# Install a data broker

## Cloud Sync

NetApp
June 24, 2022

# Table of Contents

# 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

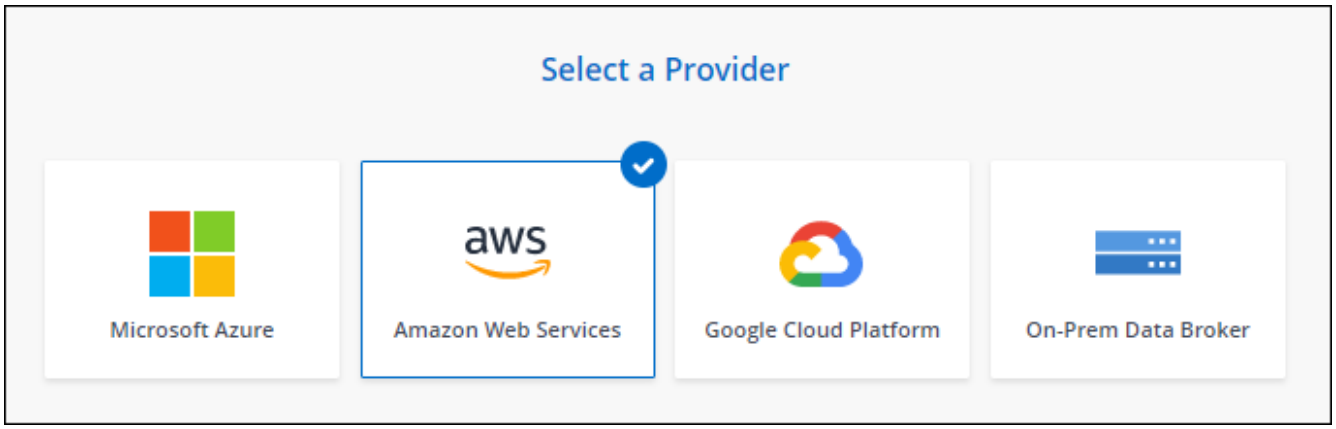
There are a few ways to create a new data broker. These steps describe how to install a data broker in AWS when creating a sync relationship.

### 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.

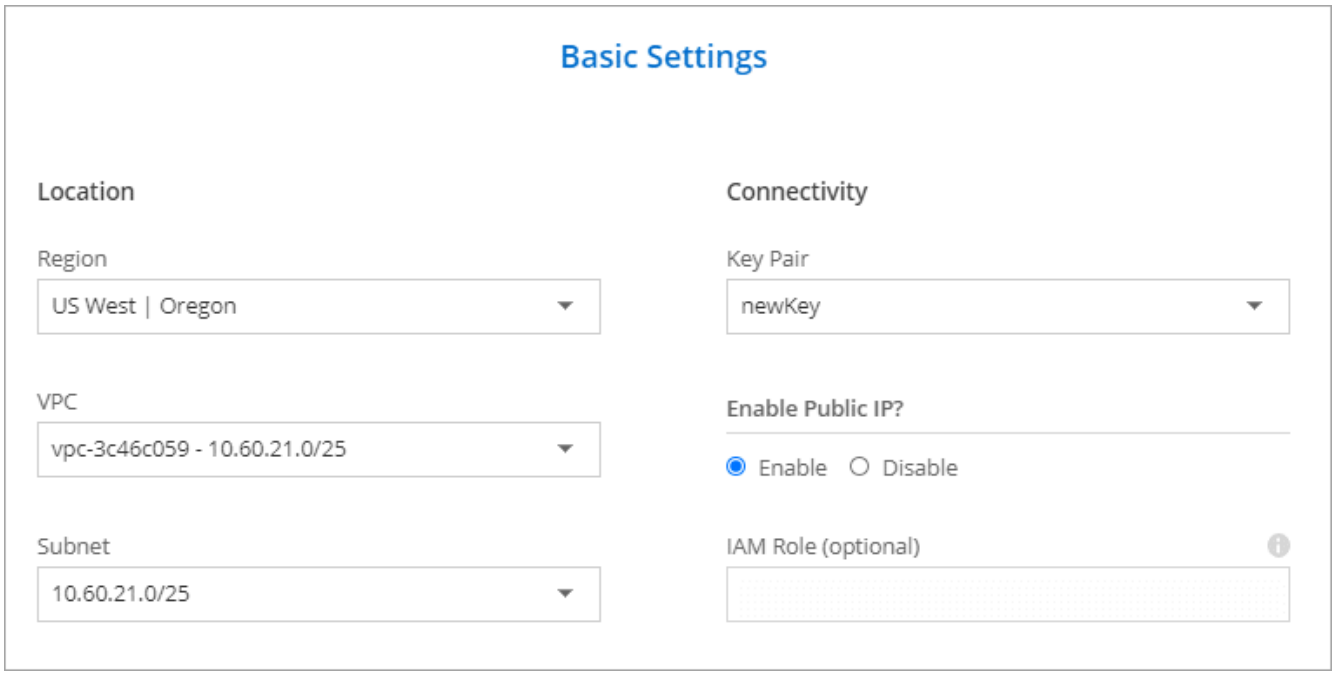
   If you'd rather not provide access keys, click the link at the bottom of the page to use a CloudFormation template instead. When you use this option, you don't need to provide credentials because you are logging in directly to AWS.

