

How to rapidly migrate your data online with AWS DataSync

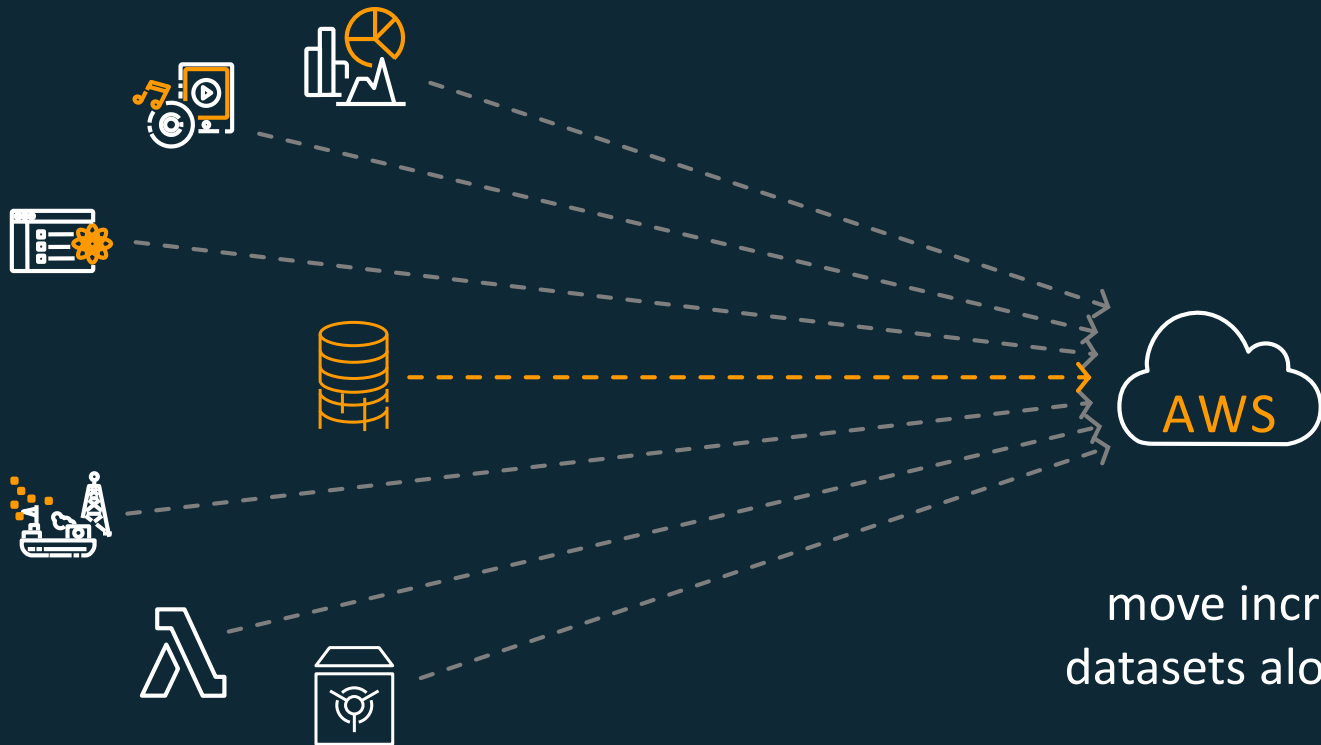
Jeff Bartley, AWS Solutions Architect

October 22nd, 2019

Agenda

- Why AWS DataSync?
- What is AWS DataSync and how does it work?
- Demo
- AWS DataSync architectural guidance
- Learn More

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?



