Get started

Cloud Sync

NetApp
April 12, 2021

Table of Contents

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 setup. The most important requirement is to verify connectivity between the data broker 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. The data broker 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 the 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

Start your free trial from NetApp Cloud Central, drag and drop your selections for the source and target, and 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.

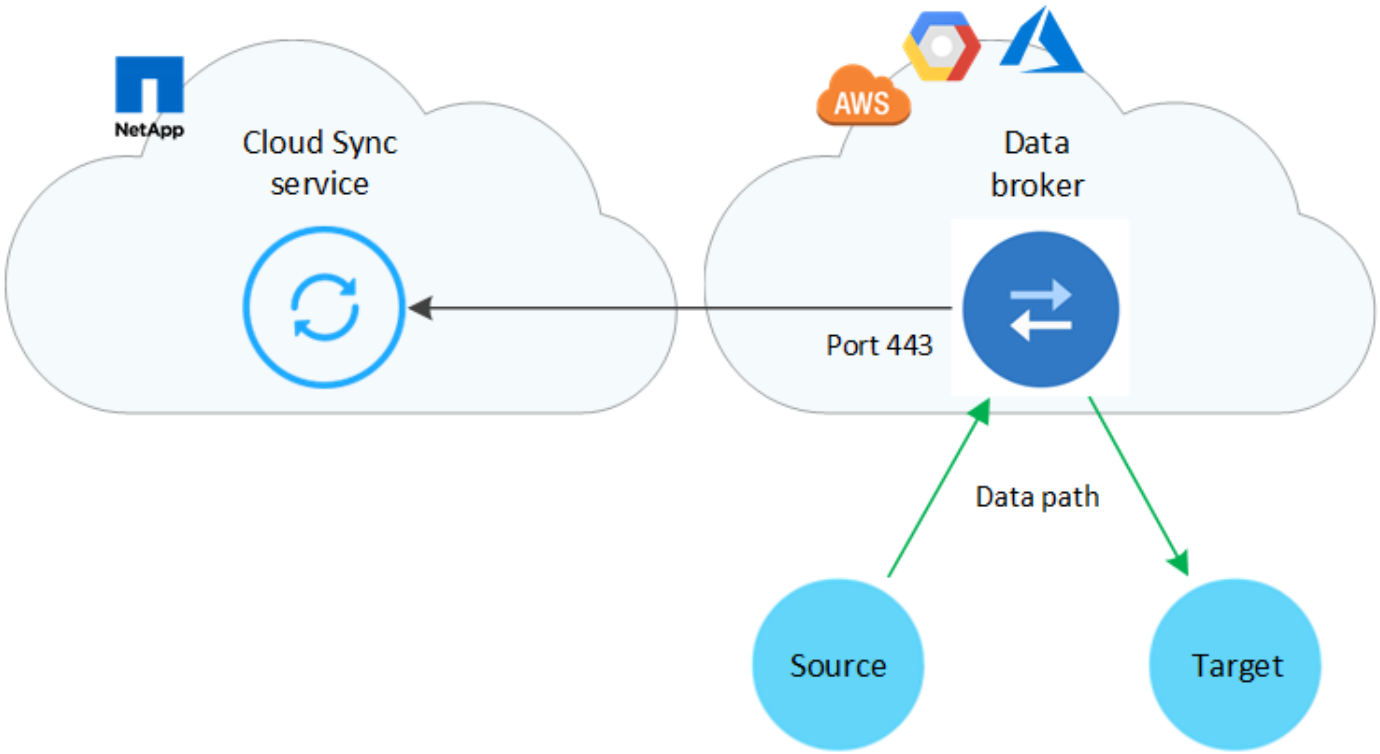
Networking overview

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

Data broker in the cloud

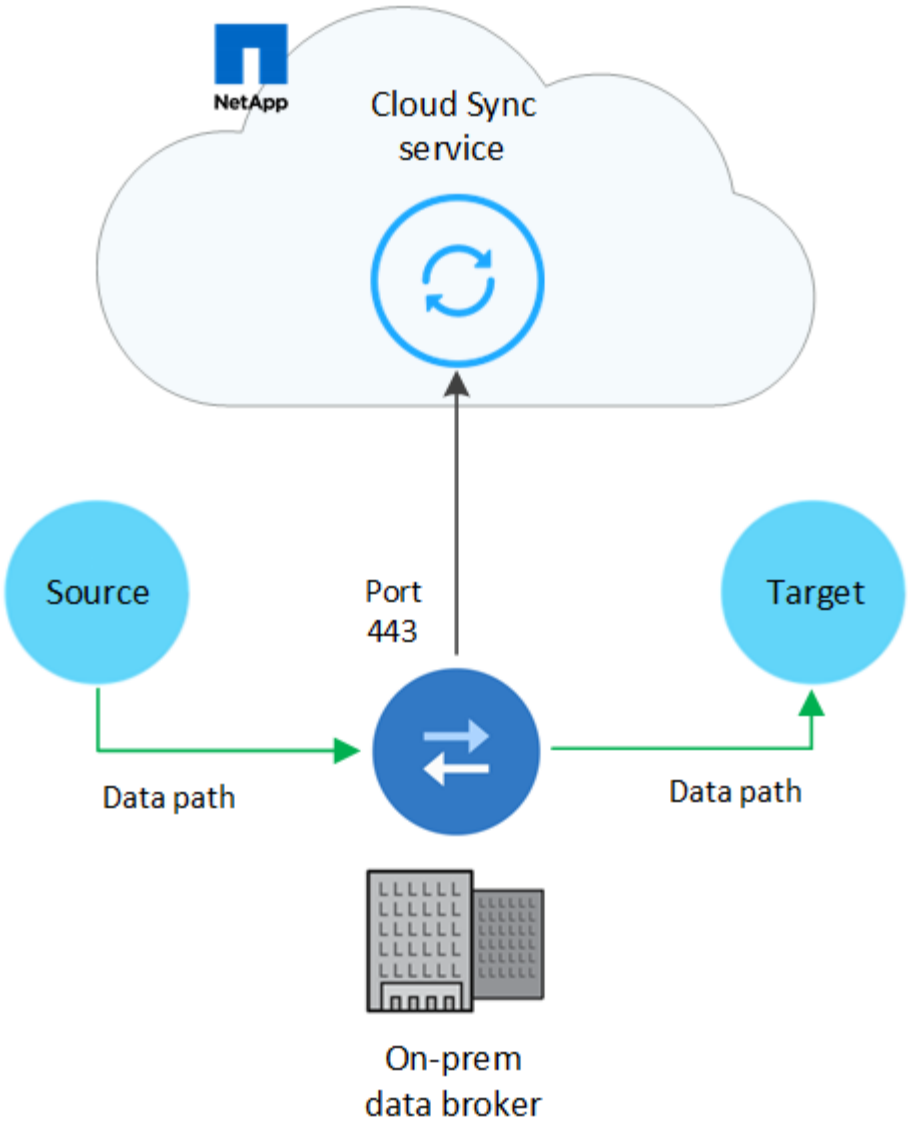
The following image shows the data broker running in the cloud, in either AWS, GCP, or Azure. The source and target can be in any location, as long as there's a connection to the data broker. For example, you might have a VPN connection from your data center to your cloud provider.