   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 *(video)*

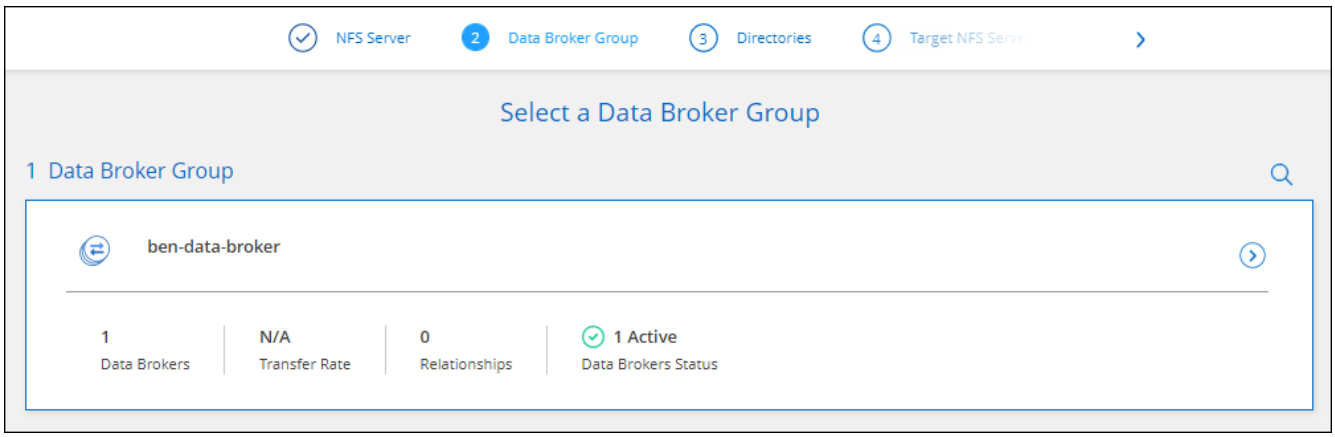
6. If you entered an AWS access key, select a location for the instance, select a key pair, choose whether to enable a public IP address, and then select an existing IAM role, or leave the field blank so Cloud Sync creates the role for you.

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




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

   The following image shows a successfully deployed instance in AWS:




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

**Result**

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

## Details about the data broker instance

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

**Instance type**

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

**vCPUs**
4

**RAM**
16 GB

**Operating system**
Amazon Linux 2

**Disk size and type**
10 GB GP2 SSD

# Creating a new data broker in Azure

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

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

## Supported Azure regions

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

## Networking requirements

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

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

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

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

## Permissions required to deploy the data broker in Azure

Ensure that the Azure user account that you use to deploy the data broker has the following permissions.

```
{
    "Name": "Azure Data Broker",
    "Actions": [
                    "Microsoft.Resources/subscriptions/read",

"Microsoft.Resources/deployments/operationstatuses/read",
                    "Microsoft.Resources/subscriptions/locations/read",
```

```
                    "Microsoft.Network/networkInterfaces/read",
                    "Microsoft.Network/virtualNetworks/subnets/read",

"Microsoft.Resources/subscriptions/resourceGroups/write",

"Microsoft.Resources/subscriptions/resourceGroups/delete",
                    "Microsoft.Resources/deployments/write",
                    "Microsoft.Resources/deployments/validate/action",

"Microsoft.Resources/deployments/operationStatuses/read",
                    "Microsoft.Resources/deployments/cancel/action",
                    "Microsoft.Compute/virtualMachines/read",
                    "Microsoft.Compute/virtualMachines/delete",
                    "Microsoft.Compute/disks/delete",
                    "Microsoft.Network/networkInterfaces/delete",
                    "Microsoft.Network/publicIPAddresses/delete",

"Microsoft.Network/networkSecurityGroups/securityRules/delete",

"Microsoft.Resources/subscriptions/resourceGroups/write",
                    "Microsoft.Compute/virtualMachines/delete",
                    "Microsoft.Network/networkSecurityGroups/write",
                    "Microsoft.Network/networkSecurityGroups/join/action",
                    "Microsoft.Compute/disks/write",
                    "Microsoft.Network/networkInterfaces/write",
                    "Microsoft.Network/virtualNetworks/read",
                    "Microsoft.Network/publicIPAddresses/write",
                    "Microsoft.Compute/virtualMachines/write",
                    "Microsoft.Compute/virtualMachines/extensions/write",
                    "Microsoft.Resources/deployments/read",
                    "Microsoft.Network/networkSecurityGroups/read",
                    "Microsoft.Network/publicIPAddresses/read",

"Microsoft.Network/virtualNetworks/subnets/join/action",
                    "Microsoft.Network/publicIPAddresses/join/action",
                    "Microsoft.Network/networkInterfaces/join/action"
    ],
    "NotActions": [],
    "AssignableScopes": [],
    "Description": "Azure Data Broker",
    "IsCustom": "true"
}
```

## Authentication method

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

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

## Creating the data broker

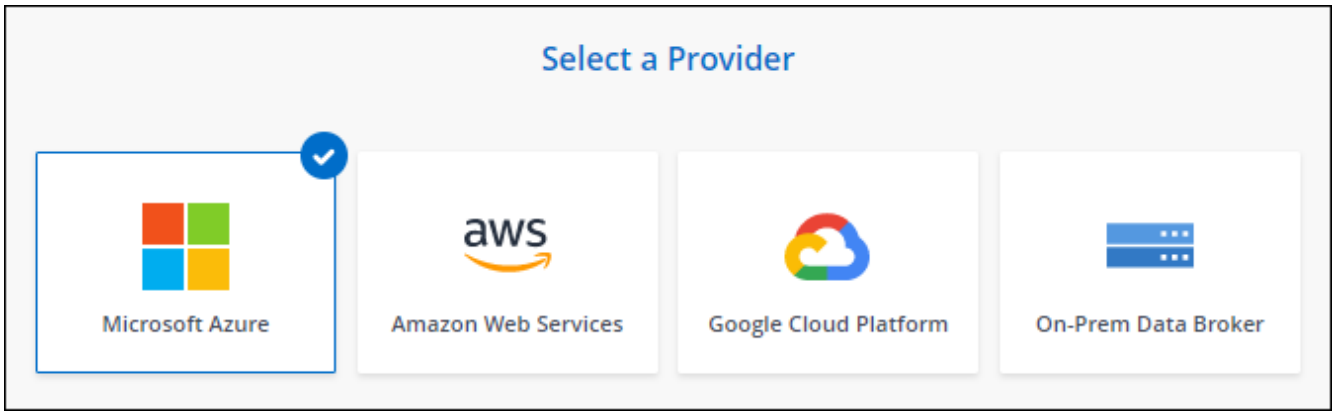
There are a few ways to create a new data broker. These steps describe how to install a data broker in Azure when you create a sync relationship.

