Amazon s3

Tamal Dey,
Dept. of CA
PESU

What's in it for you?

- What is Cloud storage?
- Types of storage
- Before Amazon S3
- What is S3?
- Benefits of S3
- Objects and Buckets
- How does Amazon S3 work
- Features of S3

What is cloud storage?

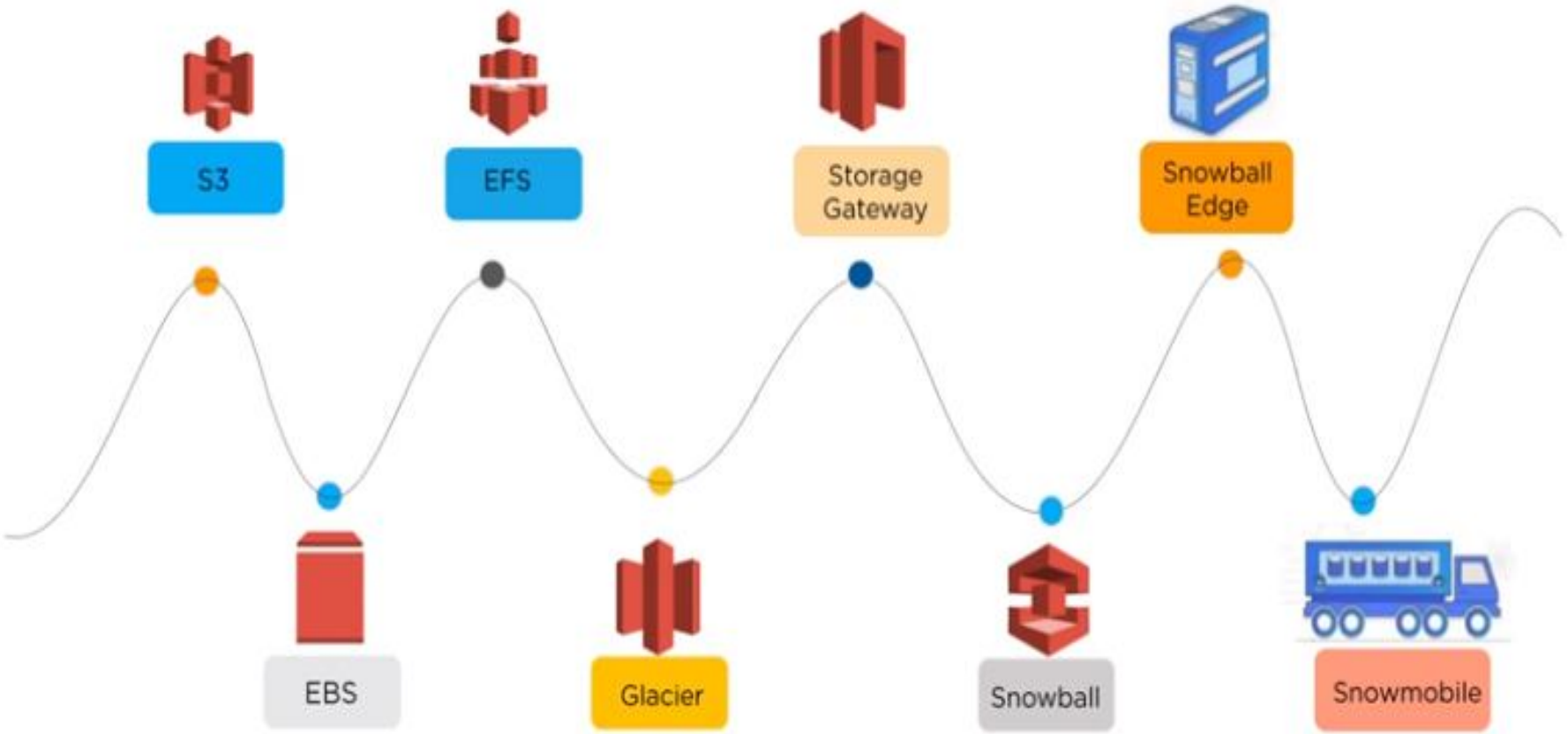
Cloud storage provides a web service where your data can be stored, accessed and easily backed up by users over the internet

