AWS re:Invent

STG204

Get your data to AWS: How to choose and use data migration services

Everett Dolgner

EMEA specialist SA manager, storage

Amazon Web Services

AWS re:Invent

aws

Agenda

Where to start

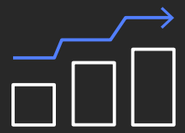
Data migration – offline and online

AWS migration options: CloudEndure, AWS DataSync, AWS Snow family

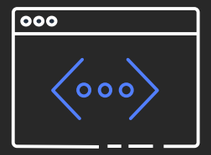
Migration case studies and lessons learned

Lightning round – 5 slides, 5 services, 5 minutes

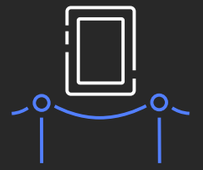
What are we not focusing on today



This is not a 300 or 400 level session



I don’t have code samples (but I can email you some)



There are no protocol level deep-dive slides (but we might go there)

Data migration key questions

Why are you migrating data to AWS?

What are you migrating?

What are the applications?

Where is the data going?

When do you need to finish?

How much data?

How much usable network capacity?

How you move the data depends on all of the above.

You migrated what?

Many ways to migrate data

- OS-based tools
- Scripts
- Open source tools
- Freeware
- Commercial purchase
- Rental

We have been moving data for decades,
but the size is increasing

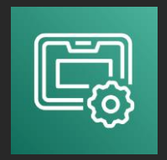
What do you need to think about

- Reporting
- Pause
- Restart
- Failure recovery/retry
- Error checking
- Error correction
- Performance
- Concurrency/parallelization

AWS data migration options

Offline transfer

Bulk data, files, objects, HDFS, databases



AWS Snowball Edge

AWS Snowmobile

Online file and object transfers

Rapid transfers

AWS DataSync

File exchanges

AWS Transfer for SFTP

Long-distance uploads and downloads

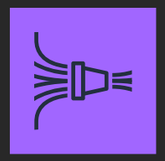
Amazon S3 Transfer Acceleration

Database and machine migration and recovery

AWS Database Migration Service

CloudEndure an AWS Company

Streaming data



Amazon Kinesis family: Data Firehose, Data Steams, Video Streams

Hybrid/Edge gateways

File, block volume, snapshot, and tape backup storage for on-premises apps

AWS Storage Gateway family

APN Partner Products

AWS storage partners

Primary Storage

Solutions that leverage file, block, object, and streamed data formats as an extension to on-premises storage

- ctera
- NASUNI
- NetApp
- panzura
- PURESTORAGE
- SoftNAS
- WEKA.IO
- zadara

Backup and Restore

Solutions that leverage Amazon S3 for durable data backup

- COHESITY
- COMMVAULT
- CLOUDBERRY LAB IS NOW MSP360
- ctera
- druva
- DELL EMC
- IBM
- N2WS
- NetApp
- rubrik
- VEEAM
- VERITAS

Archive

Solutions that leverage Amazon Glacier for durable and cost-effective long-term data backup

- COHESITY
- COMMVAULT
- IRON MOUNTAIN
- NetApp
- rubrik
- TapeArk
- VERITAS

BCDR

Solutions that utilize AWS to enable recovery strategies focused on RTO and RPO requirements

- COHESITY
- druva
- NetApp
- VERITAS
- Zerto

Consulting

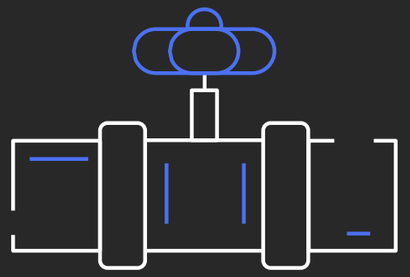
Consulting services that provide implementation capabilities in one or more core storage categories

- AHEAD
- shi
- CompuNet, Inc.
- SIRIUS
- HCL
- softchoice
- infiniti
- SUNGARD AVAILABILITY SERVICES
- ONICA
- rackspace
- PRESIDIO
- PROACT

aws.amazon.com/backup-recovery/partner-solutions

Partner who has achieved Storage Competency

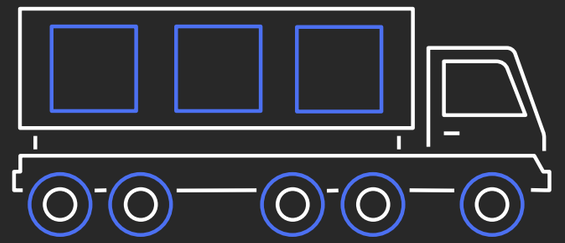
Online or offline migration



Online data migration

Data is read from the source and transferred across the network to the destination in real-time, asynchronously

Online requires enough WAN or Direct Connect bandwidth to complete the copy in the needed time

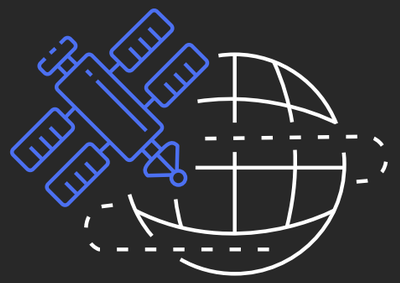


Offline data migration

Data is copied to a device that is shipped and used to load the data to the destination

Offline only requires LAN bandwidth between the storage and the device

Online or offline



Online

High bandwidth WAN or AWS Direct Connect

Bandwidth is reliable and error free

Data must be online and available during move

Enough time to complete migration

Data is in AWS as it is transferred

Offline

Low bandwidth or heavily utilized WAN or AWS Direct Connect

Bandwidth is unstable

Data can be offline during move

Data Source remote or disconnected

Time to data in AWS is days

Online or offline

Time/transfer rates/data (TB)/resources

Usable network bandwidth

	100 Mbps	1 Gbps	10 Gbps
1 TB	30 hours	3 hours	18 minutes
10 TB	12 days	30 hours	3 hours
100 TB	124 days	12 days	30 hours
1 PB	3 years	124 days	12 days
10 PB	34 years	3 years	124 days

Assumes ~25% network overhead

Online data transfer

AWS DataSync

Simplifies, automates, and accelerates your online data transfer

Fast data transfer

Up to 10 Gbps per agent (100 TB/day)

Highly parallel optimized network transfer

Scale-out agents to go faster!

