



## **Get started**

### Cloud Sync

NetApp  
April 05, 2022

# Table of Contents

- Get started ..... 1
  - Cloud Sync overview ..... 1
  - Quick start for Cloud Sync ..... 3
  - Supported sync relationships ..... 4
  - Prepare the source and target ..... 11
  - Networking overview for Cloud Sync ..... 17
  - Install a data broker ..... 20

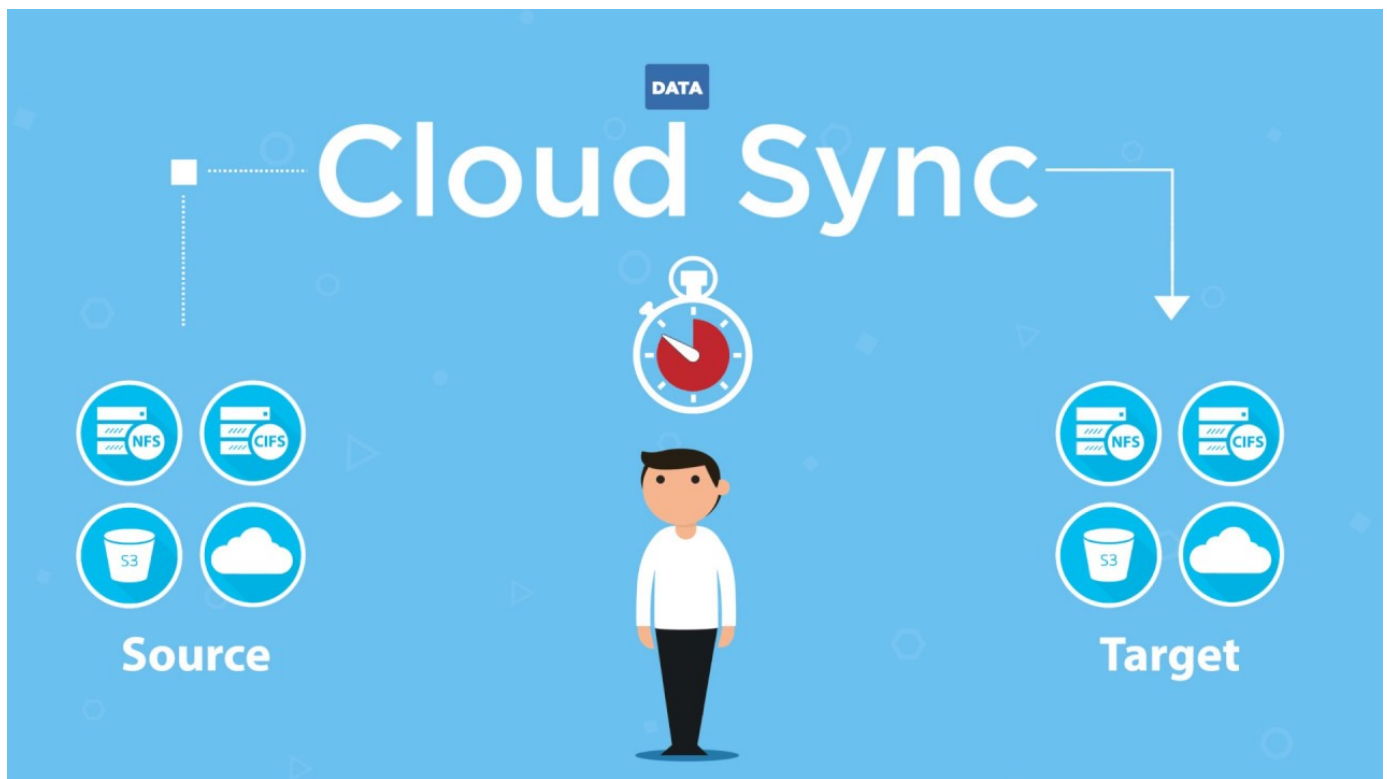
# Get started

## 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

[View the supported sync relationships.](#)

## Costs

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.

