aws INNOVATE

EVERY APPLICATION EDITION

AWS INNOVATE
EVERY APPLICATION EDITION

Moving data to AWS;
Find the right tool and process

Lily Jang

Storage Business Development Manager

AWS

Ameen Khan S

Storage Specialist Solutions Architect
AWS

Agenda

- Why migrate to AWS? Why now?

- Key questions for data movement

- AWS DataSync overview and demo

- AWS Snow Family overview

- Using AWS DataSync and AWS Snow Family together

- Key takeaways

Why migrate to AWS? Why now?

Chief Information Officers, IT Professionals

- Agility and staff productivity
- New customer experiences
- Application modernization

Why migrate to AWS? Why now?

Chief Information Officers, IT Professionals

- Agility and staff productivity
- New customer experiences
- Application modernization

Chief Security Officers, Security Professionals

- Improved security
- Operational resilience
- Scale globally quickly

Why migrate to AWS? Why now?

Chief Information Officers, IT Professionals

- Agility and staff productivity
- New customer experiences
- Application modernization

**Chief Security Officers, Security Professionals**

- Improved security
- Operational resilience
- Scale globally quickly

Infrastructure and Operations Professionals

- Data center consolidation
- Hardware/Software end of life
- Operational efficiency
- Cost reduction

aws

Why migrate to AWS? Why now?

Chief Information Officers, IT Professionals

- Agility and staff productivity
- New customer experiences
- Application modernization

Chief Security Officers, Security Professionals

- Improved security
- Operational resilience
- Scale globally quickly

Data Scientists, Engineers

- New technologies (e.g. IoT, AI/ML)
- Data transformation

Infrastructure and Operations Professionals

- Data center consolidation
- Hardware/Software end of life
- Operational efficiency
- Cost reduction
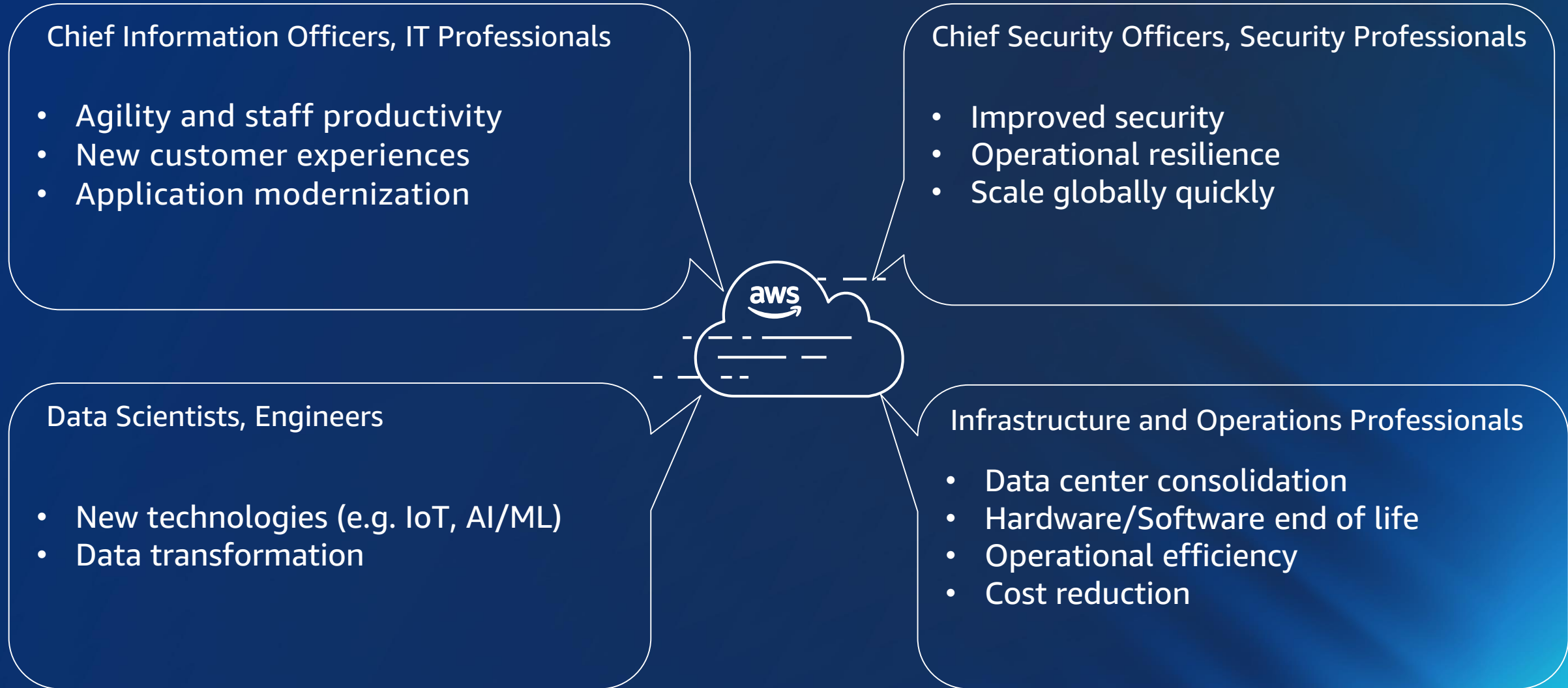
Key questions for data movement

- What datasets am I moving?

    - Use case – production data vs backup and archive data

    - Data type – virtual machines (VMs), databases, files

- How much available bandwidth do I have?