Incremental transfers

Up to 10x faster than rsync or robocopy

Easy to use

Simple data movement from NFS/SMB to Amazon S3, Amazon FSx, or Amazon EFS

Include/exclude filtering

Task scheduling

Configurable bandwidth limits

No in-cloud infrastructure to deploy, run or scale

Secure and reliable

Securely accesses AWS storage services

VPC and FIPS endpoints

Data transfer encrypted

End-to-end data validation

Automatic recovery from I/O or transmission errors

Integrated

Native support for Amazon S3, Amazon EFS, and Amazon FSx

Support for all Amazon S3 storage classes

Amazon CloudWatch metrics, logs, and events

AWS CloudTrail logs

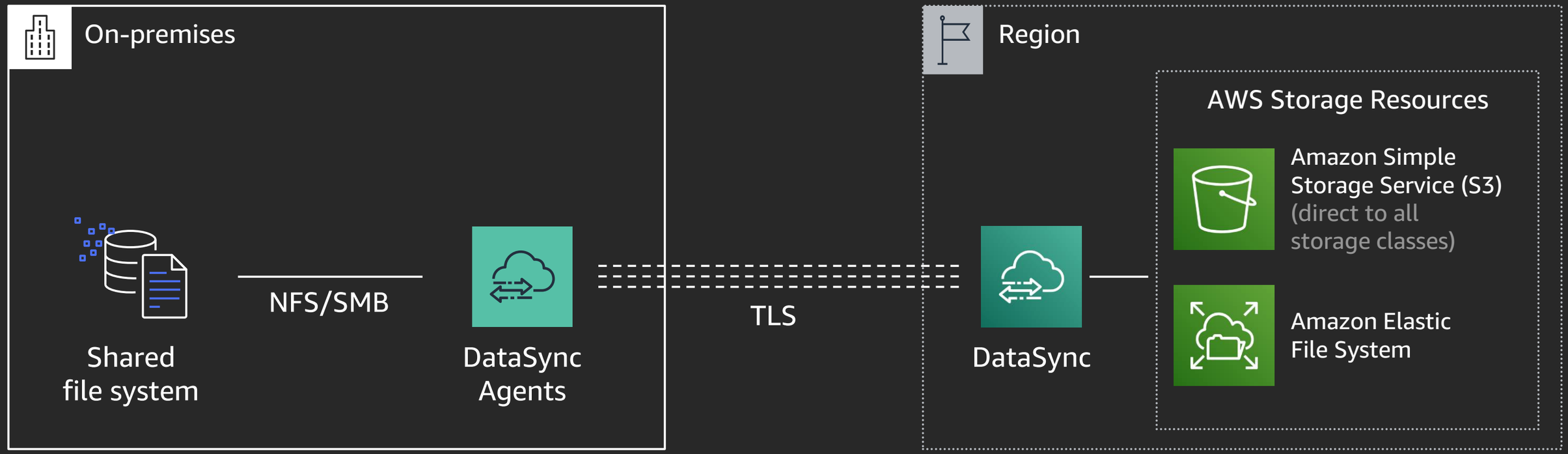
Cost-effective

Pay-as-you-go pricing—$0.0125/GB

Predictable

No minimums

How DataSync works



Deploy agent on VMware or Amazon 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 reads and writes to Amazon S3, Amazon EFS, and Amazon FSx using IAM/VPCE

DataSync tips

Bottlenecks are a moving target

The WAN might not be the biggest bottleneck

Every part of the network is critical

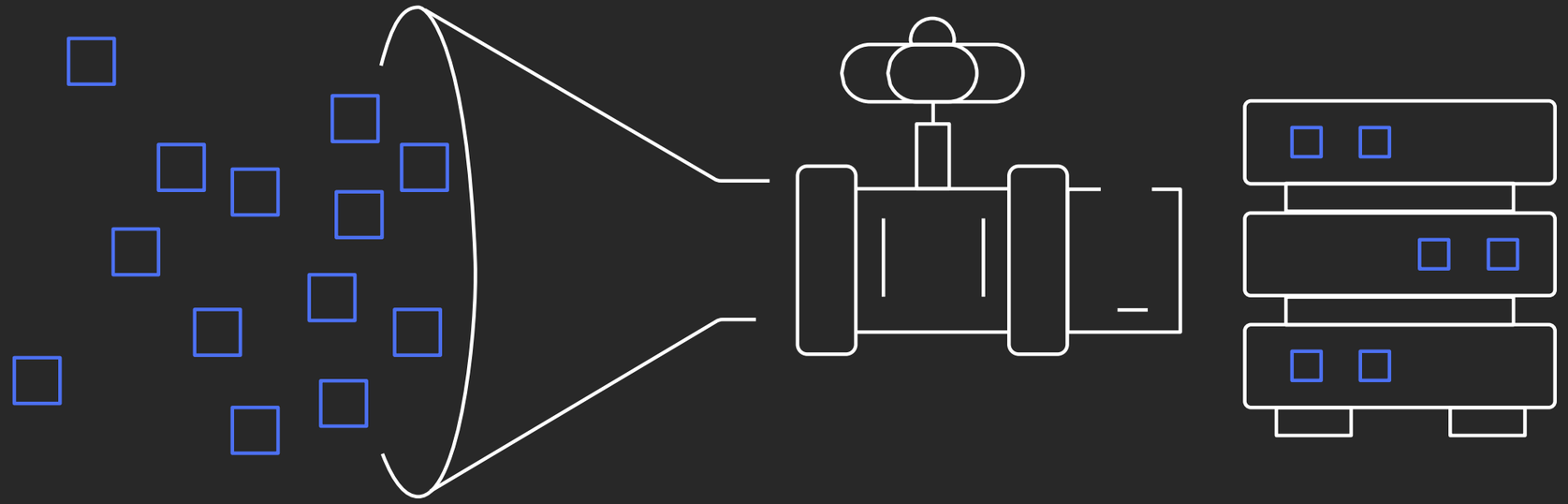
Source system configuration dictates read performance

Protocol errors are devious

To verify or not to verify… it’s a question… but not the right one

When to verify, that is the question

Will it sync?



