aws academy

CSCI 5902 Adv. Cloud Architecting
Fall 2023
Instructor: Lu Yang

Module 3 Adding a Storage Layer (Sec 1-3)
Sep 18, 2023

Housekeeping and feedback

1. First guided lab followed by Azure tutorial this Friday.
2. I will provide small treats to top 3 Kahoot players on Fridays.
3.

AWS Regions

- An AWS Region is a geographical area
- Each AWS Region consists of two or more Availability Zones
- Communication between Regions uses AWS backbone network infrastructure
- You enable and control data replication across Regions

never

Example: London Region

≥ 3 Az

3 – 6 Azs

AWS Academy Cloud Architecting

Module 3: Adding a Storage Layer

Module overview

Sections

1. The simplest architecture
2. Using Amazon S3
3. Storing data in Amazon S3
4. Moving data to and from Amazon S3
5. Choosing Regions for your architecture

Labs

- Guided Lab: Hosting a Static Website
- Challenge Lab: Creating a Static Website for the Café

Module objectives

At the end of this module, you should be able to:

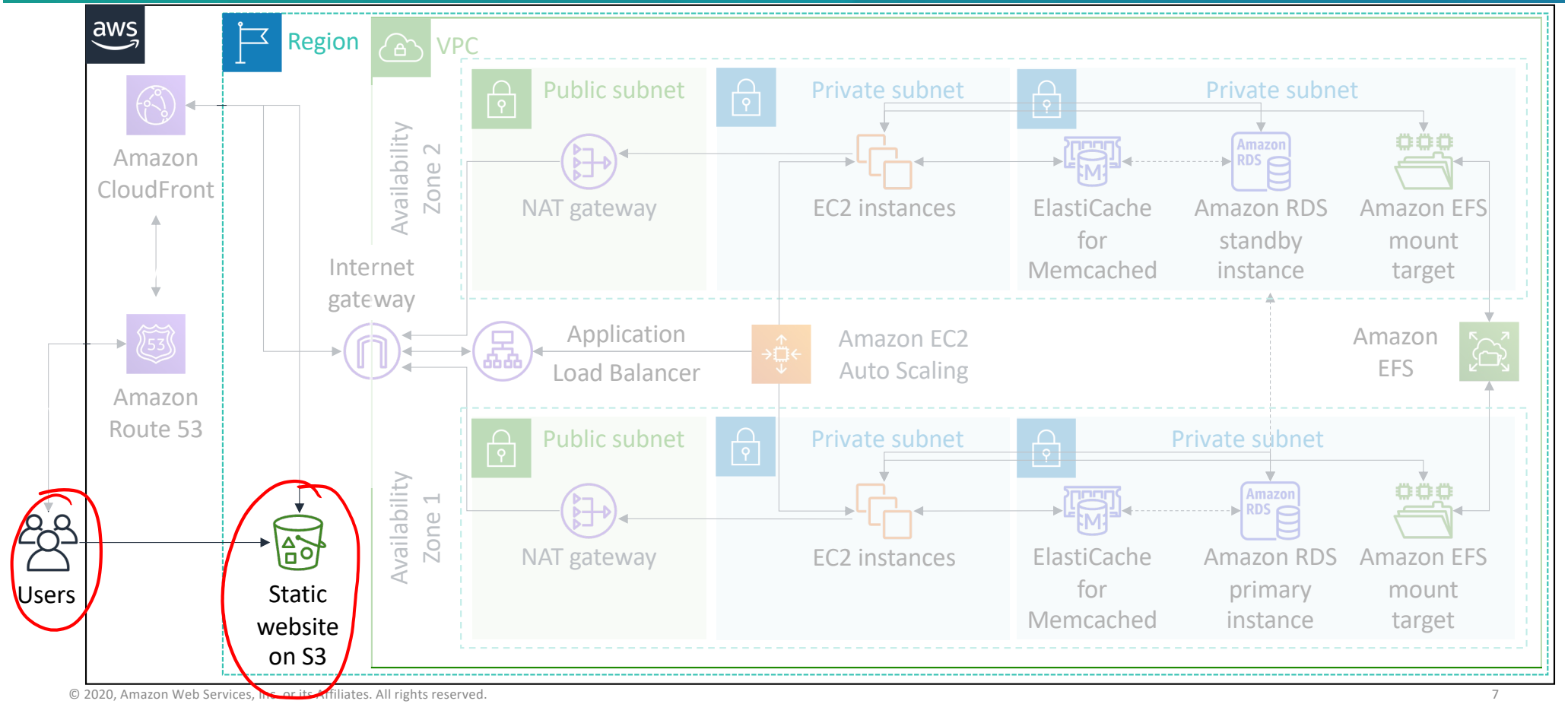
- Recognize the problems that Amazon Simple Storage Service (Amazon S3) can solve
- Describe how to store content efficiently using Amazon S3
- Recognize the various Amazon S3 storage classes and cost considerations
- Describe how to move data to and from Amazon S3
- Describe how to choose a Region
- Create a static website