What datasets am I moving? – Use case

Production data

- Essential to complete daily tasks
- Readily available

Backup and archive data

- Stored long-term for compliance or archival
- Can be used for disaster recovery

What datasets am I moving? – Use case

Production data

- Essential to complete daily tasks
- Readily available

Move data fast and seamlessly

Backup and archive data

- Stored long-term for compliance or archival
- Can be used for disaster recovery

Optimize cost and increase security

What datasets am I moving? – Data type

	Structured data, virtual machines	Databases	Unstructured data, files
Use cases			
Online			
Hybrid, offline			
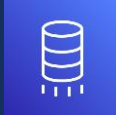
What datasets am I moving? – Data type

Use cases	Structured data, virtual machines	Databases	Unstructured data, files
	• Migrate structured data to AWS from on premises		
Online	AWS Application Migration Service (AWS MGN)		
Hybrid, offline			
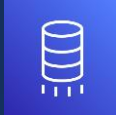
What datasets am I moving? – Data type

	Structured data, virtual machines	Databases	Unstructured data, files
Use cases	• Migrate structured data to AWS from on premises	• Migrate or replicate from commercial databases • Continuous replication keeping source and target databases in sync	
Online	**AWS Application Migration Service (AWS MGN)**	**AWS Database Migration Service (AWS DMS)**	
Hybrid, offline			

What datasets am I moving? – Data type

	Structured data, virtual machines	Databases	Unstructured data, files
Use cases	• Migrate structured data to AWS from on premises	• Migrate or replicate from commercial databases • Continuous replication keeping source and target databases in sync	• Frequent transfers of small files from external or 3rd parties (ex: bank statements) • Remove management of SFTP/FTPS/ FTP/AS2 servers
Online	AWS Application Migration Service (AWS MGN)	AWS Database Migration Service (AWS DMS)	AWS Transfer Family AWS DataSync
Hybrid, offline			

What datasets am I moving? – Data type

	Structured data, virtual machines	Databases	Unstructured data, files
Use cases	• Migrate structured data to AWS from on premises	• Migrate or replicate from commercial databases • Continuous replication keeping source and target databases in sync	• Frequent transfers of small files from external or 3rd parties (ex: bank statements) • Remove management of SFTP/FTPS/ FTP/AS2 servers
Online	AWS Application Migration Service (AWS MGN)	AWS Database Migration Service (AWS DMS)	AWS Transfer Family / AWS DataSync
Hybrid, offline		AWS Storage Gateway / AWS Snow Family	

How much available bandwidth do I have?

Online

- Available or reliable bandwidth
- Can deploy AWS Direct Connect quickly
- Data is online and ready to move
- Firewall configurations may be needed

Offline

- Limited or low bandwidth
- Heavily utilized WAN, air-gapped
- Data can be offline during the move
- Data center capacity for devices
- Device management

Estimate time to transfer data

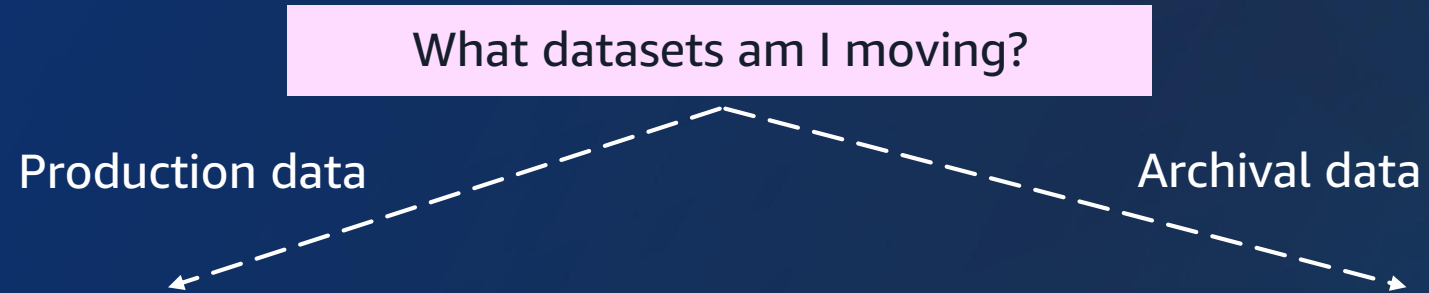|	1 Gbps	2 Gbps	5 Gbps	10 Gbps
500 TB	58 days	29 days	12 days	6 days
5 PB	2 years	1 year	116 days	58 days
10 PB	4 years	2 years	232 days	116 days