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.



Related link

Endpoints that the data broker contacts

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
AWS EFS	• AWS EFS • AWS S3 • Azure Blob • Azure NetApp Files (NFS) • Cloud Volumes ONTAP (NFS) • Cloud Volumes Service (NFS) • IBM Cloud Object Storage • Google Cloud Storage • NFS server • On-premises ONTAP cluster • StorageGRID
AWS S3	• AWS EFS • AWS S3 • Azure Blob • Azure NetApp Files • Cloud Volumes ONTAP • Cloud Volumes Service • IBM Cloud Object Storage • Google Cloud Storage • NFS server • On-premises ONTAP cluster • SMB Server • StorageGRID

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

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

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

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

Notes:

1. Cloud Sync supports sync relationships from SFTP to S3 by using the API only.
2. You can choose a specific Azure Blob storage tier when a Blob container is the target:
    - Hot storage
    - Cool storage

3. 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

Networking requirements

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

  For example, if an NFS server is in your data center and the 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, target, and data broker to use a Network Time Protocol (NTP) service. The time difference between the three components should not exceed 5 minutes.

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.

NFS server requirements

- The NFS server can be a NetApp system or a non-NetApp system. For example, Cloud Volumes ONTAP or an AFF cluster.
- The file server must allow the 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.

SMB server requirements

- The SMB server can be a NetApp system or a non-NetApp system. For example, Cloud Volumes ONTAP or an AFF cluster.
- The file server must allow the 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"

AWS S3 bucket requirements

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

Supported data broker locations for AWS 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 and GovCloud (US) regions.

Permissions required for S3 buckets in other AWS accounts

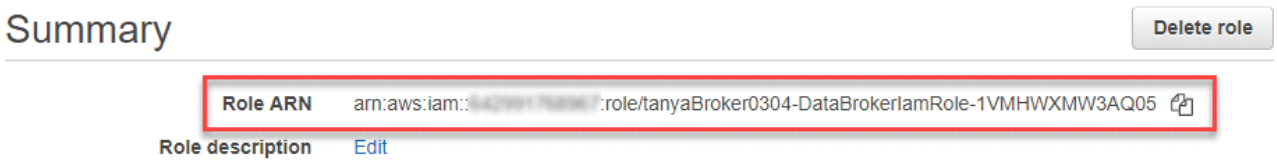
When setting up a sync relationship, you can specify an S3 bucket that resides in an AWS account that isn't associated with the data broker.

The permissions included in this JSON file must be applied to that S3 bucket so the 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 the data broker.
2. *<RoleARN>* should be replaced with one of the following:
   - If the 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 the data broker.
   - If the 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

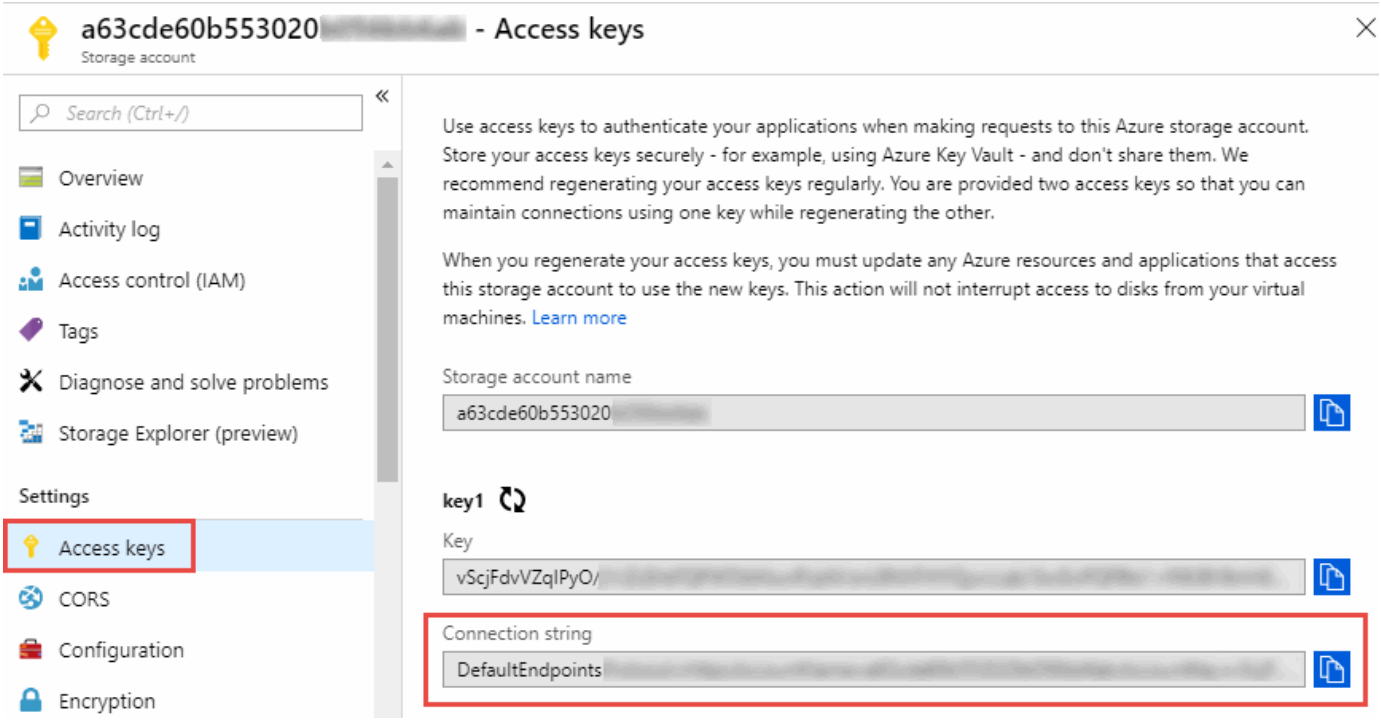
The data broker can reside in any location when a sync relationship includes Azure Blob storage.