[Learn how licensing works.](#)

## Quick start for Cloud Sync

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

1

### 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.](#)

2

### 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 Google Cloud installation](#)
- [Review Linux host installation](#)

**3****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](#).

**4****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](#).

## 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	<ul style="list-style-type: none"><li>• Amazon EFS</li><li>• Amazon FSx for ONTAP</li><li>• Amazon S3</li><li>• Azure Blob</li><li>• Azure NetApp Files</li><li>• Cloud Volumes ONTAP</li><li>• Cloud Volumes Service</li><li>• Google Cloud Storage</li><li>• IBM Cloud Object Storage</li><li>• NFS server</li><li>• On-premises ONTAP cluster</li><li>• SMB server</li><li>• StorageGRID</li></ul>

Source location	Supported target locations
Amazon FSx for ONTAP	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
Amazon S3	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Box <sup>2,3</sup></li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>

Source location	Supported target locations
Azure Blob	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
Azure NetApp Files	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB server</li> <li>• StorageGRID</li> </ul>
Box <sup>2</sup>	<ul style="list-style-type: none"> <li>• Amazon S3</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>



Source location	Supported target locations
Cloud Volumes ONTAP	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
Cloud Volumes Service	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>

Source location	Supported target locations
Google Cloud Storage	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• ONTAP S3 Storage</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
IBM Cloud Object Storage	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Box <sup>2,3</sup></li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>

Source location	Supported target locations
NFS server	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
On-prem ONTAP cluster	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
ONTAP S3 Storage	<ul style="list-style-type: none"> <li>• Google Cloud Storage</li> <li>• SMB server</li> <li>• StorageGRID</li> <li>• ONTAP S3 Storage</li> </ul>
SFTP <sup>1</sup>	S3

Source location	Supported target locations
SMB server	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• ONTAP S3 Storage</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>
StorageGRID	<ul style="list-style-type: none"> <li>• Amazon EFS</li> <li>• Amazon FSx for ONTAP</li> <li>• Amazon S3</li> <li>• Azure Blob</li> <li>• Azure NetApp Files</li> <li>• Box <sup>2,3</sup></li> <li>• Cloud Volumes ONTAP</li> <li>• Cloud Volumes Service</li> <li>• Google Cloud Storage</li> <li>• IBM Cloud Object Storage</li> <li>• NFS server</li> <li>• On-premises ONTAP cluster</li> <li>• ONTAP S3 Storage</li> <li>• SMB Server</li> <li>• StorageGRID</li> </ul>

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

## Prepare the source and target

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.

## Summary

Delete role

**Role ARN** `arn:aws:iam::142591788901:role/tanyaBroker0304-DataBrokerIamRole-1VMHXXMW3AQ05`

Role description [Edit](#)

