AWS S3
Cheat Sheet - Sem 1 2020

Amazon S3 Basics:

- It allows people to store objects/files in buckets (directories).
- They have globally unique names (must be unique across all regions all accounts).
- Buckets are defined at region level.

S3 looks like a global service but S3 is region specific.

Naming Convention:

- No uppercase, No underscore.
- 3-63 characters long.
- Not an IP.
- Must start with lowercase letter or number.
- Must not start with prefix `xmx--`.
- Must not end with suffix `--s3alias`.

Amazon S3 Objects (files):

- Have object key.
- The key is the full path.
  - s3://my-bucket/my/key/file.txt
    - bucket name.
    - object name.
  - s3://my-bucket/my/folder/file.txt
    - bucket name.
    - prefix.
    - object name.
- The key is composed of prefix + object name.
- Web: Path to an image of documents, videos, etc.
- Object Values = contents of the body.
  - Max size is 5TB/5000GB.
  - Upload using more than 5GB requires multipart upload.
- Tags = (Unicode key/value pair)
  - Tags are useful for security/IAM.
- Version ID.

Amazon S3 - Security:

- User Based:
  - IAM policies - which API calls should be allowed to a specific user.
- 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 List (BACL)
    - less common (can be disabled).

NOTE: an IAM principal can access S3 if:
- The IAM user permission allow it.
- The resource policy allows it.
- AND there is no explicit deny.

Encryption:

Encrypt the objects in Amazon S3.

S3 Bucket Policies:

- JSON Bucket Policies:
  - Resources - Bucket + Objects.
  - Effect - Allow/Deny.
  - Actions - list of API actions to allow or disallow.
  - Principal - The account/user/role to apply the policy to.

Amazon S3 Static Web Hosting:

- The website URL is based on the region.
- If you get 403 Forbidden error, make sure the bucket policy allows public access.

Amazon S3 - Versioning:

- Versioning you give is done at bucket level.
- Any file that is not versioned prior to enabling versioning will have `null` version.
- Suspending versioning does not delete the previous versions.

Amazon S3 - Replication:

- Must enable versioning in source & destination buckets.
- CRR - Cross Region Replication.
- SRR - Same Region Replication.
- Buckets can be in different accounts.
- Copying is asynchronous.

S3 Storage Classes:

- Amazon S3 Standard
- Amazon S3 Standard-Infrequent Access (IA)
- Amazon S3 One Zone - IA
- Amazon S3 Glacier Instant Retrieval
- Amazon S3 Glacier Flexible Retrieval
- Amazon S3 Glacier Deep Archive
- Amazon S3 Intelligent-Tiering

Can move between classes manually or using Amazon S3 lifecycle configurations.

- These classes are highly durable.
- Availability varies based on classes.

S3 Standard - General Purpose:

- 99.99% availability.
- Used for frequently accessed data.
- Low latency & high throughput.
- Sustain 2 concurrent facility failures.

Use cases: Big Data Analytics, Gaming etc.

S3 - Infrequent Access (IA):

- For data that is accessed less frequently but has rapid access when needed compared to Standard.

S3 - Glacier Storage:

- Low cost object storage for backup.
- Tiering Strategy = Object Retrieval cost.

Glacier Instant Retrieval:

- Millisecond Retrieval.
- Minimum storage duration - 90 days.

Glacier Flexible Retrieval:

- Retrieval time:
  - 1-5 mins - Expedited.
  - 3-5 hours - Standard.
  - 5-12 hours - Bulk (free).
- Minimum storage duration - 90 days.

Glacier Deep Archive:

- For long term storage - minimum of 180 days.
  - Standard - 12 hours.
  - Bulk - 48 hours.

S3 Intelligent-Tiering:

- Small monthly monitoring and auto tiering fee.
- Moves objects automatically between access tiers based on usage.
- No retrieval charges in this class.

Access Tiers (4):

- Frequent AT - default tier.
- Infrequent AT - 30 days.
- Archive Instant AT - 90 days.
- Archive AT (optional) - 3-5 hours.
- Deep Archive AT (optional) - 12-48 hours.

S3 Encryption:

- Four types:
  - Server Side Encryption (SSE)
  - Client Side Encryption (CSE)
- In the UI - by default when upload a file to the server encrypt the file using AES-256.
- In the CLI - about upload a file after encrypting a server side encrypt after encrypting.

SSE: Amazon Managed for S3:

- It manages the access policies of the buckets.
- Evaluates S3 Bucket policies, S3 ACLs, S3 Access point policies.

AWS KMS Service:

- Highly secure managed service to create and manage data at the edge, and regulate data with end end of life.

Integration:
- Amazon, Elastic Block Storage, Elasticache Edge Computing.
- Serverless Elastic Edge.

Data Migration with AWS Snowball:

Challenges:
- Limited Connectivity.
- Limited Bandwidth.
- High network cost.
- Shared Bandwidth.
- Connection Stability.

Basics Image:

Snowball Edge:

- Physical device used for transferring data in a out of box.
- They provide Tamper Proof.
- Snowball Edge Capabilities:
  - Storage Optimized (80TB usable).
  - Compute Optimized (210 TB of RAM/40 TB NVME).
- Provides block storage and S3 compatible object storage.

AWS Snowcone (Snowcone SSD):

- Small portable computing device.
- Rugged & secure, withstand harsh environments.
  - Snowcone - 8 TB of HDD.
  - Snowcone SSD - 14 TB of SSD.
- Use snowcone when snowball doesn't fit.
- Can be sent back to AWS offline or connect it to internet (snowcone) and AWS DMS/DS to send data.

AWS Snowmobile:

- 39 ft in length truck.
- 29 ft in length of container.
  - 1.25 - 100 PB - 1 truck 100 PB.
- Lease snowmobile for 10 PB of capacity.

Edge Computing:

- Process data while it being created on an edge location.
- Analyze a real-time data / patterns in the edge computing.
- Long term deployment options:
  - 1 to 4 years discounted pricing.

AWS OpsHub:

- To run snowball devices, we could use Command line tool.
  - It was not an out-of-box tool to manage a snowball device.
    - Unlocking/configuring devices.
    - Transferring files.
    - Launching & managing instances.
    - Launch Compatible AWS service.

Pricing:

- You pay for device usage and charge per day.
- On demand:
  - Includes a machine learning for job / storage.
  - Pay for only what used days.
- Committed Upfront:
  - Pay in advance for month/year.
  - Upto 60% discounted pricing.

Hybrid Cloud for Storage:

Storage Gateway:

- A fleet of things in cloud and in the on premises.
- Attach to backup between on-prem data & cloud in S3.

Types:

- File Gateway.
- Volume Gateway.
- Tape Gateway.