# Simple Storage Service (S3)

## S3 Introduction:

- Amazon S3 is one of the main building blocks of AWS
- It's advertised as "infinitely scaling" storage
- Many websites use Amazon S3 as a backbone
- Many AWS services use Amazon S3 as an integration as well
- We'll have a step-by-step approach to S3

## Amazon S3 Use cases:

- Backup and storage
- Disaster Recovery
- Archive
- Hybrid Cloud Storage
- Application hosting
- Data lakes & big data analytics
- Software delivery
- Static website

## Amazon S3 – Buckets:

- Amazon S3 allows people to store objects (files) in buckets (directories)
- Buckets must have a globally unique name (across all regions all accounts)
- Buckets are defined at the region level
- S3 looks like a global service but buckets are created in a region
- Naming convention
    - No Uppercase, No lowercase
    - 3-63 characters long

- Not an IP
- Must start with lowercase letter or number
- Must NOT start with the prefix
- Must NOT end with the suffix

## Amazon S3 – Objects:

- Objects (files) have a key
- The Key is the full path:
  - S3://my-bucket/my_file.txt
  - S3://my-bucket/EC2-instance_logs/25-10-2024/my_file.log
- The key is composed of prefix + object name
  - S3://my-bucket/ EC2-instance_logs/25-10-2024/my_file.log
- There's no concept of "directories" with in buckets (although the UI will trick you to think otherwise)
- Just keys with very long names that contain slashes ("/")

## Amazon S3 – Objects (cont.):

- Objects values are the content of the body:
  - Max, Object Size is 5TB (5000GB)
  - 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 – up to 10) – useful for security / lifecycle
- Version ID (if versioning is enabled)

## Amazon S3 – Security:

- User-Based:
  - IAM Policies – which API calls should be allowed for a specific user from IAM
- Resource-Based:
  - Bucket Policies – bucket wide rules from the S3 console – allows cross account
  - Object Access Control List (ACL) – finer grain (can be disable)

    - Bucket Access Control List (ACL) – less common (can be disabled)
- Note: an IAM principle can access an S3 object if
    - The user IAM permission ALLOW it OR the resource policy ALLOWS it AND there's no expect DENY
- Encryption: encrypt objects in Amazon S3 using encryption keys

## S3 Bucket Policies:

- JSON based policies
    - Resource: buckets and objects
    - Effect: Allow / Deny
    - Actions: Set of API to Allow or Deny
    - Principle: The account or user to apply the policy
- 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)

## 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)
    - http://bucket-name.s3-website-aws-region.amazonaws.com OR
    - http://bucket-name.s3-website.aws-region.amazonaws.com
- 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": 1,2,3…
- It is best practice to version your buckets

- Protect against unintended deletes (ability to restore a version)
- Easy roll back 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 and destination buckets
- Cross-Region Replication (CRR)
- Same-Region Replication (SRR)
- Buckets can be in different AWS accounts
- Copying is a 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 and test accounts
- After you enable Replication, only objects are replicated
- Optionally, you can replicate existing objects using S3 Batch Replication
  - Replicates existing objects and objects that failed replication

- For DELETE operations
  - Can replicate delete markers from source to target(optional setting)
  - Deletions with a version ID are not replicated (to avoid malicious deletes)

- There is no “chaining” of replication

- If bucket I has replication into bucket 2, which has replication into bucket 3.
- Then objects created in bucket 1 are not replicated to bucket 3.

## S3 Storage Classes:

- Amazon S3 Standard – General Purpose
- Amazon S3 Standard – Infrequent Access (IA)
- Amazon S3 One Zone-Infrequent Access
- 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 S3 Lifecycle configurations.

## S3 Durability and Availability:

- Durability:
  - High durability (99.99999999999%, 11 9's) of objects across multiple AZ
  - If you store 10,000,000 objects with amazon S3, you can on average expect to incur a loss of single object every 10,000 years
  - Same for all storage classes
- Availability:
  - Measures how readily available a service is
  - Varies depending on storage class
  - Example: S3 standard has 99.99% availability = not available 53 minutes a year

## Amazon S3 Standard – General Purpose:

- 99.99% Availability

- Used for frequent accessed data
- Low latency and high throughput
- Sustain 2 concurrent facility failures

- Use Cases: Big Data analytics, mobile & gaming applications, content distribution…

## Amazon S3 Standard – Infrequent Access (IA):

- For data that is less frequent accessed, but requires rapid access when needed.
- Lower cost that S3 Standard

- Amazon S3 Standard-Infrequent Access (S3 Standard-IA)
  - 99.9% Availability
  - Use Cases: Disaster Recovery, backup's

- Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)
  - High durability (99.9999999999%) in a single AZ; data lost when AZ is destroyed
  - 99.5% Availability
  - Use Cases: Storing secondary backup copies of on-premise data you can recreate

## Amazon S3 Glacier Storage Classes:

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

- **Amazon S3 Glacier Instant Retrieval**
  - Millisecond retrieval, great for data accesses once a quarter
  - Minimum storage duration of 90 days
- **Amazon S3 Glacier Flexible Retrieval (**formerly Amazon S3 Glacier**):**

    - Expedited (1 to 5 minutes), standard (3 to 5 hours), Bulk (5 to 12 hours) – free
    - Minimum storage duration of 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 and 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 30days
- Archive Instant Access tier (automatic): objects not accessed for 90 days
- Archive Access tier(optional): comfortable from 90 days to 700+ days
- Deep Archive Access tier (optional): config. From 180 days to 700+ days