Module 3: Adding a Storage Layer

Section 1: The simplest architecture

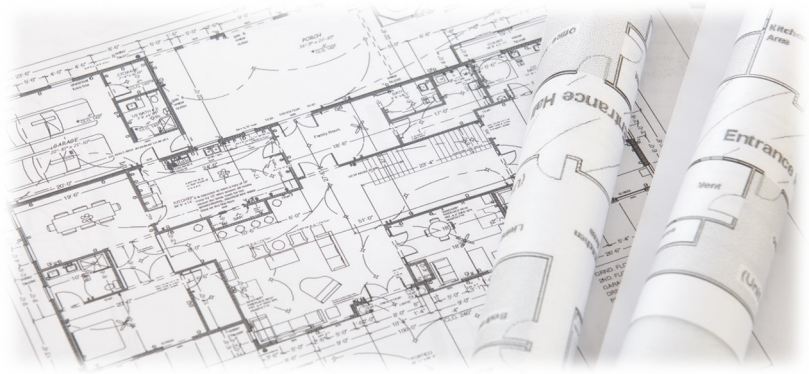
aws academy

Storage as part of a larger architecture



Café business requirement

The café has just started up. They want to establish a simple static website that provides customers with basic information about the café (including a menu, store hours, location, and more).



Module 3: Adding a Storage Layer

Section 2: Using Amazon S3

aws academy

Amazon S3

Amazon S3

An object storage service:

- It stores massive (unlimited) amounts of unstructured data
- Data files are stored as objects in a bucket that you define
- 5 TB is the maximum file size of a single object
- All objects have a REST-accessible globally unique URL (universal namespace)
- All objects have a key, version ID, value, metadata, and subresources

Questions:
Can you run an operating system or database on S3?
What is a bucket similar to?

Amazon S3 benefits

Durability

- It ensures data is not lost
- S3 Standard storage provides 11 9s (or 99.999999999%) of durability

11 9s of durability means that every year, there is a 0.000000001 percent chance of losing an object

Availability

- You can access your data when needed
- S3 Standard storage class is designed for four 9s (or 99.99%) availability

Data is stored redundantly across multiple devices in multiple facilities (>= AZs)

Scalability

- It offers virtually unlimited capacity
- Any single object of 5 TB or less

Security

- It offers fine-grained access control

Performance

- It is supported by many design patterns

Amazon S3 common usage patterns

Amazon S3

What problems can you solve by using Amazon S3?
You will now consider some use cases.

Amazon S3 use case 1: Store and distribute web content and media

Build a redundant, scalable, and highly available infrastructure that hosts video, photo, or music uploads and downloads.