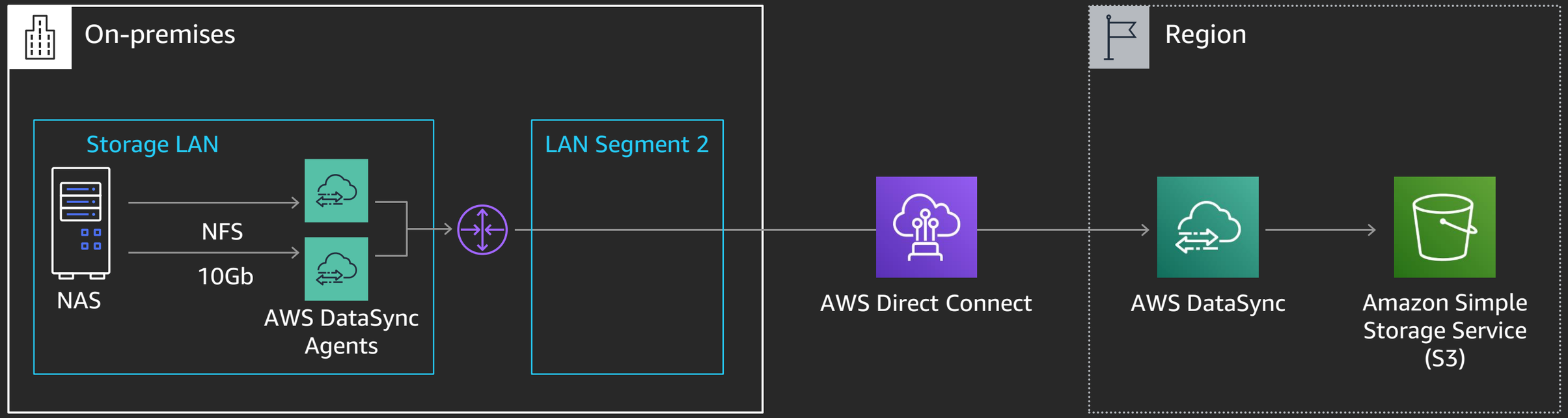
6PB of data stored on a NAS device

90 days to move it into Amazon S3

Successful previous migration with **AWS Snowball Edge**

10 Gb DX

Will it sync?



100TB per day across two DataSync agents

90 days to move 6PB

Verification after moving a share or an aggregate of directories

Will it sync? Lessons learned

The network path between the storage device, DataSync Agent, and DX is critical

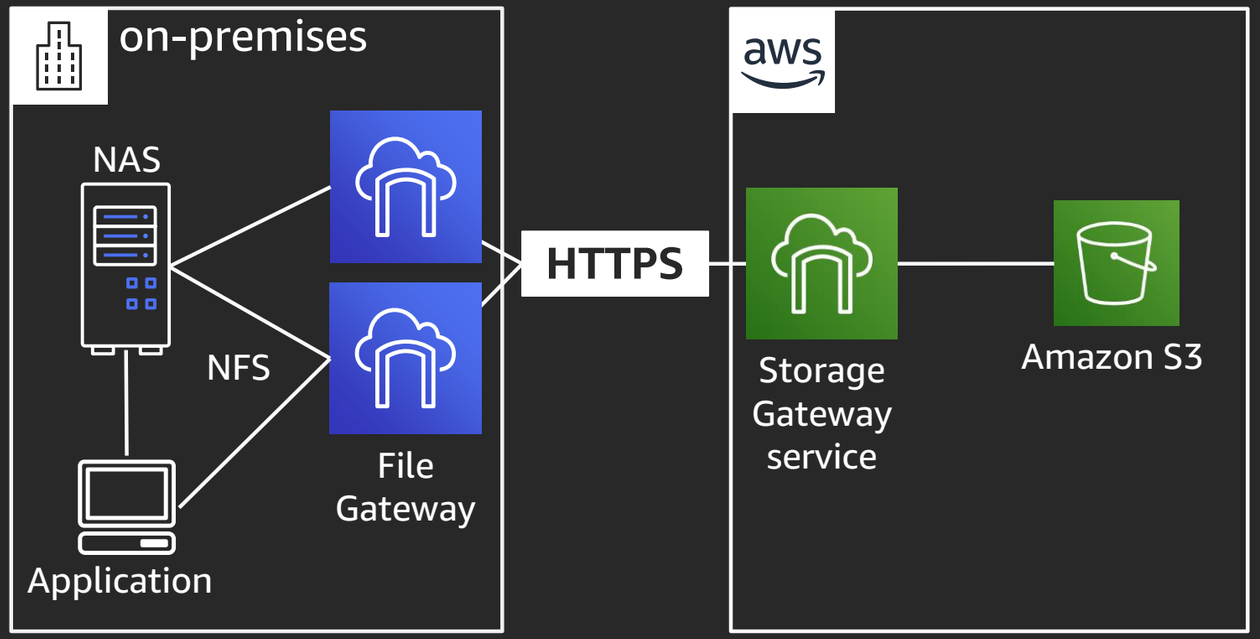
The NAS device had issues serving data fast enough

We saw multiple NFS errors

Scale DataSync Agents to fill the bandwidth, and to distribute connections across the source

AWS Storage Gateway migration

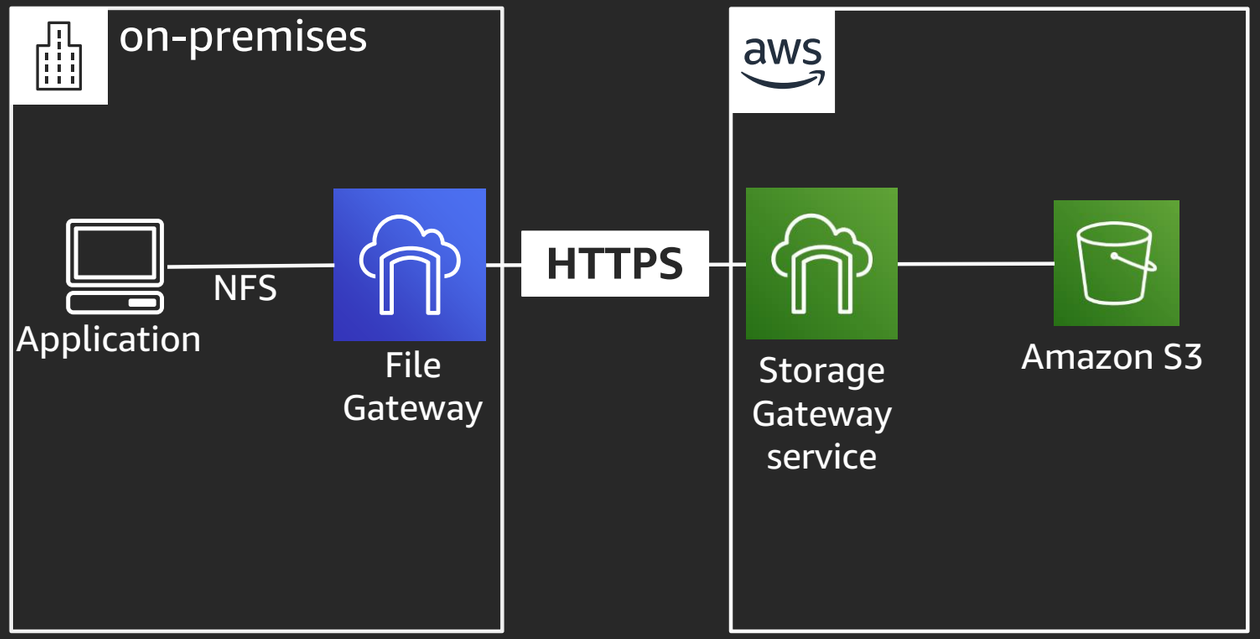
File migration with application support



- Application accesses patient data over NFS
- DX to AWS for migration
- Multiple File Gateways to fill the WAN
- Converting application to native Amazon S3
- File Gateway used to access Amazon S3 during application conversion