## Azure Blob storage requirements

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

### Supported data broker locations for Azure Blob

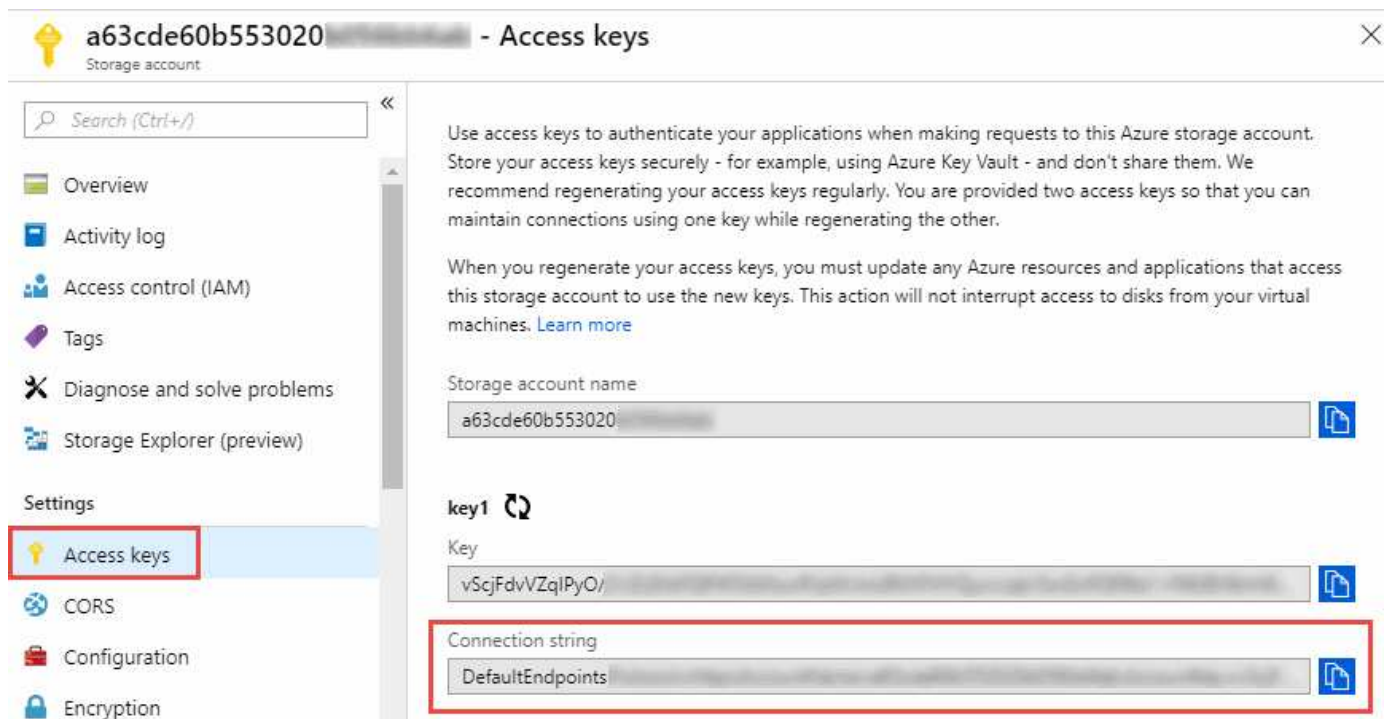
A data broker can reside in any location when a sync relationship includes Azure Blob storage.

### Supported Azure regions

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

### Connection string 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:



The screenshot shows the Azure portal interface for a storage account named 'a63cde60b553020'. The left sidebar contains navigation options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Storage Explorer (preview), Settings, Access keys (highlighted with a red box), CORS, Configuration, and Encryption. The main content area is titled 'Access keys' and includes instructions on using access keys. It also shows the 'key1' section with a 'Key' field and a 'Connection string' field. The 'Connection string' field is highlighted with a red box and contains the value 'DefaultEndpoints'.

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

Search (Ctrl+/)

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Storage Explorer (preview)

Settings

Access keys

CORS

Configuration

Encryption

**Shared access signature**

Firewalls and virtual networks

Advanced Threat Protection (pr...

Properties

Locks

Allowed services ⓘ

☒ Blob ☐ File ☐ Queue ☐ Table

Allowed resource types ⓘ

☒ Service ☒ Container ☒ Object

Allowed permissions ⓘ

☒ Read ☒ Write ☒ Delete ☒ List ☒ Add ☒ Create ☐ Update ☐ Process

Start and expiry date/time ⓘ

Start

2018-10-23 10:07:32 AM

End

2019-10-23 6:07:32 PM

(UTC-04:00) --- Current Time Zone ---

Allowed IP addresses ⓘ

for example, 168.1.5.65 or 168.1.5.65-168.1.5.70

Allowed protocols ⓘ

☒ HTTPS only ☐ HTTPS and HTTP

Signing key ⓘ

key1

**Generate SAS and connection string**

## 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 Google Cloud regions

All regions are supported.

### Permissions 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.](#)

### 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 over the required ports.
  - 111 TCP/UDP
  - 2049 TCP/UDP
  - 5555 TCP/UDP
- 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 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.

