NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SkillsFuture Career Transition Program

Cloud Infrastructure Engineering - AWS Storage

Nanyang Technological University
Skills Union

2022/2023

S3 Overview

- S3 → Simple Storage Service
- Allows users 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 underscore
  - 3-63 characters long
  - Not an IP
  - Must start with lowercase letter or number

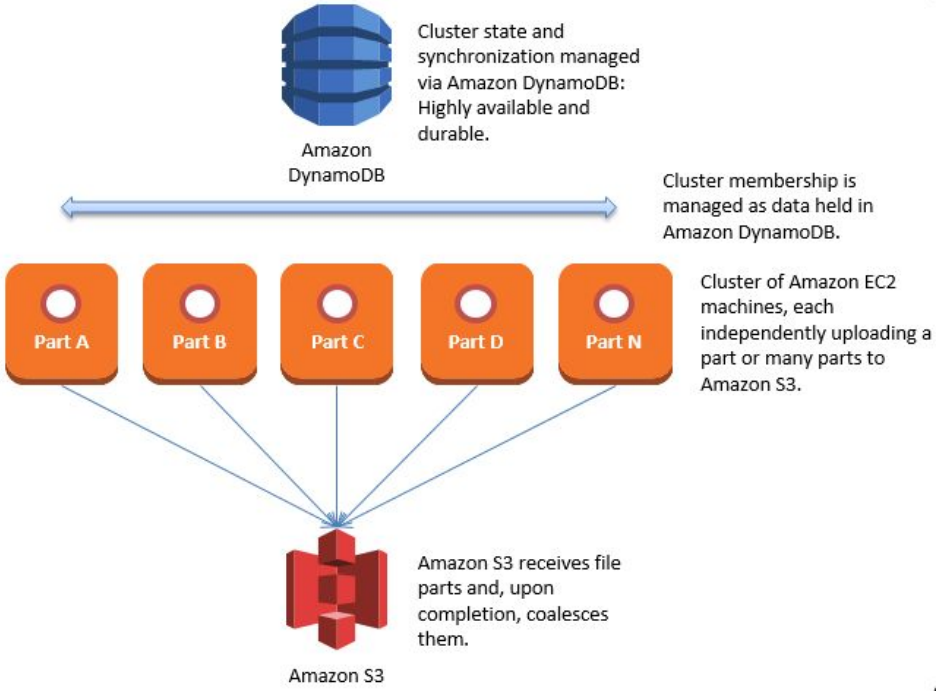
S3

- Objects (files) have a Key
- The key is the FULL path:
  - s3://my-bucket/my_file_name.pdf
  - s3://my-bucket/my_folder1/my_subfolder1/a_new_folder/my_file.txt
- The key is composed of prefix + object name
  - s3://my-bucket/my_folder1/my_subfolder1/a_new_folder/my_file.txt
- There's no concept of "directories" within buckets (although the UI will trick you to think otherwise)
- Just keys with very long names that contain slashes ("/")

S3

- Object 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)

S3 - Multi-Part Upload



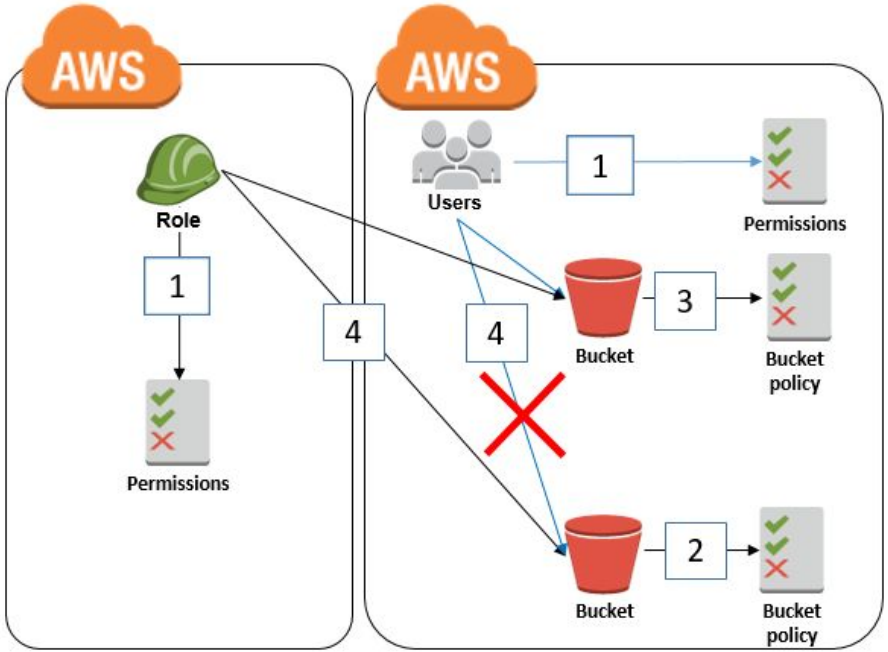
S3 Use Cases

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

S3 Security

- User based
  - IAM policies - which API calls should be allowed for a specific user from IAM console
- Resource Based
  - Bucket Policies - bucket wide rules from the S3 console - allows cross account
  - Object Access Control List (ACL) – finer grain
  - Bucket Access Control List (ACL) – less common
- Note: an IAM principal can access an S3 object if
  - 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

