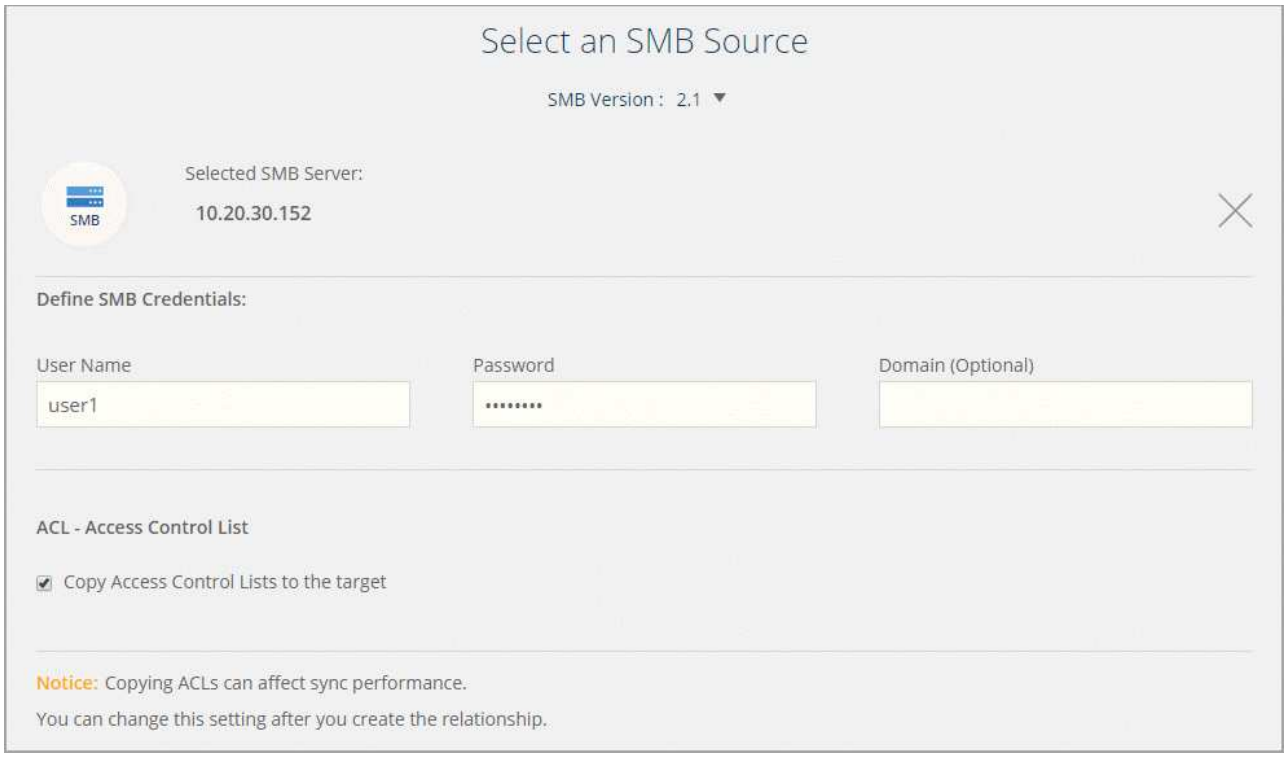
Copy ACLs from an SMB server by enabling a setting when you create a relationship or after you create a relationship.

What you'll need

This feature works with *any* type of data broker: the AWS, Azure, Google Cloud Platform, or on-prem data broker. The on-prem data broker can run any supported operating system.

Steps for a new relationship

1. From Cloud Sync, click **Create New Sync**.
2. Drag and drop **SMB Server** to the source, choose an SMB server or object storage as the target, and click **Continue**.
3. On the **SMB Server** page:
    a. Enter a new SMB server or select an existing server and click **Continue**.
    b. Enter credentials for the SMB server.
    c. Select **Copy Access Control Lists to the target** and click **Continue**.



4. Follow the remaining prompts to create the sync relationship.

   When you copy ACLs from SMB to object storage, you can choose to copy the ACLs to the object's tags or on the object's metadata, depending on the target. For Azure and Google Cloud Storage, only the metadata option is available.

   The following screenshot shows an example of the step where you can make this choice.

Steps for an existing relationship

1. Hover over the sync relationship and click the action menu.
2. Click **Settings**.
3. Select **Copy Access Control Lists to the target**.
4. Click **Save Settings**.

Result

When syncing data, Cloud Sync preserves the ACLs between the source and target SMB shares, or from a source SMB share to object storage.

Manually copying ACLs between SMB shares

You can manually preserve ACLs between SMB shares by using the Windows robocopy command.

Steps

1. Identify a Windows host that has full access to both SMB shares.
2. If either of the endpoints require authentication, use the **net use** command to connect to the endpoints from the Windows host.

   You must perform this step before you use robocopy.

3. From Cloud Sync, create a new relationship between the source and target SMB shares or sync an existing relationship.
4. After the data sync is complete, run the following command from the Windows host to sync the ACLs and ownership:

   ```
   robocopy /E /COPY:SOU /secfix [source] [target] /w:0 /r:0 /XD ~snapshots
   /UNILOG:"[logfilepath]
   ```

   Both *source* and *target* should be specified using the UNC format. For example: \\<server>\<share>\<path>

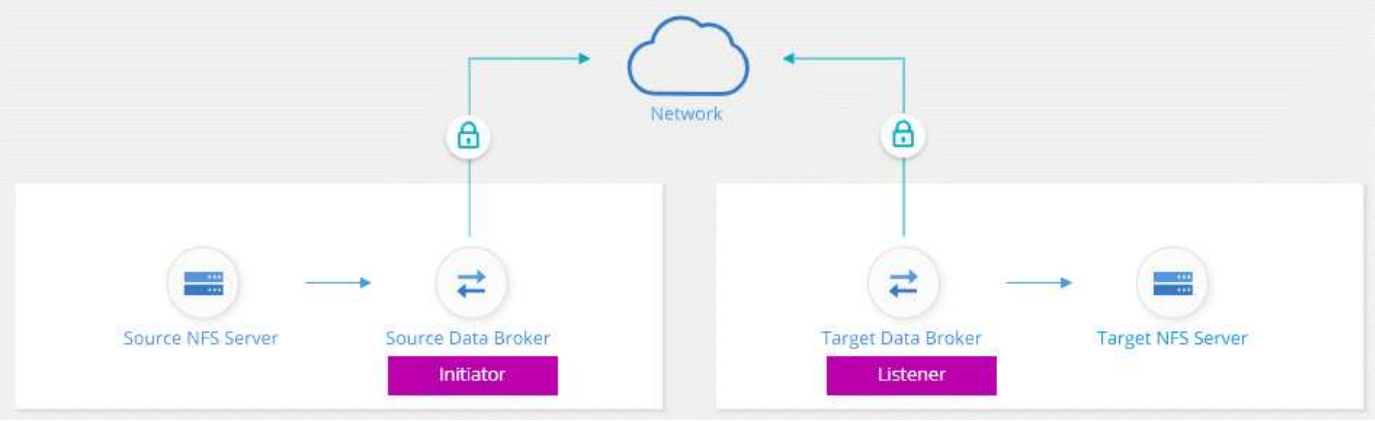
Syncing NFS data using data-in-flight encryption

If your business has strict security policies, you can sync NFS data using data-in-flight encryption. This feature is supported from an NFS server to another NFS server and from Azure NetApp Files to Azure NetApp Files.

For example, you might want to sync data between two NFS servers that are in different networks. Or you might need to securely transfer data on Azure NetApp Files across subnets or regions.

How data-in-flight encryption works

Data-in-flight encryption encrypts NFS data when it's sent over the network between two data brokers. The following image shows a relationship between two NFS servers and two data brokers:



One data broker functions as the *initiator*. When it's time to sync data, it sends a connection request to the other data broker, which is the *listener*. That data broker listens for requests on port 443. You can use a different port, if needed, but be sure to check that the port is not in use by another service.

For example, if you sync data from an on-premises NFS server to a cloud-based NFS server, you can choose which data broker listens for the connection requests and which sends them.

Here's how in-flight encryption works:

1. After you create the sync relationship, the initiator starts an encrypted connection with the other data broker.
2. The source data broker encrypts data from the source using TLS 1.3.
3. It then sends the data over the network to the target data broker.
4. The target data broker decrypts the data before sending it to the target.
5. After the initial copy, the service syncs any changed data every 24 hours. If there is data to sync, the process starts with the initiator opening an encrypted connection with the other data broker.

   If you prefer to sync data more frequently, you can change the schedule after you create the relationship.

Supported NFS versions

- For NFS servers, data-in-flight encryption is supported with NFS versions 3, 4.0, 4.1, and 4.2.
- For Azure NetApp Files, data-in-flight encryption is supported with NFS versions 3 and 4.1.

Proxy server limitation

If you create an encrypted sync relationship, the encrypted data is sent over HTTPS and isn't routable through a proxy server.