## 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 over the required ports.
  - 139 TCP
  - 445 TCP
  - 137-138 UDP
- 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

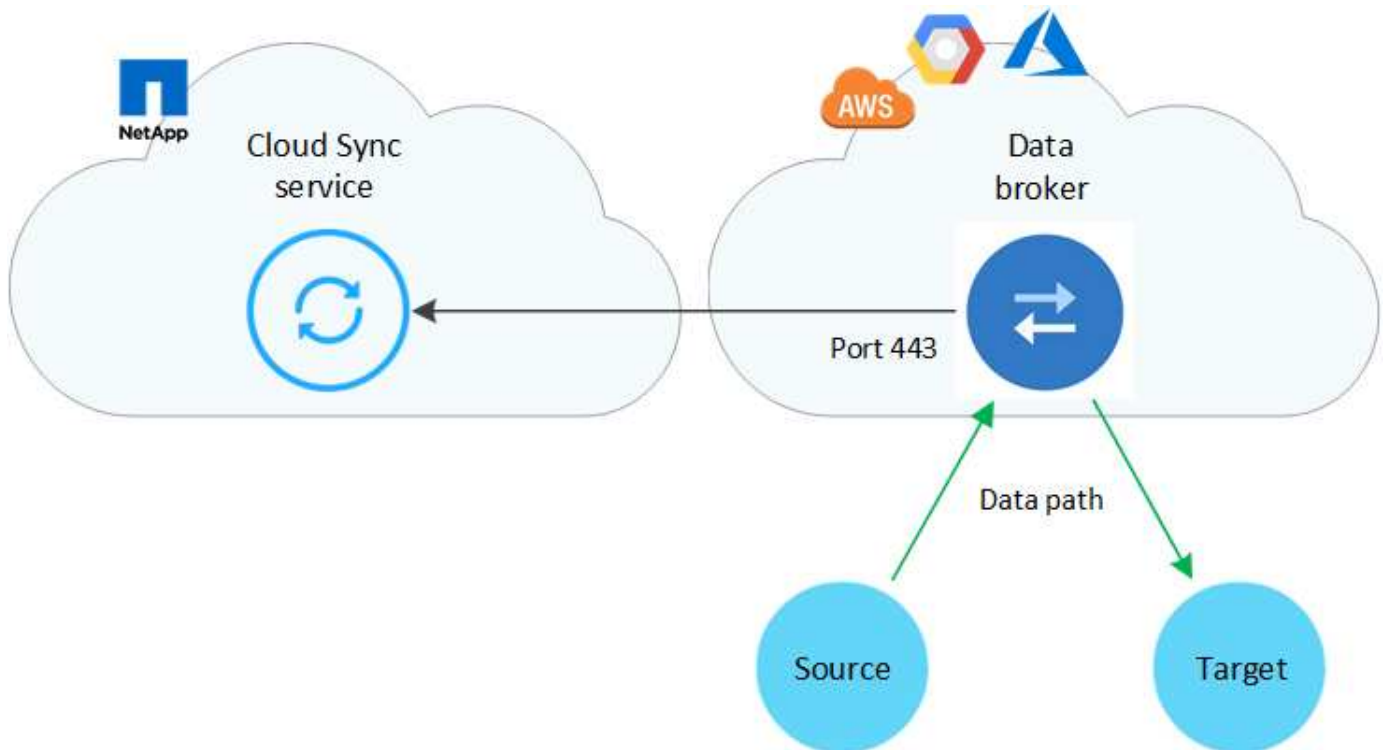
A data broker group consists of one or more data brokers installed in the cloud or on your premises.

#### Data broker in the cloud

The following image shows a data broker running in the cloud, in either AWS, Google Cloud, 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.



