Session 4 -AWS Storage

AWS Certified Solutions Architect – Associate

Albeda Siddique, AWS Solutions Architect

Aug 13, 2021

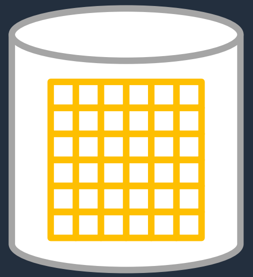
Agenda

- Introduction
- Storage Primer
- Block Storage
- Object Store
- Shared File Systems
- On-Premises Storage Integration
- CloudFormation

0

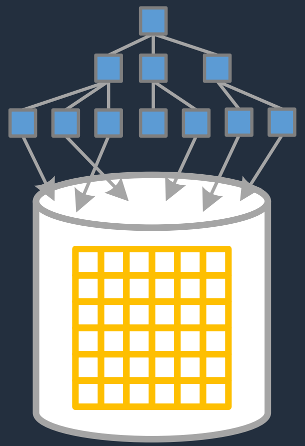
Storage Primer

Block vs File vs Object



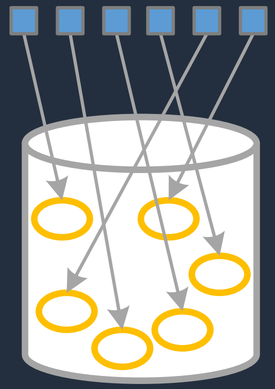
Block Storage

Raw Storage
Data organized as an array of unrelated blocks
Communication via dedicated Storage Area Network (SAN)
Ex: Hard Disks



File Storage

Data is stored in file
Every file is arranged in logical hierarchy (eg, folder, sub-folder etc).
Uses Network Attached Storage (NAS)
Ex: Windows File Servers



Object Storage

Data is stored in the form of object, which also includes metadata
Uses API Access to data
Policy-based, etc.
Ex: S3

AWS Storage Offerings

OBJECT

Amazon
Simple Storage Service
(s3)

(pictures, video, static website
more read, less write)

BLOCK

Amazon
Elastic Block Store
(EBS)

(Database – fast read & write)

FILE

Amazon
Elastic File
System (EFS)

Amazon FSx for
Windows File Server

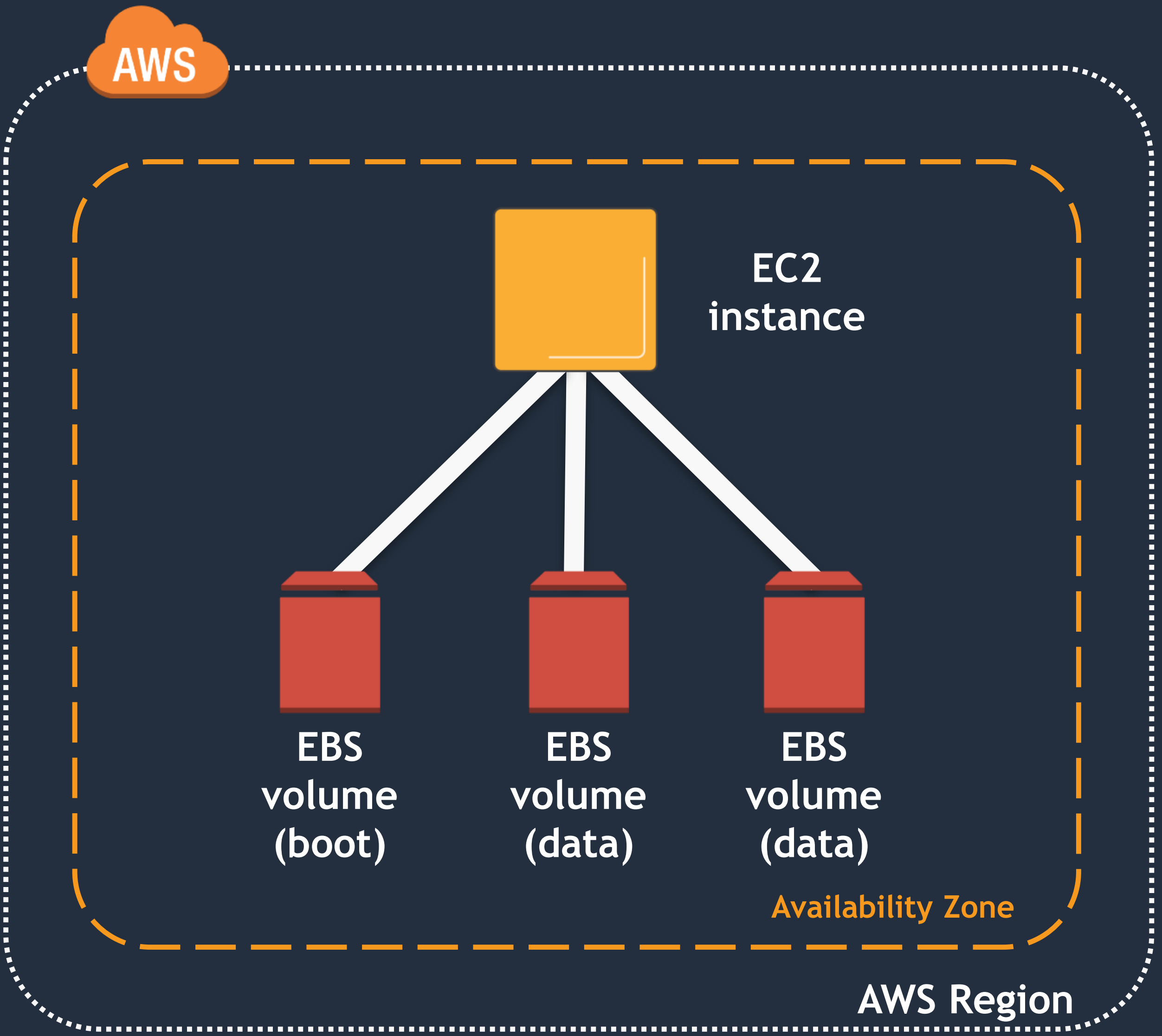
Amazon FSx
for Lustre

(File sharing – across instances)

①

Block Storage

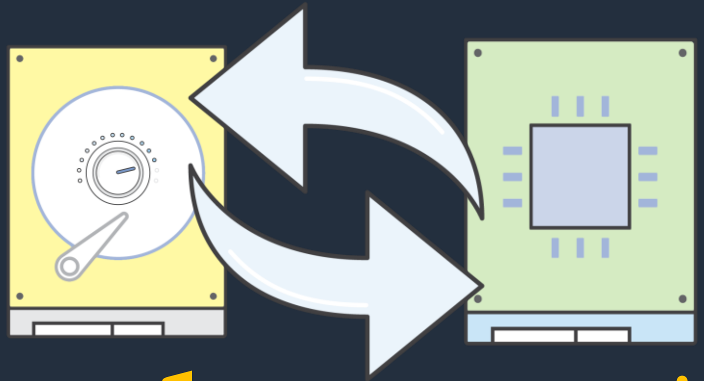
What is Amazon EBS?



- Block storage as a service (persistent)
- Volumes attach to an instance, in same AZ and replicated in its own AZ to protect component failure
- Multi-Attach is supported for SSD (io1-certain region and io2- all regions)
- Many volumes can attach to an instance
- Separate boot and data volumes
  - Can be detached from an instance and attached to a different one
- Delete On Termination – Default for Root Volume
- Volumes can be encrypted
  - During EC2 instance launch
  - From Uncrypted to Encrypted (using snapshot)

Elastic Volumes: Features

Increase volume size



sc1 magnetic

Change volume type

Increase/decrease provisioned IOPS

EBS volume types

SSD		HDD	
General-purpose (gp2/gp3) (provisioned iops < 16000)	Provisioned IOPS (io1/io2) (provisioned iops > 16000)	Throughput-optimized HDD(st1)	Cold HDD (sc1)
NoSQL databases	**Relational databases**	**Big data, analytics**	**File, media**
Transactional workloads, low-latency applications	I/O-intensive database applications	Large datasets and large I/O sizes	Less frequently accessed workloads with large, cold datasets
Cassandra, MongoDB, CouchDB	MySQL, SQL Server, PostgreSQL, SAP, Oracle	Kafka, Splunk, Hadoop, data warehousing	Transcoding, encoding, rendering

