Amazon S3 :-

- It is advertised as "infinitely scaling" storage

Amazon S3 Use Cases :-

Backup & storage
Disaster recovery
Archive
Hybrid Cloud storage
Application hosting
Media hosting
Data lakes & big data analytics
Software delivery updates
Static website

* Buckets :-

① Amazon S3 allows people to store objects (files) in "buckets" (directories)
② Buckets must have a globally unique name (across all regions all accounts)
③ Buckets are defined at region level.
④ S3 looks like a global service but buckets are created in region.

• Naming Convention for S3 bucket :→

→ No uppercase, No underscore
→ 3-63 characters long
→ Not an IP
→ Must start with lowercase letter or number
→ must Not start with the prefix xn---
→ Must not end with the suffix -s3alias

* Amazon S3 - Objects :-

- files stored in s3 buckets are called as objects.
- Objects (files) have a key
- The **key** is the **full Path**
  - → s3://my-bucket/my_file.txt
  - → s3://my-bucket/my-folder/my_file.txt
- This key is composed of prefix + obj name

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

prefix obj. name

- There is no concept of "directories" within buckets (although the UI will trick you to think otherwise)

- Just keys with very long names that contain slashes("/").

- Object values are the content of the body:
  - → max. object size is 5TB (5000 GB)
  - → If uploading more than 5GB, must use "multi-part upload".

- Metadata (list of text key/value pairs - system or user metadata)
- Tags (Unicode key/value pair - upto 10)
  - useful for security/lifecycle
- Version ID (if versioning is enabled)

Amazon S3 - Security :-

① User - Based :-
- IAM Policies - which API calls should be allowed for a specific user from IAM

② Resource - Based :-
- Bucket Policies - bucket wide rules from the S3 console - allows cross account
- Object Access Control List (ACL) :- finer grain (can be disabled)
- Bucket Access Control (ACL) :- less common (can be disabled)

⊙ Note : an IAM principal can access an S3 object
- The user IAM permissions ALLOW it OR the resource policy ALLOWS it.
- AND there's no explicit DENY

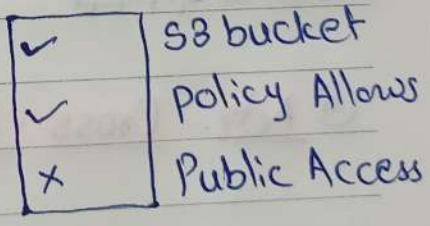
⊙ Encryption :- encrypt objects in Amazon S3 using encryption keys

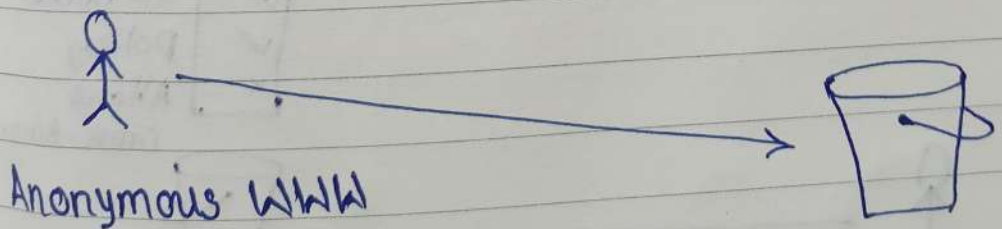
S3 Bucket Policies :-

- JSON based Policies
  - Resources: buckets & Objects
  - Effect : Allow / Deny
  - Actions : Set of API to Allow or Deny
  - Principal : The account or user to apply the policy to

