AWS Solution Architect

Storage Services

A Day in the Life of a Cloud Architect

You work as an architect for an e-commerce company that primarily deals with data. Your company is seeking cloud services to assist them in better managing their data while maintaining its security and availability. You've been asked to help them in meeting the following requirements:

- Setup a database that is scalable, has high data availability, secure.
- Preserve, retrieve, and restore every version of every object stored in the database system. They also want to make certain that data is encrypted.
- Evaluate data using standard SQL.
- Upgrade the volume type to improve or decrease the database's performance depending on their needs.

To achieve all the above along with some additional features, you will be learning a few concepts in this lesson that will help you find solutions for the above-given scenario.

Learning Objectives

By the end of this lesson, you will be able to:

- Create Amazon S3 buckets
- Enable versioning in Amazon S3 buckets
- Demonstrate static web hosting using Amazon S3
- Share Amazon S3 buckets between multiple accounts
- Back up and sync data in Amazon S3
- Transfer files from Amazon S3 to on-premises storage

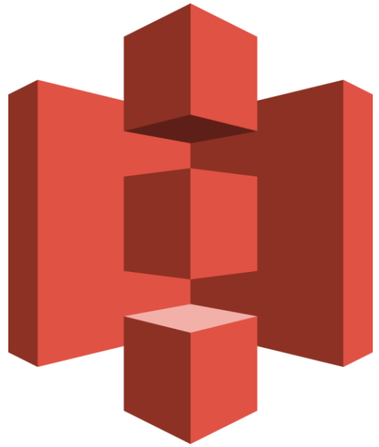
TECHNOLOGY

Introduction to Amazon S3

simplilearn

What Is Amazon S3?

Amazon Simple Storage Service (Amazon S3) is a web-based storage service offered by AWS for archive and back up data online. Users can store, retrieve, and protect any amount of data, at any time, from anywhere.



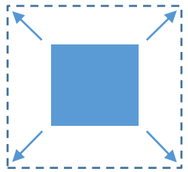
Amazon S3

Features of Amazon S3

Some of the features of Amazon S3 are:

01 **Availability**

02 **Scalability**



03 **Durability**

04 **Performance**

05 **Cost-efficiency**

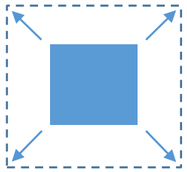
06 **Security**

Features of Amazon S3

Amazon S3 is designed for 99.99% availability. Users can choose the AWS region to store their data for optimizing latency and minimizing costs.

01 **Availability**

02 **Scalability**



03 **Durability**

04 **Performance**

05 **Cost-efficiency**

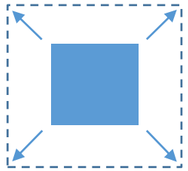
06 **Security**

Features of Amazon S3

Amazon S3 allows users to store data per their needs and scales the storage as required.

01 **Availability**

02 **Scalability**



03 **Durability**

04 **Performance**

05 **Cost-efficiency**

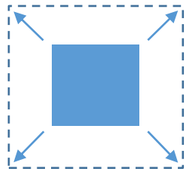
06 **Security**

Features of Amazon S3

Amazon S3 stores data redundantly across multiple facilities and multiple devices in each facility making it extremely durable.

01 **Availability**

02 **Scalability**



03 **Durability**

04 **Performance**

05 **Cost-efficiency**

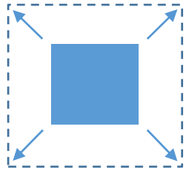
06 **Security**

Features of Amazon S3

Amazon S3 supports multipart uploads, which maximizes network throughput and resilience.

01 **Availability**

02 **Scalability**



03 **Durability**

04 **Performance**

05 **Cost-efficiency**

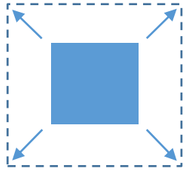
06 **Security**

Features of Amazon S3

Amazon S3 uses a pay-per-use model that allows users to store large data sets at a very low cost.

01 Availability

02 Scalability



03 Durability

04 Performance

05 Cost-efficiency

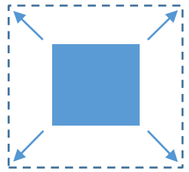
06 Security

Features of Amazon S3

Amazon S3 supports SSL (Secure Sockets Layer) data transfer and data encryption once the data is uploaded.

01 **Availability**

02 **Scalability**



03 **Durability**

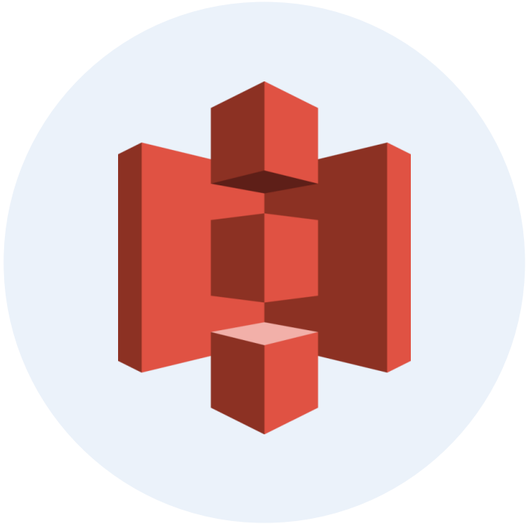
04 **Performance**

05 **Cost-efficiency**

06 **Security**

Components of Amazon S3

Components of Amazon S3



01

Buckets:
Buckets are containers for storing and organizing data of any format in Amazon S3.

02

Keys:
Keys are unique identifiers for each piece of data stored in Amazon S3 buckets.

03

Regions:
Regions are geographical locations where Amazon stores the S3 buckets created by its users.

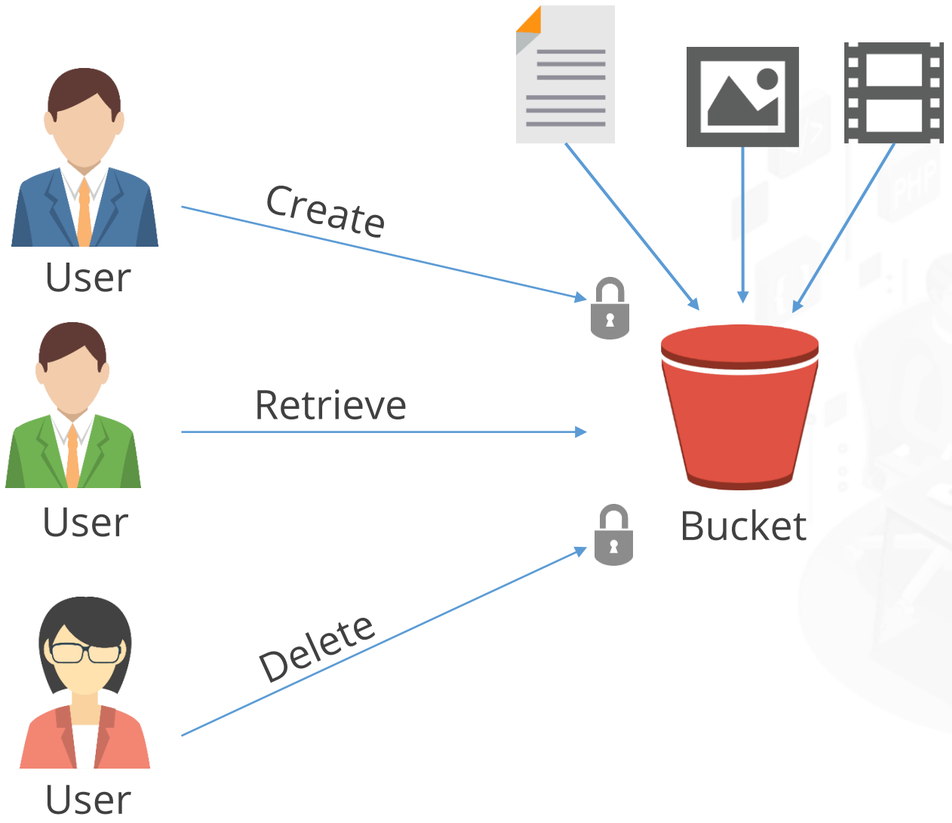
Components of Amazon S3

More about Amazon S3 buckets:

01 Buckets can be used to store text files, images, videos, and more.

02 Any number of objects can be stored in a bucket and the total bucket size is 5TB.

03 Access to each bucket action can be controlled by the user.



Amazon S3 Storage Classes

Amazon S3 Storage Classes

Amazon S3 comes in the following range of storage classes:

- Amazon S3 Standard
- Amazon S3 Intelligent-Tiering
- Amazon S3 Standard-Infrequent Access
- Amazon S3 One Zone-Infrequent Access
- Amazon S3 Glacier
- Amazon S3 Glacier Deep Archive

Amazon S3 Standard

Features of Amazon S3 Standard

- Stores objects that are frequently accessed with high availability and durability
- Has low latency and high throughput performance
- Is backed with Amazon S3 service-level agreement for availability
- Is ideal for dynamic websites, cloud and mobile applications, and file storage
- Has S3 Lifecycle management for an automatic transition of objects between S3 storage classes

Amazon S3 Intelligent-Tiering

Features of Amazon S3 Intelligent-Tiering

- Stores objects in two access tiers optimized for frequent and infrequent access
- Is designed for 99.99% availability and 11 9’s of durability
- Has the same low latency and high throughput performance of S3 Standard
- Includes a small monthly monitoring and auto-tiering fee
- Is ideal for long-lived data with unpredictable access patterns
- Moves objects automatically between two access tiers based on changing access patterns

Amazon S3 Standard-Infrequent Access (IA)

Features of Amazon S3 Standard-Infrequent Access

- Stores objects that are accessed less frequently
- Is designed for 99.99% availability and 11 9’s of durability
- Has the same low latency and high throughput performance of S3 Standard
- Has a low price per GB of storage
- Is ideal for long-term storage, backups, and as a data store for disaster recovery files

Amazon S3 One Zone-Infrequent Access (IA)

Features of Amazon S3 One Zone-Infrequent Access

- Stores objects that are accessed less frequently but require rapid access when needed
- Stores data in a single AZ and costs 20% less than S3 Standard-IA
- Has a low price per GB of storage
- Is ideal for customers who want a lower cost option for infrequently accessed data
- Is good for storing secondary backup copies of on-premises data

Amazon S3 Glacier

Features of Amazon S3 Glacier

- Is a secure, durable, and low-cost storage class for archiving rarely accessed data
- Provides three configurable retrieval options varying from minutes to hours
- Has a lower per GB storage price and per GB retrieval fee
- Offers a secure vault lock feature that enforces compliance
- Ensures data is resilient in case of an entire Availability Zone (Az) destruction
- Is ideal for database backups, compliance data, or audit log files that are rarely accessed

Amazon S3 Glacier Deep Archive

Features of Amazon S3 Glacier Deep Archive

- Is the lowest cost storage class supporting long-term retention and digital preservation
- Stores data that may be accessed once or twice in a year
- Is used for backup and disaster recovery
- Ensures data is replicated and stored in at least three geographically-dispersed AZs
- Ensures data is retrieved within 12 hours
- Is good for customers from highly regulated industries who retain data for 7-10 years

Performance Chart

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB

Performance across the S3 storage classes

Source: https://aws.amazon.com/s3/storage-classes/

Performance Chart

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours
Storage type	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes

Performance across the S3 storage classes

Source: https://aws.amazon.com/s3/storage-classes/

Create an S3 Storage Bucket

Duration: 05 mins

Problem Statement:

You have been asked to create an S3 storage bucket using an AWS account.