Migration of active application data



Transfers for timely in-cloud processing

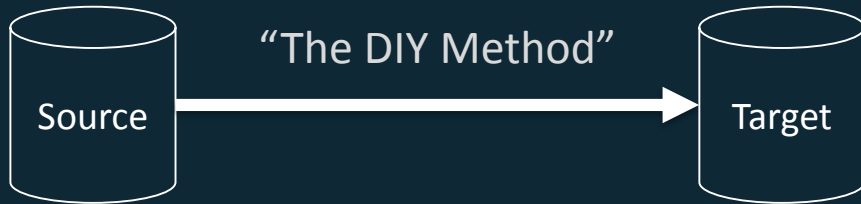


Archive to free up on-premises storage capacity



Replication for data protection and recovery

How do organizations solve this today?



Consider:

- Transfer protocol?
- Ample compute for data transfer?
- Compression algorithm?
- Preserve metadata?
- Individual or multiple hosts?
- Security?
- Data transfer validation?

“Prior to DataSync, we once **spent \$30K** to build tools to **validate data** in S3 was identical to the source dataset.”

~Medical devices manufacturer

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



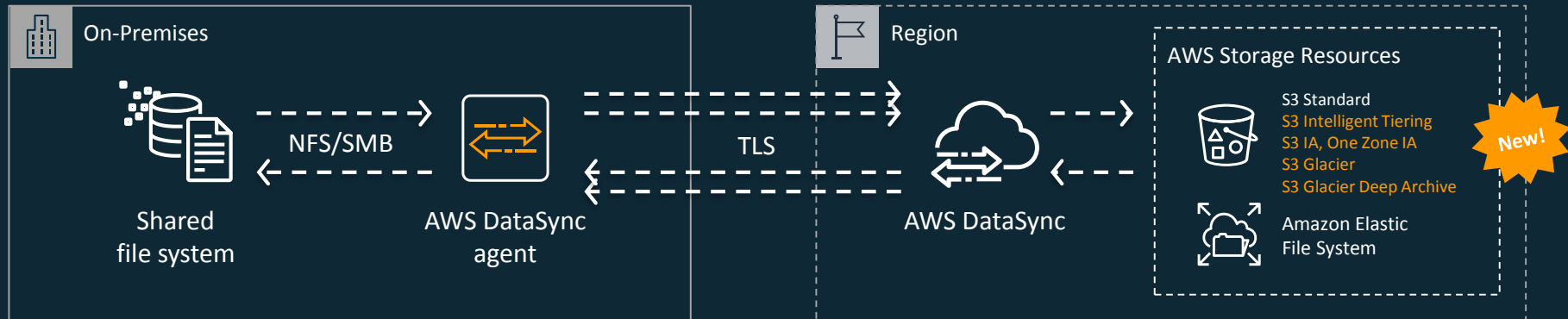
Cloud integrated



Cost-effective

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