Use AWS Snow Family

- Offline migration
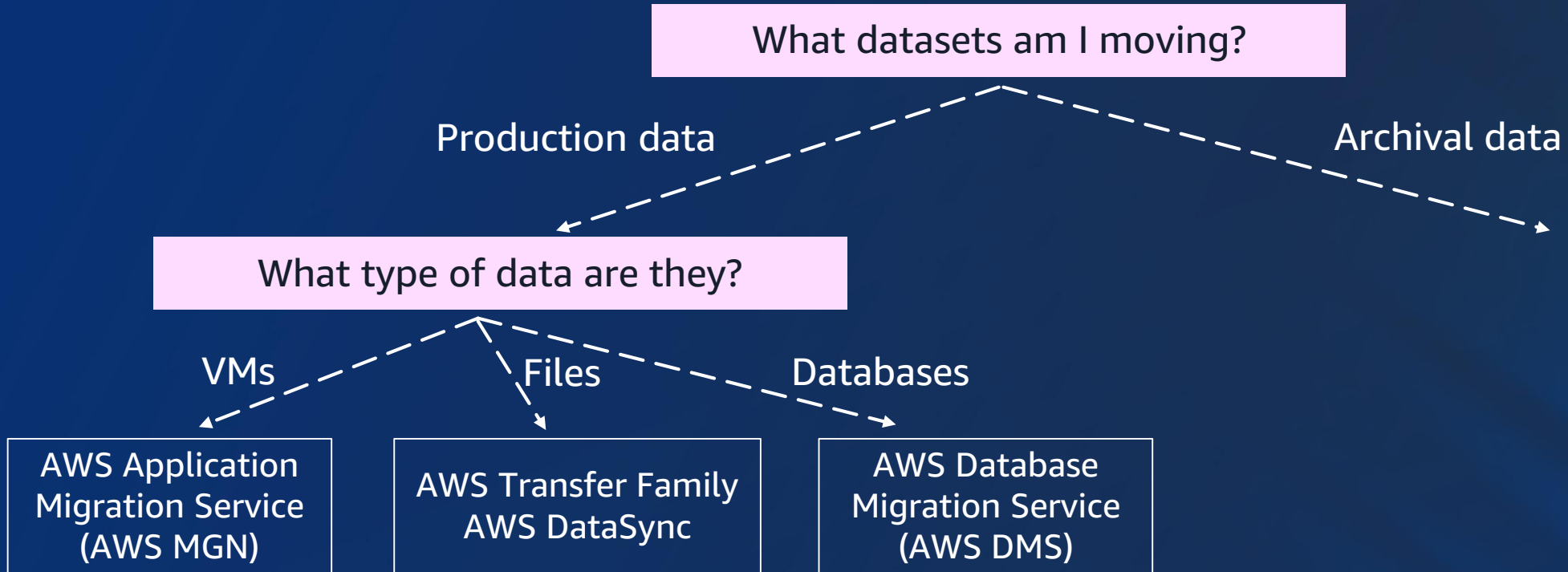- Bandwidth limitations
- Flexible timeline

Use AWS DataSync

- Online migration
- Available bandwidth
- Strict timelines

Data migration decision tree

What datasets am I moving?

Production data   Archival data

Data migration decision tree

What datasets am I moving?

Production data · Archival data

What type of data are they?

VMs · · · · · · · · · · Files · · · · · · · · · · Databases

AWS Application
Migration Service
(AWS MGN)

AWS Transfer Family
AWS DataSync

AWS Database
Migration Service
(AWS DMS)

aws

Data migration decision tree

What datasets am I moving?

Production data → What type of data are they?

Archival data → Do I want online or offline transfer?

Online → What type of data are they?

What type of data are they?
- VMs → AWS Application Migration Service (AWS MGN)
- Files → AWS Transfer Family / AWS DataSync
- Databases → AWS Database Migration Service (AWS DMS)

Do I want online or offline transfer?
- Offline → AWS Snow Family
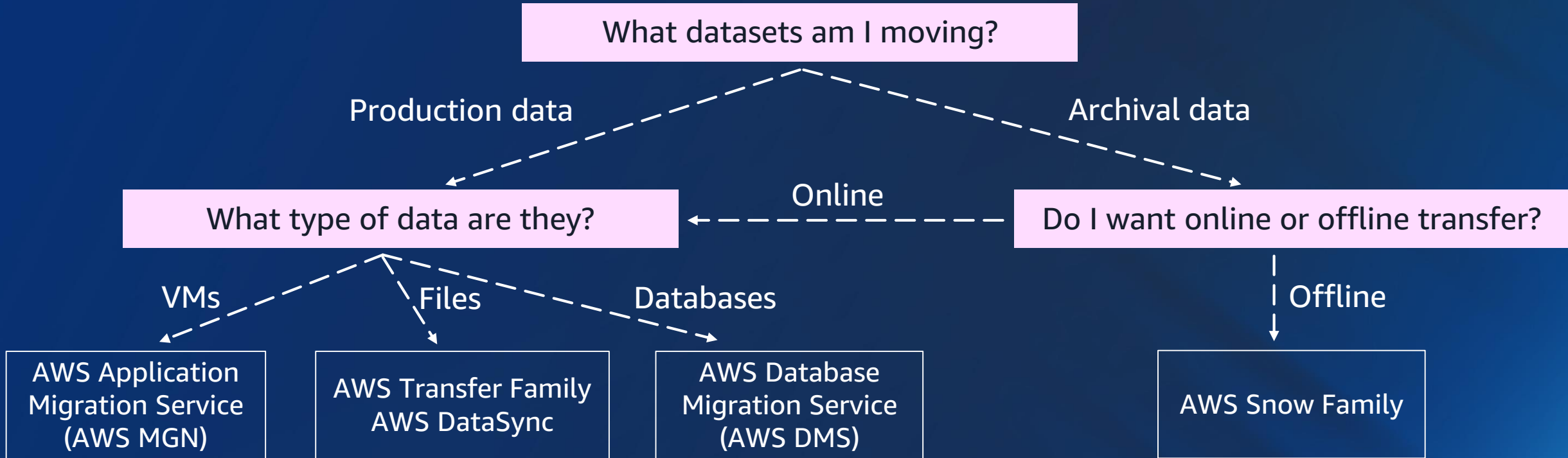
Data migration decision tree
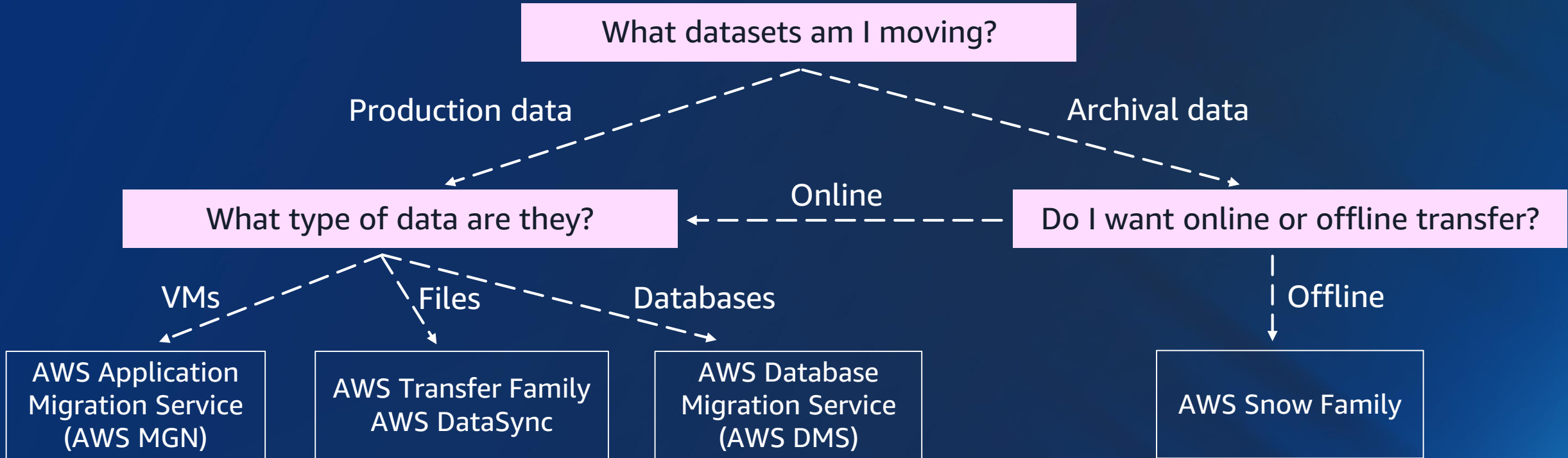
AWS DataSync overview

