

AWS DataSync

Overview

Features

Pricing

Getting Started

Resources

FAQs

Customers

[Products](#) / [Migration & Transfer](#) / [AWS DataSync](#) / ...

AWS DataSync FAQs

General

Q: What is AWS DataSync?

A: AWS DataSync is an online data transfer service that simplifies, automates, and accelerates copying large amounts of data between on-premises storage systems and AWS Storage services, as well as between AWS Storage services. DataSync can copy data between Network File System (NFS) shares, Server Message Block (SMB) shares, Hadoop Distributed File Systems (HDFS), self-managed object storage, Google Cloud Storage, Azure Files, [AWS Snowcone](#), Amazon Simple Storage Service ([Amazon S3](#)), Amazon Elastic File System ([Amazon EFS](#)) file systems, [Amazon FSx for Windows File Server](#) file systems, [Amazon FSx for Lustre](#) file systems, and [Amazon FSx for OpenZFS](#) file systems.

Q: Why should I use AWS DataSync?

A: AWS DataSync allows you to copy large datasets with millions of files, without having to build custom solutions with open source tools, or license and manage expensive commercial network acceleration software. You can use DataSync to migrate active data to AWS, archive data to free up on-premises storage capacity, replicate data to AWS for business continuity, or transfer data to the cloud for analysis and processing.

Q: What problem does AWS DataSync solve for me?

A: AWS DataSync reduces the complexity and cost of online data transfer, making it simple to transfer datasets between on-premises, edge, or other cloud storage and AWS Storage services, as well as between AWS Storage services. DataSync connects to existing storage systems and data sources with standard storage protocols (NFS, SMB), as an HDFS client, or using the Amazon S3 API. It uses a purpose-built network protocol and scale-out architecture to accelerate data transfer between storage systems and AWS services. DataSync automatically scales and handles moving files and objects, scheduling data transfers, monitoring the progress of transfers, encryption, verification of data transfers, and notifying customers of any issues. With DataSync you pay only for the amount of data copied, with no minimum commitments or upfront fees.

Q: Where can I move data to and from?

A: DataSync supports the following storage location types: Network File System (NFS) shares, Server Message Block (SMB) shares, Hadoop Distributed File Systems (HDFS), self-managed object storage, Google Cloud Storage, Azure Files, AWS Snowcone, Amazon Simple Storage Service (Amazon S3), Amazon Elastic File System (Amazon EFS) file systems, Amazon FSx for Windows File Server file systems, Amazon FSx for Lustre file systems, and Amazon FSx for OpenZFS file systems.

Q: How do I use AWS DataSync to migrate data to AWS?

A: You can use AWS DataSync to migrate data located on premises, at the edge, or in other clouds to Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, and Amazon FSx for OpenZFS. Configure DataSync to make an initial copy of your entire dataset, and schedule subsequent incremental transfers of changing data until the final cut-over from on-premises to AWS. DataSync includes encryption and integrity validation to help make sure your data arrives securely, intact, and ready to use. To minimize impact on workloads that rely on your network connection, you can schedule your migration to run during off-hours, or limit the amount of network bandwidth that DataSync uses by configuring the built-in bandwidth throttle. DataSync preserves metadata between storage systems that have similar metadata structures, enabling a smooth transition of end users and applications to using your target AWS Storage service. Read the storage blog, "[Migrating storage with AWS DataSync](#)," to learn more about migration best practices and tips.

Q: How do I use AWS DataSync to archive cold data?

A: You can use AWS DataSync to move cold data from on-premises storage systems directly to durable and secure long-term storage, such as [Amazon S3 Glacier Flexible Retrieval](#) (formerly S3 Glacier) or [Amazon S3 Glacier Deep Archive](#). Use DataSync's [filtering functionality](#) to exclude copying temporary files and folders or copying only a subset of files from your source location. You can select the most cost-effective storage service for your needs: transfer data to any [S3 storage class](#), or use DataSync with EFS Lifecycle Management to store data in [Amazon EFS Infrequent Access storage class \(EFS IA\)](#). Use the built-in task scheduling functionality to regularly archive data that should be retained for compliance or auditing purposes, such as logs, raw footage, or electronic medical records.

Q: How do I use AWS DataSync to replicate data to AWS for business continuity?

A: With AWS DataSync, you can periodically replicate files into any Amazon S3 storage classes, or send the data to Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, or Amazon FSx for OpenZFS for a standby file system. Use the built-in task scheduling functionality to ensure that changes to your dataset are regularly copied to your destination storage. Read this AWS Storage blog to [learn more about data protection using AWS DataSync](#).

Q: How do I use AWS DataSync for recurring transfers between on-premises and AWS for ongoing workflows?

A: You can use AWS DataSync for ongoing transfers from on-premises systems into or out of AWS for processing. DataSync can help speed up your critical hybrid cloud storage workflows in industries that need to move active files into AWS quickly. This includes machine learning in life sciences, video production in media and entertainment, big data analytics in financial services, and seismic research in oil and gas. DataSync provides timely delivery to ensure dependent processes are not delayed. You can specify [exclude filters](#), [include filters](#), or [both](#), to determine which files, folders or objects gets transferred each time your task runs.

