

Introducing AWS DataSync: simplify, automate,
and accelerate your online data transfer

Agenda

Why did we build AWS DataSync?

What is AWS DataSync?

How do start using AWS DataSync?

Cox Automotive use case

Demo

As more and more **critical workloads** move to the cloud...



...you need to
move
increasingly
large datasets
along with them

What use cases need online data transfer?



Online data transfer



Migration of active
application data



Transfers for timely
in-cloud processing



Replication for
data protection
and recovery

How do you solve this today?

Solution #1: Build DIY from open source tools

But you have to worry about ...

End-to-end security Optimizing and managing bandwidth

Developing and debugging scripts Handling failures and retry

Network routing Installing and updating software Reporting and logging

Infrastructure scale-out Running in-cloud infrastructure

Data integrity & validation Integrating with cloud services

... it's complicated and time consuming

How do you solve this today?

Solution #2 License commercial software

But these are not managed or AWS-integrated...

Software license and maintenance costs

Deploy in-cloud infrastructure Install, update, and patch software

Manage secure access to AWS storage Integrate with AWS CloudTrail for audit

Integrate with Amazon CloudWatch for metrics, logging, and events

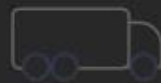
Configure networking, VPCs, network ACLs, and security groups

... it's complicated and costly

AWS data transfer & hybrid storage



Online data transfer



Offline data transfer



Hybrid storage

Edge locations for S3 enabled applications



Amazon S3 Transfer Acceleration

Load streaming data into Amazon S3



Amazon Kinesis Data Firehose

Managed file transfers into Amazon S3



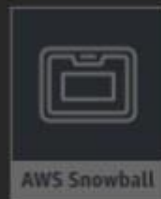
AWS Transfer for SFTP

Online transfer of active data



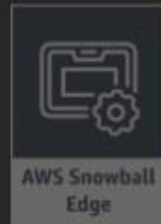
AWS DataSync

Ship static data into and out of Amazon S3



AWS Snowball

Storage and compute in disconnected environments



AWS Snowball Edge

Access AWS storage from on-premises



AWS Storage Gateway

What is AWS DataSync?

Online transfer service that simplifies, automates, and accelerates moving data between on-premises storage and AWS



Fast data transfer



Easy to use



Secure and reliable



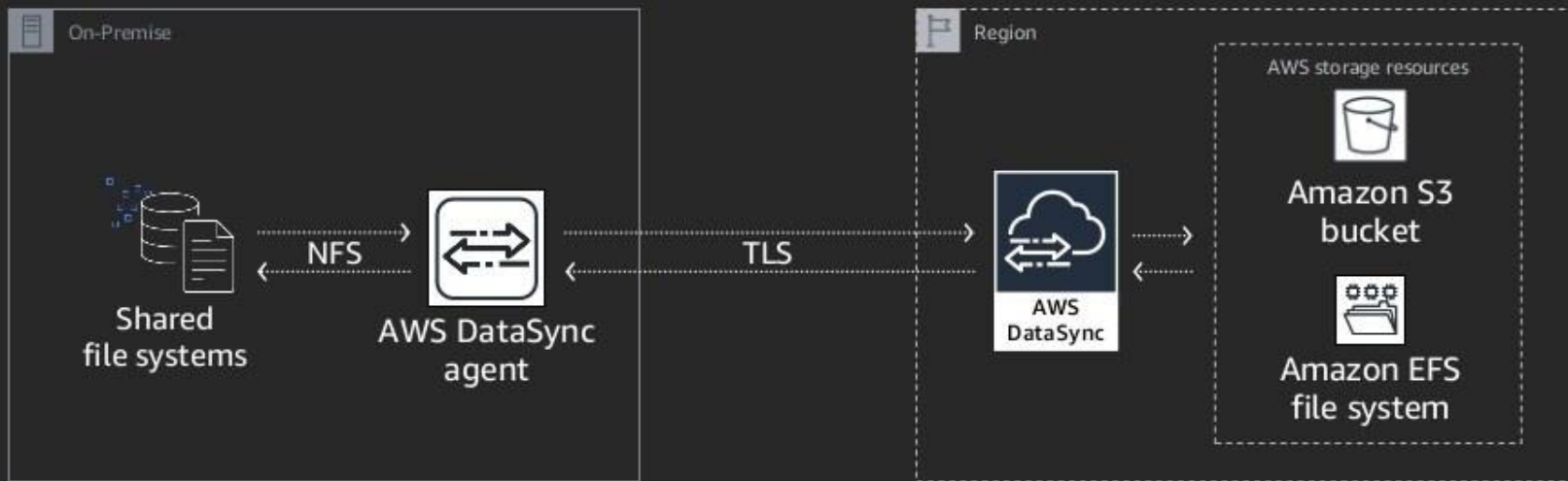
Cloud integrated



Cost-effective

Combines the speed and reliability of *network acceleration* software with the cost-effectiveness of *open source tools*