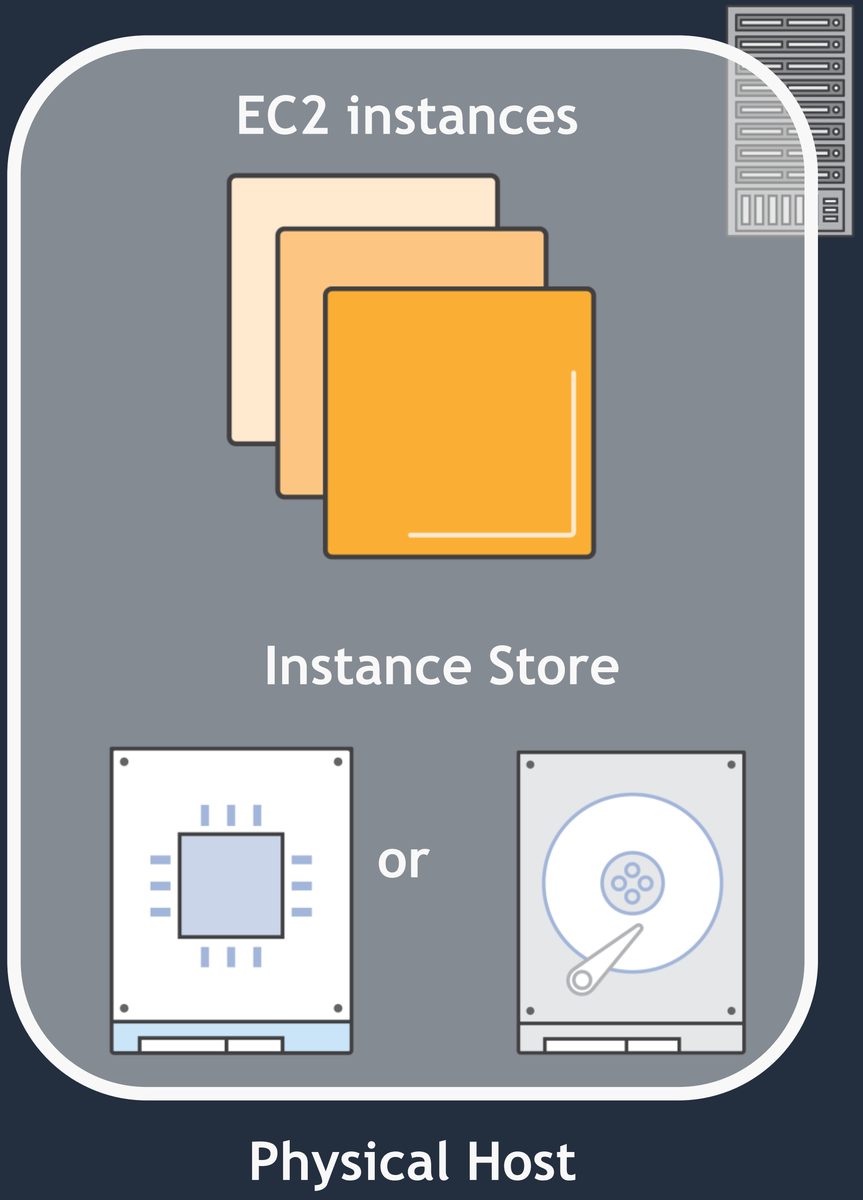
EBS Snapshots

- Point-in-time snapshots of volume blocks
- Stored in Amazon S3 and accessed via EBS APIs
- Subsequent snapshots are incremental
- Deleting snapshot only removes data exclusive to that snapshot
- Move EC2 and EBS to another AZ
    - Create a snapshot of EBS
    - Create an AMI image of that snapshot
    - Launch that AMI in different AZ
- Snapshot of encrypted volume will be encrypted automatically

Why use EBS Snapshots

- Replicate volumes across Availability Zones
- Copy to another region for Disaster Recovery
- Backup critical data
- Capture production data for test/dev
- Create machine images (AMIs)
- Copy and share EBS volumes

What is Amazon EC2 instance store?



- Local to instance
- Non-persistent data store
- Only specify during instance launch
- Data is lost
  - When disk fails
  - Instance stops
  - Instance terminates
- SSD or HDD

Poll 1

Which condition must be met to attach an EBS volume to an EC2 instance ? (SELECT 1)

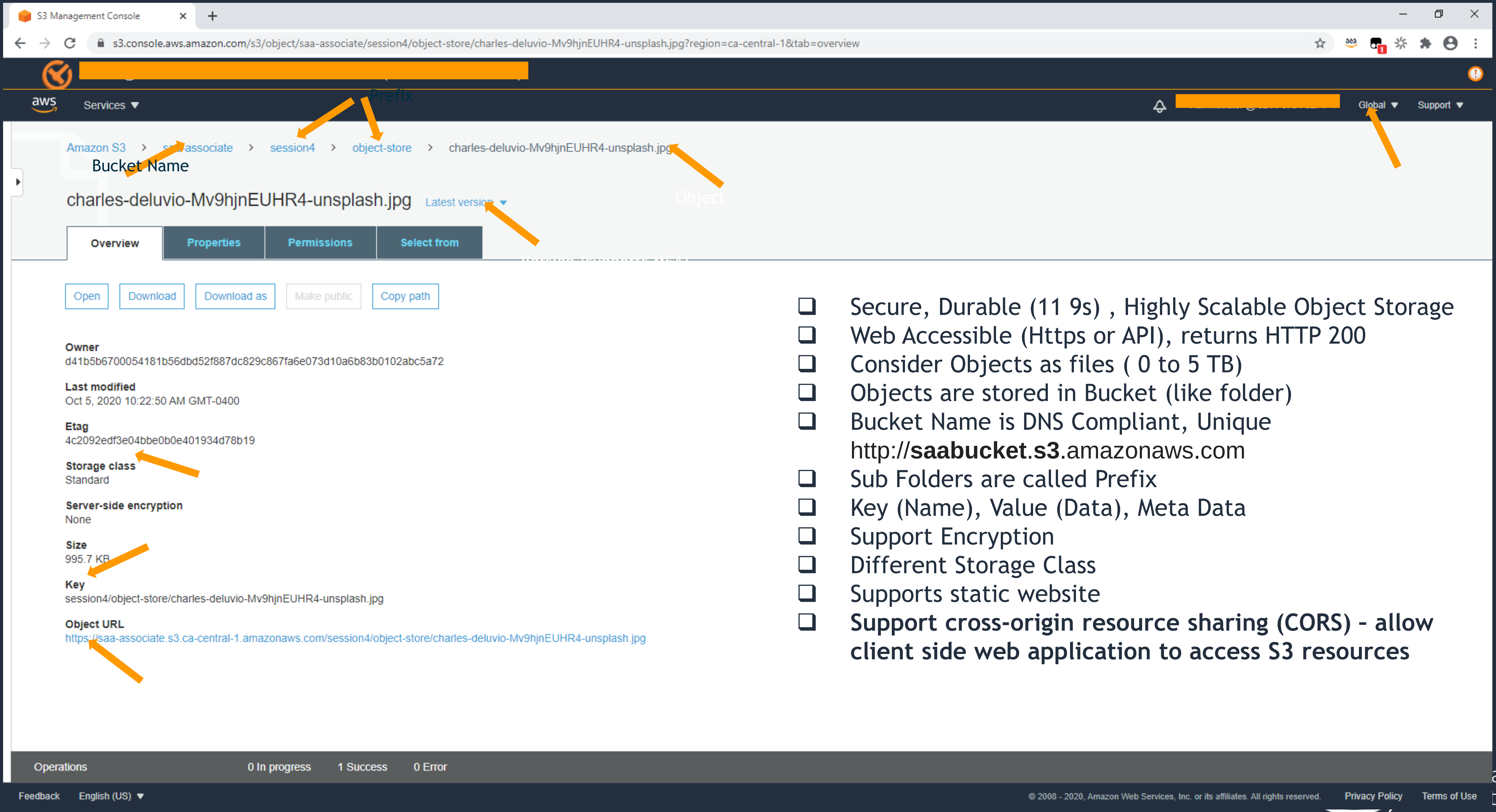
Poll 2

Which of the following Block Volume Type is the preferred choice for Relational Database (SELECT 1)?

2

Object Stores

Amazon Simple Storage Service (S3)


S3 Management Console
s3.console.aws.amazon.com/s3/object/saa-associate/session4/object-store/charles-deluvio-Mv9hjnEUHR4-unsplash.jpg?region=ca-central-1&tab=overview
[redacted]
aws
Services ▼
[redacted]
Global ▼
Support ▼
Prefix
Amazon S3 › saa-associate › session4 › object-store › charles-deluvio-Mv9hjnEUHR4-unsplash.jpg
Bucket Name
Object
charles-deluvio-Mv9hjnEUHR4-unsplash.jpg
Latest version ▼
Version (supports MFA)
Overview
Properties
Permissions
Select from
Open
Download
Download as
Make public
Copy path
Operations
0 In progress
1 Success
0 Error
Feedback
English (US) ▼
© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use

Owner
d41b5b6700054181b56dbd52f887dc829c867fa6e073d10a6b83b0102abc5a72

Last modified
Oct 5, 2020 10:22:50 AM GMT-0400

Etag
4c2092edf3e04bbe0b0e401934d78b19

Storage class
Standard

Server-side encryption
None

Size
995.7 KB

Key
session4/object-store/charles-deluvio-Mv9hjnEUHR4-unsplash.jpg

Object URL
https://saa-associate.s3.ca-central-1.amazonaws.com/session4/object-store/charles-deluvio-Mv9hjnEUHR4-unsplash.jpg

- ❑ Secure, Durable (11 9s) , Highly Scalable Object Storage
- ❑ Web Accessible (Https or API), returns HTTP 200
- ❑ Consider Objects as files (0 to 5 TB)
- ❑ Objects are stored in Bucket (like folder)
- ❑ Bucket Name is DNS Compliant, Unique
  http://**saabucket.s3**.amazonaws.com
- ❑ Sub Folders are called Prefix
- ❑ Key (Name), Value (Data), Meta Data
- ❑ Support Encryption
- ❑ Different Storage Class
- ❑ Supports static website
- ❑ **Support cross-origin resource sharing (CORS) - allow client side web application to access S3 resources**

and
tion

Your choice of object storage classes

Frequent ← *Access Frequency* → *Infrequent*

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 Intelligent-Tiering

- Data with changing access patterns
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

S3 Standard-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 One Zone-IA

- Re-creatable, less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier Deep Archive

- Archive data
- Select 12 or 48 hours
- ≥ 3 AZ
- $0.00099/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

Bucket Versioning helps protect your data

With Bucket Versioning,

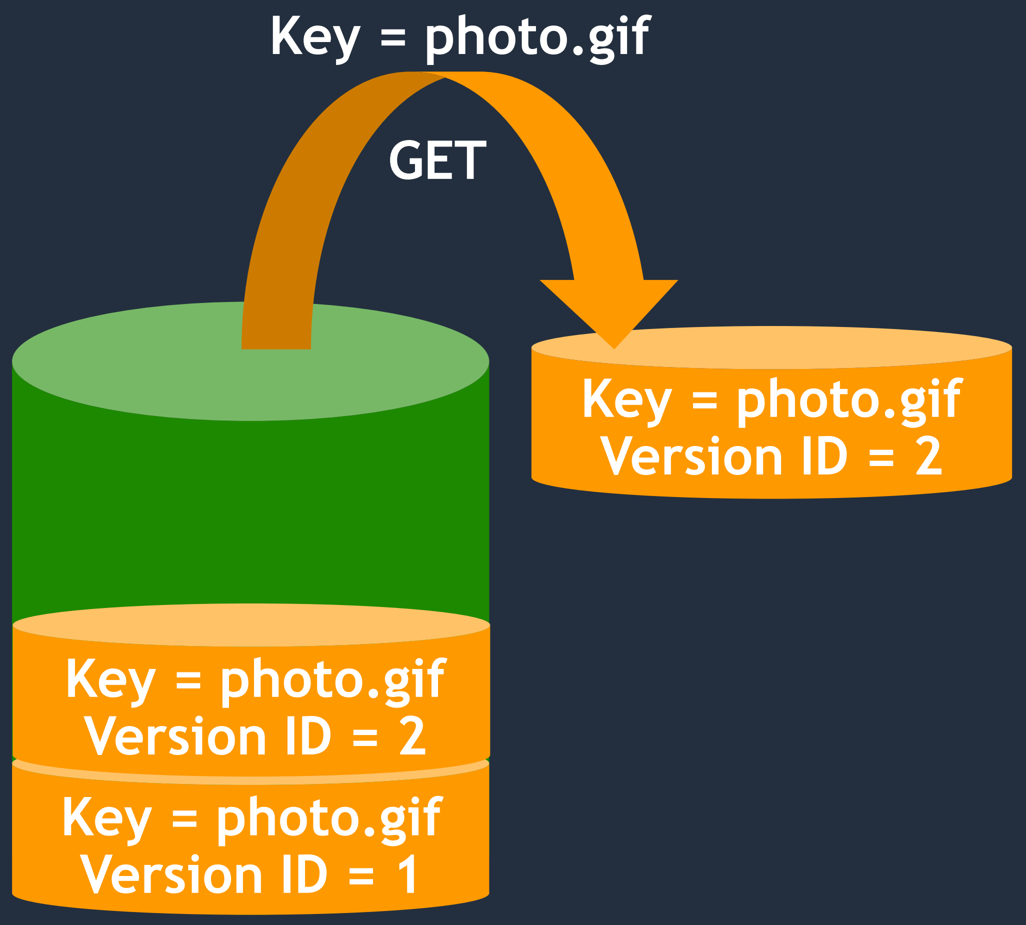
- Create a new version with every upload
- Previous versions are retained
- Once enabled, you can't delete it. You can only suspend it.
- You can also enable MFA on a version delete