AWS Storage Gateway migration

File migration with application support



- Application accesses patient data over NFS
- DX to AWS for migration
- Multiple File Gateways to fill the WAN
- Converting application to native Amazon S3
- File Gateway used to access Amazon S3 during application conversion

Storage Gateway migration learnings

- Concern that the data is the same in Amazon S3 as it was on-premises
- To verify data, EC2 instances recalculated the hash for each file
- File Gateway does a hashed PUT to guarantee data is the same in Amazon S3 as cache

Simplify and accelerate migration with CloudEndure

Flexible

Migrate from any source

Wide range of OS, application, and database support

Option to migrate back

Reliable

Robust, predictable, non-disruptive continuous replication

Short cutover windows with minimal downtime

Highly secure for regulated environments

Highly automated

Minimal skill set required to operate

Easy, non-disruptive tests prior to cutover

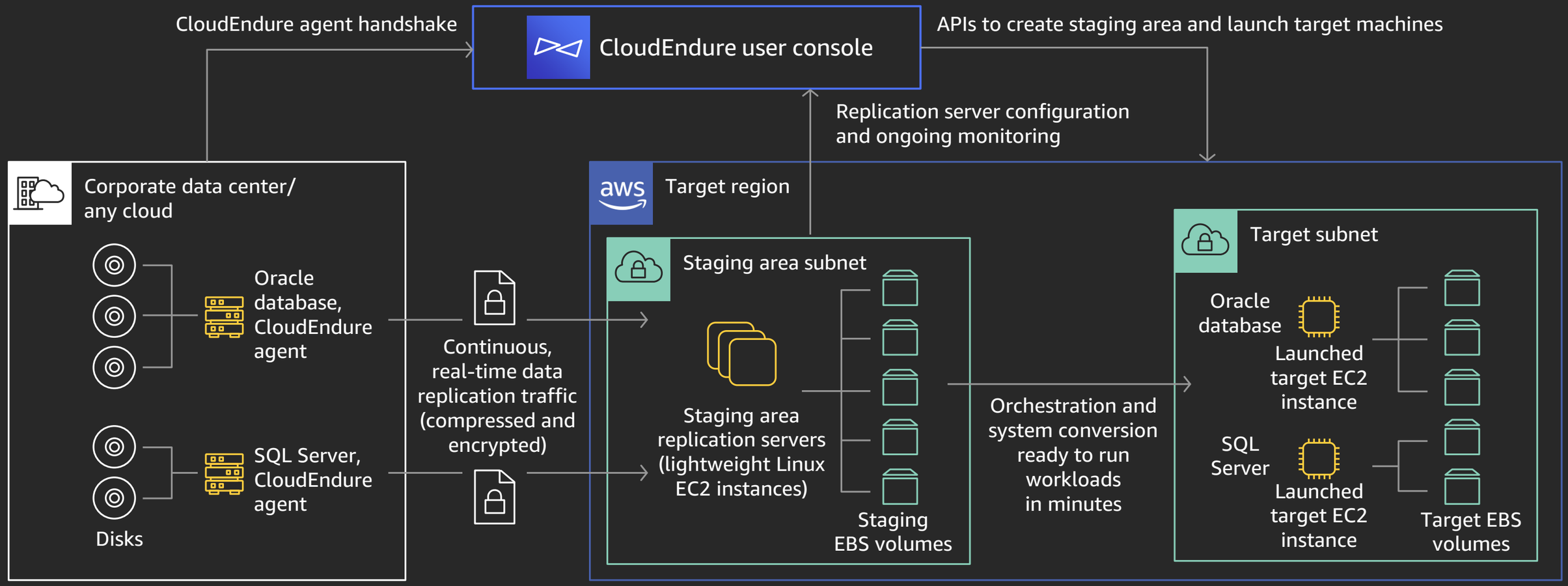
Easily plugs into migration factories and cloud COEs

- Designed for rapid, large-scale migrations
- Simple setup lets you start in minutes
- Same highly automated process for any workload
  - Common workloads include databases such as Microsoft SQL Server, Oracle, and MySQL, and enterprise applications such as SAP
- Eliminates complexity and reduces risk
- Migrate with minimal business disruption

How CloudEndure migration works

CloudEndure continuously replicates any application or database from any source into AWS

Business outcome: Allow self-service, rapid, reliable migrations with minimal business disruption



Wide platform support

Any application

Any database

X86 operating systems

Source infrastructure

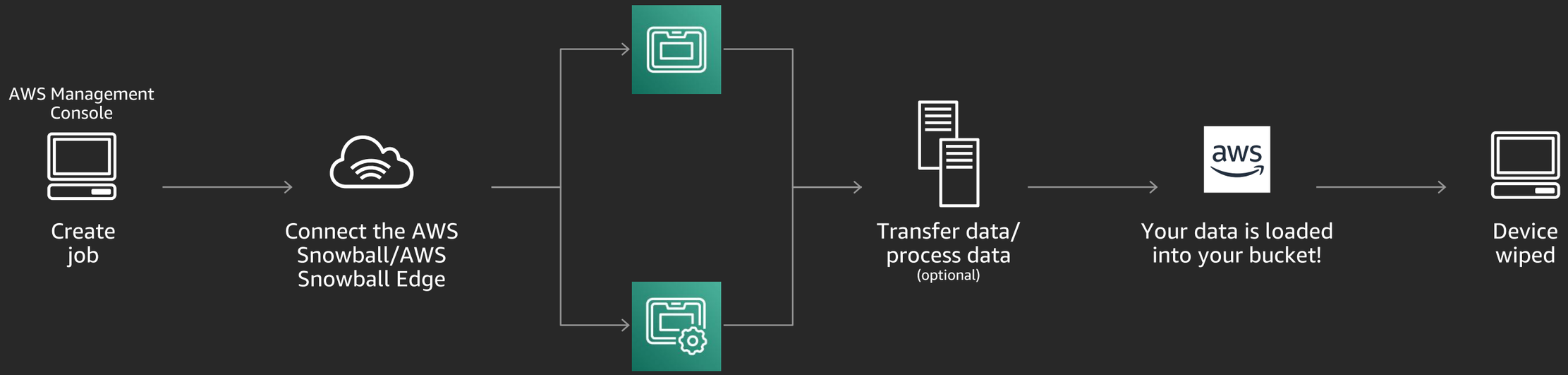
* See documentation or contact Support for a complete list