How AWS DataSync works



Deploy on-premises agent for fast access to local storage



Data transfer over the WAN using purpose-built protocol



Service in AWS writes or reads data from AWS storage services



Managed from AWS Console or Command Line Interface (CLI)

Transfers at speeds up to 10 Gbps per agent



**Fast data
transfer**

Purpose-built transfer protocol

Multithreaded design scales across multiple agents

Data reduction through incremental transfers, inline compression, and sparse file detection

Optimized read and write to Amazon S3 and Amazon EFS

Configurable throughput limits

Simplifies management of data movement



Easy to use

Setup and manage in the AWS Console,
CLI, or SDK

No infrastructure in AWS to deploy or manage

Agent updates and patches are fully-managed

Metadata preserved between storage systems
and services

Transfer your data with confidence



**Secure and
reliable**

Encryption in-transit with TLS1.2

Supports AWS KMS encryption at-rest for
AWS services

Accesses AWS services within your VPCs

Data validation in-transit and at-rest

Automatic recovery from I/O errors
or transmission failures

AWS management, identity, and compliance



Cloud
integrated

Monitor with Amazon CloudWatch

Detailed CloudWatch Logs track data movement

Control and audit usage with AWS Identity and Access Management (IAM) and AWS CloudTrail

PCI-DSS compliant and HIPAA eligible



Spend less money



**Cost-
effective**

Pay \$0.04 per-GB of data copied, with no minimum commitments or upfront fees

E.g., transferring 100 GB into EFS costs \$4

AWS infrastructure is fully managed and scaled based on usage

On-premises agents automatically updated and patched

How do I start using AWS DataSync?

As simple as 1-2-3 ...

Deploy on-premises agent and
connect to your AWS account



VMware ESXi
8 vCPU
32 GB memory

Create a task by configuring
storage locations and options



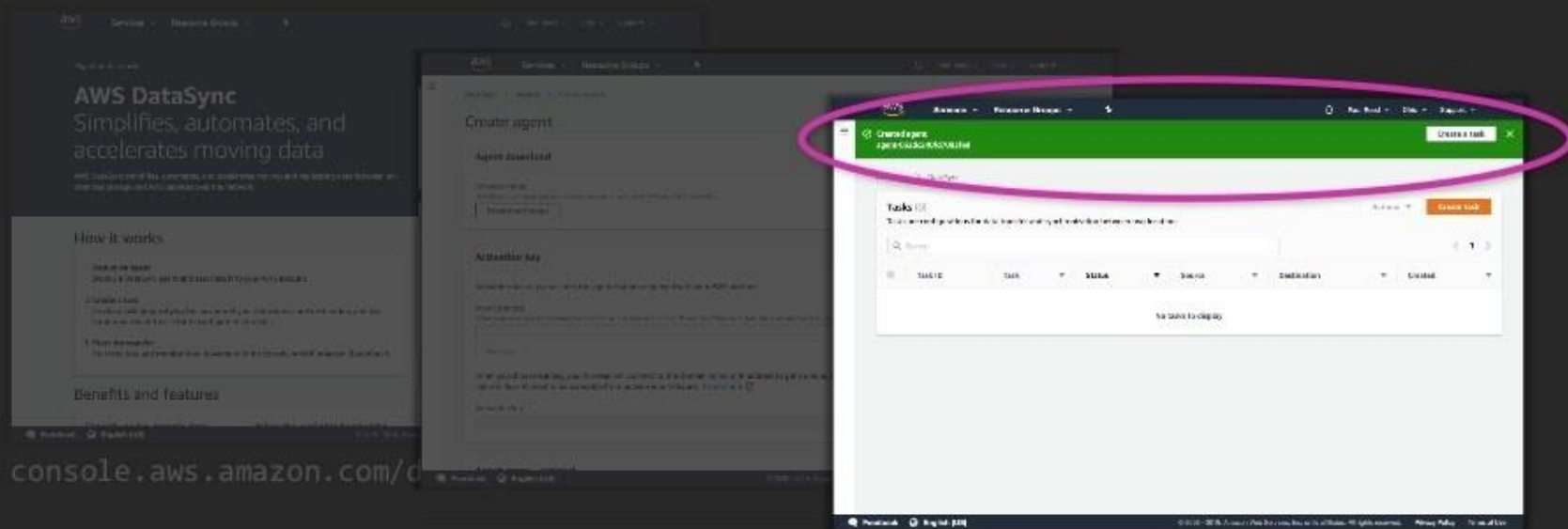
On-premises NFS file
systems, S3 buckets,
or EFS file systems