What is AWS DataSync?

Online data migration service
that simplifies, automates, and accelerates
copying file and object data to and from AWS storage

What is AWS DataSync?

Online data migration service
that simplifies, automates, and accelerates
copying file and object data to and from AWS storage

Fast data transfer

- Highly optimized, parallel network transfer (up to 100 TB/day)
- Transfers only incremental changes

What is AWS DataSync?

Online data migration service
that simplifies, automates, and accelerates
copying file and object data to and from AWS storage

Fast data transfer

- Highly optimized, parallel network transfer (up to 100 TB/day)
- Transfers only incremental changes

Easy to use

- Schedule transfers
- Throttle bandwidth
- Filter by file name patterns

aws

What is AWS DataSync?

Online data migration service
that simplifies, automates, and accelerates
copying file and object data to and from AWS storage

Fast data transfer

- Highly optimized, parallel network transfer (up to 100 TB/day)
- Transfers only incremental changes

Easy to use

- Schedule transfers
- Throttle bandwidth
- Filter by file name patterns

Secure and reliable

- End-to-end encryption
- End-to-end data verification
- Virtual private cloud (VPC) endpoints with AWS PrivateLink

Fully managed

- Integrates with AWS management and monitoring tools
- Direct transfer into all Amazon S3 storage classes

What is AWS DataSync?

Online data migration service
that simplifies, automates, and accelerates
copying file and object data to and from AWS storage

Fast data transfer

- Highly optimized, parallel network transfer (up to 100 TB/day)
- Transfers only incremental changes

Easy to use

- Schedule transfers
- Throttle bandwidth
- Filter by file name patterns

Secure and reliable

- End-to-end encryption
- End-to-end data verification
- Virtual private cloud (VPC) endpoints with AWS PrivateLink
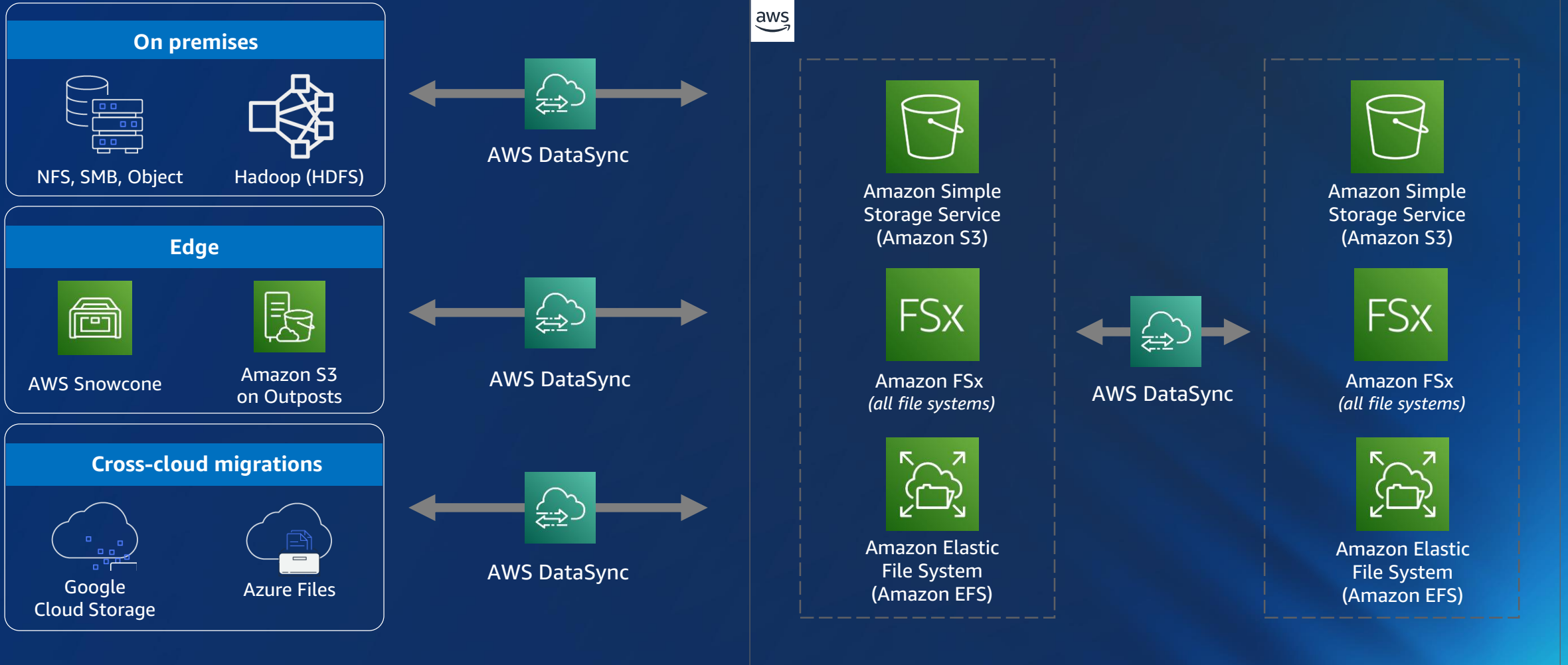
Fully managed

