



Concepts

Cloud Sync

NetApp
April 18, 2022

Table of Contents

- Concepts 1
 - Licensing overview 1
 - Data privacy 2
 - Cloud Sync technical FAQ 2

Concepts

Licensing overview

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.

Marketplace subscription

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

Hourly subscription

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

- [View pricing in Azure](#)
- [View pay-as-you-go pricing in AWS](#)

Annual subscription

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

[View annual pricing in AWS](#)

Licenses from NetApp

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

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

[Learn how to purchase licenses and add them to Cloud Sync.](#)

License terms

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

- Customers are entitled to leverage the BYOL license for a term not to exceed one year from the date of delivery.
- Customers are entitled to leverage the BYOL license to establish and not to exceed a total of 20 individual connections between a source and a target (each a "sync relationship").
- A customer's entitlement expires at the conclusion of the one-year license term, irrespective as to whether Customer has reached the 20 sync relationship limitation.

- In the event the Customer chooses to renew its license, unused sync relationships associated from the previous license grant DO NOT roll over to the license renewal.

Data privacy

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

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

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

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

Cloud Sync technical FAQ

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

Getting started

The following questions relate to getting started with Cloud Sync.

How does Cloud Sync work?

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

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

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

How does the 14-day free trial work?

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

How much does Cloud Sync cost?

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

Service charges

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

- [View pay-as-you-go pricing in AWS](#)
- [View annual pricing in AWS](#)
- [View pricing in Azure](#)

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

[Learn more about licenses.](#)



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

Resource charges

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

How is Cloud Sync billed?

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

Can I use Cloud Sync outside the cloud?

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

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

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

What's a data broker group?

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

How do I access Cloud Sync?

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

Supported sources and targets

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

Which sources and targets does Cloud Sync support?

Cloud Sync supports many different types of sync relationships. [View the entire list.](#)

What versions of NFS and SMB does Cloud Sync support?

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

[Learn more about sync requirements.](#)

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

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

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

What about storage tiers for Azure Blob storage?

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

- Hot storage
- Cool storage

Do you support Google Cloud storage tiers?

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

- Standard
- Nearline
- Coldline
- Archive

Networking

The following questions relate to networking requirements for Cloud Sync.

What are the networking requirements for Cloud Sync?

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

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

For more details, [review networking requirements.](#)

Can I use a proxy server with the data broker?

Yes.

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

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

Data synchronization

The following questions relate to how data synchronization works.

How often does synchronization occur?

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

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

What is the minimum sync schedule?

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

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

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

[Learn how to change the settings for a sync relationship.](#)

What if I have a very large dataset?

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

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

Security

The following questions related to security.

Is Cloud Sync secure?

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

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

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

- An AWS Direct Connect, Azure ExpressRoute, or Google Cloud Interconnect connection, which is non-internet routed (and can only communicate with the cloud networks that you specify)

- A VPN connection between your on-premises gateway device and your cloud networks
- For extra secure data transfer with S3 buckets, Azure Blob storage, or Google Cloud Storage, an Amazon Private S3 Endpoint, Azure Virtual Network service endpoints, or Private Google Access may be established.

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

Is data encrypted by Cloud Sync?

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

Cloud Sync cannot encrypt SMB data itself.

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

Permissions

The following questions relate to data permissions.

Are SMB data permissions synced to the target location?

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



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

[Learn how to copy ACLs between SMB shares.](#)

Are NFS data permissions synced to the target location?

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

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

Object storage metadata

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

- Amazon S3 → Amazon S3 ¹
- Amazon S3 → StorageGRID
- StorageGRID → Amazon S3
- StorageGRID → StorageGRID
- StorageGRID → Google Cloud Storage
- Google Cloud Storage → StorageGRID ¹

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

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

Performance

The following questions relate to Cloud Sync performance.

What does the progress indicator for a sync relationship represent?

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

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

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

Limiting the number of concurrent transfers can help. [Contact us for help](#).

Why am I experiencing low performance with Azure NetApp Files?

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

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

[Learn more about Azure NetApp Files service levels and throughput.](#)

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

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

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

How many data brokers are required in a group?

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

Multiple factors can impact data transfer performance. The overall sync performance might be impacted due to

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

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

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

Deleting things

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

What happens if I delete my Cloud Sync relationship?

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

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

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

[Learn how to change the settings for a sync relationship.](#)

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

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

Troubleshooting

[NetApp Knowledgebase: Cloud Sync FAQ: Support and Troubleshooting](#)

Data broker deep dive

The following question relates to the data broker.

Can you explain the architecture of the data broker?

Sure. Here are the most important points:

- The data broker is a node.js application running on a Linux host.
- Cloud Sync deploys the data broker as follows:
 - AWS: From an AWS CloudFormation template
 - Azure: From Azure Resource Manager
 - Google: From Google Cloud Deployment Manager
 - If you use your own Linux host, you need to manually install the software

- The data broker software automatically upgrades itself to the latest version.
- The data broker uses AWS SQS as a reliable and secure communication channel and for control and monitoring. SQS also provides a persistency layer.
- You can add additional data brokers to a group to increase transfer speed and add high availability. There is service resiliency if one data broker fails.

Copyright Information

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

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

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

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

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

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

Trademark Information

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