Hybrid Cloud Storage with AWS Storage Gateway & Amazon S3

Paul Reed, Principal Product Manager

August 29, 2018

Why are we here?

**You have data …
and you want to store it in the cloud and access it on-premises**
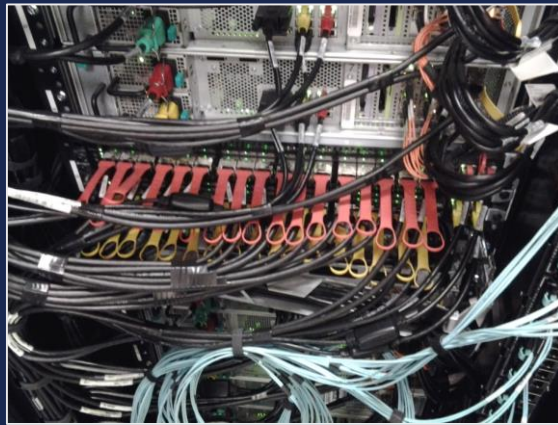

We'll be talking about data that is:

- **Connected to on-premises applications**
- **Part of continuous hybrid workflows**
- **Backup for critical systems**
- **In bulk storage (e.g., NAS filers)**

aws

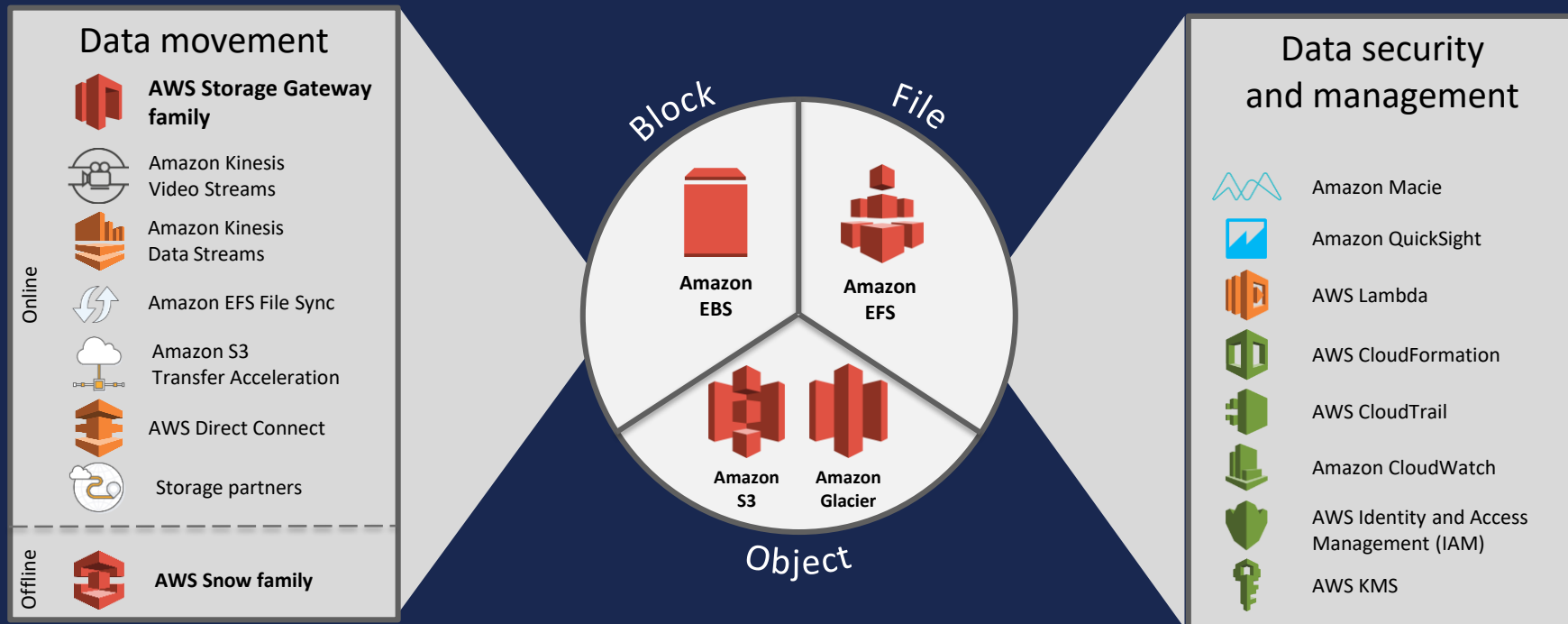# Why move data storage to the cloud?

Retirement

Management

Cost

Complete set of data building blocks

Data movement

Online

AWS Storage Gateway family

Amazon Kinesis Video Streams

Amazon Kinesis Data Streams

Amazon EFS File Sync

Amazon S3 Transfer Acceleration

AWS Direct Connect

Storage partners

Offline

AWS Snow family

Block / File / Object

Block — Amazon EBS

File — Amazon EFS

Object — Amazon S3, Amazon Glacier

Data security and management

Amazon Macie

Amazon QuickSight

AWS Lambda

AWS CloudFormation

AWS CloudTrail

Amazon CloudWatch

AWS Identity and Access Management (IAM)

AWS KMS

aws

So, how can you make storage hybrid?

Gateways

AWS Storage Gateway enables a spectrum of hybrid use cases

Analytics | File services | Machine learning | Data processing
Data distribution | Backup | DR | Archive | Migration

AWS Storage Gateway

Files
(NFS / SMB)

Volumes
(iSCSI)

Tapes
(iSCSI VTL)

Storage Gateway

Gateway

HTTPS

Service

Amazon S3

Amazon EBS
snapshots

Amazon Glacier

vmware®

Hyper-V

Amazon EC2

AWS CloudTrail

IAM

Amazon CloudWatch

AWS Key Management Service
(AWS KMS)

Your data center

AWS Region

AWS
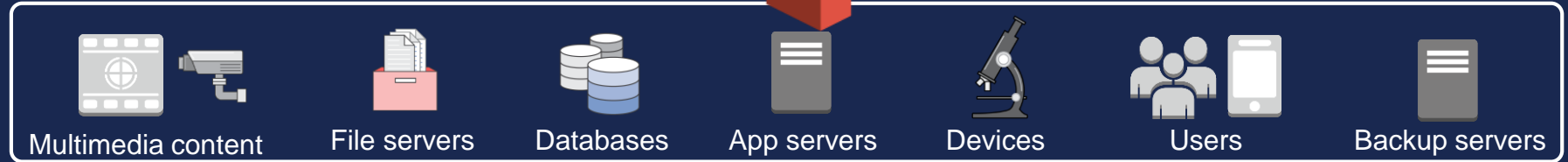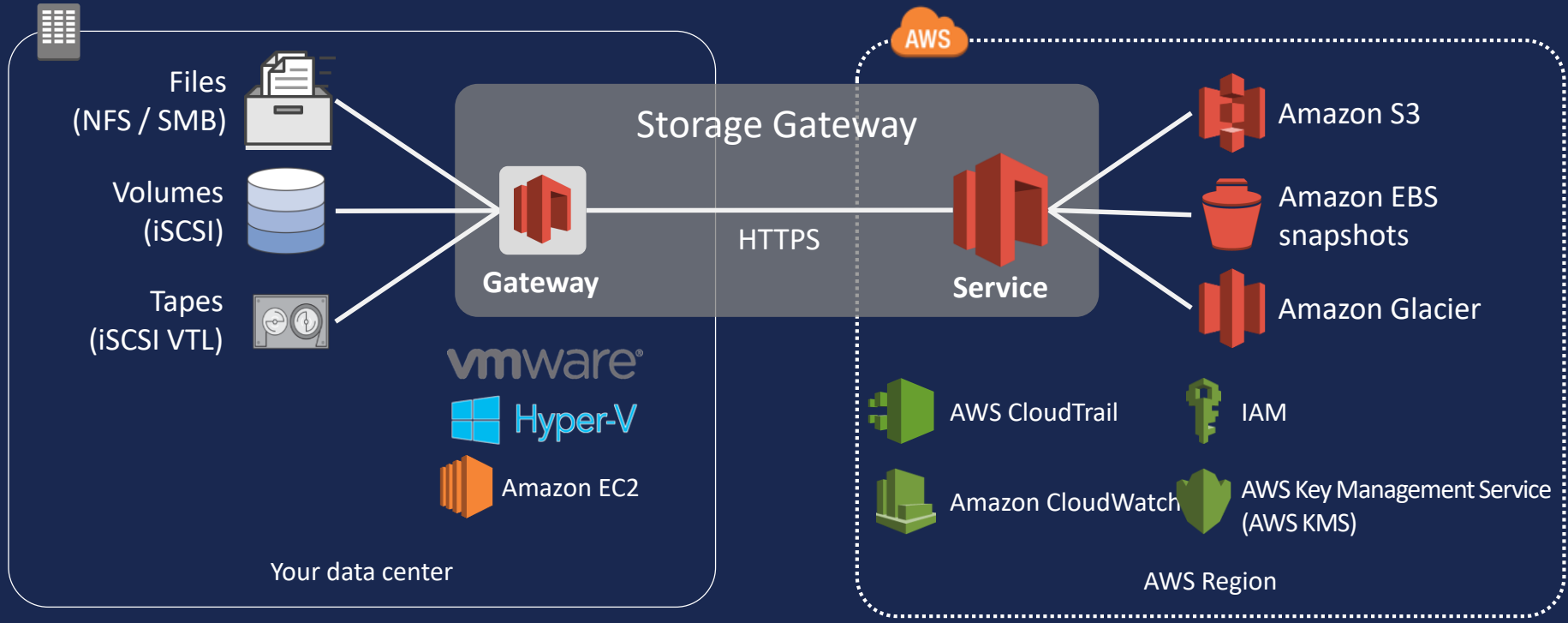
AWS Storage Gateway family