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

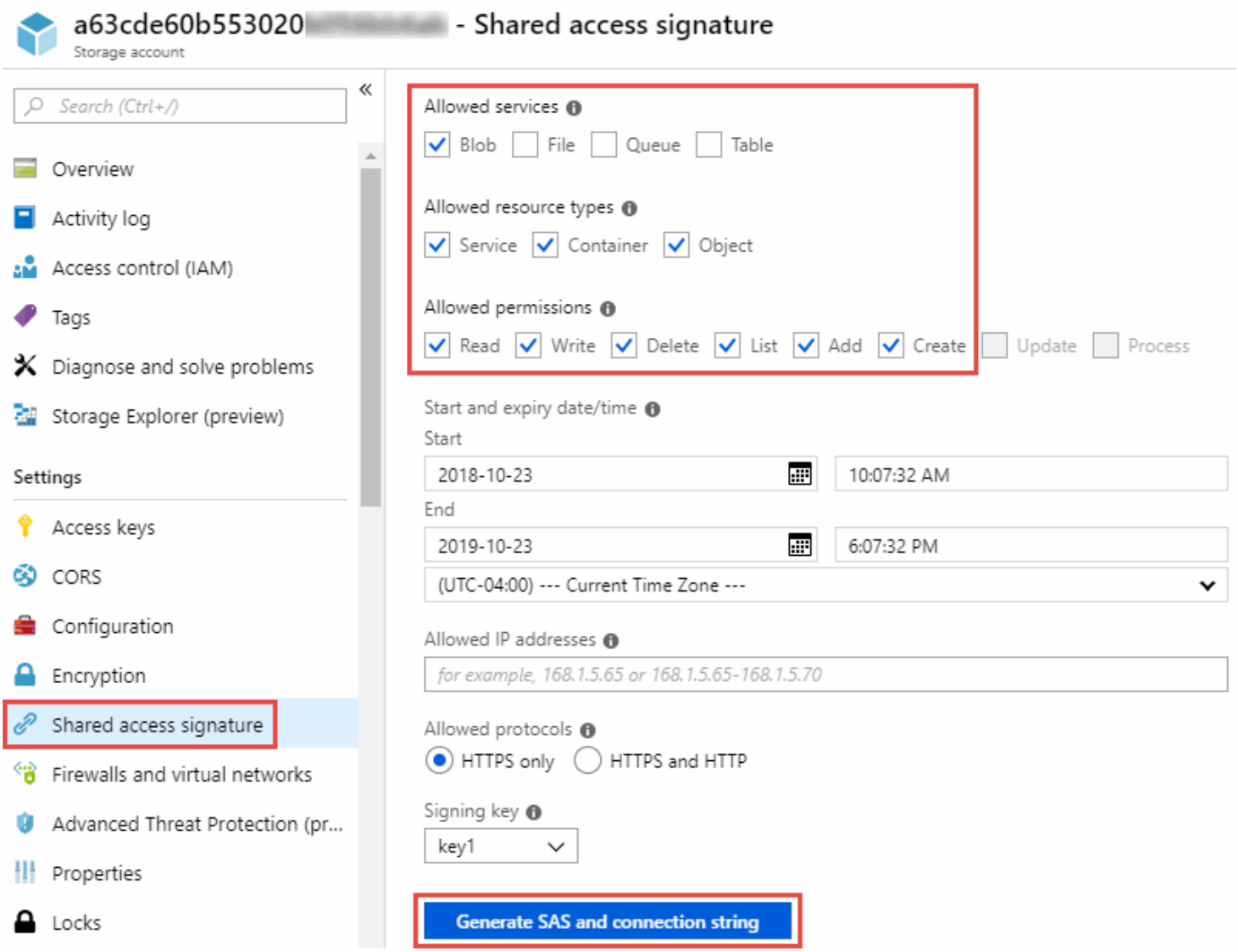
When creating a sync relationship between an Azure Blob container and an NFS or SMB server, you need to provide Cloud Sync with the storage account connection string:



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



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

Supported GCP regions

All regions are supported.

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.

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

Endpoints that are required for Cloud Sync

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

The data broker contacts the following endpoints:

Endpoints	Purpose
olcentgbl.trafficmanager.net:443	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.
rpm.nodesource.com:443 registry.npmjs.org:443 nodejs.org:443	To contact repositories for updating Node.js, npm, and other 3rd party packages used in development.
tgz.pm2.io:443	To access a repository for updating PM2, which is a 3rd party package used to monitor Cloud Sync.
sqs.us-east-1.amazonaws.com:443 kinesis.us-east-1.amazonaws.com:443	To contact the AWS services that Cloud Sync uses for operations (queuing files, registering actions, and delivering updates to the data broker).
s3.*region*.amazonaws.com:443 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.
cf.cloudsync.netapp.com:443 repo.cloudsync.netapp.com:443	To contact the Cloud Sync service.
support.netapp.com:443	To contact NetApp support when using a BYOL license for sync relationships.