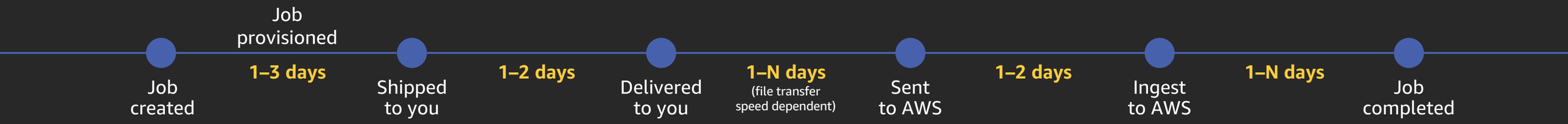
Offline data transfer

AWS Snow family for data collection and movement

	Snowball	Snowball Edge	Snowmobile
Migration size	Up to petabytes, offline		Up to exabytes offline
Form factor	Rugged 8.5 G impact cases that are rain and dust resistant, e-ink label for shipping automation		45-foot container, scheduled delivery
Security	Encryption, tamper detection		Encryption, security staff, GPS tracking, video surveillance, alarms, etc.
Capacity	42 TB or 72TB usable	82 TB usable	<100 PB
Compute		Amazon EC2 or AWS IoT Greengrass processing to use applications or functions to load or pre-process data	

AWS Snowball Edge import workflow

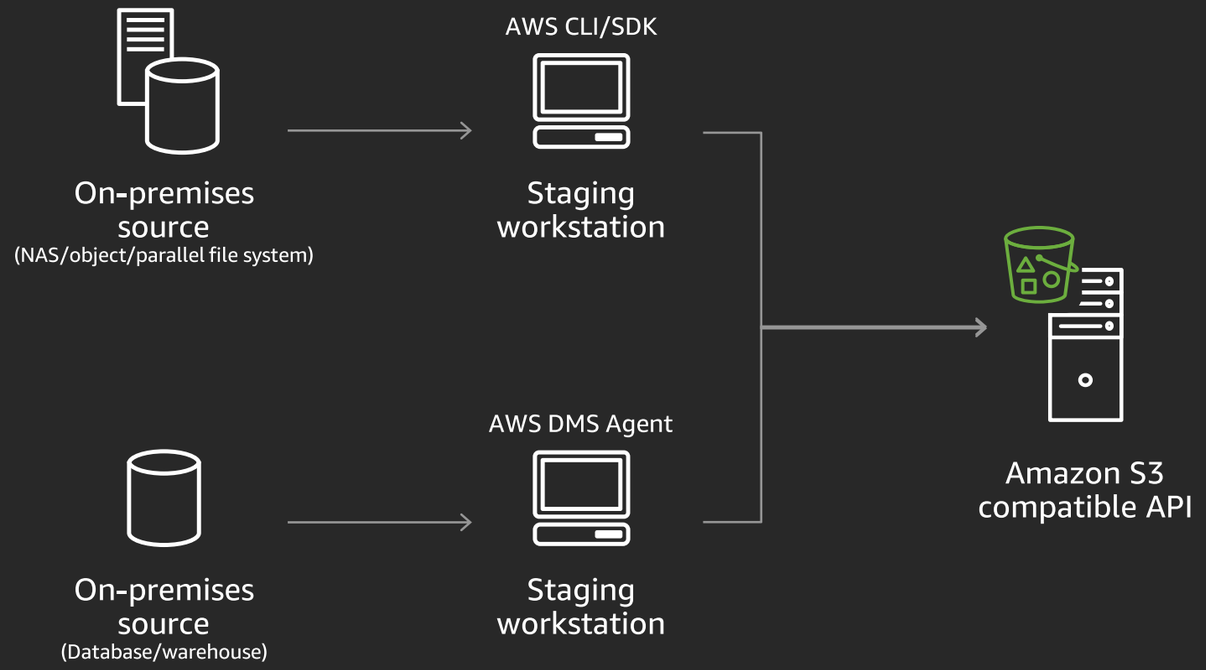




Large-scale migrations with Snowball Edge
Learned from multi-petabyte customer migrations

Run a proof of concept (POC)

- Early discovery and remediation of environmental issues
- Sets more realistic migration and edge compute timelines
- Deploy staging workstations
- Ensure low network latencies (<1ms)
- Ensure larger files (>5MB)
- Benchmark & optimize data transfer (target 300-500 MBps)



Plan devices and scheduling with your account team/TAM before ordering jobs

Resources

- White paper: AWS Snowball Edge data migration guide
- Blog: Data migration best practices with Snowball Edge

Large-scale migration strategies

Methods
Presenting data in a consumable manner by AWS CLI (staging or direct transfer)