What you'll need to get started

Be sure to have the following:

- Two NFS servers that meet source and target requirements or Azure NetApp Files in two subnets or regions.
- The IP addresses or fully qualified domain names of the servers.
- Network locations for two data brokers.

  You can select an existing data broker but it must function as the initiator. The listener data broker must be a *new* data broker.

  If you want to use an existing data broker group, the group must have only one data broker. Multiple data brokers in a group aren't supported with encrypted sync relationships.

  If you have not yet deployed a data broker, review the data broker requirements. Because you have strict security policies, be sure to review the networking requirements, which includes outbound traffic from port 443 and the internet endpoints that the data broker contacts.

  - Review AWS installation
  - Review Azure installation

- Review GCP installation
- Review Linux host installation

Syncing NFS data using data-in-flight encryption

Create a new sync relationship between two NFS servers or between Azure NetApp Files, enable the in-flight encryption option, and follow the prompts.

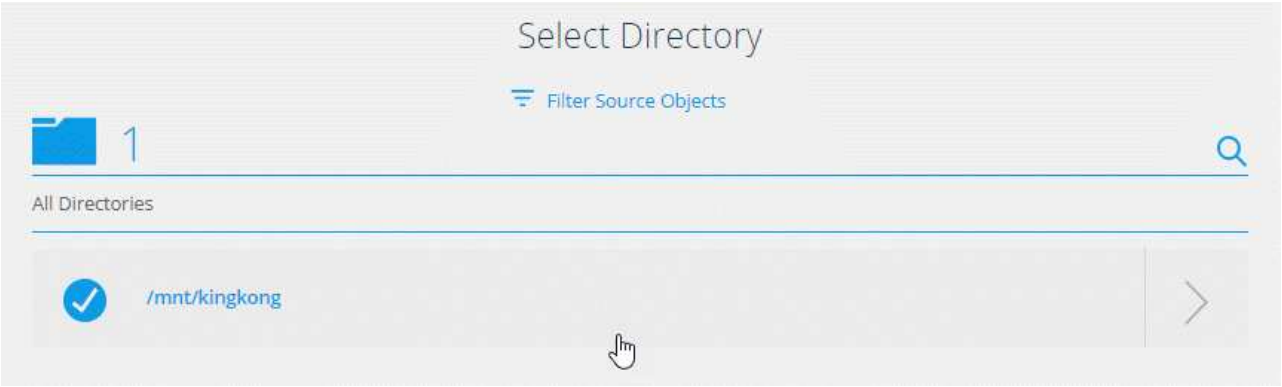
Steps

1. Click **Create New Sync**.
2. Drag and drop **NFS Server** to the source and target locations or **Azure NetApp Files** to the source and target locations and select **Yes** to enable data-in-flight encryption.
3. Follow the prompts to create the relationship:
   a. **NFS Server/Azure NetApp Files**: Choose the NFS version and then specify a new NFS source or select an existing server.
   b. **Define Data Broker Functionality**: Define which data broker *listens* for connection requests on a port and which one *initiates* the connection. Make your choice based on your networking requirements.
   c. **Data Broker**: Follow the prompts to add a new source data broker or select an existing data broker.

      Note the following:

      - If you want to use an existing data broker group, the group must have only one data broker. Multiple data brokers in a group aren't supported with encrypted sync relationships.
      - If the source data broker acts as the listener, then it must be a new data broker.
      - If you need a new data broker, Cloud Sync prompts you with the installation instructions. You can deploy the data broker in the cloud or download an installation script for your own Linux host.
   d. **Directories**: Choose the directories that you want to sync by selecting all directories, or by drilling down and selecting a subdirectory.

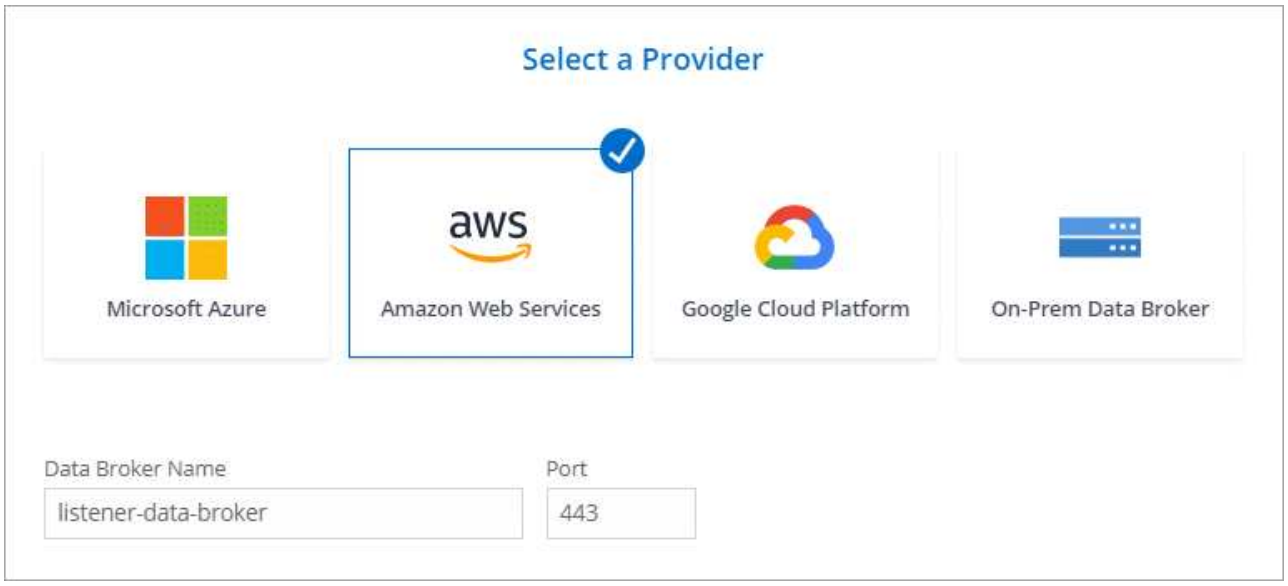
      Click **Filter Source Objects** to modify settings that define how source files and folders are synced and maintained in the target location.



   e. **Target NFS Server/Target Azure NetApp Files**: Choose the NFS version and then enter a new NFS target or select an existing server.
   f. **Target Data Broker**: Follow the prompts to add a new source data broker or select an existing data broker.

      If the target data broker acts as the listener, then it must be a new data broker.

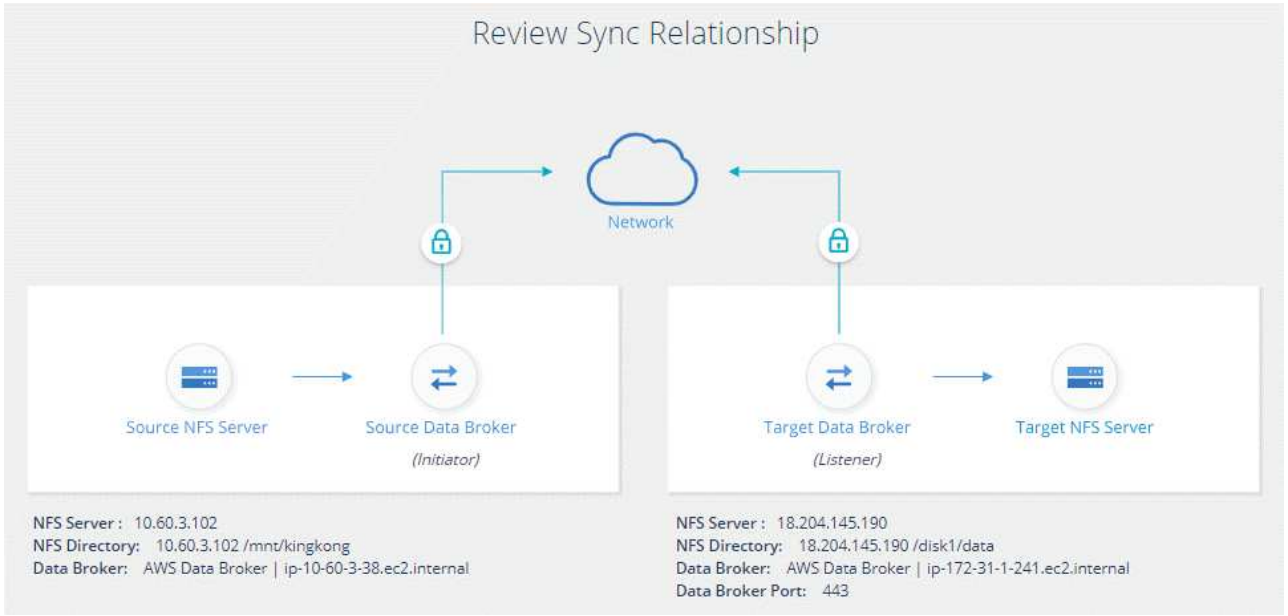
Here's an example of the prompt when the target data broker functions as the listener. Notice the option to specify the port.



g. **Target Directories**: Select a top-level directory, or drill down to select an existing subdirectory or to create a new folder inside an export.

h. **Settings**: Define how source files and folders are synced and maintained in the target location.

i. **Review**: Review the details of the sync relationship and then click **Create Relationship**.



