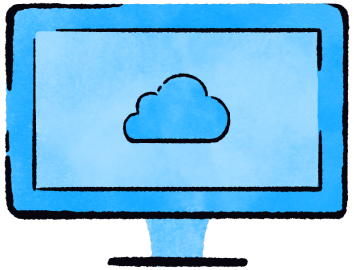
Amazon Simple Storage Service (S3): Overview and Introduction

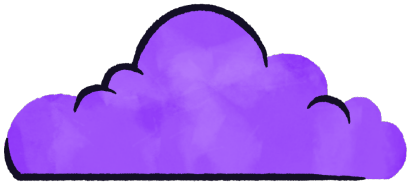
Andru Estes
Principal Author
andru-estes

What Is Amazon Simple Storage Service (S3)?

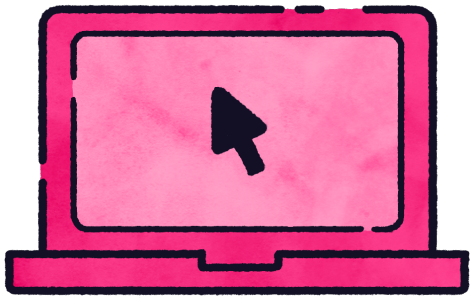
Amazon S3 Overview



S3 provides secure, durable, highly scalable object storage

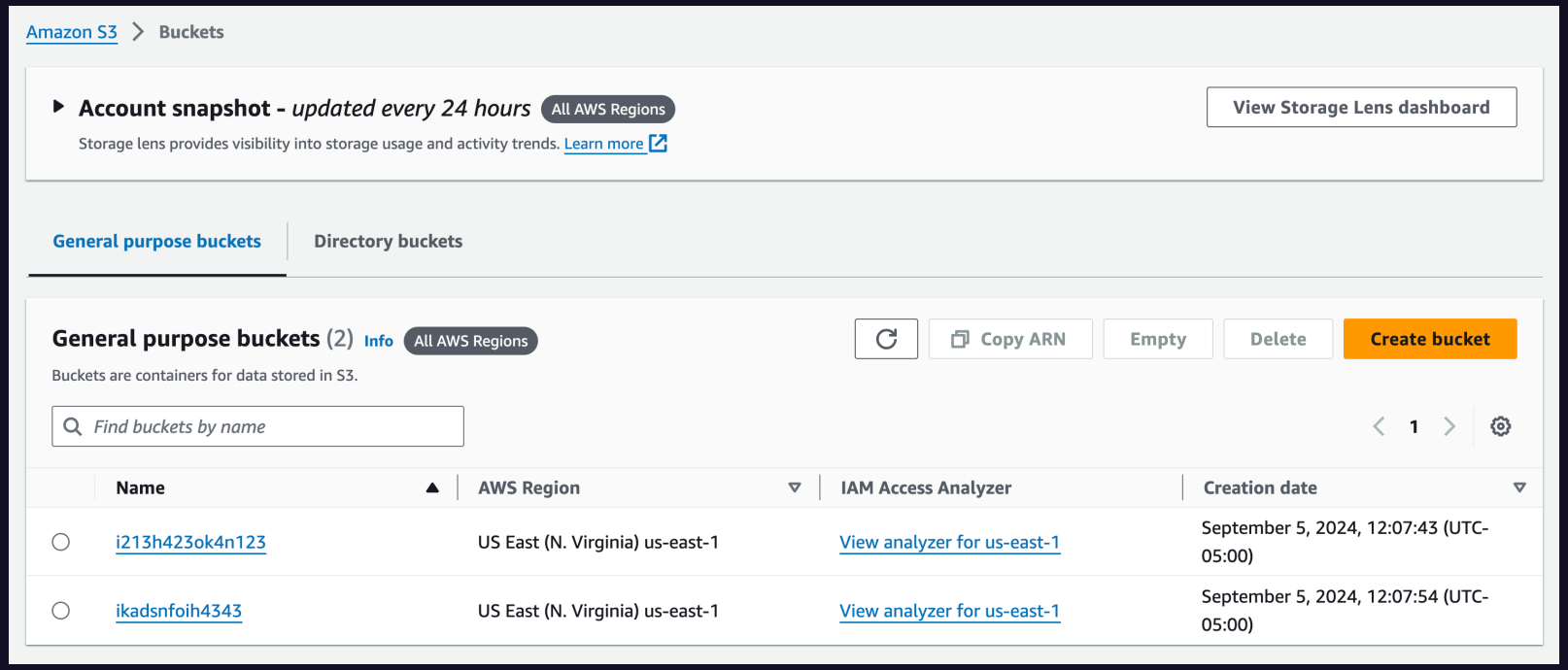


Store and retrieve any amount of data from anywhere at a very low cost



Easy to use, with a simple web service interface

Amazon S3 Console Example



Simple and easy to use console to browse your resources!

