

# Service Name

AWS DataSync

## Service Overview

AWS DataSync is an online data transfer service that simplifies and accelerates data migrations to AWS and facilitates data movement between on-premises storage, edge locations, other cloud providers, and AWS Storage services. DataSync automates data-transfer processes and infrastructure requirements for high performance and secure data transfers. It uses a purpose-built network protocol and parallel, multi-threaded architecture to accelerate transfers while ensuring end-to-end security through encryption and data integrity validation. DataSync supports various storage systems including NFS, SMB, HDFS, self-managed object storage, cloud providers like Azure and Google Cloud, and AWS services like S3, EFS, and FSx.

## Key Use Cases

- **Data Migration:** Transferring active datasets from on-premises or other cloud providers to AWS storage services for cloud adoption
- **Data Archiving:** Moving cold data from on-premises storage to durable and secure long-term storage classes like S3 Glacier Flexible Retrieval or S3 Glacier Deep Archive
- **Data Replication:** Copying data to AWS for business continuity, disaster recovery, and creating backup copies across regions
- **Hybrid Cloud Workflows:** Enabling recurring data transfers between on-premises and AWS for ongoing processing and analysis
- **Data Lake Building:** Transferring large datasets to AWS for analytics, machine learning, and data processing workloads

## Features

- **Purpose-built Network Protocol:** Uses AWS-designed transfer protocol for optimized data movement with parallel, multi-threaded architecture
- **Network Optimization:** Performs incremental transfers, in-line compression, sparse file detection, and bandwidth controls
- **Multi-cloud Support:** Supports data movement between on-premises, AWS, and other cloud providers including Azure, Google Cloud, and object storage systems
- **Enhanced and Basic Modes:** Enhanced mode offers higher performance and parallel processing, while Basic mode supports broader compatibility
- **Security and Encryption:** Encrypts data in transit using TLS 1.3, supports Kerberos authentication, and ensures data integrity validation

- **Metadata Preservation:** Preserves file permissions, timestamps, ownership, and other metadata during transfers
- **Scheduling and Automation:** Built-in scheduling mechanism for periodic data transfer tasks and automated transfer management
- **Monitoring and Auditing:** Provides JSON-formatted task reports, CloudWatch integration, and CloudTrail auditing capabilities

## Value Proposition

AWS DataSync eliminates the complexity and cost of building custom data transfer solutions or licensing expensive commercial network acceleration software. It provides up to 10 times faster transfer speeds than open-source tools while automating data transfer processes, infrastructure management, and ensuring security:

- DataSync offers a pay-as-you-go pricing model with no upfront costs, supports virtually unlimited numbers of files, and handles encryption, verification, and monitoring automatically
- The service integrates seamlessly with AWS storage services and provides granular bandwidth controls to minimize impact on production systems

## Service Integration

DataSync integrates deeply with core AWS storage services including Amazon S3 (all storage classes), Amazon EFS, Amazon FSx for Windows File Server, Amazon FSx for Lustre, Amazon FSx for OpenZFS, and Amazon FSx for NetApp ONTAP. It leverages AWS IAM for access control, AWS Secrets Manager for credential storage, Amazon CloudWatch for monitoring and metrics, AWS CloudTrail for auditing, and supports AWS PrivateLink for private connectivity. DataSync also integrates with AWS Backup for centralized backup policies and works with data analytics services like AWS Glue, Amazon Athena, and Amazon QuickSight for transferred data analysis.

## Onboarding and Offboarding

Getting started with DataSync involves signing up for an AWS account, creating an IAM user with appropriate permissions, and choosing the transfer method based on source and destination. For on-premises transfers, customers deploy and activate a DataSync agent on VMware ESXi, KVM, Nutanix AHV, Microsoft Hyper-V, or Amazon EC2:

- For cloud-to-cloud transfers, no agent may be required
- Customers then create source and destination locations, configure transfer tasks with options like bandwidth controls and scheduling, and start the transfer using the DataSync console, CLI, or API

Getting Started Guide: <https://docs.aws.amazon.com/datasync/latest/userguide/setting-up.html>

## Data Removal

DataSync does not store customer data permanently - it only transfers data between source and destination locations. To offboard, customers can delete DataSync tasks, locations, and agents through the console or API. Any transferred data remains in the destination storage service and must be deleted separately according to that service's procedures. Customers should also remove any associated IAM roles and policies created specifically for DataSync operations.

## Backup/Restore and Disaster Recovery

DataSync enables backup and disaster recovery by replicating data to AWS storage services across regions. It supports automated replication schedules for data protection and can transfer data to various S3 storage classes including Glacier for long-term archival. DataSync integrates with AWS Backup for centralized backup management and can replicate Amazon EFS and FSx file systems to different AWS regions for disaster recovery purposes.

### Backup Capabilities

DataSync provides backup capabilities through scheduled recurring transfers and integrates with AWS Backup service for centralized policy management. It supports transferring data to Amazon S3 storage classes optimized for backup scenarios, including S3 Glacier Instant Retrieval, S3 Glacier Flexible Retrieval, and S3 Glacier Deep Archive. DataSync can perform incremental backups by transferring only changed data.

### Disaster Recovery

DataSync supports disaster recovery by enabling cross-region replication of data to AWS storage services. It can replicate Amazon EFS file systems, FSx file systems, and other data to different AWS regions. DataSync does not directly integrate with AWS Elastic Disaster Recovery but serves as a data replication tool that supports broader disaster recovery strategies by ensuring data availability across regions.

### Recovery Objectives

DataSync helps customers meet RPO objectives through scheduled incremental transfers that can run as frequently as every hour, minimizing data loss. For RTO objectives, DataSync's high-speed transfer capabilities and automation reduce the time needed to restore data to recovery locations. The service's ability to transfer data at speeds up to 10 times faster than traditional tools helps achieve aggressive recovery time targets.

## Service Constraints

Service Quotas: <https://docs.aws.amazon.com/datasync/latest/userguide/datasync-limits.html>

FAQ: <https://aws.amazon.com/datasync/faqs/>

## Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/datasync/latest/userguide/what-is-datasync.html>

User Guide: <https://docs.aws.amazon.com/datasync/latest/userguide/setting-up.html>

API Reference:

[https://docs.aws.amazon.com/datasync/latest/userguide/API\\_Reference.html](https://docs.aws.amazon.com/datasync/latest/userguide/API_Reference.html)

## Pricing Overview

Please see the AWS UK G Cloud 15 Pricing Document accompanying this service in the Digital Marketplace.

Public Pricing: <https://aws.amazon.com/datasync/pricing/>

## Cost Optimization

DataSync offers several cost optimization features including pay-as-you-go pricing with no upfront costs, bandwidth throttling to control network usage, incremental transfers that only move changed data, and built-in compression to reduce data transfer volumes. Enhanced mode provides better performance per dollar with parallel processing. Customers can optimize costs by scheduling transfers during off-peak hours, using appropriate S3 storage classes for destinations, and leveraging AWS Direct Connect to reduce data transfer charges for large migrations.

## Savings Considerations

Main cost savings include eliminating the need for expensive commercial network acceleration software, reducing operational overhead through automation, and minimizing network bandwidth usage through compression and incremental transfers. DataSync's speed improvements (up to 10x faster than open-source tools) reduce the time and resources needed for data migration projects. The service's ability to transfer data directly to cost-effective S3 storage classes like Glacier can significantly reduce long-term storage costs compared to maintaining on-premises infrastructure.