How AWS DataSync works



Agent deployed on VMware or EC2 for efficient access to local NFS/SMB server



Highly parallel transfers using optimized network protocol



Fully-managed service scales to send or receive data from agent

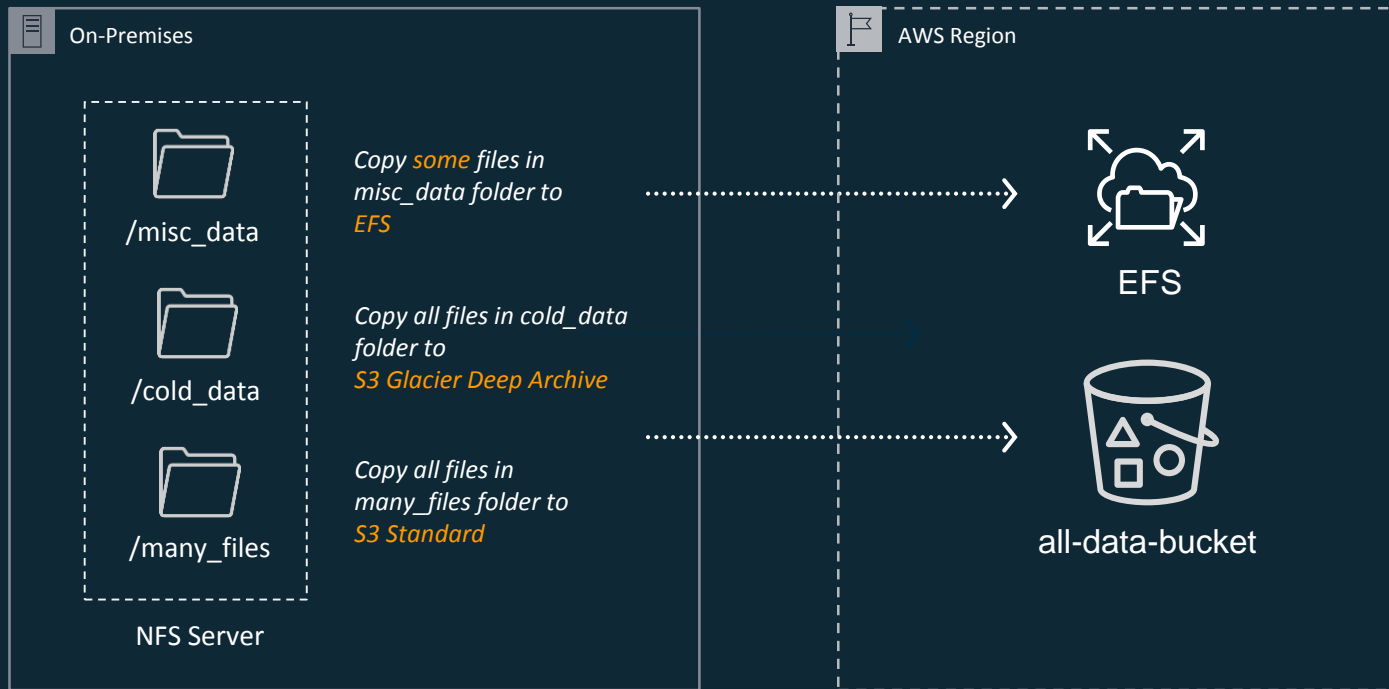


Optimized to read and writes to S3 and EFS using IAM/VPCE

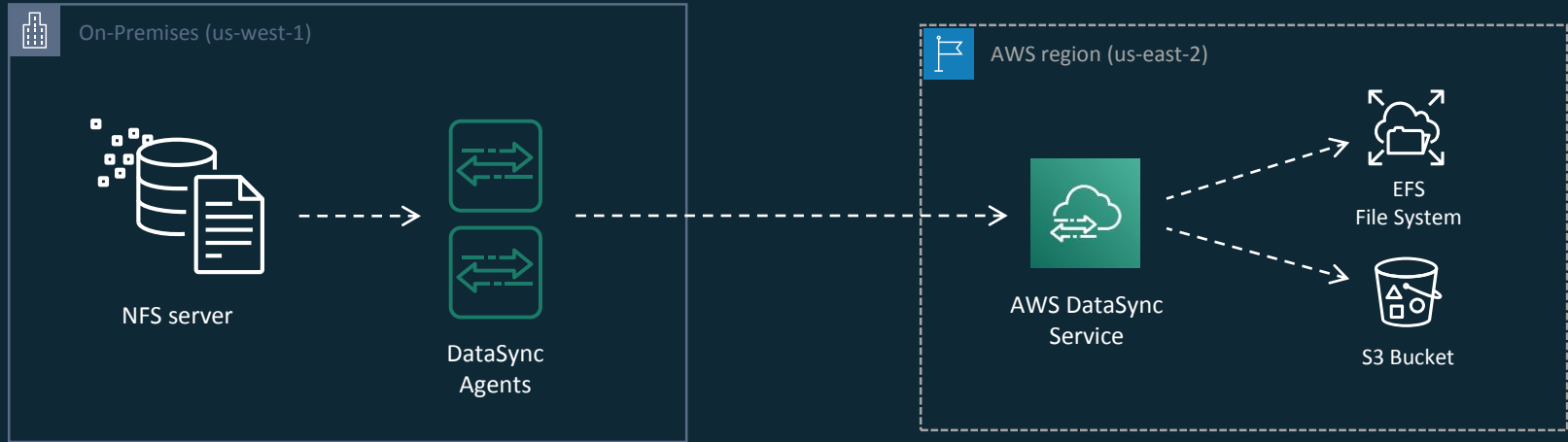
Demo



Demo scenario



Demo architecture



Start transferring files quickly

1

**Deploy on-premises agent
and connect to your
AWS account**



Agents are activated in the region where data will be transferred

2

**Create a task by
configuring storage
locations and options**



Locations can be on-premises NFS or SMB file systems, S3 buckets, or EFS file systems