Amazon Simple Storage Service (S3) Concepts

S3 is a global service that deploys Regional resources

Store unlimited volume of data and number of objects (*scalable*)

Objects can be up to a maximum of 5 terabytes in size

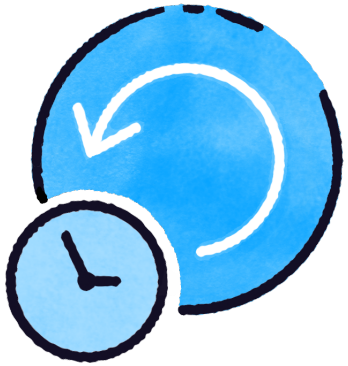
Cannot be used to run an operating system or database

Great for backups, archiving, media hosting, data lakes, and static websites

You can upload any file type that you can think of to Amazon S3.

Files like photos, company documents, video files, log files, ZIP archives, etc.

Amazon S3 Durability and Availability



Availability

Built for 99.95% – 99.99% service availability, depending on the S3 tier

Durability

Designed for 99.999999999% durability for data stored in S3 (eleven 9s)

Amazon S3 Buckets

Amazon S3 Bucket Concepts

Buckets are the containers for storing all objects within S3

All AWS accounts share the S3 namespace, so each bucket name must be globally unique

Even though bucket names are global, the buckets themselves are Regional

How Safe Is Your Data?

S3 is a safe place to store your files

S3 Standard storage class redundantly stores objects on multiple devices across a minimum of three Availability Zones in a Region

The service handles concurrent device failures by quickly detecting and repairing any lost redundancy

Image Source: https://unsplash.com/

Strong Read-After-Write Consistency

Read-After-Write

After a write or overwrite of an object (PUT), any subsequent read request immediately receives the latest object

Strong Consistency

For list operations, you can immediately perform a listing of the objects in a bucket with all changes reflected

Amazon S3 Bucket Naming Convention Requirements

Bucket names must be between **3** and **63** characters long.
Bucket names can consist only of lowercase letters, numbers, dots (.), and hyphens (-).
Bucket names must begin and end with a letter or number.
Bucket names must not contain two adjacent periods.
Bucket names must not be formatted as an IP address (**192.168.5.4**).
Buckets used with Amazon S3 Transfer Acceleration can't have dots (.) in their names.

Amazon S3 Bucket Naming Convention Requirements

Bucket names must not start with the prefix **xn--**.
Bucket names must not start with the prefix **sthree-**.
Bucket names must not start with the prefix **sthree-configurator**.
Bucket names must not start with the prefix **amzn-s3-demo-**.
Bucket names must not end with the suffix **-s3alias**. This suffix is reserved for access point alias names.
Bucket names must not end with the suffix **--ol-s3**. This suffix is reserved for Object Lambda Access Point alias names.
Bucket names must not end with the suffix **.mrap**. This suffix is reserved for Multi-Region Access Point names.
Bucket names must not end with the suffix **--x-s3**. This suffix is reserved for directory buckets.

When you upload a file to an S3 bucket, you will receive an HTTP 200 code if the upload was successful.

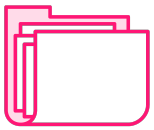
Amazon S3 Objects

S3 objects are your files (data) that you upload to your Amazon S3 buckets.

Objects are stored in your buckets and offer a flat path, similar to, but not the same as, a typical folder structure.

S3 Objects Concepts

Each object has a Key (*name*) and a Value (*object data*)



Prefixes are meant to serve as human-friendly naming structures

An object Key is the full path, including any prefixes

Objects can have multiple versions and metadata (*content-type*, *last-modified*)