Q: Can I use AWS DataSync to copy data from other public clouds to AWS?

A: Yes. Using AWS DataSync, you can copy data from Google Cloud Storage using the S3 API, or from [Azure Files using the SMB protocol](#). Simply deploy the DataSync agent in your cloud environment or on Amazon EC2, create your source and destination locations, and then start your task to begin copying data. Learn more about using DataSync to copy data from Google Cloud Storage or from Azure Files.

Q: Can I use AWS DataSync to build my data lake?

A: Yes. With AWS DataSync, you can easily build your data lake, by automating the transfer of on-premises datasets or data in other clouds to Amazon S3. DataSync enables a simple and fast transfer of your entire data set using standard storage protocols (NFS, SMB), as an HDFS client, or the Amazon S3 API. After transferring your initial dataset, you can schedule subsequent transfers of new data to AWS. DataSync includes encryption and integrity validation to help make sure your data arrives securely, intact, and ready to use. To minimize impact on workloads that rely on your network connection, you can schedule transfer tasks to run during off-hours, or limit the amount of network bandwidth that DataSync uses by configuring the built-in bandwidth throttle. Once your data lands in Amazon S3, you can use native AWS services to run big data analytics, artificial intelligence (AI), machine learning (ML), high-performance computing (HPC) and media data processing applications to gain insights from your unstructured data sets. Read the [AWS data lake storage web page](#) to learn more about building and leveraging your data lake.

Q: How do I use AWS DataSync to transfer data between AWS Storage services?

A: You can use DataSync to transfer files or objects between Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, or Amazon FSx for OpenZFS within the same AWS account. You can transfer data between AWS services in the same AWS Region, between services in different Commercial AWS Regions except for China, or between AWS GovCloud (US-East and US-West) Regions. This does not require deploying a DataSync agent, and can be configured end to end using the AWS DataSync console, AWS Command Line Interface (CLI), or AWS Software Development Kit (SDK).

Q: Can I use AWS DataSync to migrate to Amazon WorkDocs?

A: Yes. AWS DataSync accelerates a required step for Amazon WorkDocs Migration Service by [automating file upload to the Amazon S3 bucket that is used for the migration](#). DataSync makes it easier and faster to migrate home directories and department shares to WorkDocs. To learn more about using DataSync for migrations to WorkDocs, read the blog '[Migrating network file shares to Amazon WorkDocs using AWS DataSync](#).'

Usage

Q: How do I get started with AWS DataSync?

A: You can transfer data using AWS DataSync with a few clicks in the [AWS Management Console](#) or through the [AWS Command Line Interface \(CLI\)](#). To get started, follow these 3 steps:

1. To transfer data between on-premises, edge, or other cloud storage systems and AWS Storage services, deploy an agent - Deploy a DataSync agent and associate it to your AWS account via the Management Console or API. The agent will be used to access your NFS server, SMB file share, Hadoop cluster, or self-managed or cloud object storage to read data from it or write data to it. Deploying an agent is not required to transfer data between AWS Storage services within the same AWS account.

2. Create a data transfer task - Create a task by specifying the location of your data source and destination, and any options you want to use to configure the transfer, such as the desired task schedule.

3. Start the transfer - Start the task and monitor data movement in the console or with [Amazon CloudWatch](#).

Q: How do I deploy an AWS DataSync agent?

A: You deploy an AWS DataSync agent to your on-premises hypervisor, in your public cloud environment, or in [Amazon EC2](#). To copy data to or from an on-premises file server, you download the agent virtual machine image from the AWS Console and deploy to your on-premises VMware ESXi, Linux Kernel-based Virtual Machine (KVM), or Microsoft Hyper-V hypervisor. When a DataSync agent is used, the agent must be deployed so that it can access your file server using the NFS, SMB protocol, access NameNodes and DataNodes in your Hadoop cluster, or access your self-managed object storage using the [Amazon S3 API](#). To set up transfers between your S3 on AWS Outposts buckets and S3 buckets in AWS Regions, [deploy the agent on your Outpost](#). To set up transfers between your AWS Snowcone device and AWS storage, use the [DataSync agent AMI that comes pre-installed on your device](#).

Deploying an agent is not required to transfer data between AWS Storage services within the same AWS account. To copy data to or from a self-managed in-cloud file server, or between AWS Storage services in different AWS accounts, you launch an Amazon EC2 instance using a DataSync agent AMI.

Q: What are the resource requirements for the AWS DataSync agent?

A: You can find the minimum required resources to run the agent [here](#).

Q: How do I start an AWS DataSync data transfer task?