Partitioning
Dividing large set of files into manageable chunks

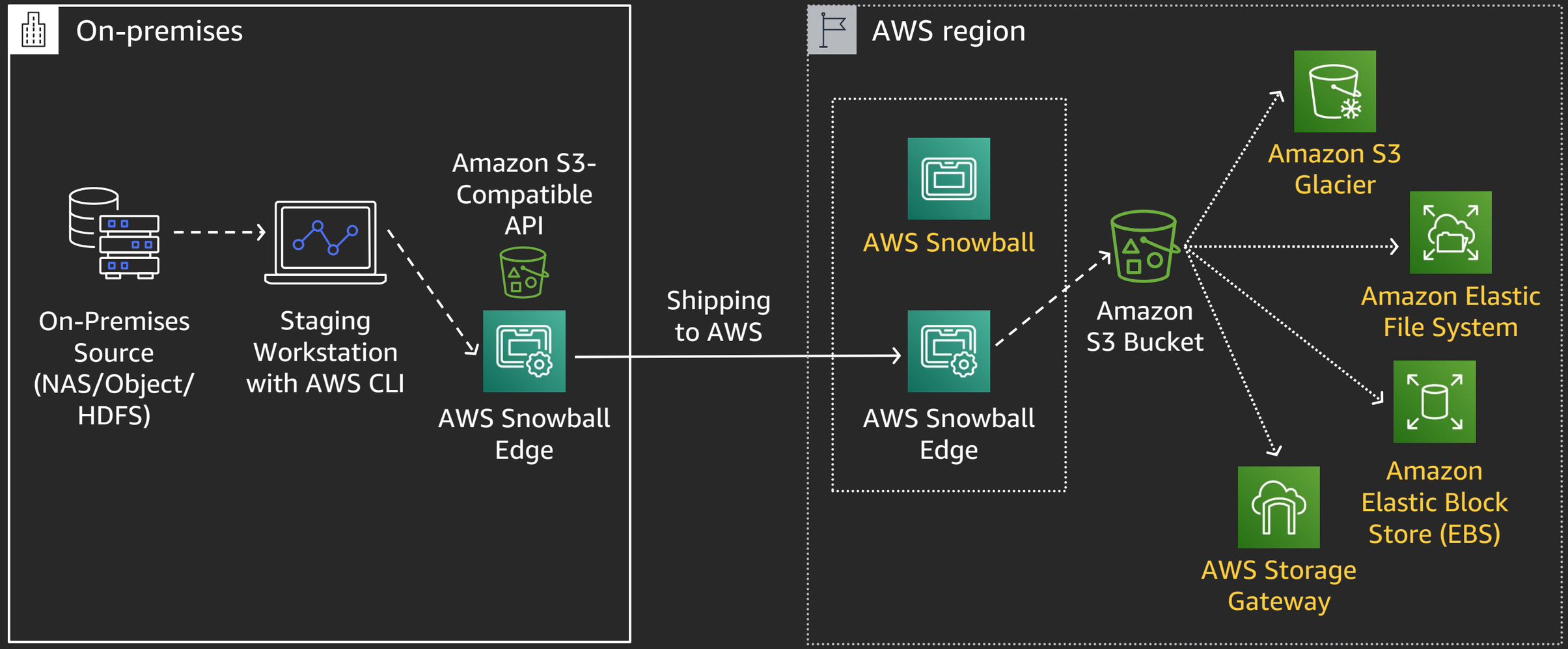
Job pipeline
Maintaining number of devices in order/shipment/at site

Batching
Packaging a subset of files for small file optimization or meta-data preservation

Parallelization
Running simultaneous transfers

Workflow optimizations
Administration, data transfer, verification, erasure (in case of staged data), shipping, import, and verification in Amazon S3