- Integrates with AWS management and monitoring tools
- Direct transfer into all Amazon S3 storage classes

Cost effective

- Pay per GB transferred
- No minimum charge

What can you do with AWS DataSync?

On premises

NFS, SMB, Object

Hadoop (HDFS)

AWS DataSync

Edge

AWS Snowcone

Amazon S3 on Outposts

AWS DataSync

Cross-cloud migrations

Google Cloud Storage

Azure Files

AWS DataSync

aws

Amazon Simple Storage Service (Amazon S3)

Amazon FSx *(all file systems)*

Amazon Elastic File System (Amazon EFS)

AWS DataSync

Amazon Simple Storage Service (Amazon S3)

Amazon FSx *(all file systems)*

Amazon Elastic File System (Amazon EFS)

AWS DataSync Discovery

Gain insights into storage utilization

- ✓ Automated data collection
- ✓ Dashboards for aggregated views of data
- ✓ Find underutilized resources

Receive recommendations for AWS storage services

- ✓ Select the right storage for your use case
- ✓ Optimize your AWS storage configuration
- ✓ Meet your performance needs while minimizing costs

Simplify migration planning

- ✓ Minimize time, effort, and costs
- ✓ Use estimated costs to inform your budget
- ✓ Validate assumptions before migrating

Learn more at
aws.amazon.com/datasync/discovery

TMAP Mobility transferred 2.4 PB of Hadoop data using AWS DataSync

Problem

TMAP Mobility (South Korea) needed to transfer 2.4 PB of on premises Cloudera HDFS to a new data lake in Amazon S3 within 2 months. The new data lake would support an existing mobile application with more than 5 million daily active users.

The new data lake project would leverage real-time analytics and batch processing with an AWS managed services–based analytics pipeline.

TMAP MOBILITY

Read more: AWS Storage Blog

TMAP Mobility transferred 2.4 PB of Hadoop data using AWS DataSync

Read more: AWS Storage Blog

Problem

TMAP Mobility (South Korea) needed to transfer 2.4 PB of on premises Cloudera HDFS to a new data lake in Amazon S3 within 2 months. The new data lake would support an existing mobile application with more than 5 million daily active users.

The new data lake project would leverage real-time analytics and batch processing with an AWS managed services–based analytics pipeline.

Solution

TMAP Mobility selected AWS DataSync as their petabyte-scale data movement solution. Integrated capabilities such as task monitoring, data encryption, data integrity validation, and native support for HDFS, further reduced project risk and simplified operations.

TMAP Mobility transferred 2.4 PB of Hadoop data using AWS DataSync

Problem

TMAP Mobility (South Korea) needed to transfer 2.4 PB of on premises Cloudera HDFS to a new data lake in Amazon S3 within 2 months. The new data lake would support an existing mobile application with more than 5 million daily active users.

The new data lake project would leverage real-time analytics and batch processing with an AWS managed services–based analytics pipeline.

Solution

TMAP Mobility selected AWS DataSync as their petabyte-scale data movement solution. Integrated capabilities such as task monitoring, data encryption, data integrity validation, and native support for HDFS, further reduced project risk and simplified operations.

Outcome

Using AWS DataSync, TMAP Mobility was able to meet their timeline for migrating their data into their new Amazon S3 data lake. As a fully managed service, AWS DataSync made it easy for operators, with no prior experience with the service, to configure and perform the migration within a short 2 month time window.