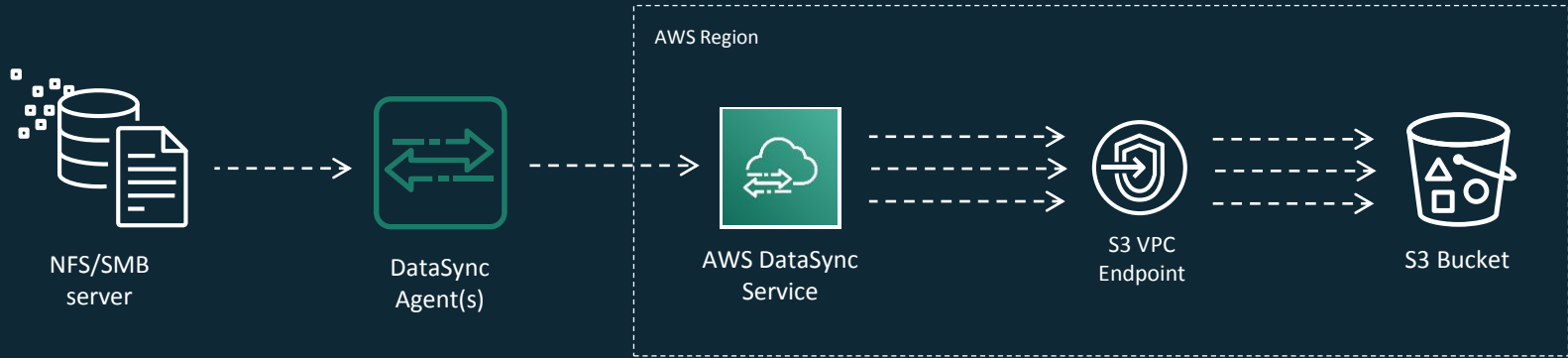
3

**Start the task and
monitor progress**



Monitor in the console or through Amazon CloudWatch

