AWS Solution Architect

Storage Services

Learning Objectives

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

- Create Amazon S3 buckets to store data and objects in the Cloud
- Enable versioning in Amazon S3 buckets to enable data protection
- Demonstrate static web hosting using Amazon S3 to optimize the cost and enable scalability
- Share Amazon S3 buckets between multiple accounts to centralize data storage within an organization

Learning Objectives

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

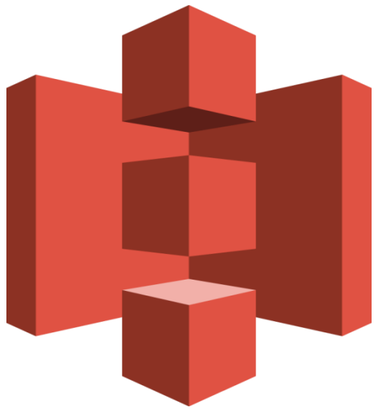
- Backup and sync data with Amazon S3 to provide data durability and redundancy
- Transfer files from Amazon S3 to on-premises storage to perform data analysis on-premises

TECHNOLOGY

Introduction to Amazon S3

What Is Amazon S3?

Amazon Simple Storage Service (Amazon S3) is a web-based storage service offered by AWS for archiving and backing up data online.



Amazon S3

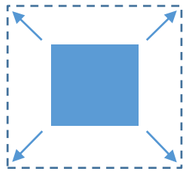
Users can store, retrieve, and protect any amount of data at any time from anywhere.

Amazon S3: Features

It offers a range of features for storage solutions, including:

Availability

Scalability



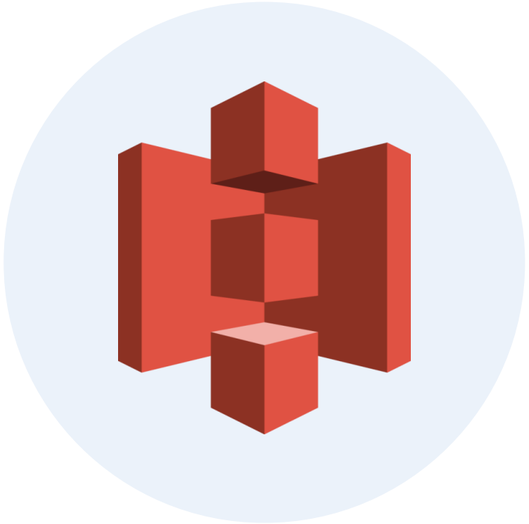
Durability

Performance

Cost-efficiency

Security

Amazon S3: Components



01

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

02

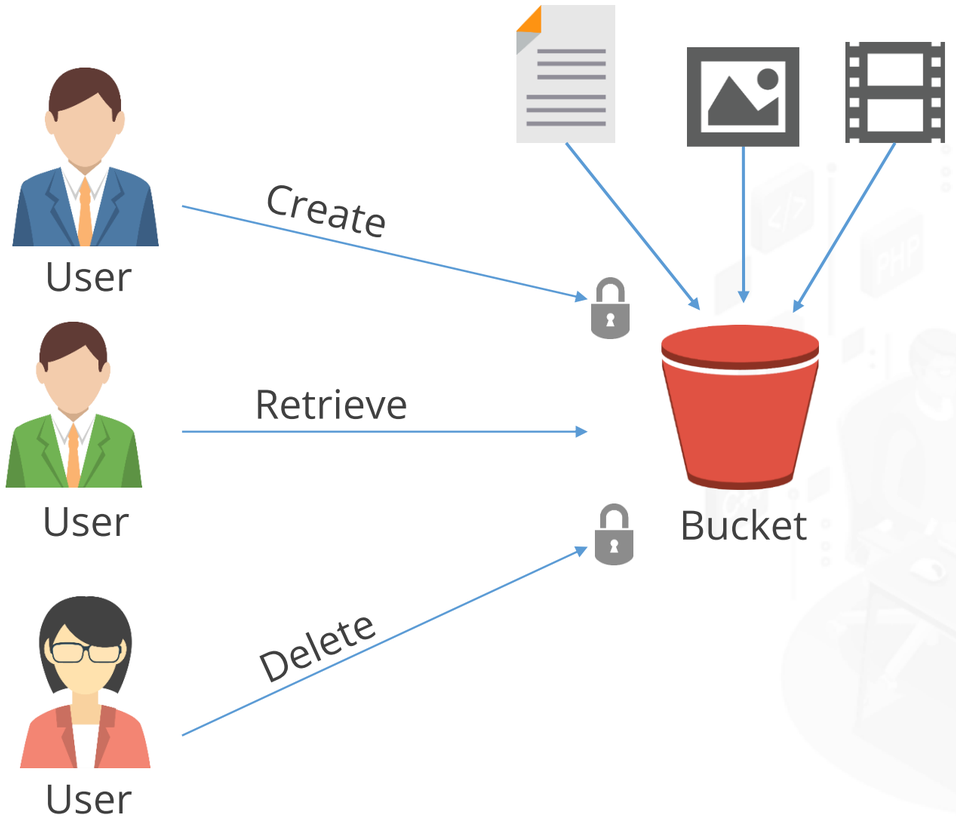
Keys:
They are a unique identifier for a piece of data stored in Amazon S3 buckets.