Hybrid storage service enabling applications to seamlessly use AWS storage

File Gateway

Store and access objects in Amazon S3 from file-based applications with local caching

Volume Gateway

Block storage on-premises backed by cloud storage with local caching, Amazon EBS snapshots, and clones

Tape Gateway

Drop-in replacement for physical tape infrastructure backed by cloud storage with local caching

aws

AWS Storage Gateway key features

Standard storage protocols
AWS storage accessible without applications needing to be modified

Fully managed cache
Frequently used data cached locally for low-latency access

Optimized data transfer
Secure upload of changed data and downloads requested data

Durable storage
On-premises application data natively stored in Amazon S3, Amazon Glacier, and Amazon EBS snapshots

aws

File Gateway for hybrid cloud file workloads

File Gateway

Store and access objects in Amazon S3 from file-based applications with local caching

Now supports SMB!

Customer Premises

Application Server — NFS v3/v4 SMB v2/v3 — File Gateway — HTTPS — Objects in your S3 buckets

AWS

Per file share options

Restrict access by client IP (NFS) or Active Directory (SMB) users/groups

POSIX permissions for object-level access*

Read-only/read-write

* Compatible subset of NTFS

## Per S3 bucket options

IAM role

Object storage class

Encryption with AWS KMS

Guess MIME type, requester pays, bucket owner ACL, etc.

aws

File Gateway: Fully managed cache for low latency access

Some of your data cached on gateway

All of your data durably stored in Amazon S3

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

Objects in your S3 bucket

AWS Region
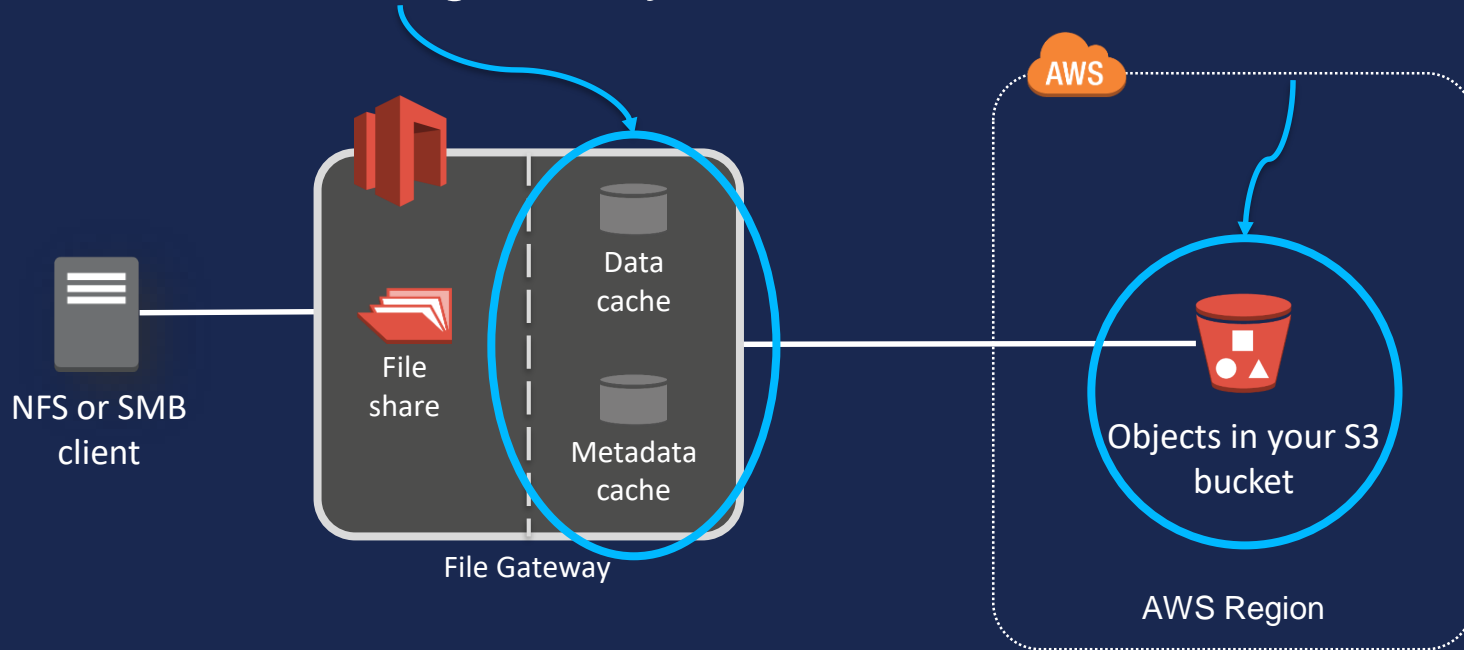
File Gateway: Data cache for file reads and writes

Parts of files cached based on what the client reads/writes

Data cached until space needed for more recently accessed files

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

Objects in your S3 bucket

AWS Region
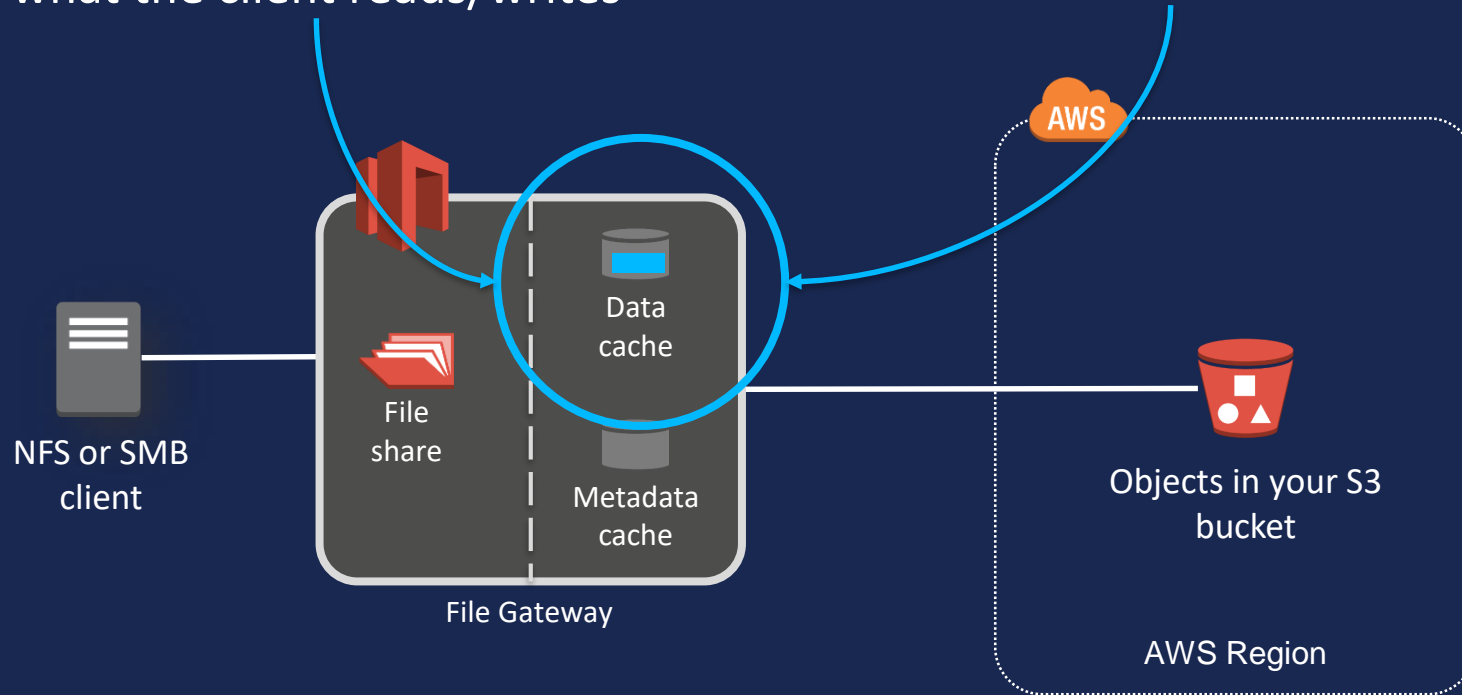
File Gateway: Data is read from the cache