ASSISTED PRACTICE

Assisted Practice: Guidelines

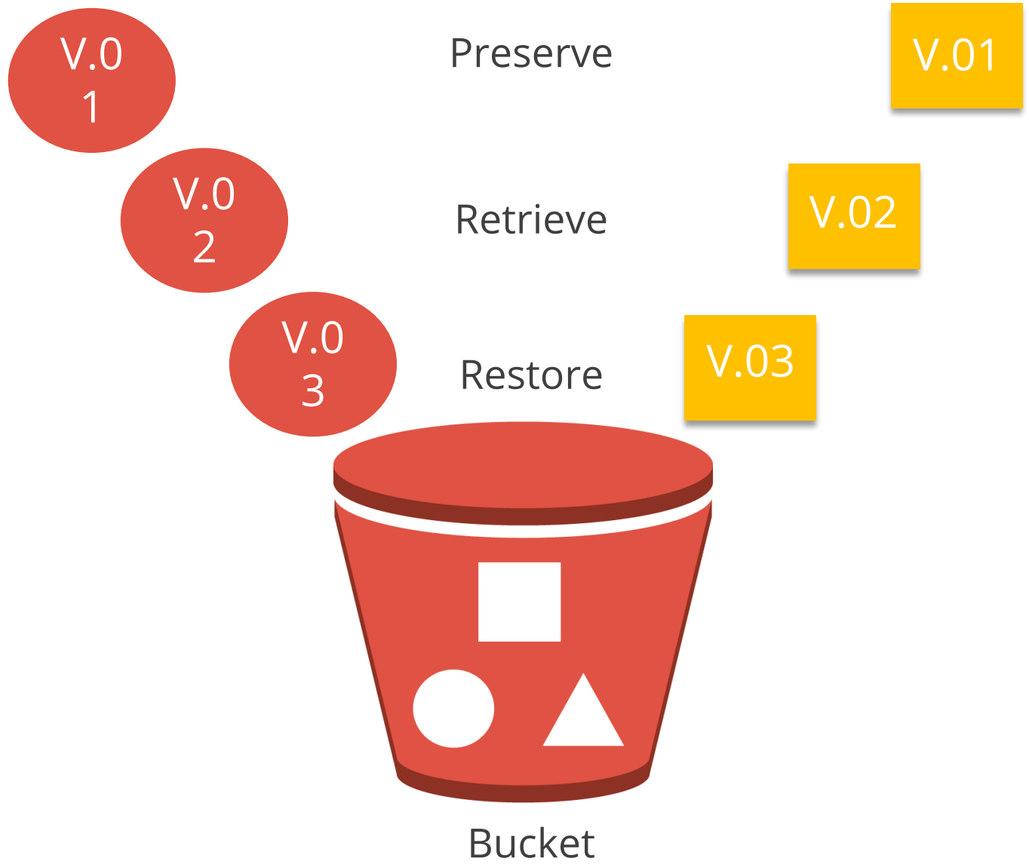
Steps to be followed:

1. Create an S3 bucket from the AWS Management Console

Version Control in Amazon S3

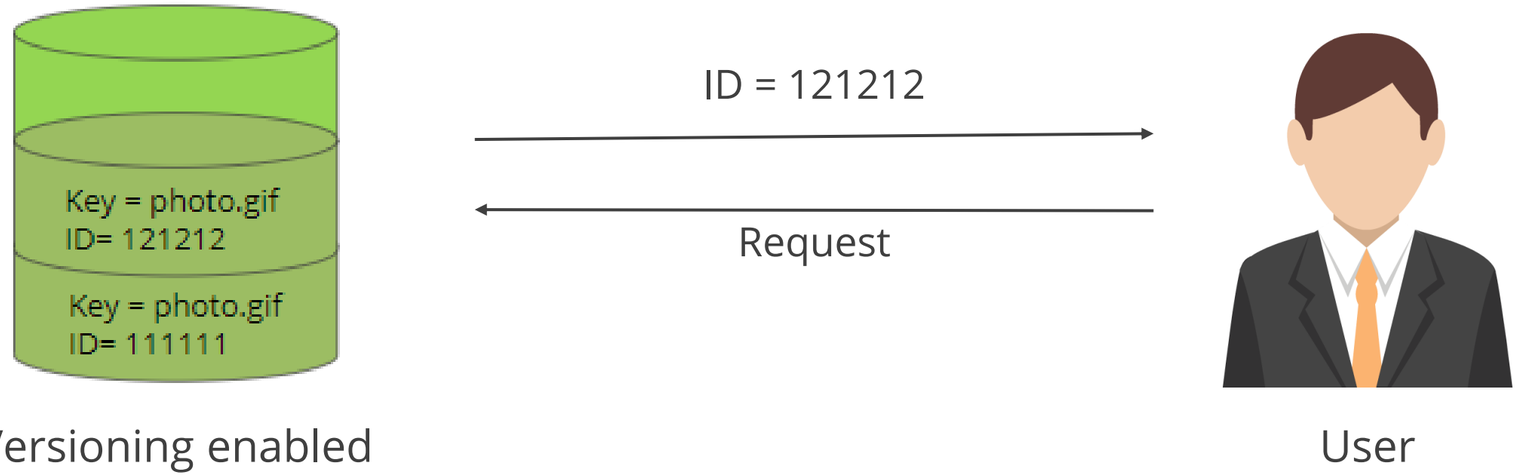
Version Control

Versioning is used to keep and maintain multiple variants of an object in the same bucket. It is used to preserve, retrieve, and restore earlier versions of every object stored in Amazon S3 buckets.



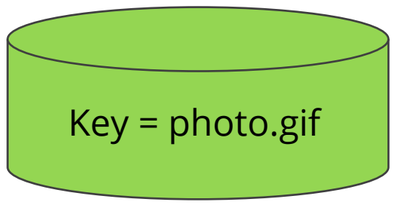
Version Control

Versioning allows you to recover your files from accidental deletion or overwrites.

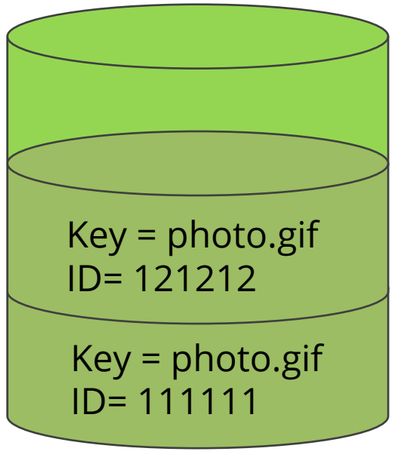


Version Control States

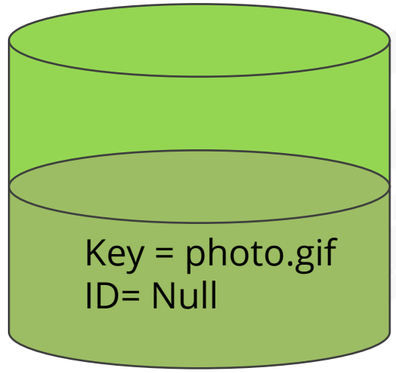
S3 buckets can be in one of the three states of versioning: unversioned, versioning enabled, and versioning suspended.



Unversioned



Versioning enabled

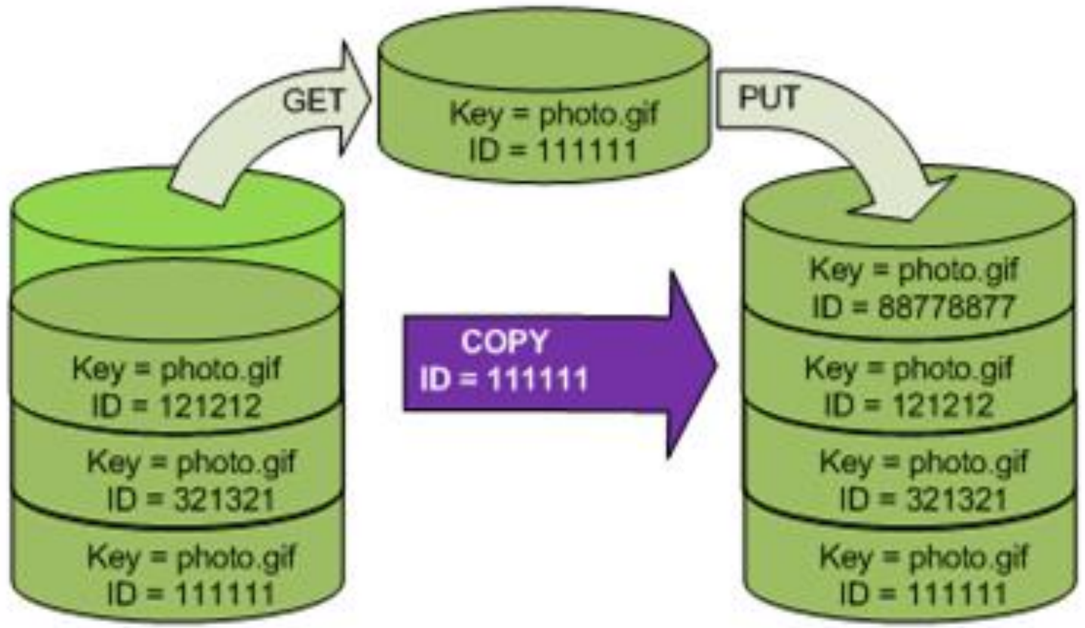


Versioning suspended

Restoring a Previous Version

There are two approaches to restore a previous version of an object:

1. Copy the previous version of an object into the bucket



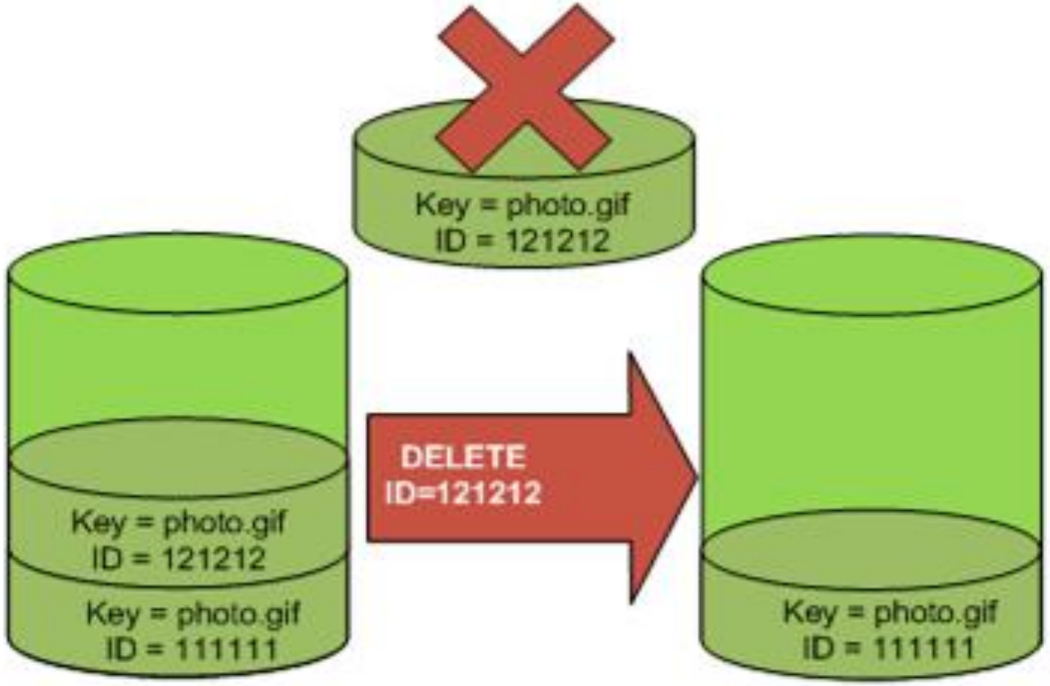
Versioning enabled

Versioning enabled

Source: https://docs.aws.amazon.com/AmazonS3/latest/dev/RestoringPreviousVersions.html

Restoring a Previous Version

2. Delete the current version of an object



Versioning enabled

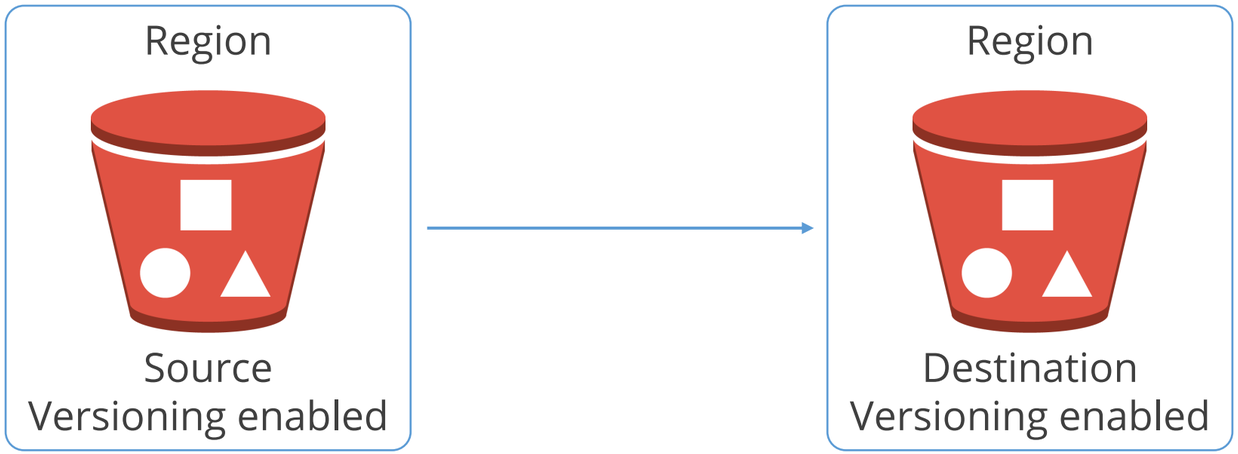
Versioning enabled

Source: https://docs.aws.amazon.com/AmazonS3/latest/dev/RestoringPreviousVersions.html

Cross-Region Replication

Cross-region replication is a bucket-level feature that enables automatic, asynchronous copying of objects across buckets in different AWS regions.

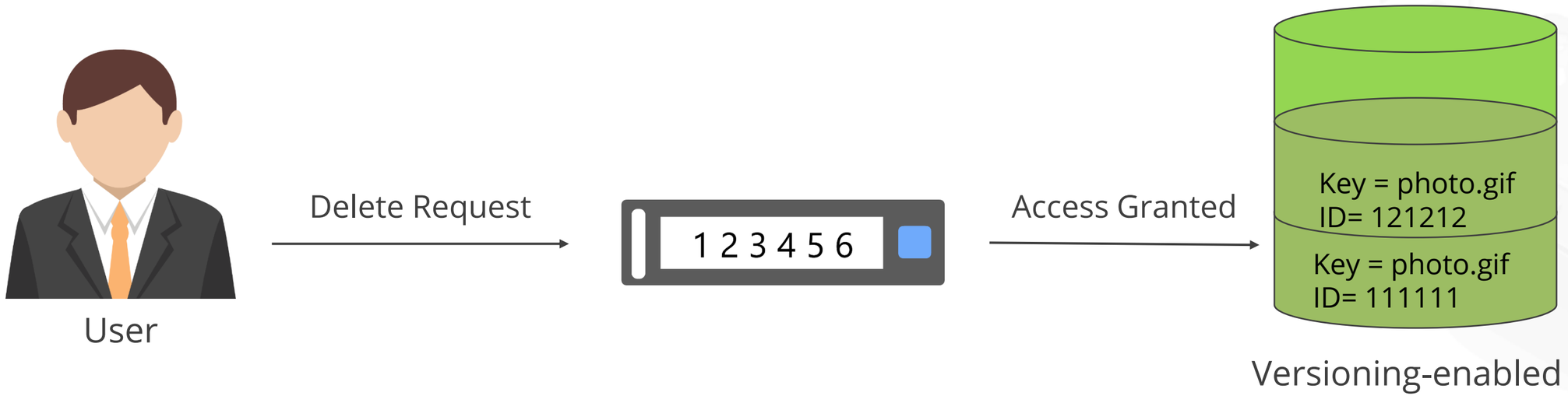
For cross-region replication to work, versioning needs to be enabled on both the source and destination buckets.



MFA Delete

Amazon S3 allows users to protect their data by enabling multi-factor authentication (MFA) delete.

This provides additional authentication for operations, such as changing the versioning state of a bucket permanently and deleting an object version.



Configuring S3 Bucket Versioning

Duration: 05 mins

Problem Statement:

You have been asked to configure versioning for an S3 bucket.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Enabling bucket versioning for an S3 bucket
2. Uploading files to an S3 bucket
3. Reuploading the same files to the S3 bucket to check if the versioning is enabled

TECHNOLOGY

S3 Lifecycle Management

S3 Lifecycle Management

Amazon S3 Lifecycle needs to be configured to manage a user's object so that these are stored cost effectively throughout their lifecycles.

An S3 Lifecycle configuration is a set of rules that define actions that Amazon S3 applies to a group of objects. There are two types of actions:

01 Transition actions

02 Expiration actions

Amazon S3 Actions

01 Transition actions

These actions define when objects transition to another storage class. For example, a user might choose to transition objects to S3 Standard-IA storage class 30 days after creating them or archive objects to the S3 Glacier Flexible Retrieval storage class one year after creating these.

Note

There are costs associated with S3 Lifecycle transition requests.

Amazon S3 Actions

02 Expiration actions

These actions define when objects expire. Amazon S3 deletes expired objects on a user's behalf.

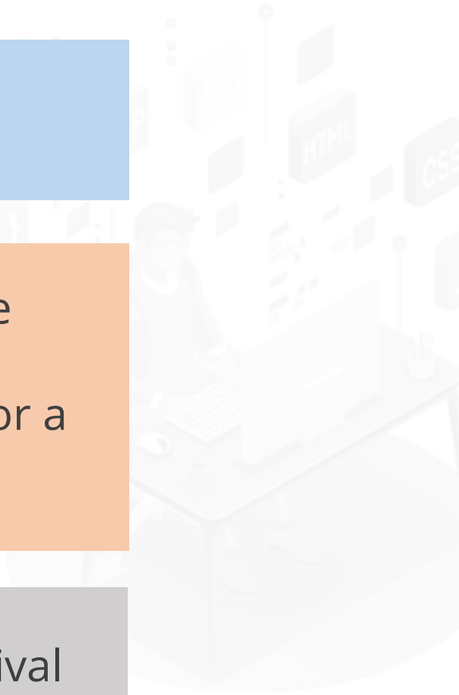
Note

Lifecycle expiration costs depend on when you choose to expire objects.

Managing Object Lifecycle

S3 Lifecycle configuration rules for objects that have a well-defined lifecycle can be managed using the following examples:

01	If a user uploads periodic logs to a bucket, the user's application might need them for a week or a month. After that, they can delete them.
02	Some documents are frequently accessed for a limited period. At some point, the users might not need real-time access to these, but users' organizations or regulations might require the users to archive them for a specific period. After that, they can delete these documents.
03	Users might upload some types of data to Amazon S3 primarily for archival purposes.



Creating a Lifecycle Configuration

1. An S3 Lifecycle configuration is an XML file that consists a set of rules with predefined actions that one wants Amazon S3 to perform on objects during their lifetime.

2. Users can also configure the lifecycle by using the Amazon S3 console, REST API, AWS SDKs, and the AWS Command Line Interface (AWS CLI).

3. Amazon S3 provides a set of REST API operations for managing the Lifecycle configuration on a bucket. Amazon S3 stores the configuration as a Lifecycle sub-resource that is attached to the user's bucket.

Age Based Retention

Duration: 05 mins

Problem Statement:

You have been asked to create a lifecycle rule in the S3 bucket to delete expired objects.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create a lifecycle rule in S3 bucket

S3 Intelligent-Tiering Archive Configuration

Duration: 05 mins

Problem Statement:

You have been asked to create an Intelligent-Tiering Archive configuration for an S3 bucket.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create an S3 bucket
2. Create the Intelligent-Tiering Archive configuration

Access Controls

Access Controls

Access control lists are used to grant basic read and write permissions to other AWS accounts. By default, the account which owns the object, has access to it, and can grant other users access to it through access control lists.

1

Object Ownership can be used to change the default behavior of an access control list.

2

When the access control lists are disabled, the user as the bucket owner automatically owns every object in their bucket.