A: AWS DataSync copies data when you initiate a task via the [AWS Management Console](#) or [AWS Command Line Interface \(CLI\)](#). Each time a task runs, it scans the source and destination for changes, and performs a copy of any data and metadata differences between the source to the destination. You can configure which characteristics of the source are used to determine what

changed, define [filters to include and exclude specific files or folders](#), and control if files or objects in the destination should be overwritten when changed in the source or deleted when not found in the source.

Q: How does AWS DataSync ensure my data is copied correctly?

A: As AWS DataSync transfers and stores data, it performs integrity checks to ensure the data written to the destination matches the data read from the source. Additionally, an optional verification check can be performed to compare source and destination at the end of the transfer. DataSync will calculate and compare full-file checksums of the data stored in the source and in the destination. You can check either the entire dataset or just the files or objects that DataSync transferred.

Q: How can I monitor the status of data being transferred by AWS DataSync?

A: You can use the AWS Management Console or CLI to monitor the status of data being transferred. Using Amazon CloudWatch Metrics, you can see the [number of files and amount of data which has been copied](#). You can also [enable logging of individual files to CloudWatch Logs](#), to identify what was transferred at a given time, as well as the results of the content integrity verification performed by DataSync. This simplifies monitoring, reporting, and troubleshooting, and enables you to provide timely updates to stakeholders. You can find additional information, such as transfer progress, in the AWS Management Console or CLI.

Q: Can I filter the files and folders that AWS DataSync transfers?

A: Yes. You can specify an exclude filter, an include filter, or both to limit which files, folders, or objects are transferred each time a task runs. Include filters specify the file paths or object keys that should be included when the task runs and limits the scope of what is scanned by DataSync on the source and destination. Exclude filters specify the file paths or object keys that should be excluded from being copied. If no filters are configured, each time a task runs it will transfer all changes from the source to the destination. When creating or updating a task, you can configure both exclude and include filters. When starting a task, you can override filters configured on the task. Read this [AWS storage blog to learn more about using common filters with DataSync](#) for more detail.

Q: Can I configure AWS DataSync to transfer on a schedule?

A: Yes. You can schedule your tasks using the AWS DataSync Console or AWS Command Line Interface (CLI), without needing to write and run scripts to manage repeated transfers. Task scheduling automatically runs tasks on the schedule you configure, with hourly, daily, or weekly options provided directly in the Console. This enables you to ensure that changes to your dataset are automatically detected and copied to your destination storage.

Q: Does AWS DataSync preserve the directory structure when copying files?

A: Yes. When transferring files, AWS DataSync creates the same directory structure on the destination as on the source location's structure.

Q: What happens if an AWS DataSync task is interrupted?

A: If a task is interrupted, for instance, if the network connection goes down or the AWS DataSync agent is restarted, the next run of the task will transfer missing files, and the data will be complete and consistent at the end of this run. Each time a task is started it performs an incremental copy, transferring only the changes from the source to the destination.

Q: Can I use AWS DataSync with AWS Direct Connect?

A: Yes. You can use AWS DataSync with your Direct Connect link to access public service endpoints or [private VPC endpoints](#). When using VPC endpoints, data transferred between the DataSync agent and AWS services does not traverse the public internet or need public IP addresses, increasing the security of data as it is copied over the network.

Q: Does AWS DataSync support VPC endpoints or AWS PrivateLink?

A: Yes. You can [use VPC endpoints](#) to ensure data transferred between your AWS DataSync agent, either deployed on-premises or in-cloud, doesn't traverse the public internet or need public IP addresses. Using VPC endpoints increases the security of your data by keeping network traffic within your [Amazon Virtual Private Cloud \(Amazon VPC\)](#). VPC endpoints for DataSync are powered by [AWS PrivateLink](#), a highly available, scalable technology that enables you to privately connect your VPC to supported AWS services.

Q: How do I configure AWS DataSync to use VPC endpoints?

A: To use VPC endpoints with AWS DataSync, you create an [AWS PrivateLink](#) interface VPC endpoint for the DataSync service in your chosen VPC, and then choose this endpoint elastic network interface (ENI) when creating your DataSync agent. Your agent will connect to this ENI to activate, and subsequently all data transferred by the agent will remain within your configured VPC. You can use either the [AWS DataSync Console](#), AWS Command Line Interface (CLI), or AWS SDK, to configure VPC endpoints. To learn more, see [Using AWS DataSync in a Virtual Private Cloud](#).

Moving to and from Amazon S3

Q: Can I copy my data into Amazon S3 Glacier Flexible Retrieval (formerly S3 Glacier), Amazon S3 Glacier Deep Archive, or other S3 storage classes?

A: Yes. When configuring an S3 bucket for use with AWS DataSync, you can select the S3 storage class that DataSync uses to store objects. DataSync supports storing data directly into S3 Standard, S3 Intelligent-Tiering, S3 Standard-Infrequent Access (S3 Standard-IA), S3 One Zone-Infrequent Access (S3 One Zone-IA), Amazon S3 Glacier Flexible Retrieval, and Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive). More information on [Amazon S3 storage classes](#) can be found in the [Amazon Simple Storage Service Developer Guide](#).