Endpoints	Purpose
fedoraproject.org:443	To install 7z on the data broker virtual machine during installation and updates. 7z is needed to send AutoSupport messages to NetApp technical support.
sts.amazonaws.com:443	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.

Web browser endpoints

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

logs.cloudsync.netapp.com:443

Install the data broker

Installing the data broker in AWS

When you create a new data broker, choose the AWS Data Broker option 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 and GovCloud (US) 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.

Installing the data broker

You can install a data broker in AWS when you create a sync relationship.

Steps

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

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

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

   If you already have a data broker, you'll need to click the  icon first.

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.

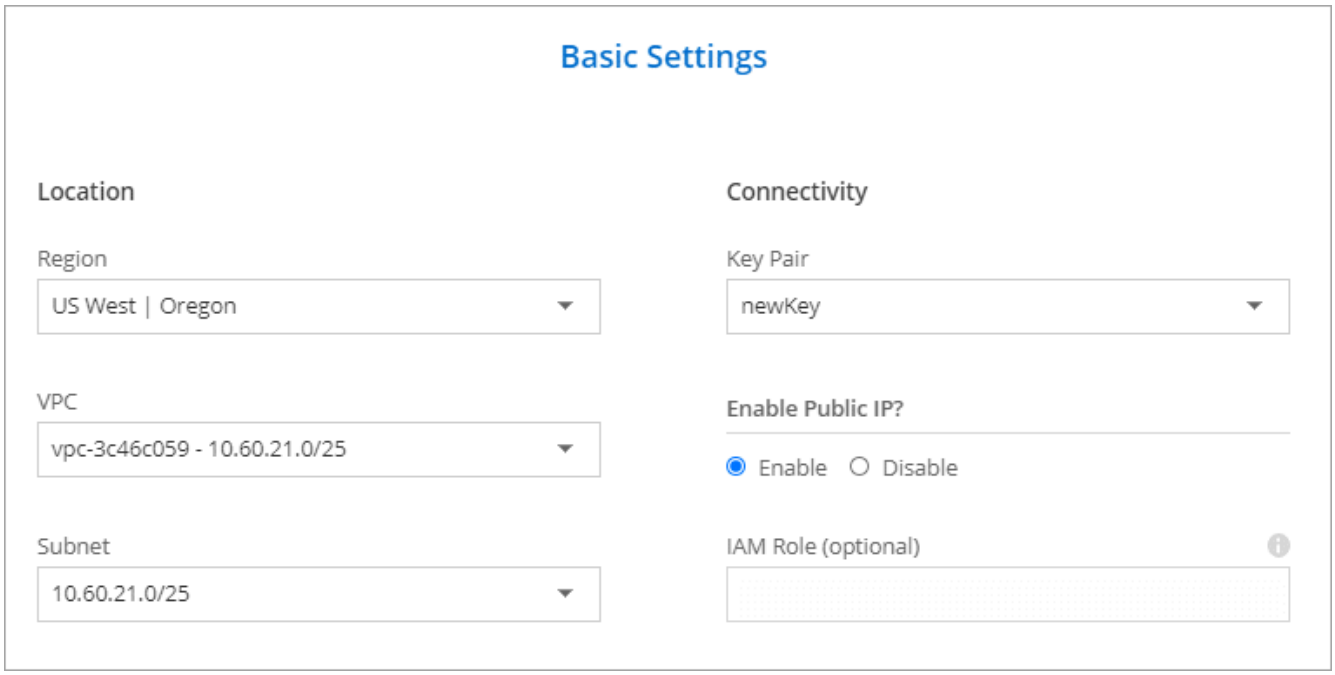
   If you'd rather not provide access keys, click the link at the bottom of the page to use a CloudFormation template instead. When you use this option, you don't need to provide credentials because you are logging in directly to AWS.

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

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

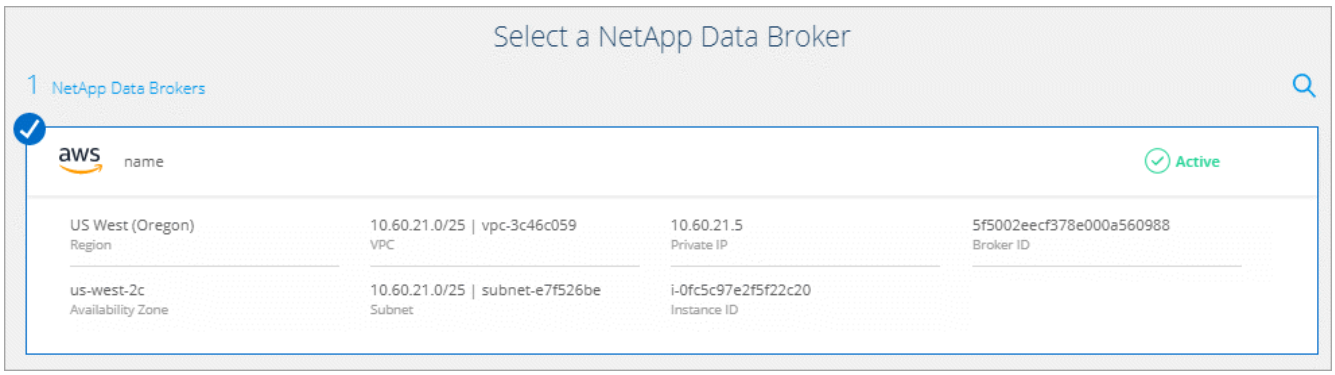
6. If you entered an AWS access key, select a location for the instance, select a key pair, choose whether to enable a public IP address, and then select an existing IAM role, or leave the field blank so Cloud Sync creates the role for you.

   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 with additional sync relationships.

Installing the data broker in Azure

When you create a new data broker, choose the Azure Data Broker option 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.

Installing the data broker

You can install a data broker in Azure when you create a sync relationship.