3

Access control for a user's data is based on IAM policies, S3 bucket policies, VPC endpoint policies, and AWS Organizations' service control policies.

Sharing an S3 Bucket Between Multiple Accounts

Sharing an S3 bucket between multiple accounts can depend on the type of access that a user wants to provide.

The methods that can be used to grant cross-account access for sharing objects between multiple accounts are:

- Resource-based policies and AWS IAM policies
- Resource-based access control list (ACL) and IAM policies
- Cross-account IAM roles

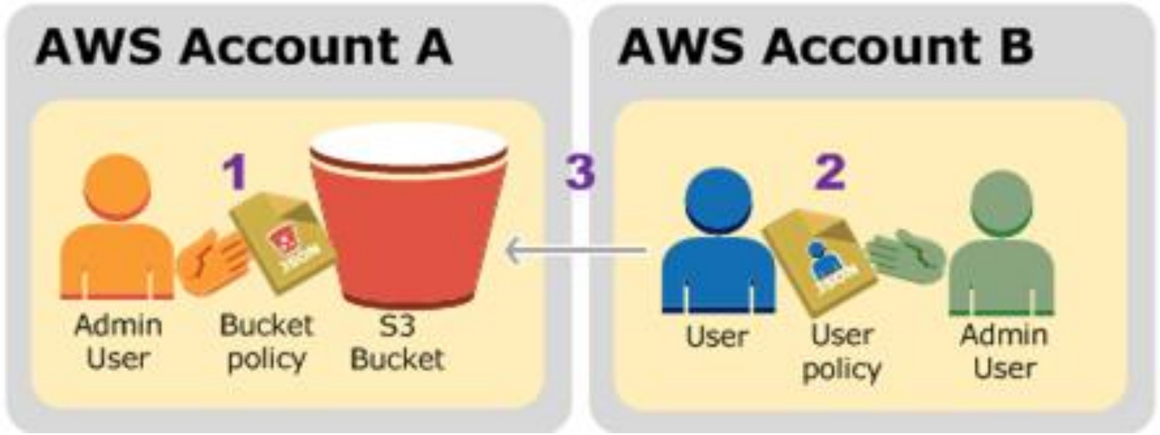
Sharing an S3 Bucket Between Multiple Accounts

The steps to share an S3 bucket between multiple accounts are:

1. Account A admin user attaches a bucket policy granting cross-account permissions to Account B.

2. Account B admin user attaches a user policy authorizing the permissions it received from Account A.

3. User in Account B verifies the permissions by accessing the bucket owned by Account A.

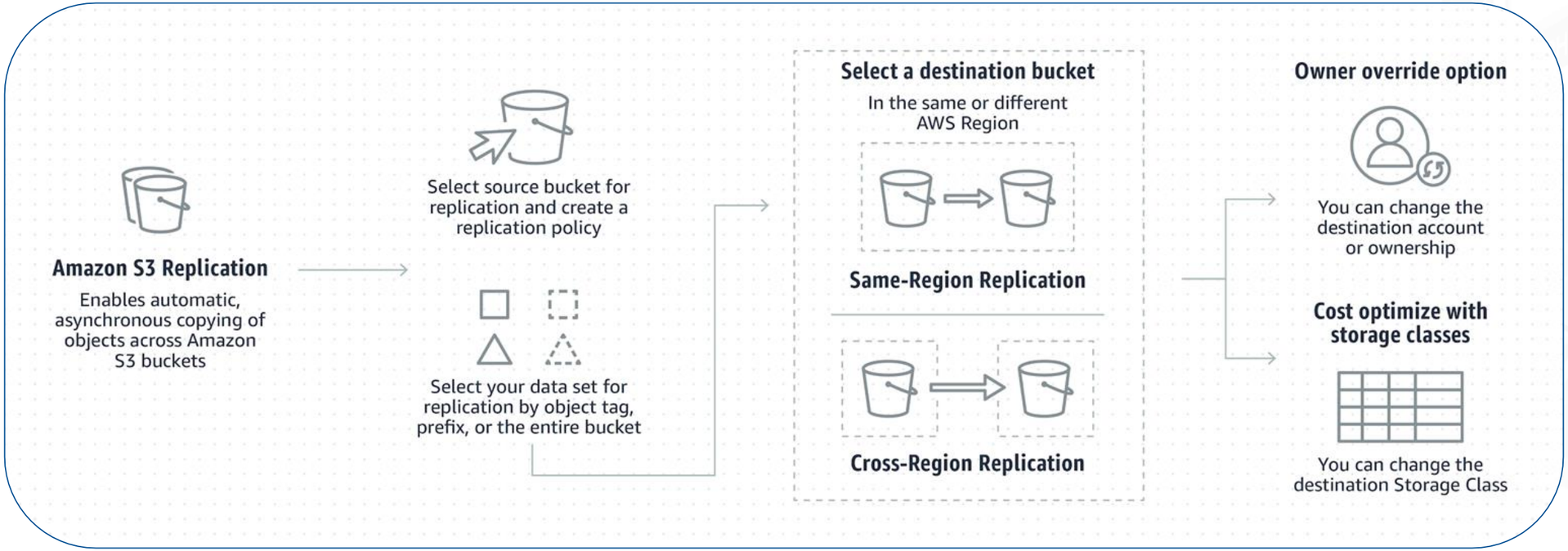


Source: https://docs.aws.amazon.com/AmazonS3/latest/dev/example-walkthroughs-managing-access-example2.html

Amazon S3 Replication

Amazon S3 Replication

Amazon S3 Replication is an elastic, fully managed, and low-cost feature that replicates objects between buckets by providing great flexibility and functionality in cloud storage.

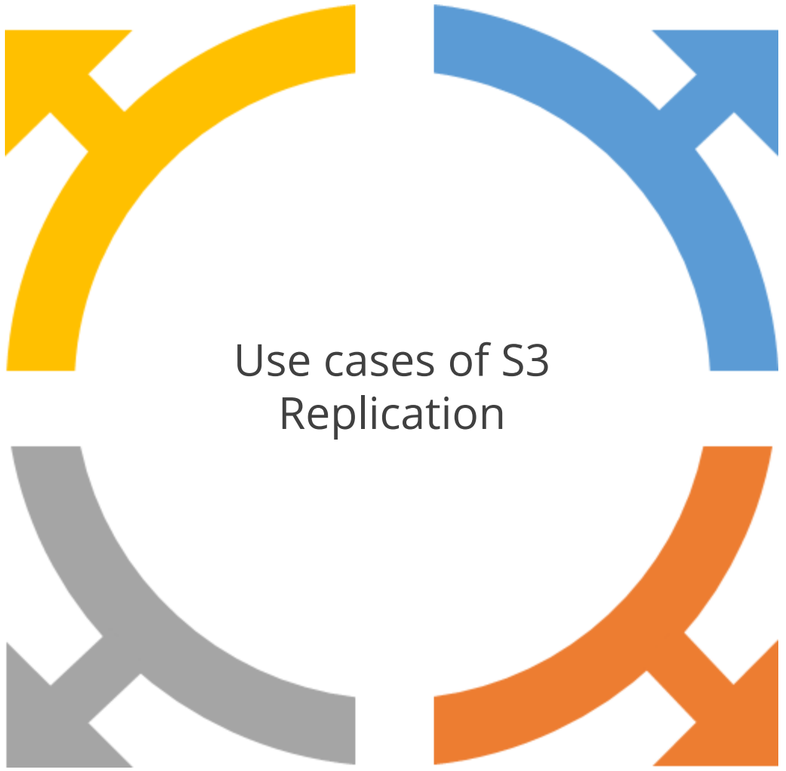


Working of S3 Replication

Source: https://aws.amazon.com/s3/features/replication/

S3 Replication Use Cases

Replicate objects while retaining metadata

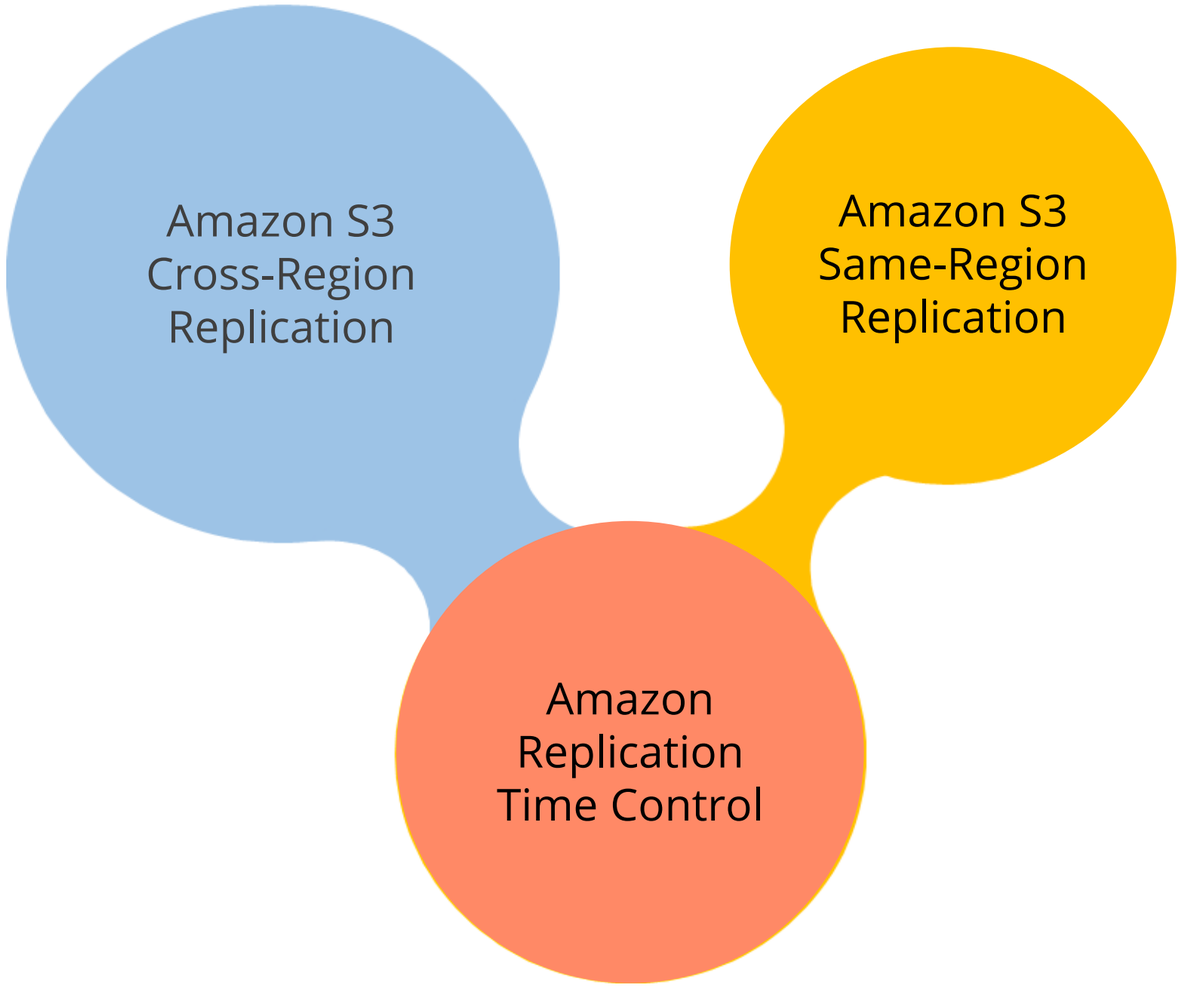


Replicate objects to more cost-effective storage classes

Replicate objects within 15 minutes

Maintain object copies under different ownership

Types of S3 Replications



Types of S3 Replications

Amazon S3 Cross-Region Replication (CRR)