Objects smaller than the minimum charge capacity per object will be stored in S3 Standard. For example, folder objects, which are zero-bytes in size and hold only metadata, will be stored in S3 Standard. Read about [considerations when working with Amazon S3 storage classes](#) in our documentation, and for more information on minimum charge capacities see [Amazon S3 Pricing](#).

Q: Can I copy data out of S3 Standard-IA and S3 One Zone-IA storage classes?

A: Yes. When using S3 as the source location for an AWS DataSync task, the service will retrieve all objects from the bucket which need to be copied to the destination. Retrieving objects from S3 Standard-IA and S3 One Zone-IA storage will incur a retrieval fee based on the size of the objects. Read about [considerations when working with Amazon S3 storage classes](#) in our documentation.

Q: Can I copy data out of Amazon S3 Glacier Flexible Retrieval (formerly S3 Glacier) and Amazon S3 Glacier Deep Archive?

A: When using S3 as the source location for an AWS DataSync task, the service will attempt to retrieve all objects from the bucket which need to be copied to the destination. Retrieving objects which are archived in the S3 Glacier Flexible Retrieval or S3 Glacier Deep Archive storage class results in an error. Any errors retrieving archived objects will be logged by DataSync and will result in a failed task completion status. Read about [considerations when working with Amazon S3 storage classes](#) in our documentation.

Q: How does AWS DataSync access my Amazon S3 bucket?

A: AWS DataSync assumes an IAM role that you provide. The policy you attach to the role determines which actions the role can perform. DataSync can auto generate this role on your behalf or you can [manually configure a role](#).

Q: How does AWS DataSync convert files and folders to or from objects in Amazon S3?

A: When files or folders are copied to Amazon S3, there is a one-to-one relationship between a file or folder and an object. File and folder timestamps and POSIX permissions, including user ID, group ID, and permissions, are stored in [S3 user metadata](#). For NFS shares, file metadata stored in S3 user metadata is fully interoperable with [File Gateway](#), providing on-premises file-based access to data stored in Amazon S3 by AWS DataSync.

When DataSync copies objects that contain this user metadata back to an NFS server, the file metadata is restored. Symbolic links and hard links are also restored when copying back from NFS to S3.

When copying from an SMB file share, default POSIX permissions are stored in S3 user metadata. When copying back to an SMB file share, ownership is set based on the user that was configured in DataSync to access that file share, and default permissions are assigned.

When copying from HDFS, file and folder timestamps, user and group ownership, and POSIX permissions are stored in S3 user metadata. When copying from Amazon S3 back to HDFS, file and folder metadata are restored.

Learn more about how [DataSync stores files and metadata](#) in our documentation.

Q: What object metadata is preserved when transferring objects between self-managed object storage and Amazon S3?

A: When transferring objects between self-managed object storage and Amazon S3, DataSync copies objects together with object metadata and tags.

Q: What object metadata is preserved when transferring objects between Amazon S3 buckets?

A: When transferring objects between Amazon S3 buckets, DataSync copies objects together with object metadata and tags. DataSync does not copy other object information such as object ACLs or prior object versions.

Q: Which Amazon S3 request and storage costs apply when using S3 storage classes with AWS DataSync?

A: Some S3 storage classes have behaviors that can affect your cost, such as data retrieval, minimum storage capacities, and minimum storage durations. DataSync automates management of data to address these factors, and provides settings to minimize data retrieval.

To avoid minimum capacity charge per object, AWS DataSync automatically stores small objects in S3 Standard. To minimize data retrieval fees, you can configure DataSync to verify only files that were transferred by a given task. To avoid minimum storage duration charges, DataSync has

controls for overwriting and deleting objects. Read about [considerations when working with Amazon S3 storage classes](#) in our documentation.

Q: Can I copy object data to and from Amazon S3 buckets on AWS Outposts?

A: Yes. You can copy objects between Amazon S3 on AWS Outposts and Amazon S3 buckets in AWS Regions. AWS DataSync copies objects together with object metadata and object tags. For DataSync to access your Amazon S3 on Outposts buckets, [deploy a DataSync EC2 agent on your Outpost](#).

When using DataSync with Amazon S3 on Outposts, you can only transfer data to and from Amazon S3 buckets in AWS Regions. You can learn more about supported sources and destinations for DataSync tasks in our [documentation](#).

Moving to and from Amazon EFS

Q: How does AWS DataSync access my Amazon EFS file system?

A: AWS DataSync accesses your Amazon EFS file system using the NFS protocol. The DataSync service mounts your file system from within your VPC from Elastic Network Interfaces (ENIs) managed by the DataSync service. DataSync fully manages the creation, use, and deletion of these ENIs on your behalf. You can choose to mount your EFS file system using a mount target or an [EFS Access Point](#).

Q: Can I use AWS DataSync with all Amazon EFS storage classes?