Manage versions:

  - Lifecycle expirations

Versioning required for:

  - Cross-Region Replication
  - S3 Object Lock



Automate data management
Lifecycle policies

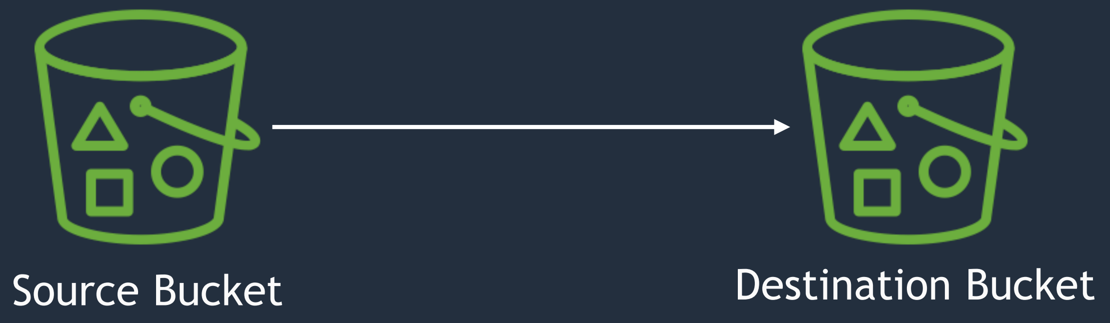
Lifecycle policies

- Automatic tiering and cost controls
- Includes two possible actions:
  - Transition: Standard to Standard - IA or Amazon Glacier based on object age you specified
  - Expiration: deletes objects after specified time

- Set policies by bucket, prefix, or tags
- Set policies for current version or non-current versions

Replication with ownership override helps create backups and set up DR plans

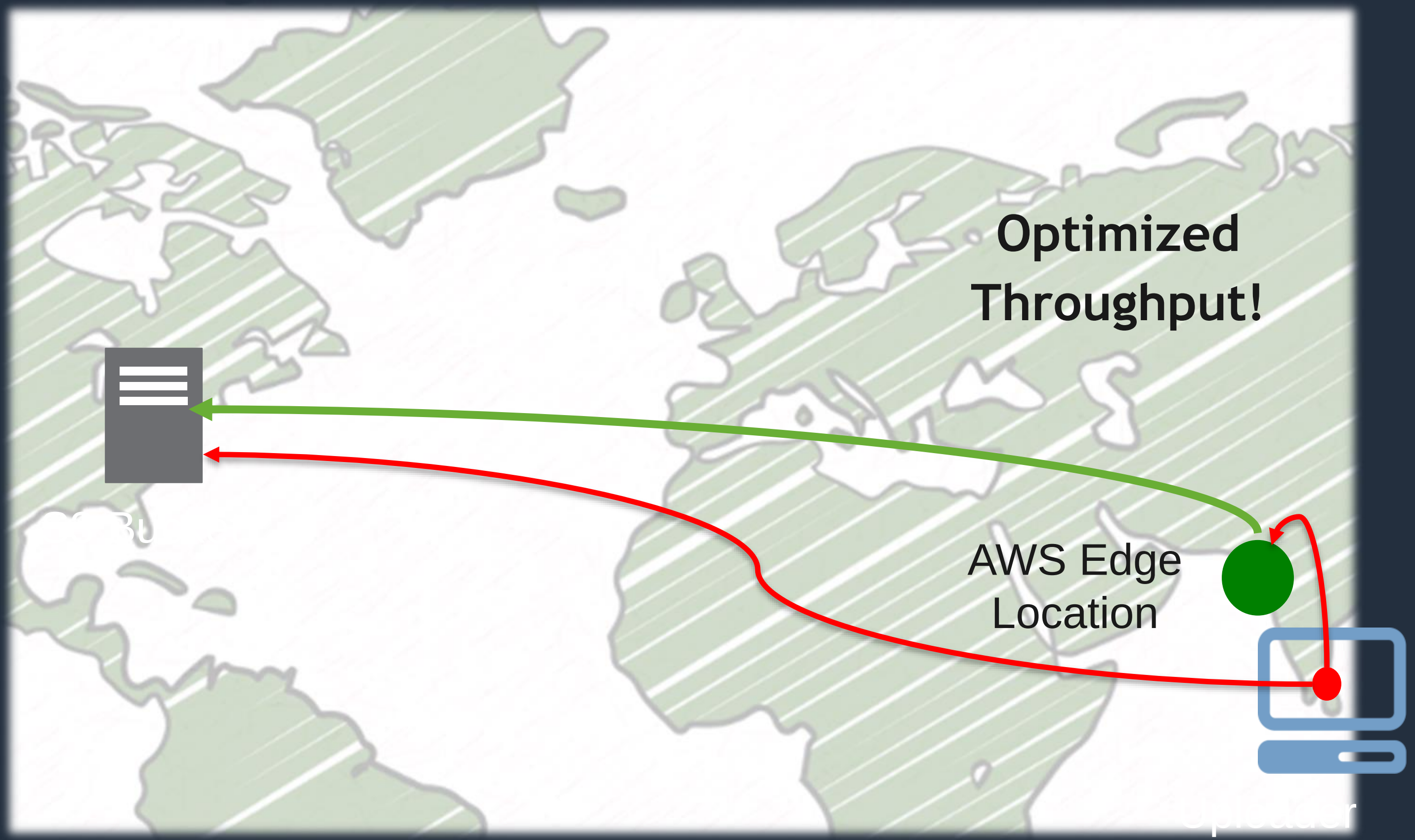
Amazon S3 Replication automatically copies your data to the same or different AWS region



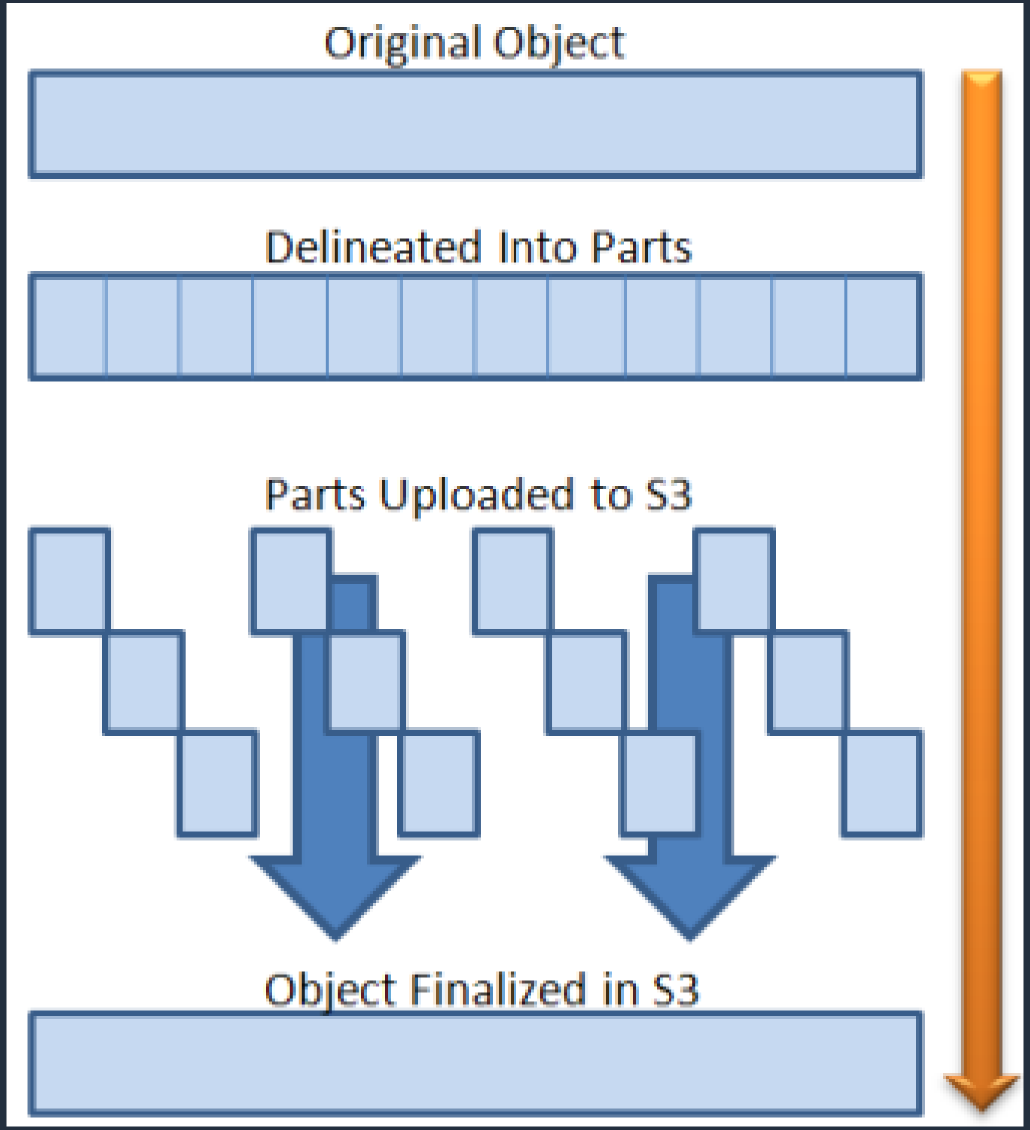
- Filter by prefix, object tag or a combination of both
- Replicate to any S3 storage class including S3 Glacier
- Change ownership of replica objects using the ownership override feature
- New feature – supports for multiple destination buckets