03

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

Amazon S3: Buckets

- They can store text files, images, videos, and more.
- Any number of objects can be stored in a bucket, and the total bucket size is 5TB.
- The user can control access to each bucket option.



Amazon S3: Storage Classes

Amazon S3 has 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

Performance Chart

Below is the performance chart across the S3 storage classes:

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

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

Performance Chart

Below is the performance chart across the S3 storage classes:

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

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

Creating an S3 Bucket

Duration:10 min

Problem Statement:

You have been assigned a task to create an S3 bucket using an AWS account for storing and managing data.

Outcome:

You will be able to implement cost estimates for your Terraform templates, providing you with valuable insights into the potential costs associated with your infrastructure resources.

Note: Refer to the demo document for detailed steps:
01_Creating_an_S3_Bucket

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Create an S3 bucket

Quick Check

You need to store large amounts of data in Amazon S3, and you want to minimize storage costs for data that is infrequently accessed. Which S3 storage class should you choose?

A. S3 Standard

B. S3 Intelligent-Tiering

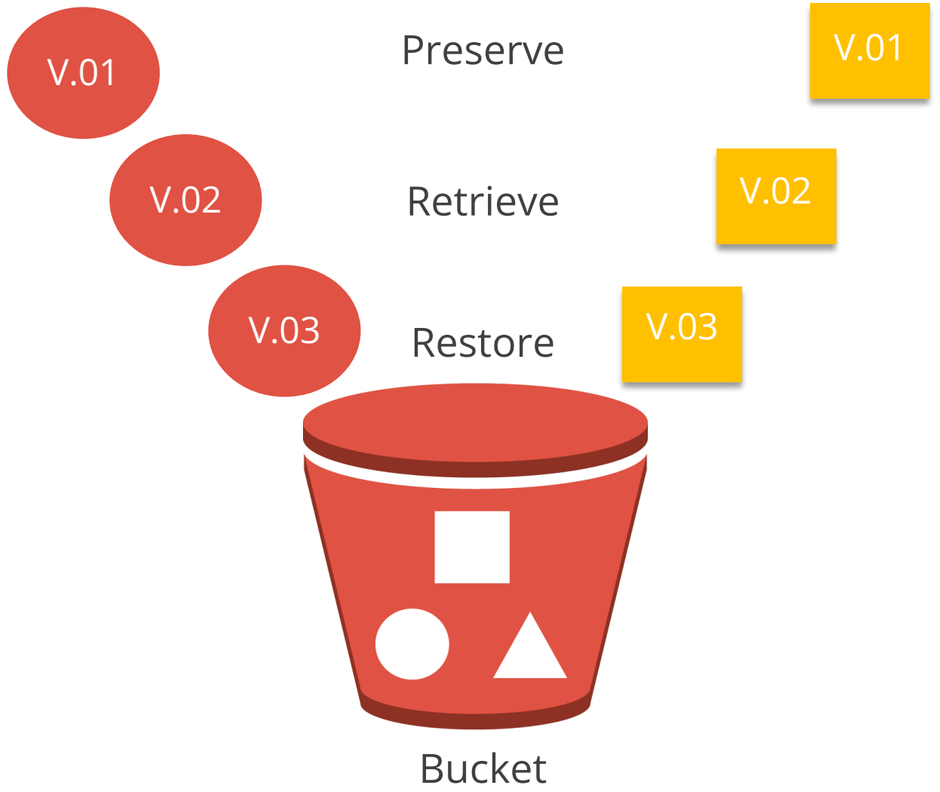
C. S3 Glacier

D. S3 Standard-IA (Infrequent Access)

Version Control in Amazon S3

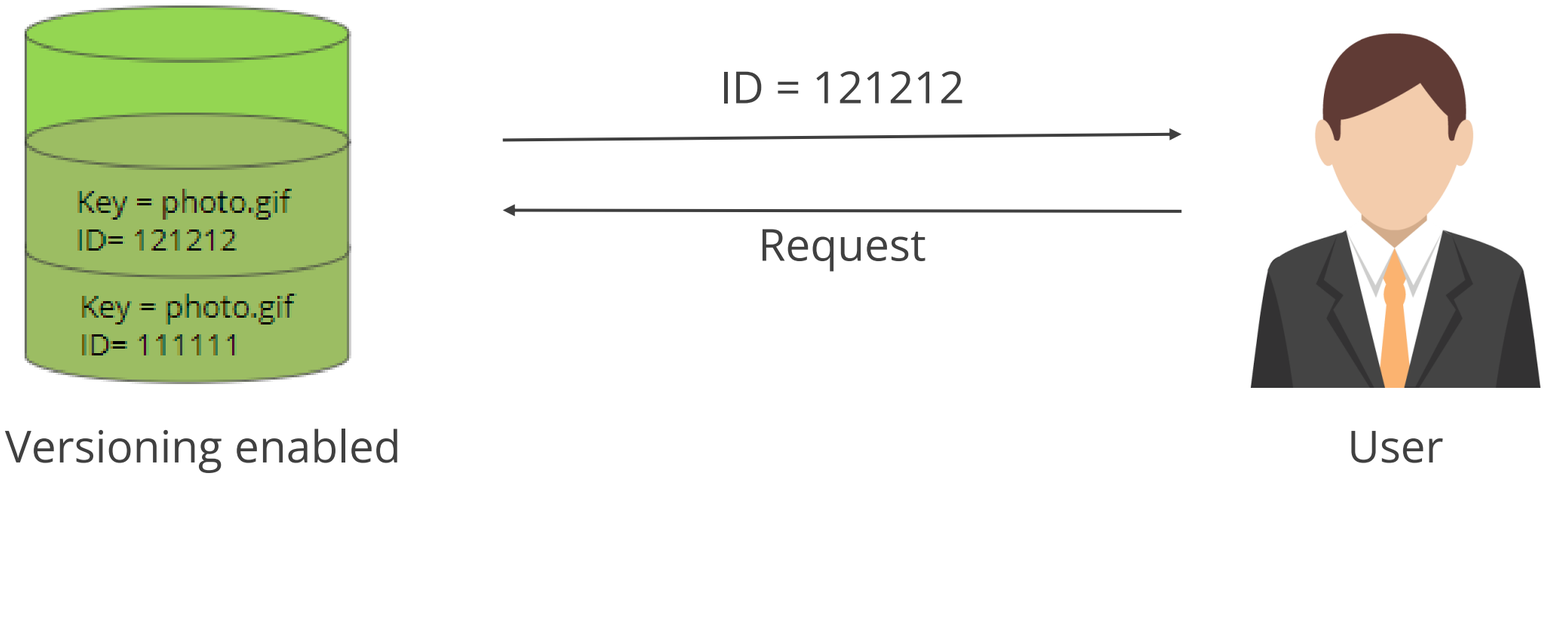
Version Control

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



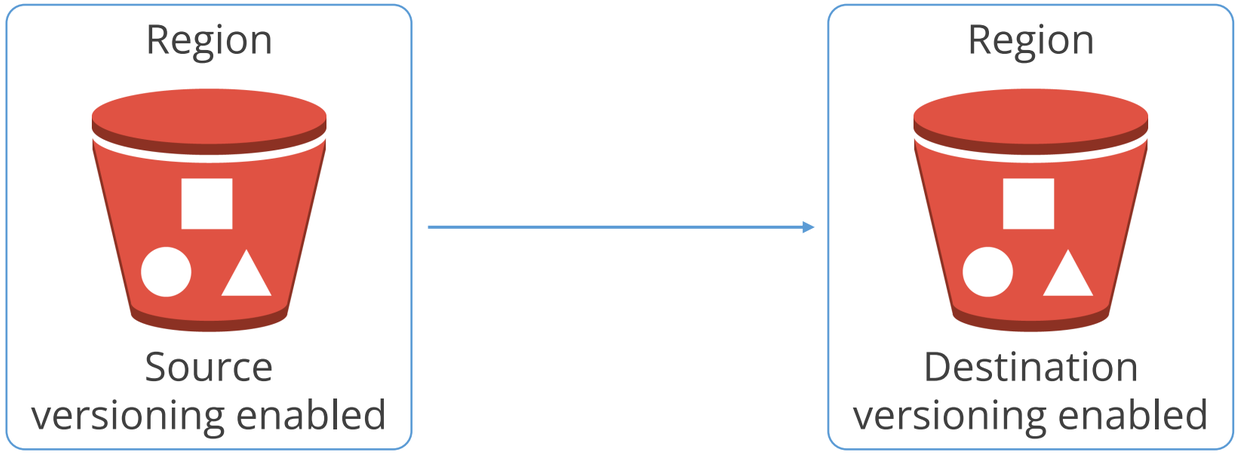
Version Control

Versioning allows users to recover their files from accidental deletions or overwrites.



Cross-Region Replication

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



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

Configuring S3 Bucket Version

Duration:10 min

Problem Statement:

You have been assigned a task to demonstrate the process of configuring S3 bucket versioning for version control and file management.

Outcome:

You will be able to successfully configure S3 bucket versioning and manage file uploads to handle multiple versions of the same data.