Clients reads data from cache

Ranged GET from Amazon S3 if required data isn't cached

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

Objects in your S3 bucket

AWS Region
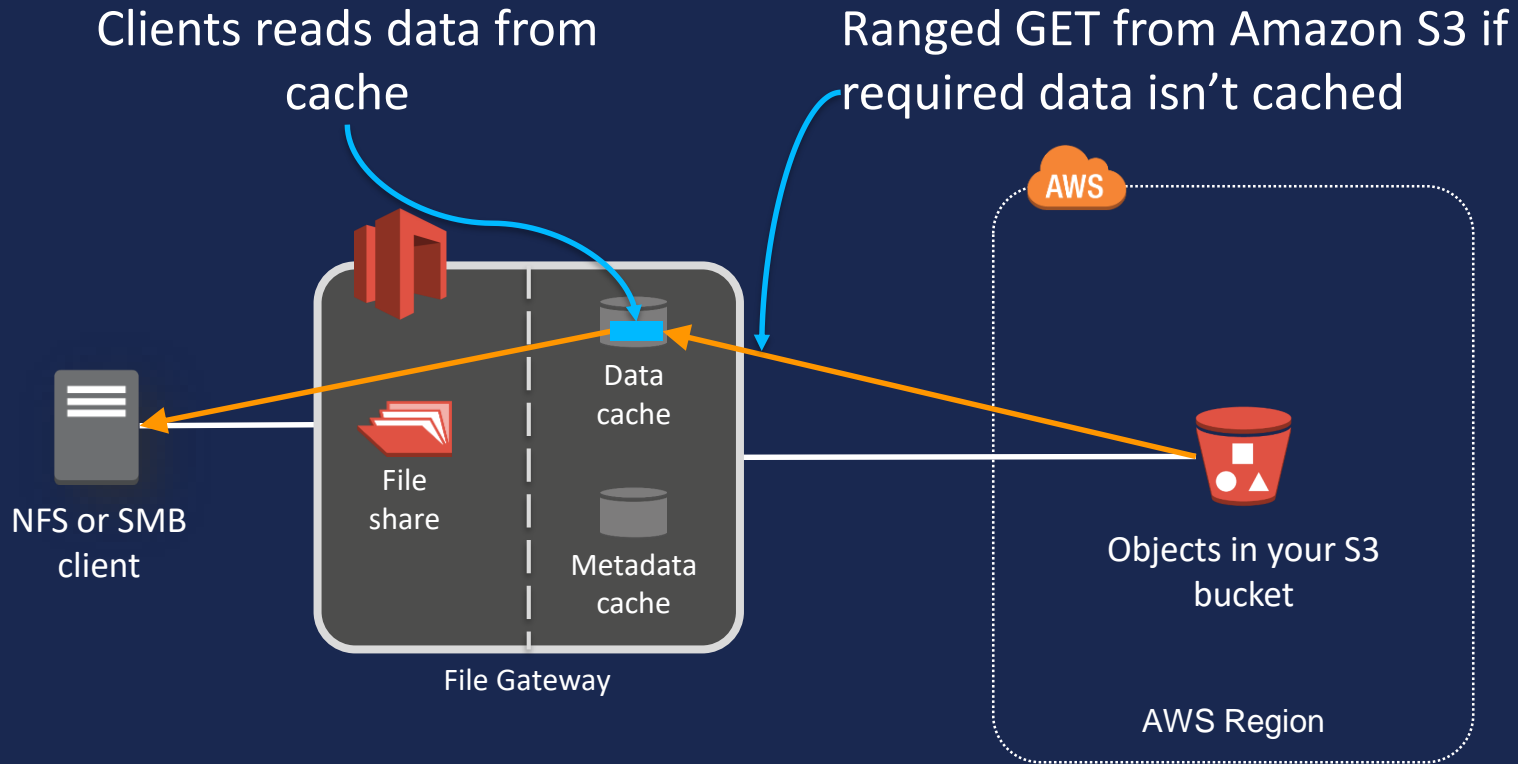
File Gateway: Data is written to cache then uploaded

Write to cache first

Parallel multi-part PUT to Amazon S3 of just the changed data

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

Objects in your S3 bucket

AWS Region
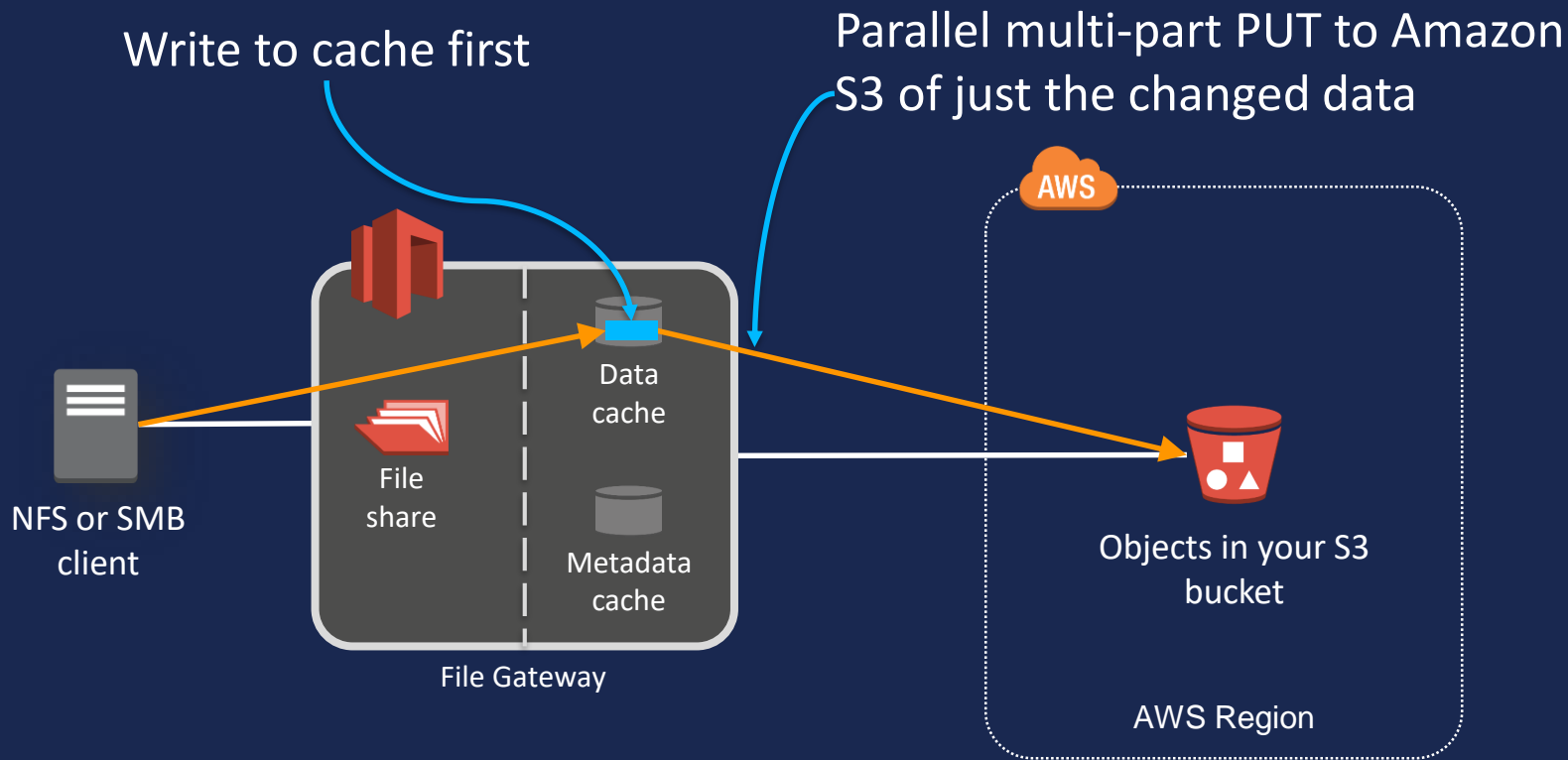
File Gateway: How do I know my data is uploaded?