Start the task and
monitor progress

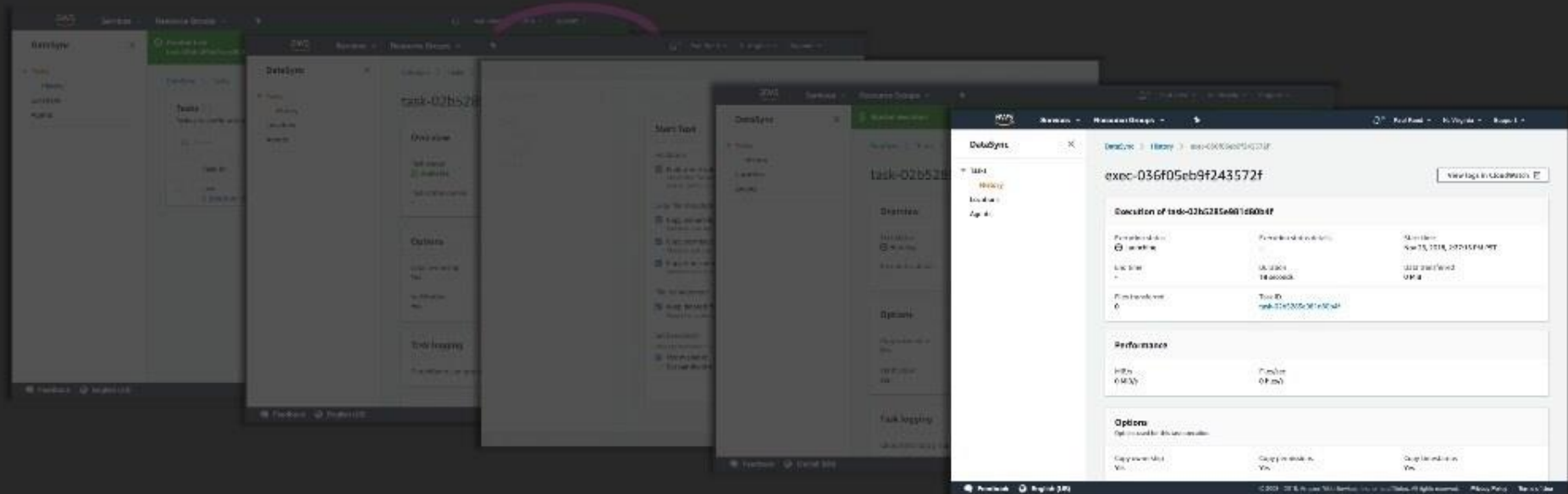


Monitor in the
console or through
Amazon CloudWatch

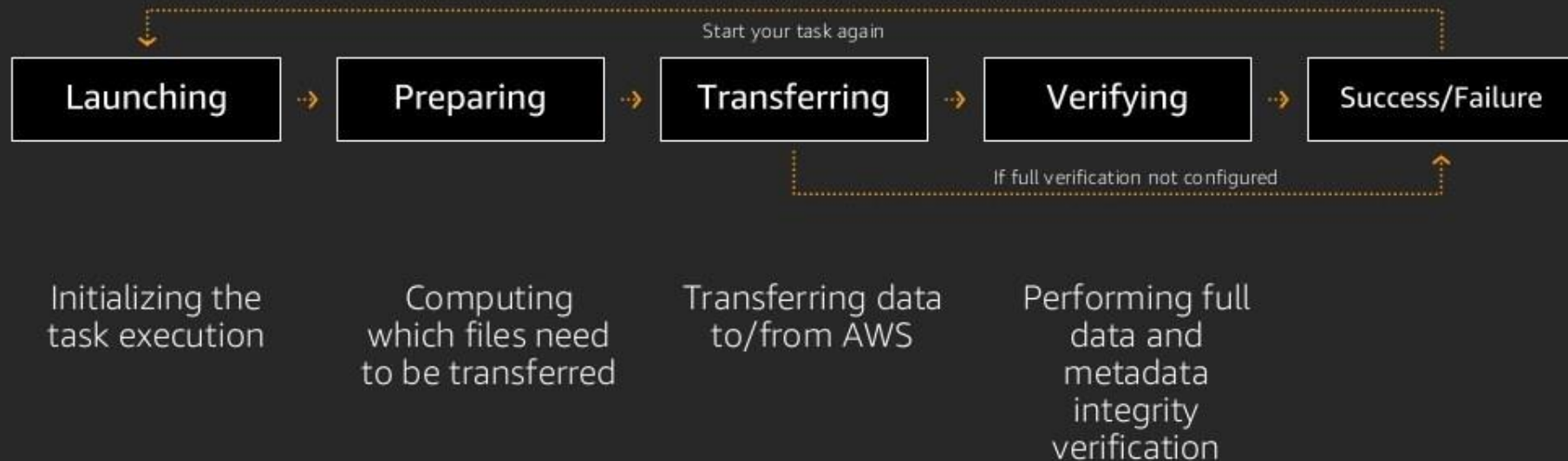
Deploy the agent and connect to your AWS account



Start and monitor the task



Task execution status



Task options

Configured with the task, but can be overridden each execution



File-level **validation** of source and destination after execution



Which **file metadata** is copied and used to determine changes



Mirror **files deleted** in source to destination



Set **bandwidth limits** when sharing network

Options

Validation

☒ **Enable verification**

Check files for consistency between source and destination data at the end of the transfer

Copy file metadata

☒ **Copy ownership**

Maintain user and group ID

☒ **Copy permissions**

Maintain existing permissions

☒ **Copy timestamps**

Maintain access time and modification time

File management

☒ **Keep deleted files**

Keep files in destination even when deleted from source

Set bandwidth

Allocate maximum bandwidth to be utilized by this task

☒ **Use available**

☐ **Set bandwidth (MB/s)**

Common use patterns

Migrate active application data



Existing on-premise data

Large data in a constant state of flux, with no natural break or stopping point for a one-time transfer.



Mirrored to S3 objects or an EFS file system

Phased migration enables testing and application testing and validation before cutover

Migrate **active data** to AWS confidently with efficient **incremental transfers** and end-to-end **data validation**

