Simple Storage Service – S3

Ponnam Phani Krishna
PONNAM.PHANI@GMAIL.COM

Simple Storage Service – S3

Amazon Simple Storage Service (Amazon S3) is an object-based storage service that offers industry-leading scalability, data availability, security, and performance. This means customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as data lakes, websites, mobile applications, backup and restore, archive enterprise applications, IoT devices and big data analytics.

Amazon S3 provides easy-to-use management features so you can organize your data and configure finely tuned access controls to meet your specific business, organizational, and compliance requirements.

Amazon S3 is designed for **99.999999999% of durability**, and stores data for millions of applications for companies all around the world.

S3 Benefits:

- Easily manae data and access controls
- Industry leading Performance, scalability, availability, and durability
- Wide range of cost-effective storage classes
- Unmatched security, compliance and audit capabilities
- Most supported cloud storage service

Use cases:

- Backup and Restore
- Disaster Recovery (DR)
- Archive
- Data lakes and big data analytics
- Hybrid Cloud storage
- Cloud-native applications

Amazon S3 Concepts:

This section describes key concepts and terminology you need to understand to use Amazon S3 effectively.

- Buckets
- Objects
- Keys
- Amazon S3 data consistency model

Buckets:

A bucket is a container for objects stored in Amazon S3. Every object is contained in a bucket. Every object will get a url to access them.

While creating the buckets we need to assign a name for the bucket and it must follow some rules.

The following rules apply for naming buckets in amazon S3:

- Bucket name must be unique across globally
- Bucket names must be between 3 and 63 characters long.
- Bucket names can consist only of lowercase letters, numbers, dots (.), and hyphens (-).
- Bucket names must begin and end with a letter or number.
- Bucket names must not be formatted as an IP address (for example, 192.168.5.4)
- Buckets used with Amazon S3 Transfer Acceleration can't have dots (.) in their names.

Example bucket names:

The following example bucket names are valid and follow the recommended naming guidelines:

- ***docexamplebucket1***
- ***log-delivery-march-2020***
- ***my-hosted-content***

The following example bucket names are valid but not recommended for uses other than static website hosting:

- ***docexamplewebsite.com***
- ***www.docexamplewebsite.com***
- ***my.example.s3.bucket***

Objects: Objects are the fundamental entities stored in amazon S3, ex: files,folders etc..

Keys: A key is the unique identifier for an object with in a bucket. Every object in abucket has exactly one key. The combination of a bucket, key and version ID uniquely identify each object.

Amazon S3 data consistency model: Amazon S3 provides strong read-after-write consistency for PUTs and DELETEs of objects in your Amazon S3 bucket in all AWS Regions. This applies to both writes to new objects as well as PUTs that overwrite existing objects and DELETEs

Amazon S3 achieves high availability by replicating data across multiple servers within AWS data centers. If a PUT request is successful, your data is safely stored.

Bucket configurations have an eventual consistency model. Specifically: If you delete a bucket and immediately list all buckets, the deleted bucket might still appear in the list.

Amazon S3 Features:

Following are the important features of Amazon S3:

- Storage Classes
- Bucket policies
- Versioning
- Static Website Hosting
- Versioning
- Encryption
- Transfer Acceleration
- Object Lock
- Cross Region Replication
- Life Cycle rules

Storage Classes:

Amazon S3 offers a range of storage classes designed for different use cases. Following are the list of storage classes available on Amazon S3:

- S3 Standard
- S3 Intelligent-Tiering
- S3 Standard-IA
- S3 One Zone-IA
- Reduced Redundancy Storage
- Glacier
- Glacier Deep Archive

S3 Standard: S3 standard is the default storage class. if you don't specify the storage class when you upload an object, Amazon S3 assigns the S3 standard storage class.

- S3 Standard storage class is designed for frequently accessed Data and there is no retrieval fees for data accessing.
- S3 standard is designed to provide 99.999999999% of Durability
- S3 standard is designed to provide 99.99% of availability
- Data can be maintained in >=3 Availability zones
- There is no Min storage duration

S3 Intelligent-Tiering: S3 Intelligent-Tiering is an Amazon S3 storage class designed to optimize storage costs by automatically moving data to the most cost-effective access tier, without operational overhead. S3 Intelligent-tiering is the perfect storage class when you want to optimize storage costs for data that has unknown or changing access patterns.

Objects that are uploaded or transitioned to S3 Intelligent-Tiering are automatically stored in the *Frequent Access* tier. S3 Intelligent-Tiering works by monitoring access patterns and then moving the objects that have not been accessed in 30 consecutive days to the *Infrequent Access* tier.