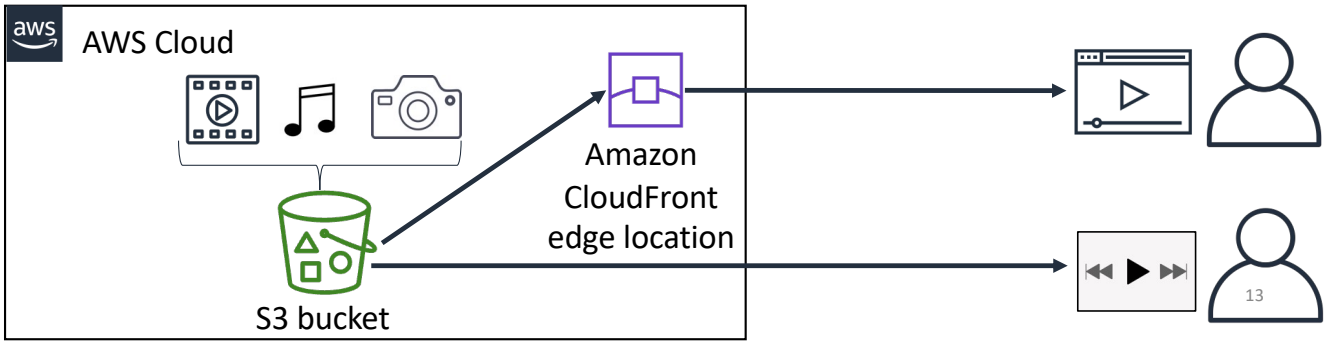
- Universal Namespace
- Example of URLs

  https://<bucket-name>.s3.amazonaws.com

  https://<bucket-name>.s3.amazonaws.com/video.mp4

- Uploading Files

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



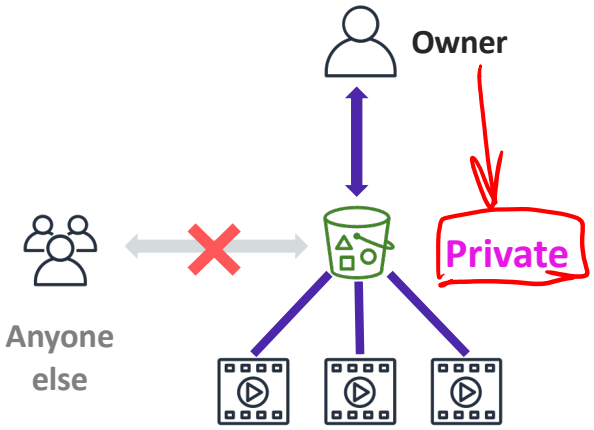
Securing Amazon S3 buckets and objects

- Newly created S3 buckets and objects are private and protected by default
- When use cases must share Amazon S3 data –
  - Manage and control the data access
  - Follow the principle of least privilege
- Tools and options for controlling access to Amazon S3 data –
  - Block Public Access feature: It is enabled on new buckets by default, simple to manage
  - IAM policies: A good option when the user can authenticate using IAM
  - Bucket policies: You can define access to a specific object or bucket
  - Access control lists (ACLs): A legacy access control mechanism
  - S3 Access Points: You can configure access with names and permissions specific to each application
  - Presigned URLs: You can grant time-limited access to others with temporary URLs
  - AWS Trusted Advisor bucket permission check: A free feature

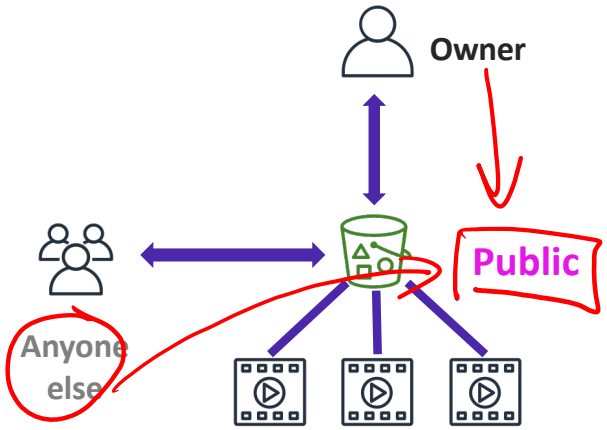
Three general approaches to configuring access

Configure the appropriate security settings for your use case on the bucket and objects.

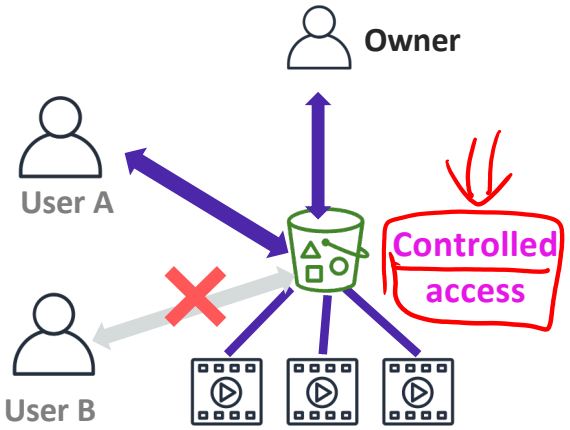
Default
Amazon S3 security settings



Public access
Amazon S3 security settings



Access policy applied to
Amazon S3 security settings



Consider encrypting objects in Amazon S3

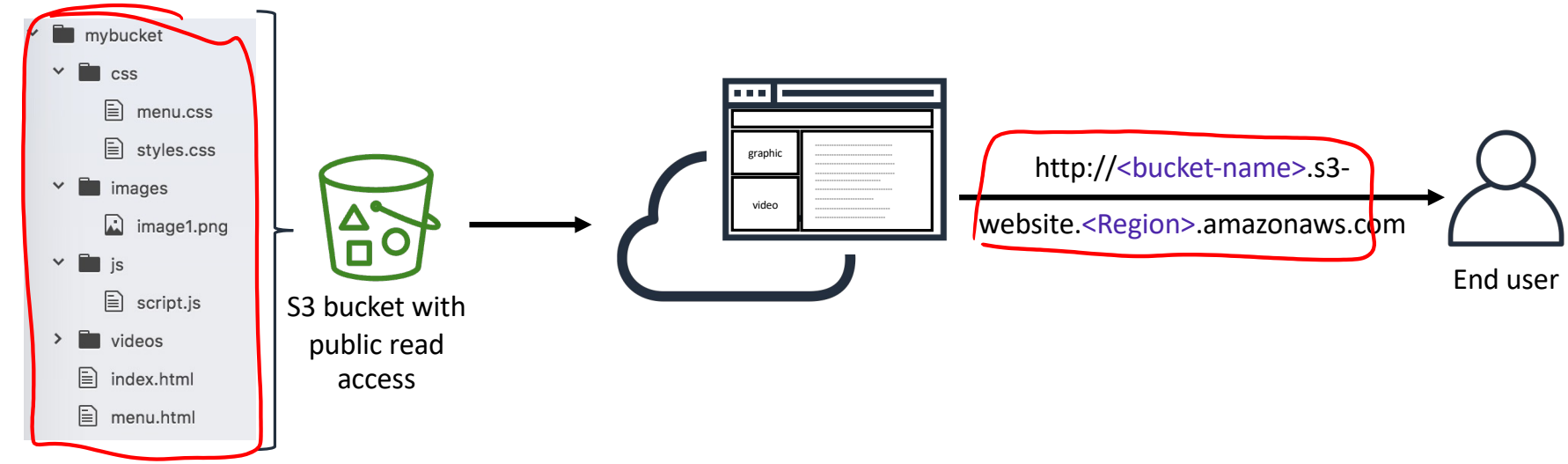
- Encryption encodes data with a secret key, which makes it unreadable
  - Only users who have the secret key can decode the data
  - Optionally, use AWS Key Management Service (AWS KMS) to manage secret keys