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

S3 URI

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

The S3 bucket name

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

Prefix

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

Another prefix

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

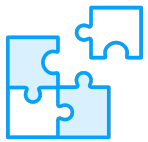
The object name

S3 Object Key Example

s3://pluralsight-bucket/dev/2024/test_data.json

This is the entire Key for the object

Amazon S3 Multipart Upload



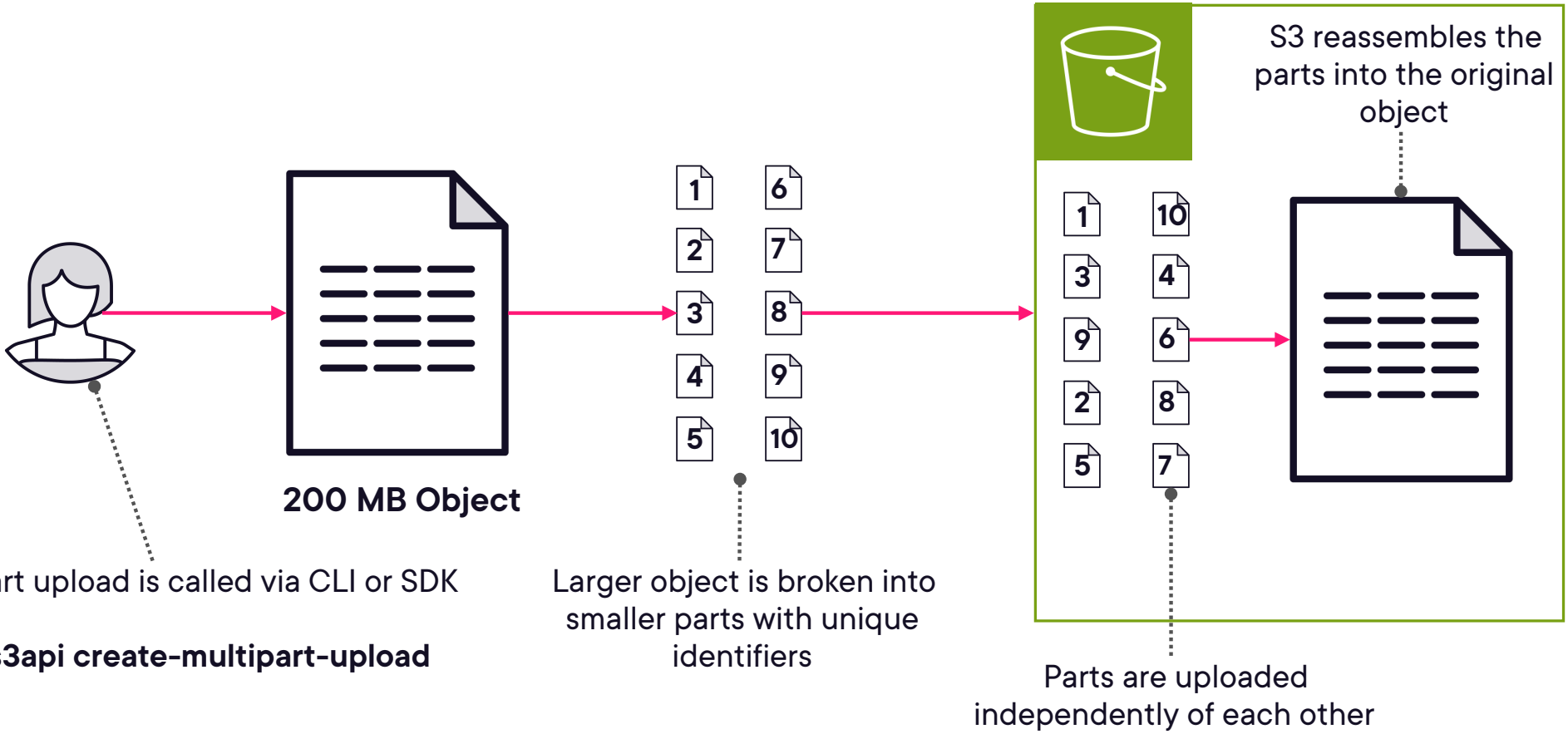
Feature that allows users to upload large objects in smaller, more manageable parts

Object parts are uploaded independently of one another, and if any part fails, you can retry the upload without affecting other parts

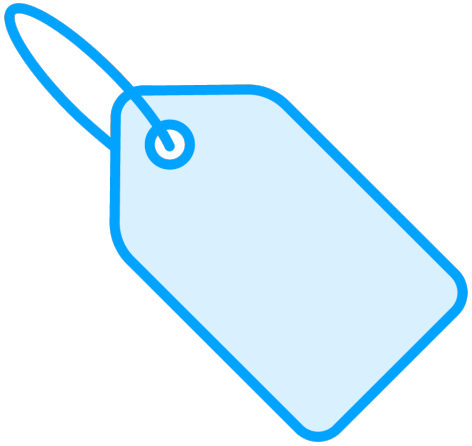
Use Case 1: Uploading very large files, especially objects over 5GB in size

Use Case 2: Uploading objects over an unstable or poor network connection

Multipart Upload Diagram

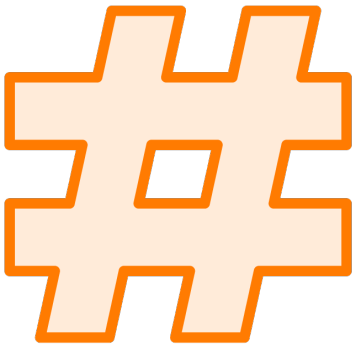


Object Tagging and Metadata



Tagging

Easily tag your objects using key-value pairs of useful information



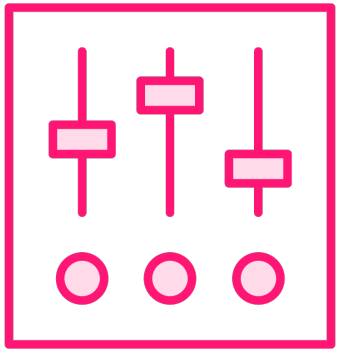
Metadata

Objects contain data about the object data you are storing (e.g., content-type, last-modified, etc.)

Amazon S3 Storage Classes

S3 Storage Classes Overview

Each object has a storage class associated with it



There are many storage classes for the objects that you store

Depends on your use case and performance access requirements

All of the classes offer high durability

S3 Storage Classes for Frequently Accessed Objects

S3 Standard

Default class that offers the default SLA and pricing (*99.99% availability & 99.999999999% durability*)

Express One Zone

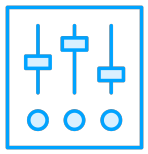
High-performance, single-zone storage class offering single-digit millisecond data access

S3 Standard is suitable for most workloads.

It is best for general purpose workloads.

S3 Standard use cases include websites, content distribution, mobile/gaming applications, and big data analytics storage.

