Simple Storage Service (S3)

Kiran

Table of Contents

1. Simple Storage Service (S3)

1.1. Introduction

Simple Storage Service (S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance.

S3 is designed for 99.999999999% (11 9's) of durability, meaning the possibility of data loss is very very very low.

S3 stores multiple copies of any objects (files) in multiple AZs thus ensuring high levels of durability.

1.2. Advantages of S3

1. Unlimited Storage for objects: You can upload as many objects as you like into S3. Each object can be as large as 5 GB.
2. Fine grained access control: Limit who can view, update or manage the objects you upload.
3. Life Cycle Management: Move your objects across various storage classes to save costs.
4. Multiple ways to access: S3 could be accessed through aws console, S3 api and using aws SDK.

1.3. S3 Objects

A S3 bucket is a logical unit to upload your files and folders.

S3 bucket name is globally unique, and the namespace is shared by all AWS accounts. This means that after a bucket is created, the name of that bucket cannot be used by another AWS account in any AWS Region until the bucket is deleted. You should not depend on specific bucket naming conventions for availability or security verification purposes.

For bucket naming guidelines, see Bucket naming rules .

A bucket can be created in a region of your choice.

The selection of the bucket depends on the following factors:

1. Regulatory Requirements (Data Protection laws such as HIPPA, GDPR, SOX etc)
2. Proximity to end users to decrease latency.

1.4. Create a S3 Bucket.

1. Go to `S3 Service`
2. Enter the details a. Name - Name should be Globally Unique (Refer the previous para)

   b. Choose Region

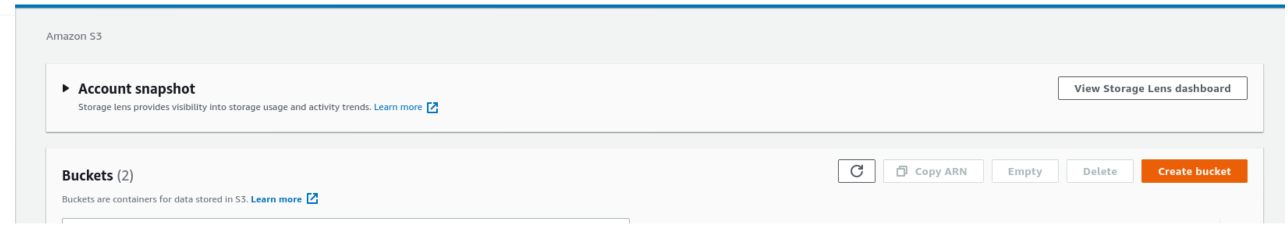
   c. The third option lets you copy settings from an existing bucket.

   d. Block all public access - Uncheck this only if you want anonymous users to have access.

   e. Bucket Version - You can keep multiple versions of an object with versioning. Read more about it here

   f. Tags

   g. Default Encryption - server-side encryption, encrypts an object before saving it to disk and decrypts it when you download the objects. Read about Protecting data using server-side encryption.