TMAP MOBILITY

Read more: AWS Storage Blog

Demo:
From HDFS to Amazon S3
using AWS DataSync

Demo
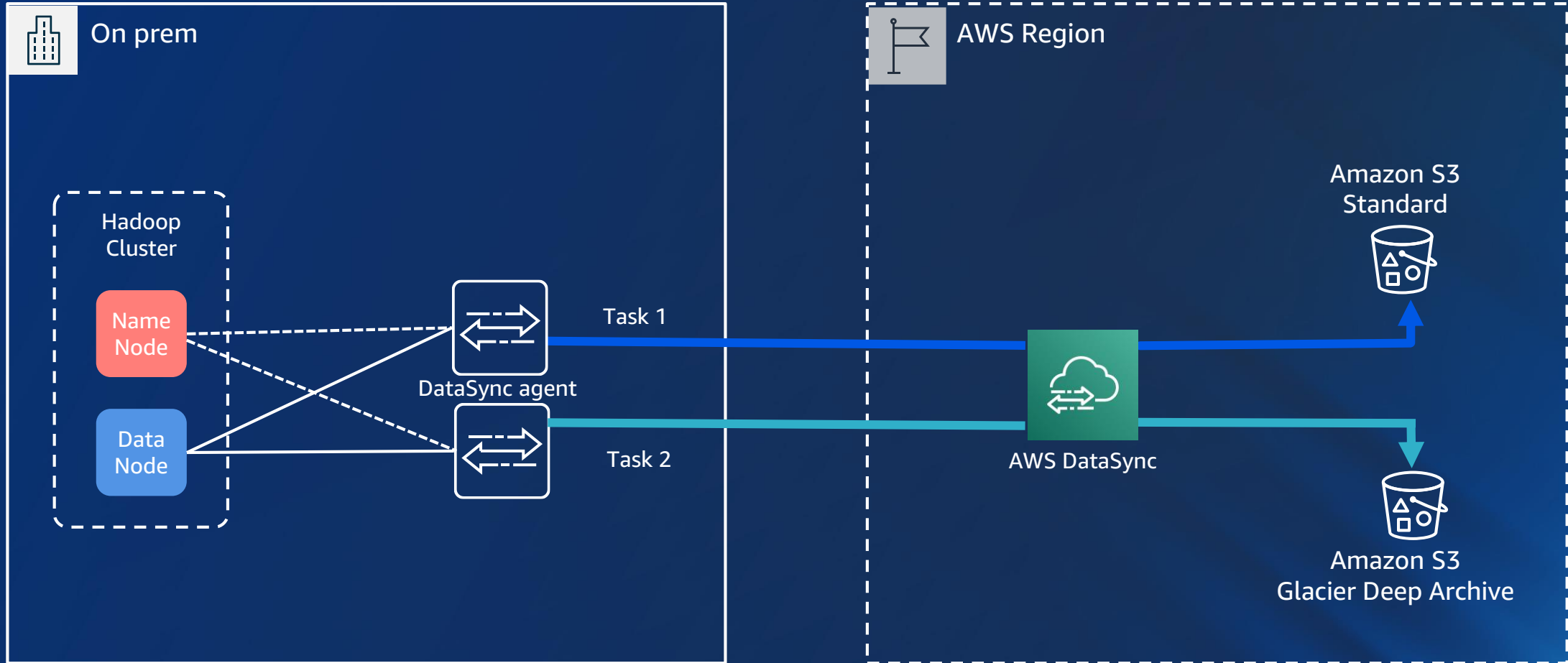
AWS Snow Family overview

What is AWS Snow Family?

Data movement

Edge computing

Customer sites

AWS Snowcone

AWS Snowball Edge

AWS Snowmobile

aws

AWS Snowball Edge

AWS Snowcone

Edge sites

- Set of services and functionality on the partially to fully disconnected and secure edge
- Enables data collection, data processing, and data movement from the edge to AWS