Faster upload over long distances S3 Transfer Acceleration

- Change your endpoint: saa-bucket.s3-accelerate.amazonaws.com
- No Changes to your code
- No firewall changes or client software
- Longer distance, larger files, more benefit
- AWS global edge locations
- Try it at S3speedtest.com



Faster upload of large objects
Parallelize PUTs with multipart uploads

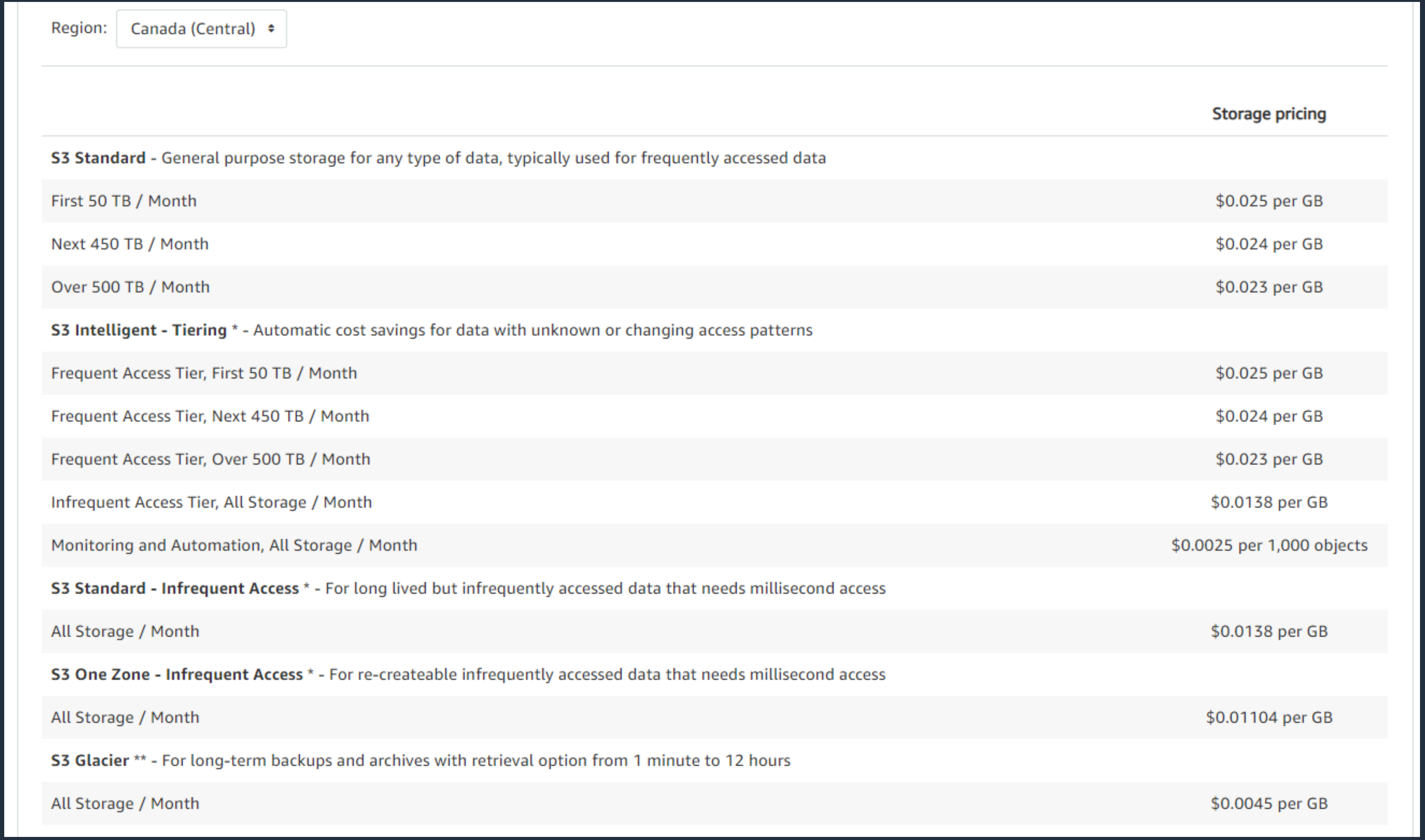


- Recommended for files over 500 MB, Required for files over 5 GB
- Increase aggregate throughput by parallelizing PUTs on high-bandwidth networks
- Increase resiliency to network errors; fewer large restarts on error-prone networks

Best Practice

Amazon S3 Pricing

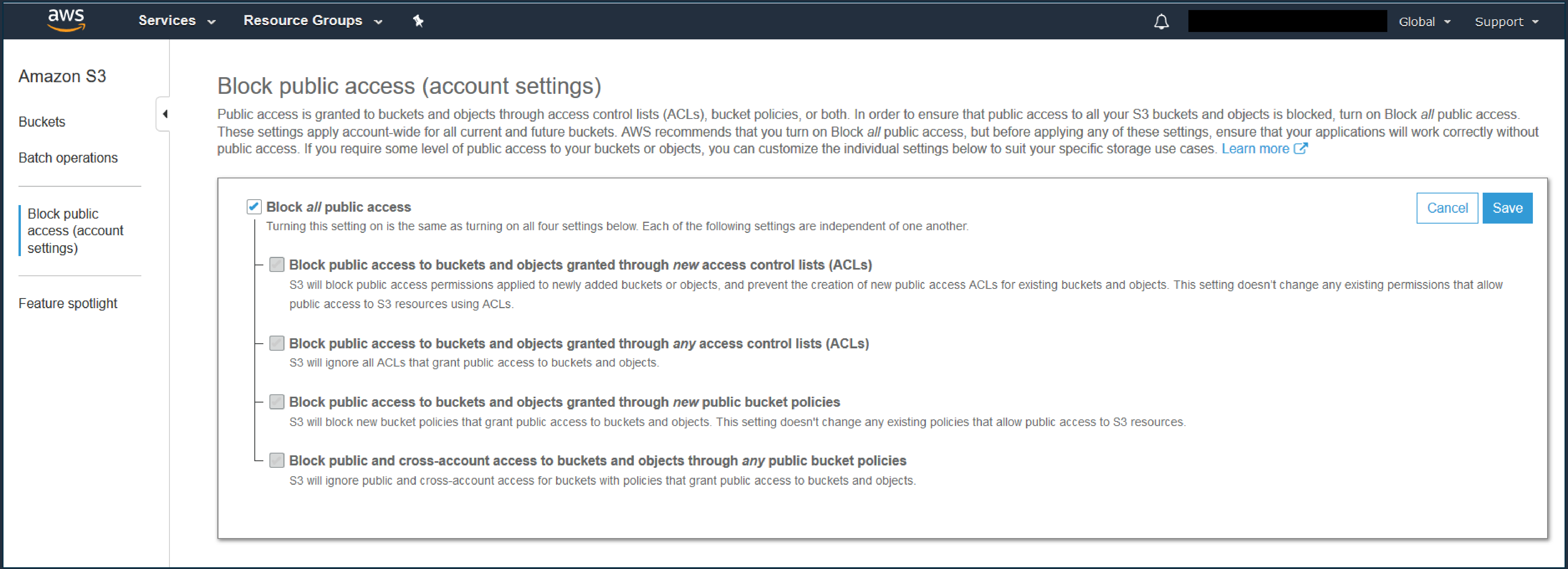
- Storage
- Number of Requests
- Data Transfer Pricing (no charge for inbound, only outbound)
- Storage Management Pricing
- Transfer Acceleration
- Cross Region Replication



Region: Canada (Central)

	Storage pricing
S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data	
First 50 TB / Month	$0.025 per GB
Next 450 TB / Month	$0.024 per GB
Over 500 TB / Month	$0.023 per GB
S3 Intelligent - Tiering * - Automatic cost savings for data with unknown or changing access patterns	
Frequent Access Tier, First 50 TB / Month	$0.025 per GB
Frequent Access Tier, Next 450 TB / Month	$0.024 per GB
Frequent Access Tier, Over 500 TB / Month	$0.023 per GB
Infrequent Access Tier, All Storage / Month	$0.0138 per GB
Monitoring and Automation, All Storage / Month	$0.0025 per 1,000 objects
S3 Standard - Infrequent Access * - For long lived but infrequently accessed data that needs millisecond access	
All Storage / Month	$0.0138 per GB
S3 One Zone - Infrequent Access * - For re-createable infrequently accessed data that needs millisecond access	
All Storage / Month	$0.01104 per GB
S3 Glacier ** - For long-term backups and archives with retrieval option from 1 minute to 12 hours	
All Storage / Month	$0.0045 per GB

Block public access: Console view



Block public access (account settings)

