Install the data broker

Cloud Sync

NetApp
June 05, 2020
# Table of Contents

Install the data broker .................................................................................................................. 1  
Endpoints that the data broker contacts ...................................................................................... 1  
Installing the data broker in AWS ............................................................................................... 2  
Installing the data broker in Azure ............................................................................................. 4  
Installing the data broker in Google Cloud Platform .................................................................... 6  
Installing the data broker on a Linux host ................................................................................... 8
Install the data broker

Endpoints that the data broker contacts

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. If you need to limit outbound connectivity, refer to the following list of endpoints when configuring your firewall for outbound traffic.

The data broker contacts the following endpoints:

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

When you create a sync relationship, 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 creating the AWS CloudFormation stack.

**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 **AWS Data Broker**.
   
   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**.
   
   The AWS CloudFormation Console loads in a separate browser tab. The Cloud Sync service uses a CloudFormation template to quickly provision the data broker instance. The Cloud Sync service does not have access to your AWS credentials because you are logging in directly to AWS.
5. Follow the instructions to install the data broker.
   
   To use your own IAM role, enter the name of the role when you specify details for the AWS CloudFormation stack:

   ```
   IAM role name (Optional)
   The name of an existing IAM role to use with NetApp Data Broker
   ```

   The following video shows how to launch the data broker instance:

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

   The instance takes approximately 5 to 10 minutes to deploy. You can monitor the progress from the AWS CloudFormation console, or you can return to the Cloud Sync service website, which automatically refreshes when the instance is available.

6. After the data broker is available, click **Continue** in Cloud Sync.
   
   The following image shows a successfully deployed instance in AWS:
7. 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 sync relationship, 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
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.

   
   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.
   
   The deployment page loads in a separate browser tab.

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. Click **Deploy** and keep the page open until the deployment is complete. The process can take up to 7 minutes.

8. In Cloud Sync, click **Continue** once the data broker is available.

9. 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.

## Installing the data broker in Google Cloud Platform

When you create a sync relationship, 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 following GCP permissions are in place for the service account that you use to create the data broker:

- Compute Admin
- Deployment Manager Editor
- Service Account User
- Storage Admin

**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 GCP Data Broker.

   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.

The instructions load in a separate browser tab.

5. Follow the instructions to install the data broker in GCP. The instructions are repeated here for your convenience.
   a. In the GCP Console, click Activate Cloud Shell.
   b. Run the following command:

   ```bash
   gcloud deployment-manager deployments create --template <TEMPLATE> <DEPLOYMENT_NAME> --automatic-rollback-on-error --properties "network:<NETWORK>,subnetwork:<SUBNETWORK>,region:<REGION>,zone:<ZONE>,project:<PROJECT>,service-account:<SERVICE_ACCOUNT>"
   ```

   For help with this command, refer to Google Cloud SDK documentation. Note the following about the --template parameter:

   **--template**
   Cloud Sync displays the URI of the template file when you follow the prompts to deploy the GCP Data Broker. That URI isn't repeated here because the link is generated dynamically and can be used only once. Follow steps 1-4 to obtain the URI from Cloud Sync.

6. 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.

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

**Result**
You have 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 sync relationship, 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 18.04 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

**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 **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 in a separate browser tab. You’ll need to follow these instructions—they include a unique link to download the installer.

5. On Step 3 of the instructions page, select whether to enable access to **AWS**, **Google Cloud**, or both.

6. Follow Step 4 on the instructions page 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.

a. Select an installation option: **No proxy**, **Use proxy server**, or **Use proxy server with authentication**.

b. 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.**
c. 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.

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

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

### Demo video

Watch the following video to learn how to install the data broker when S3 is the source or target.

[https://docs.netapp.com/us-en/cloudsync/media/video_on_prem_data_broker.mp4](https://docs.netapp.com/us-en/cloudsync/media/video_on_prem_data_broker.mp4) (video)
Copyright Information

Copyright © 2020 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.