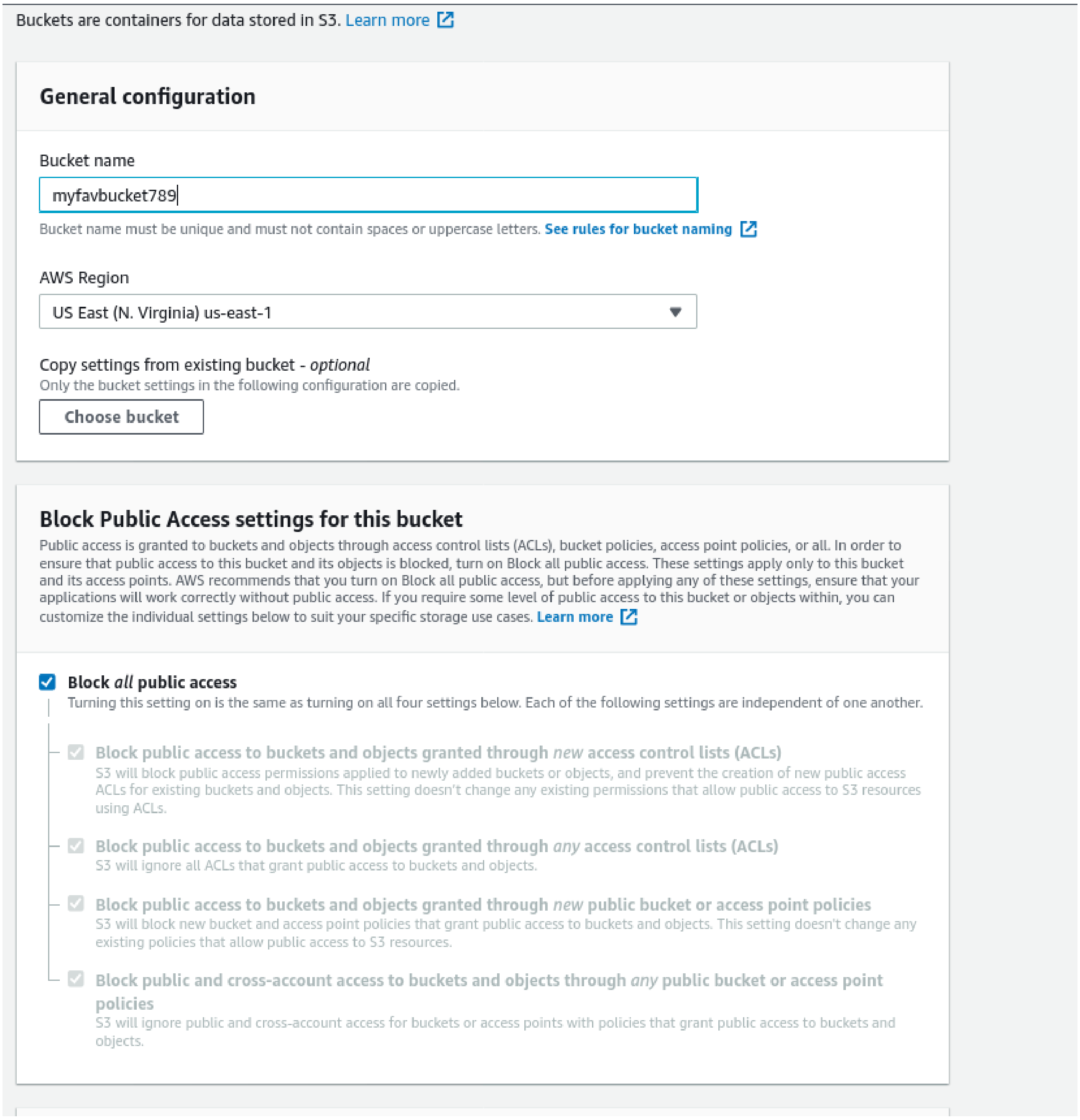
Amazon S3

Account snapshot

Storage lens provides visibility into storage usage and activity trends. Learn more

Buckets (2)

Buckets are containers for data stored in S3. Learn more



Buckets are containers for data stored in S3. Learn more

General configuration

Bucket name

myfavbucket789

Bucket name must be unique and must not contain spaces or uppercase letters. **See rules for bucket naming**

AWS Region

US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. **Learn more**

☑ **Block *all* public access**

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

- ☑ **Block public access to buckets and objects granted through *new* access control lists (ACLs)**

  S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.