- Replicates data automatically between buckets across different AWS Regions
- Configures to a source S3 bucket and replicates objects into a destination bucket in another AWS Region
- Replicates data at the bucket level, a shared prefix level, or an object level using S3 object tags
- Provides lower latency data access in different geographic regions

Types of S3 Replications

Amazon S3 Same-Region Replication (SRR)

- Replicates data automatically between buckets within the same AWS Region
- Replicates data at the bucket level, a shared prefix level, or an object level using S3 object tags
- Addresses data sovereignty and compliance requirements
- Changes account ownership for the replicated objects to protect accidental data deletion
- Collects logs from various S3 buckets for in-region processing, or configures live replication

Types of S3 Replications

Amazon Replication Time Control

- Replicates 99.99% of new objects stored in Amazon S3 within 15 minutes
- Meets compliance or business requirements for data replication, and provides visibility into S3 Replication activity
- Provides S3 Replication metrics that monitor:
  - The total number of S3 API operations that are pending replication
  - The total size of objects pending replication
  - The maximum replication time to the destination AWS Region

Replication of Objects from Source to Destination Bucket

Duration: 10 mins

Problem Statement:

You have been asked to replicate an object from a source bucket to a destination bucket using replication rules.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create two S3 buckets for source and destination
2. Add a file to the source bucket
3. Create a replication rule from the source bucket
4. Replicate the object file from the source bucket to the destination bucket
5. Deleting and verifying the replicated object file

Performance and Storage Gateway

Performance Guidance for S3 File Gateway

Performance guidance is basically provisioning of hardware for your S3 File Gateway Virtual Machine.

The cache disk size must be tuned to the size of the active working set for best performance.

Performance Guidance for S3 File Gateway

The table below shows the example of S3 File Gateway performance on Linux clients:

Example configurations	Protocol	Write throughput (file sizes 1 GB)	Cache hit read throughput	Cache miss read throughput
Root disk: 80 GB, io1 SSD, 4,000 IOPS	NFSv3 - 1 thread	110 MiB/sec (0.92 Gbps)	590 MiB/sec (4.9 Gbps)	310 MiB/sec (2.6 Gbps)
Cache disk: 512 GiB cache, io1, 1,500 provisioned IOPS	NFSv3 - 8 threads	160 MiB/sec (1.3 Gbps)	590 MiB/sec (4.9 Gbps)	335 MiB/sec (2.8 Gbps)
Minimum network performance: 10 Gbps	NFSv4 - 8 threads	160 MiB/sec (1.3 Gbps)	590 MiB/sec (4.9 Gbps)	335 MiB/sec (2.8 Gbps)
CPU: 16 vCPU | RAM: 32 GB	SMBV3 - 1 thread	115 MiB/sec (1.0 Gbps)	325 MiB/sec (2.7 Gbps)	255 MiB/sec (2.1 Gbps)

Performance Guidance for S3 File Gateway

The table below shows the example of S3 File Gateway performance on Linux clients:

Example configurations	Protocol	Write throughput (file sizes 1 GB)	Cache hit read throughput	Cache miss read throughput
NFS protocol recommended for Linux	SMBV3 - 8 threads	190 MiB/sec (1.6 Gbps)	590 MiB/sec (4.9 Gbps)	335 MiB/sec (2.8 Gbps)
Storage Gateway Hardware Appliance	NFSv3 - 1 thread	265 MiB/sec (2.2 Gbps)	590 MiB/sec (4.9 Gbps)	310 MiB/sec (2.6 Gbps)
Minimum network performance: 10 Gbps	NFSv3 - 8 threads	385 MiB/sec (3.1 Gbps)	590 MiB/sec (4.9 Gbps)	335 MiB/sec (2.8 Gbps)
	NFSv4 - 8 threads	385 MiB/sec (3.1 Gbps)	590 MiB/sec (4.9 Gbps)	335 MiB/sec (2.8 Gbps)

Performance Guidance for S3 File Gateway

The table below shows the example of S3 File Gateway performance on Windows clients:

Example configurations	Protocol	Write throughput (file sizes 1 GB)	Cache hit read throughput	Cache miss read throughput
Root disk: 80 GB, io1 SSD, 4,000 IOPS	SMBV3 - 1 thread	150 MiB/sec (1.3 Gbps)	180 MiB/sec (1.5 Gbps)	20 MiB/sec (0.2 Gbps)
Cache disk: 512 GiB cache, io1, 1,500 provisioned IOPS	SMBV3 - 8 threads	190 MiB/sec (1.6 Gbps)	335 MiB/sec (2.8 Gbps)	195 MiB/sec (1.6 Gbps)
Minimum network performance: 10 Gbps	NFSv3 - 1 thread	95 MiB/sec (0.8 Gbps)	130 MiB/sec (1.1 Gbps)	20 MiB/sec (0.2 Gbps)
CPU: 16 vCPU | RAM: 32 GB	NFSv3 - 8 threads	190 MiB/sec (1.6 Gbps)	330 MiB/sec (2.8 Gbps)	20 MiB/sec (0.2 Gbps)

Performance Guidance for S3 File Gateway

The table below shows the example of S3 File Gateway performance on Windows clients:

Example configurations	Protocol	Write throughput (file sizes 1 GB)	Cache hit read throughput	Cache miss read throughput
SMB protocol recommended for Windows	NFSv3 - 8 threads	190 MiB/sec (1.6 Gbps)	330 MiB/sec (2.8 Gbps)	190 MiB/sec (1.6 Gbps)
Storage Gateway Hardware Appliance	SMBV3 - 1 thread	230 MiB/sec (1.9 Gbps)	255 MiB/sec (2.1 Gbps)	20 MiB/sec (0.2 Gbps)
Minimum network performance: 10 Gbps	SMBV3 - 8 threads	835 MiB/sec (7.0 Gbps)	475 MiB/sec (4.0 Gbps)	195 MiB/sec (1.6 Gbps)
	NFSv3 - 1 thread	135 MiB/sec (1.1 Gbps)	185 MiB/sec (1.6 Gbps)	20 MiB/sec (0.2 Gbps)

Optimizing Gateway Performance

Gateway performance can be optimized by adding resources to the user's gateway.

Following are the ways in which gateway performance can be optimized by adding resources to gateway:

- Use higher-performance disks
- Add CPU resources to the gateway host
- Back gateway virtual disks with separate physical disks

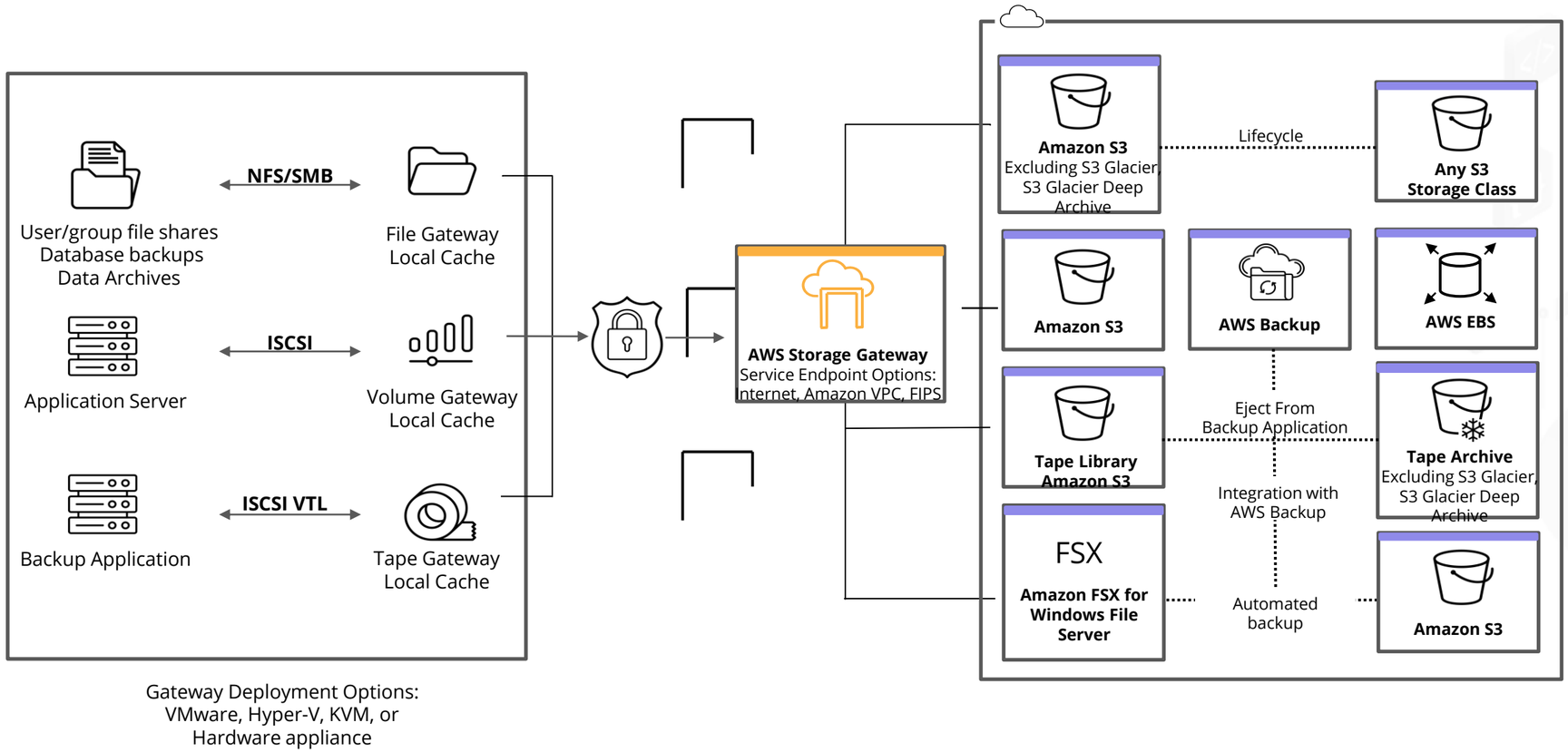
Optimizing Gateway Performance

Following are the ways in which gateway performance can be optimized by adding resources to the Application Environment:

- Increase the bandwidth between the application server and gateway
- Add CPU resources to the application environment

Storage Gateway

AWS Storage Gateway is a set of hybrid cloud storage services that provide on-premises access to virtually unlimited cloud storage.