Benefits

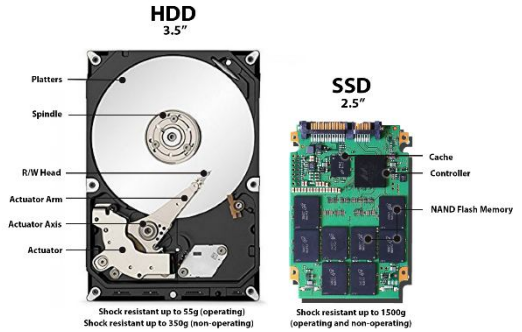
Cloud storage is

- Reliable
- Scalable
- Secure

Types of Storage in AWS



Types of Storage in AWS



- **S3** (Simple Storage Service) – cloud storage
- **EBS** (Elastic Block Store)– similar C drive or E drive (SSD drives attached to instances)
- **EFS** (Elastic File System) - shared file systems (multiple systems)
- **Glacier** – archiving solution (low cost back up)[store infrequently used data, or "cold data."]
- **Storage gateway** – safely moving data from on-premises to cloud, hybrid cloud storage
- **Snowball** – data import and export system (h/w given to premise for data storage securely)
- **Snowmobile** – massive data centre on mobile. You can transfer up to 100PB per Snowmobile

Before Amazon S3

MAINTAINING YOUR OWN REPOSITORY IS EXPENSIVE AND TIME CONSUMING

Factors that make a repository expensive and time consuming are:

- To purchase hardware and software components
- Hiring a team of experts for maintenance
- Lack of scalability based on your requirements
- Data security requirements

What is S3?

Amazon S3 (Simple Storage Service) provides object storage which is built for storing and recovering any amount of information or data from anywhere over the internet

Know about S3

- S3- simple storage service - It provides **object storage** service
- Amazon S3 provides storage through web service interface
- It is designed for developers where web-scale computing can be easier for them
- You cannot install anything on S3
- It can store files upto 5 TB in size
- durability(99.999999999%)
- 99.99% availability, expected loss of 0.000000001% of objects
- S3 is cheap
- S3 is a regional service (any region you can opt)
- Lot of security provision

	Approx. Bytes	Actual Bytes	Approx. Bits	Typical file/media
1B	1	1	8	Text email, SMS
1KB	1000B = 10^3	1024B = 2^{10}	8 x 10^3	Word document
1MB	1000KB = 10^6	1024KB = 2^{20}	8 x 10^6	Digital photo
1GB	1000MB = 10^9	1024MB = 2^{30}	8 x 10^9	DVD
1TB	1000GB = 10^{12}	1024GB = 2^{40}	8 x 10^{12}	Hard disk
1PB	1000TB = 10^{15}	1024TB = 2^{50}	8 x 10^{15}	Cloud?

Highlights of S3

What is object and bucket?

An object consists of data, key(assigned name) and metadata

A bucket stores objects

When data is added to the bucket, Amazon S3 creates a unique version ID and allocates it to the object

For Example:

Object: folder/Penguins.jpg → Key(name)
Bucket: simplilearn → Version ID
Link Address: https://s3.amazonaws.com/simplilearn/folder/Penguins.jpg

How does it work?

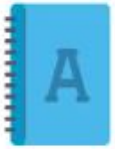
- ✓ When files are uploaded to the bucket, the user will specify the type of S3 storage class to be used for those specific objects
- ✓ Later, users can define features to the bucket like bucket policy, lifecycle policies, versioning control etc.

Storage Classes

- **Standard**
  - For frequently accessed data. Stores object data redundantly across multiple geographically separated Availability Zones
- **Standard-IA**
  - For infrequently accessed data. Stores object data redundantly across **multiple** geographically separated Availability Zones. Minimum **30-day** retention period and minimum **128 KB** object size.
- **One Zone-IA**
  - For infrequently accessed data. Stores object data in **only one** Availability Zone at a lower price than Standard-IA. Minimum **30-day** retention period and minimum **128 KB** object size
- **Glacier**
  - low-cost cloud storage service to move infrequently accessed data
- **Reduced redundancy**
  - For **frequently** accessed data. Stores noncritical, reproducible data at lower levels of redundancy than Standard.
- To define the storage classes – Go to objects – click properties