COX
AUTOMOTIVE™

Transfer data for timely in-cloud processing



Data generated on-premises

e.g., DNA sequencers, video production, GIS or seismographic data

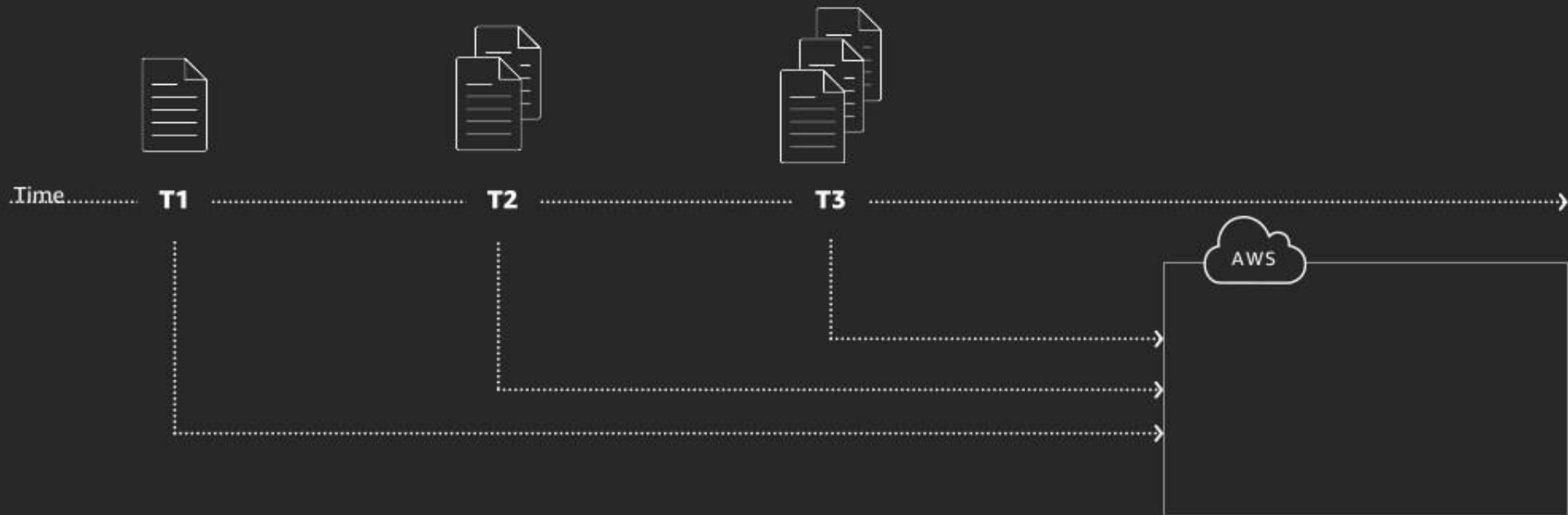
Data processed in-cloud

Amazon CloudWatch Events trigger analysis on data stored natively in AWS storage services

Automated and **accelerated** data transfer, **securely** and **reliably** copies data to Amazon S3 or Amazon EFS for **timely** analysis

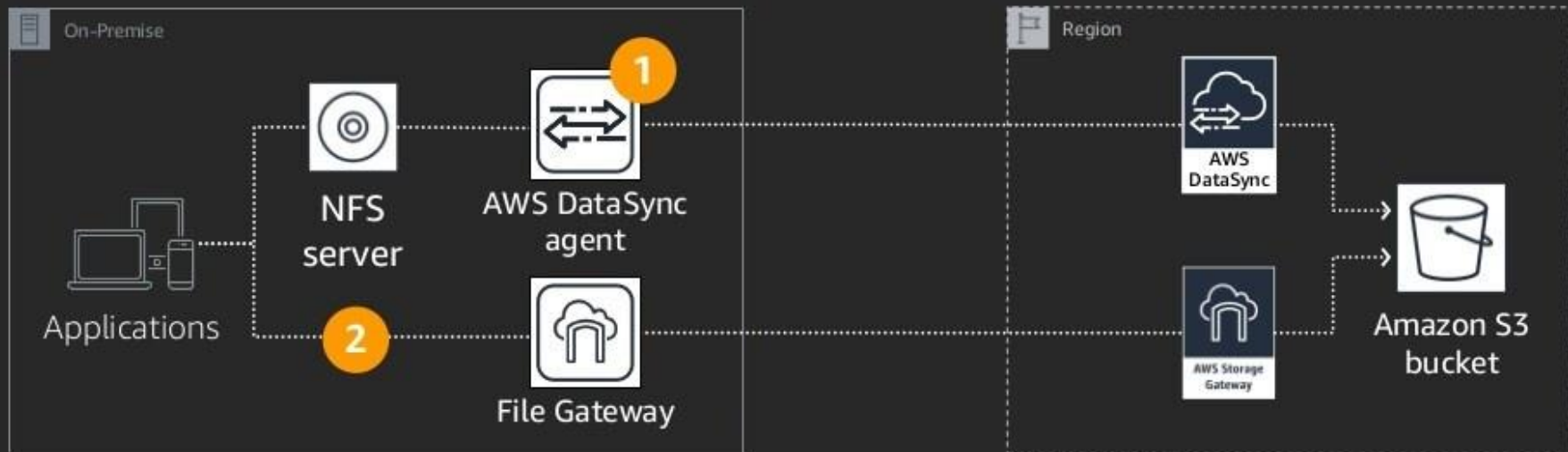


Replicate for data protection and recovery



Periodic **incremental copies** reduce bandwidth and **minimize loss** in the event of an on-premises failure

Reduce infrastructure with AWS DataSync and AWS Storage Gateway



- ➊ DataSync migrates to Amazon S3. Incremental copies keeps Amazon S3 up to date
- ➋ Latency-sensitive applications can access data in S3 through File Gateway
Can reduce on-premises storage

AWS DataSync and File Gateway working together



AWS DataSync

Copy on-premises files to
Amazon S3 or EFS

- Client to existing NFS server
- Accelerated data transfer
- Copies into Amazon S3 or Amazon EFS