Gateway Deployment Options:
VMware, Hyper-V, KVM, or
Hardware appliance

S3 Encryption

Server-side Encryption (SSE-S3)

Server-side Encryption protects data at rest where as Amazon S3 encrypts each object with a unique key.

- Amazon S3 server-side encryption uses one of the strongest block ciphers available to encrypt the data which is the 256-bit Advanced Encryption Standard (AES-256).
- There are no additional charges for using server-side encryption with Amazon S3-managed keys (SSE-S3).
- A bucket policy can be used to encrypt all the objects that are stored in a bucket.

Server-side Encryption (SSE-S3)

Below is an example that shows the bucket policy denying permissions to upload an object unless the request includes the x-amz-server-side-encryption header to request server-side encryption:

Example

```
{
  "Version": "2012-10-17",
  "Id": "PutObjectPolicy",
  "Statement": [
    {
      "Sid": "DenyIncorrectEncryptionHeader",
      "Effect": "Deny",
      "Principal": "*",
      "Action": "s3:PutObject",
      "Resource": "arn:aws:s3:::awsexamplebucket1/*",
      "Condition": {
        "StringNotEquals": {
          "s3:x-amz-server-side-encryption": "AES256"
        }
```

Server-side Encryption (SSE-S3)

Below is an example that shows the bucket policy denying permissions to upload an object unless the request includes the x-amz-server-side-encryption header to request server-side encryption:

Example

```
    {
      "Sid": "DenyUnencryptedObjectUploads",
      "Effect": "Deny",
      "Principal": "*",
      "Action": "s3:PutObject",
      "Resource":
"arn:aws:s3:::awsexamplebucket1/*",
      "Condition": {
        "Null": {
          "s3:x-amz-server-side-encryption": "true"
        }
      }
    }
```

API Support for Server-side Encryption

To request server-side encryption using the object creation REST APIs, provide the x-amz-server-side-encryption request header.

Following are the Amazon S3 APIs which support the request header:

- **PUT operations:** Specify the request header when uploading data using the PUT API.
- **Initiate multipart upload:** Specify the header in the initiate request when uploading large objects using the multipart upload API.
- **COPY operations:** When you copy an object, you have both the source object and target object.

Server-side Encryption with AWS KMS (SSE-KMS)

When an object is created, the use of server-side encryption with AWS Key Management Service (AWS KMS) keys to encrypt the data is specified. This encryption is known as SSE-KMS.

- Encryption can be applied either by uploading a new object or copying an existing object.
- SSE-KMS can be specified when using the Amazon S3 console, REST API operations, AWS SDKs, and AWS Command Line Interface (AWS CLI).

Server-side Encryption with SSE-C

Server-side encryption with customer-provided encryption keys (SSE-C) allows a user to set their encryption keys which only encrypt the object data.

- Amazon S3 manages the encryption as it writes to disks and decryption when objects are accessed.
- When an object is uploaded, Amazon S3 uses the encryption key the user provides to apply AES-256 encryption to the user's data and remove the encryption key from the memory.

Protecting Data using Client-side Encryption

Client-side encryption is the act of encrypting a user's data locally to ensure its security as it passes to the Amazon S3 service.

To enable the client-side encryption, we have the following options:

- Use a key stored in AWS Key Management Service (AWS KMS)
- Use a key that is stored within a user's application

The Amazon S3 service receives only the user's encrypted data and does not play a role in encrypting or decrypting it.

AWS Encryption SDK

The AWS Encryption SDK is a client-side encryption library that is separate from the language-specific SDKs.

- The AWS SDK is not tied to Amazon S3 and can be used to encrypt or decrypt the data to be stored anywhere.
- The AWS Encryption SDK and Amazon S3 encryption clients are not compatible because they produce ciphertexts with different data formats.

AWS Encryption SDK

Some of the AWS SDKs which support client-side encryption are as follows:

- AWS SDK for .NET
- AWS SDK for GO
- AWS SDK for Java
- AWS SDK for PHP
- AWS SDK for Ruby
- AWS SDK for C++

Server-side Encryption Using S3 and KMS

Duration: 10 mins

Problem Statement:

You have been asked to create a server-side encryption using S3 and KMS.

ASSISTED PRACTICE

Assisted Practice: Guidelines

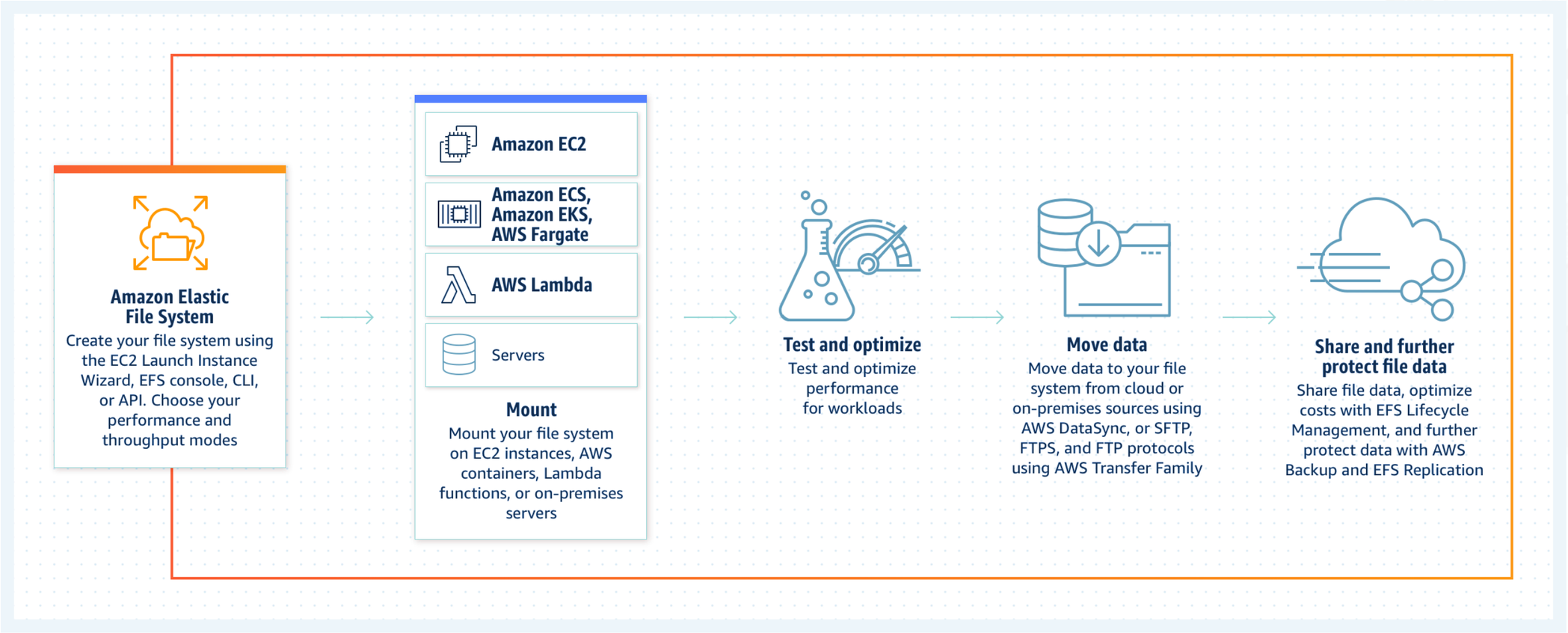
Steps to be followed:

1. Create an S3 bucket by enabling SSE-S3 encryption
2. Create a KMS
3. Create an S3 bucket by enabling SSE-KMS encryption

Elastic File System

Amazon Elastic File System

Amazon Elastic File System (Amazon EFS) provides a simple, scalable, fully manageable elastic NFS file system for use with AWS Cloud services and on-premises resources.



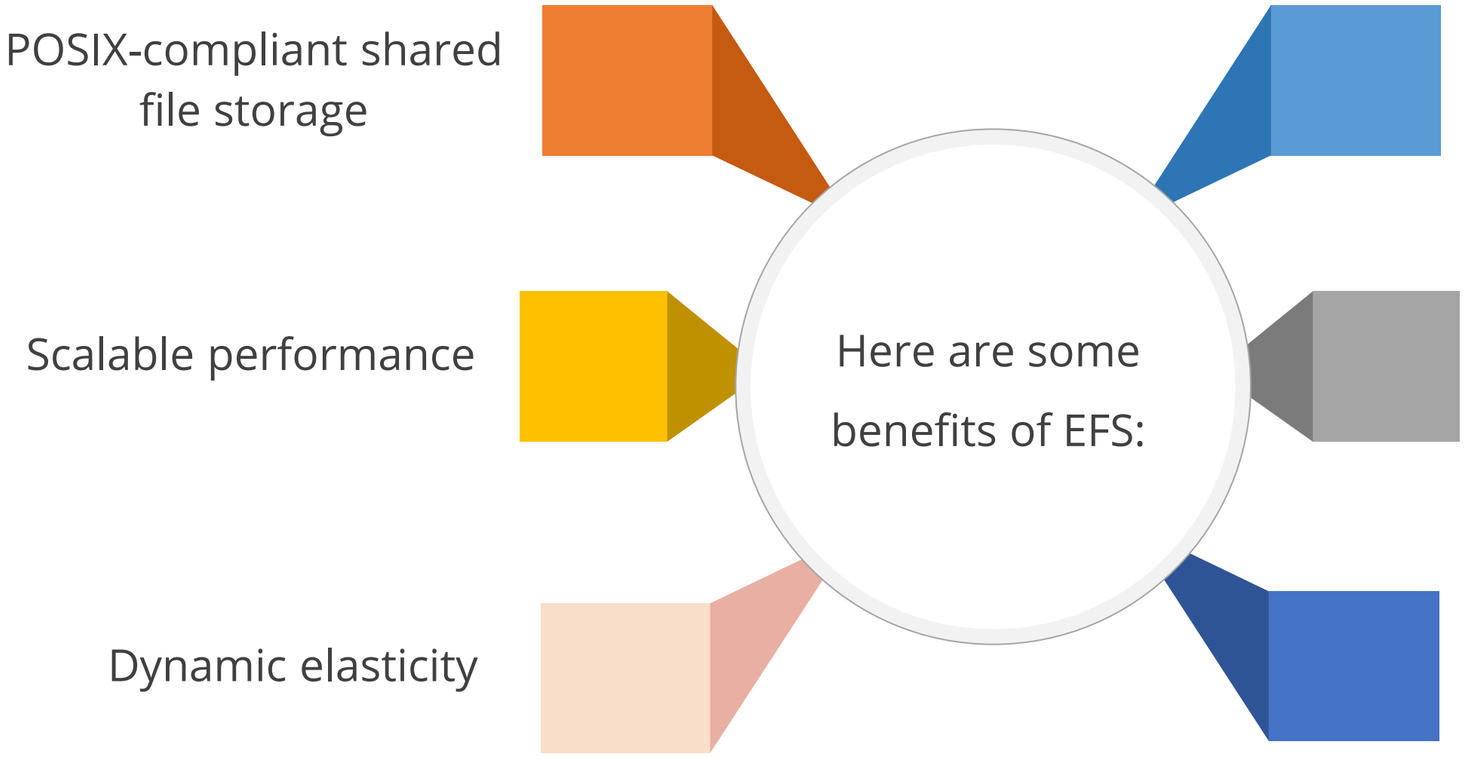
Elastic File System Overview

The data in EFS is stored and accessed from all the availability zones in the AWS region.