- Encryption of S3:
  - Encryption at Rest (Server-Side Encryption: SSE)
    - SSE-S3 (S3-managed keys, using AES 256-bit)
      - On the bucket, enable this feature by selecting the Default encryption option
      - Amazon S3 encrypts objects before it saves the objects to disk, and decrypts the objects when you download them
    - SSE-KMS (AWS Key Management Service-managed keys, external from S3)
    - SSE-C (Customer-provided keys)

  - Encryption at Rest (Client-Side Encryption: CSE)
    - Encrypt data on the client side and upload the encrypted data to Amazon S3
    - In this case, you manage the encryption process

  - Encryption in Transit
    - SSL/TLS
    - HTTPS

Amazon S3 use case 2: Host static websites



Example objects stored in the bucket configured for website hosting

Supports static content including HTML files, images, videos, and client-side scripts.

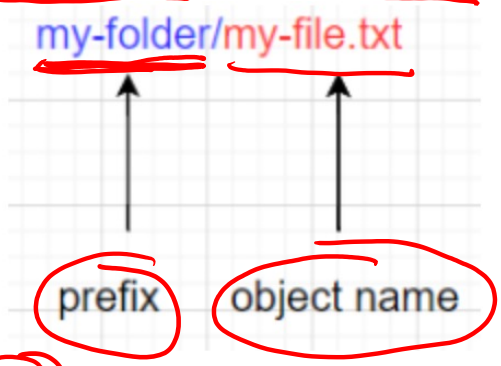
~~server side scripts~~

Amazon S3 use case 2: Host static websites

- Object prefix

  A prefix is a string of characters at the beginning of the object name.
- For example:

  s3://my-bucket/my-folder/my-file.txt



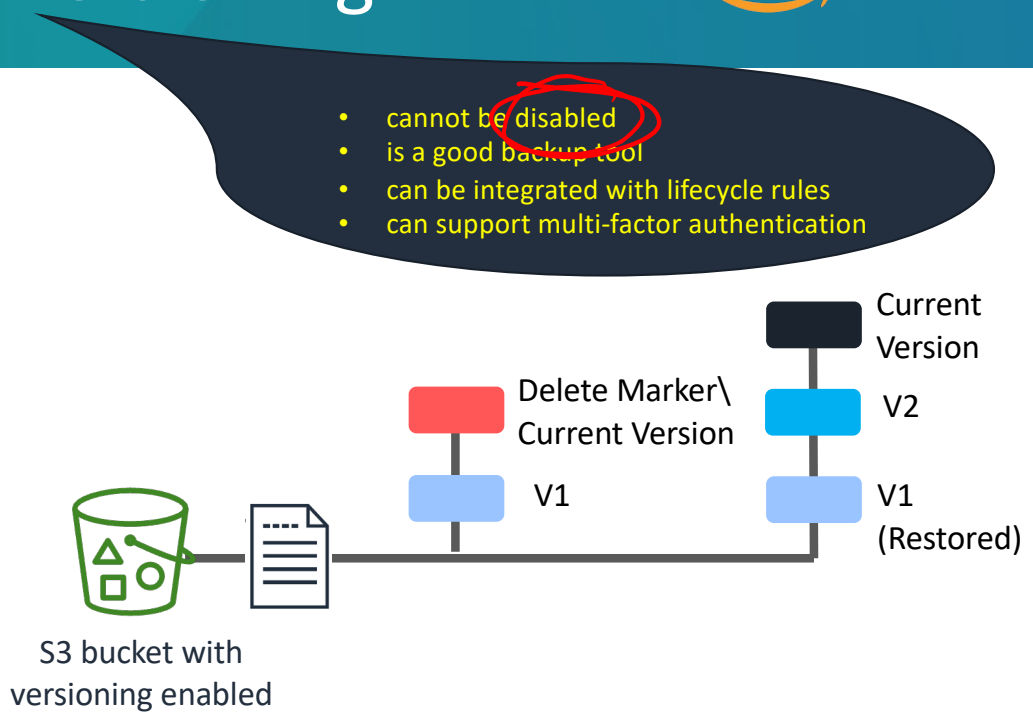
In simple words, key = prefix + object name

https://www.educative.io/answers/what-do-keys-mean-in-amazon-s3

Amazon S3 best practice: Versioning

- Protects against accidental overwrites and deletes with no performance penalty
- Generates a new version with every upload
- Enables easy retrieval of deleted objects or rollback to previous versions
- Three possible states of an S3 bucket –