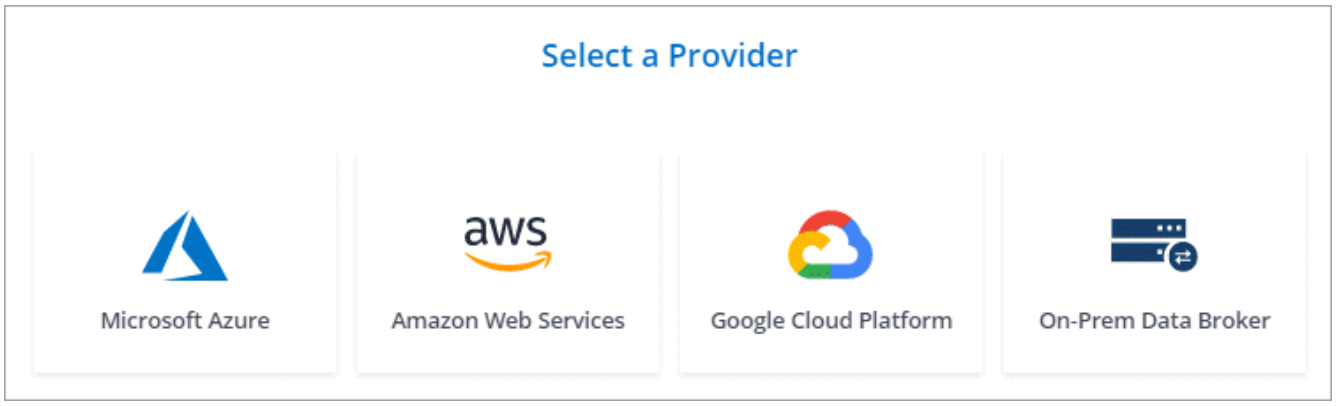
Steps

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

   Complete the pages until you reach the **Data Broker** page.

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

   If you already have a data broker, you'll need to click the + icon first.

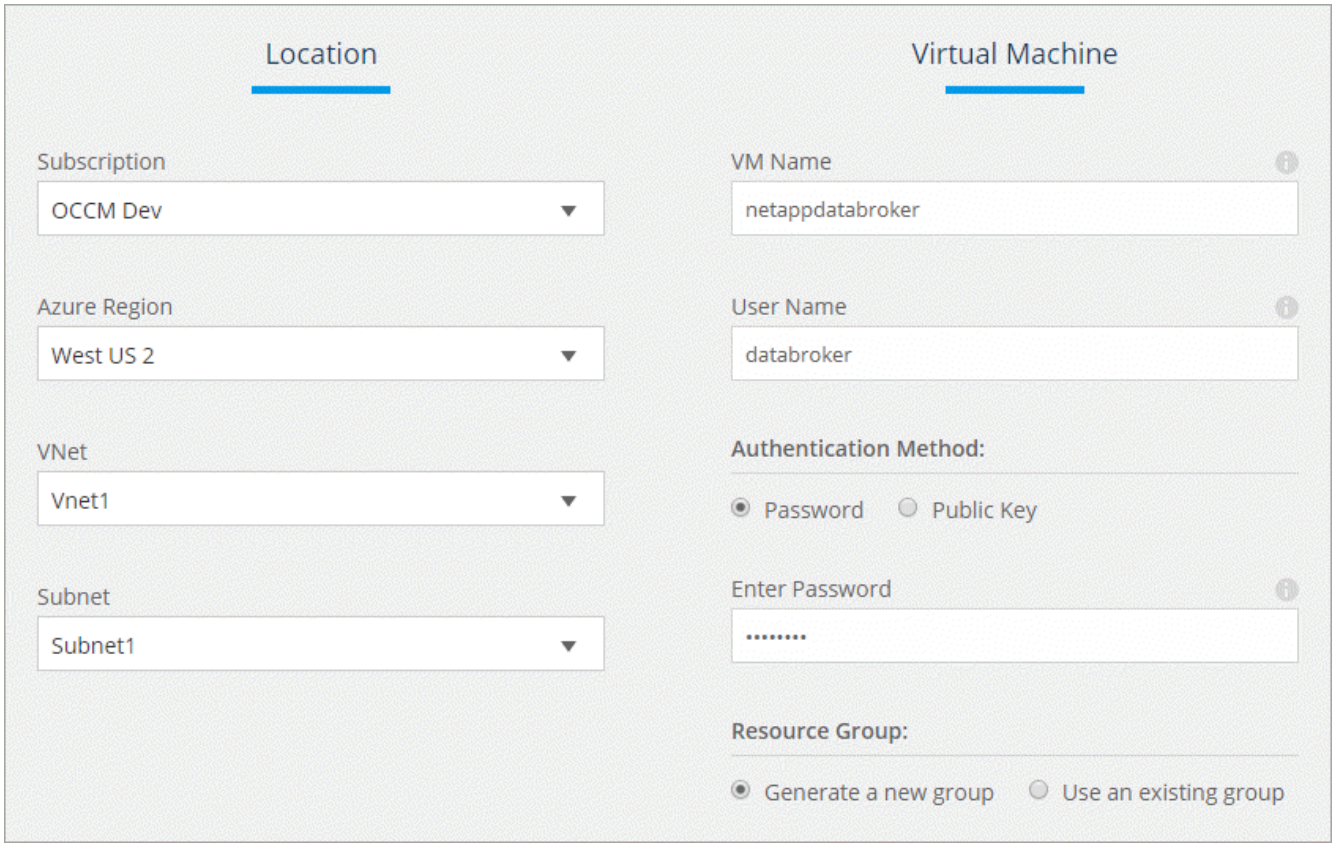
   

4. Enter a name for the data broker and click **Continue**.

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.