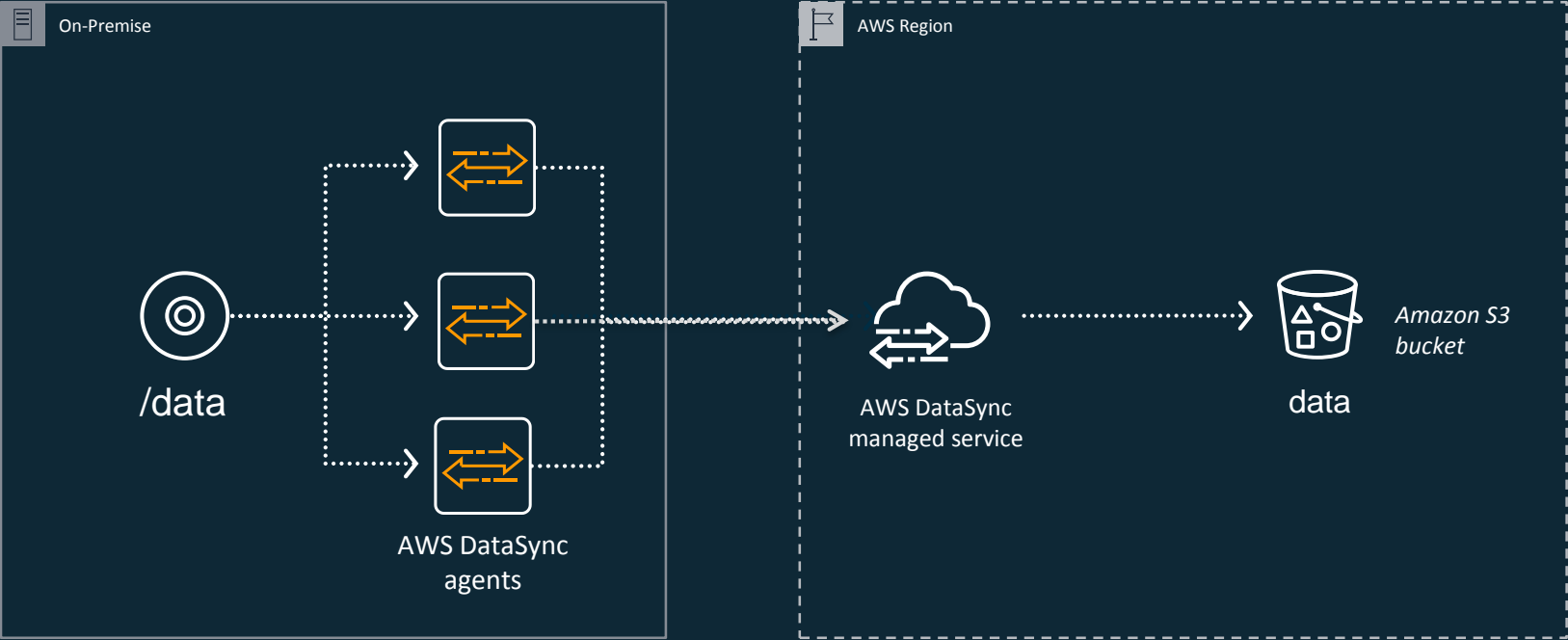
DataSync is optimized for S3



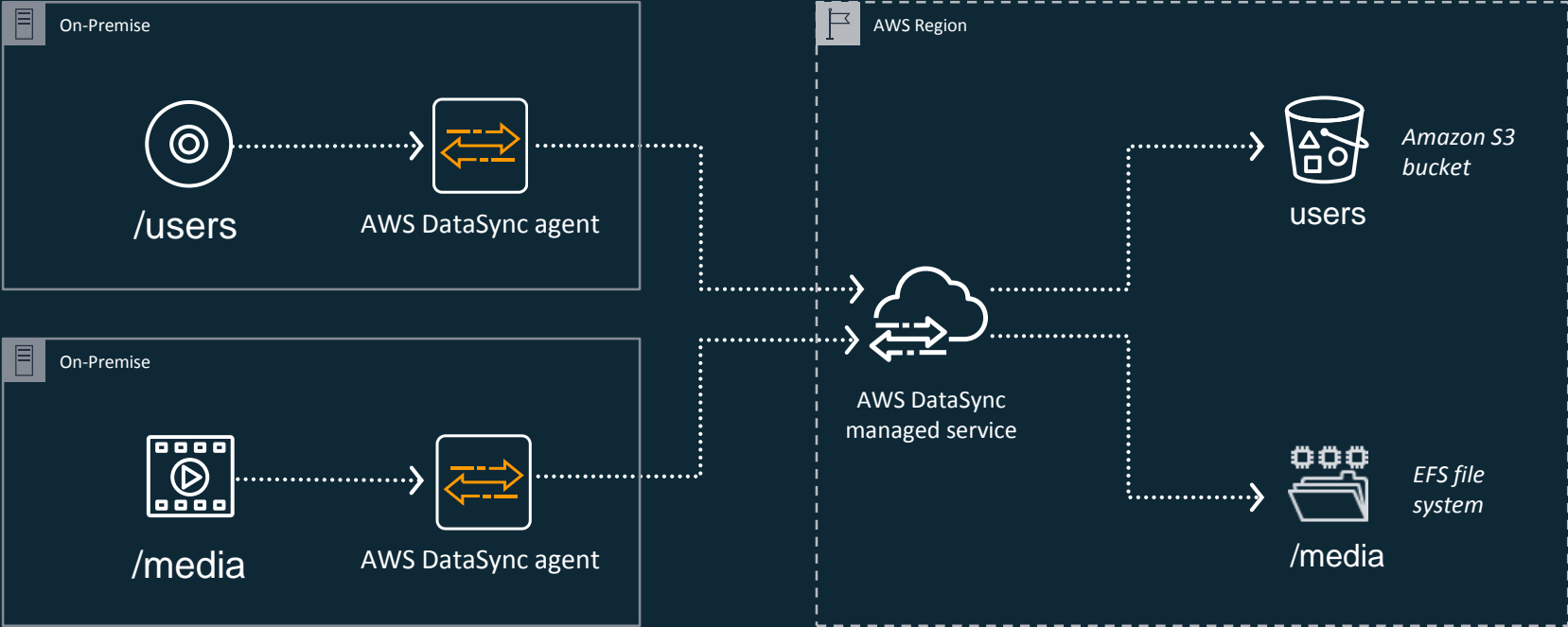
AWS DataSync takes care of:

- Multi-part transfers
- Exponential back off
- MD5 PUTs
- MIME detection
- Horizontal scaling
- Request parallelization
- Direct to storage classes
- Folder creation
- S3 VPC endpoints
- Retry logic
- Cost optimization
- IAM roles

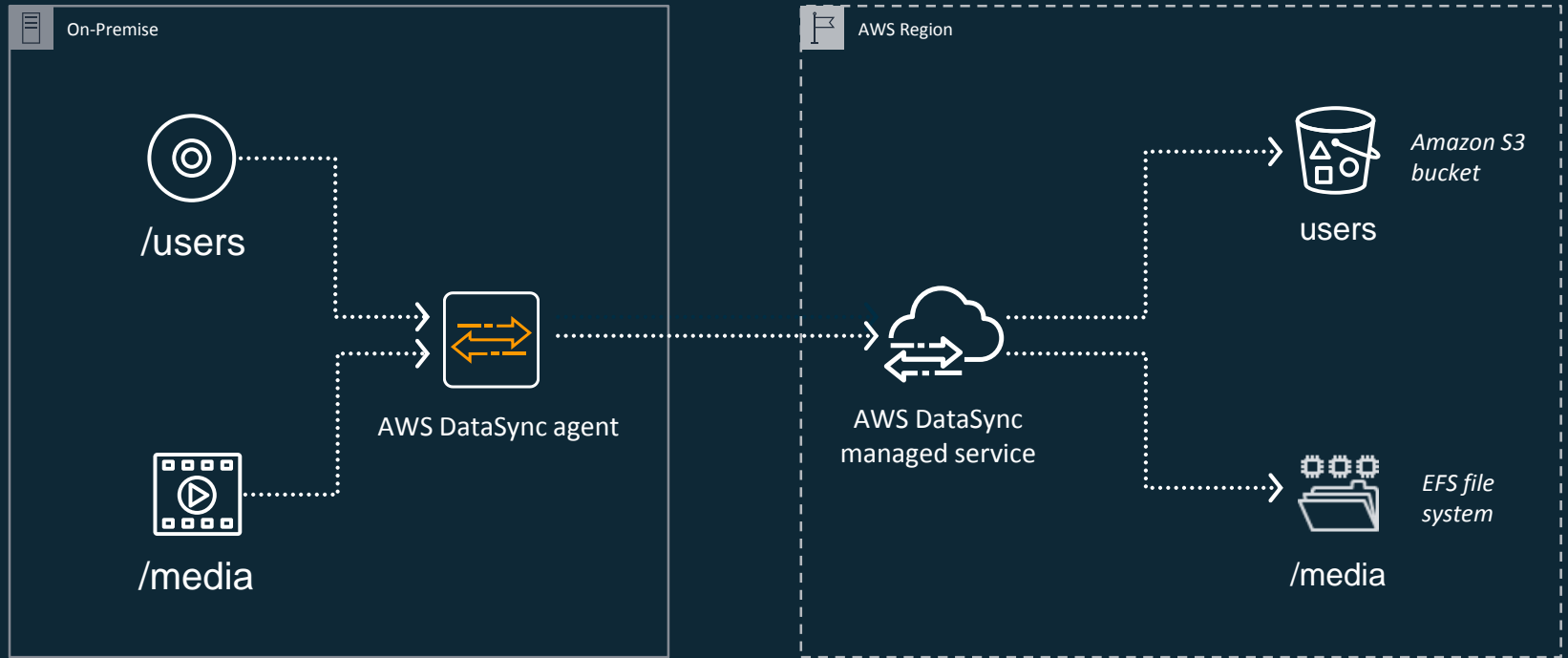
Multiple agents can transfer from a single location