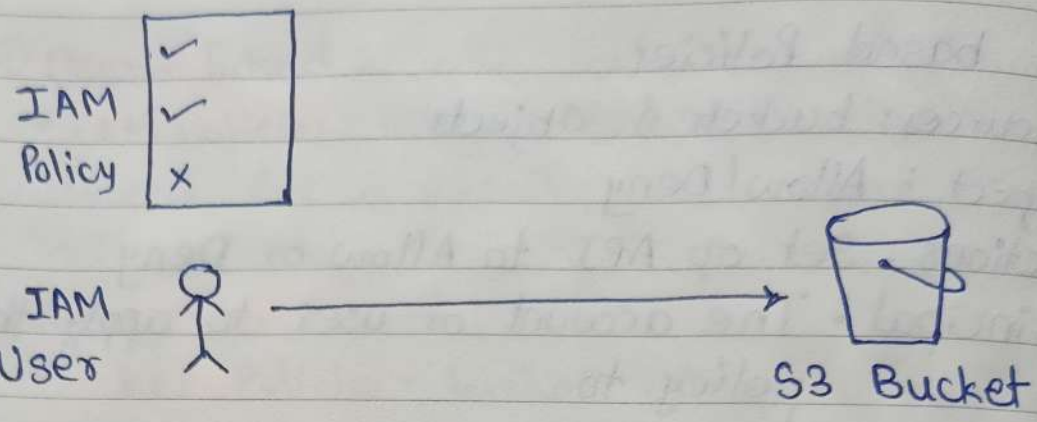
- Use S3 bucket for Policy to
  - Grant public access to the bucket
  - Force objects to be encrypted at upload
  - Grant access to another account (cross Account)

① ex: Public Access :- Use Bucket Policy

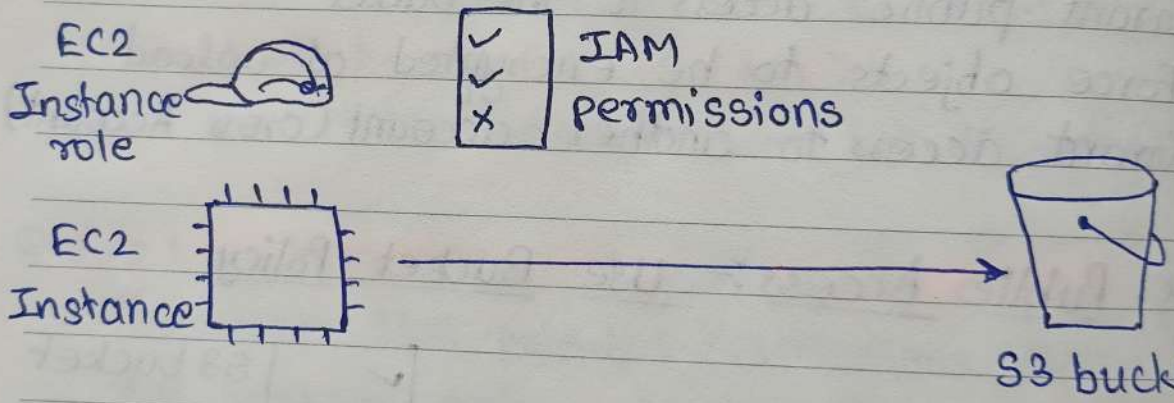




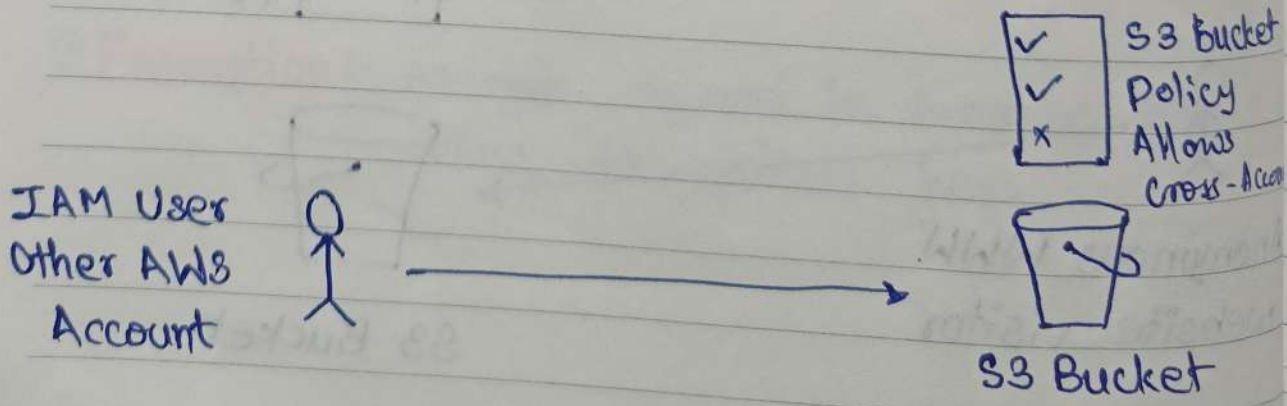
② Ex: User Access to S3 - IAM Permissions