Result

Cloud Sync starts creating the new sync relationship. When it's done, click **View in Dashboard** to view details about the new relationship.

Setting up a data broker group to use an external HashiCorp Vault

When you create a sync relationship that requires Amazon S3, Azure, or Google Cloud credentials, you need to specify those credentials through the Cloud Sync user interface or API. An alternative is to set up the data

broker group to access the credentials (or *secrets*) directly from an external HashiCorp Vault.

This feature is supported through the Cloud Sync API with sync relationships that require Amazon S3, Azure, or Google Cloud credentials.

1 Prepare the vault

Prepare the vault to supply credentials to the data broker group by setting up the URLs. The URLs to the secrets in the vault must end with *Creds*.

2 Prepare the data broker group

Prepare the data broker group to fetch credentials from the external vault by modifying the local config file for each data broker in the group.

3 Create a sync relationship using the API

Now that everything is set up, you can send an API call to create a sync relationship that uses your vault to get the secrets.

Preparing the vault

You'll need to provide Cloud Sync with the URL to the secrets in your vault. Prepare the vault by setting up those URLs. You need to set up URLs to the credentials for each source and target in the sync relationships that you plan to create.

The URL must be set up as follows:

```
/<path>/<requestid>/<endpoint-protocol>Creds
```

Path

The prefix path to the secret. This can be any value that's unique to you.

Request ID

A request ID that you need to generate. You'll need to provide the ID in one of the headers in the API POST request when you create the sync relationship.

Endpoint protocol

One of the following protocols, as defined in the post relationship v2 documentation: S3, AZURE, or GCP (each must be in uppercase).

Creds

The URL must end with *Creds*.

Examples

The following examples show URLs to secrets.

Example for the full URL and path for source credentials

http://example.vault.com:8200/my-path/all-secrets/hb312vdasr2/S3Creds

As you can see in the example, the prefix path is */my-path/all-secrets/*, the request ID is *hb312vdasr2* and the source endpoint is S3.

Example for the full URL and path for target credentials

http://example.vault.com:8200/my-path/all-secrets/n32hcbnejk2/AZURECreds

The prefix path is */my-path/all-secrets/*, the request ID is *n32hcbnejk2*, and the target endpoint is Azure.

Preparing the data broker group

Prepare the data broker group to fetch credentials from the external vault by modifying the local config file for each data broker in the group.

Steps

1. SSH to a data broker in the group.
2. Edit the local.json file that resides in /opt/netapp/databroker/config.
3. Set enable to **true** and set the config parameter fields under *external-integrations.hashicorp* as follows:

enabled

- Valid values: true/false
- Type: Boolean
- Default value: false
- True: The data broker gets secrets from your own external HashiCorp Vault
- False: The data broker stores credentials in its local vault

url

- Type: string
- Value: The URL to your external vault

path

- Type: string
- Value: Prefix path to the secret with your credentials

Reject-unauthorized

- Determines if you want the data broker to reject unauthorized external vault
- Type: Boolean
- Default: false

auth-method

- The authentication method that the data broker should use to access credentials from the external vault
- Type: string
- Valid values: “aws-iam” / “role-app” / "gcp-iam"

role-name

- Type: string
- Your role name (in case you use aws-iam or gcp-iam)

Secretid & rootid

- Type: string (in case you use app-role)

Namespace

- Type: string
- Your namespace (X-Vault-Namespace header if needed)

4. Repeat these steps for any other data brokers in the group.

Example for aws-role authentication

```
{
            "external-integrations": {
                    "hashicorp": {
                            "enabled": true,
                            "url": "https://example.vault.com:8200",
                            "path": ""my-path/all-secrets",
                            "reject-unauthorized": false,
                            "auth-method": "aws-role",
                            "aws-role": {
                                  "role-name": "my-role"
                            }
                    }
          }
}
```

Example for gcp-iam authentication

```
{
"external-integrations": {
    "hashicorp": {
      "enabled": true,
      "url": http://ip-10-20-30-55.ec2.internal:8200,
      "path": "v1/secret",
      "namespace": "",
      "reject-unauthorized": true,
      "auth-method": "gcp-iam",
      "aws-iam": {
        "role-name": ""
      },
      "app-role": {
        "root_id": "",
        "secret_id": ""
      },
"gcp-iam": {
          "role-name": "my-iam-role"
      }
    }
  }
}
```

Setting up permissions when using gcp-iam authentication

If you're using the *gcp-iam* authentication method, then the data broker must have the following GCP permission:

```
- iam.serviceAccounts.signJwt
```

Learn more about GCP permission requirements for the data broker.

Creating a new sync relationship using secrets from the vault

Now that everything is set up, you can send an API call to create a sync relationship that uses your vault to get the secrets.

Post the relationship using the Cloud Sync REST API.

```
Headers:
Authorization: Bearer <user-token>
Content-Type: application/json
x-account-id: <accountid>
x-netapp-external-request-id-src: request ID as part of path for source
credentials
x-netapp-external-request-id-trg: request ID as part of path for target
credentials
Body: post relationship v2 body
```

- To obtain a user token and your Cloud Central account ID, refer to this page in the documentation.
- To build a body for your post relationship, refer to the relationships-v2 API call.

Example

Example for the POST request:

```
url: https://api.cloudsync.netapp.com/api/relationships-v2
headers:
"x-account-id": "CS-SasdW"
"x-netapp-external-request-id-src": "hb312vdasr2"
"Content-Type": "application/json"
"Authorization": "Bearer eyJhbGciOiJSUzI1NiIsInR5cCI6IkpXVCIsImtpZCI6Ik…"
Body:
{
"dataBrokerId": "5e6e111d578dtyuu1555sa60",
"source": {
        "protocol": "s3",
        "s3": {
                "provider": "sgws",
                "host": "1.1.1.1",
                "port": "443",
                "bucket": "my-source"
     },
"target": {
        "protocol": "s3",
        "s3": {
                "bucket": "my-target-bucket"
        }
    }
}
```

Managing sync relationships

You can manage sync relationships at any time by immediately syncing data, changing schedules, and more.

Performing an immediate data sync

Rather than wait for the next scheduled sync, you can press a button to immediately sync data between the source and target.

Steps

1. From the **Dashboard**, navigate to the sync relationship and click 
2. Click **Sync Now** and then click **Sync** to confirm.

Result

Cloud Sync starts the data sync process for the relationship.

Accelerating sync performance

Accelerate the performance of a sync relationship by adding an additional data broker to the group that manages the relationship. The additional data broker must be a *new* data broker.

How this works

If the data broker group manages other sync relationships, then the new data broker that you add to the group also accelerates the performance of those sync relationships.

For example, let's say you have three relationships:

- Relationship 1 is managed by data broker group A
- Relationship 2 is managed by data broker group B
- Relationship 3 is managed by data broker group A

You want to accelerate the performance of relationship 1 so you add a new data broker to data broker group A. Because group A also manages sync relationship 3, the sync performance for relationship is automatically accelerated as well.

Steps

1. Ensure that at least one of the existing data brokers in the relationship are online.
2. From the **Dashboard**, navigate to the sync relationship and click 
3. Click **Accelerate**.
4. Follow the prompts to create a new data broker.

Result

Cloud Sync adds the new data broker to the group. The performance of the next data sync should be accelerated.

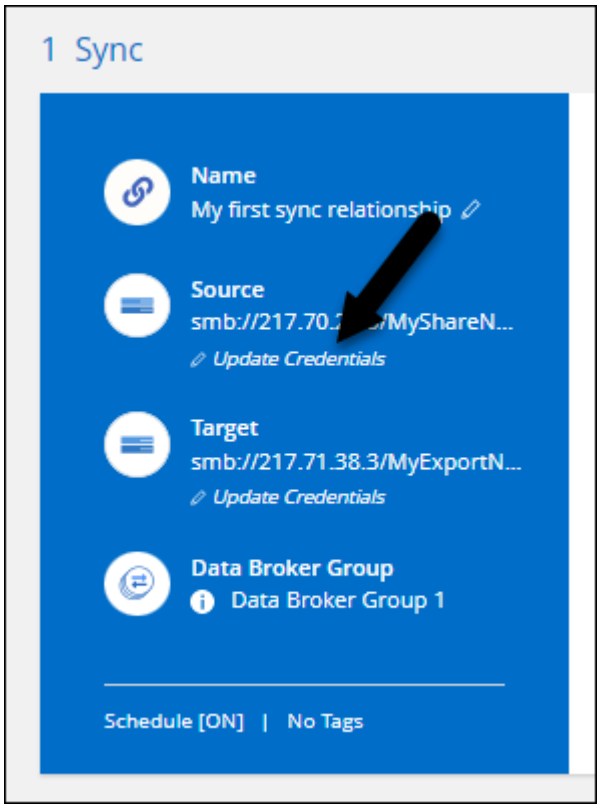
Updating credentials

You can update the data broker with the latest credentials of the source or target in an existing sync relationship. Updating the credentials can help if your security policies require you to update credentials on a periodic basis.