- ☑ **Block public access to buckets and objects granted through *any* access control lists (ACLs)**

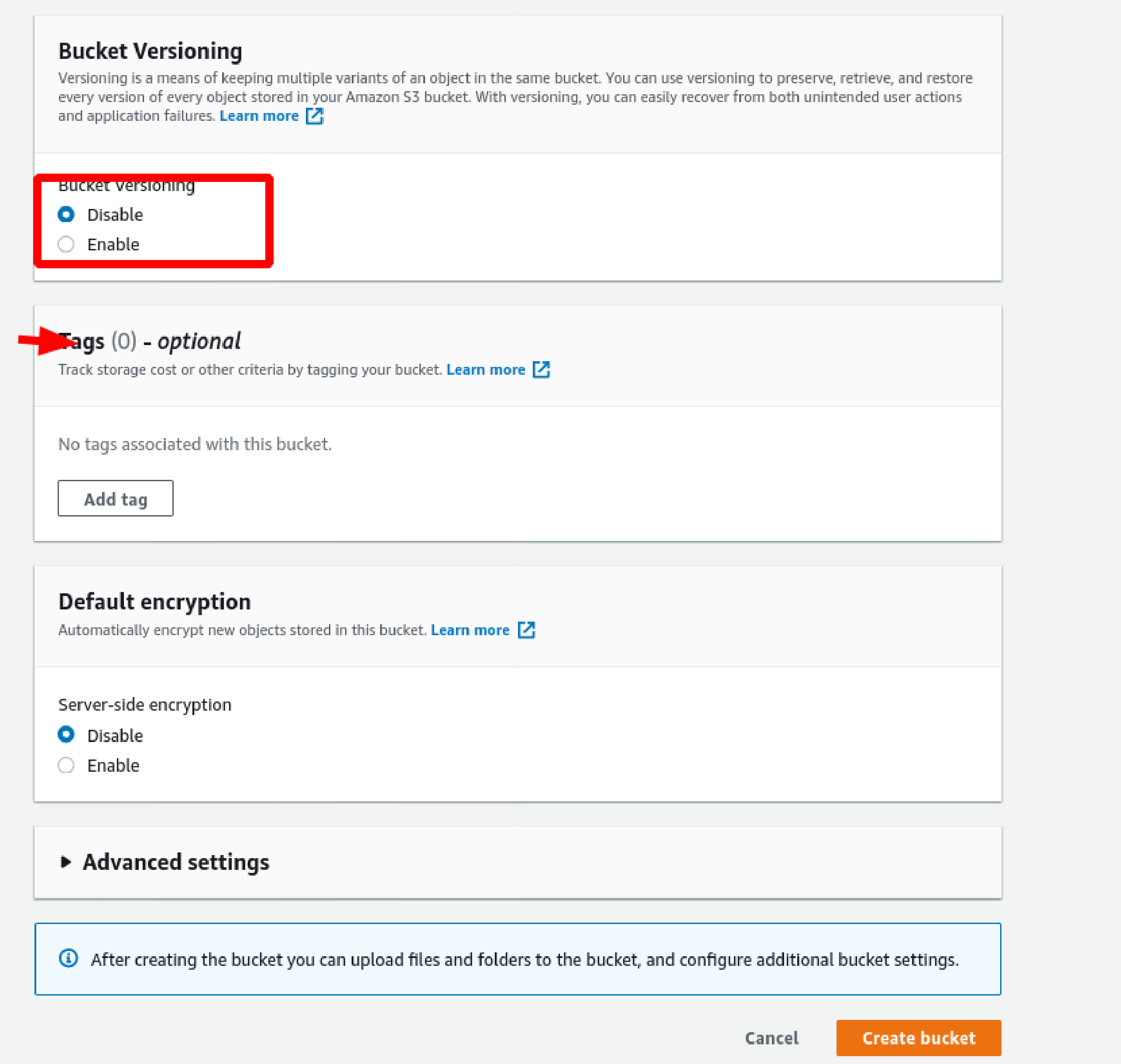
  S3 will ignore all ACLs that grant public access to buckets and objects.

- ☑ **Block public access to buckets and objects granted through *new* public bucket or access point policies**

  S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.

- ☑ **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**

  S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.



Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. Learn more

Bucket Versioning

● Disable

○ Enable

Tags (0) - *optional*

Track storage cost or other criteria by tagging your bucket. Learn more

No tags associated with this bucket.