NotifyWhenUploaded API

Amazon CloudWatch Events

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

AWS

Objects in your S3 bucket

AWS Region

AWS Lambda

Amazon Athena

Amazon Rekognition

Amazon EMR

Amazon QuickSight
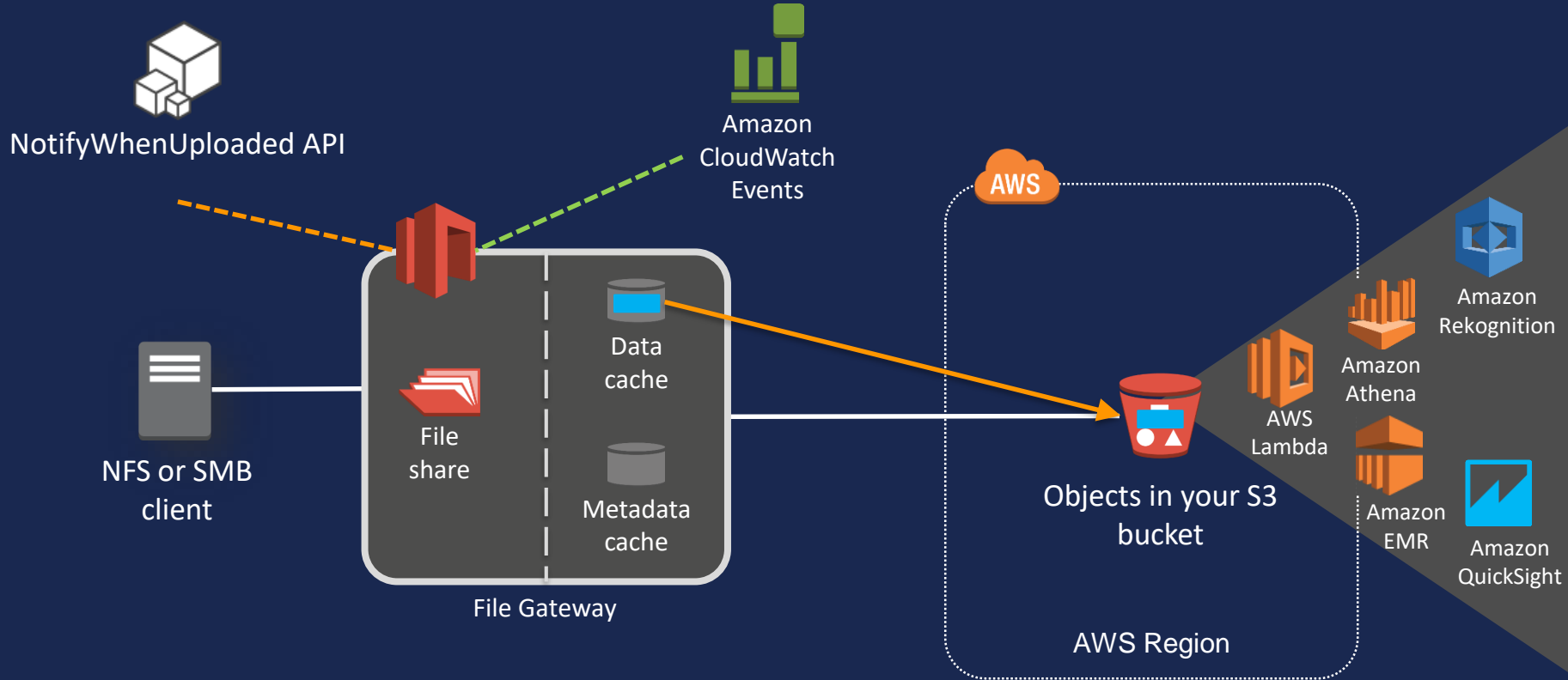
File Gateway: Metadata cache for folder operations


```
C:\> dir D:\foo
$ ls /foo/*
```

```
s3:ListBucket /foo/*
```

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

Objects in your S3 bucket

AWS Region
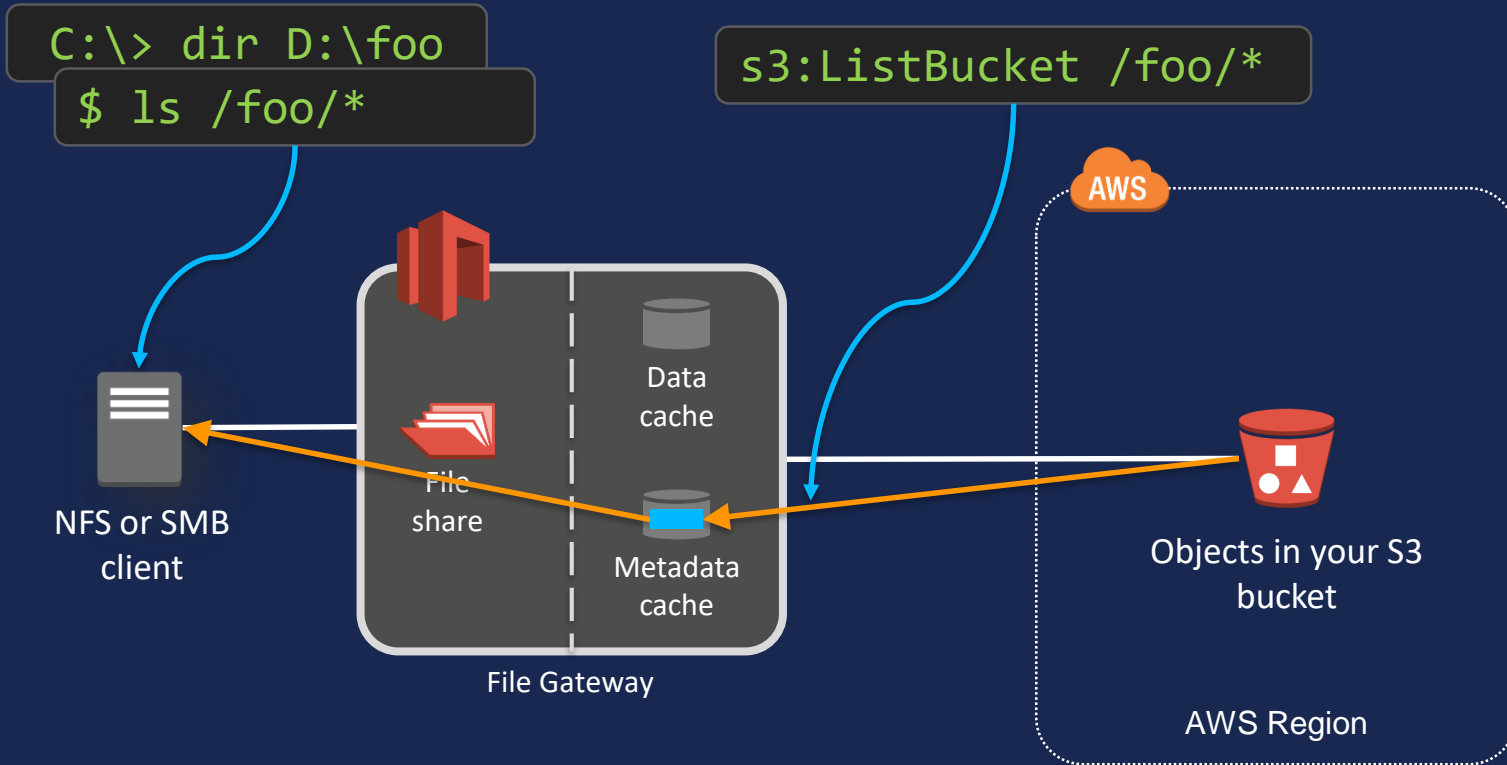
File Gateway: Updating the metadata cache