Installing the data broker in Google Cloud Platform

When you create a new data broker, choose the GCP Data Broker option to deploy the data broker software on a new virtual machine 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 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.*
```

Installing the data broker

You can install a data broker in GCP when you create a sync relationship.

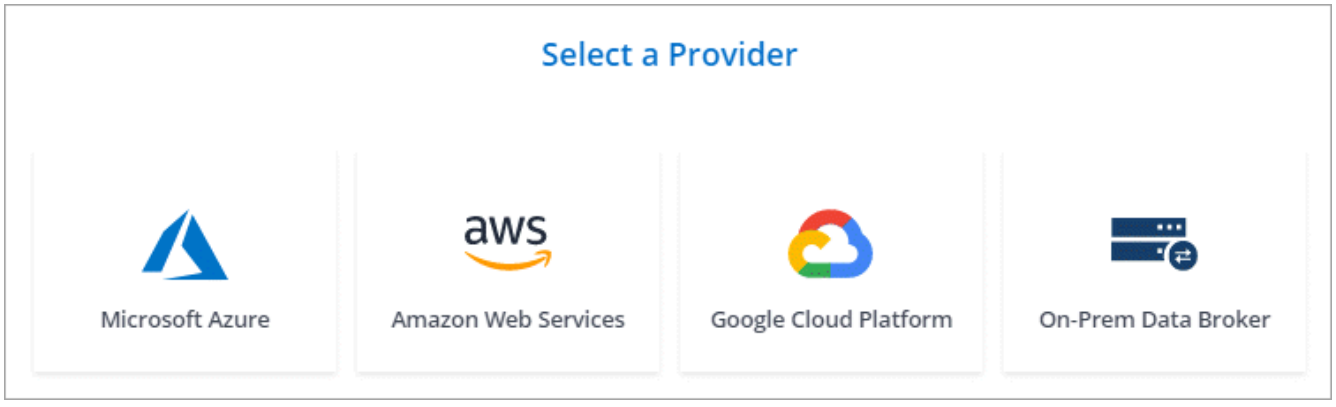
Steps

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

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

3. On the **Data Broker** page, click **Create Data Broker** and then select **Google Cloud Platform**.

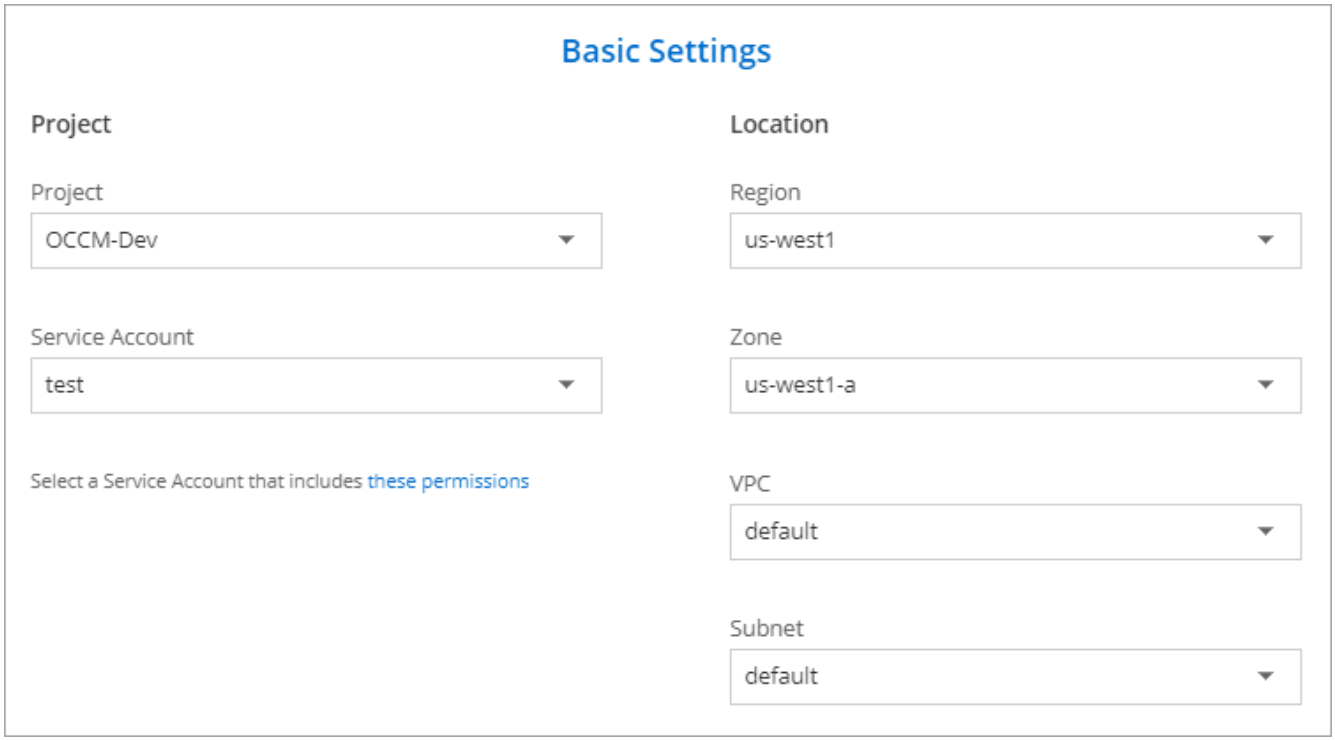
   If you already have a data broker, you'll need to click the + icon first.

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.



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 GCP and created a new sync relationship. You can use this data broker with additional sync relationships.

Installing the data broker on a Linux host

When you create a new data broker, 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
  - 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.

- **OpenSSL**: OpenSSL must be installed on the Linux host.

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.

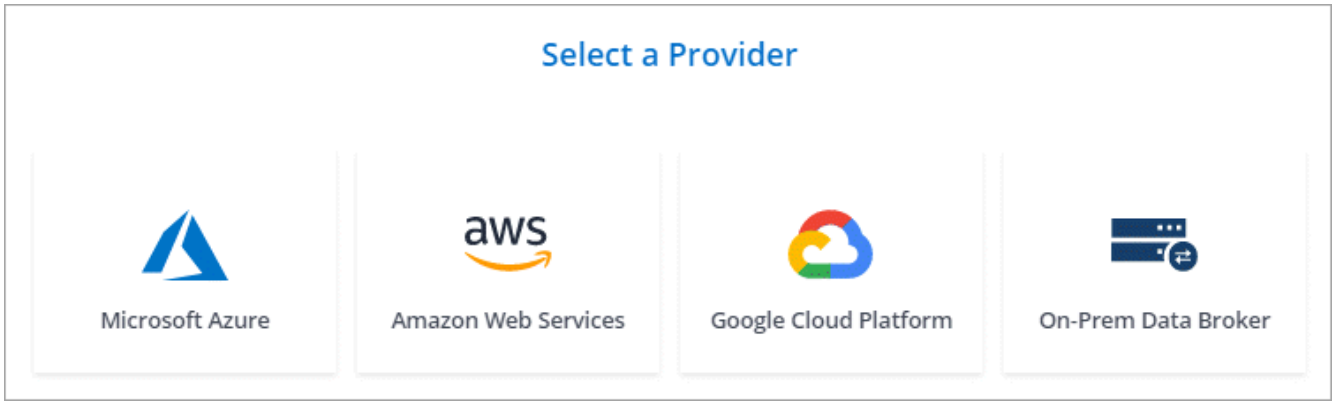
Steps

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

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

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

   If you already have a data broker, you'll need to click the + icon first.



   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.

Creating a sync relationship

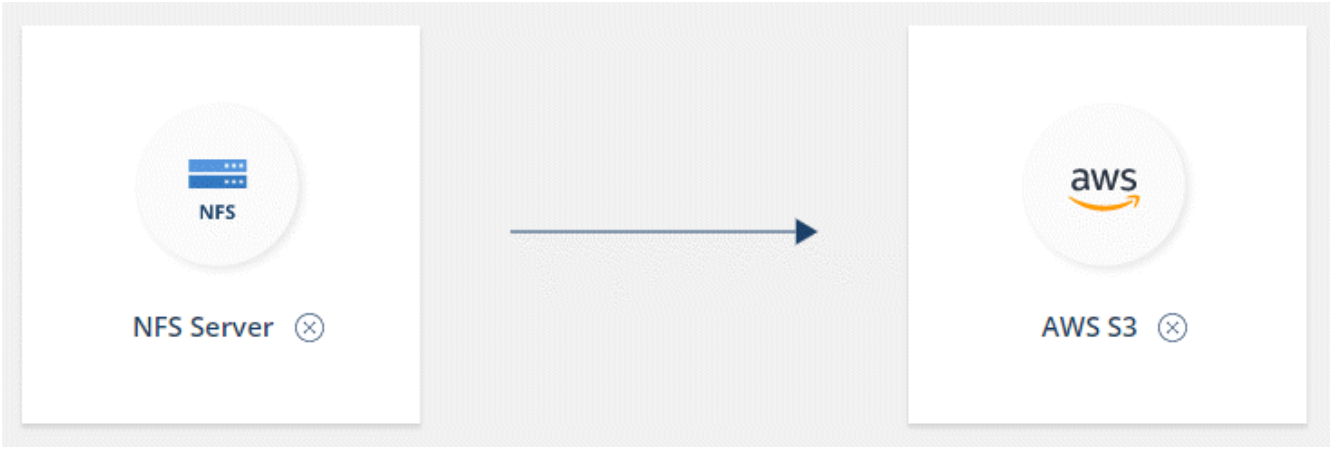