Note: Refer to the demo document for detailed steps:
02_Configuring_S3_Bucket_Version

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Enable bucket versioning for an S3 bucket
2. Upload files to the S3 bucket
3. Reupload the same files to the S3 bucket

Quick Check

You have enabled versioning on your S3 bucket and accidentally deleted an important object. How can you recover the deleted object?

A. Use S3 Lifecycle policies to restore the object

B. Enable Cross-Region Replication to recover the object

C. Restore the object from a previous version

D. Re-upload the object from a local backup

TECHNOLOGY

S3 Lifecycle Management

S3 Lifecycle Management

S3 Lifecycle configuration is done to manage a user's objects cost-effectively throughout their lifecycles.

An S3 Lifecycle configuration is a set of rules defining actions applied to a group of objects. There are two types of actions:

01 Transition actions

02 Expiration actions

Managing Object Lifecycle

S3 Lifecycle configuration rules help manage objects by:

01	Deleting periodic logs after a set time (e.g., a month)
02	Archiving frequently-accessed documents when real-time access is no longer needed
03	Retaining data uploaded for long-term archival

Creating a Lifecycle Configuration

1. An S3 Lifecycle configuration is an XML file containing rules and predefined actions for Amazon S3 to perform on objects during their lifetime.

2. Users can configure the lifecycle through the Amazon S3 console, REST API, AWS SDKs, or the AWS Command Line Interface (AWS CLI).

3. Amazon S3 manages the lifecycle configuration of a bucket by storing it as a lifecycle sub-resource attached to that bucket.

Configuring S3 Intelligent-Tiering Archive

Duration:10 min

Problem Statement:

You have been assigned a task to set up an S3 Intelligent-Tiering Archive configuration to optimize storage costs by automatically transitioning objects to cost-effective access tiers based on changing access patterns.

Outcome:

You will be able to successfully configure S3 bucket versioning and manage file uploads to handle multiple versions of the same data.

Note: Refer to the demo document for detailed steps:
03_Configuring_S3_Intelligent_Tiering_Archive

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Create a new Intelligent-Tiering Archive configuration

Configuring Age-Based Retention

Duration:10 min

Problem Statement:

You have been assigned a task to create a lifecycle rule in S3 bucket to delete expired objects.

Outcome:

You will be able to successfully create and apply a lifecycle rule in the S3 bucket to automatically delete expired objects, optimizing storage management.

Note: Refer to the demo document for detailed steps:
04_Configuring_Age_Based_Retention

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Create an S3 bucket lifecycle rule

Access Controls

Access control lists (ACLs) grant basic read and write permissions to other AWS accounts. By default, the account that owns an object can access it and grant other users access through ACLs.

1

Users can use Object Ownership to change the default behavior of an ACL.

2

When users disable the ACLs, the bucket owner automatically owns every object in the bucket.

3

User access control 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 a user wants to provide.

The methods 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

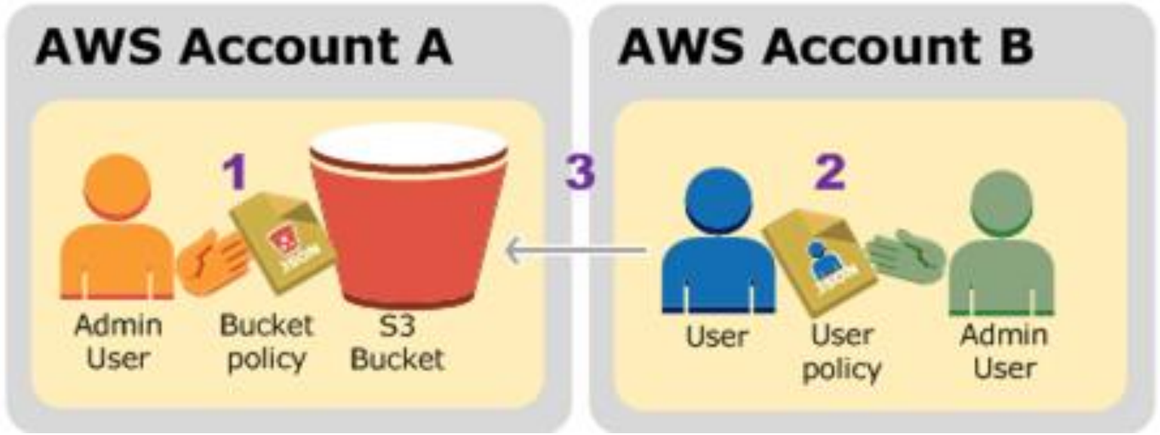
Sharing an S3 Bucket Between Multiple Accounts

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

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. A 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

Quick Check

You have a large dataset in S3 that is frequently accessed for the first month and rarely accessed afterward. You want to automatically transition these objects to a cheaper storage class after 30 days. Which feature should you use?