RefreshCache API

NFS or SMB client

File share

Data cache

Metadata cache

File Gateway

s3[ill]stBucket

Amazon CloudWatch Events

AWS

New object

Objects in your S3 bucket

AWS Region

aws
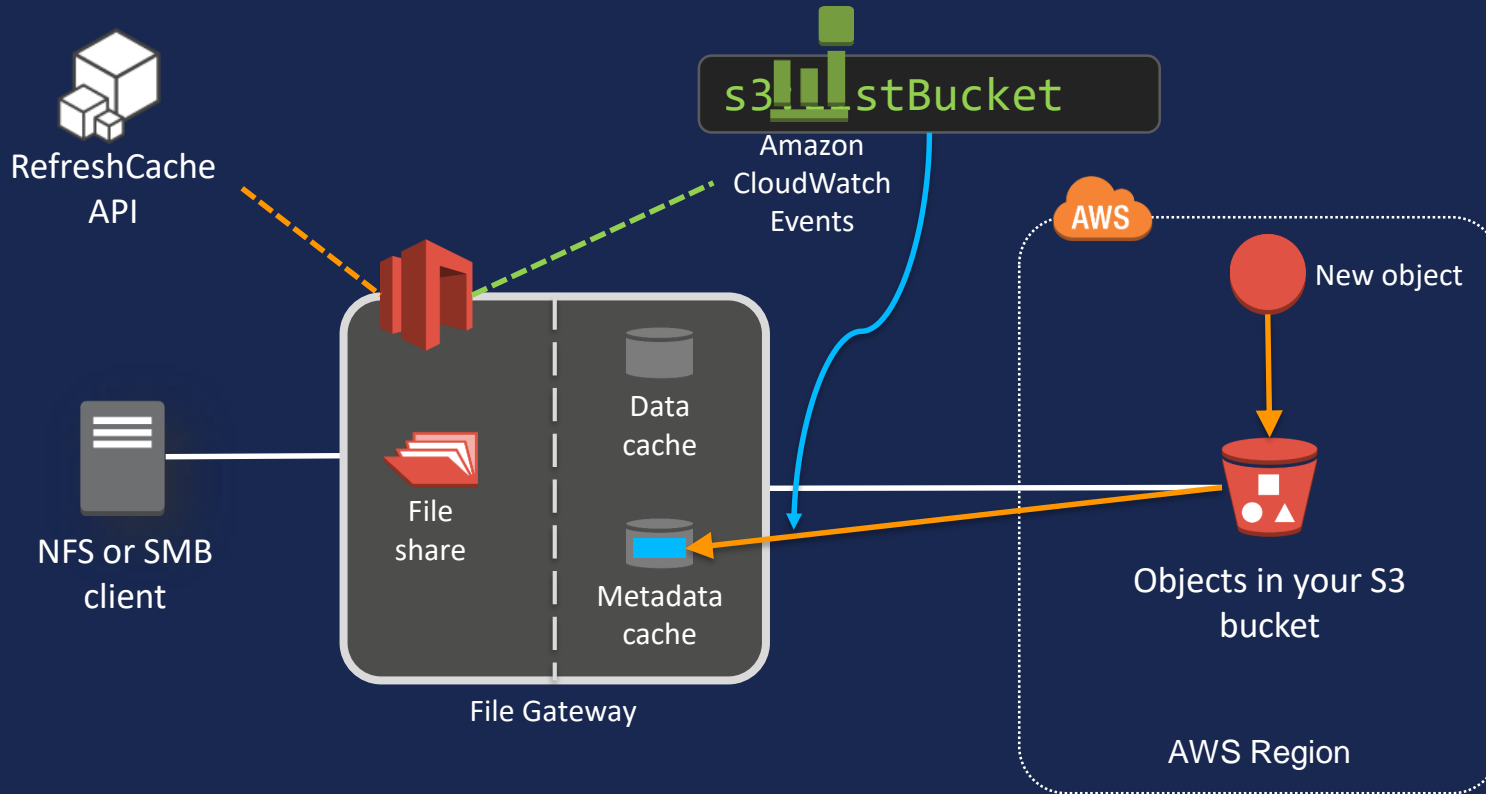
Hybrid File Storage Use Cases

Hybrid file use case: Low-latency content editing

Los Angeles office

Editing workstation

Editing workstation

File gateway

1

AWS

2

Content repository in Amazon S3
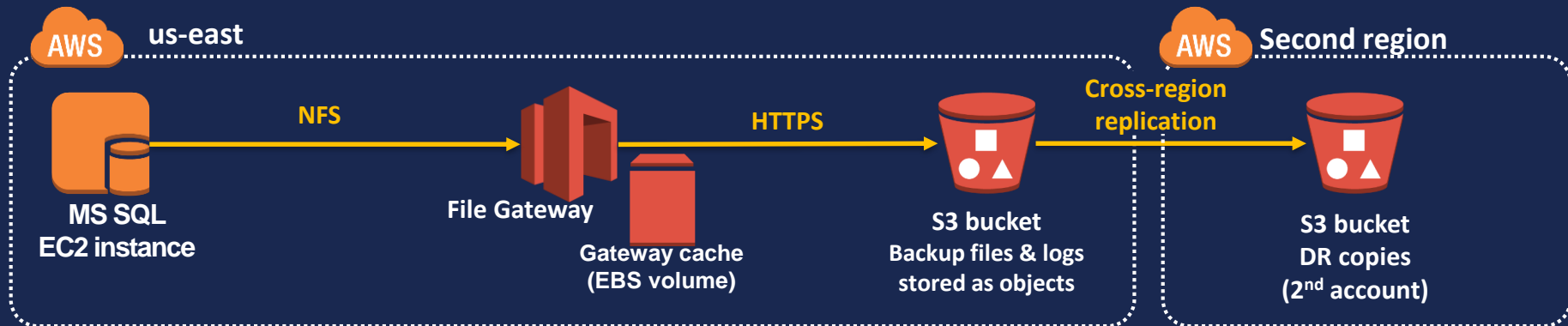
AWS Region
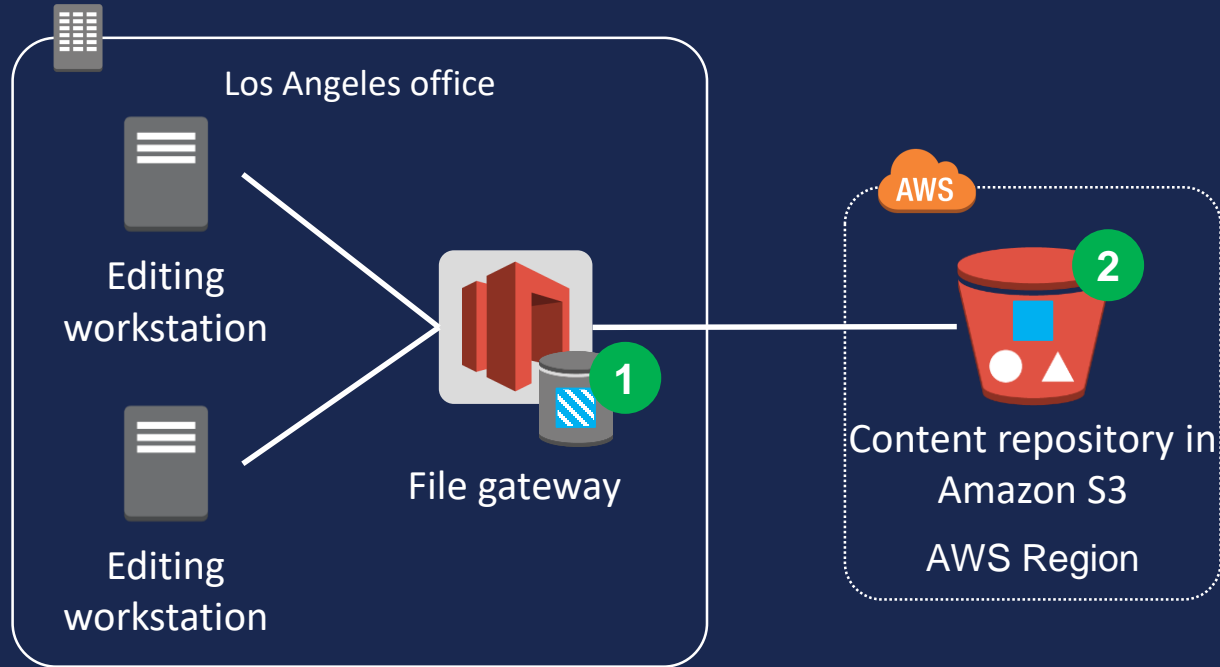
1 Low latency access for data actively being edited