Updating credentials is supported with any source or target that Cloud Sync requires credentials for: Azure Blob, Box, IBM Cloud Object Storage, StorageGRID, ONTAP S3 Storage, SFTP, and SMB servers.

Steps

1. From the **Sync Dashboard**, go to a sync relationship that requires credentials and then click **Update Credentials**.



2. Enter the credentials and click **Update**.

   A note about SMB servers: if the domain is new, then you'll need to specify it when you update the credentials. If the domain hasn't changed, then you don't need to enter it again.

   If you entered a domain when you created the sync relationship, but you don't enter a new domain when you update the credentials, then Cloud Sync will keep using the original domain that you provided.

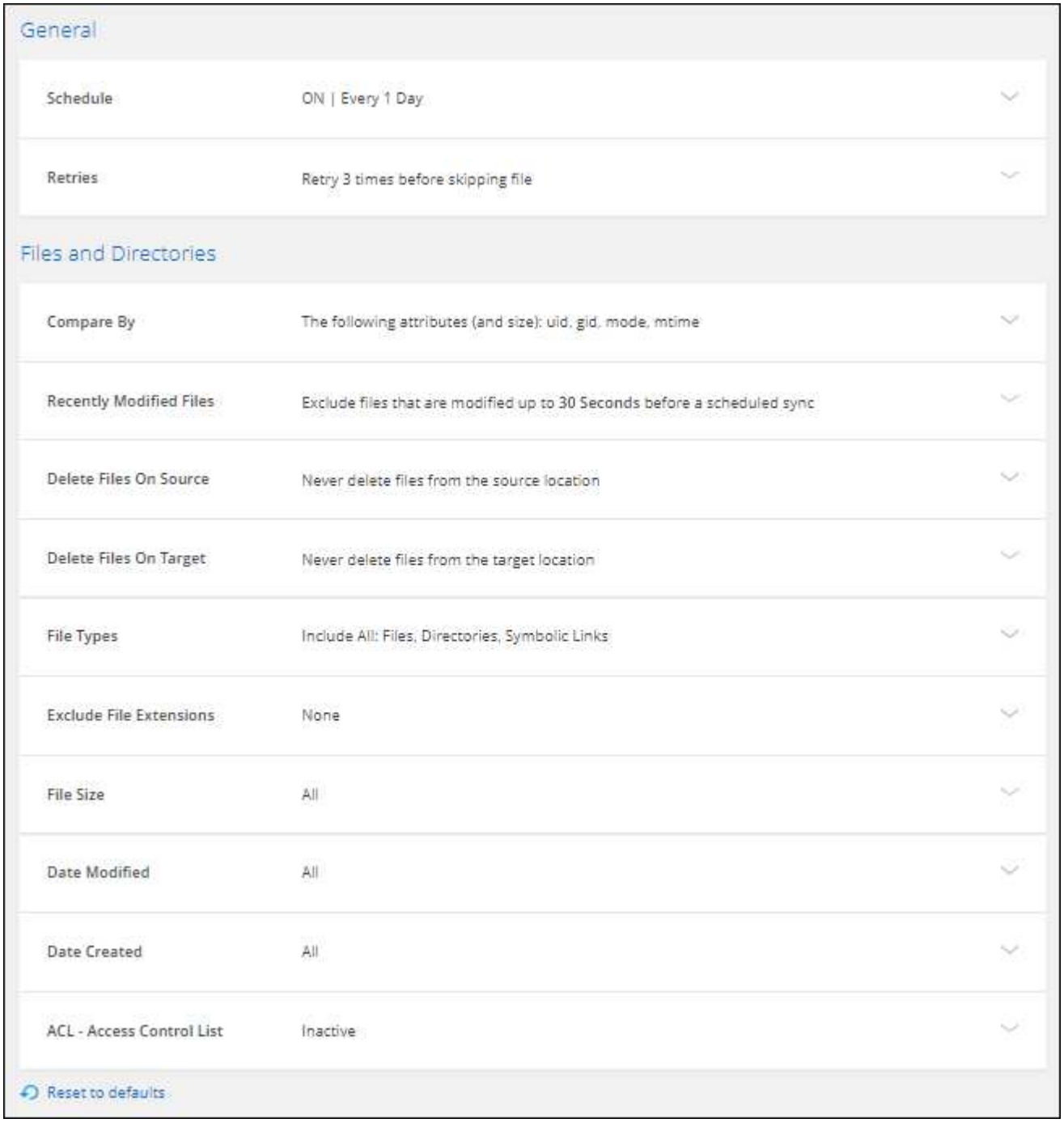
Result

Cloud Sync updates the credentials on the data broker. It can take up 10 minutes until the data broker starts using the updated credentials for data syncs.

Changing the settings for a sync relationship

Modify settings that define how source files and folders are synced and maintained in the target location.

1. From the **Dashboard**, navigate to the sync relationship and click 
2. Click **Settings**.
3. Modify any of the settings.



Here's a brief description of each setting:

Schedule

Choose a recurring schedule for future syncs or turn off the sync schedule. You can schedule a relationship to sync data as often as every 1 minute.

Retries

Define the number of times that Cloud Sync should retry to sync a file before skipping it.

Compare By

Choose whether Cloud Sync should compare certain attributes when determining whether a file or directory has changed and should be synced again.

Even if you uncheck these attributes, Cloud Sync still compares the source to the target by checking the paths, file sizes, and file names. If there are any changes, then it syncs those files and directories.

You can choose to enable or disable Cloud Sync from comparing the following attributes:

- **mtime**: The last modified time for a file. This attribute isn't valid for directories.
- **uid**, **gid**, and **mode**: Permission flags for Linux.

Copy for Objects

You can't edit this option after you create the relationship.

Recently Modified Files

Choose to exclude files that were recently modified prior to the scheduled sync.

Delete Files on Source

Choose to delete files from the source location after Cloud Sync copies the files to the target location. This option includes the risk of data loss because the source files are deleted after they're copied.

If you enable this option, you also need to change a parameter in the local.json file on the data broker. Open the file and update it as follows:

```
{
"workers":{
"transferrer":{
"delete-on-source": true
}
}
}
```

Delete Files on Target

Choose to delete files from the target location, if they were deleted from the source. The default is to never deletes files from the target location.

File Types

Define the file types to include in each sync: files, directories, and symbolic links.

Exclude File Extensions

Specify file extensions to exclude from the sync by typing the file extension and pressing **Enter**. For example, type *log* or *.log* to exclude *.log files. A separator isn't required for multiple extensions. The following video provides a short demo:

► https://docs.netapp.com/us-en/occm//media/video_file_extensions.mp4 *(video)*

File Size

Choose to sync all files regardless of their size or just files that are in a specific size range.

Date Modified

Choose all files regardless of their last modified date, files modified after a specific date, before a specific date, or between a time range.

Date Created

When an SMB server is the source, this setting enables you to sync files that were created after a specific date, before a specific date, or between a specific time range.

ACL - Access Control List

Copy ACLs from an SMB server by enabling a setting when you create a relationship or after you create a relationship.

4. Click **Save Settings**.

Result

Cloud Sync modifies the sync relationship with the new settings.

Deleting relationships

You can delete a sync relationship, if you no longer need to sync data between the source and target. This action doesn't delete the data broker group (or the individual data broker instances) and it does not delete data from the target.

Steps

1. From the **Dashboard**, navigate to the sync relationship and click 
2. Click **Delete** and then click **Delete** again to confirm.

Result

Cloud Sync deletes the sync relationship.

Manage data broker groups

A data broker group syncs data from a source location to a target location. At least one data broker is required in a group for each sync relationship that you create. Manage data broker groups by adding a new data broker to a group, by viewing information about groups, and more.

Data broker groups

Grouping data brokers together can help improve the performance of sync relationships.

Determining the number of data brokers in a group

In many cases, a single data broker can meet the performance requirements for a sync relationship. If it doesn't, you can accelerate sync performance by adding additional data brokers to the group. But you should first check other factors that can impact sync performance. Learn more about how to determine when multiple data brokers are required.

Groups can manage several relationships

A data broker group can manage one or more sync relationships at a time.

For example, let's say you have three relationships:

- Relationship 1 is managed by data broker group A
- Relationship 2 is managed by data broker group B
- Relationship 3 is managed by data broker group A

You want to accelerate the performance of relationship 1 so you add a new data broker to data broker group A. Because group A also manages sync relationship 3, the sync performance for relationship is automatically accelerated as well.

New data brokers only

You can only add new data brokers to a group. You can't add existing data brokers to a group.

Add a new data broker

There are several ways to create a new data broker:

- When creating a new sync relationship

  Learn how to create a new data broker when creating a sync relationship.

- From the **Manage Data Brokers** page by clicking **Add New Data Broker** which creates the data broker in a new group
- From the **Manage Data Brokers** page by creating a new data broker in an existing group

Things you should know

- You can't add data brokers to a group that manages an encrypted sync relationship.
- If you want to create a data broker in an existing group, the data broker must be an on-prem data broker or the same type of data broker.