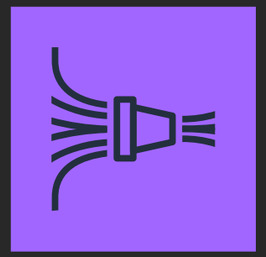
Snowball data transfer workflow



Lightning round

Lightning round

AWS Transfer for SFTP



Amazon Kinesis Data Firehose

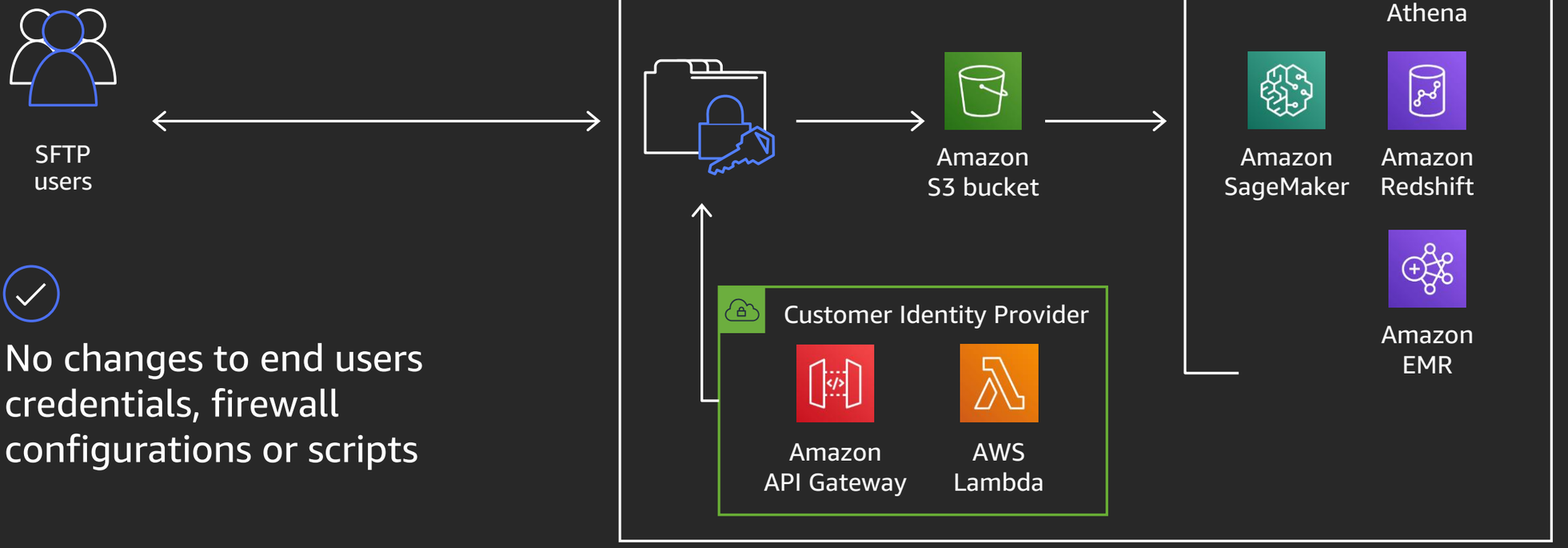
Amazon S3 Transfer Acceleration

AWS Storage Gateway

AWS Database Migration Service

AWS SFTP

No need to manage SFTP infrastructure



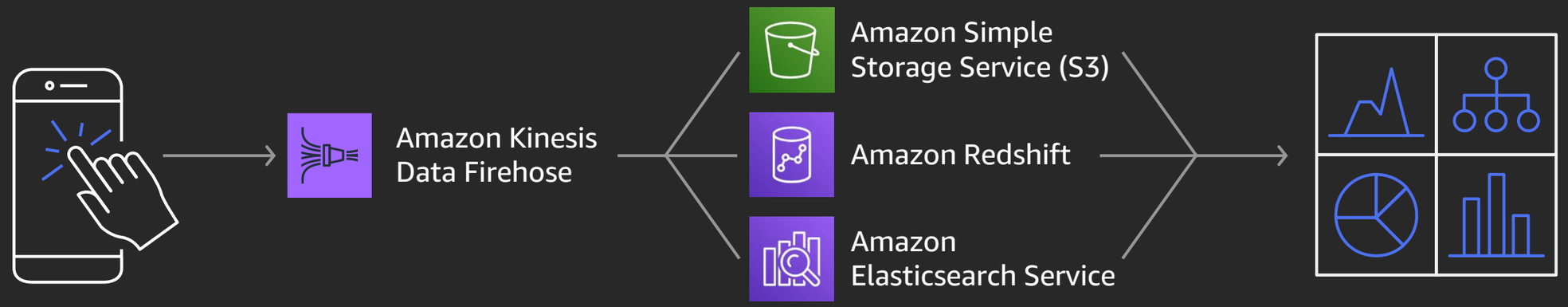
No changes to end users credentials, firewall configurations or scripts

Modernize your workflow using cloud native services

Amazon Kinesis Firehose

- Load streaming data into data stores and analytics tools
- Enable near real-time analytics with existing BI tools and dashboards
- Automatically scales to match throughput of data
- Batch, compress, and encrypt data before sending it to the cloud

HOW IT WORKS



Capture and submit streaming data to Kinesis Data Firehose

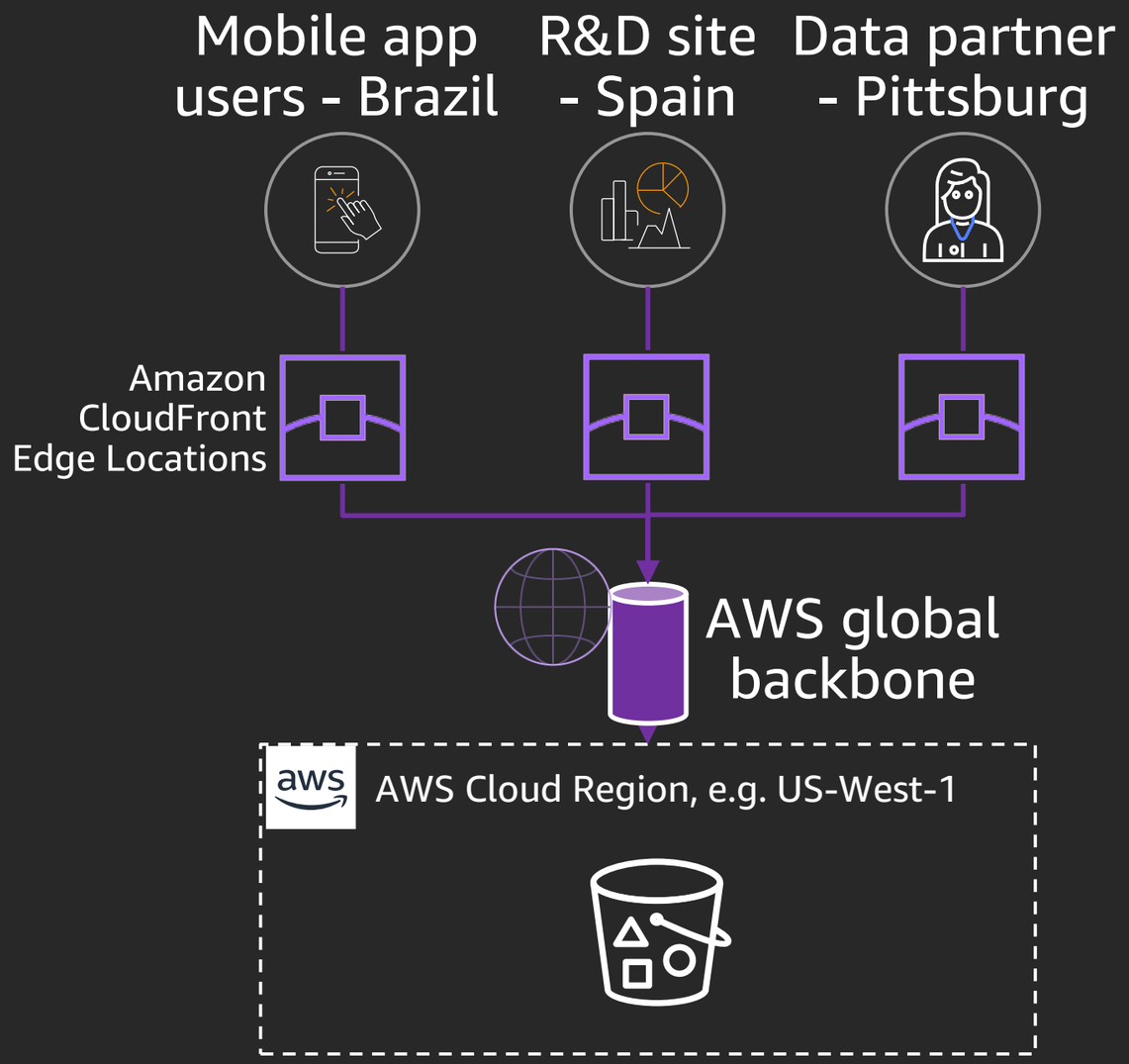
Kinesis Data Firehose transforms and loads streaming data into Amazon S3, Amazon Redshift, and Amazon ES domains

Analyze streaming data using your favorite BI tools

Amazon S3 Transfer Acceleration (S3TA)

Faster long-distance Amazon S3 uploads & downloads

- Speed up transfers of large objects to/from Amazon S3 buckets over long distances
- S3TA routes puts & gets to closest AWS Edge location and over AWS backbone
- Enable per bucket and use “s3-accelerate” endpoint domain names
  - *bucketname*.s3-accelerate.amazonaws.com, or
  - for IPv6 *bucketname*.s3-accelerate.dualstack.amazonaws.com