2 Changes are all durably stored in master repository in Amazon S3
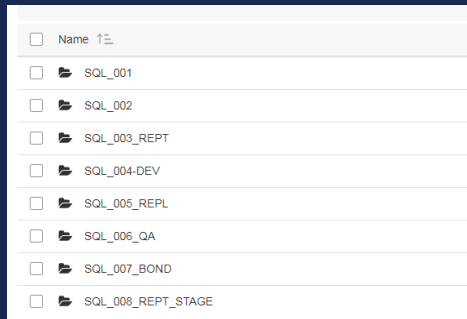
aws

Hybrid file use case: Content distribution

1 Application in Seattle writes files, which are uploaded to Amazon S3 by gateway

2 After refresh cache, files are visible to applications in Boston
Local cache improves access performance

Hybrid file use case: SQL Server backups


```
mount -o nolock
[gateway-IP]:/[bucket-
name]/[sql-instance]
[WindowsDriveLetter]:
```

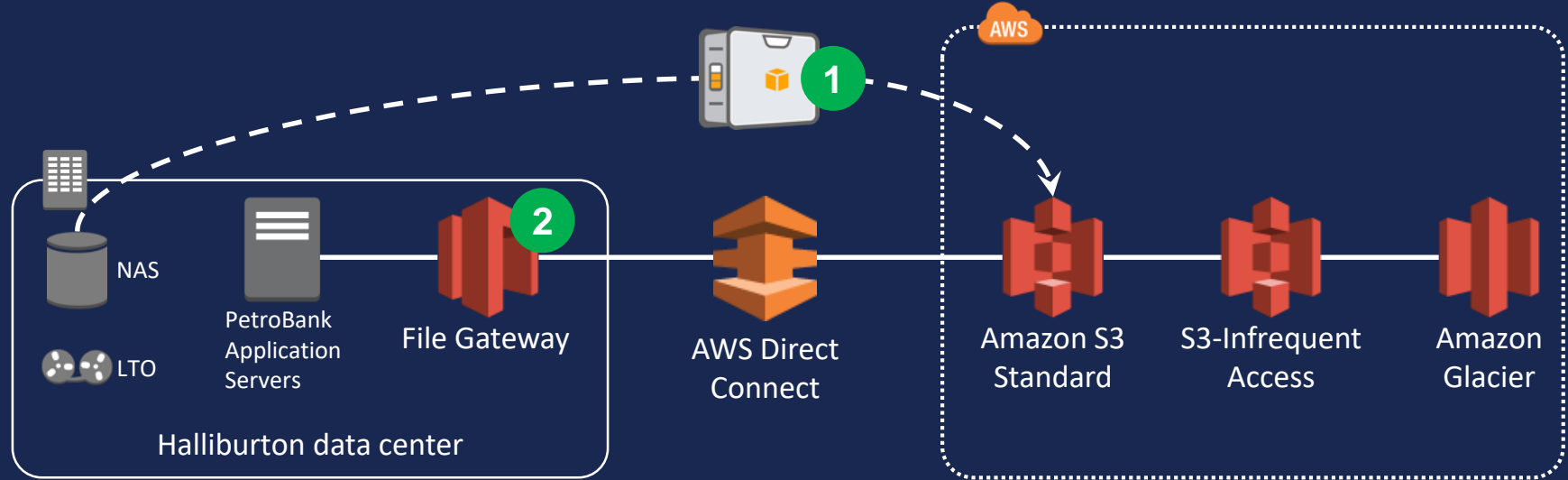|	Name ↑≡
☐	📁 SQL_001
☐	📁 SQL_002
☐	📁 SQL_003_REPT
☐	📁 SQL_004-DEV
☐	📁 SQL_005_REPL
☐	📁 SQL_006_QA
☐	📁 SQL_007_BOND
☐	📁 SQL_008_REPT_STAGE

Hybrid file use case: Active archive migration

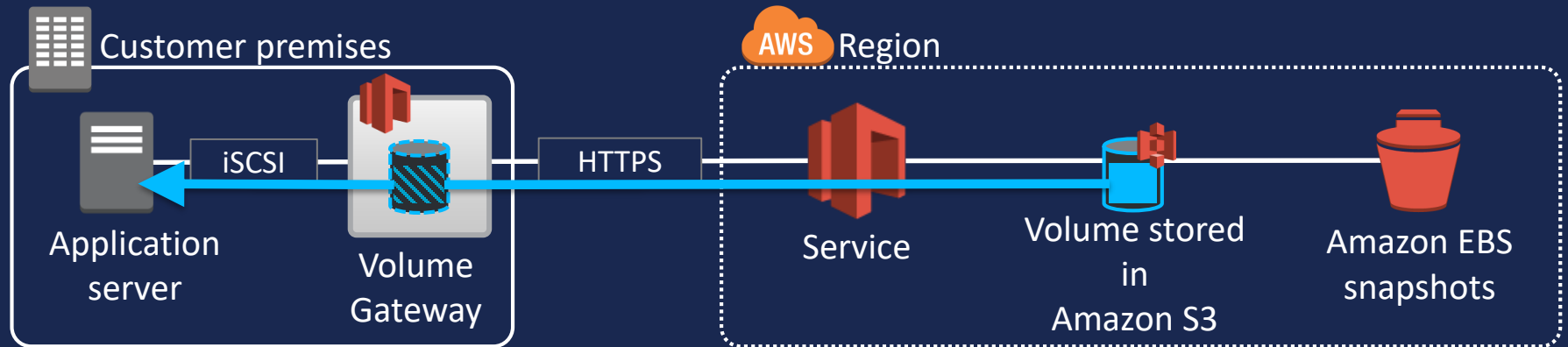Cost-effective storage in AWS with local data access

HALLIBURTON
Landmark

1 Use Snowball to ship data from on-premises offline archives

2 Online access to all data through gateway
Minimal on-premises storage reduces cost
Time-to-data reduced by days or weeks

aws

Hybrid volumes for recovery & migration with Volume Gateway and Amazon EBS snapshots
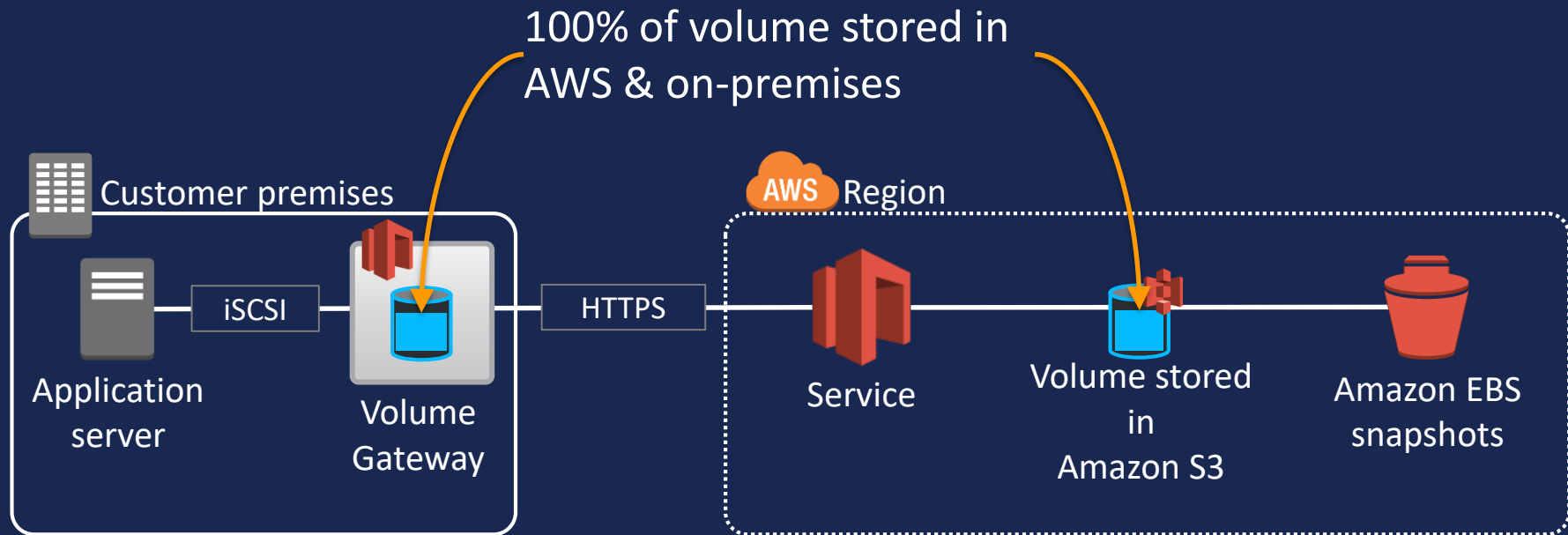
aws

Volume Gateway

Cloud-backed block storage presented on-premises

- Tier snapshots or whole volumes to the cloud to reduce SAN/NAS mgt.