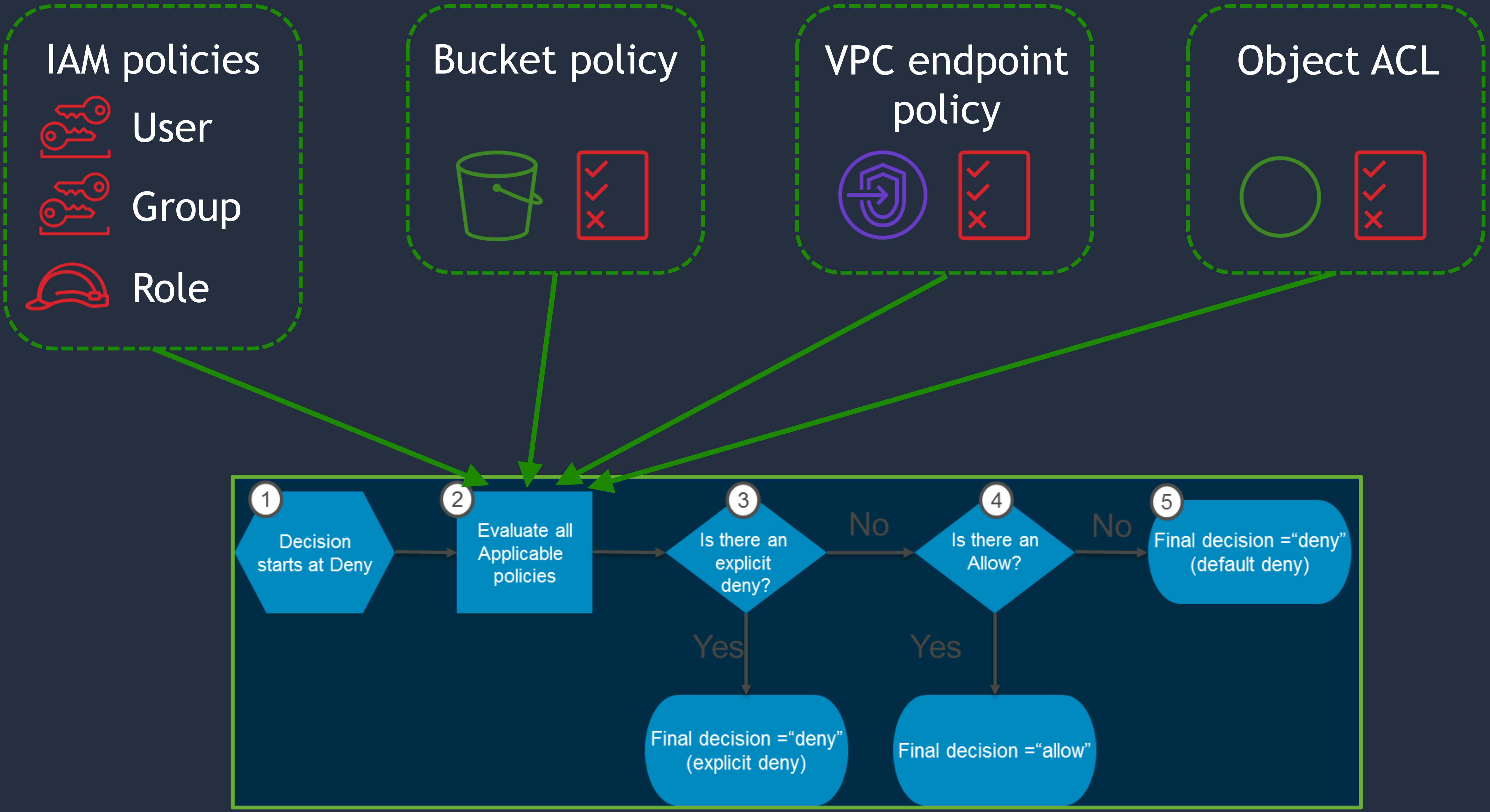
Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, or both. In order to ensure that public access to all your S3 buckets and objects is blocked, turn on Block *all* public access. These settings apply account-wide for all current and future buckets. AWS recommends that you turn on Block *all* public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to your buckets or objects, you can customize the individual settings below to suit your specific storage use cases. Learn more

☑ **Block *all* public access**

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

- ☐ **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
  S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
- ☐ **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
  S3 will ignore all ACLs that grant public access to buckets and objects.
- ☐ **Block public access to buckets and objects granted through *new* public bucket policies**
  S3 will block new bucket policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
- ☐ **Block public and cross-account access to buckets and objects through *any* public bucket policies**
  S3 will ignore public and cross-account access for buckets with policies that grant public access to buckets and objects.

Amazon S3: Access policy processing



IAM user policy

User policy allows this particular user to PUT and GET objects into the reinventbucket

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
       "Effect":"Allow-write-and-read",
         "Action":[
            "s3:PutObject",
          "s3:GetObject",
         ],

"Resource":"arn:aws:s3:::reinventbucket/*"
      }
   ]
}
```

Amazon S3 bucket policy

Bucket policy allows principal from AWS Account 1111111111 to read objects from reinventbucket, but condition limits it to objects that have a specific Tag value

```
{
    "Version": "2012-10-17",
    "Id": "123",
    "Statement": [
{
     "Sid": "Allowing Read Permission",
     "Effect": "Allow",
     "Principal": {"AWS":"1111111111"},
     "Action": ["s3:GetObject"],
     "Resource": ["arn:aws:s3:::
             reinventbucket /*"],
     "Condition": {"StringEquals":
{"s3:ExistingObjectTag/Project": "X"}}
     }
    ]
}
```

Amazon S3 encryption support

Encryption in transit

HTTPS/TLS

Encryption at rest

Server side

- SSE-S3 (Amazon S3 managed keys) – AES 256
- SSE-KMS (AWS Key Management Service)
- SSE-C (customer-provided keys)

Client side

- Encrypt with the AWS Encryption SDK

S3 Object Lock

If you want to …

- Meet regulatory requirements that require you to store immutable data
- Add another layer of protection against object changes and deletion

… use Amazon S3 Object Lock to store store objects using a write-once-read-many (WORM) model on Amazon S3

Amazon S3 object lock modes

Two modes of protection

Compliance mode

- Intended for **compliance**
- Deletes disallowed, even for root account

Governance mode

- Intended for **data protection**
- Enables privileged delete of WORM-protected objects
- Protects against account compromise and rogue actors
- Retention can be changed to compliance mode

Poll 3

Which S3 Object Class provides immediate access to data ? (SELECT 2)?

Poll 4

User wants to keep an Application Logs in S3 for one month and then purge AUTOMATICALLY. What S3 feature will enable it? (SELECT 1)

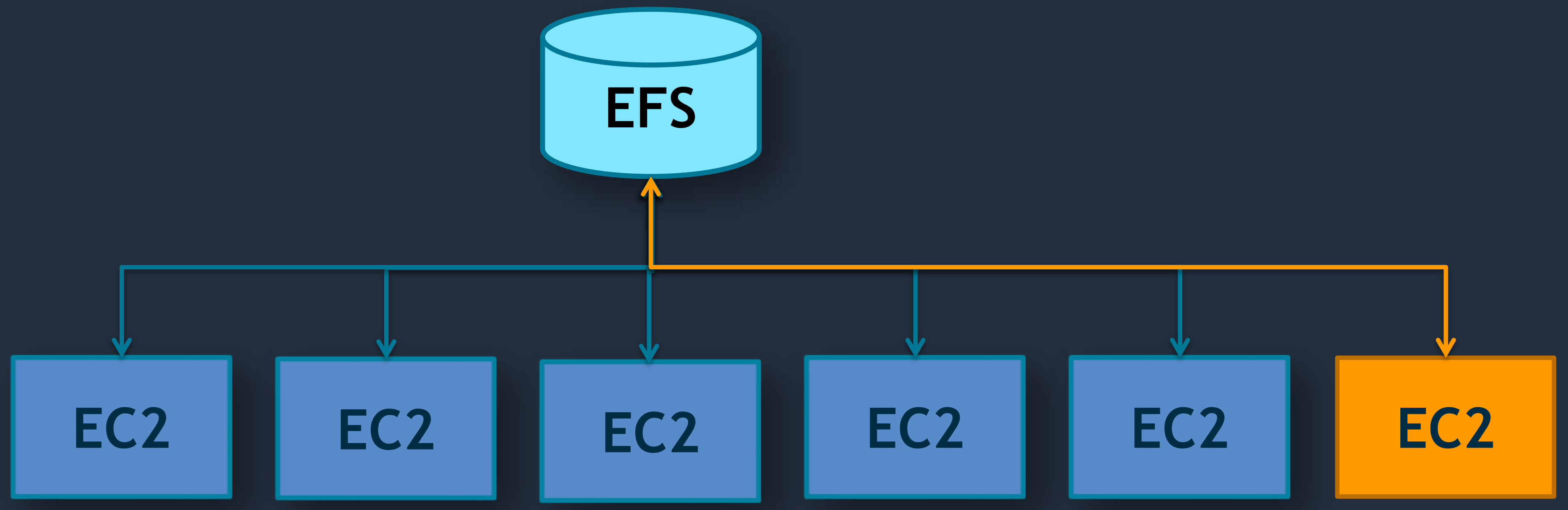
3

Shared file system

Elastic File System (EFS)

- Fully managed NFS file system for EC2 instances
- Sharable across thousands of instances
- Elastically grows to petabyte scale
- Highly available (regional, whereas EBS is AZ based) and durable
- Provides standard file system semantics
- Delivers performance for a wide variety of workloads
- NFS v4–based
- Accessible from on-premise servers

EFS - Mounting



EFS DNS Name

availability-zone.*file-system-id*.efs.*aws-region*.amazonaws.com

Mount on machine