Default encryption

Automatically encrypt new objects stored in this bucket. Learn more

Server-side encryption

● Disable

○ Enable

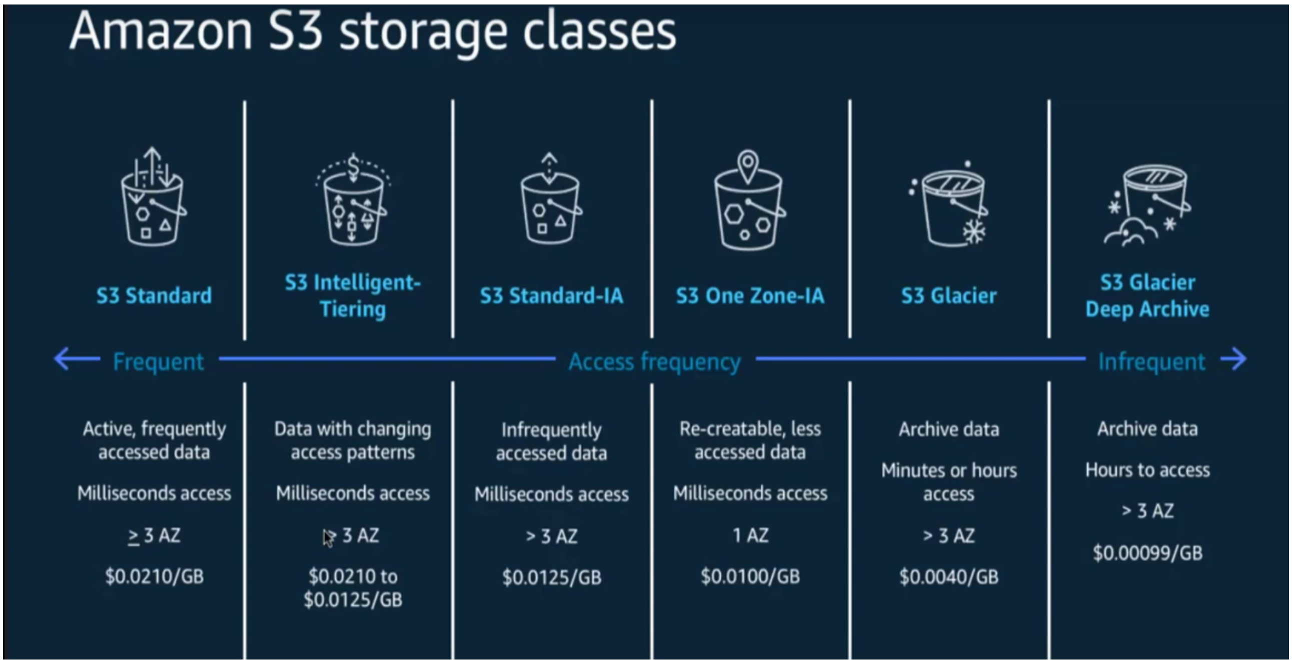
After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.

1.5. S3 Storage Classes

Amazon S3 offers a range of storage classes designed for different use cases.

S3 offers 6 Storage Classes. You can choose one of them based on the frequency with which you access objects.

Picking the right storage class helps you optimize costs



Amazon S3 storage classes

S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	S3 Glacier Deep Archive
Active, frequently accessed data	Data with changing access patterns	Infrequently accessed data	Re-creatable, less accessed data	Archive data	Archive data
Milliseconds access	Milliseconds access	Milliseconds access	Milliseconds access	Minutes or hours access	Hours to access
≥ 3 AZ	≥ 3 AZ	> 3 AZ	1 AZ	> 3 AZ	> 3 AZ
$0.0210/GB	$0.0210 to $0.0125/GB	$0.0125/GB	$0.0100/GB	$0.0040/GB	$0.00099/GB

1.5.1. S3 Standard: For general-purpose storage of frequently accessed data

Key Features:

Low latency and high throughput performance. Millisecond access.

Designed for durability of 99.999999999% of objects across multiple Availability Zones.