A: Yes. You can use AWS DataSync to copy files into Amazon EFS and configure [EFS Lifecycle Management](#) to migrate files that have not been accessed for a set period of time to the Infrequent Access (IA) storage class.

Q: How do I use AWS DataSync with Amazon EFS file system resource policies?

A: You can use both IAM identity policies and resource policies to control client access to Amazon EFS resources in a way that is scalable and optimized for cloud environments. When you create a DataSync location for your EFS file system, you can specify an IAM role that DataSync will assume when accessing EFS. You can then use [EFS file system policies](#) to configure access for the IAM role. Because DataSync mounts EFS file systems as the root user, your IAM policy must allow the following action: `elasticfilesystem:ClientRootAccess`.

Q: Can I use AWS DataSync to replicate my Amazon EFS file system to a different AWS Region?

A: Yes. In addition to the built-in [replication](#) provided by Amazon EFS, you can also use AWS DataSync to schedule periodic replication of your Amazon EFS file system to a second Amazon EFS file system within the same AWS account. This capability is available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Q: What metadata is preserved when copying data between an NFS share and Amazon EFS, or between two Amazon EFS file systems?

A: AWS DataSync copies file and folder timestamps and POSIX permissions, including user ID, group ID, and permissions. You can learn more and see the complete list of copied metadata in our [documentation](#).

Q: What metadata is preserved when copying data between HDFS and Amazon EFS?

A: AWS DataSync copies file and folder timestamps and POSIX permissions and applies default values for user ID and group ID. You can learn more and see the complete list of copied metadata in our [documentation](#).

Moving to and from Amazon FSx for Windows File Server

Q: How does AWS DataSync access my Amazon FSx file system?

A: AWS DataSync accesses your Amazon FSx file system using the SMB protocol, authenticating with the username and password you configure in the AWS Console or CLI. The DataSync service mounts your file system from within your VPC from Elastic Network Interfaces (ENIs) managed by the DataSync service. DataSync fully manages the creation, use, and deletion of these ENIs on your behalf.

Q: What Windows metadata is transferred when copying between an SMB share to Amazon FSx for Windows File Server file system, or between two Amazon FSx file systems?

A: AWS DataSync copies Windows metadata, including file timestamps, file owner, standard file attributes, NTFS discretionary access lists (DACLS), and NTFS system access control lists (SACLs). You can learn more and see the complete list of copied metadata in our [documentation](#).

Q: Can I use AWS DataSync to replicate my Amazon FSx for Windows File Server file system to a different AWS Region?

A: Yes. You can use AWS DataSync to schedule periodic replication of your Amazon FSx for Windows File Server file system to a second file system within the same AWS account. This

capability is available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Moving to and from Amazon FSx for Lustre

Q: How does AWS DataSync access my Amazon FSx for Lustre file system?

A: When you create a DataSync task to copy to or from your FSx for Lustre file system, the DataSync service will create Elastic Network Interfaces (ENIs) in the same VPC and subnet where your file system is located. DataSync uses these ENIs to access your FSx for Lustre file system using the Lustre protocol as the root user. When you create a DataSync location resource for your FSx for Lustre file system, you can specify up to five security groups to apply to the ENIs and configure outbound access from the DataSync service. The security groups must be configured to allow outbound traffic on the [network ports required by FSx for Lustre](#). The security groups on your FSx for Lustre file system should be configured to allow inbound access from the security groups you assigned to the DataSync location resource for your FSx for Lustre file system.

Q: What metadata is preserved when either copying data between an NFS share or Amazon EFS file system and Amazon FSx for Lustre, or between two Amazon FSx for Lustre file systems?

A: AWS DataSync copies file and folder timestamps and POSIX permissions, including user ID, group ID, and permissions. You can learn more and see the complete list of copied metadata in our [documentation](#).

Q: Can I use AWS DataSync to migrate data from one FSx for Lustre file system to another?

A: Yes. You can use AWS DataSync to copy from your FSx for Lustre file system to a second file system within the same AWS account. This capability is available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Q: Can I use AWS DataSync to replicate my Amazon FSx for Lustre file system to a different AWS Region?

A: Yes. You can use AWS DataSync to schedule periodic replication of your Amazon FSx for Lustre file system to a second file system within the same AWS account. This capability is available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Q: Will DataSync copy the striping or layout settings when copying from one Amazon FSx for Lustre file system to another?

A: No. Files are written using the file layout and striping configuration on the destination's file system.

Moving data to and from Amazon FSx for OpenZFS

Q: How does AWS DataSync access my Amazon FSx for OpenZFS file system?

A: When you create a DataSync task to copy to or from your FSx for OpenZFS file system, the DataSync service will create Elastic Network Interfaces (ENIs) in the same VPC and subnet where your file system is located. DataSync uses these ENIs to access your FSx for OpenZFS file system using the OpenZFS protocol as the root user. When you create a DataSync location resource for your FSx for OpenZFS file system, you can specify up to five security groups to apply to the ENIs and configure outbound access from the DataSync service. The security groups must be configured to allow outbound traffic on the [network ports required by FSx for OpenZFS](#). The security groups on your FSx for OpenZFS file system should be configured to allow inbound access from the security groups you assigned to the DataSync location resource for your FSx for OpenZFS file system.