A. S3 Versioning

B. S3 Lifecycle Management

C. S3 Cross-Region Replication

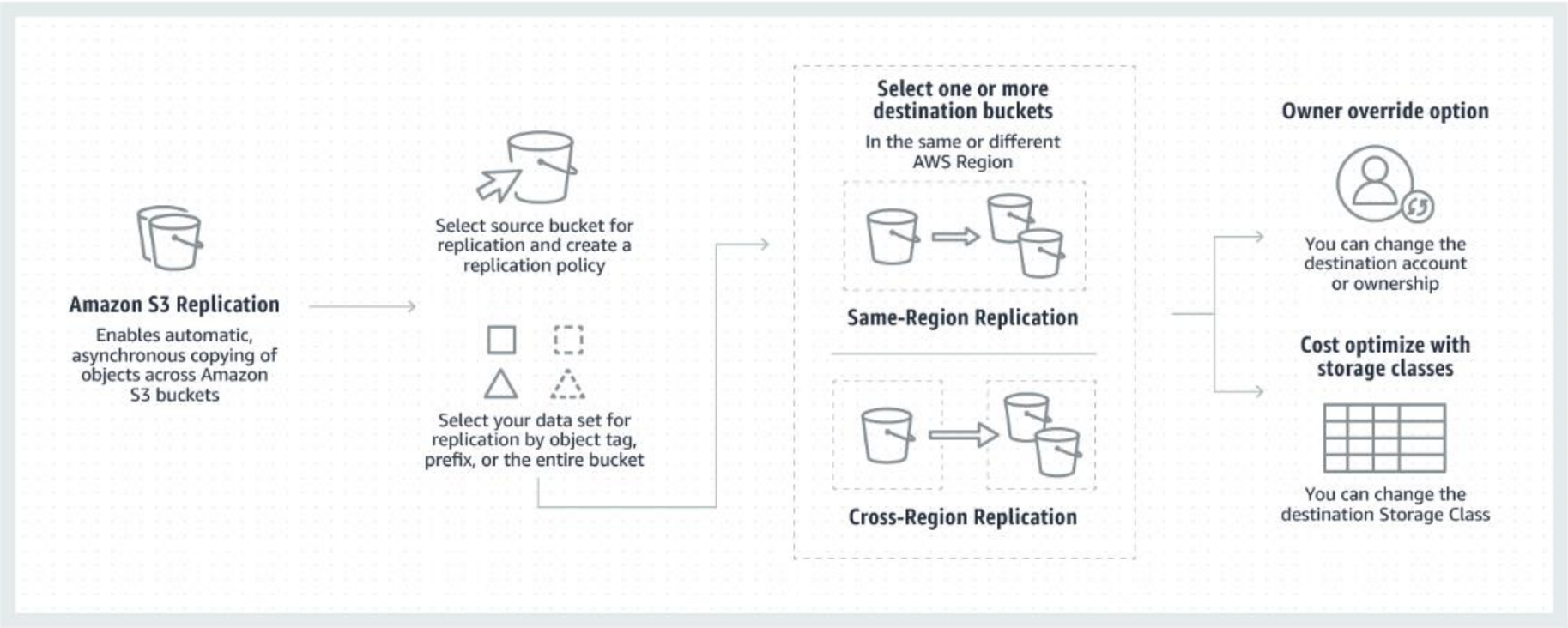
D. S3 Intelligent-Tiering

TECHNOLOGY

Amazon S3 Replication

Amazon S3 Replication

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



Working of S3 Replication

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

S3 Replication Use Cases

These are the various use cases of S3 replication:

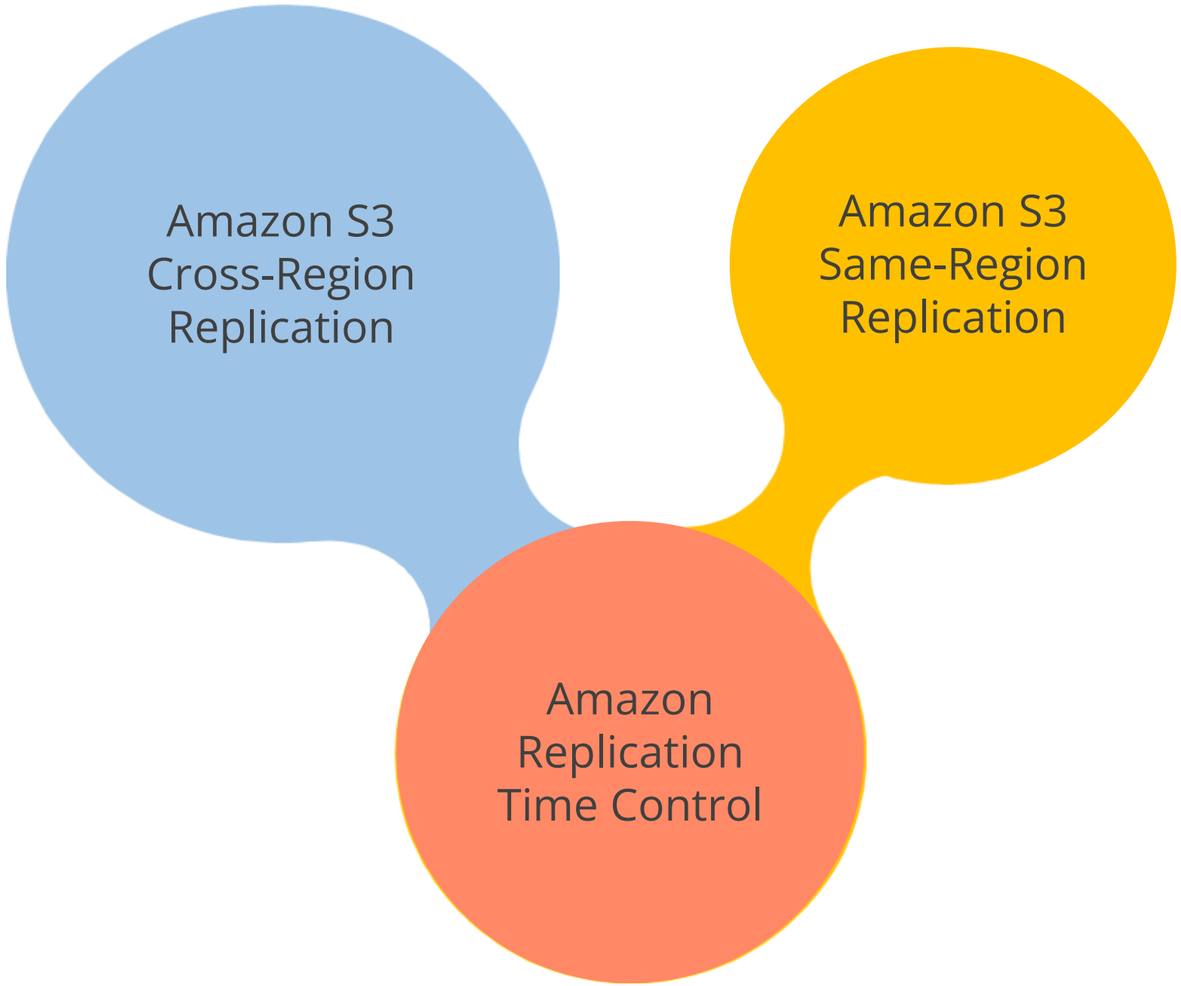
Replicate objects while retaining metadata

Replicate objects to more cost-effective storage classes

Urgent requirement of replicated buckets

Maintain object copies under different ownership

S3 Replications: Types



Types of S3 Replications

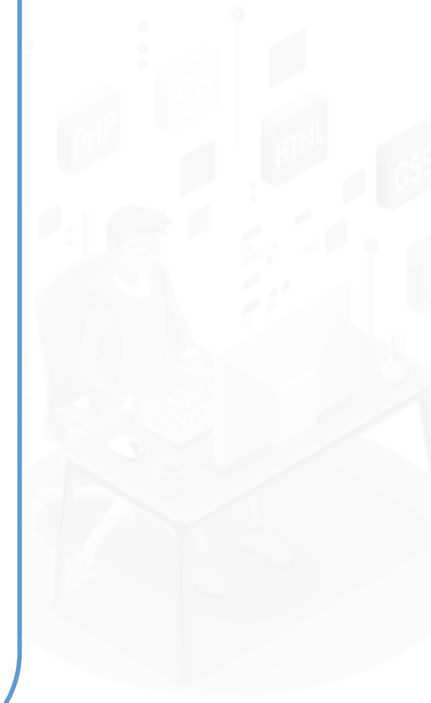
Amazon S3 Cross-Region Replication (CRR)

- Replicates data automatically between buckets across different AWS Regions
- Configures 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)

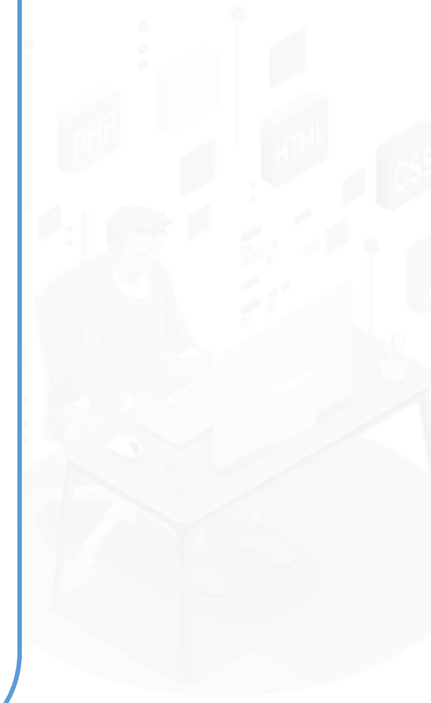
- Replicates data automatically between buckets within the same AWS Region
- Replicates data at the bucket level, a shared prefix level, or an object level using S3 object tags
- Addresses data sovereignty and compliance requirements
- Changes account ownership for the replicated objects to protect against 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:
  I. The total number of S3 API operations that are pending replication
  II. The total size of objects pending replication
  III. The maximum replication time to the destination AWS Region



Implementing Object Replication Between S3 Buckets

Duration:10 min

Problem Statement:

You have been assigned a task to demonstrate the replication of an object from a source to a destination S3 bucket using replication rules.