When you create a sync relationship, the Cloud Sync service copies files from the source to the target. After the initial copy, the service syncs any changed data every 24 hours.

The steps below provide an example that shows how to set up a sync relationship from an NFS server to an S3 bucket.

Steps

1. Go to NetApp Cloud Central.
2. Sign up or log in and then start a free trial of Cloud Sync.
3. After you log in, review details about using the service after the free trial ends, and then click **OK**.
4. On the **Select Source & Target** 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.



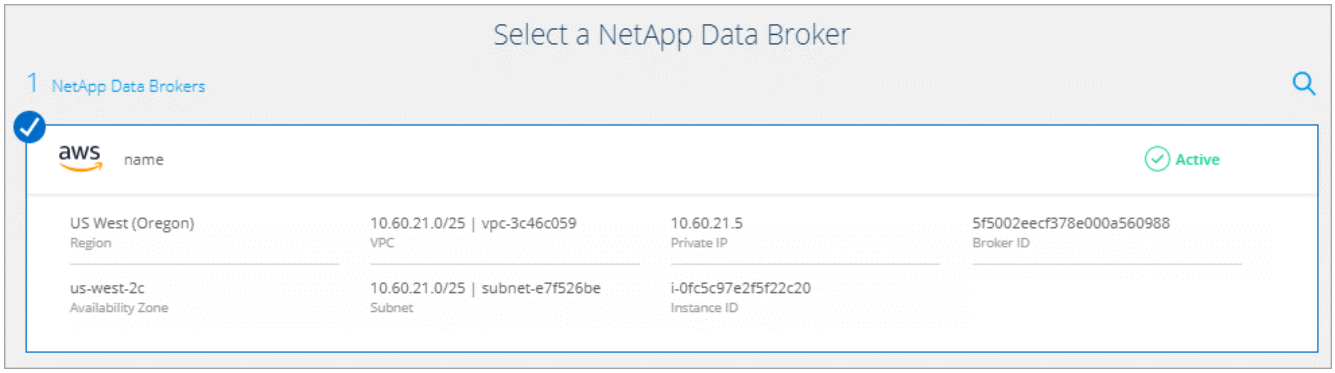
5. Review the details about how the service works and then click **Continue**.
6. On the **NFS Server** page, enter the IP address or fully qualified domain name of the NFS server that you want to sync to AWS.
7. On the **Data Broker** page, follow the prompts to create a data broker virtual machine 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:

- Installing the data broker in AWS
- Installing the data broker in Azure
- Installing the data broker in GCP
- Installing the data broker on a Linux host

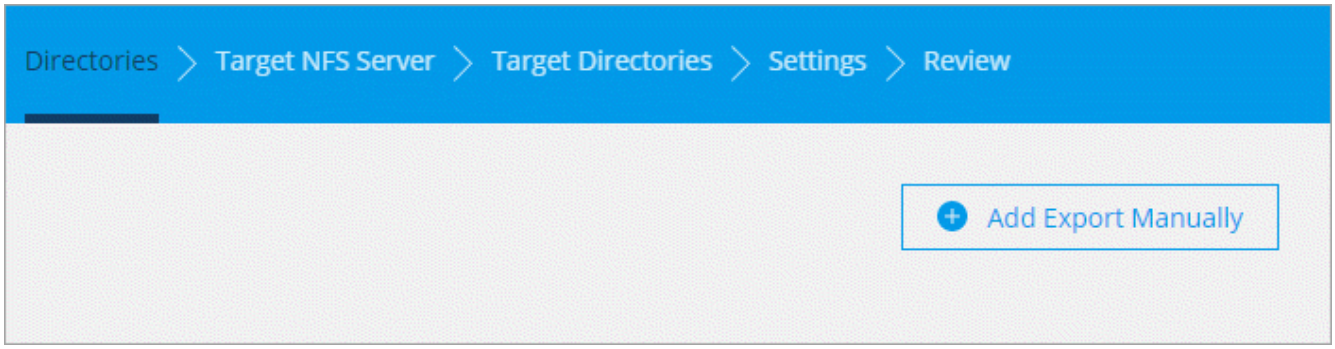
8. After you install the data broker, click **Continue**.

   The following image shows a successfully deployed data broker in AWS:



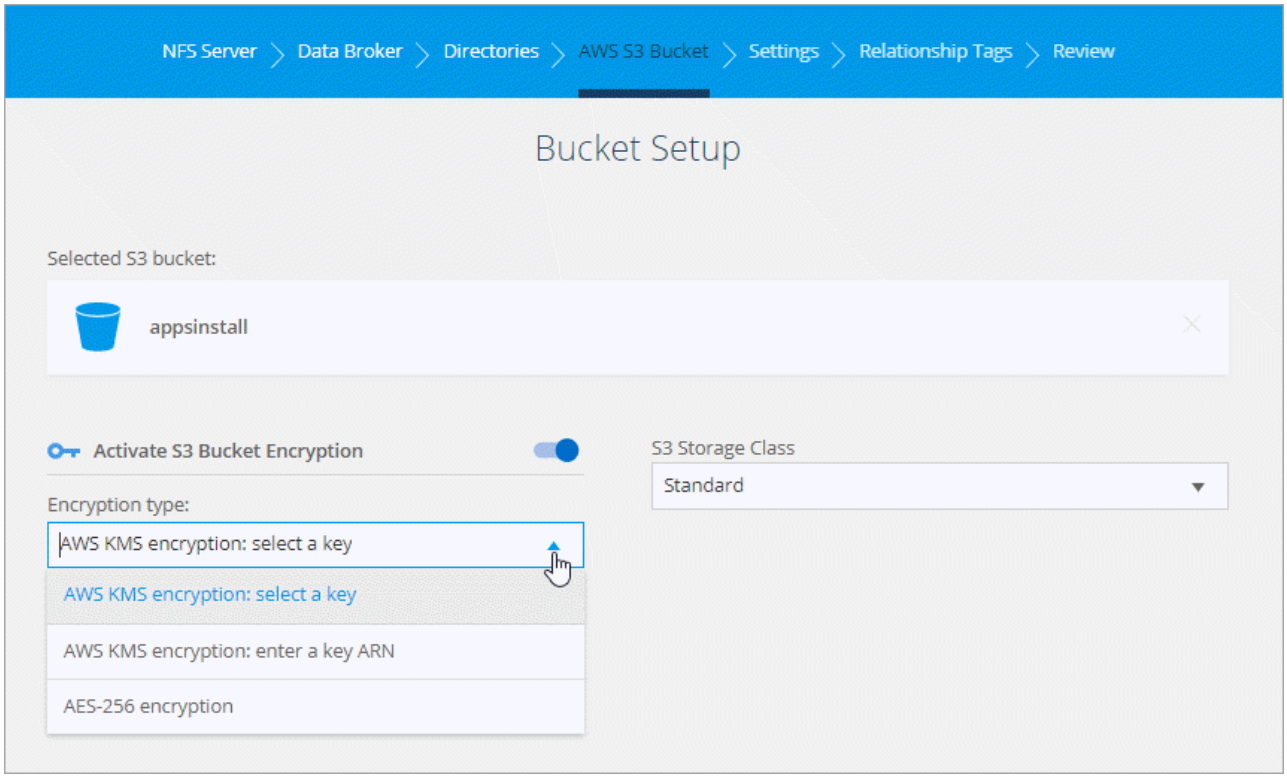
9. On the **Directories** page, select a top-level directory or subdirectory.

   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.

10. 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.
11. 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.



12. 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.

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 change the parameter named *workers.transferrer.delete-on-source* to **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.

Object tagging

When AWS S3 is the target in a sync relationship, Cloud Sync tags S3 objects with metadata that's relevant to the sync operation. You can disable tagging of S3 objects, if it's not desired in your environment. There's no impact to Cloud Sync if you disable tagging—-Cloud Sync just stores the sync metadata in a different way.

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/cloudsync/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.

13. On the **Relationship Tags** page, enter up to 9 relationship tags and then click **Continue**.

    The Cloud Sync service assigns the tags to each object that it syncs to the S3 bucket.

14. Review the details of the sync relationship and then click **Create Relationship**.
15. After the Cloud Sync service successfully creates the relationship, click **View in Dashboard** to view details about the data sync relationship.

Result

Cloud Sync starts syncing data between the source and target.

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 use licenses from NetApp 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 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. Go to the License Settings page.
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/cloudsync/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. Go to the License Settings page.
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/cloudsync/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.

Steps

1. Purchase a license by contacting NetApp.
2. Go to the License Settings page and add the license.