  For example, if a group includes an AWS data broker, then you can create an AWS data broker or on-prem data broker in that group. You can't create an Azure data broker or GCP data broker because they aren't the same data broker type.

Steps to create a data broker in a new group

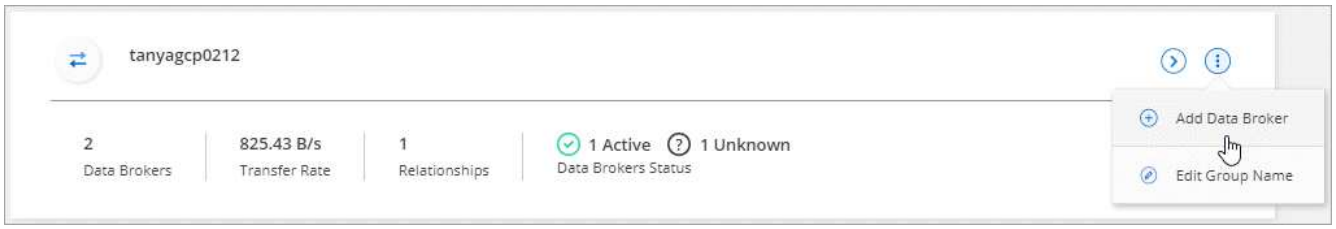
1. Click **Sync > Manage Data Brokers**.
2. Click **Add New Data Broker**.
3. Follow the prompts to create the data broker.

   For help, refer to the following pages:

   - Create a data broker in AWS
   - Create a data broker in Azure
   - Create a data broker in Google Cloud
   - Installing the data broker on a Linux host

Steps to create a data broker in an existing group

1. Click **Sync > Manage Data Brokers**.
2. Click the action menu and select **Add Data Broker**.



3. Follow the prompts to create the data broker in the group.

   For help, refer to the following pages:

   - Create a data broker in AWS
   - Create a data broker in Azure
   - Create a data broker in Google Cloud
   - Installing the data broker on a Linux host

Security recommendations

To ensure the security of your data broker machine, NetApp recommends the following:

- SSH should not permit X11 Forwarding
- SSH should not permit TCP connection forwarding
- SSH should not permit tunnels
- SSH should not accept client environment variables

These security recommendations can help prevent unauthorized connections to the data broker machine.

View a data broker's configuration

You might want to view details about a data broker to identify things like its host name, IP address, available CPU and RAM, and more.

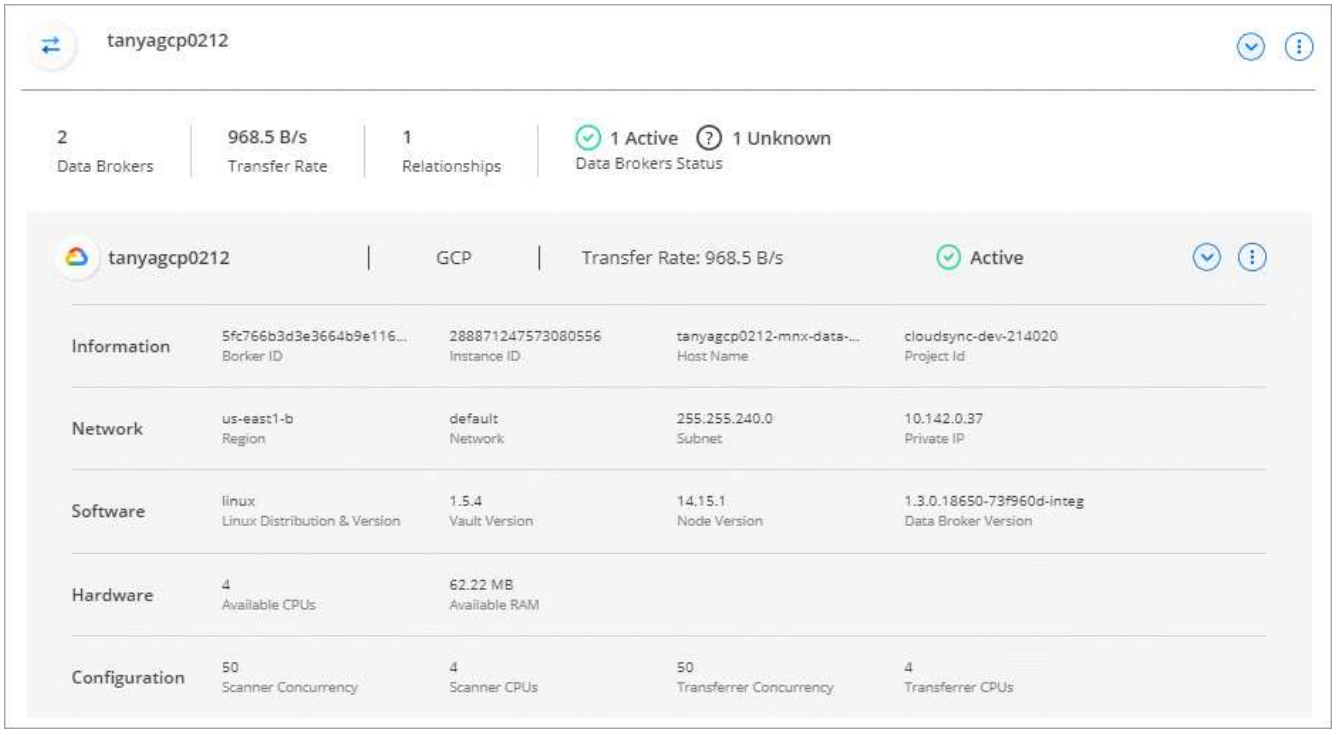
Cloud Sync provides the following details about a data broker:

- Basic information: Instance ID, host name, etc.
- Network: Region, network, subnet, private IP, etc.
- Software: Linux distribution, data broker version, etc.
- Hardware: CPU and RAM
- Configuration: Details about the data broker's two kinds of main processes—scanner and transferrer

> The scanner scans the source and target and decides what should be copied. The transferrer does the actual copying. NetApp personnel might use these configuration details to suggest actions that can optimize performance.

Steps

1. Click **Sync > Manage Data Brokers**.
2. Click to expand the list of data brokers in a group.
3. Click to view details about a data broker.



Remove a data broker from a group

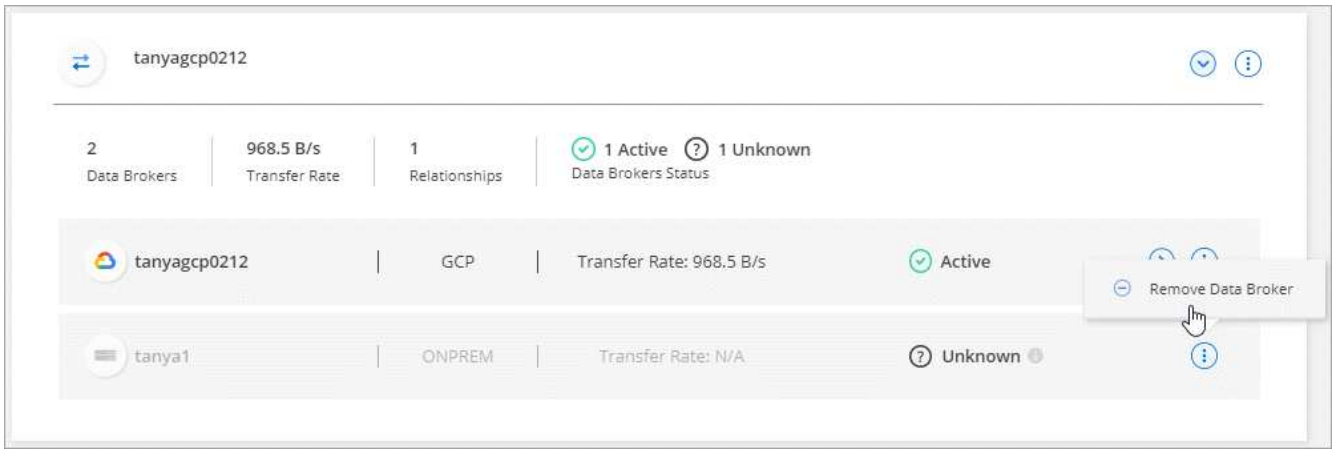
You might remove a data broker from a group if it's no longer needed or if the initial deployment failed. This action only deletes the data broker from Cloud Sync's records. You'll need to manually delete the data broker and any additional cloud resources yourself.

Things you should know

- Cloud Sync deletes a group when you remove the last data broker from the group.
- You can't remove the last data broker from a group if there is a relationship using that group.

Steps

1. Click **Sync > Manage Data Brokers**.
2. Click to expand the list of data brokers in a group.
3. Click the action menu for a data broker and select **Remove Data Broker**.



