Chapter 4: S3

Rahul Daware

May 11, 2018

What is S3?

S3 provides developers and IT teams with secure, durable, highly scalable object storage. Amazon S3 is easy to use, with a simple web services interface to store and retrieve any amount of data from anywhere on the web. S3 is a safe place to store your files. It is object based storage. The date is spread across multiple devices and facilities.

S3 - The Basics

- S3 is object based - i.e. allows you to upload files
- Files can be from 0 bytes to 5 TB.
- There is unlimited storage.
- Files are stored in buckets. (Buckets are folder in cloud. Each bucket has universal namespace. Each bucket will have a DNS address. Bucket names have to unique.
- S3 is a universal namespace. i.e. names must be unique globally.
- URL look like this : https://s3-eu-west-1.amazonaws.com/acloudguru
- When you upload a file to S3, you will receive a HTTP 200 if the upload was successful
- Built for 99.99% availability for the S3 platform
- Amazon guarantee 99.9% availability
- Amazon guarantees 11 9s durability for S3 information
- Tiered Storage available
- Lifecycle Management
- Versioning
- Encryption
- Secure your data using Access Control Lists and Bucket Policies

Data Consistency Model for S3

- Read after write consistency for PUTS of new objects
- Eventual consistency for overwrite PUTS and DELETES (can take some time to propogate)

S3 is a simple key-value store

S3 is object based. Objects consists of the following:

- Key (This is simply name of the object
- Value ((This is simply the data and is made up of a sequence of bytes
- Version ID (Important for versioning)
- Metadata (Data about data you are storing)
- Subresources
  - Access Control Lists
  - Torrent

S3 - Storage Tiers/Classes

- **S3 Standard** : 99.99% availability 11 9s durability, stored redundantly across multiple facilities, and is designed to sustain the loss of 2 facilities concurrently.
- **S3 -IA** : (Infrequently Accessed): For data that is accessed less frequently, but requires rapid access when needed. Lower fee than S3, but you are charged a retrieval fee.
- **S3 One Zone - IA** Want a lower cost option for infrequently accessed data, but do not require the multiple availability zone data resilience.
- **Glacier** Very cheap, but used for archival only. Expedited, standard or bulk. A standard retrievel time takes 3-5 hours.

S3 - Charges

Charged for:

- Storage (Per GB basis)
- Requests (No. of requests)
- Storage Management Pricing (Charged for tags on data)
- Data Transfer Pricing (Transferring data from one region to other)
- Transfer Acceleration : Amazon S3 Transfer acceleration enables fast, easy and secure transfers of files over long distances between your end users and an S3 bucket. Transfer acceleration takes advantage of Amazon CloudFront's globally distributed edge locations. AS the data arrives at an edge location, data is routed to Amazon S3 over an optimized network path.

S3 Exam Tips for S3 101

- Remember that S3 is object based. i.e. allows you to upload files
- Files can be from 0 bytes to 5 TB
- There is unlimited storage.
- Files are stored in buckets
- S3 is a universal namespace. That is names must be unique globally
- Read afte write consistency for PUTS of new objects
- Eventual Consistency for overwrite PUTS and DELETES (can take some time to propogate)
- Storage Classes/Tiers:
  - S3 (durable, immediately available, frequently accessed)
  - S3 -IA (Durable, immediately available, infrequently accessed)
  - S3 One Zone - IA (event chaeaper than IA, but onle in one AZ)
  - Glacier - Achived data, where you cn wait 3-5hours before accessing
- Remember core fundamentals of S3 object - Key (name), Value (data) , Version ID, Metadata, Subresources (ACL, Torrent)
- Object based storage only (for files)
- Not suitable to install an operating system on
- Successful uploads give HTTP 200 response

Create an S3 Bucket - Exam Tips

- Buckets are a univeral name space
- Upload an object to S3 receive a HTTP 200 code
- S3, S3 -IA, S3 - One Zone IA, S3 Reduced Redundancy Storage
- Encryption
  - Client Side Encryption
  - Server Side Encryption with Amazon S3 Managed Keys (SSE-S3), Server Side Encryption with KMS (SSE-KMS), Server Side Encryption with Customer provided keys (SSE-C)
- Control access to buckets using either a bucket ACL or using bucket policies
- By Default buckets are private and all objects stored inside them are private

S3 Versioning - Exam Tips

- Stores all versions of an object (including all writes and even if you delete an object)
- Great backup tool
- Once enabled, versioning cannot be disabled, only suspended
- Integrates with lifecycle rules
- Versioning's MFA Delete capability, which uses multi-factor authentication, can be used to provide an additional layer of security

S3 - Cross Region Replication Exam Tips

- Versioning must be enabled on both the source and destination buckets
- Regions must be unique
- Files in an existing bucket are not replicated automatically. All the subsequent updated files will be replicated automatically
- You cannot replicate to multiple buckets or use daisy chaining (at this time)
- Delete markers are replicated
- Deleting individual versions or delete markers will not be replicated
- Understand what cross region replication is at high level

S3 Lifecycle Management Lab

- Can be used in conjunction with versioning
- Can be applied to current versions and previous versions
- Following actions can now be done:
  - Transition to the standard -infrequent access storage class (minimum file size should be 128 KB and 30 days after the creation date)
  - Archive to the glacier storage class (30 days after IA, if relevant)
  - Permanently Delete

Notes

- Read the S3 FAQs before taking the exam. It comes up a lot.