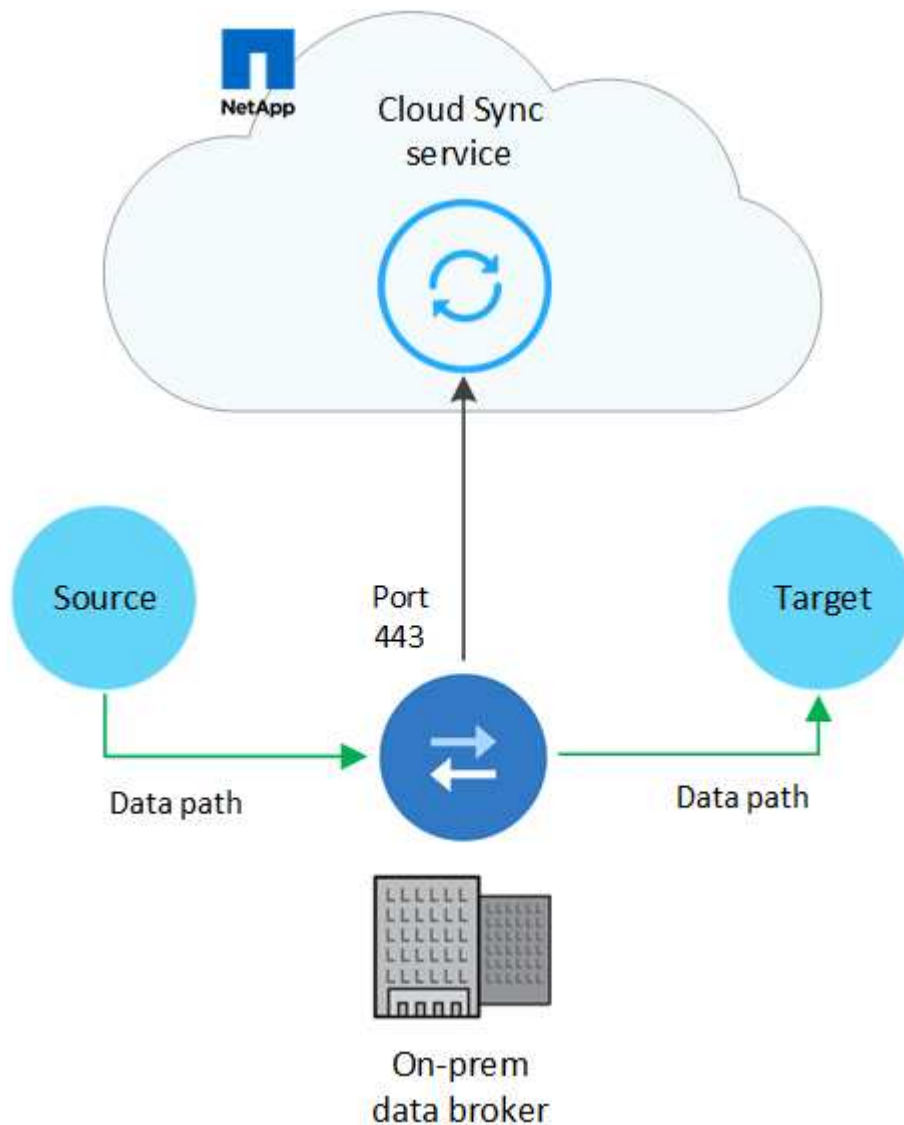
When Cloud Sync deploys the data broker in AWS, Azure, or Google Cloud, it creates a security group that enables the required outbound communication.



#### 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
<a href="https://olcentgbl.trafficmanager.net">https://olcentgbl.trafficmanager.net</a>	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.
<a href="https://rpm.nodesource.com">https://rpm.nodesource.com</a> <a href="https://registry.npmjs.org">https://registry.npmjs.org</a> <a href="https://nodejs.org">https://nodejs.org</a> :	To contact repositories for updating Node.js, npm, and other 3rd party packages used in development.
<a href="https://tgz.pm2.io">https://tgz.pm2.io</a>	To access a repository for updating PM2, which is a 3rd party package used to monitor Cloud Sync.
<a href="https://sqs.us-east-1.amazonaws.com">https://sqs.us-east-1.amazonaws.com</a> <a href="https://kinesis.us-east-1.amazonaws.com">https://kinesis.us-east-1.amazonaws.com</a>	To contact the AWS services that Cloud Sync uses for operations (queuing files, registering actions, and delivering updates to the data broker).
<a href="https://s3.region.amazonaws.com">https://s3.region.amazonaws.com</a>  For example: <a href="https://s3.us-east-2.amazonaws.com:443">s3.us-east-2.amazonaws.com:443</a> <a href="#">See AWS documentation for a list of S3 endpoints</a>	To contact Amazon S3 when a sync relationship includes an S3 bucket.
<a href="https://s3.us-east-1.amazonaws.com">https://s3.us-east-1.amazonaws.com</a>	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.
<a href="https://cf.cloudsync.netapp.com">https://cf.cloudsync.netapp.com</a> <a href="https://repo.cloudsync.netapp.com">https://repo.cloudsync.netapp.com</a>	To contact the Cloud Sync service.
<a href="https://support.netapp.com">https://support.netapp.com</a>	To contact NetApp support when using a BYOL license for sync relationships.
<a href="https://fedoraproject.org">https://fedoraproject.org</a>	To install 7z on the data broker virtual machine during installation and updates. 7z is needed to send AutoSupport messages to NetApp technical support.
<a href="https://sts.amazonaws.com">https://sts.amazonaws.com</a>	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.
<a href="https://cloudmanager.cloud.netapp.com">https://cloudmanager.cloud.netapp.com</a> <a href="https://netapp-cloud-account.auth0.com">https://netapp-cloud-account.auth0.com</a>	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](https://logs.cloudsync.netapp.com:443)