Resilient against events that impact an entire Availability Zone.

Designed for 99.99% availability over a given year.

Backed with the Amazon S3 Service Level Agreement for availability.

Supports SSL for data in transit and encryption of data at rest.

S3 Lifecycle management for automatic migration of objects to other S3 Storage Classes

1.5.2. S3 Intelligent-Tiering: For data with unknown or changing access

patterns. S3 Standard-IA is for data that is accessed less frequently, but requires rapid access when needed. S3 Standard-IA offers the high durability, high throughput, and low latency of S3 Standard, with a low per GB storage price and per GB retrieval fee. This combination of low cost and high performance make S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery files. S3 Storage Classes can be configured at the object level and a single bucket can contain objects stored across S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA. You can also use S3 Lifecycle policies to automatically transition objects between storage classes without any application changes.

Key Features:

Automatically optimizes storage costs for data with changing access patterns.

Stores objects in four access tiers, optimized for frequent, infrequent, archive, and deep archive access.

Frequent and Infrequent Access tiers have same low latency and high throughput performance of S3 Standard.

Activate optional automatic archive capabilities for objects that become rarely accessed.

Archive access and deep Archive access tiers have same performance as Glacier and Glacier Deep Archive.

Designed for durability of 99.999999999% of objects across multiple Availability Zones.

Designed for 99.9% availability over a given year.

Backed with the Amazon S3 Service Level Agreement for availability.

Small monthly monitoring and auto-tiering fee.

No operational overhead, no retrieval fees, no additional tiering fees apply when objects are moved between access tiers within the S3 Intelligent-Tiering storage class

1.5.3. S3 Standard-Infrequent Access (Standard-IA): S3 Standard-IA is for

data that is accessed less frequently, but requires rapid access when needed. S3 Standard-IA offers the high durability, high throughput, and low latency of S3 Standard, with a low per GB storage price and per GB retrieval fee. This combination of low cost and high performance make S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery files. S3 Storage Classes can be configured at the object level and a single bucket can contain objects stored across S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA. You can also use S3 Lifecycle policies to automatically transition objects between storage classes without any application changes.

Key Features:

Same low latency and high throughput performance of S3 Standard.

Designed for durability of 99.999999999% of objects across multiple Availability Zones.

Resilient against events that impact an entire Availability Zone.

Data is resilient in the event of one entire Availability Zone destruction.

Designed for 99.9% availability over a given year.

Backed with the Amazon S3 Service Level Agreement for availability.

Supports SSL for data in transit and encryption of data at rest.

S3 Lifecycle management for automatic migration of objects to other S3 Storage Classes.

1.5.4. One Zone-Infrequent Access (One Zone-IA):

S3 One Zone-IA is for data that is accessed less frequently, but requires rapid access when needed. Unlike other S3 Storage Classes which store data in a minimum of three Availability Zones (AZs), S3 One Zone-IA stores data in a single AZ and costs 20% less than S3 Standard-IA. S3 One Zone-IA is ideal for customers who want a lower-cost option for infrequently accessed data but do not require the availability and resilience of S3 Standard or S3 Standard-IA. It's a good choice for storing secondary backup copies of on-premises data or easily re-creatable data. You can also use it as cost-effective storage for data that is replicated from another AWS Region using S3 Cross-Region Replication. S3 One Zone-IA offers the same high durability, high throughput, and low latency of S3 Standard, with a low per GB storage price and per GB retrieval fee. S3 Storage Classes can be configured at the object level, and a single bucket can contain objects stored across S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA. You can also use S3 Lifecycle policies to automatically transition objects between storage classes without any application changes.

S3 Glacier: S3 Glacier is a secure, durable, and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions. To keep costs low yet suitable for varying needs, S3 Glacier provides three retrieval options that range from a few minutes to hours. You can upload objects directly to S3 Glacier, or use S3 Lifecycle policies to transfer data between any of the S3 Storage Classes for active data (S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA, and S3 One Zone-IA) and S3 Glacier.