4. Click **Remove Data Broker**.

Result

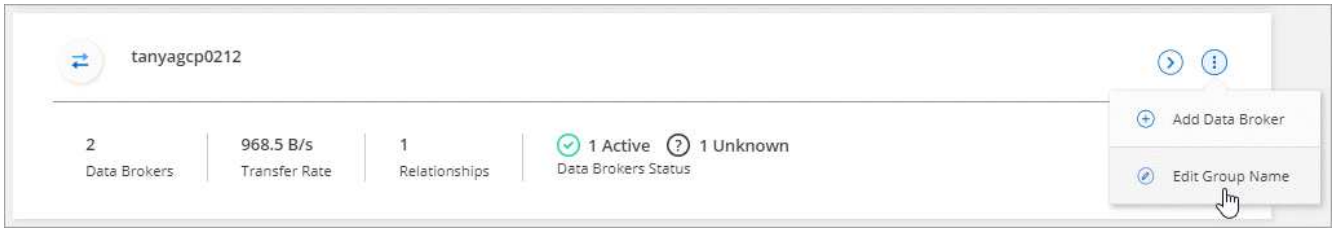
Cloud Sync removes the data broker from the group.

Edit a group's name

Change the name of a data broker group at any time.

Steps

1. Click **Sync > Manage Data Brokers**.
2. Click the action menu and select **Edit Group Name**.



3. Enter a new name and click **Save**.

Result

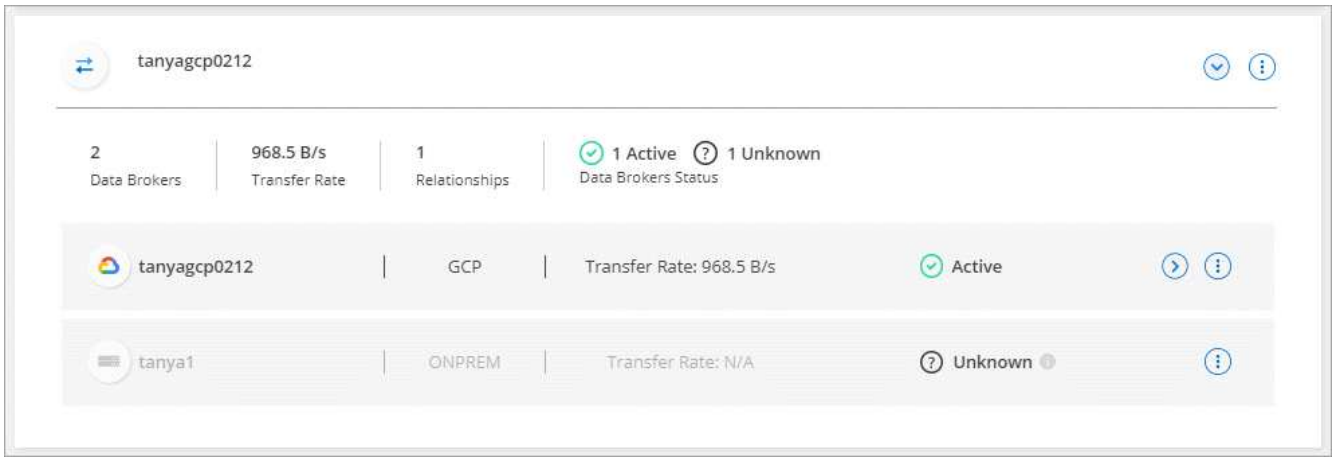
Cloud Sync updates the name of the data broker group.

Address issues with a data broker

Cloud Sync displays a status for each data broker that can help you troubleshoot issues.

Steps

1. Identify any data brokers that have a status of "Unknown" or "Failed."



2. Hover over the icon to see the failure reason.
3. Correct the issue.

   For example, you might need to simply restart the data broker if it's offline, or you might need to remove data broker if the initial deployment failed.

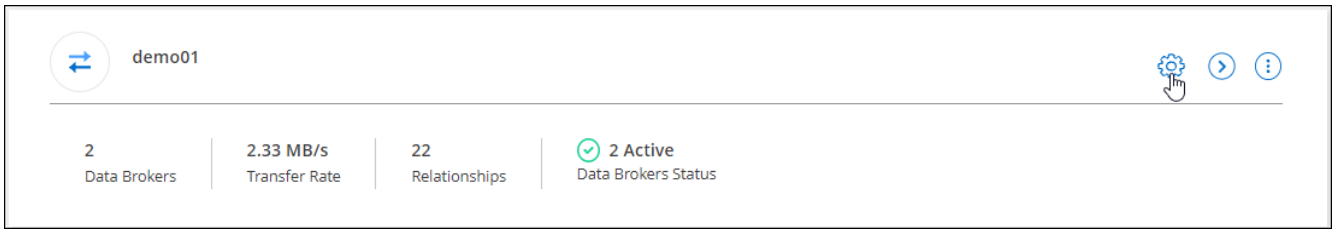
Define a unified configuration for a data broker group

If a sync relationship encounters errors during the sync process, unifying the concurrency of the data broker group can help to decrease the number of sync errors. Be aware that changes to the group's configuration can affect performance by slowing down the transfer.

We don't recommend changing the configuration on your own. You should consult with NetApp to understand when to change the configuration and how to change it.

Steps

1. Click **Manage Data Brokers**.
2. Click the Settings icon for a data broker group.



3. Change the settings as needed and then click **Unify Configuration**.

   Note the following:

   - You can pick and choose which settings to change—you don't need to change all four at once.
   - After a new configuration is sent to a data broker, the data broker automatically restarts and uses the new configuration.
   - It can take up to a minute until this change takes place and is visible in the Cloud Sync interface.
   - If a data broker isn't running, it's configuration won't change because Cloud Sync can't communicate with it. The configuration will change after the data broker restarts.

- After you set a unified configuration, any new data brokers will automatically use the new configuration.

Creating and viewing reports to tune your configuration

Create and view reports to get information that you can use with the help of NetApp personnel to tune a data broker's configuration and improve performance.

Each report provides in-depth details about a path in a sync relationship. For example, the report for a file system shows how many directories and files there are, the distribution of file size, how deep and wide the directories are, and more.

Creating reports

Each time that you create a report, Cloud Sync scans the path and then compiles the details into a report.

Steps

1. Click **Sync > Reports**.

   The paths (source or target) in each of your sync relationships display in a table.

2. In the **Reports Actions** column, go to a specific path and click **Create**, or click the action menu and select **Create New**.
3. When the report is ready, click the action menu and select **View**.

   Here's a sample report for a file system path.

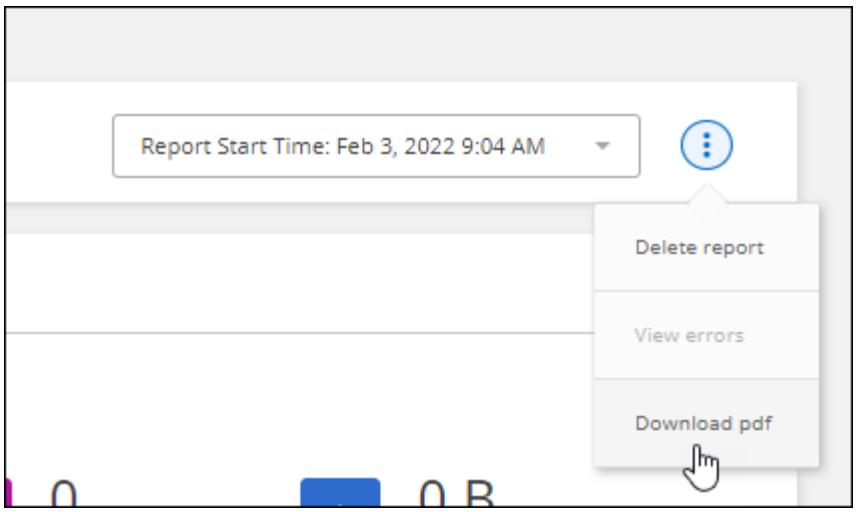
   And here's a sample report for object storage.

Downloading reports

You can download a report in PDF so that you can view it offline or share it.

Steps

1. Click **Sync > Reports**.
2. In the **Reports Actions** column, click the action menu and select **View**.
3. In the top right of the report, click the action menu and select **Download pdf**.



Viewing report errors

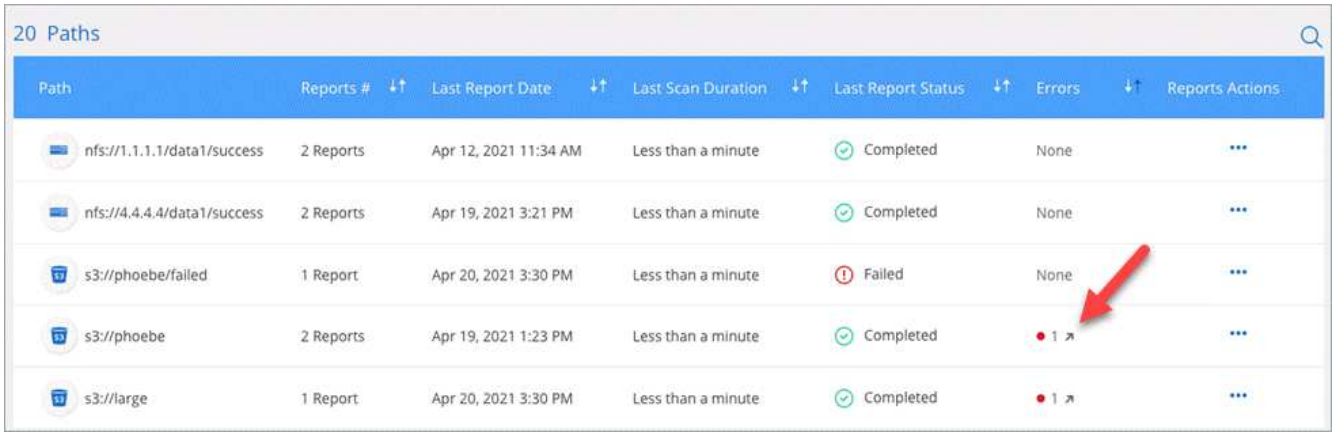
The Paths table identifies whether any errors are present in the most recent report. An error identifies an issue that Cloud Sync faced when scanning the path.