The users only pay for what they use with EFS storage.

EFS supports Network File System Version 4.0 and 4.1 protocol (NFSv4).

Benefits of Elastic File System



EFS vs. EBS

The difference between EFS and EBS are as follows:

Feature	EFS	EBS
Storage Size	No limitations	Maximum 16 TiB
Storage Type	Object storage	Block storage
Performance	Scalable	Hardly scalable
File Size Limitation	Maximum file size 47.9 TiB	No limitation
Data Throughput	Default throughput of 3 GB	SSD- and HDD-backed storage types
Data Access	Can be accessed concurrently	Limited to single EC2 instance
Availability Zone (AZ) Failure	Can survive one AZ failure	Cannot withstand AZ failure without snapshots

Amazon FSx

Amazon FSx

The Amazon FSx family of services makes it easy to launch, run, and scale shared storage powered by popular commercial and open-source file systems. The Amazon FSx consists of two categories:

FSx windows File server

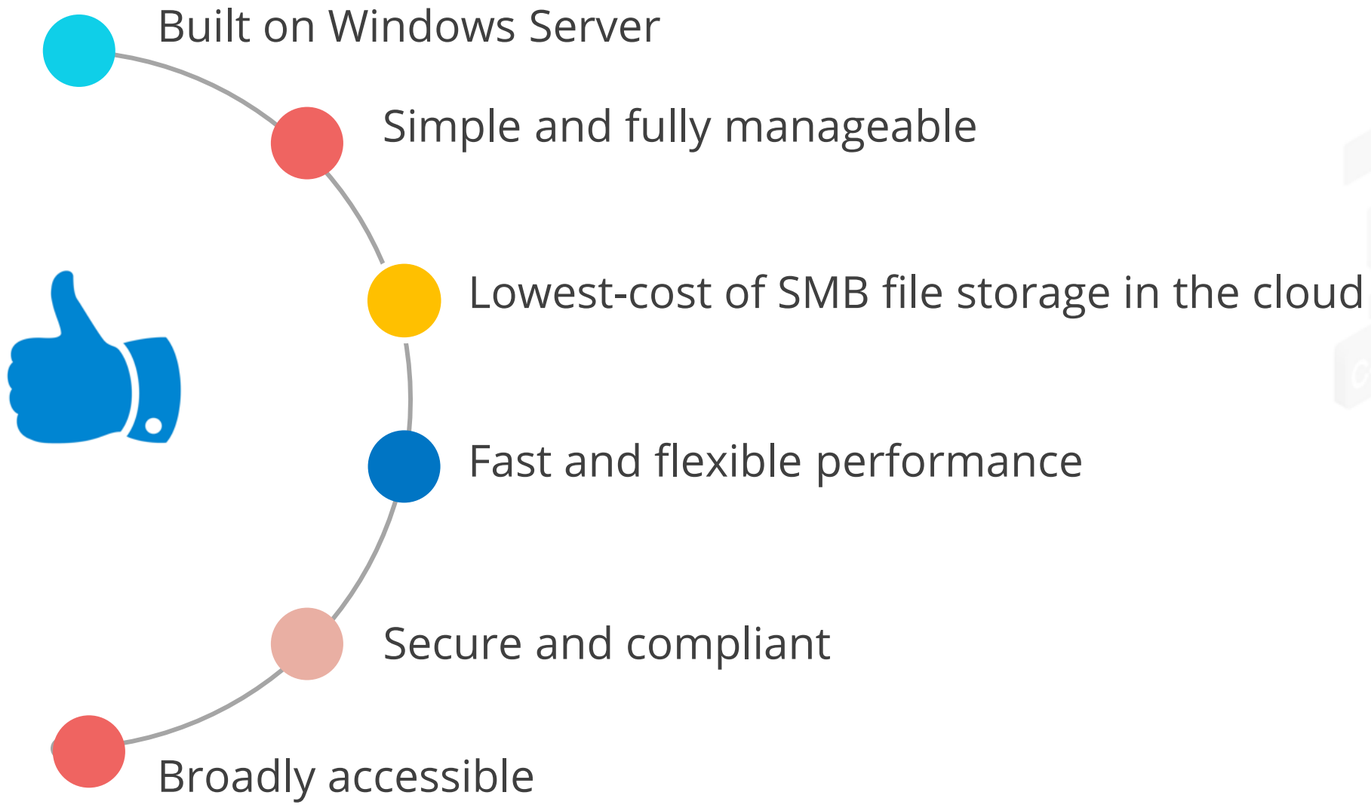
- The FSx for the windows file system is a fully native managed windows file system and it is easily integrated with the whole AWS environment.
- Amazon FSx for Windows File Server provides fully managed, highly reliable, and scalable file storage that is accessible over the industry-standard Server Message Block (SMB) protocol.

FSx windows File server

- It is built on Windows Server, delivering a wide range of administrative features such as user quotas, end-user file restore, and Microsoft Active Directory (AD) integration.
- It offers single-AZ and multi-AZ deployment options, fully managed backups, and encryption of data at rest and in transit.

Amazon FSx for Windows

Here are a few benefits of FSx for windows:



FSx Lustre

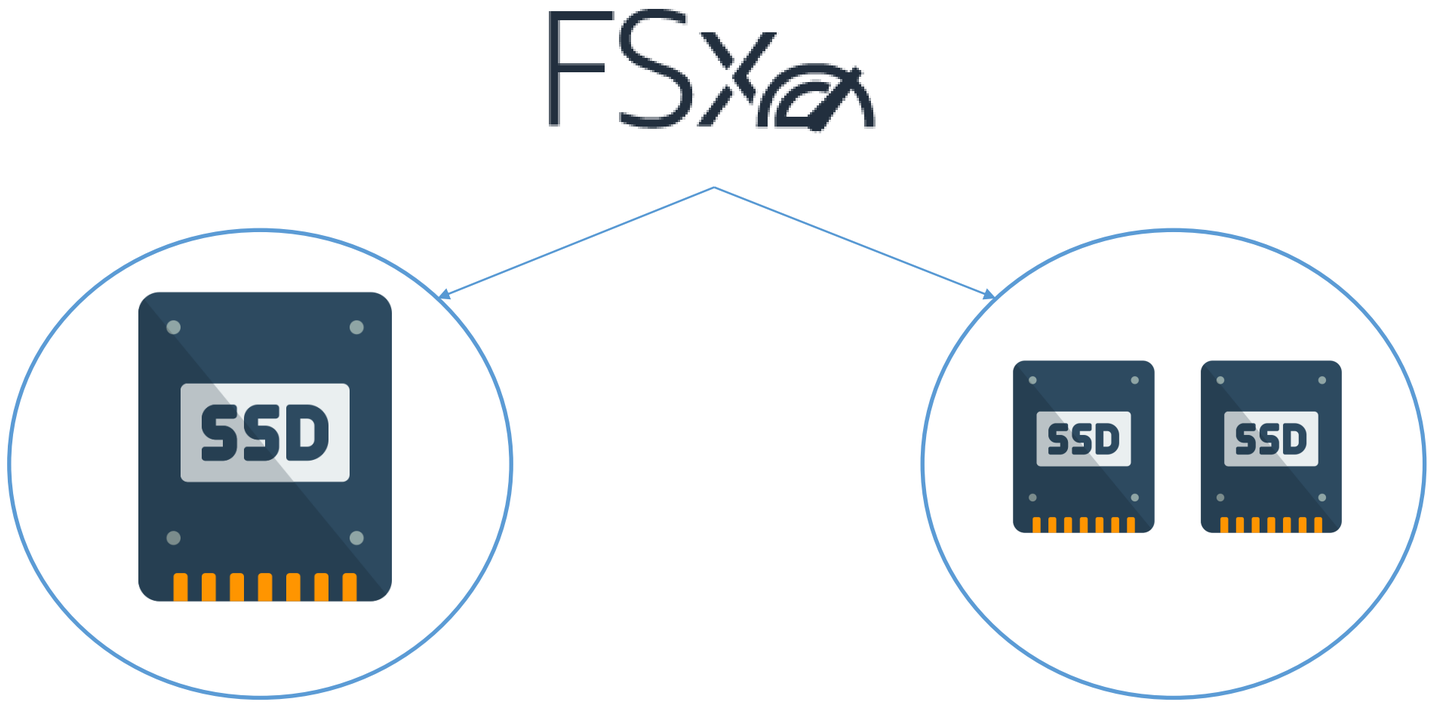
- FSx is integrated Natively and supports fast processing up to 100+ GB/s.
- The open-source Lustre file system is designed for applications that require fast storage that can keep up with your computing performance.
- Amazon FSx enables you to use Lustre file systems for any workload where storage speed matters.

FSx Lustre

- Amazon FSx for Lustre makes it easy and cost-effective to launch and run the world's most popular high-performance file system.
- Amazon FSx for Lustre integrates with Amazon S3, making it easy to use the Lustre file system to access data sets.