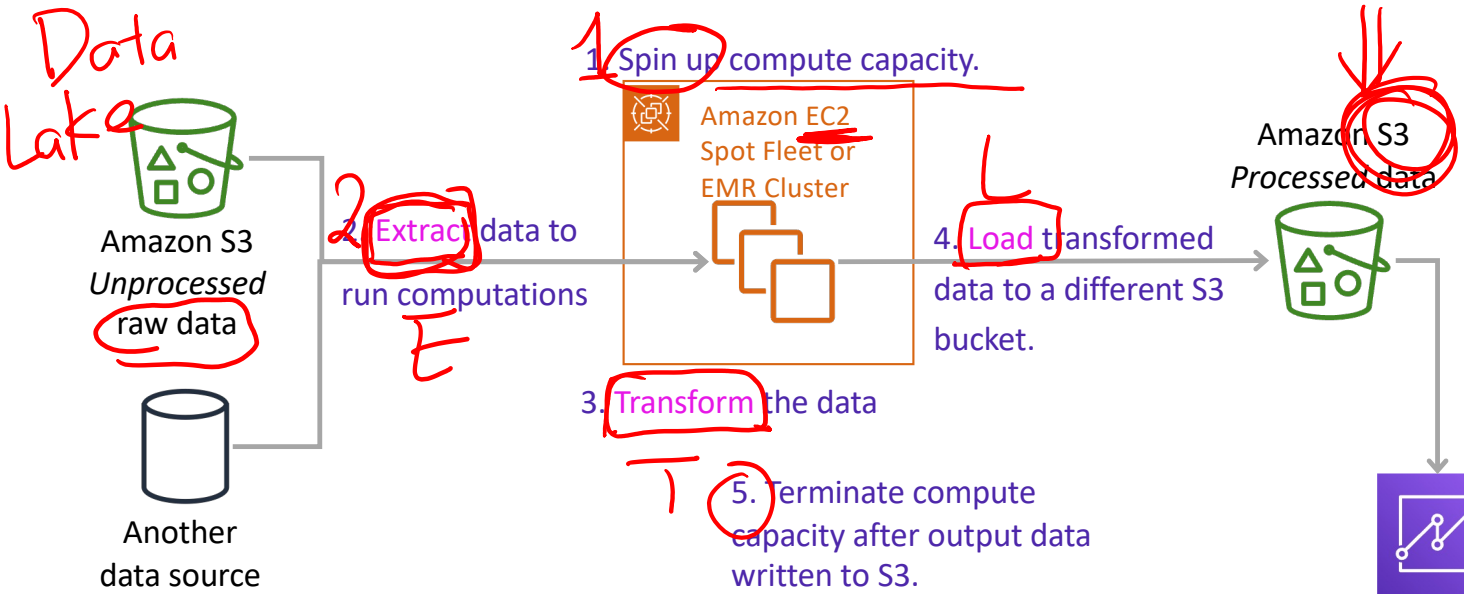
1. *Default*: Versioning not enabled
2. Versioning-enabled
3. Versioning-suspended