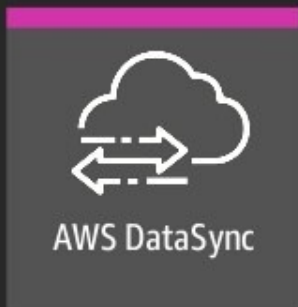


File Gateway

On-premises access to data
in Amazon S3

- NFS or SMB server
- Local caching
- Backed by Amazon S3

AWS DataSync demo—architecture



Demo architecture and requirements

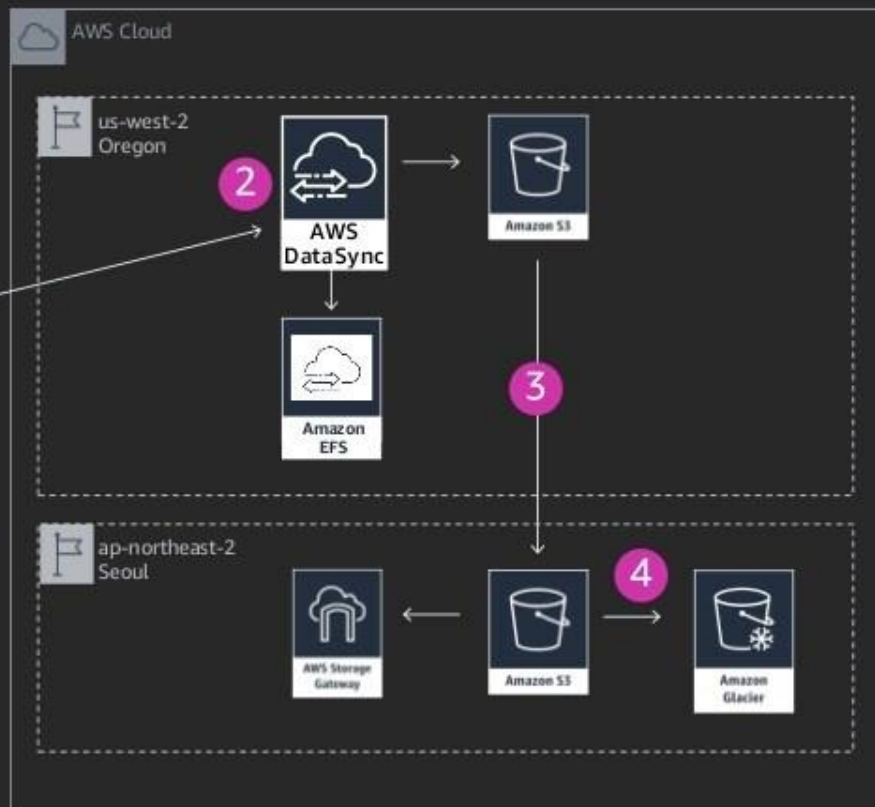
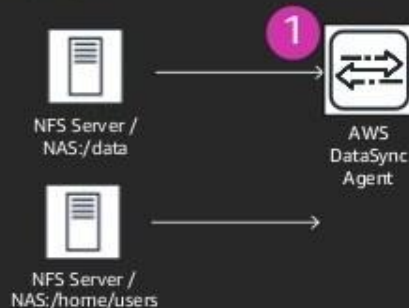
- NAS Server
- Directory1 to S3 (us-west-2)
- Directory2 to EFS (us-west-2)
- Replicate S3 data to another region for archive and partner access with Storage Gateway
- 2 x 10Gb DirectConnect
- AWS DataSync Agent

Demo architecture



Los Angeles

On-premises architecture





1. AWS DataSync moves files from NAS to Amazon S3 us-west-1
2. AWS DataSync writes data to Amazon S3 and Amazon Elastic File System (Amazon EFS)
3. S3 Cross Region Replication (CRR) moves data to another region within another account
4. S3 Lifecycle Policy writes data to Amazon Glacier
5. File access provided by AWS Storage Gateway (file gateway)

AWS DataSync

Simplifies, automates, and accelerates your online data transfer

 **Migrate** active application data

 **Transfer** data for timely processing

 **Replicate** for data protection and recovery



Transfers up to 10 Gbps per agent



Simple data movement to S3 or EFS



Secure and reliable transfers



AWS integrated



Pay as you go

Combines the speed and reliability of **network acceleration** software with the cost-effectiveness of **open source tools**

Available in 10 AWS regions



- US East (N. Virginia)
- US East (Ohio)
- US West (N. California)
- US West (Oregon)
- EU (Frankfurt)
- EU (Ireland)
- Asia Pacific (Seoul)
- Asia Pacific (Singapore)
- Asia Pacific (Sydney)
- Asia Pacific (Tokyo)



Get started at
<http://aws.amazon.com/datasync>