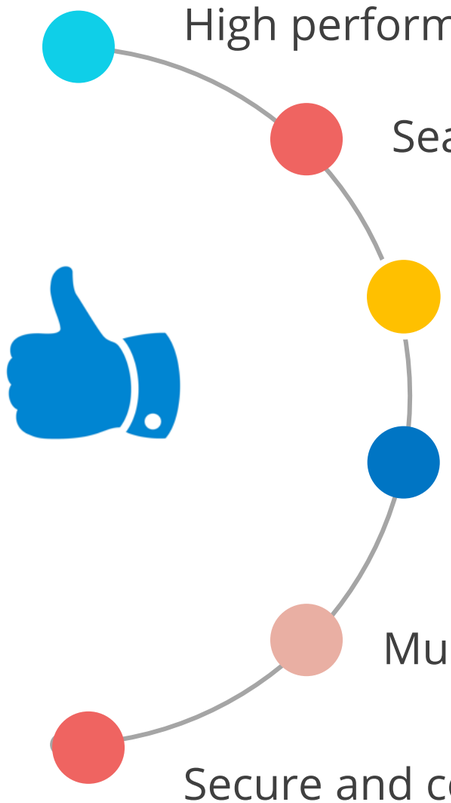
FSx Lustre Deployment Options

FSx Lustre provides an option for deployment: scratch File system and persistent File system



Amazon FSx for Lustre

Here are a few benefits of FSx for Lustre:



- High performance and scalability
- Seamless access to S3 data
- Simple and fully manageable
- Access from Linux applications running anywhere
- Multiple deployment options
- Secure and compliant

Create and Mount EFS on a Linux Server

Duration: 20 mins

Problem Statement:

You have been asked to create an EFS with two AWS instances and mount it on a Linux server.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create and customize an EFS
2. Create a Security Group to configure network access
3. Create two AWS instances to access the EFS
4. Install EFS on the Linux server instance that is created

Creation and Deletion of Customer Managed Key (KMS)

Problem Statement:

Duration: 10 mins

You have been asked to create and delete a customer managed key.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create a customer managed key
2. Delete a customer managed key

Unauthorized Link Access

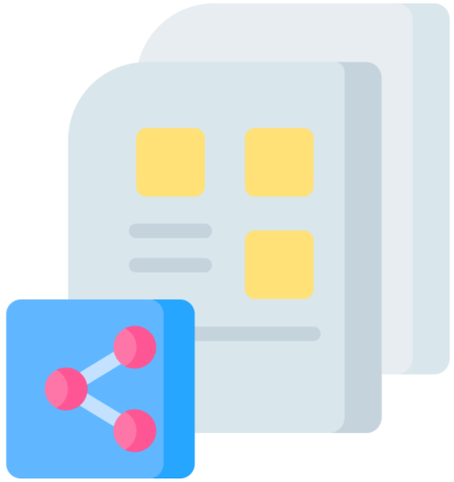
Unauthorized Link Access

When a user tries to use Amazon s3 to host a static website, there are chances of getting an Access Denied error.

To avoid or rectify the error, there are a few steps that need to be followed.

Unauthorized Link Access

Objects in the bucket must be publicly accessible:



- Only publicly accessible content is supported by the s3 static website endpoint.
- Open the URL of the object in a web browser to see if it is visible to the public in the s3 bucket.

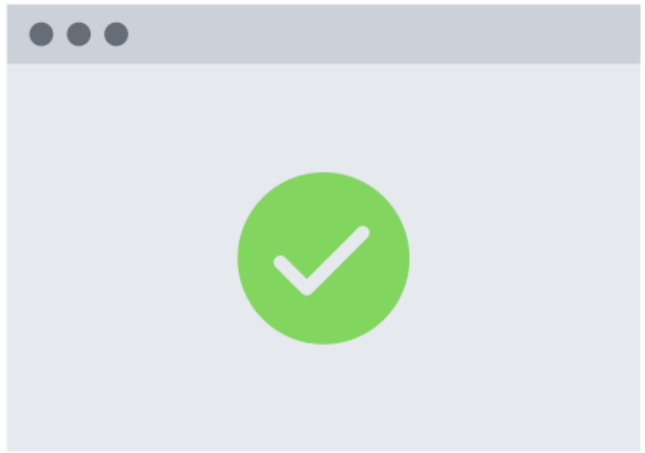
Unauthorized Link Access

S3 bucket policy must allow user to access the s3:GetObject action:

- Make sure that there are no deny statements that restrict the public to the s3:GetObject operation by reviewing the bucket policy
- Make sure there is no competing explicit refuse statement, even if the bucket policy contains an explicitly allow statement for s3:GetObject
- An explicit refuse statement will always take precedence over an explicit allow statement

Unauthorized Link Access

The AWS account that owns the bucket must also own the object:



The AWS account that owns the bucket must also control the objects to grant public read access.

Unauthorized Link Access

Objects in the bucket should not be AWS KMS-encrypted:

- Anonymous queries are not supported by AWS KMS.
- So, to serve things using the Amazon s3 static website endpoint, KMS encryption must be removed from the objects.

Unauthorized Link Access

Objects that are requested must exist in the s3 bucket:

- An Access Denied error for missing objects is returned if the user making the request lacks the s3:ListBucket access.
- To determine if an object is present in the bucket, use the head-object AWS CLI tool.

Unauthorized Link Access

Amazon s3 Block Public Access must be disabled on the bucket:

- Settings for Amazon s3 Block Public Access can be applied to specific buckets or AWS accounts.
- Verify that neither the s3 bucket nor the AWS account has any Amazon s3 Block Public Access settings enabled.

AWS Glue Data

AWS Glue is a serverless data integration service that makes it easy to discover, prepare, and combine data for analytics, machine learning, and application development.

Faster data integration

Different groups across your organization can use AWS Glue to work together on data integration tasks reducing the time it takes to analyze your data

Automate your data integration at scale

use AWS Glue to easily run and manage thousands of ETL jobs or to combine and replicate data across multiple data stores using SQL.

No servers to manage

AWS Glue runs in a serverless environment. It provisions, configures, and scales the resources required to run your data integration jobs.

Creation of a Glue Data Catalog

Duration: 10 mins

Problem Statement:

You have been asked to create and set up a Glue Data Catalog.

ASSISTED PRACTICE

Assisted Practice: Guidelines

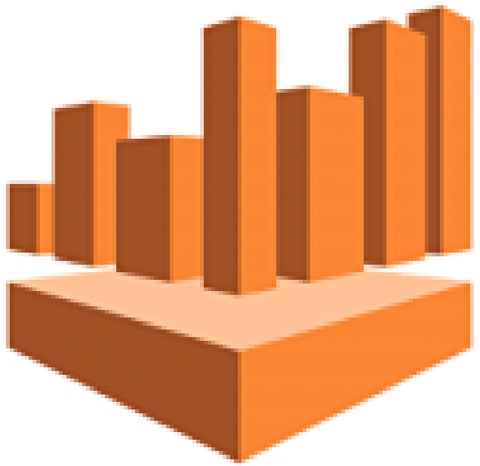
Steps to be followed:

1. Create two AWS S3 buckets
2. Create and execute queries in Athena

Amazon Athena

Amazon Athena

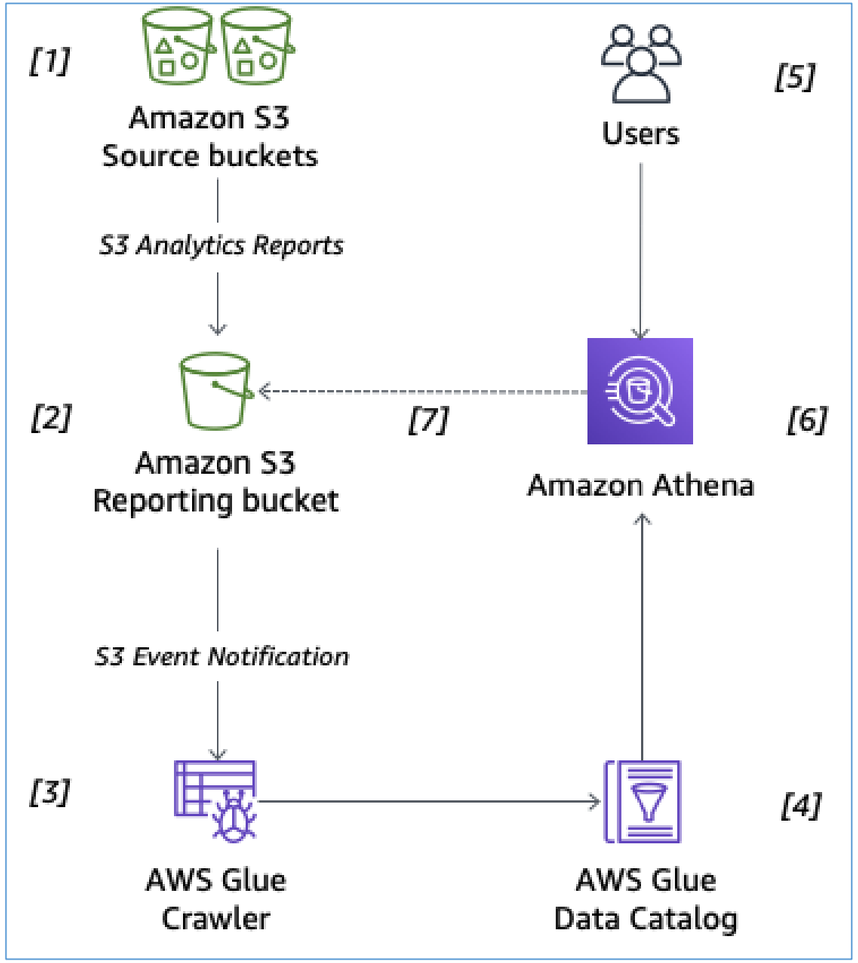
Amazon Athena is an interactive query service that allows easy analysis of data in Amazon S3 using standard SQL. It is serverless so there is no infrastructure to manage, and users pay only for the queries that they run.



Amazon Athena

Amazon Athena

Steps to query S3 analytics data with Amazon Athena:



Source: https://aws.amazon.com/blogs/storage/query-amazon-s3-analytics-data-with-amazon-athena/

Query with Athena

Duration: 15 mins

Problem Statement:

You have been asked to run a query in Athena to perform some operations on a specific S3 bucket.

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed:

1. Create a VPC endpoint
2. Create a Data Glue Catalog

Key Takeaways

- The lifecycle configuration rule can be implemented in the S3 bucket to delete expired objects.
- The objects in the S3 bucket can be moved from a source bucket to a destination bucket by applying the Replication rule.
- Server-side Encryption can be enabled to an S3 bucket using Key Management Service (KMS).
- EFS can be mounted on a Linux Server, and FSx can be mounted on a Windows Server with the help of a directory service.
- Amazon Athena is an interactive query service that allows easy data analysis in Amazon S3 using standard SQL.

Implementation of S3 Buckets for Web Content Hosting

Duration: 25 mins

Project agenda: To host content on the static website using an S3 bucket

Description:

You have been asked to deliver HTML content on a static website using an S3 bucket.

Perform the following:

1. Create an S3 bucket
2. Generate Bucket Policy
3. Enabling static website hosting by uploading index.html
4. Testing and verifying the hosted content

LESSON-END PROJECT

TECHNOLOGY

Thank You

simplilearn