Amazon S3 use case 3: Data store for computation and analytics

Data store for computation and large-scale analytics

Example data integration and preparation pattern



Financial transaction analysis

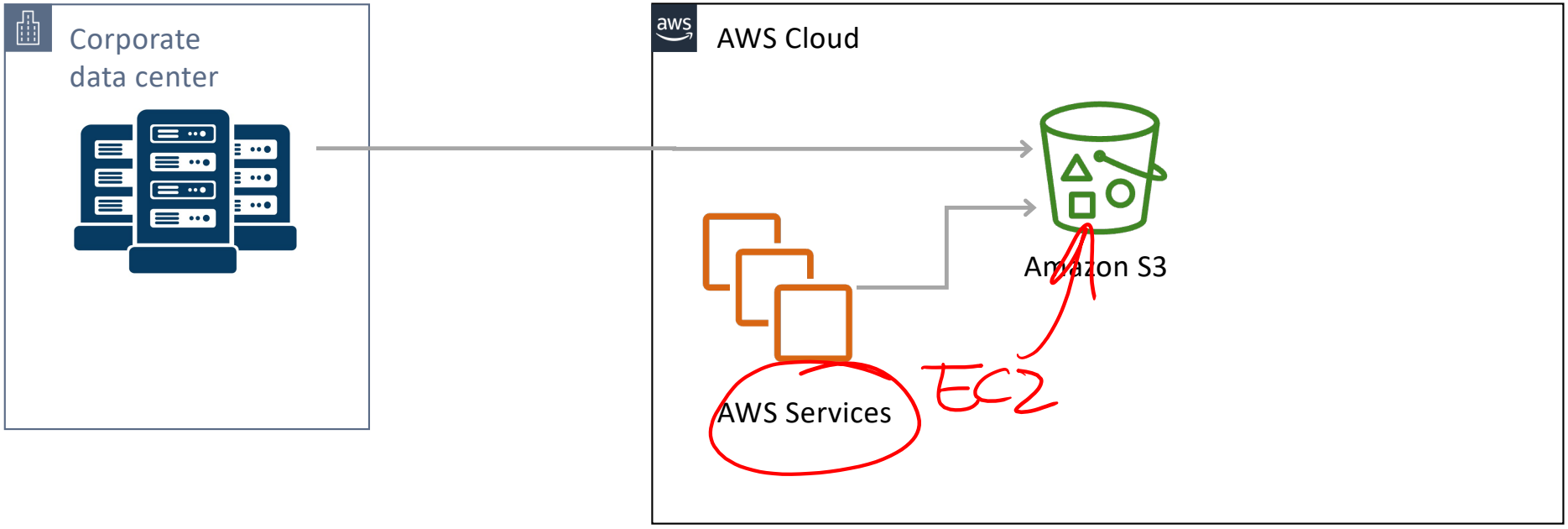
Clickstream analytics

Media transcoding

Amazon S3 use case 4: Back up and archive critical data

aws academy

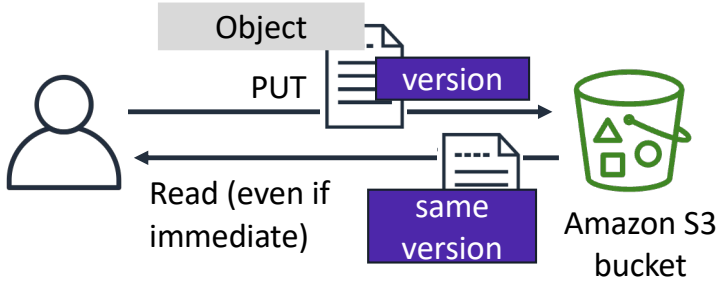
Amazon S3 as a data backup solution



Amazon S3 data strong consistency model

eventually consistent

- Amazon S3 is **strongly consistent** for all new and existing *objects* in all Regions
  - Provides strong read-after-write consistency for all GET, LIST, and PUT operations on objects in S3 buckets
  - The consistency model offers an advantage for big data workloads



Section 2 key takeaways

- Buckets must have a globally unique name and are defined at the Region level
- Buckets are private and protected by default
- Amazon S3 security can be configured with IAM policies, bucket policies, access control lists, S3 access points, and presigned URLs
- Amazon S3 is strongly consistent for all new and existing objects in all Regions
- 5 TB is the maximum size of a single object
- S3 use cases
  - Store and distribute web content and media
  - Host static websites
  - Data store for computation and analytics
  - Back up and archive critical data

Module 3: Adding a Storage Layer

Section 3: Storing data in Amazon S3

aws academy

Amazon S3 classes (1/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

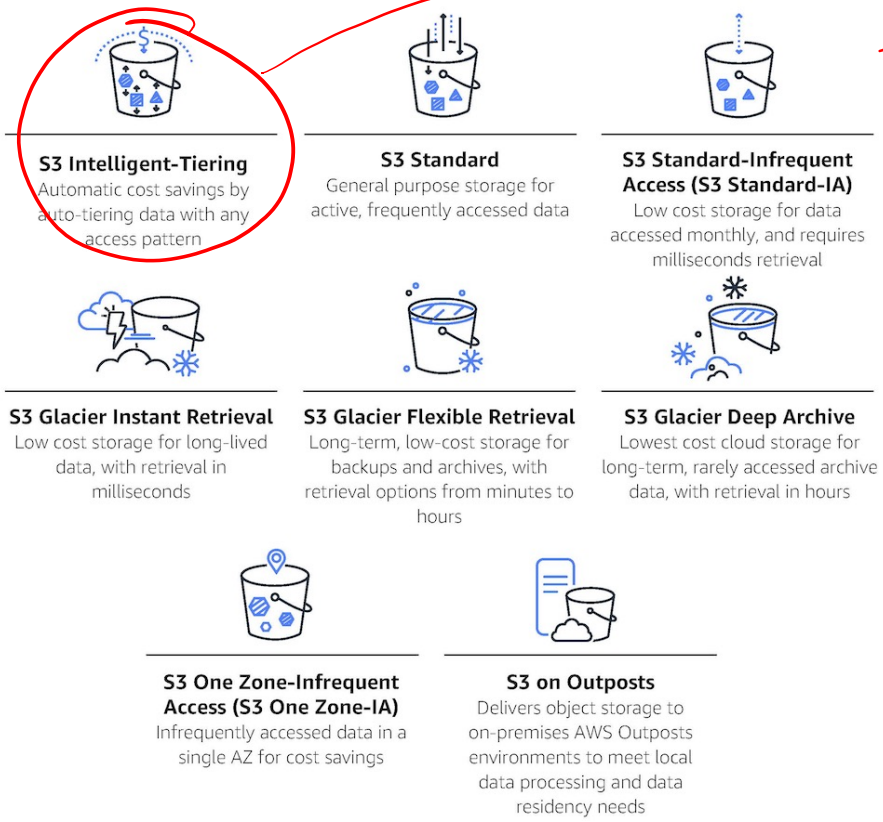
S3 Standard
General purpose storage for active, frequently accessed data with millisecond access