Storage class in Amazon S3 with a "School" use case

Amazon S3 Standard for frequent data access	Suitable for a use case where the latency should be low Example: Frequently accessed data will be the data of students' attendance, which should be retrieved quickly
Amazon S3 Standard for infrequent data access	Can be used where the data is long lived and less frequently accessed Example: Students' academic record will not be needed on a daily basis, but if they have any requirement, their details should be retrieved quickly
Amazon Glacier	Can be used where the data has to be archived and high performance is not required Example: Ex-student's old record (like admission fee), will not be required on a daily basis and even if it is necessary, low latency is not needed



Students

Student's old record

Storage class in Amazon S3 with a “School” use case

One Zone-IA Storage Class	Can be used where the data is infrequently accessed and stored in a single region Example: Student’s report card is not used on a daily basis and stored in a single availability region (i.e., school)
Amazon S3 Standard Reduced Redundancy storage	Suitable for a use case where the data is non critical and reproduced quickly Example: Books in the library are non critical data and can be replaced if lost

Student’s report card

Library Books

Storage Class Summary

Amazon S3 Standard for frequent data access	Amazon S3 Standard for infrequent data access	Amazon Glacier
• For frequently accessed data • It is a default storage class • Can be used for cloud applications, dynamic websites, content distribution, gaming applications, and Big data analytics	• For infrequently accessed data • Demands rapid access • Suitable for backups, disaster recovery and lifelong storage of data	• Suitable for archiving data where data access is infrequent • Vault-lock feature provides a long term data storage • Provides the lowest cost availability

Storage Classes Comparison in Amazon S3

Storage Class	Durability	Availability	SSL support	First byte latency	Lifecycle Management Policies
STANDARD	99.999999999%	99.99%	Yes	Milliseconds	Yes
STANDARD_IA	99.999999999%	99.99%	Yes	Milliseconds	Yes
ONEZONE_IA	99.999999999%	99.5%	Yes	Milliseconds	Yes
GLACIER	99.999999999%	99.99%	Yes	Minutes or Hours	Yes
RRS	99.99%	99.99%	Yes	Milliseconds	Yes

Life Cycle Management

... and after 60 days, it is moved to Glacier

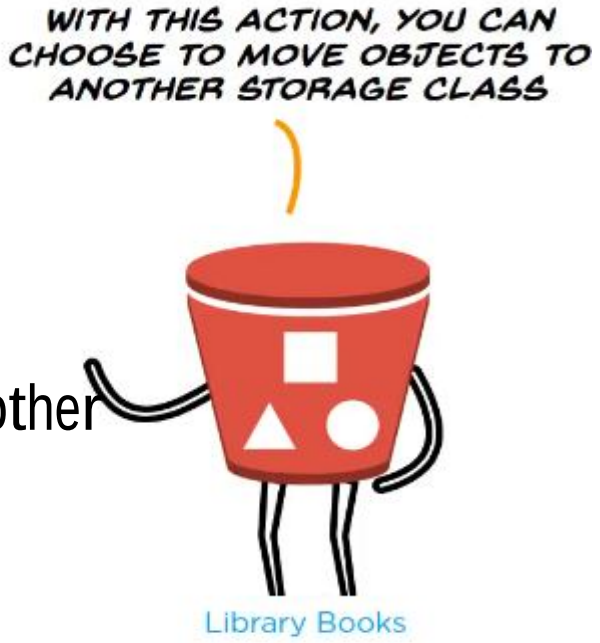
Transition actions

THIS LIFECYCLE MANAGEMENT HELPS YOU TO AUTOMATICALLY MIGRATE YOUR DATA TO LOWER COST STORAGE AS YOUR DATA AGES

You can configure S3 to move your data between various storage classes on a defined schedule

Life Cycle Management

In lifecycle management, Amazon S3 applies a set of rules that define actions to a group of objects



- **Transition actions**
  - Move data from one storage class and another storage class
- **Expiration actions**
  - Expiration date can be fixed for any object
- To do Life cycle management go to bucket and click to Management tab and add the transition action and expiration action.