③ EC2 instance Access - Use IAM Roles



④ Adv. Cross-Account Access - Use Bucket Policy



* Bucket Settings for Block Public Access

+ These settings were created to prevent company data leaks.
+ If you know your bucket should never be public, leave these on
+ Can be set at the account level.

* Amazon S3 - Static Website Hosting

- S3 can host static websites and have them accessible on the internet
- The website URL will be (depending on the region)
-
- If you get a 403 forbidden error; make sure the bucket policy allows public reads

Amazon S3 - Versioning

→ You can version your files in Amazon S3.
→ It is enabled at the bucket level
→ Same key overwrite will change the "version":
→ It is best practice to version your buckets
  • Protect against unintended deletes
    (ability to restore a version)
  • Easy rollback to previous version.
→ Notes :-
  • Any file that is not versioned prior to enabling versioning will have version "null"
  • Suspending versioning does not delete the previous versions.

* Amazon S3 - Replication (CRR & SRR) :→

→ Must enable versioning in source & destination buckets

① → Cross - Region Replication (CRR)

② → Same - Region Replication (SRR)

→ Buckets can be in different AWS Accounts

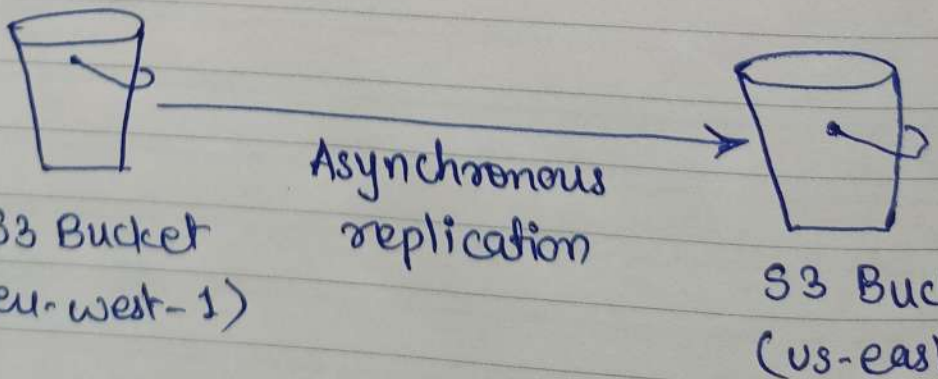
→ Copying is asynchronous

→ Must give proper IAM Permission to S3

• Use Cases :

CRR - Compliance, lower latency access, replication across accounts

SRR - log aggregation, live replication between production & test accounts



S3 Durability & Availability

* **Durability** :-

High durability (99.99999999, 11 9's) of objects across multiple AZ.

- If you store 10,000,000 objects with S3, you can on avg expect to incur a loss of one object once ~~year~~ every 10,000 year.
- Same for all storage classes.

* **Availability** :-

Measures how readily available service is.

→ Varies depending on storage class

→ ex: S3 std. has 99.99% availability
= not available 53 min. a year.