- Flexible recoveries in-cloud or on-premises with snapshots and clones

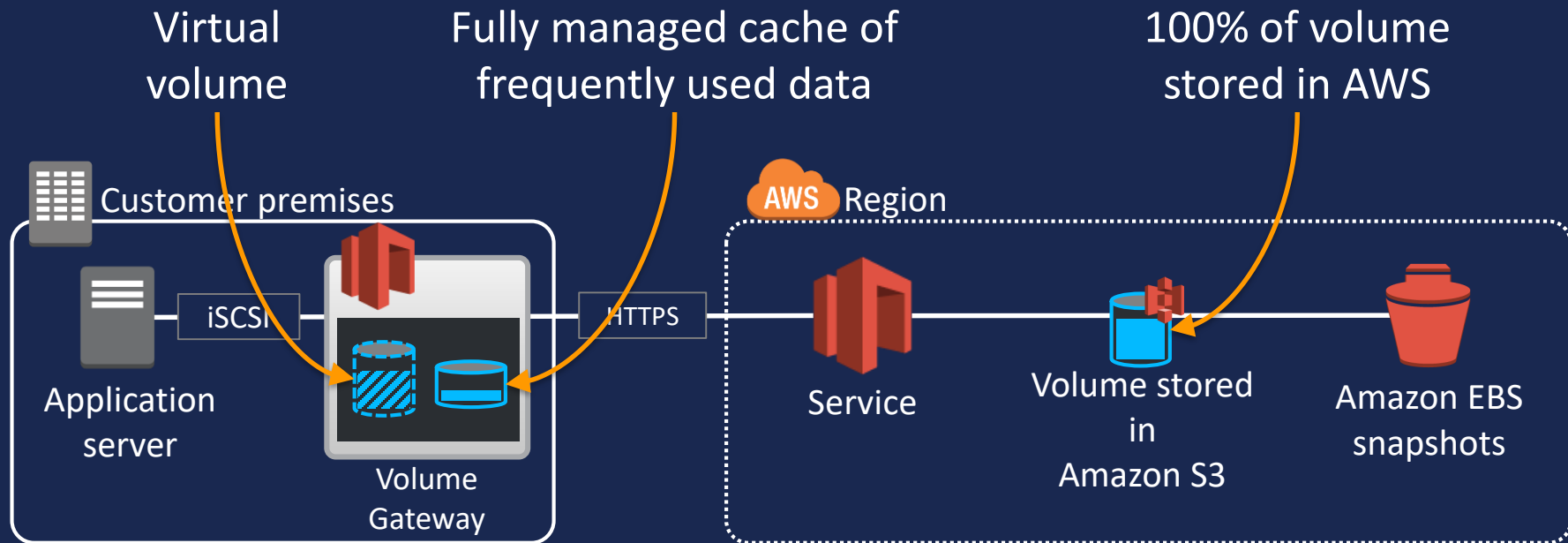- Common uses: backup and restore, disaster recovery (DR), data migrations

Customer premises

iSCSI

HTTPS

Application server

Volume Gateway

AWS Region

Service

Volume stored in Amazon S3

Amazon EBS snapshots

aws

Volume Gateway: Stored mode

Low-latency access to all your data with point-in-time backup to the cloud through Amazon EBS snapshots

100% of volume stored in AWS & on-premises

Customer premises

Application server

iSCSI

Volume Gateway

HTTPS

AWS Region

Service

Volume stored in Amazon S3

Amazon EBS snapshots

aws

Volume Gateway: Cached mode

Reduce on-premises storage, caching frequently used data local to your application, with 100% of your data in the cloud

Virtual volume

Fully managed cache of frequently used data

100% of volume stored in AWS

Customer premises

AWS Region

iSCSI

HTTPS

Application server

Volume Gateway

Service

Volume stored in Amazon S3

Amazon EBS snapshots

aws

Volume Gateway: Snapshots and clones for data protection

EBS snapshots

Point-in-time backups of a stored or caches volumes

Created on-demand or on a configurable schedule

Restore either as an Amazon EBS or a Storage Gateway volume

Volume clones

Instant real-time copy of a cached volume

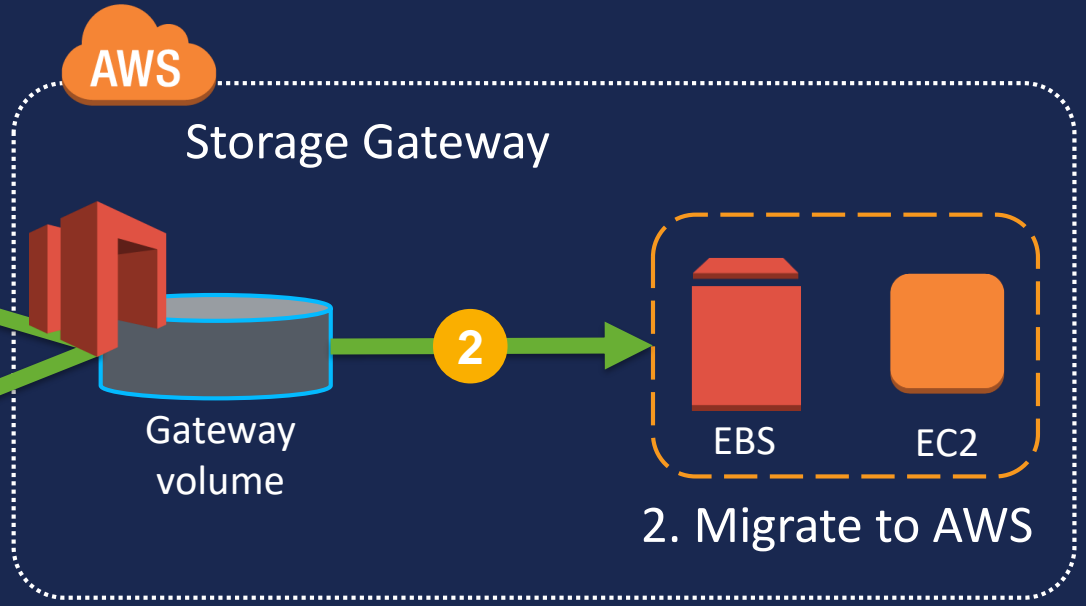Represents current state of volume stored in AWS

Restore as a Storage Gateway volume

aws

Flexible data backup, recovery & migration use cases

1. Restore to your data center

AWS

Storage Gateway

Gateway volume

EBS   EC2

2. Migrate to AWS

1

2

3

3. Recover to a Disaster Recovery (DR) site

aws

Hybrid use case: Protect & recover volumes in-cloud

STEMCELL™
TECHNOLOGIES

Canada's largest biotech firm

"**It made no sense to keep buying big disk siloes, especially as we opened up new global offices, and now we can recover in the cloud from a snapshot if we ever had to.**"

**Adam Leggett
IT manager**

- Data sovereignty required local hot files & tape archives in 10 global offices
- Volume Gateway eliminated 50-hour backup windows and tape archive systems
- Cut on-premises; storage capex 40%; reduced RTO from 48 hours to 10 minutes
- Meets cloud strategy while retaining local ownership and data sovereignty
- Enabled data center exit in next 6–12 months