Key Features:

Designed for durability of 99.999999999% of objects across multiple Availability Zones.

Data is resilient in the event of one entire Availability Zone destruction.

Supports SSL for data in transit and encryption of data at rest.

Low-cost design is ideal for long-term archive.

Configurable retrieval times, from minutes to hours.

S3 PUT API for direct uploads to S3 Glacier, and S3 Lifecycle management for automatic migration of objects.

1. S3 Glacier Deep Archive:

   Glacier deep archive is lowest-cost storage class and supports long-term retention and digital preservation for data that may be accessed once or twice in a year. It is designed for customers - particularly those in highly-regulated industries, such as the Financial Services, Healthcare, and Public Sectors - that retain data sets for 7-10 years or longer to meet regulatory compliance requirements. S3 Glacier Deep Archive can also be used for backup and disaster recovery use cases, and is a cost-effective and easy-to-manage alternative to magnetic tape systems, whether they are on-premises libraries or off-premises services. S3 Glacier Deep Archive complements Amazon S3 Glacier, which is ideal for archives where data is regularly retrieved and some of the data may be needed in minutes. All objects stored in S3 Glacier Deep Archive are replicated and stored across at least three geographically-dispersed Availability Zones, protected by 99.999999999% of durability, and can be restored within 12 hours.

   Key Features:

   Designed for durability of 99.999999999% of objects across multiple Availability Zones.

   Lowest cost storage class designed for long-term retention of data that will be retained for 7-10 years.

   Ideal alternative to magnetic tape libraries.

   Retrieval time within 12 hours.

   S3 PUT API for direct uploads to S3 Glacier Deep Archive, and S3 Lifecycle management for automatic migration of objects.

   If you have data residency requirements that can'tbe met by an existing AWS Region, you can use the `S3 Outposts` storage class to store your S3 data on-premises.

1.6. S3 Lifecycle Management.

Amazon S3 also offers capabilities to manage your data throughout its lifecycle. Once an S3 Lifecycle policy is set, your data will automatically transfer to a different storage class without any changes to your application.

An S3 Lifecycle 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 when objects transition to another Using Amazon S3 storage classes. For example, you might choose to transition objects to the S3 Standard-IA storage class 30 days after you created them, or archive objects to the S3 Glacier storage class one year after creating them.

There are costs associated with the lifecycle transition requests. For pricing information, see Amazon S3 pricing

Expiration actions: Define when objects expire. Amazon S3 deletes expired objects on your behalf.

The lifecycle expiration costs depend on when you choose to expire objects. For more information, see Expiring objects.

1.6.1. Create Lifecycle:

1. Create the bucket you'd like to set up the Lifecycle configuration.
2. Click on Management
3. Click on `Create lifecycle rule`

4, Enter the Name for the Lifecycle rule

1. Choose the Filter type a. Prefix : You can add a string that matches the prefix of multiple objects b. Object Tags: You could choose to move objects with certain tags
2. Select one or more from the below Lifecycle rule actions: a. Transition current versions of objects between storage classes.

   b. Transition previous versions of objects between storage classes.

   c. Transition previous versions of objects between storage classes

   d. Transition previous versions of objects between storage classes.

   e. Delete expired delete markers or incomplete multipart uploads.

Below is an example of a Life Cycle configuration.

It is applied on all objects beginning with `americasbusinessreport`

As per the `Transition current versions of objects between storage classes` rule, the object will be moved to `Standard-IA` after 30 days of creation of the object and then to `Glacier` after 180 days of creation of the object and then to `Glacier Deep Archive` after 365 days.

1.7. Access Control

1.7.1. Key Concepts

- Principal: An IAM entity which needs permission.
- Resource: An S3 object such as a Bucket, a file or a folder.
- ARN: Amazon resource Names or ARNs uniquely identify AWS resources. We require an ARN when you need to specify a resource unambiguously across all of AWS, such as in IAM policies, Amazon Relational Database Service (Amazon RDS) tags, and API calls. More information on ARN