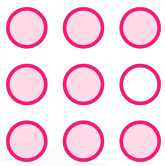
S3 Storage Classes for Infrequently Accessed Objects



S3 offers classes designed for long-lived and infrequently accessed data

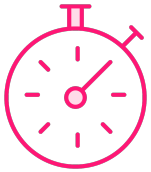
Keep in mind that there is a retrieval fee for this data, so it is best used for things like backups and older data

S3 Standard-IA: Stores objects redundantly across multiple AZs and could require rapid access when it is needed. Good for backups.

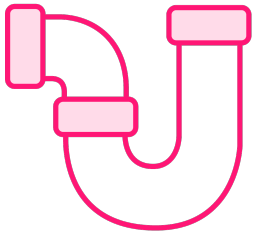


S3 One Zone-IA: Stores objects redundantly in one AZ, making it less expensive, but also less resilient. ~20% cheaper than S3 Standard-IA.

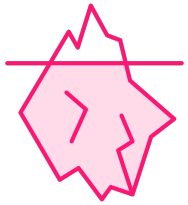
S3 Storage Classes for Rarely Accessed Objects



S3 Glacier Instant Retrieval: Use for long-term data that's rarely accessed and requires milliseconds retrieval. Allows for real-time access if needed.



S3 Glacier Flexible Retrieval: Archives where portions of the data might need to be retrieved in minutes. Not available for real-time access.



S3 Glacier Deep Archive: Use for archiving data that rarely needs to be accessed. Not available for real-time access. Up to 48 hours to retrieve.

S3 Storage Classes for Rarely Accessed Objects

These storage classes come with a minimum storage duration requirement!

S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
90 days	90 days	180 days

Don’t know if you will be accessing your data frequently or infrequently?

S3 Storage Classes for Unknown Access Patterns

S3 Intelligent-Tiering is used to optimize storage costs by automatically moving data to the most cost-effective access tier with no performance impacts.

No retrieval fees for using this feature, but you pay for monitoring and auto-tiering.

Frequent Access

The default tier when using S3 Intelligent-Tiering

Infrequent Access

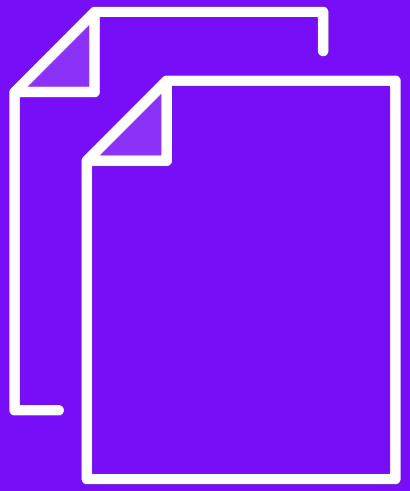
Objects that have not been accessed in 30 consecutive days

Archive Instant Access

Objects that have not been accessed for 90 consecutive days

Each one of the discussed classes is a valid cost-saving approach! It comes down to the specific requirements in the scenarios.

Amazon S3 Versioning



Amazon S3 Versioning

You can enable versioning in S3 buckets, so you can have multiple versions of an object within S3.

S3 Bucket States

Once versioning has been enabled, you cannot set it back to unversioned. You cannot turn it off!

You can only suspend versioning!