# 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

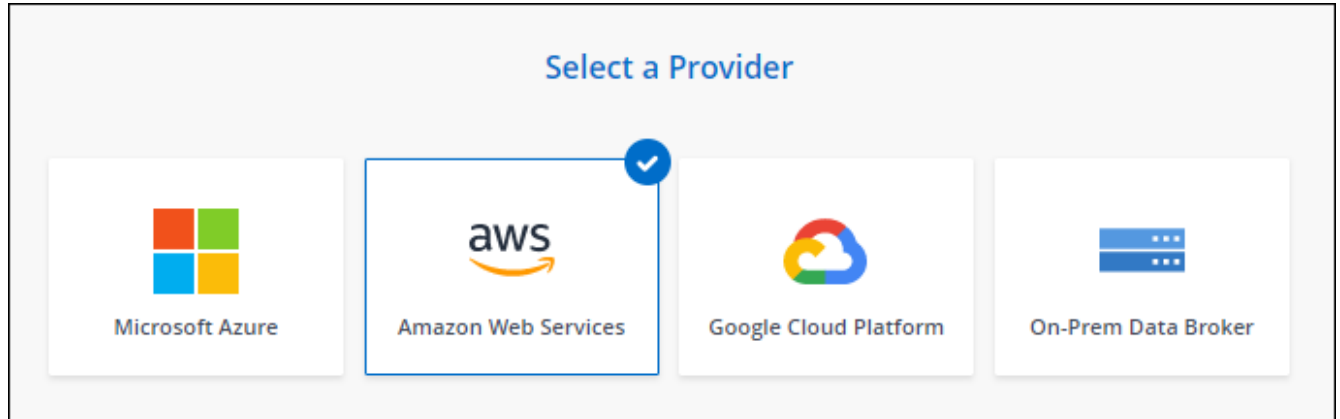
There are a few ways to create a new data broker. These steps describe how to install a data broker in AWS when creating a sync relationship.

## 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/cloud-manager-sync//media/video\\_cloud\\_sync.mp4](https://docs.netapp.com/us-en/cloud-manager-sync//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](#).