### Steps

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

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

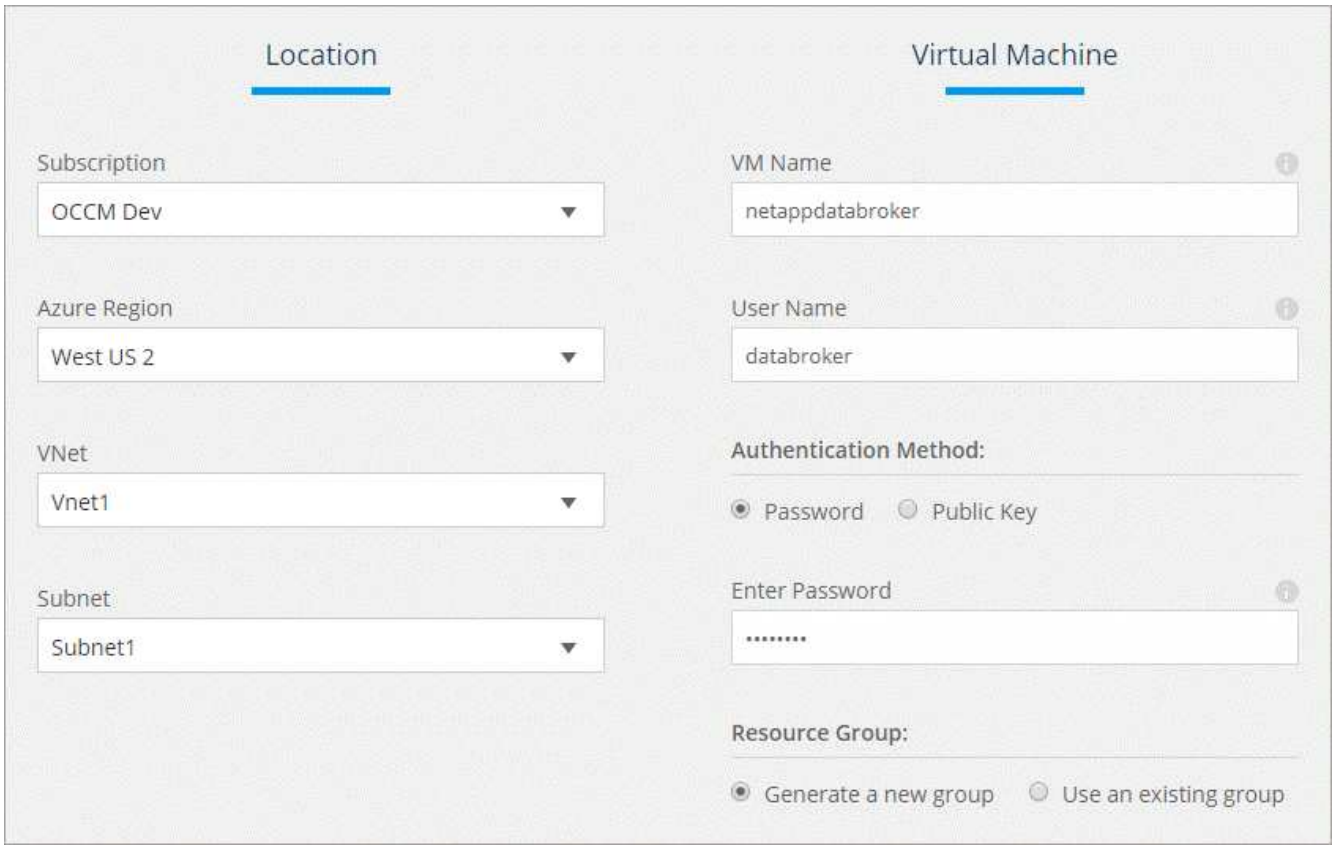
3. On the **Data Broker Group** page, click **Create Data Broker** and then select **Microsoft Azure**.




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

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

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




7. Specify a proxy configuration, if a proxy is required for internet access in the VNet.
8. Click **Continue** and keep the page open until the deployment is complete.

   The process can take up to 7 minutes.