For example, a report might contain permission denied errors. This type of error can affect Cloud Sync's ability to scan the entire set of files and directories.

After you view the list of errors, you can then address the issues and run the report again.

Steps

1. Click **Sync > Reports**.
2. In the **Errors** column, identify whether any errors are present in a report.
3. If errors are present, click the arrow next to the number of errors.



4. Use the information in the error to correct the issue.

   After you resolve the issue, the error shouldn't appear the next time that you run the report.

Deleting reports

You might delete a report of it contained an error that you fixed, or if the report is related to a sync relationship that you removed.

Steps

1. Click **Sync > Reports**.
2. In the **Reports Actions** column, click the action menu for a path and select **Delete last report** or **Delete all reports**.
3. Confirm that you want to delete the report or reports.

Uninstalling the data broker

If needed, run an uninstall script to remove the data broker and the packages and directories that were created when the data broker was installed.

Steps

1. Log in to the data broker host.
2. Change to the data broker directory: `/opt/netapp/databroker`
3. Run the following commands:

```
chmod +x uninstaller-DataBroker.sh
./uninstaller-DataBroker.sh
```

4. Press ‘y’ to confirm the uninstallation.

Cloud Sync APIs

The Cloud Sync capabilities that are available through the web UI are also available through the RESTful API.

Getting started

To get started with the Cloud Sync API, you need to obtain a user token and your Cloud Central account ID. You’ll need to add the token and account ID to the Authorization header when making API calls.

Steps

1. Obtain a user token from NetApp Cloud Central.

```
POST https://netapp-cloud-account.auth0.com/oauth/token
Header: Content-Type: application/json
Body:
{
                "username": "<user_email>",
                "scope": "profile",
                "audience": "https://api.cloud.netapp.com",
                "client_id": "UaVhOIXMWQs5i1WdDxauXe5Mqkb34NJQ",
                "grant_type": "password",
                "password": "<user_password>"
}
```

2. Obtain your Cloud Central account ID.

```
GET https://api.cloudsync.netapp.com/api/accounts
Headers: Authorization: Bearer <user_token>
Content-Type: application/json
```

This API will return a response like the following:

```
[
  {
    "accountId": "account-JeL97Ry3",
    "name": "Test"
  }
]
```

3. Add the user token and account ID in the Authorization header of each API call.

Example

The following example shows an API call to create a data broker in Microsoft Azure. You would simply replace <user_token> and <accountId> with the token and ID that you obtained in the previous steps.

```
POST https://api.cloudsync.netapp.com/api/data-brokers
Headers: Authorization: Bearer <user_token>
Content-Type: application/json
x-account-id: <accountId>
Body: { "name": "databroker1", "type": "AZURE" }
```

What should I do when the token expires?

The user token from NetApp Cloud Central has an expiration date. To refresh the token, you need to call the API from step 1 again.

The API response includes an "expires_in" field that states when the token expires.

API reference

Documentation for each Cloud Sync API is available from https://api.cloudsync.netapp.com/docs.

Using list APIs

List APIs are asynchronous APIs, so the result does not return immediately (for example: `GET /data-brokers/{id}/list-nfs-export-folders` and `GET /data-brokers/{id}/list-s3-buckets`). The only response from the server is HTTP status 202. To get the actual result, you must use the `GET`

`/messages/client` API.

Steps

1. Call the list API that you want to use.
2. Use the `GET /messages/client` API to view the result of the operation.
3. Use the same API by appending it with the ID that you just received: `GET http://api.cloudsync.netapp.com/api/messages/client?last=<id_from_step_2>`

   Note that the ID changes each time that you call the `GET /messages/client` API.

Example

When you call the `list-s3-buckets` API, a result is not immediately returned:

```
GET http://api.cloudsync.netapp.com/api/data-brokers/<data-broker-id>/list-s3-buckets
Headers: Authorization: Bearer <user_token>
Content-Type: application/json
x-account-id: <accountId>
```

The result is HTTP status code 202, which means the message was accepted, but was not processed yet.

To get the result of the operation, you need to use the following API:

```
GET http://api.cloudsync.netapp.com/api/messages/client
Headers: Authorization: Bearer <user_token>
Content-Type: application/json
x-account-id: <accountId>
```

The result is an array with one object that includes an ID field. The ID field represents the last message that the server sent. For example:

```
[
    {
        "header": {
            "requestId": "init",
            "clientId": "init",
            "agentId": "init"
        },
        "payload": {
            "init": {}
        },
        "id": "5801"
    }
]
```

You would now make the following API call using the ID that you just received:

```
GET
http://api.cloudsync.netapp.com/api/messages/client?last=<id_from_step_2>
Headers: Authorization: Bearer <user_token>
Content-Type: application/json
x-account-id: <accountId>
```

The result is an array of messages. Inside each message is a payload object, which consists of the name of the operation (as key) and its result (as value). For example:

```
[
    {
        "payload": {
            "list-s3-buckets": [
                {
                    "tags": [
                        {
                            "Value": "100$",
                            "Key": "price"
                        }
                    ],
                    "region": {
                        "displayName": "US West (Oregon)",
                        "name": "us-west-2"
                    },
                    "name": "small"
                }
            ]
        },
        "header": {
            "requestId": "f687ac55-2f0c-40e3-9fa6-57fb8c4094a3",
            "clientId": "5beb032f548e6e35f4ed1ba9",
            "agentId": "5bed61f4489fb04e34a9aac6"
        },
        "id": "5802"
    }
]
```


Cloud Sync technical FAQ

This FAQ can help if you're just looking for a quick answer to a question.

Getting started

The following questions relate to getting started with Cloud Sync.

How does Cloud Sync work?

Cloud Sync uses the NetApp data broker software to sync data from a source to a target (this is called a *sync relationship*).

A data broker group controls the sync relationships between your sources and targets. After you set up a sync relationship, Cloud Sync analyzes your source system and breaks it up into multiple replication streams to push to your selected target data.

After the initial copy, the service syncs any changed data based on the schedule that you set.

How does the 14-day free trial work?

The 14-day free trial starts when you sign up for the Cloud Sync service. You're not subject to NetApp charges for Cloud Sync relationships you create for 14 days. However, all resource charges for any data brokers that you deploy still applies.

How much does Cloud Sync cost?

There are two types of costs associated with using Cloud Sync: service charges and resource charges.

Service charges

For pay-as-you-go pricing, Cloud Sync service charges are hourly, based on the number of sync relationships that you create.

- View pay-as-you-go pricing in AWS
- View annual pricing in AWS
- View pricing in Azure

Cloud Sync licenses are also available through your NetApp representative. Each license enables 20 sync relationships for 12 months.

Learn more about licenses.

Cloud Sync relationships are free for Cloud Volumes Service and Azure NetApp Files.

Resource charges

The resource charges are related to the compute and storage costs for running the data broker in the cloud.

How is Cloud Sync billed?

There are two ways to pay for sync relationships after your 14-day free trial ends. The first option is to subscribe from AWS or Azure, which enables you to pay-as-you-go or to pay annually. The second option is to purchase licenses directly from NetApp.

Can I use Cloud Sync outside the cloud?

Yes, you can use Cloud Sync in a non-cloud architecture. The source and target can reside on-premises and so can the data broker software.

Note the following key points about using Cloud Sync outside of the cloud:

- For on-premises synchronization, a private Amazon S3 bucket is available through NetApp StorageGRID.
- A data broker group needs an internet connection to communicate with the Cloud Sync service.
- If you don't purchase a license directly from NetApp, you will need an AWS or Azure account for the PAYGO Cloud Sync service billing.

What's a data broker group?

Each data broker belongs to a data broker group. Grouping data brokers together helps improve the performance of sync relationships.

How do I access Cloud Sync?

Cloud Sync is available from Cloud Manager in the **Sync** tab.

Supported sources and targets

The following questions related to the source and targets that are supported in a sync relationship.

Which sources and targets does Cloud Sync support?

Cloud Sync supports many different types of sync relationships. View the entire list.

What versions of NFS and SMB does Cloud Sync support?

Cloud Sync supports NFS version 3 and later, and SMB version 1 and later.