Delivers high durability, availability, and performance

Designed for frequent access, low latency, and high throughput

Data lakes, cloud-native applications, websites, and content distribution

Amazon S3 classes (2/8)



S3 Intelligent-Tiering

Only cloud storage class with automatic cost savings by moving objects between cost-optimized access tiers

Automatically moves objects between the optimal access tiers	Optimizes storage costs based on access patterns	Data lakes and applications with changing or unknown access patterns

For a small monthly object monitoring and automation charge, S3 Intelligent-Tiering monitors access patterns and automatically moves objects that have not been accessed to lower-cost access tiers. S3 Intelligent-Tiering automatically stores objects in three access tiers: one tier that is optimized for frequent access, a 40% lower-cost tier that is optimized for infrequent access, and a 68% lower-cost tier optimized for rarely accessed data. S3 Intelligent-Tiering monitors access patterns and moves objects that have not been accessed for 30 consecutive days to the Infrequent Access tier and after 90 days of no access to the Archive Instant Access tier. For data that does not require immediate retrieval, you can set up S3 Intelligent-Tiering to monitor and automatically move objects that aren't accessed for 180 days or more to the Deep Archive Access tier to realize up to 95% in storage cost savings.

Amazon S3 classes (3/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

S3 Standard-Infrequent Access (S3 Standard-IA)
Lower cost storage for data accessed monthly, with milliseconds retrieval

Infrequently accessed data with rapid retrieval and the durability, availability, and performance of S3 Standard

Low-latency and high throughput of S3 Standard, with a low per GB storage price and per GB retrieval fee

Long-term storage, backups, and disaster recovery

Amazon S3 classes (4/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

S3 Glacier Instant Retrieval

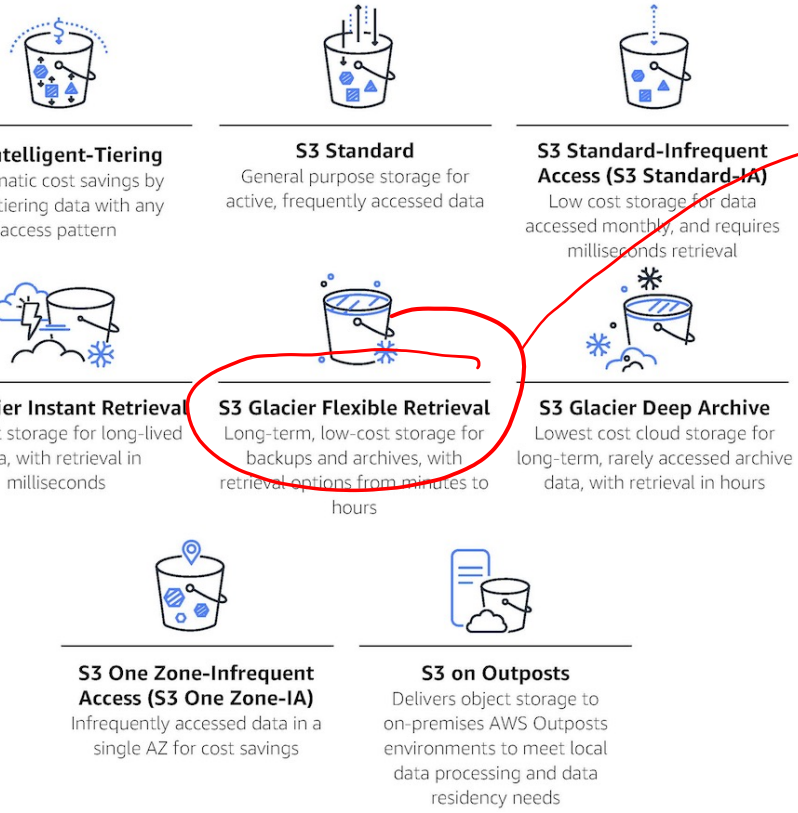
Lowest cost storage for long-term archive data, with milliseconds retrieval

Designed for rarely accessed, long-term data that requires immediate retrieval

Save up to 68% on storage costs compared with using the S3 Standard-Infrequent Access storage class

Long-term digital preservation of data that is accessed once per quarter

Amazon S3 classes (5/8)



S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

S3 Glacier Flexible Retrieval

Archive or backup data with secure, durable, and low-cost storage

Low-cost, durable archive with low retrieval fees

3 retrieval speeds: expedited (1–5 mins), standard (3–5 hours), and bulk (12 hours)

Data archive with query-in-place capabilities to learn from data-at-rest

Amazon S3 classes (6/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

S3 Glacier Deep Archive
Lowest-cost cloud storage for long-term archives at about $1 per TB/month

Eliminate on-premises tape libraries and the need for hardware refresh cycles

2 retrieval speeds: standard (within 12 hours) and bulk (within 48 hours)