### Basic Settings

**Location**

Region  

US West | Oregon
▼

VPC  

vpc-3c46c059 - 10.60.21.0/25
▼

Subnet  

10.60.21.0/25
▼

**Connectivity**

Key Pair  

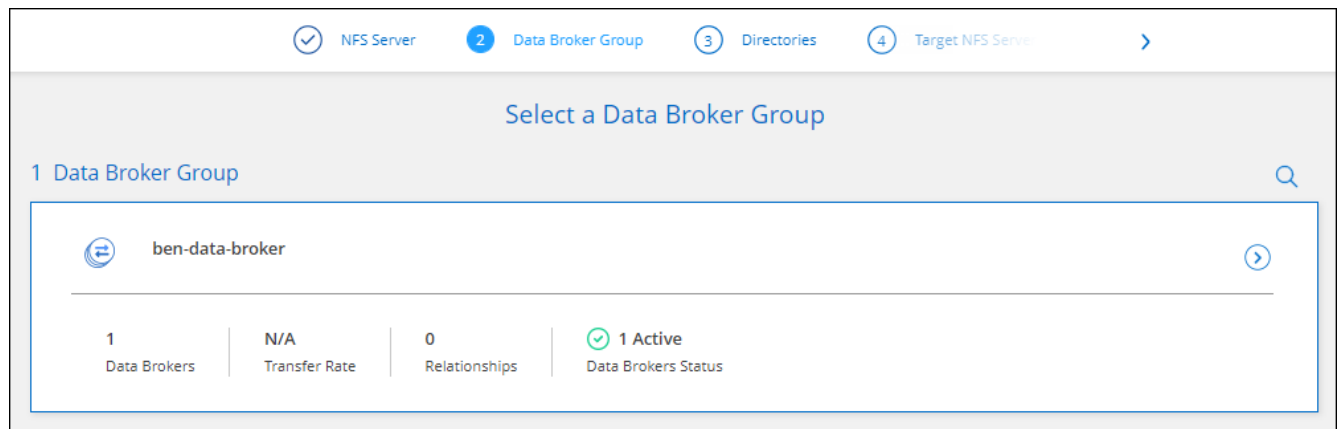
newKey
▼

**Enable Public IP?**  
☒ Enable ☐ Disable

IAM Role (optional) ?

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**

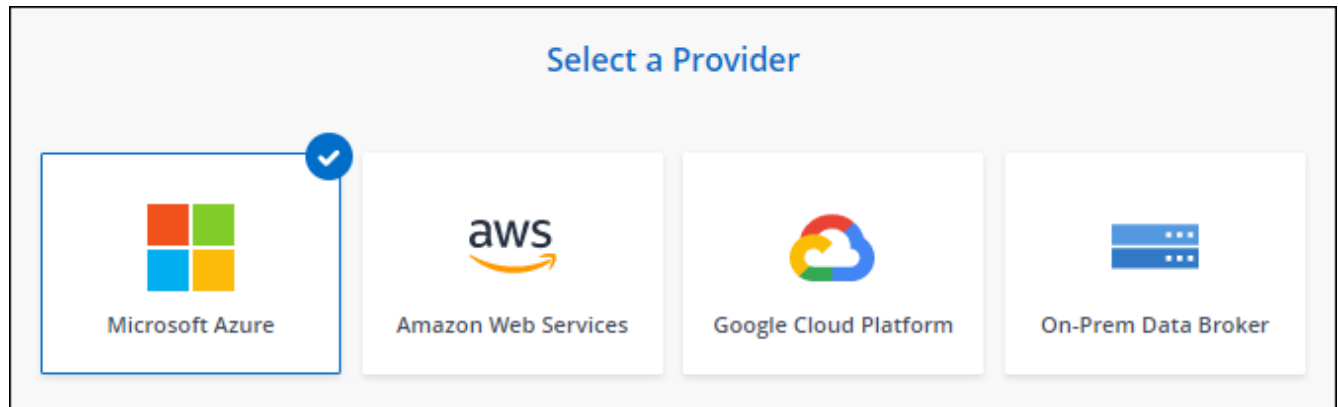
There are a few ways to create a new data broker. These steps describe how to install a data broker in Azure when you create a sync relationship.

**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.

The screenshot shows a configuration page with two main sections: 'Location' and 'Virtual Machine'. The 'Location' section has four dropdown menus: 'Subscription' (OCCM Dev), 'Azure Region' (West US 2), 'VNet' (Vnet1), and 'Subnet' (Subnet1). The 'Virtual Machine' section has a text input for 'VM Name' (netappdatabroker), a text input for 'User Name' (databroker), a section for 'Authentication Method' with 'Password' selected, a password input field, and a section for 'Resource Group' with 'Generate a new group' selected.