S3 Storage Classes

① Standard
② Intelligent -Tiering
③ Std.-IA
④ One Zone IA
⑤ Glacier - Instant Retrieval
⑥ Glacier - Ins flexible Retrival
⑦ Glacier - Deep Archive

- Can move between classes manually or using S3 Lifecycle configurations.

① S3 std. - General Purpose

① 99.99 % availability
② Used for frequently accessed data
③ Low latency & high throughput
④ Sustain 2 Concurrent facility failures

- Use Cases: Big data analytics, mobile & gaming applications, Content distribution

② S3 Storage Classes - Infrequent Access

- for ddta that is less frequently accessed, but requires rapid access when needed
- Lower cost than S3 bucket

ⓐ Amazon S3 std. Infrequent Access ⇒
- 99.9 % availability
- Use case: disaster recovery, backups

ⓑ Amazon S3 One-Zone-Infrequent Access ⇒
- High durability in single AZ; data lost when AZ is destroyed.
- 99.5 % availability
- Use case:- Secondary backup copies of on-premise data

③ Amazon S3 Glacier Storage classes

① Low-cost storage meant for archiving / backup
② Pricing: price for storage + object retrieval cost

- **Amazon S3 Glacier Instant Retrieval**
  - Millisecond retrieval, great for data accessed once a quarter
  - Minimum storage duration of 90 days

- **Amazon S3 Glacier Flexible Retrieval** (formerly Amazon S3 Glacier) :-
  - Expedited (1 to 5 minutes), Std. (3 to 5 hrs) bulk (5 to 12 hrs) - free
  - Minimum storage duration 90 days.

- **Amazon S3 Glacier Deep Archive** :-
  -
  - for long term storage
  - Standard (12 hours), bulk (48 hours)
  - minimum storage duration of 180 days.

S3 Intelligent - Tiering :-

+ Small monthly monitoring & Auto-tiering fee
+ Moves Objects automatically between Access tiers based on usage
+ There are no retrieval charges in S3 intelligent - tiering.

@ Frequent Access Tier (Automatic) - default tier

ⓑ Infrequent Access Tier (Automatic) - Objects not accessed for 30 days.

ⓒ Archive Instant Access Tier (automatic) - objects not accessed for 90 days

ⓓ Archive Access Tier (Optional) - Configurable from 90 days to 700+ days

ⓔ Deep Archive Access Tier (Optional) - Config. from 180 days to 700+ days

S3 Encryption :→

① Server - Side Encryption (Default) :-

Once file/object is uploaded to S3 bucket then it encryption will be performed on server side..

② Client - Side Encryption :-

User performs encryption of file & then uploads it to the S3 bucket

IAM Access Analyzer for S3 :-

+ Ensures that only intended ppl have access to your S3 buckets
+ ex: publicly accessible bucket, bucket shared with other AWS account...
+ evaluates S3 bucket Policies, S3 ACL, S3 Access Point policies
+ Powered by IAM Access Analyser

Shared Responsibility Model for S3

⇒ AWS :→

① Infrastructure (global security, durability, availability, sustain concurrent loss of data in two facilities.)
② Configuration & Vulnerability analysis
③ Compliance validation.

⇒ User :→

① S3 Versioning
② S3 bucket Policies
③ S3 Replication setup
④ Logging & Monitoring
⑤ S3 Storage Classes
⑥ Data encryption at rest & in transit

AWS Snow family :→

* Highly-secure, portable device to collect & process data at the edge, & migrate data into & out of AWS.

+ Data Migration : ① Snowcone
② Snowball edge
③ Snow mobile

+ Edge Migration : ① Snowcone
② Snowball edge
③ Snow mobile

AWS Snow Family : Offline devices to perform data migrations.

If it takes more than a week to transfer data over the n/w, use snowball devices.