Learn more about sync requirements.

When Amazon S3 is the target, can the data be tiered to a specific S3 storage class?

Yes, you can choose a specific S3 storage class when AWS S3 is the target:

- Standard (this is the default class)
- Intelligent-Tiering
- Standard-Infrequent Access
- One Zone-Infrequent Access
- Glacier
- Glacier Deep Archive

What about storage tiers for Azure Blob storage?

You can choose a specific Azure Blob storage tier when a Blob container is the target:

- Hot storage
- Cool storage

Do you support Google Cloud storage tiers?

Yes, you can choose a specific storage class when a Google Cloud Storage bucket is the target:

- Standard
- Nearline
- Coldline
- Archive

Networking

The following questions relate to networking requirements for Cloud Sync.

What are the networking requirements for Cloud Sync?

The Cloud Sync environment requires that a data broker group is connected with the source and the target through the selected protocol or object storage API (Amazon S3, Azure Blob, IBM Cloud Object Storage).

In addition, a data broker group needs an outbound internet connection over port 443 so it can communicate with the Cloud Sync service and contact a few other services and repositories.

For more details, review networking requirements.

Can I use a proxy server with the data broker?

Yes.

Cloud Sync supports proxy servers with or without basic authentication. If you specify a proxy server when you deploy a data broker, all HTTP and HTTPS traffic from the data broker is routed through the proxy. Note that non-HTTP traffic such as NFS or SMB can't be routed through a proxy server.

The only proxy server limitation is when using data-in-flight encryption with an NFS or Azure NetApp Files sync relationship. The encrypted data is sent over HTTPS and isn't routable through a proxy server.

Data synchronization

The following questions relate to how data synchronization works.

How often does synchronization occur?

The default schedule is set for daily synchronization. After the initial synchronization, you can:

- Modify the sync schedule to your desired number of days, hours, or minutes
- Disable the sync schedule
- Delete the sync schedule (no data will be lost; only the sync relationship will be removed)

What is the minimum sync schedule?

You can schedule a relationship to sync data as often as every 1 minute.

Does the data broker group retry when a file fails to sync? Or does it timeout?

A data broker group doesn't timeout when a single file fails to transfer. Instead, the data broker group retries 3 times before skipping the file. The retry value is configurable in the settings for a sync relationship.

Learn how to change the settings for a sync relationship.

What if I have a very large dataset?

If a single directory contains 600,000 files or more, contact us so that we can help you configure the data broker group to handle the payload. We might need to add additional memory to the data broker group.

Note that there's no limit to the total number of files in the mount point. The extra memory is required for large directories with 600,000 files or more, regardless of their level in the hierarchy (top directory or subdirectory).

Security

The following questions related to security.

Is Cloud Sync secure?

Yes. All Cloud Sync service networking connectivity is done using Amazon Simple Queue Service (SQS).

All communication between the data broker group and Amazon S3, Azure Blob, Google Cloud Storage, and IBM Cloud Object Storage is done through the HTTPS protocol.

If you're using Cloud Sync with on-premises (source or destination) systems, here's a few recommended connectivity options:

- An AWS Direct Connect, Azure ExpressRoute, or Google Cloud Interconnect connection, which is non-internet routed (and can only communicate with the cloud networks that you specify)
- A VPN connection between your on-premises gateway device and your cloud networks
- For extra secure data transfer with S3 buckets, Azure Blob storage, or Google Cloud Storage, an Amazon Private S3 Endpoint, Azure Virtual Network service endpoints, or Private Google Access may be established.

Any of these methods establishes a secure connection between your on-premises NAS servers and a Cloud Sync data broker group.

Is data encrypted by Cloud Sync?

- Cloud Sync supports data-in-flight encryption between source and target NFS servers. Learn more.
- For SMB, Cloud Sync supports SMB 3.0 and 3.11 data that you've encrypted on the server side. Cloud Sync copies the encrypted data from the source to the target where the data remains encrypted.

  Cloud Sync cannot encrypt SMB data itself.

- When an Amazon S3 bucket is the target in a sync relationship, you can choose whether to enable data encryption using AWS KMS encryption or AES-256 encryption.

Permissions

The following questions relate to data permissions.

Are SMB data permissions synced to the target location?

You can set up Cloud Sync to preserve access control lists (ACLs) between a source SMB share and a target SMB share, and from a source SMB share to object storage (except for ONTAP S3).

Cloud Sync doesn't support copying ACLs from object storage to SMB shares.

Learn how to copy ACLs between SMB shares.

Are NFS data permissions synced to the target location?

Cloud Sync automatically copies NFS permissions between NFS servers as follows:

- NFS version 3: Cloud Sync copies the permissions and the user group owner.
- NFS version 4: Cloud Sync copies the ACLs.

Object storage metadata

Cloud Sync copies object storage metadata from the source to the target for the following types of sync relationships:

- Amazon S3 → Amazon S3 [1]
- Amazon S3 → StorageGRID
- StorageGRID → Amazon S3
- StorageGRID → StorageGRID
- StorageGRID → Google Cloud Storage
- Google Cloud Storage → StorageGRID [1]
- Google Cloud Storage → IBM Cloud Object Storage [1]
- Google Cloud Storage → Amazon S3 [1]
- Amazon S3 → Google Cloud Storage
- IBM Cloud Object Storage → Google Cloud Storage
- StorageGRID → IBM Cloud Object Storage
- IBM Cloud Object Storage → StorageGRID
- IBM Cloud Object Storage → IBM Cloud Object Storage

[1] For these sync relationships, you need to enable a setting when you create the sync relationship.

Performance

The following questions relate to Cloud Sync performance.

What does the progress indicator for a sync relationship represent?

The sync relationship shows the throughput of the data broker group's network adapter. If you accelerated sync performance by using multiple data brokers, then the throughput is the sum of all traffic. This throughput refreshes every 20 seconds.

I'm experiencing performance issues. Can we limit the number of concurrent transfers?

If you have very large files (multiple TiBs each), it can take a long time to complete the transfer process and performance might be impacted.

Limiting the number of concurrent transfers can help. Contact us for help.

Why am I experiencing low performance with Azure NetApp Files?

When you sync data to or from Azure NetApp Files, you might experience failures and performance issues if the disk service level is Standard.

Change the service level to Premium or Ultra to enhance the sync performance.

Learn more about Azure NetApp Files service levels and throughput.

Why am I experiencing low performance with Cloud Volumes Service for AWS?

When you sync data to or from a cloud volume, you might experience failures and performance issues if the level of performance for the cloud volume is Standard.

Change the Service level to Premium or Extreme to enhance the sync performance.

How many data brokers are required in a group?

When you create a new relationship, you start with a single data broker in a group (unless you selected an existing data broker that belongs to an accelerated sync relationship). In many cases, a single data broker can meet the performance requirements for a sync relationship. If it doesn't, you can accelerate sync performance by adding additional data brokers to the group. But you should first check other factors that can impact sync performance.

Multiple factors can impact data transfer performance. The overall sync performance might be impacted due to network bandwidth, latency, and network topology, as well as the data broker VM specs and storage system performance. For example, a single data broker in a group can reach 100 MB/s, while disk throughput on the target might only allow 64 MB/s. As a result, the data broker group keeps trying to copy the data, but the target can't meet the performance of the data broker group.

So be sure to check the performance of your networking and the disk throughput on the target.

Then you can consider accelerating sync performance by adding an additional data brokers to a group to share the load of that relationship. Learn how to accelerate sync performance.

Deleting things

The following questions relate to deleting sync relationships and data from sources and targets.

What happens if I delete my Cloud Sync relationship?

Deleting a relationship stops all future data syncs and terminates payment. Any data that was synced to the target remains as-is.

What happens if I delete something from my source server? Is it removed from the target too?

By default, if you have an active sync relationship, the item deleted on the source server is not deleted from the target during the next synchronization. But there is an option in the sync settings for each relationship, where you can define that Cloud Sync will delete files in the target location if they were deleted from the source.

Learn how to change the settings for a sync relationship.

What happens if I delete something from my target? Is it removed from my source too?

If an item is deleted from the target, it will not be removed from the source. The relationship is one-way—from source to target. On the next sync cycle, Cloud Sync compares the source to the target, identifies that the item is missing, and Cloud Sync copies it again from the source to the target.

Troubleshooting

NetApp Knowledgebase: Cloud Sync FAQ: Support and Troubleshooting

Data broker deep dive

The following question relates to the data broker.

Can you explain the architecture of the data broker?

Sure. Here are the most important points:

- The data broker is a node.js application running on a Linux host.
- Cloud Sync deploys the data broker as follows:
  - AWS: From an AWS CloudFormation template
  - Azure: From Azure Resource Manager
  - Google: From Google Cloud Deployment Manager
  - If you use your own Linux host, you need to manually install the software
- The data broker software automatically upgrades itself to the latest version.
- The data broker uses AWS SQS as a reliable and secure communication channel and for control and monitoring. SQS also provides a persistency layer.
- You can add additional data brokers to a group to increase transfer speed and add high availability. There is service resiliency if one data broker fails.