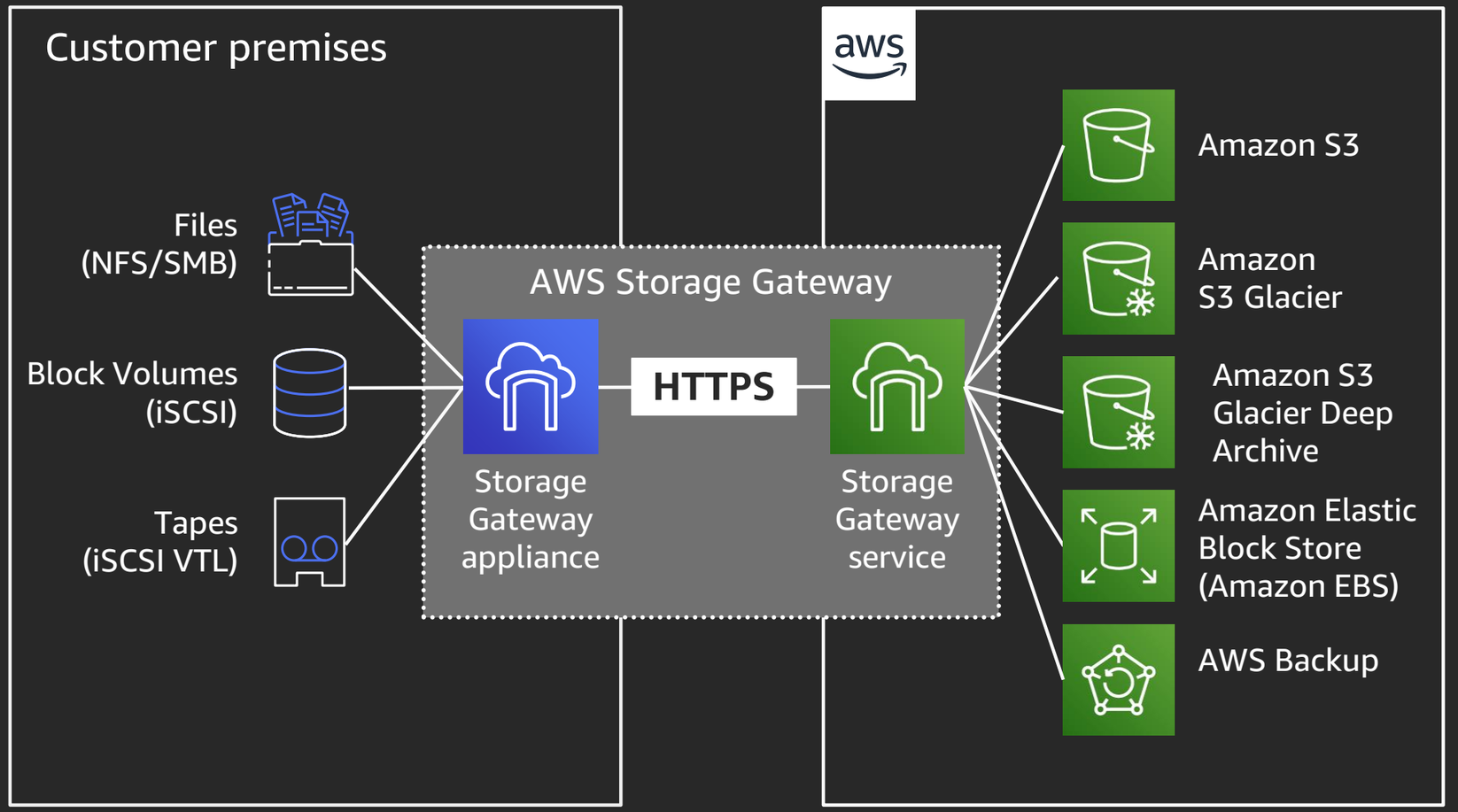


Good for centralized buckets serving

- Mobile & web apps. with distributed users
- Distributed sites with Amazon S3-compatible apps
- Large data exchanges with trusted partners

AWS Storage Gateway

On-premises access to virtually unlimited cloud storage

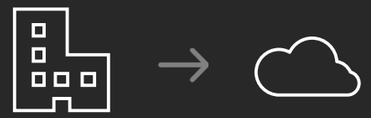


- **NEW:** High-availability support for VMware deployments
- Local VM or hardware appliance
- Low-latency caching
- Multi-protocol
- Managed from AWS console with native integrations
- Optimized data transfer

AWS Database Migration Service (AWS DMS)

Migrating databases to AWS

100,000+
databases migrated

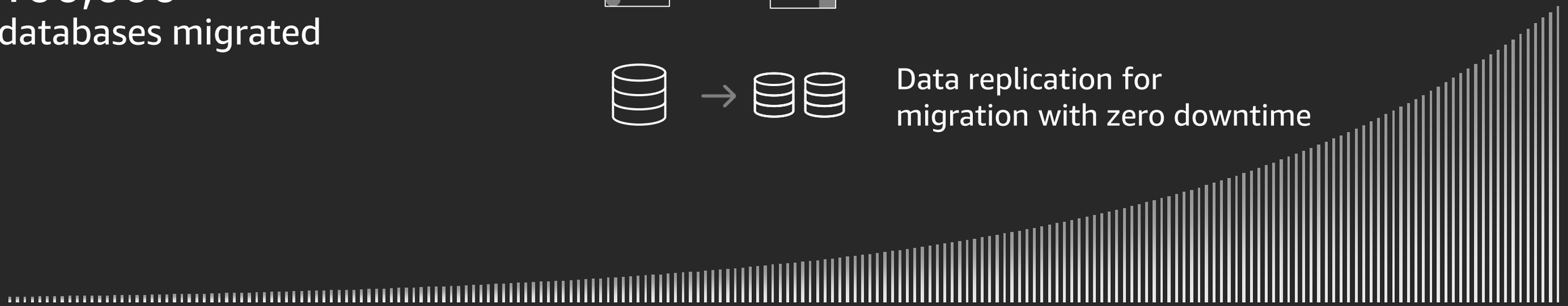


Migrate between on-premises and AWS

Migrate between databases

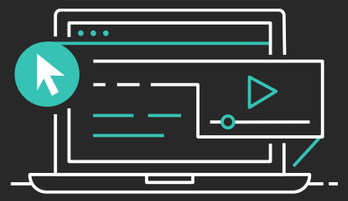
Automated schema conversion

Data replication for
migration with zero downtime



Learn storage with AWS Training and Certification

Resources created by the experts at AWS to help you build cloud storage skills



45+ free digital courses cover topics related to cloud storage, including:

- Amazon S3
- AWS Storage Gateway
- Amazon S3 Glacier
- Amazon Elastic File Storage (Amazon EFS)
- Amazon Elastic Block Storage (Amazon EBS)

Classroom offerings, like Architecting on AWS, feature AWS expert instructors and hands-on activities

Visit aws.amazon.com/training/path-storage/

aws training and certification

Thank you!

AWS re:Invent

aws

Please complete the session survey in the mobile app.