* Data Migration with AWS Snow family

Challenges Current :-

1. Limited Connectivity
2. Limited bandwidth
3. High n/w cost
4. Shared bandwidth (can't max the line)
5. Connection stability

	Time to Transfer		
	100 Mbps	1 Gbps	10 Gbps
10 TB	12 days	30 hrs	3 hrs
100 TB	124 days	12 days	30 hrs
1 PB	3 yrs	124 days	12 days

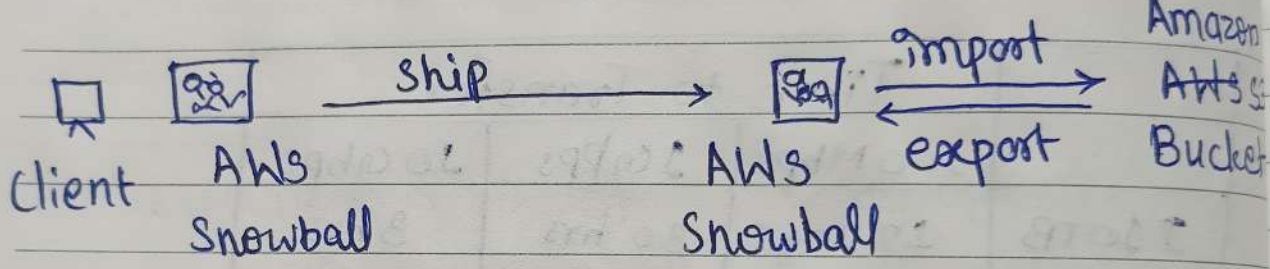
→ If it takes more than a week to transfer over the network, use Snowball devices!

Diagrams :-

① Direct upload to S3 :-

Client ——10Gbit/s——→ Amazon S3 bucket

② With Snow family :-



① Client, place order for AWS Snowball device
② Once its received data will be stored loaded snowball and ship to the Amazon facility
③ Then ~~import~~ they will plug device to the facility where data needs to be imported or exported in Amazon S3 bucket

* Snowball Edge (for data transfers) :-

① Physical data transport solution: move TBs or PBs of data in or out of AWS.

② Alternative to moving data over the n/w (and paying n/w fees)

③ Pay per ~~data~~ transfer job

④ Provide block storage & S3-compatible object storage.

⑤ Snowball Edge Storage Optimized

- 80 TB of HDD ~~capacity or 28 TB NVMe~~ capacity for block vol & S3 compatible object storage.

⑥ Snowball Edge Compute Optimized

- 42 TB of HDD or 28 TB NVMe capacity for block vol. & S3 compatible obj. storage

⑦ Use cases large data cloud migrations, DC decommission, disaster recovery.

* AWS Snowcone & Snowcone SSD :→

① Small, portable computing anywhere, rugged & secure, withstands harsh env.

② Light (4.5 pounds, 2.1 kg)

③ Device used for edge computing, storage & data transfer

④ Snowcone - 8 TB of HDD storage

⑤ Snowcone SSD - 14 TB of SSD storage

⑥ Use snowcone where snowball does not fit (space - constrained env)

⑦ Must provide your own battery / cables

⑧ Can be sent back to AWS offline, or connect it to internet & use AWS datasync to send data.

⑧ Upto 24 TB. Online and Offline

* AWS Snowmobile :- (remember Truck)

① Transfer exabytes of data (1EB = 1000TB = 1,000,000TBs)

② Each ~~snowball~~ snowmobile has 100PB of capacity (use multiple in parallel

③ High Security: Temp controlled, GPS, 24/7 video surveillance

④ Better than snowball if you transfer more than 10 PB.

① Snowcone & Snowcone SSD :- 8TB HDD
14TB HDD

② Snowball Edge Storage Optimized:-
upto Petabytes.

③ Snowmobile :- upto exabytes

* Snow Family - Usage Process

① Request snowball devices from the AWS console for delivery

② Install the snowball client / AWS OpsHub on your servers

③ Connect the snowball to your servers & copy files using the client

④ Ship back the device when you're done (goes to the right AWS facility)

⑤ Data will be loaded into S3 bucket

⑥ Snowball is completely wiped

* What is edge Computing ?