Outcome:

You will be able to successfully configure and execute object replication between S3 buckets using defined replication rules.

Note: Refer to the demo document for detailed steps:
05_Object_Replication_Between_S3_Buckets

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

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 to destination bucket
5. Delete and verify the replicated object file

Quick Check

You need to ensure that objects uploaded to your S3 bucket in one region are automatically copied to a bucket in another region for disaster recovery purposes. Which S3 feature should you use?

A. S3 Versioning

B. S3 Cross-Region Replication (CRR)

C. S3 Transfer Acceleration

D. S3 Lifecycle Management

S3 Encryption

Server-Side Encryption (SSE-S3)

Server-side encryption protects data by encrypting each object using a unique key.

Server-Side Encryption on Amazon S3

- Amazon S3 uses the 256-bit Advanced Encryption Standard (AES-256), one of the strongest block ciphers, to encrypt data.
- Users can use server-side encryption with Amazon S3-managed keys (SSE-S3) to incur no additional charges.
- Users can use a bucket policy to encrypt all objects in a bucket.

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

SSE-KMS encrypts data using AWS KMS keys, enhancing the security of the data.

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

AWS Encryption SDK

Some of the AWS SDKs that 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++

Demonstrating Server-Side Encryption Using S3 and KMS

Duration:10 min

Problem Statement:

You have been assigned a task to demonstrate the utilization of Amazon S3 buckets with different server-side encryption options: SSE-S3 and SSE-KMS.

Outcome:

You will be able to successfully demonstrate the implementation of server-side encryption for Amazon S3 buckets using both SSE-S3 and SSE-KMS options.

Note: Refer to the demo document for detailed steps:
06_Demonstrating_Server_Side_Encryption_Using_S3_and_KMS

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Create an S3 bucket with SSE-S3 encryption
2. Create a Key Management Service (KMS) key
3. Create an S3 bucket with SSE-KMS encryption

Quick Check

You want to encrypt your data at rest in S3 using a key managed by AWS, and you also need detailed audit trails of key usage. Which encryption method should you use?

A. Client-side Encryption with your own keys

B. Server-side Encryption with S3-Managed Keys (SSE-S3)

C. Server-side Encryption with Customer-Provided Keys (SSE-C)

D. Server-side Encryption with AWS KMS-Managed Keys (SSE-KMS)

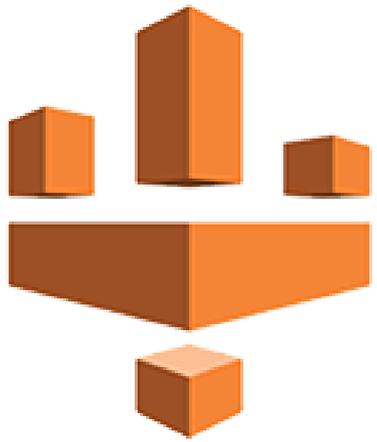
Unauthorized Link Access

Unauthorized Link Access

When a user attempts to use Amazon S3 to host a static website, there is a possibility of encountering an **Access Denied** error.

AWS Glue Data

AWS Glue is a serverless data integration service that simplifies the process of discovering, preparing, and combining data for various purposes, such as analytics, machine learning, and application development.



AWS Glue Data: Features

Faster data integration

Accelerated data integration enables diverse teams within the organization to collaboratively use AWS Glue, streamlining data tasks and expediting data analysis.

Data integration automation at scale

Utilize AWS Glue for effortless management of countless ETL jobs and data consolidation across various stores via SQL.

Serverless

AWS Glue operates serverlessly, handling resource provisioning, configuration, and scaling for your data integration tasks.

Building a Glue Data Catalog

Duration:15 min

Problem Statement:

You have been assigned a task to create a Glue Data Catalog using AWS Glue for seamless organization and efficient cataloging.

Outcome:

You will be able to successfully create a Glue Data Catalog in AWS Glue, facilitating seamless organization and efficient data cataloging.