```
sudo mount -t nfs4 mount-target-DNS:/  ~/efs-mount-point
```

EFS Lifecycle Management

- EFS offers both Standard and Infrequent Access (IA) storage classes
- With Lifecycle Management enabled, EFS automatically moves files not accessed for certain days from the Standard storage class to the EFS IA storage class. You can specify one of the following life cycle policies
  - After 7 days
  - After 14 days
  - After 30 days
  - After 60 days
  - After 90 days

Amazon FSx for Windows

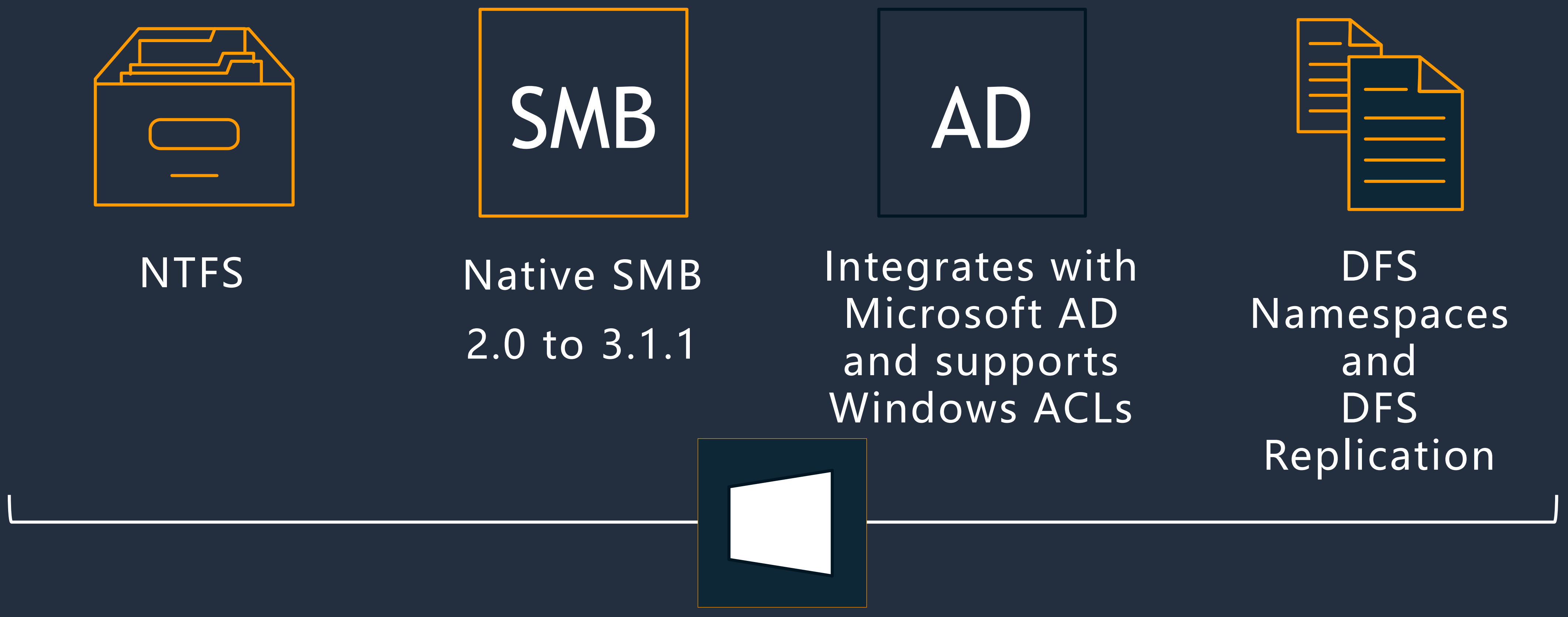
Fully managed Windows file systems …

… built on Windows Server

Integrated with AWS

Native Windows compatibility and features

Native Windows compatibility



NTFS

Native SMB

2.0 to 3.1.1

Integrates with Microsoft AD and supports Windows ACLs

DFS Namespaces and DFS Replication

Windows Server

Amazon FSx for Lustre

Fully managed Lustre file system for compute-intensive workloads

Massively scalable performance

Seamless access to your data repositories

Simple and fully managed

Native file system interface

Cost-optimized for compute-intensive workloads

Secure and compliant

④

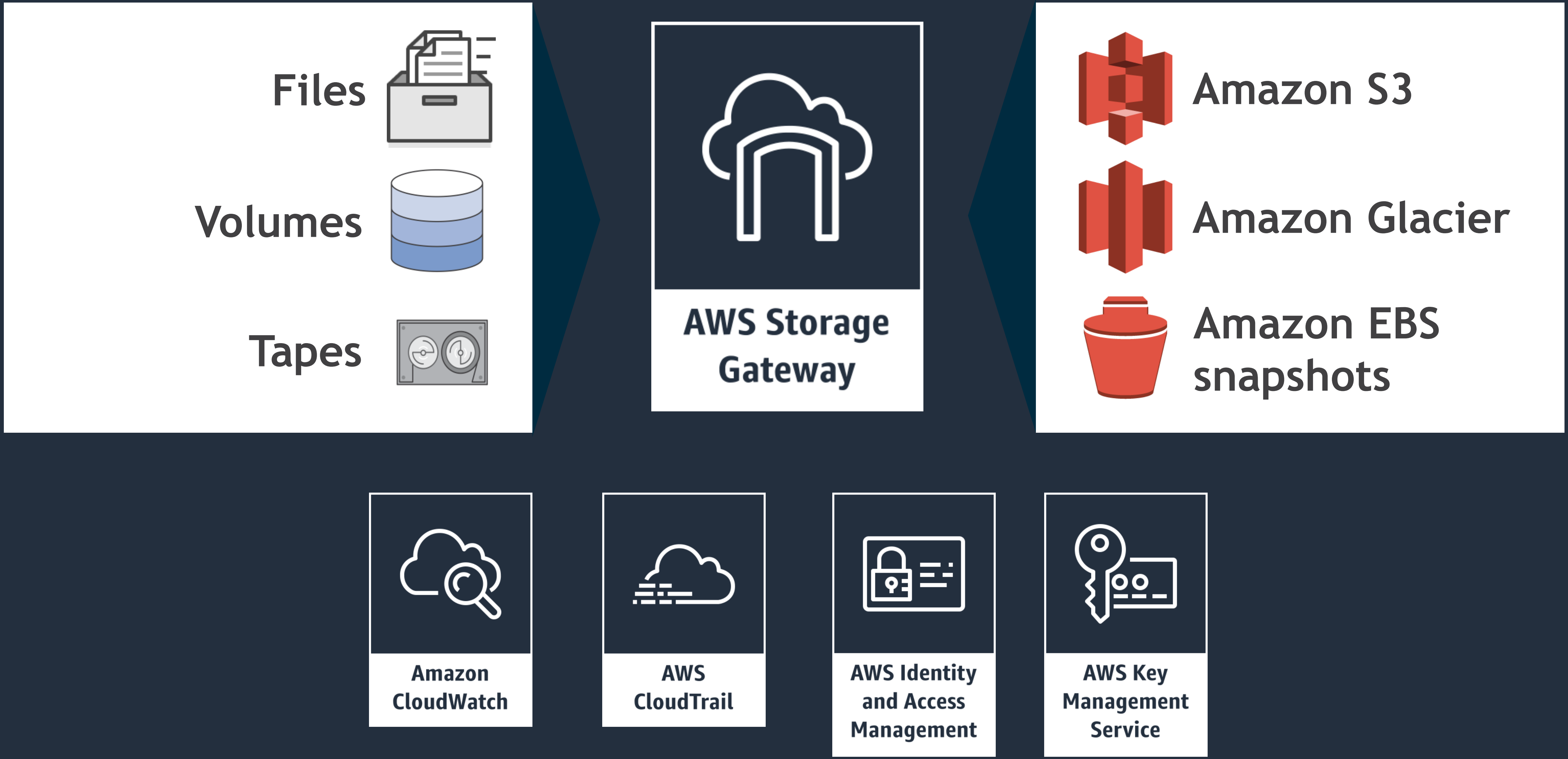
On-Premises Storage Integration

Many Options for Data Transfer

- AWS Direct Connect
- Amazon Kinesis Firehose
- Amazon Kinesis Data Streams
- Amazon Kinesis Video Streams
- Amazon S3 Transfer Acceleration
- AWS Storage Gateway
- AWS Database Migration Service
- AWS Snowball
- AWS Snowball Edge
- AWS Snowmobile
- AWS DataSync
- AWS Transfer for SFTP

Storage Gateway hybrid storage solutions

Enables using standard storage protocols to access AWS storage services



Files

Volumes

Tapes

AWS Storage Gateway

Amazon S3

Amazon Glacier

Amazon EBS snapshots

Amazon CloudWatch

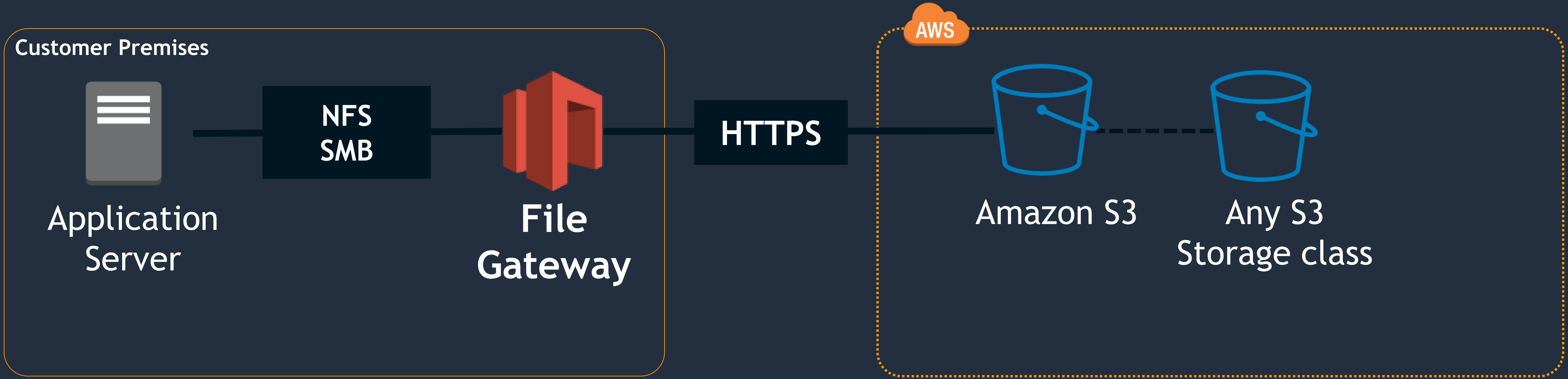
AWS CloudTrail

AWS Identity and Access Management

AWS Key Management Service

File gateway

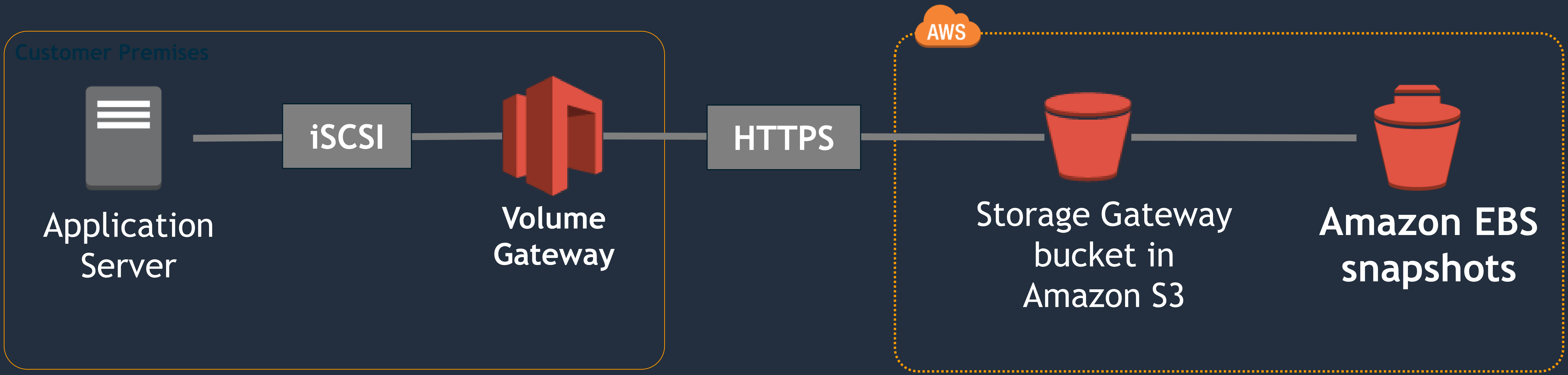
On-premises file storage maintained as objects in Amazon S3



- Data stored and retrieved from your S3 buckets
- One-to-one mapping from files-to-objects
- File metadata stored in object metadata
- Bucket access managed by IAM role you own and manage
- Use S3 Lifecycle Policies, versioning, or CRR to manage data

Volume gateway

On-premises volume storage backed by Amazon S3 with EBS snapshots



Block storage in S3 accessed via the volume gateway