S3 Versioning Concepts

Enabled at the bucket level, not the object level

All writes and deletes of an object are marked as versions

General best practice for protection against accidental deletions

Objects present before it is enabled have a “null” version

Can be integrated with lifecycle rules and supports MFA

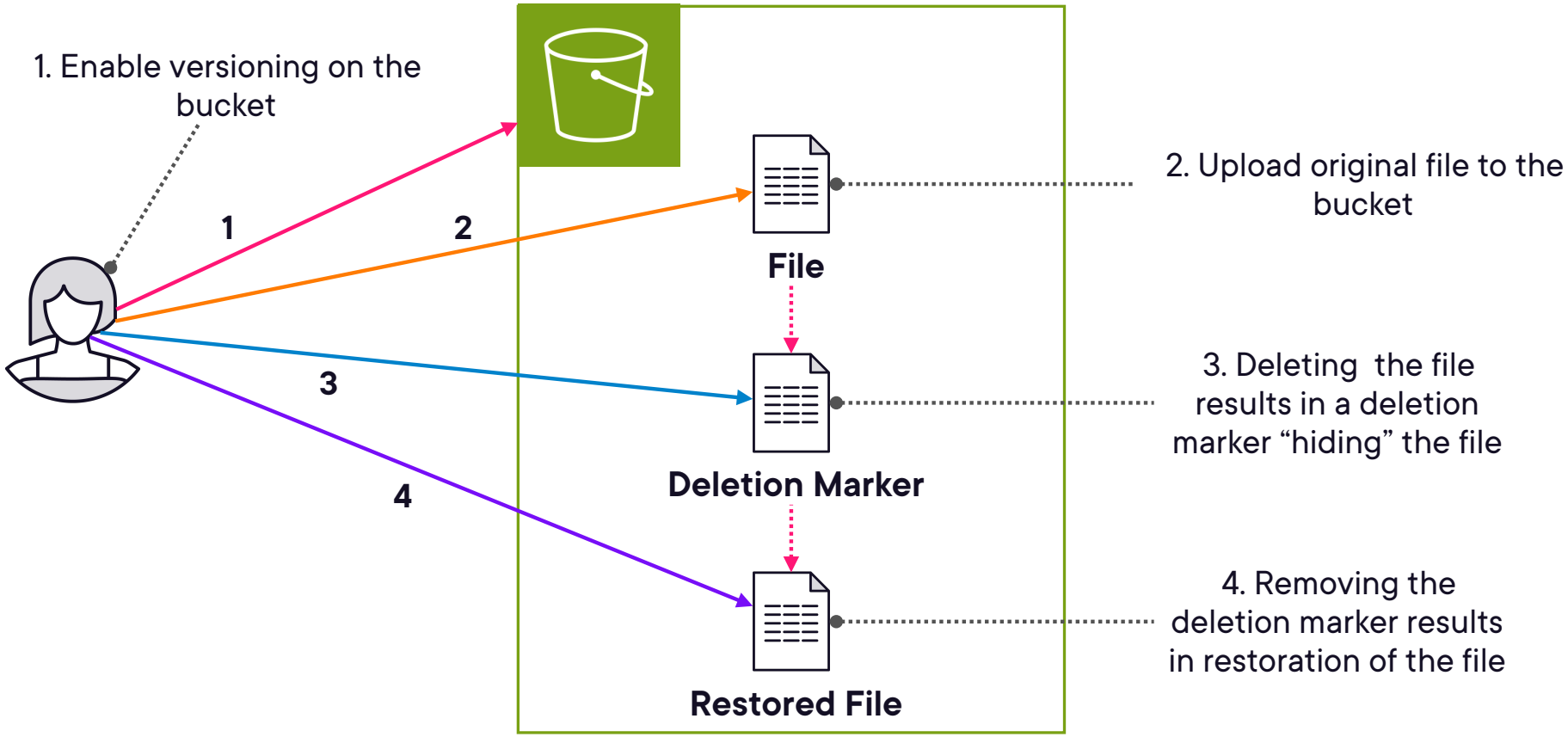
Remember that you only suspend versioning

Amazon S3 Deletion Markers

Instead of deleting objects, Amazon S3 inserts a delete marker, which becomes the object version

To restore the object to the previous version after deletion, you simply remove the delete marker

Demo: Enabling S3 Versioning and Recovering a Deleted File



Amazon S3 Object Lifecycles

You can manually move your objects between different storage classes.

Lifecycle configurations automate moving your objects between the different storage tiers, helping to maximize cost effectiveness.

Amazon S3 Lifecycle Rules

Rules defined to automatically manage the lifecycle of objects in a bucket based on your specified criteria.

This offers Transition Actions and Expiration Actions

Examples:

- Transition objects from S3 Standard to S3 Infrequent Access after 30 days
- Expire objects after 366 days

Rules can be used for specific object names, prefixes, and even object tags

Image Source: https://unsplash.com/

Lifecycle rules work for previous/noncurrent versions of objects as well.

This means you can transition versions of objects, and you can expire versions of objects.