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

**Result**

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

### Getting a message about needing admin consent?

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

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

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

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

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

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

### Details about the data broker VM

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

**VM type**
Standard DS4 v2

**vCPUs**
8

**RAM**
28 GB

**Operating system**
CentOS 7.7

**Disk size and type**
64 GB Premium SSD

# Creating a new data broker in Google Cloud

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

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

## Supported 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

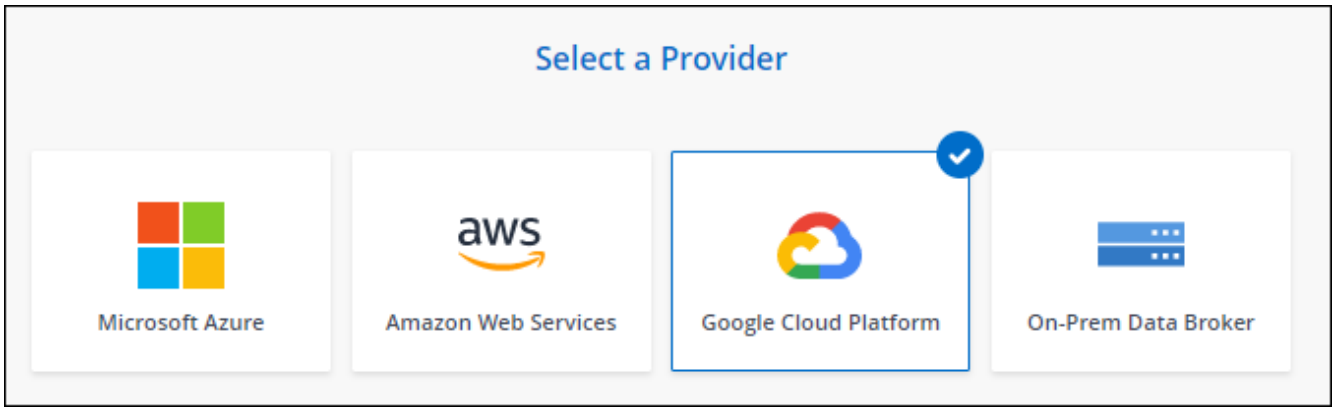
There are a few ways to create a new data broker. These steps describe how to install a data broker in Google Cloud when you create a sync relationship.