Two types – Cached and Stored Volume

- Cached – S3 as the primary storage and cache (attached disk storage) is used for frequently accessed data (partial).
- Stored Volume – primary data is stored locally and then asynchronously backing upto S3 as EBS snapshots. Low latency to entire dataset

Backup on-premises volumes to EBS snapshots

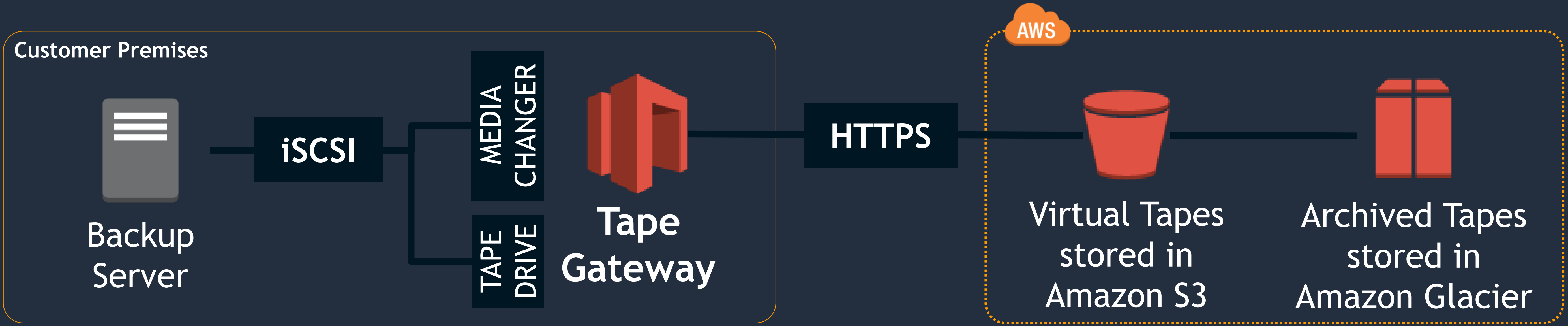
Data compressed in-transit and at-rest

Create on-premises volumes from EBS snapshots

Up to 1PB of total volume storage per gateway

Tape gateway

Virtual tape storage in Amazon S3 and Glacier with VTL management



Virtual tape storage in S3 and Glacier accessed via tape gateway

Data compressed in-transit and at-rest

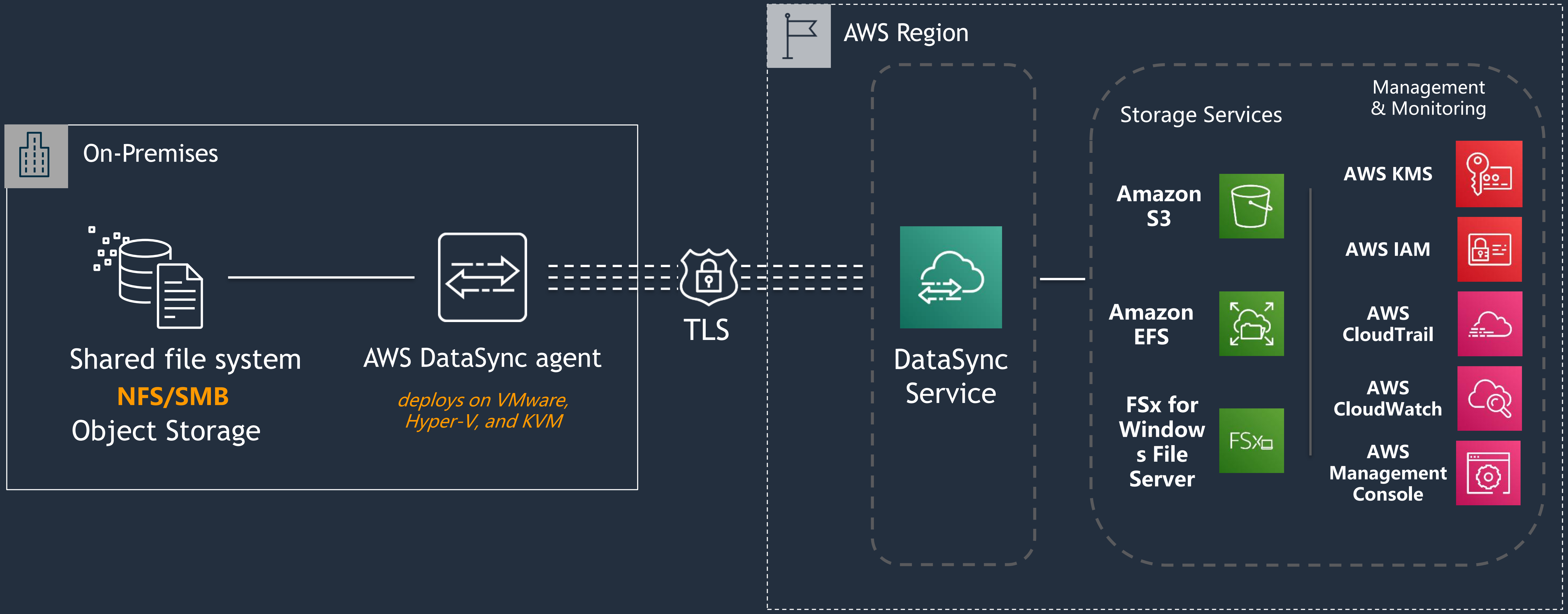
Unlimited virtual tape storage, with up to 1PB of tapes active in library

Supports leading backup applications:

VERITAS DELL EMC veeam Hewlett Packard Enterprise Microsoft System Center Data Protection Manager 2012 arcserve

AWS DataSync

Simplifies, automates, and accelerates data transfer to or from AWS

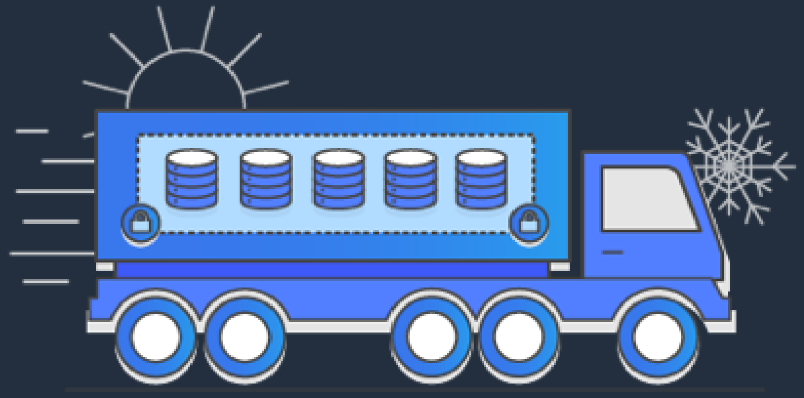


Amazon Snowball & Snowball Edge

- TERAbyte scale data transport
- Uses secure appliances
- Faster than Internet for significant data sets
- Import into S3
- Snowball Edge allows onboard storage and compute
- Snowball edge can be clustered

Amazon Snowmobile

https://www.youtube.com/watch?v=8vQmTZTq7nw



Poll 5

Which is the Correct Storage for an Application that needs NFS file System, shared across many Linux EC2 instances (SELECT 1) ?

Poll 6

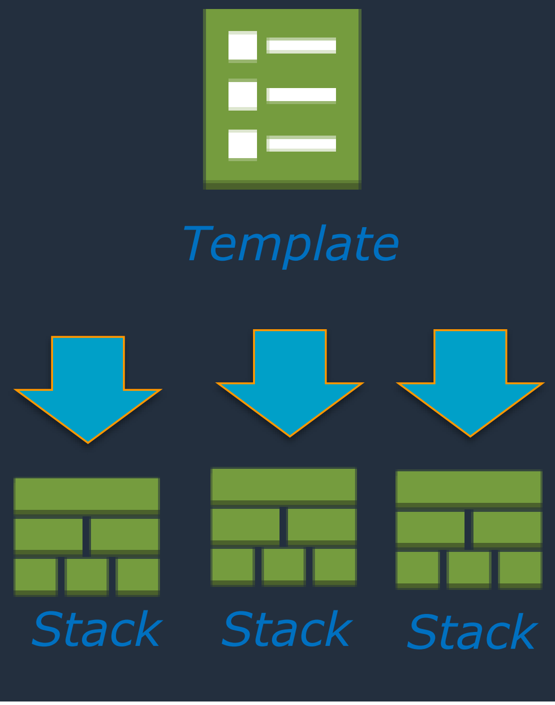
Choose the Preferred AWS Storage option that integrates on-prem infrastructure when low latency to ALL DATA is a must (SELECT 1) ?

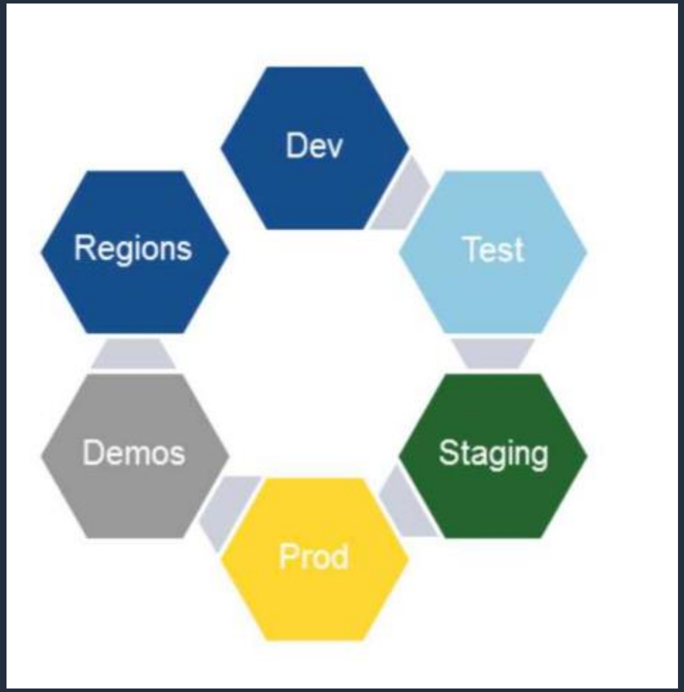
5

AWS CloudFormation

AWS CloudFormation

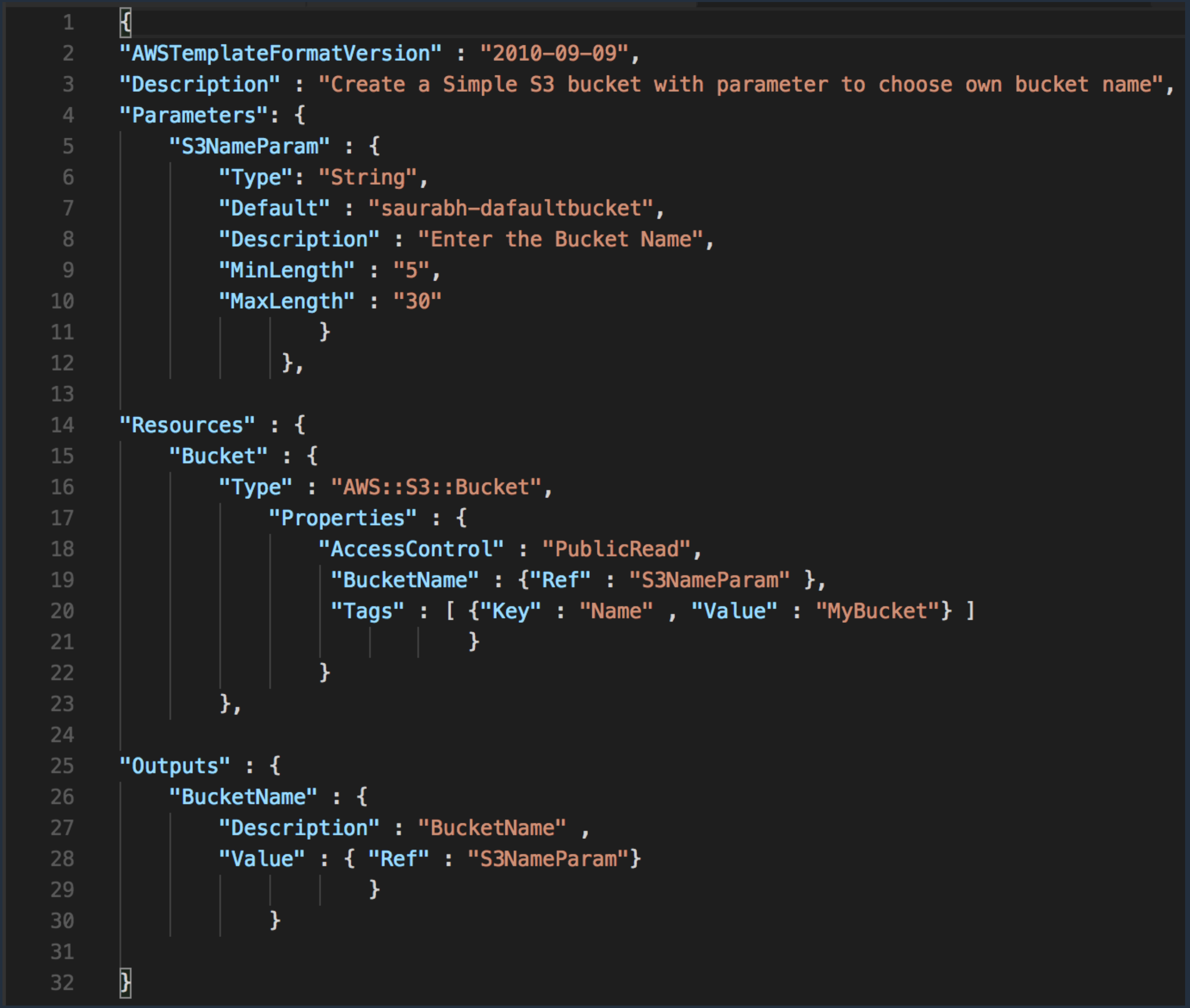
- Simplified way to create and manage a collection of AWS resources (JSON/YAML format template)
- Single source of truth to deploy the whole stack
- Infrastructure that you can replicate, re-deploy, and re-purpose
- Control versioning on your infrastructure and your application together
- Service rolls back to the last good state on failures
- API calls are in parallel, manages dependencies/relationship
- FREE – you only pay for resources





AWS CloudFormation syntax

- JSON – JavaScript object notation
- Attribute-value pairs
- Similar to XML