Note: Refer to the demo document for detailed steps:
09_Building_a_Glue_Data_Catalog

ASSISTED PRACTICE

Assisted Practice: Guidelines

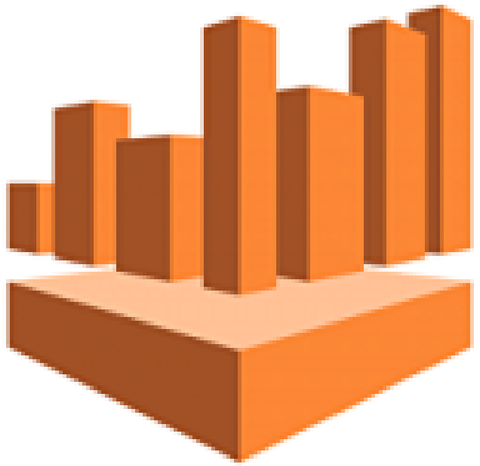
Steps to be followed are:

1. Create a VPC endpoint
2. Create a Glue Data Catalog
3. Add a data source to crawlers

Amazon Athena

Amazon Athena

Amazon Athena is an interactive query service that enables easy data analysis of information stored in Amazon S3 using standard SQL.

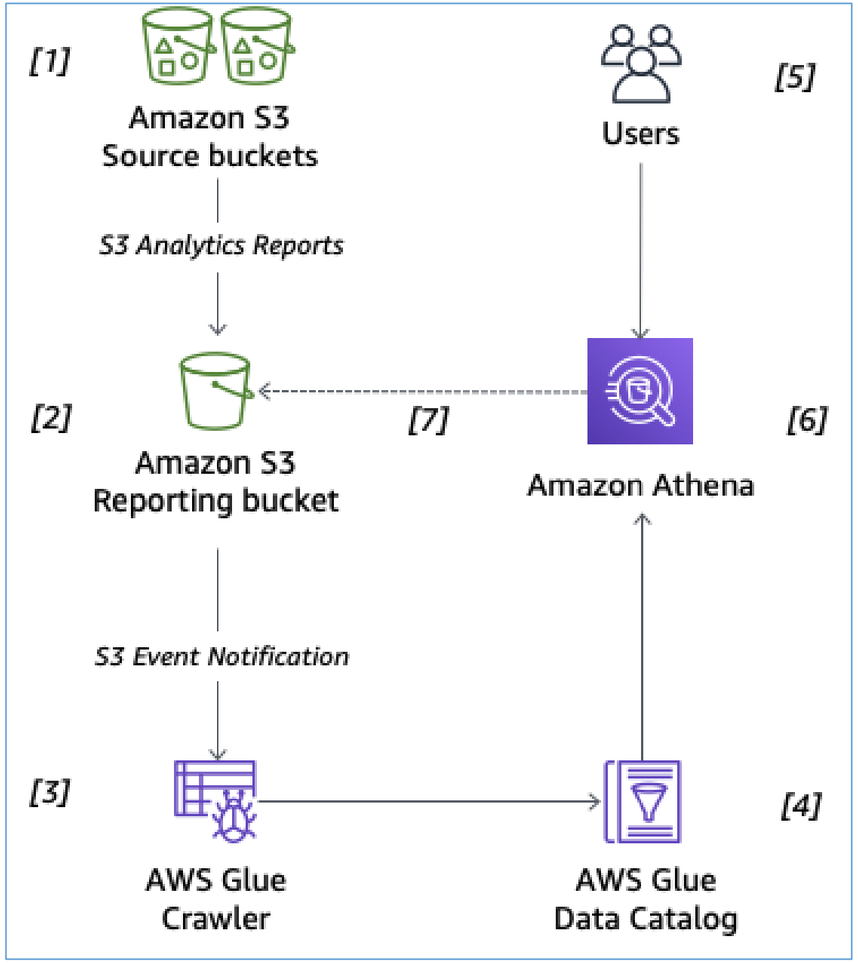


Amazon Athena

It operates in a serverless manner, eliminating the need for users to manage any underlying infrastructure.

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/

Querying S3 Bucket Operations with Athena

Duration:15 min

Problem Statement:

You have been assigned a task to execute a query within AWS Athena for performing operations on a designated S3 bucket.

Outcome:

You will be able to successfully execute a query within AWS Athena to analyze and manage operations on a designated S3 bucket.

Note: Refer to the demo document for detailed steps:
09_Querying_S3_Bucket_Operations_with_Athena

ASSISTED PRACTICE

Assisted Practice: Guidelines

Steps to be followed are:

1. Create S3 buckets
2. Execute queries in Athena

Quick Check

You need to run SQL queries on your data stored in Amazon S3 without setting up any servers or data warehouses. Which AWS service should you use?

A. Amazon Redshift

B. Amazon RDS

C. Amazon Athena

D. Amazon EMR

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 in an S3 bucket using the 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.

Deploying Static Website with S3 Buckets

Duration: 30 mins

Project agenda: To demonstrate the implementation of S3 buckets for hosting static website content

Description: The organizational admin's goal is to deliver HTML content through a static website. This objective is achieved by utilizing the capabilities of an S3 bucket for seamless content deployment.

Perform the following:

1. Create an S3 bucket
2. Generate a bucket Policy
3. Enabling static website hosting
4. Test and verify the hosted website content

LESSON-END PROJECT

TECHNOLOGY

Thank You

simplilearn