**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**.

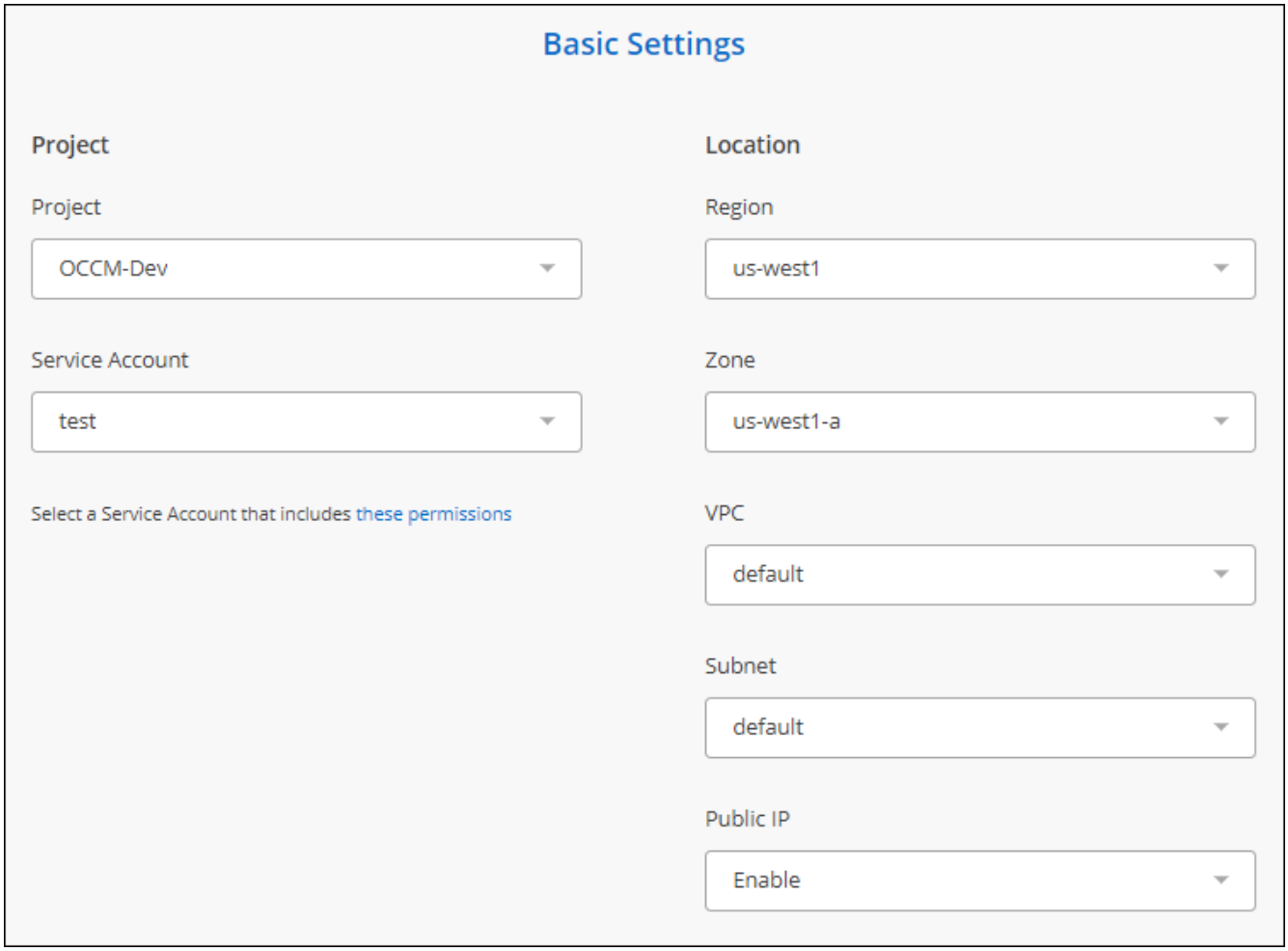



4. Enter a name for the data broker and click **Continue**.
5. If you're prompted, log in with your Google account.

   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.




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

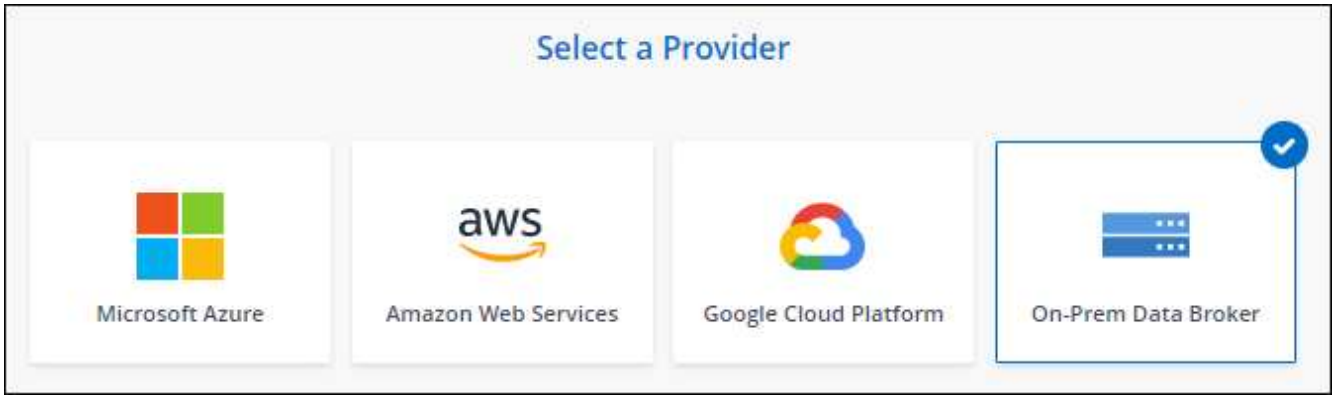
You can install a data broker on a Linux host when you create a sync relationship.

**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.