Bucket Policy

- ✓ Bucket policy is an IAM policy where you can allow and deny permission to your Amazon S3 resources
- ✓ With bucket policy, you also define security rules that apply to more than one file within a bucket
- ✓ For example: If you do not want a user to access the "Simplilearn" bucket, then with the help of JSON script you can set permissions

- Create a bucket policy
  - Suppose you want to deny particular user
  - First generate the policy as a JSON file
  - Goto https://awspolicygen.s3.amazonaws.com/policygen.html
  - Or type "aws policy generator" in Google.
    - Effect – Deny
    - Principal – action (give * for all)
    - Service – amazon s3
    - Actions – click all
    - ARN – copy from aws s3 bucket properties
    - Add condition – specify the user = or <>
    - Add statement
    - Generate policy -> A JSON script will be generated.
    - Copy the JSON file and paste in the bucket policy.

Data Protection

- ✓ Amazon S3 provides IT teams a highly durable, protected and scalable infrastructure designed for object storage

S3 Data Protection Techniques

- ✓ Amazon S3 provides IT teams a highly durable, protected and scalable infrastructure designed for object storage
- ✓ Amazon S3 protects your data using 2 methods:
  - ❑ Data Encryption and
  - ❑ Versioning

Data encryption

Versioning

Data Encryption

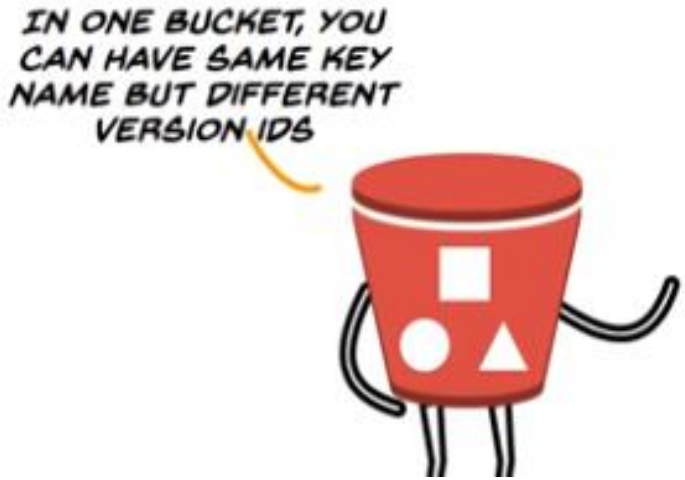
- ✓ It refers to protection of data while it's being transmitted and at rest
- ✓ Data Encryption can happen in two ways:

Client-Side Encryption - Data encryption at rest

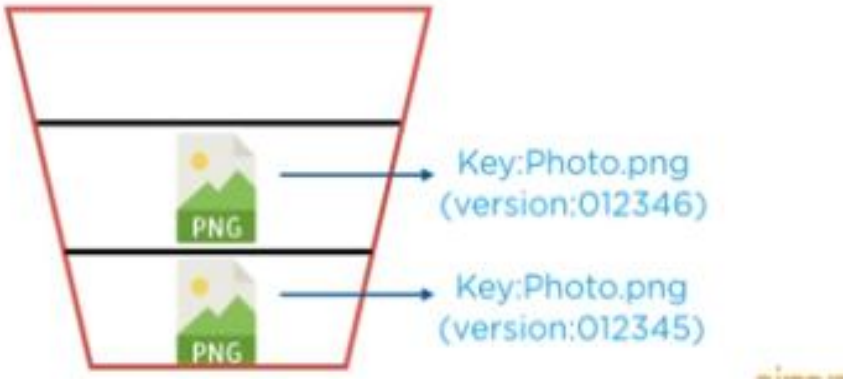
Server-Side Encryption - Data encryption in motion

Versioning

- ✓ It can be utilized to preserve, recover and restore early versions of every object you store in your Amazon S3 bucket
- ✓ Unintentional erase or overwriting of objects can be easily regained with versioning