Multiple agents can transfer from different locations



Single agent can transfer from multiple locations



DataSync architectural guidance

- One agent can fill a 10Gbps network link
- Supports Private Link – don't have to use the Internet to move data
- Transfers to / from AWS
- DataSync can copy directly to low-cost S3 storage classes
- Transfer data via multiple DataSync agents to the same target
- Use DataSync for EFS-EFS migration
- Tasks can be queued to run on the same agent
- Multiple sources transfer via a single agent

Learn more...



aws.amazon.com/datasync

A screenshot of the AWS DataSync product page. The page features the AWS logo at the top left, navigation links for Products, Solutions, Pricing, Documentation, Learn, Partner Network, AWS Marketplace, and Explore More. A prominent orange button says "Create an AWS Account". The main heading is "AWS DataSync" with the subtext "Easily transfer data to and from AWS up to 10x faster". A yellow button below the heading says "Get started with AWS DataSync". The page includes a detailed paragraph about the service's capabilities and a video thumbnail titled "AWS DataSync - Automate & accelerate online data transfer (1:31)".

aws

Contact Sales Support English My Account Create an AWS Account

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

AWS DataSync Overview Features Pricing Getting Started FAQs Blogs

AWS DataSync

Easily transfer data to and from AWS up to 10x faster

Get started with AWS DataSync

AWS DataSync is a data transfer service that makes it easy for you to automate moving data between on-premises storage and Amazon S3 or Amazon Elastic File System (Amazon EFS). DataSync automatically handles many of the tasks related to data transfers that can slow down migrations or burden your IT operations, including running your own instances, handling encryption, managing scripts, network optimization, and data integrity validation. You can use DataSync to transfer data at speeds up to 10 times faster than open-source tools. DataSync uses an on-premises software agent to connect to your existing storage or file systems using the Network File System (NFS) protocol, so you don't have to write scripts or modify your applications to work with AWS APIs. You can use DataSync to copy data over AWS Direct Connect or internet links to AWS. The service enables one-time data migrations, recurring data processing workflows, and automated replication for data protection and recovery. Getting started with DataSync is easy: Deploy the DataSync agent on premises, connect it to a file system or storage array, select Amazon EFS or S3 as your AWS storage, and start moving data. You pay only for the data you copy.

AWS DataSync

AWS DataSync - Automate & accelerate online data transfer (1:31)

Blogs: aws.amazon.com/datasync/blogs/

GitHub: [aws-samples/aws-datasync-migration-workshop](https://github.com/aws-samples/aws-datasync-migration-workshop)

Thank you!