Q: What metadata is preserved when either copying data between an NFS share or Amazon EFS file system and Amazon FSx for OpenZFS, or between two Amazon FSx for OpenZFS file systems?

A: AWS DataSync copies file and folder timestamps and POSIX permissions, including user ID, group ID, and permissions. You can learn more and see the complete list of copied metadata in our [documentation](#).

Q: Can I use AWS DataSync to migrate data from one FSx for OpenZFS file system to another?

A: Yes. You can use AWS DataSync to copy from your FSx for OpenZFS file system to a second file system within the same AWS account. This capability is available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Q: Can I use AWS DataSync to replicate my Amazon FSx for OpenZFS file system to a different AWS Region?

A: Yes. You can use AWS DataSync to schedule periodic replication of your Amazon FSx for OpenZFS file system to a second file system within the same AWS account. This capability is

available for both same-region and cross-region deployments, and does not require using a DataSync agent.

Transferring to and from AWS Snowcone

Q: How do I transfer data between AWS Snowcone and AWS storage services?

A: The DataSync agent is pre-installed on your [Snowcone](#) device as an AMI. To transfer data online to AWS, connect the AWS Snowcone device to the external network and use [AWS OpsHub](#) or the CLI to launch the DataSync agent AMI. Activate the agent using the AWS Management Console or CLI, and set up your online data transfer task between AWS Snowcone's NFS store, and Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, or Amazon FSx for Lustre.

Performance

Q: How fast can AWS DataSync copy my file system to AWS?

A: The rate at which AWS DataSync can copy a given dataset is a function of amount of data, I/O bandwidth achievable from the source and destination storage, network bandwidth available, and network conditions. For data transfer between on premises and AWS Storage services, a single DataSync task is capable of fully utilizing a 10 Gbps network link.

Q: Can I control the amount of network bandwidth that an AWS DataSync task uses?

A: Yes. You can control the amount of network bandwidth that AWS DataSync will use by [configuring the built-in bandwidth throttle](#). You can increase or decrease this limit while your data transfer task is running. This enables you to minimize impact on other users or applications who rely on the same network connection.

Q: How can I monitor the performance of AWS DataSync?

A: AWS DataSync generates [Amazon CloudWatch Metrics to provide granular visibility into the transfer process](#). Using these metrics, you can see the number of files and amount of data which has been copied, as well as file discovery and verification progress. You can see [CloudWatch Graphs](#) with these metrics directly in the DataSync Console.

Q: Will AWS DataSync affect the performance of my source file system?

A: Depending on the capacity of your on-premises file store, and the quantity and size of files to be transferred, AWS DataSync may affect the response time of other clients when accessing the

same source data store, because the agent reads or writes data from that storage system.

[Configuring a bandwidth limit](#) for a task will reduce this impact by limiting the I/O against your storage system.

Security and compliance

Q: Is my data encrypted while being transferred and stored?

A: Yes. All data transferred between the source and destination is encrypted via Transport Layer Security (TLS), which replaced Secure Sockets Layer (SSL). Data is never persisted in AWS DataSync itself. The service supports using [default encryption for S3 buckets](#), [Amazon EFS file system encryption of data at rest](#), and [Amazon FSx For Windows File Server encryption at rest and in transit](#).

Q: How does AWS DataSync access my NFS server or SMB file share?

A: AWS DataSync uses an agent that you deploy into your IT environment or into Amazon EC2 to access your files through the NFS or SMB protocol. This agent connects to DataSync service endpoints within AWS, and is securely managed from the AWS Management Console or CLI.

Q: How does AWS DataSync access HDFS on my Hadoop cluster?

A: AWS DataSync uses an agent that you deploy into your IT environment or into Amazon EC2 to access your Hadoop cluster. The DataSync agent acts as an HDFS client and communicates with the NameNodes and DataNodes in your clusters. When you start a task, DataSync queries the primary NameNode to determine the locations of files and folders on the cluster. DataSync then communicates with the DataNodes in the cluster to copy files and folders to, or from, HDFS.

Q: How does AWS DataSync access my self-managed or cloud object storage?

A: AWS DataSync uses an agent that you deploy into your IT environment or into Amazon EC2 to access your objects using the Amazon S3 API. This agent connects to DataSync service endpoints within AWS, and is securely managed from the AWS Management Console or CLI.

Q: Does AWS DataSync require setting up a VPN to connect to my destination storage?

A: No. When copying data to or from your premises, there is no need to setup a VPN/tunnel or allow inbound connections. Your AWS DataSync agent can be configured to route through a firewall using standard network ports. You can also [deploy DataSync within your Amazon Virtual Private Cloud \(Amazon VPC\)](#) using VPC endpoints. When using VPC endpoints, data transferred

between the DataSync agent and AWS services does not need to traverse the public internet or need public IP addresses.