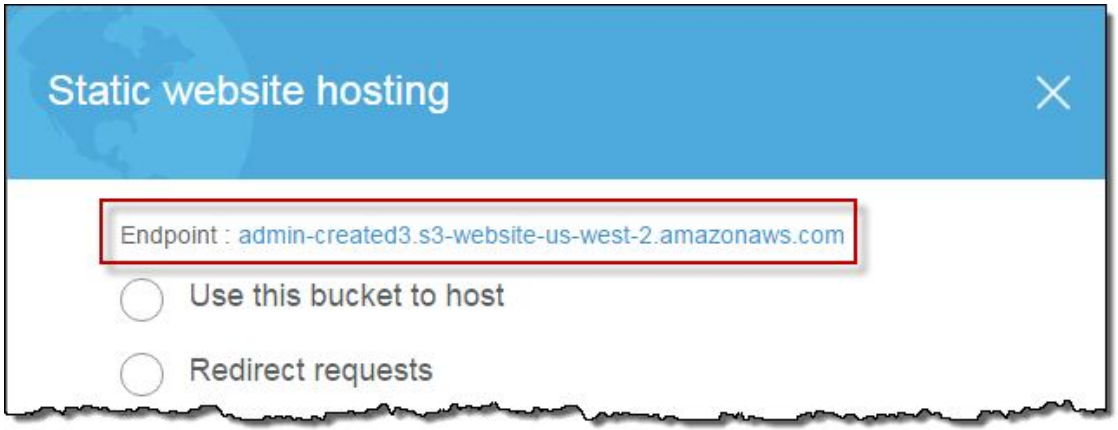


For Example:



Static Webpage Hosting

- Click on Bucket Properties and Enable Static Webpage Hosting
- Click on the Static Webpage Hosting menu and write two file names (index.html and error.html) and save
- Create and Edit index.html and error.html in your local machine and upload the files on the bucket with public access permission
- Click on the endpoint link on Static Webpage Hosting menu to get the page view



Cross-Region Replication

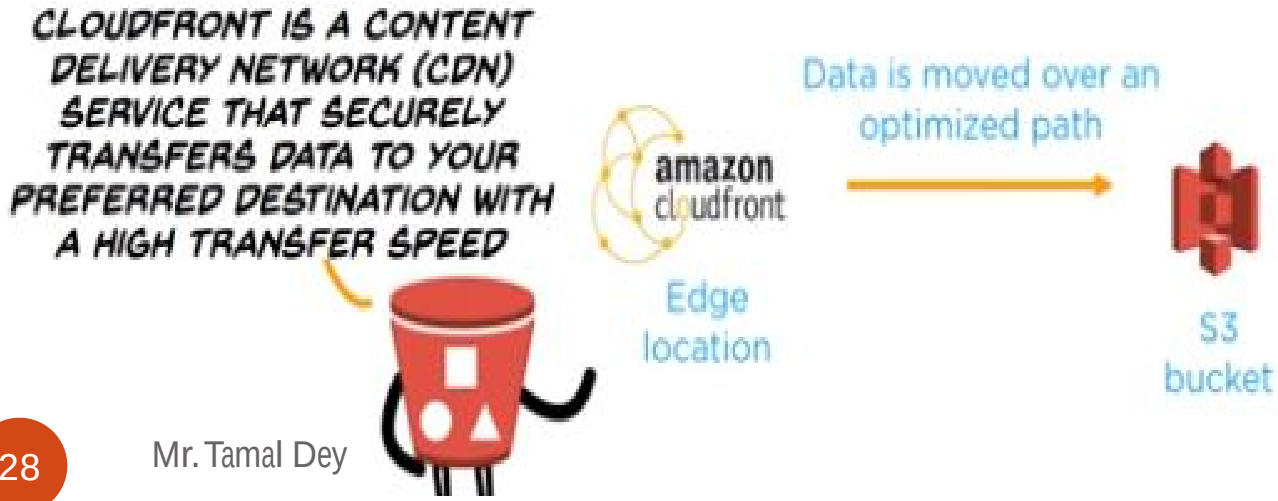
Cross-Region Replication provides automatic copying of every object uploaded to your buckets (source bucket and destination bucket) in different AWS regions

Note: Versioning must be turned on to enable CRR

- Before doing CRR both buckets should have versioning enabled.
- Create a destination bucket in a different region
- Under Bucket -> properties -> Replicate

Accelerated Transfer

- ✓ It enables fast, easy and secure transfers of files over long distances between your client and S3 bucket
- ✓ The edge locations around the world provided by Amazon CloudFront are taken advantage by transfer acceleration
- ✓ It works via carrying data over an optimized network bridge that keeps running between the AWS Edge Location (closest region to your clients) and your Amazon S3 bucket



- Cloud front helps to copy the bucket from one region to another region by serveral intermediate copies to the nearest region and not directly to the destination region.
- This is useful, when the source and destination regions are farther.
- Under bucket properties ->Enable accelerated transfer

Reading resources

- https://www.youtube.com/watch?v=XGcoeEyt2UM