AWS Snow Family data migration workflow

AWS Snow Family
management console

Job
created &
provisioned

AWS Snow Family data migration workflow

AWS Snow Family management console

Manage AWS Snow Family devices via AWS OpsHub or AWS CLI

Device shipped

Data copied
1-N* days

1-3 days

1-2⁺ days

Job created & provisioned

Device arrives

Courier transport service

Copy data via >1GB/s local network

AWS Snow Family data migration workflow

AWS Snow Family management console

Manage AWS Snow Family devices via AWS OpsHub or AWS CLI

Courier transport service

Secure erasure using NIST 800-88 standards

Device shipped

Data copied
1-N* days

Device arrives at AWS

1-3 days

1-2⁺ days

1-2⁺ days

1-N days

Job created & provisioned

Device arrives

Device sent to AWS

Job complete

Courier transport service

Copy data via >1GB/s local network

Your data begins ingesting into your Amazon S3 bucket

Using AWS DataSync and Snow Family together

Large migrations with AWS Snowball and AWS DataSync

On premises storage systems

Initial transfer with AWS Snowball devices

Incremental transfers using AWS DataSync

AWS DataSync
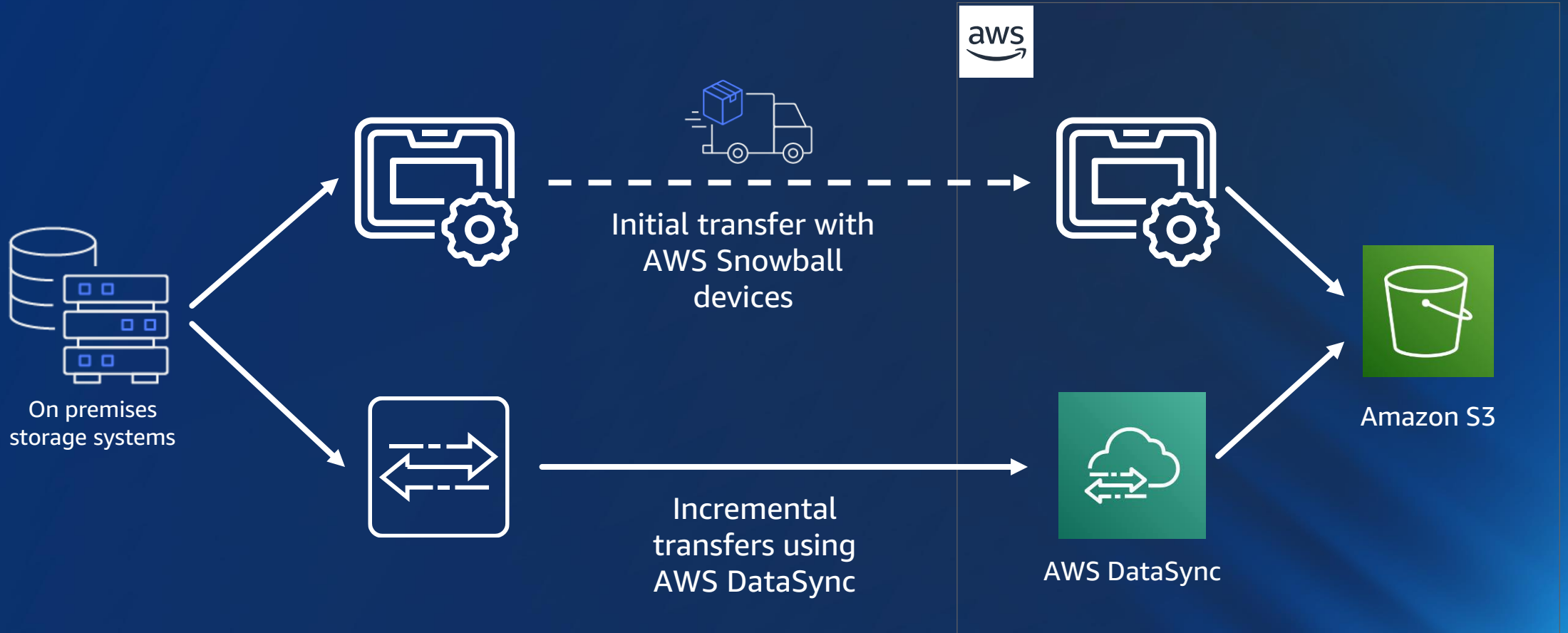
Amazon S3

Ongoing workflows with AWS Snowcone and AWS DataSync

On premises

Vertical-specific equipment

Wireless or wired connection

Sensor and machine data gathering

AWS Snowcone

Local NFS storage

DataSync agent (EC2 instance)

Ship to AWS

AWS Cloud

AWS storage resources

Amazon S3

Amazon EFS

Amazon FSx for Windows File Server

AWS DataSync service
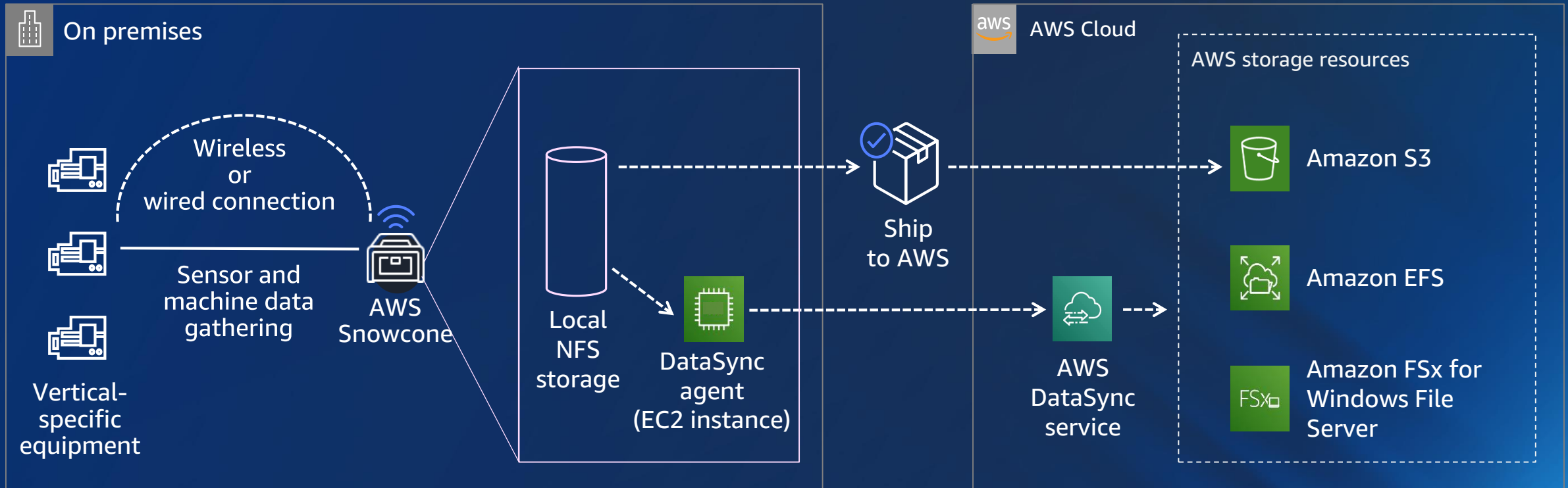
**Demo:
File transfer to Amazon FSx for
NetApp ONTAP with AWS DataSync**