Q: How do my AWS DataSync agents securely connect to AWS?

A: Your AWS DataSync agent connects to DataSync service endpoints within your chosen [AWS Region](#). You can choose to have the agent connect to public internet facing endpoints, Federal Information Processing Standards (FIPS) validated endpoints, or endpoints within one of your VPCs. Activating your agent securely associates it with your AWS account. To learn more, see [Choose a Service Endpoint](#) and [Activate Your Agent](#).

Q: How is my AWS DataSync agent patched and updated?

A: Updates to the agent VM, including both the underlying operating system and the AWS DataSync software packages, are automatically applied by AWS once the agent is activated. Updates are applied non-disruptively when the agent is idle and not executing a data transfer task.

Q: Which compliance programs does AWS DataSync support?

A: AWS has the longest-running compliance program in the cloud. AWS is committed to helping customers navigate their requirements. AWS DataSync has been assessed to meet global and industry security standards. DataSync complies with [PCI DSS](#), [ISO 9001](#), [27001](#), [27017](#), and [27018](#); [SOC 1, 2, and 3](#); in addition to being [HIPAA eligible](#). DataSync is also authorized in the AWS US East/West Regions under FedRAMP Moderate and in the AWS GovCloud (US) Regions under FedRamp High. That makes it easier for you to verify our security and meet your own obligations. For more information and resources, visit our [compliance pages](#). You can also go to the [Services in Scope by Compliance Program page](#) to see a full list of services and certifications.

Q: Is AWS DataSync PCI compliant?

A: Yes. AWS DataSync is [PCI-DSS](#) compliant, which means you can use it to transfer payment information. You can download the PCI Compliance Package in [AWS Artifact](#) to learn more about how to achieve PCI Compliance on AWS.

Q: Is AWS DataSync HIPAA eligible?

A: Yes. AWS DataSync is HIPAA eligible, which means if you have a [HIPAA BAA in place with AWS](#), you can use DataSync to transfer protected health information (PHI).

Q: Does AWS DataSync have FedRAMP JAB Moderate Provisional Authorization in the AWS US East/West?

A: Yes. AWS DataSync has received a Provisional Authority to Operate (P-ATO) from the Joint Authorization Board (JAB) at the Federal Risk and Authorization Management Program (FedRAMP) Moderate baseline in the US East/West Regions. If you are a federal or commercial customer, you can use AWS DataSync in the AWS East/West Region's authorization boundary with data up to the moderate impact level.

Q: Does AWS DataSync have FedRAMP JAB High Provisional Authorization in the AWS GovCloud (US) Regions?

A: Yes. AWS DataSync has received a Provisional Authority to Operate (P-ATO) from the Joint Authorization Board (JAB) at the Federal Risk and Authorization Management Program (FedRAMP) High baseline in the US GovCloud Region. If you are a federal or commercial customer, you can use AWS DataSync in the AWS GovCloud (US) Region's authorization boundary with data up to the high impact level.

When to choose AWS DataSync

Q: How is AWS DataSync different from using command line tools such as rsync or the Amazon S3 command line interface?

A: AWS DataSync fully automates and accelerates moving large active datasets to AWS. It is natively integrated with Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, Amazon FSx for OpenZFS, [Amazon CloudWatch](#), and [AWS CloudTrail](#), which provides seamless and secure access to your storage services, as well as detailed monitoring of the transfer.

DataSync uses a purpose-built network protocol and scale-out architecture to transfer data. For data transfer between on premises and AWS Storage services, a single DataSync task is capable of fully utilizing a 10 Gbps network link.

DataSync fully automates the data transfer. It comes with retry and network resiliency mechanisms, network optimizations, built-in task scheduling, monitoring via the DataSync API and Console, and CloudWatch metrics, events and logs that provide granular visibility into the transfer process. DataSync performs data integrity verification both during the transfer and at the end of the transfer.

DataSync provides end to end security, and integrates directly with AWS storage services. All data transferred between the source and destination is encrypted via TLS, and access to your AWS storage is enabled via built-in AWS security mechanisms such as IAM roles. DataSync with VPC endpoints are enabled to ensure that data transferred between an organization and AWS

does not traverse the public internet, further increasing the security of data as it is copied over the network.

Q: To transfer objects between my buckets, when do I use AWS DataSync, when do I use S3 Replication, and when do I use S3 Batch Operations?

A: AWS provides multiple tools to copy objects between your buckets.

Use AWS DataSync for ongoing data distribution, data pipelines, and data lake ingest, as well as for consolidating or splitting data between multiple buckets.

Use [S3 Replication](#) for continuous replication of data to a specific destination bucket.

Use [S3 Batch Operations](#) for large-scale batch operations on S3 objects, such as to copy objects, set object tags or access control lists (ACLs), initiate object restores from Amazon S3 Glacier Flexible Retrieval (formerly S3 Glacier), invoke an AWS Lambda function to perform custom actions using your objects, manage S3 Object Lock legal hold, or manage S3 Object Lock retention dates.