```
{
"AWSTemplateFormatVersion" : "2010-09-09",
"Description" : "Create a Simple S3 bucket with parameter to choose own bucket name",
"Parameters": {
    "S3NameParam" : {
        "Type": "String",
        "Default" : "saurabh-dafaultbucket",
        "Description" : "Enter the Bucket Name",
        "MinLength" : "5",
        "MaxLength" : "30"
                }
             },

"Resources" : {
    "Bucket" : {
        "Type" : "AWS::S3::Bucket",
            "Properties" : {
                "AccessControl" : "PublicRead",
                 "BucketName" : {"Ref" : "S3NameParam" },
                 "Tags" : [ {"Key" : "Name" , "Value" : "MyBucket"} ]
                            }
                }
        },

"Outputs" : {
    "BucketName" : {
        "Description" : "BucketName" ,
        "Value" : { "Ref" : "S3NameParam"}
                    }
            }

}
```

Additional

- S3 FAQ –
  https://aws.amazon.com/s3/faqs/

- Amazon S3 Storage Class
  https://aws.amazon.com/s3/storage-classes/

- S3 Versioning
  https://docs.aws.amazon.com/AmazonS3/latest/dev/ObjectVersioning.html

Practice Question 7

A company currently stores data for on-premises applications on local drives. The chief technology officer wants to reduce hardware costs by storing the data in Amazon S3 but does not want to make modifications to the applications. To minimize latency, **FREQUENTLY** accessed data should be available locally.

What is a reliable and durable solution for a solutions architect to implement that will reduce the cost of local storage? (SELECT ONE)

A) Deploy an SFTP client on a local server and transfer data to Amazon S3 using AWS Transfer for SFTP.

B) Deploy an AWS Storage Gateway volume gateway configured in cached volume mode.

C) Deploy an AWS DataSync agent on a local server and configure an S3 bucket as the destination.

D) Deploy an AWS Storage Gateway volume gateway configured in stored volume mode.

Practice Question 7

The answer to the question is…

A company currently stores data for on-premises applications on local drives. The chief technology officer wants to reduce hardware costs by storing the data in Amazon S3 but does not want to make modifications to the applications. To minimize latency, frequently accessed data should be available locally.

What is a reliable and durable solution for a solutions architect to implement that will reduce the cost of local storage? (SELECT ONE)

A) Deploy an SFTP client on a local server and transfer data to Amazon S3 using AWS Transfer for SFTP.

B) Deploy an AWS Storage Gateway volume gateway configured in cached volume mode.

C) Deploy an AWS DataSync agent on a local server and configure an S3 bucket as the destination.

D) Deploy an AWS Storage Gateway volume gateway configured in stored volume mode.

Practice Question 8

An analytics company is planning to offer a site analytics service to its users. The service will require that the users’ webpages include a JavaScript script that makes authenticated GET requests to the company’s Amazon S3 bucket.

What must a solutions architect do to ensure that the script will successfully execute? (SELECT ONE)

A) Enable cross-origin resource sharing (CORS) on the S3 bucket.

B) Enable S3 versioning on the S3 bucket.

C) Provide the users with a signed URL for the script.

D) Configure a bucket policy to allow public execute privileges.

Practice Question 8

The answer to the question is…

An analytics company is planning to offer a site analytics service to its users. The service will require that the users’ webpages include a JavaScript script that makes authenticated GET requests to the company’s Amazon S3 bucket.

What must a solutions architect do to ensure that the script will successfully execute? (SELECT ONE)

A) Enable cross-origin resource sharing (CORS) on the S3 bucket.

B) Enable S3 versioning on the S3 bucket.

C) Provide the users with a signed URL for the script.

D) Configure a bucket policy to allow public execute privileges.

Practice Question 9

You are auditing charge of S3 buckets for your company. There are multiple buckets, each is separated based on the type of data it is holding and the level of security required for that data. You are concerned of losing data on several buckets that you have and you want to safeguard from accidental deletion. Which configuration will meet this requirement? (SELECT ONE)

A) Archive sensitive data to Amazon Glacier using Life Cycle Rule

B) Configure cross-account access with an IAM Role prohibiting object deletion in the bucket and enable Cross Region Replication

C) Enable versioning on the bucket and multi-factor authentication delete as well.

D) Signed URLs to all users to access the bucket.

Practice Question 9

The answer to the question is…

You are auditing charge of S3 buckets for your company. There are multiple buckets, each is separated based on the type of data it is holding and the level of security required for that data. You are concerned of losing data on several buckets that you have and you want to safeguard from accidental deletion. Which configuration will meet this requirement? (SELECT ONE)

A) Archive sensitive data to Amazon Glacier using Life Cycle Rule

B) Configure cross-account access with an IAM Role prohibiting object deletion in the bucket and enable Cross Region Replication

C) Enable versioning on the bucket and multi-factor authentication delete as well.

D) Signed URLs to all users to access the bucket.