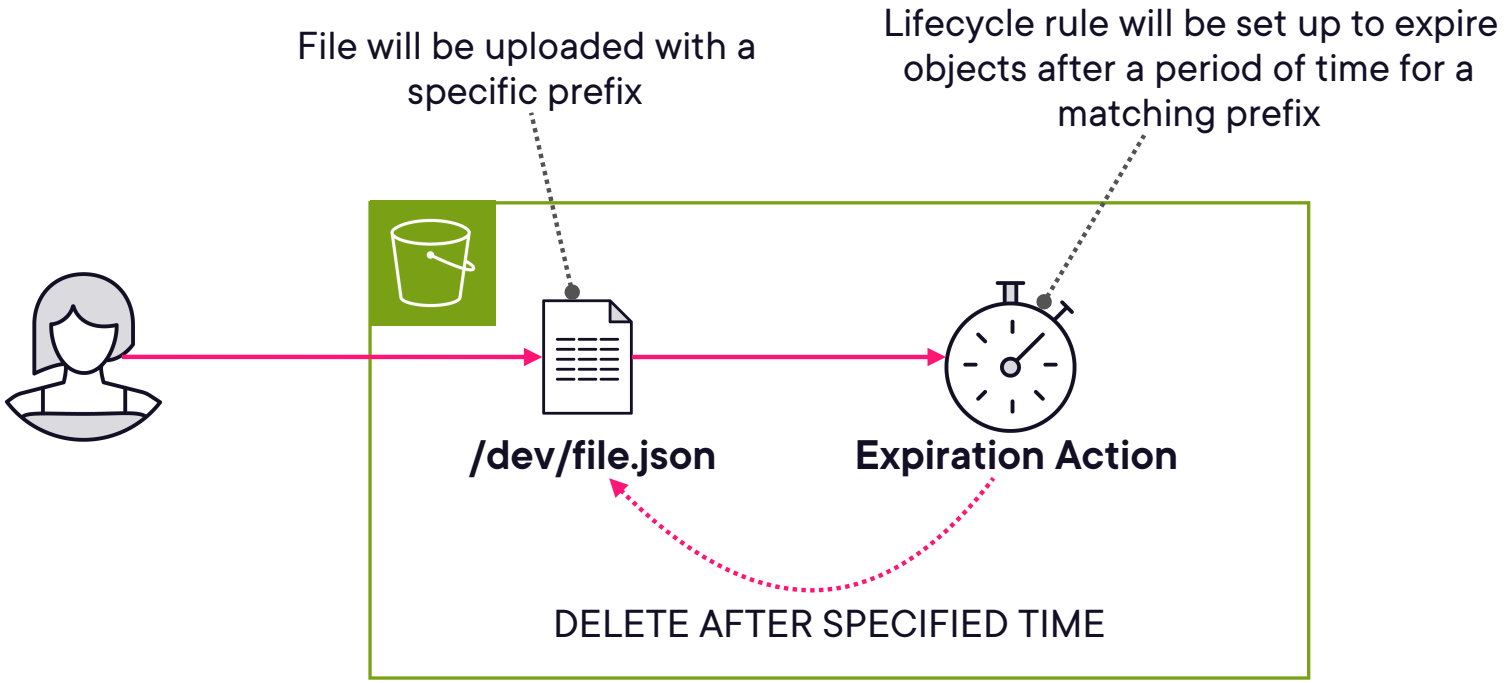
Three Scenarios for Lifecycle Rules

Transitioning noncurrent versions of objects to S3 Standard IA

Transition compliance documents to S3 Glacier Deep Archive after 365 days

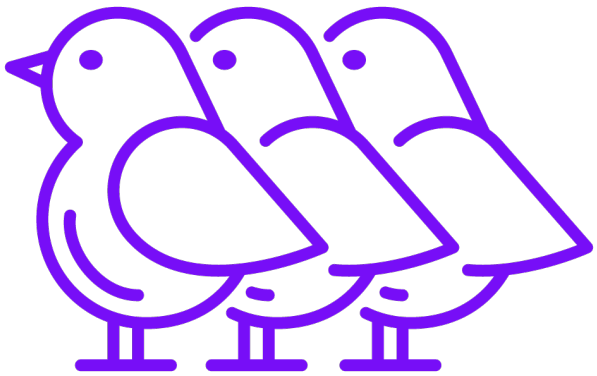
Expire development data 30 days after the object was created

Demo: Transitioning Objects Using Lifecycle Policies



Amazon S3 Bucket Replication

S3 Replication Overview



You can replicate objects from one bucket to another bucket

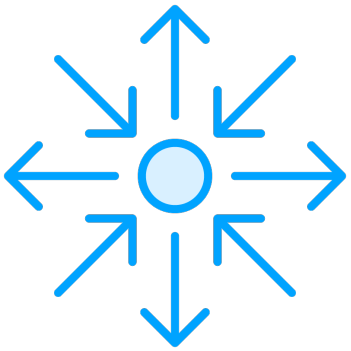
Versioning must be enabled on both the source and destination buckets

Existing objects in a bucket are not replicated automatically

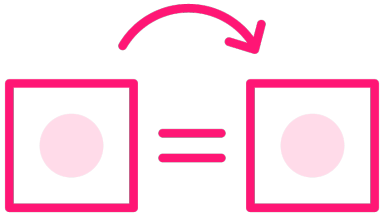
Once replication is turned on, all subsequent updated objects will be replicated automatically