Demo

On premises

/mnt/data/

Local NFS storage

DataSync agent

Application server

/mnt/ontap/

Task to move data

AWS

AWS DataSync service

Amazon FSx for NetApp ONTAP
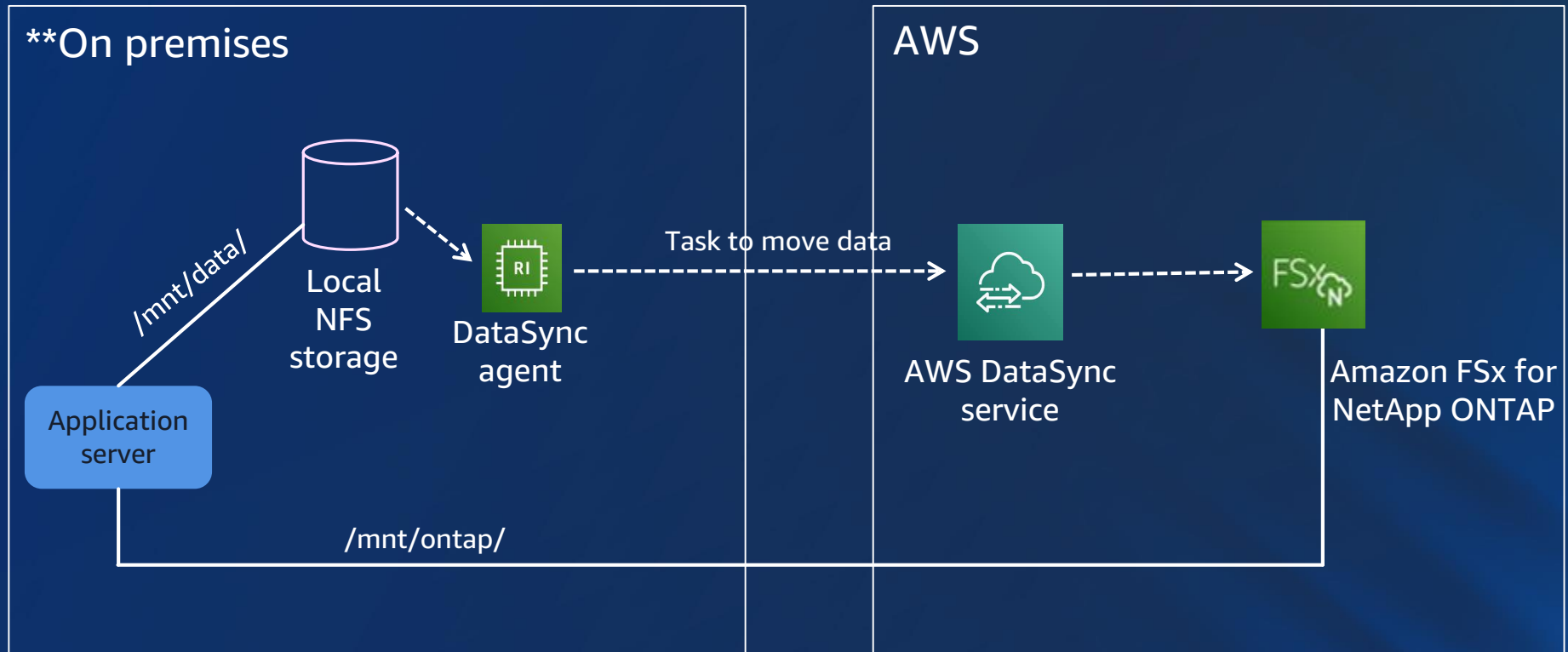
**Simulated as on prem in AWS

Migrating data with AWS Storage Competency Partners

Production data		Backup and archive data		
Online migration	Hybrid cloud	Online migration	Offline migration	Backup, disaster recovery, archive

Production data — Online migration:
komprise
NetApp
wanDISCO
VERITAS

Production data — Hybrid cloud:
ctera
komprise
NASUNI
panzura

Backup and archive data — Online migration:
Tape Ark
IRON MOUNTAIN

Backup and archive data — Offline migration:
IRON MOUNTAIN
Supports via
AWS Snow Family

Backup and archive data — Backup, disaster recovery, archive:
CLUMIO   COMMVAULT
COHESITY   druva
IBM   model
rubrik   veeAM
VERITAS   Zerto

Learn more at
aws.amazon.com/backup-recovery/partner-solutions

Wrap up

Key takeaways

- Evaluate your goals and resources

- Understand the data types you want to move

- Understand the available bandwidth for your data transfer

- Set realistic migration and edge compute timelines

- Leverage the AWS team and AWS partners

Resources

Cost efficiently migrate billions of small files with AWS Snowball

Migrating Azure Blob Storage to Amazon S3 using AWS DataSync

Simplify and accelerate your data migration using AWS DataSync Discovery

Visit the AWS Data Resource Hub

A modern data strategy can help you manage, act on, and react to your data so you can make better decisions, respond faster, and uncover new opportunities. Dive deeper with these resources today.

- Maximizing business value with data: 6 essential data-driven use cases for leaders
- The ultimate guide to developing an end-to-end data strategy
- 5 key elements of a data-driven culture
- A builder's guide to AWS analytics services
- … and more!

https://tinyurl.com/data-hub-aws

Visit resource hub

AWS Training and Certification

Access 600+ free digital courses with AWS Skill Builder

Focus on the cloud skills and services that are most relevant to you across 30+ AWS solutions, including digital self-paced learning plans and ramp-up guides

- Build your future in the AWS Cloud at your own pace

- Advance your skills and knowledge with learning plans

- Validate your cloud expertise with AWS Certification

LEARN YOUR WAY SKILLBUILDER.AWS

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apj-marketing@amazon.com

twitter.com/AWSCloud

facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

Thank you!

Lily Jang

yjang@amazon.com

Ameen Khan S

ameenamz@amazon.com

aws