- Process data while its being created on a edge location
  - A truck on a road, a ship on the sea, a mining station underground.

- These locations may have
  - Limited / no internet access
  - Limited / no easy access to computing power

- We setup a snowball edge / snowcone device to do edge computing

- Use cases of edge computing :-
  - Process data
  - Machine learning at the edge
  - Transcoding media streams

- Eventually (if we need be) we can ship back the device to AWS (for transferring data for example)

Snow family - Edge Computing :-

① Snowcone & Snowcone SSD (smaller)
- 2 CPU, 4 GB of memory, wired or wireless
- USB-C Power using a cord or the optional battery

② Snowball Edge - Compute Optimized
- 104 vCPUs, 416 GiB of RAM
- Optional GPU (Useful for video processing or machine learning)
- 24 TB NVMe or 42 TB HDD usable storage
- Storage clustering available (up to 16 nodes

③ Snowball edge - Storage Optimized
- up to 40 vCPUs, 80 GiB of RAM, 80 TB stor
-

- All: can run EC2 instances & AWS Lamba funct (using AWS IoT Greengrass)

- Long-term deployment options: 1 & 3 years discounted pricing

* AWS OpsHub :→

- Historically, to use snow family devices, you needed a CLI (

- Today, you can use AWS OpsHub (a software you install on your computer / laptop) to manage your snow family device.

- Unlocking & configuring single or clustered devices
- Transferring files
- Launching & managing instances running on snow family devices
- Monitor device metrics (storage capacity, active instances on your device)
- Launch compatible AWS services on your devices.
  (ex. Amazon EC2 instances, ~~on your devices~~ AWS DataSync, N/w file system (NFS)

* Snowball Edge Pricing

→ You pay for device usage & data transfer out of AWS.
→ Data transfer IN to amazon S3 is 0.00 per

- On-Demand :-
  + Includes a one-time service fee per job which includes:
    - 10 days of usage for snowball Edge storage Optimized 80TB
    - 15 days of usage for snowball edge storage optimized 210 TB
  + Shiping days are not counted towards the included 10 or 15 days
  + Pay per day for any additional days

- Committed Upfront :-
  + Pay in advance for monthly, 1-Yr & 3 years of usage (Edge computing)
  + Up to 62% discounted pricing

Storage Gateway :→

Hybrid Cloud for Storage

- AWS is pushing for "hybrid cloud"
  - Part of your infrastructure is on-premises
  - Part of your infrastructure is on the cloud

- This can be due to -
  - long cloud migration
  - Security requirements
  - complaince req.
  - IT strategy

- S3 is a proprietary storage technology (unlike EFS / NFS), so how do you exppse the S3 data on-premise?

- AWS Storage Gateway!

* AWS Storage Cloud Native Options

Block :- Amazon EBS
EC2 Instance store

File :- Amazon EFS

Object :- Amazon S3
Glacier

Amazon Storage Gateway :→

- Bridge between on-premise data & cloud data in S3.

- Hybrid storage service to allow on-premise to seamlessly use the AWS Cloud

- Use Cases : distraster recovery, backup & restore , tiered storage

- Types of Storage Gateway :- ① File Gateway
② Volume Gateway
③ Tape Gateway

- No need to know the types at the exam

Amazon S3 - Summary

① Buckets vs Objects : global unique name, tied to a region

② S3 Security : IAM policy, S3 Bucket Policy (public access), S3 Encryption

③ S3 Websites : host a static website on Amazon S3

④ S3 Versioning : multiple versions for files, prevents accidental deletes.

⑤ S3 Replication : same-region or cross-region, must enable versioning

⑥ S3 Storage Classes : Standard, IA, IZ-IA, Intelligent, Glacier (Instant, Flexible, Deep)

⑦ Snow family :- import data onto S3 through a physical device, edge computing

⑧ OpsHub :- desktop application to manage Snow family devices.

⑨ Storage Gateway: hybrid solution to extend on-premises storage to S3.