Deleting individual versions and delete markers will not be replicated

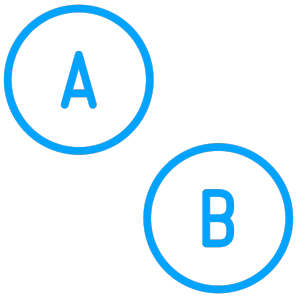
S3 Replication Concepts



Cross-Region Replication



Same-Region Replication



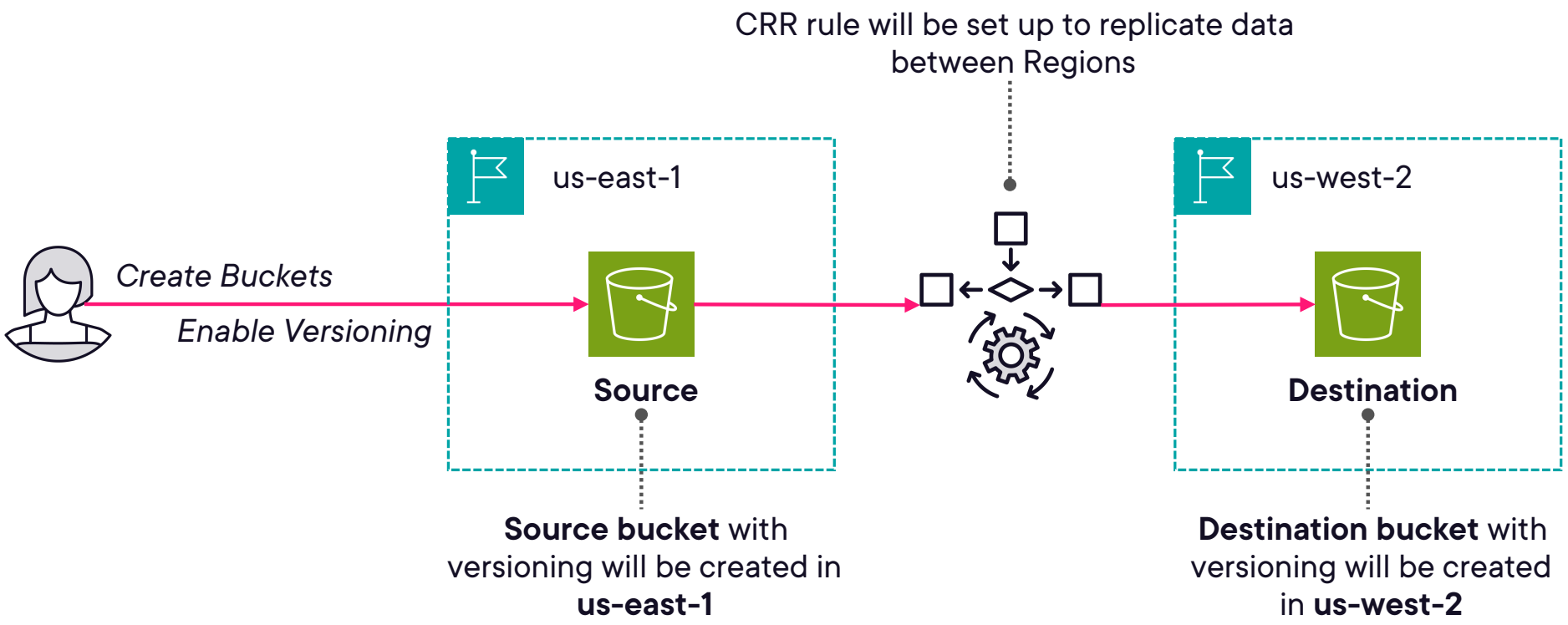
Buckets do not have to be in the same account

IAM permissions are required for S3 replication to work

S3 Replication is an asynchronous process!

This means objects will not be immediately available in the destination bucket.

Demo: Implement Cross-Region Replication in S3



Module Summary and Exam Tips

Amazon S3 Exam Tips

Object-based storage service that allows you to upload an unlimited amount of files/data

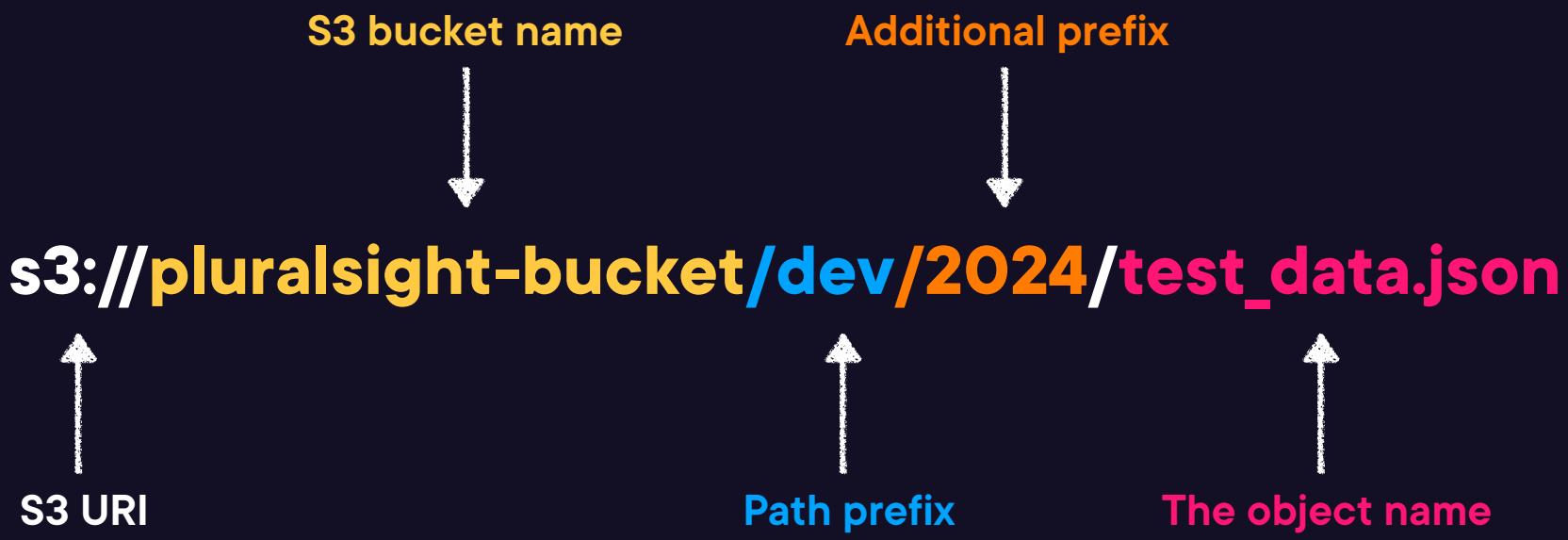
Built for **99.95% – 99.99%** service availability, depending on the S3 tier

Designed for **99.999999999%** durability for data stored in S3 (*eleven 9s*)

Files can be from 0 bytes to 5 TB

Not suitable to install an operating system or run a database on

Reviewing Object Keys



Reviewing Object Keys

Here are some possible prefix matches for this object:

1. *.json
2. /dev/*
3. /dev/2024/*

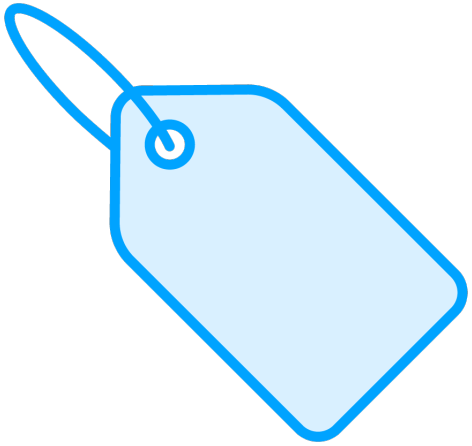
s3://pluralsight-bucket/dev/2024/test_data.json

The following tables are just for easy reference if needed in the future! We won't cover them in detail here.

	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%	99.9%	99.9%	99.9%
Availability Zone(s)	≥3	≥3	≥3	1	≥3	≥3	≥3
Min Capacity Charge per Object	N/A	N/A	128KB	128KB	128KB	40KB	40KB
Minimum Storage Duration Charge	N/A	30 days	30 days	30 days	90 days	90 days	180 days
Retrieval Fee	N/A	N/A	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved
Storage Type	Object	Object	Object	Object	Object	Object	Object
Lifecycle Transitions	Yes	Yes	Yes	Yes	Yes	Yes	Yes

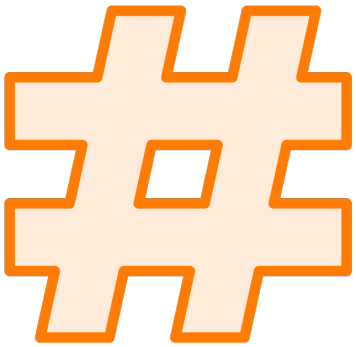
Storage Class	Availability and Durability	AZ(s)	Use Case
S3 Standard	99.99% Availability 11 9's Durability	≥3	Suitable for most workloads (e.g., websites, content distribution, mobile and gaming applications, and big data analytics).
S3 Standard-Infrequent Access	99.9% Availability 11 9's Durability	≥3	Long-term, infrequently accessed critical data (e.g., backups, data store for disaster recovery files, etc.).
S3 One Zone-Infrequent Access	99.5% Availability 11 9's Durability	1	Long-term, infrequently accessed, non-critical data.
S3 Intelligent-Tiering	99.9% Availability 11 9's Durability	≥3	Unknown or unpredictable access patterns.
S3 Glacier Instant Retrieval	99.9% Availability 11 9's Durability	≥3	Provides long-term data archiving with instant retrieval time for your data.
S3 Glacier Flexible Retrieval	99.99% Availability 11 9's Durability	≥3	Ideal storage class for archive data that does not require immediate access but needs the flexibility to retrieve large sets of data at no cost, such as backup or disaster recovery use cases. Can be minutes or up to 12 hours.
S3 Glacier Deep Archive	99.99% Availability 11 9's Durability	≥3	Cheapest storage class and designed for customers that retain datasets for 7-10 years or longer to meet customer needs and regulatory compliance requirements. The standard retrieval time is 12 hours, and the bulk retrieval time is 48 hours.

Reviewing Object Tagging and Metadata



Tagging

Easily tag your objects using key-value pairs of useful information



Metadata

Objects contain data about the object data you are storing (e.g., content-type, last-modified, etc.)

Remember the S3 Versioning Bucket States

You cannot disable versioning, you can only suspend it!

Objects present before versioning was enabled receive a “null” version.

Amazon S3 Versioning Exam Tips

Versioning can be integrated with lifecycle rules

Versioning can be set up to require and support multi-factor authentication

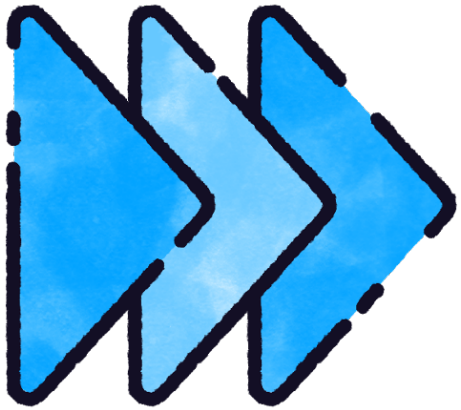
Amazon S3 Deletion Markers Exam Tips

Amazon S3 inserts a delete marker, which becomes the object version

To restore the object to the previous version, you simply remove the delete marker

Multipart uploads allow users to upload large objects in smaller, more manageable parts.

S3 Lifecycle Rules Exam Tips



Automates moving objects between different storage tiers, or expiration

Transition Actions and Expiration Actions

Can be used in conjunction with versioning and can be applied to current versions or previous versions

Exam scenario use cases:

- Transitioning noncurrent versions of objects to S3 Standard IA
- Transition compliance documents/logs to S3 Glacier Deep Archive after 365 days
- Expire development data 30 days after the object was created

S3 Replication Exam Tips

You can replicate objects from one bucket to another (*same or cross-Region*)

Already existing objects in buckets are not replicated automatically

Delete markers are not replicated by default