Long-term digital preservation for data that is accessed once per year

Amazon S3 classes (7/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Performance of S3 Standard-IA stored in a single AZ at 20% of the cost

Low latency, high throughput storage for infrequently accessed objects that require rapid retrieval

Saves storage costs for data that doesn't require multi-AZ redundancy

Ideal for secondary backups and workloads with easily re-creatable data

Amazon S3 classes (8/8)

S3 Intelligent-Tiering
Automatic cost savings by auto-tiering data with any access pattern

S3 Standard
General purpose storage for active, frequently accessed data

S3 Standard-Infrequent Access (S3 Standard-IA)
Low cost storage for data accessed monthly, and requires milliseconds retrieval

S3 Glacier Instant Retrieval
Low cost storage for long-lived data, with retrieval in milliseconds

S3 Glacier Flexible Retrieval
Long-term, low-cost storage for backups and archives, with retrieval options from minutes to hours

S3 Glacier Deep Archive
Lowest cost cloud storage for long-term, rarely accessed archive data, with retrieval in hours

S3 One Zone-Infrequent Access (S3 One Zone-IA)
Infrequently accessed data in a single AZ for cost savings

S3 on Outposts
Delivers object storage to on-premises AWS Outposts environments to meet local data processing and data residency needs

Amazon S3 on Outposts

Delivers object storage to your on-premises AWS Outposts environment to meet local data processing and data residency needs

Makes it easy to store, secure, tag, retrieve, report on, and control access to the data on your Outpost

Satisfy demanding performance needs by keeping data close to on-premises applications

Provides on-premises object storage to minimize data transfers and buffer from network variations

Performance across the S3 storage classes

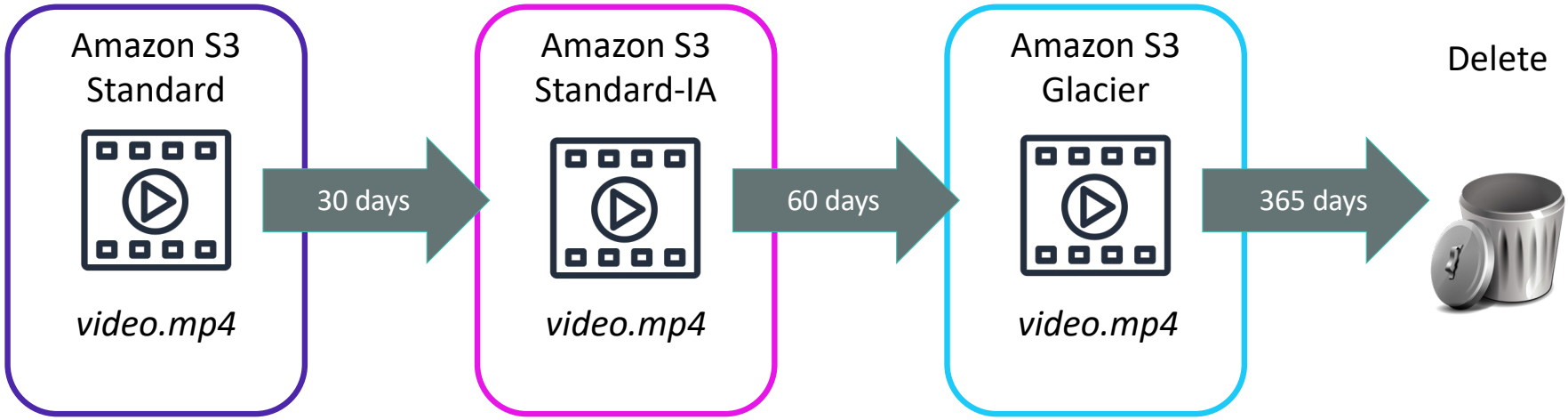
	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (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
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

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

Amazon S3 lifecycle policies

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

Configure an **Amazon S3 Lifecycle Policy** to **automate** deleting or moving objects based on age.



Amazon S3 costs

Pay only for use, including:
GBs of objects stored (per month). Different pricing per *Region* and per *storage class*.
Transfer OUT to other Regions or the internet.
PUT, COPY, POST, LIST, GET, SELECT, lifecycle transition, data retrieval requests.

No charge for:
Data transfers IN from the internet to Amazon S3.
Transfers between S3 buckets or from Amazon S3 to any services in the same AWS Region.
Transfer OUT to Amazon CloudFront.
DELETE and CANCEL requests.

Section 3 key takeaways

- Amazon S3 storage classes include –
  - S3 Standard
  - S3 Standard-IA
  - S3 One Zone-IA
  - S3 Intelligent-Tiering
  - S3 Glacier
  - S3 Glacier Deep Archive
- An Amazon S3 lifecycle policy can delete or move objects to less expensive storage classes based on age
- Transferring data in from the internet to Amazon S3 is free, but transferring out to other Regions or to the internet incurs a fee

Thank you and Kahoot!

aws academy