You can also choose to activate one or both of the archive access tiers. After you activate one or both of the archive access tiers, S3 Intelligent-Tiering automatically moves objects that haven't been accessed for 90 consecutive days to the *Archive Access* tier, and after 180 consecutive days of no access, to the *Deep Archive Access* tier.

If the objects are accessed later, the objects are moved back to the *Frequent Access* tier. There are no retrieval fees, so you won't see unexpected increases in storage bills when access patterns change.

- S3 Intelligent-Tiering is designed for Long-Lived data with changing or unknow access patterns
- Provides 99.999999999% of durability
- Provides 99.9% of availability
- Data can be maintained in >=3 Availability Zones
- Minimum object storage duration must be 30 Days
- Monitoring and automation fees per object apply, No retrieval fees.

S3 Standard-IA & S3 One Zone-IA:

The **S3 Standard-IA** and **S3 One Zone-IA** storage classes are designed for long-lived and infrequently accessed data. (IA stands for *infrequent access*.) S3 Standard-IA and S3 One Zone-IA objects are available for millisecond access (similar to the S3 Standard storage class). Amazon S3 charges a retrieval fee for these objects, so they are most suitable for infrequently accessed data.

These storage classes Differs as follows:

S3 Standard-IA : Amazon S3 stores the object data redundantly across multiple geographically separated Availability Zones (similar to the S3 Standard storage class). S3 Standard-IA objects are resilient to the loss of an Availability Zone. This storage class offers greater availability and resiliency than the S3 One Zone-IA class. It offers **99.9% Availability**

S3 One Zone-IA: Amazon S3 stores the object data in only one Availability Zone, which makes it less expensive than S3 Standard-IA. However, the data is not resilient to the physical loss of the Availability Zone resulting from disasters, such as earthquakes and floods. The S3 One Zone-IA storage class is as durable as Standard-IA, but it is less available and less resilient. It offers **99.5% Availability**

Reduced Redundancy Storage (RRS): The Reduced Redundancy Storage (RRS) class is designed for noncritical, reproducible data that can be stored with les redundancy than the S3 standard storage class.

We recommend that you not use this storage class. The S3 Standard storage class is more cost effective.

It offers 99.99% Durability and 99.99% availability

The S3 Glacier and S3 Glacier Deep Archive storage classes are designed for low-cost data archiving. These storage classes offer the same durability and resiliency as the S3 standard storage class.

S3 Glacier: Use for archives where portions of the data might need to be retrieved in minutes. Data stored in the S3 Glacier storage class has a minimum storage duration period of 90 days and can be accessed in as little as 1-5 minutes using expedited retrieval.

S3 Glacier Deep Archive: Use for archiving data that rarely needs to be accessed. Data stored in the S3 Glacier Deep Archive storage class has a minimum storage duration period of 180 days and a default retrieval time of 12 hours.

We can use the below mentioned 3 retrieval methods to retrieve the data from Glacier storage:

- Expedited Retrieval
- Standard Retrieval
- Bulk Retrieval

Bucket Policies:

Bucket policies provide centralized access control to buckets and objects based on a variety of conditions, including Amazon S3 operations, requesters, resources, and aspects of the request (for example, IP address).

The permissions attached to a bucket apply to all of the bucket's objects that are owned by the bucket owner account.

Versioning:

You can use versioning to keep multiple copies of an object in the same bucket. You can use the S3 Versioning feature to preserve, retrieve, and restore every version of every object stored in your buckets. With versioning you can recover more easily from both unintended user actions and application failures.

Versioning-enabled buckets can help you recover objects from accidental deletion or overwrite. For example, if you delete an object, Amazon S3 inserts a delete marker instead of removing the object permanently. The delete marker becomes the current object version. If you overwrite an object, it results in a new object version in the bucket. You can always restore the previous version.

After enabling the versioning on your S3 bucket, we cannot disable it. It allows only to suspend the versioning, which keeps all existing versions as is.

Encryption: Encryption can be used to secure & protect the data while in-transit (as it travels to and from Aazon S3) and at rest (while it is stored on disks in amazon S3 data centers). You can protect data in transit using SSL/TLS or client-side encryption.

You have the following options for protecting data at rest in amazon S3:

- **Server-Side Encryption** – Request Amazon S3 to encrypt your object before saving it on disks in its data centers and then decrypt it when you download the objects.
- **Client-Side Encryption** – Encrypt data client-side and upload the encrypted data to Amazon S3. In this case, you manage the encryption process, the encryption keys, and related tools.