aws

Migrate tape backup workflows to AWS with Tape Gateway

Tape Gateway: Drop-in replacement for tape backups

- Emulates a tape library; virtual tapes on Amazon S3 and Amazon Glacier
- Works with common backup apps, to support existing backup workflows
- Low-cost: Predictable costs and reduced management
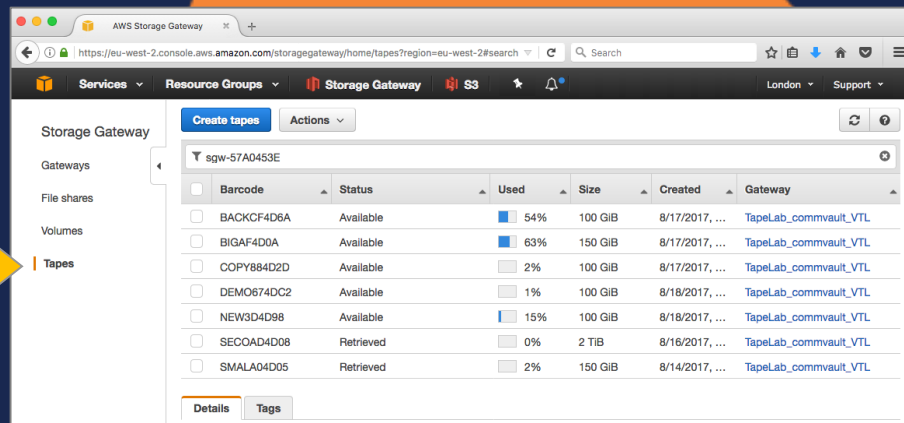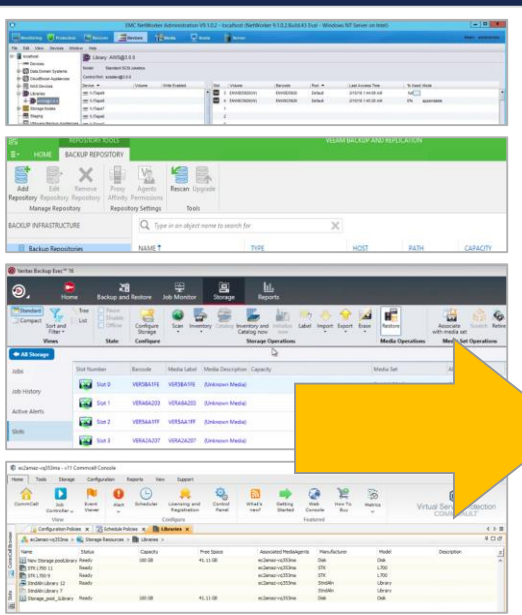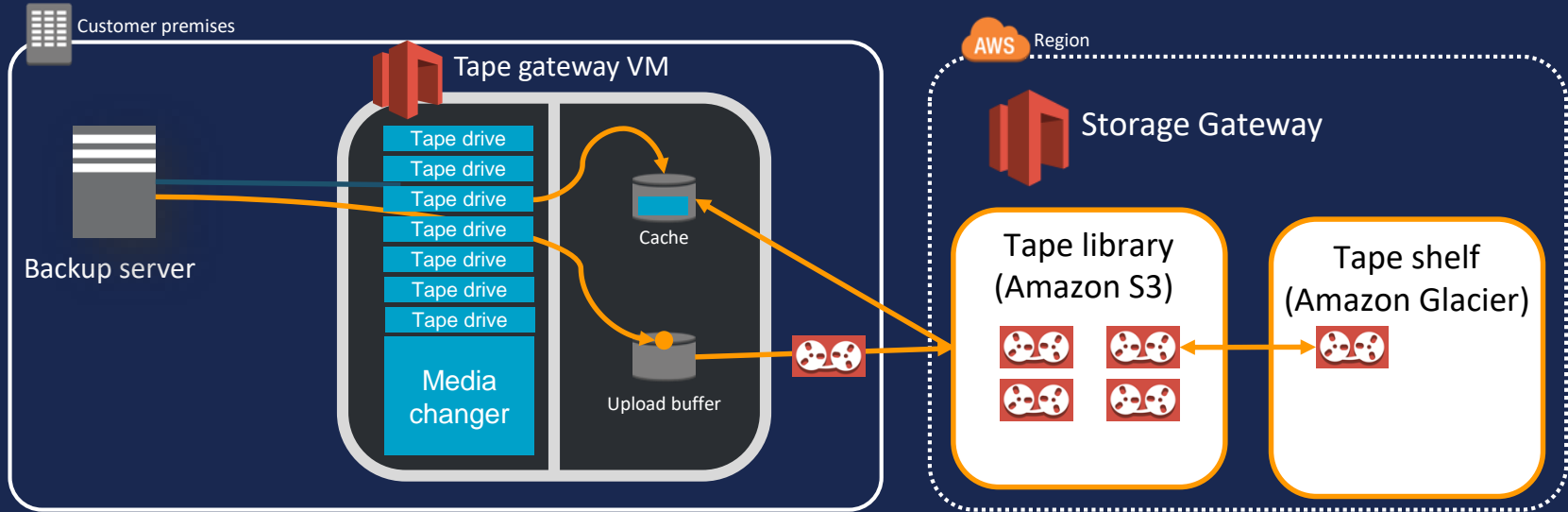
Tape gateway: Cost comparison, a customer example

Every month, 100 TB written, and 1 tape retrieval; 1-year retention

Tape gateway
$34.5 K / year

On-premises tape
$56.8K / year

Admin time

Data transfer out and tape retrieval

Data written via gateway

Virtual tapes archived in Amazon Glacier

Tape library maintenance

Admin time

Robotic tape library

Tape media costs

Offsite archive service

aws
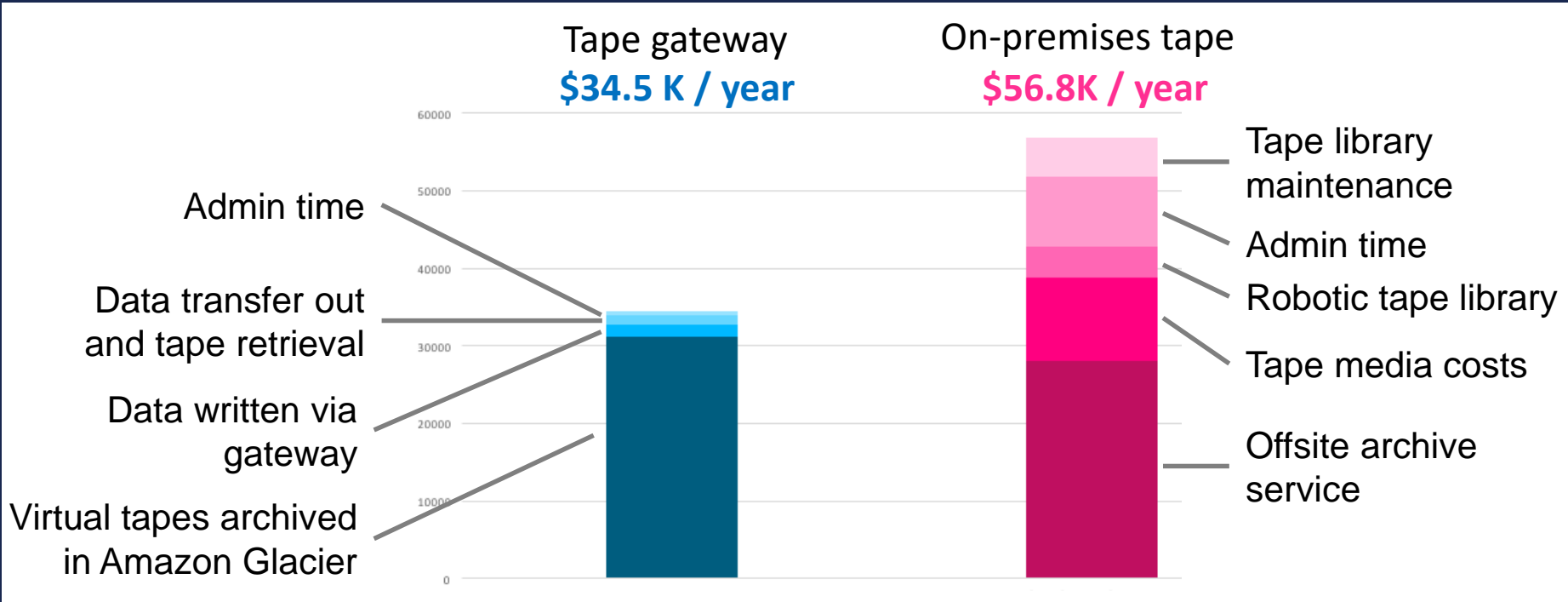
Tape gateway: How the VTL works

- Emulates a physical tape device with a media changer and tape drives
- Scalable: Virtually unlimited tape storage in AWS
- Virtual tapes written to Amazon S3: Data is in S3 when tape is in virtual library drive or slot
- Ejected virtual tapes are marked "read only" and moved to tape shelf on Amazon Glacier
- Recovery: Retrieve tapes to library (3–5 hours) and read data to same or different gateway

Backup, archive, and disaster recovery

Cost effective storage in AWS with local or cloud restore

"Using AWS Storage Gateway,
we switched from physical to virtual tape backup
simply by dropping the gateway's virtual appliance into
our existing Veeam workflow.

Setting it all up took three hours, at most.
We can now provision virtual tapes on AWS
with the click of a button."

Summary: AWS Storage Gateway enables you to …

Reduce on-premises storage infrastructure with AWS

Move file, block, and tape storage to the cloud

Enable a range of backup and hybrid workloads

aws

Questions?

Thank you.

aws