Q: When do I use AWS DataSync and when do I use AWS Snowball Edge?

A: AWS DataSync is ideal for online data transfers. You can use DataSync to migrate active data to AWS, transfer data to the cloud for analysis and processing, archive data to free up on-premises storage capacity, or replicate data to AWS for business continuity.

[AWS Snowball Edge](#) is ideal for offline data transfers, for customers who are bandwidth constrained, or transferring data from remote, disconnected, or austere environments.

Q: When do I use AWS DataSync and when do I use AWS Storage Gateway?

A: Use AWS DataSync to migrate existing data to Amazon S3, and subsequently use the File Gateway configuration of [AWS Storage Gateway](#) to retain access to the migrated data and for ongoing updates from your on-premises file-based applications.

You can use a combination of DataSync and File Gateway to minimize your on-premises infrastructure while seamlessly connecting on-premises applications to your cloud storage. AWS DataSync enables you to automate and accelerate online data transfers to AWS Storage services. After the initial data transfer phase using AWS DataSync, File Gateway provides your on-premises applications with low latency access to the migrated data. When using DataSync with NFS shares, POSIX metadata from your source on-premises storage is preserved, and permissions from the source storage apply when accessing your files using File Gateway.

Q: When do I use AWS DataSync, and when do I use Amazon S3 Transfer Acceleration?

A: If your applications are already integrated with the Amazon S3 API, and you want higher throughput for transferring large files to S3, you can use [S3 Transfer Acceleration](#). If you want to transfer data from existing storage systems (e.g. Network Attached Storage), or from instruments that cannot be changed (e.g. DNA sequencers, video cameras), or if you want multiple destinations, you use AWS DataSync. DataSync also automates and simplifies the data transfer by providing additional functionality, such as built-in retry and network resiliency mechanisms, data integrity verification, and flexible configuration to suit your specific needs, including bandwidth throttling, etc.

Q: When do I use AWS DataSync and when do I use AWS Transfer Family?

A: If you currently use SFTP to exchange data with third parties, [AWS Transfer Family](#) provides a fully managed SFTP, FTPS, and FTP transfer directly into and out of Amazon S3, while reducing your operational burden.

If you want an accelerated and automated data transfer between NFS servers, SMB file shares, Hadoop clusters, self-managed or cloud object storage, AWS Snowcone, Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, and Amazon FSx for OpenZFS, you can use AWS DataSync. DataSync is ideal for customers who need online migrations for active data sets, timely transfers for continuously generated data, or replication for business continuity.



Learn more about pricing

AWS DataSync has simple, predictable, usage-based pricing; you pay only for the amount of data that you copy.

[Learn more »](#)



Sign up for a free account

Instantly get access to the AWS Free Tier.

[Sign up »](#)



Start moving data in the Console

Get started building with AWS DataSync in the AWS Console.

[Sign in »](#)

Get Started with AWS

Learn how to start using AWS in minutes



What's New with AWS

Learn about the latest products, services, and more



Learn Cloud Fundamentals in 3 Hours | 9 June, 2022

Get started with AWS Cloud through step-by-step guides and video tutorials. [Register now »](#)



[Sign In to the Console](#)

Learn About AWS

[What Is AWS?](#)

[What Is Cloud Computing?](#)

[AWS Inclusion, Diversity & Equity](#)

[What Is DevOps?](#)

[What Is a Container?](#)

[What Is a Data Lake?](#)

[AWS Cloud Security](#)

[What's New](#)

Resources for AWS

[Getting Started](#)

[Training and Certification](#)

[AWS Solutions Portfolio](#)

[Architecture Center](#)

[Product and Technical FAQs](#)

[Analyst Reports](#)

[AWS Partners](#)

Developers on AWS

[Developer Center](#)

[SDKs & Tools](#)

[.NET on AWS](#)

[Python on AWS](#)

[Java on AWS](#)

[PHP on AWS](#)

[JavaScript on AWS](#)

[Blogs](#)

[Press Releases](#)

Help

[Contact Us](#)

[File a Support Ticket](#)

[Knowledge Center](#)

[AWS re:Post](#)

[AWS Support Overview](#)

[Legal](#)

[AWS Careers](#)

[Sign In to the Console](#)



Amazon is an Equal Opportunity Employer: *Minority / Women / Disability / Veteran / Gender Identity / Sexual Orientation / Age.*

Language

[عربي |](#)

[Bahasa Indonesia |](#)

[Deutsch |](#)

[English |](#)

[Español |](#)

[Français |](#)

[Italiano |](#)

[Português |](#)

[Tiếng Việt |](#)

[Türkçe |](#)

[Русский |](#)

[ไทย |](#)

[日本語 |](#)

[한국어 |](#)

[中文 \(简体\) |](#)

[中文 \(繁體\)](#)

[Privacy](#)

[|](#)

[Site Terms](#)

[|](#)

[Cookie Preferences](#)