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 Google Cloud 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 Google Cloud, 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 Google Cloud

Ensure that the Google Cloud 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

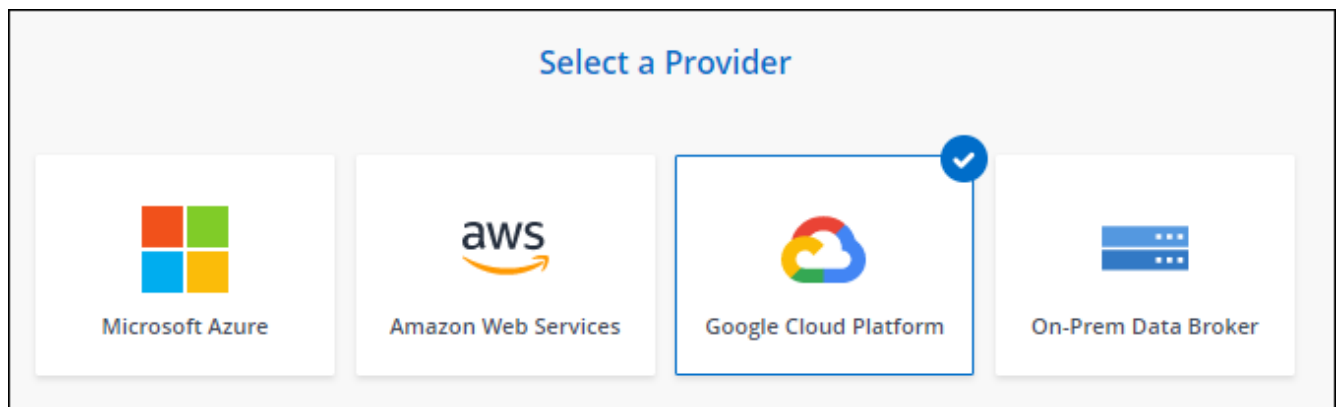
There are a few ways to create a new data broker. These steps describe how to install a data broker in Google Cloud when you create a sync relationship.

### 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.

Basic Settings

<p><b>Project</b></p> <p>Project</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">OCCM-Dev ▼</div> <p>Service Account</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">test ▼</div> <p>Select a Service Account that includes <a href="#">these permissions</a></p>	<p><b>Location</b></p> <p>Region</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">us-west1 ▼</div> <p>Zone</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">us-west1-a ▼</div> <p>VPC</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">default ▼</div> <p>Subnet</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">default ▼</div> <p>Public IP</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 10px;">Enable ▼</div>
---	---

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 Google Cloud 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 Google Cloud 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 Google Cloud 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

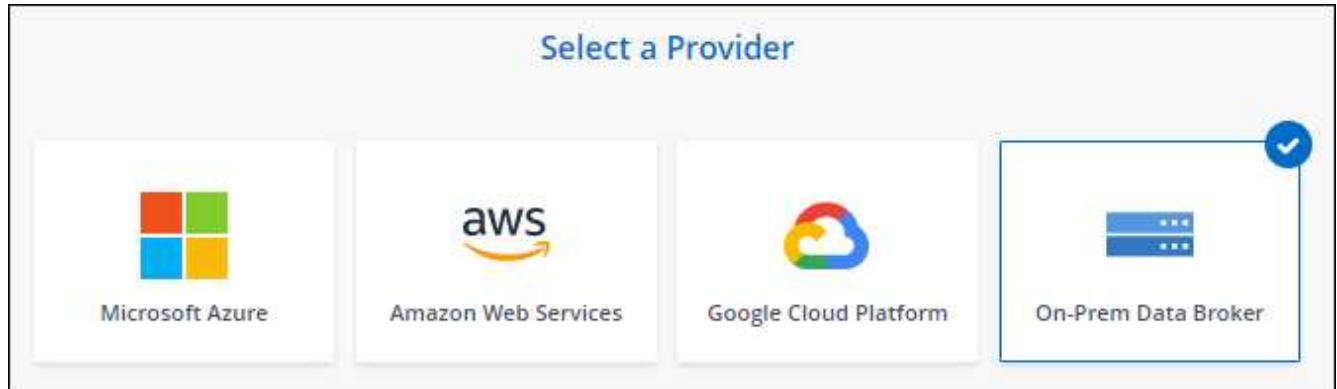
You can install a data broker on a Linux host when you create a sync relationship.

### 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 Google Cloud 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.

## Copyright Information

Copyright © 2022 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

## Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.