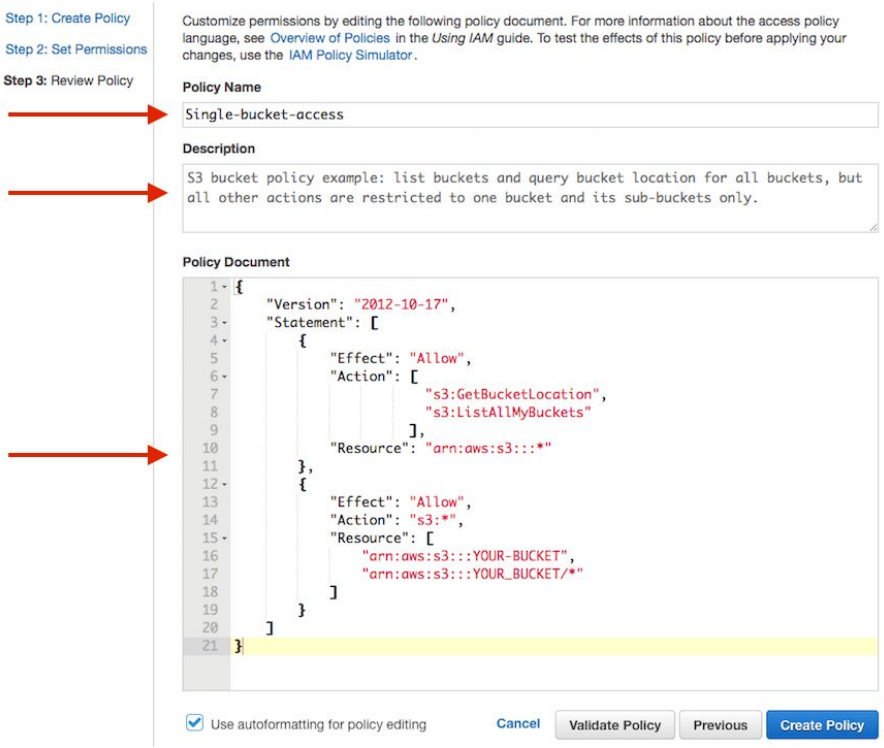
S3 Security


AWS
AWS
Role
1
Permissions
4
4
Users
1
Permissions
Bucket
3
Bucket policy
Bucket
2
Bucket policy

S3 Bucket Policies

- JSON based policies
  - Resources: buckets and objects
  - Actions: Set of API to Allow or Deny
  - Effect: Allow / 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)

S3 Bucket Policies



S3 Bucket Settings

- As much as possible, **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

S3 Bucket Settings

Block *all* public access
On

- **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
  On
- **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
  On
- **Block public access to buckets and objects granted through *new* public bucket or access point policies**
  On
- **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**
  On

S3 Websites

- S3 can host static websites and have them accessible on the www
- The website URL will be:
  - <bucket-name>.s3-website-<AWS-region>.amazonaws.com

- If you get a 403 (Forbidden) error, make sure the bucket policy allows public reads!

S3 Versioning

- You can version your files in Amazon S3
- It is enabled at the bucket level
- Same key overwrite will increment 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

S3 Access Logs

- For audit purpose, you may want to **log all access to S3 buckets**
- Any request made to S3, from any account, authorized or denied, will be logged into another S3 bucket
- That data can be analyzed using data analysis tools...
- Very helpful to come down to the root cause of an issue, or audit usage, view suspicious patterns, etc.

S3 Data Replication

- Must enable versioning in source and destination
  - Cross Region Replication (CRR)
  - Same Region Replication (SRR)
- Buckets can be in different accounts
- Copying is asynchronous
- Must give proper IAM permissions to S3

- CRR - Use cases: compliance, lower latency access, replication across accounts
- SRR – Use cases: log aggregation, live replication between production and test accounts

S3 Storage Classes

- Amazon S3 Standard - General Purpose
- Amazon S3 Standard-Infrequent Access (IA)
- Amazon S3 One Zone-Infrequent Access
- Amazon S3 Intelligent Tiering
- Amazon Glacier
- Amazon Glacier Deep Archive

S3 Durability & Availability

- Durability:
  - High durability (99.999999999%, 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 a single object once every 10,000 years
  - Same for all storage classes

- Availability:
  - Measures how readily available a service is
  - S3 standard has 99.99% availability, which means it will not be available 53 minutes a year
  - Varies depending on storage class

S3 Standard - GP

- 99.99% Availability
- Used for frequently accessed data
- Low latency and high throughput
- Sustain **2 concurrent facility failures**
- Use Cases:
  - Big Data analytics,
  - mobile & gaming applications,
  - content distribution,

S3 Standard - IA

- Suitable for data that is less frequently accessed, but requires rapid access when needed
- 99.9% Availability
- Lower cost compared to Amazon S3 Standard, but retrieval fee
- Sustain 2 concurrent facility failures

- Use Cases:
  - Data store for disaster recovery
  - Backups

S3 Intelligent Tiering

- 99.9% Availability
- Same low latency and high throughput performance of S3 Standard
- Cost-optimized by automatically moving objects between two access tiers based on changing access patterns:
  - Frequent access
  - Infrequent access
- Resilient against events that impact an entire Availability Zone

S3 One Zone - IA

- Same as IA but data is stored in a single AZ
- 99.5% Availability
- Low latency and high throughput performance
- Lower cost compared to S3-IA (by 20%)
- Use Cases:
  - Storing secondary backup copies of on-premise data, or
  - Storing data you can recreate

S3 Glacier/ Glacier Deep Dive

- Low cost object storage (in GB/month) meant for archiving / backup
- Data is retained for the longer term (years)
- Various retrieval options of time + fees for retrieval:
- Amazon Glacier – cheap:
  - Expedited (1 to 5 minutes)
  - Standard (3 to 5 hours)
  - Bulk (5 to 12 hours)
- Amazon Glacier Deep Archive – super cheap:
  - Standard (12 hours)
  - Bulk (48 hours)

S3 Comparison

	S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	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)
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved

What’s Next?