Server-Side Encryption:

- **Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3):** When you use Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3), each object is encrypted with a unique key. As an additional safeguard, it encrypts the key itself with a master key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt your data.
- **Server-Side Encryption in AWS Key Management Service (SSE-KMS):** SSE-KMS is similar to SSE-S3, but with some additional benefits and charges for using this service. We are going to create our own encryption keys using KMS Services and amazon will maintain the keys and encryption for you.
- **Server-Side Encryption with Customer-Provided Keys (SSE-C):** Using SSE-C, you manage the encryption keys and Amazon S3 manages the encryption, as it writes to disks, and decryption, when you access your objects.

S3 Transfer Acceleration: Amazon S3 Transfer Acceleration (S3-TA) can speed up content transfers to and from amazon S3 by as much as 50-500% for long-distance transfer of larger objects.

Benefits:

- Move Data faster over long distances
- Reduce network variability
- Shorten the distance to S3
- Maximize bandwidth utilization

Requirements for using Transfer Acceleration:

- The name of the busket used for transfer acceleration must not contain periods (“.”)
- Transfer acceleration must be enabled on the bucket.

After you enable the transfer acceleration on a bucket, it might take upto 20 m inutes before the data transfer speed to the bucket increases.

To access the bucket that is enabled for transfer acceleration, you must use the endpoint ***bucketname.s3-accelerate.amazonaws.com***

Static Website Hosting: You can use amazon S3 to host a static website. On a static website, individual webpages include static content.

By contrast, a *dynamic* website relies on server-side processing, including server-side scripts such as PHP, JSP, or ASP.NET. Amazon S3 does not support server-side scripting.

S3-Object Lock:

With S3 Object Lock, you can store objects using a *write-once-read-many* (WORM) model. Object Lock can help prevent objects from being deleted or overwritten for a fixed amount of time or indefinitely. You can use Object Lock to help meet regulatory requirements that require WORM storage, or to simply add another layer of protection against object changes and deletion.

Object Lock provides two ways to manage object retention: *retention periods* and *legal holds*.

- **Retention period** — Specifies a fixed period of time during which an object remains locked. During this period, your object is WORM-protected and can't be overwritten or deleted.
- **Legal hold** — Provides the same protection as a retention period, but it has no expiration date. Instead, a legal hold remains in place until you explicitly remove it. Legal holds are independent from retention periods.

Object Lock works only in versioned buckets, and retention periods and legal holds apply to individual object versions. When you lock an object version, Amazon S3 stores the lock information in the metadata for that object version. Placing a retention period or legal hold on an object protects only the version specified in the request. It doesn't prevent new versions of the object from being created.

Retention Mode:

S3 Object lock provides two retention modes:

- **Governane Mode**
- **Compliance Mode**

Governance Mode: users can't overwrite or delete an object version or alter its lock settings unless they have special permissions. With governance mode, you protect objects against being deleted by most users, but you can still grant some users permission to alter the retention settings or delete the object if necessary.

To override or remove governance-mode retention settings, a user must have the **s3:BypassGovernanceRetention** permission and must explicitly **include x-amz-bypass-governance-retention:true** as a request header with any request that requires overriding governance mode.

Compliance Mode: a protected object version can't be overwritten or deleted by any user, including the root user in your AWS account. When an object is locked in compliance mode, its retention mode can't be changed, and its retention period can't be shortened. Compliance mode helps ensure that an object version can't be overwritten or deleted for the duration of the retention period.

Legal Hold: With Object Lock you can also place a *legal hold* on an object version. Like a retention period, a legal hold prevents an object version from being overwritten or deleted. However, a legal hold doesn't have an associated retention period and remains in effect until removed. Legal holds can be freely placed and removed by any user who has the **s3:PutObjectLegalHold** permission.

Cross Region Replication:

Replication enables automatic, asynchronous copying of object across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS Account or by different accounts.

Objects may be replicated to a single destination bucket or multiple destination buckets. Destination buckets can be in different AWS regions or with in the same region as the source bucket.

By default, replication only supports copying new amazon S3 objects after it is enabled. You can use replication to copy existing objects and clone them to a different bucket, but in order to do so, you can use ***aws s3 sync*** command through amazon cli

Note: Versioning must be enabled on bit surce and destination buckets to enable cross-region replication.

Replication can help you do the following:

- Replicate objects while retaining metadata
- Replicate object into different storage classes
- Maintain object copies under different ownerships
- Keep object copies over multiple AWS Regions
- Replicate objects with in 15 minutes

Life Cycle Rules: To manage your object so that they are stored cost effectively throughout their life cycke, configure their amazon S3 lifecycle. An S3 life cycle configuration is a set of rules that define actions that amazon S3 applies to a group of objects.

There are two types of actions:

- **Transition actions:** Define which objects transition to another storage class. For example, you might choose to transition objects to S3 Standard-IA storage class 30 days after you created them
- **Expiration Actions:** Define when objects expire. Amazon S3 deletes expired objects on your behalf.