## **Simple Storage Service (S3)**

* AWS S3 is object storage designed to store and retrieve any amount of data from anywhere
* It is designed for 99.999999999% durability and 99.99% availability.
* The thing that makes AWS S3 so powerful is the features that it comes preloaded with which are simply the best.

Let’s understand this with a use-case:

Large Corp is a payments organization and has more than 1000 servers. As being PCI DSS compliant, they must retain their logs for 1 year. It has been found that every day, the payment server generates logs of 200 GB.  How to achieve this use case pertaining to the storage capacity in a cost-effective manner?

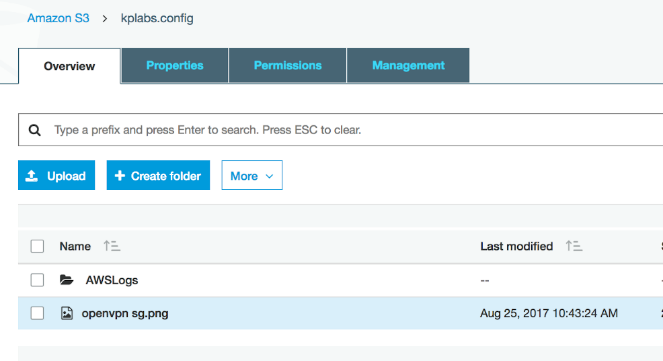


**S3 Terminology**

There are two important terminologies in AWS S3 :

* Buckets
* Objects

Buckets are like “Folders” where you can store multiple files (objects)



## **S3 Storage Classes**

S3 offers various kinds of storage classes for different use cases: -

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

**Durability vs Availability**

* Durability is percent ( % ) over one year period of time that the file which is stored in S3 will not be lost.
* Availability is percent (%) over one year period of time that the file stored in S3 will

be available.

Example:-

        For Servers, Availability is one of the key metrics and any minute of downtime is a loss.

        However, what happens if the component of the server itself fails and the server goes down?

**Understanding Every S3 Storage Class**

S3 Standard:

* Amazon S3 Standard offers high durability, availability and performance for objects stored.
* Designed for durability of 99.999999999% of objects ( eleven nines )
* Designed for 99.99% availability over a given year

Example:-

If we have 10,000 files stored in S3 ( 11 nines durability ) then you can expect to lose one file every ten million years.

**Standard In-Frequent Access:**

* Amazon S3 Standard - Infrequent Access is for data that is accessed less frequently but requires rapid access when needed.
* Designed for durability of 99.999999999% of objects
* Designed for 99.99% availability over a given year

**AWS S3 Intelligent Tiering:**

* The S3 Intelligent Tiering is primarily designed to optimize cost by automatically moving data to the most cost-effective tier.
  + General Purpose   -     Standard S3
  + Infrequent Access -    Standard IA

1TB of data stored in Standard S3  = 22.88$

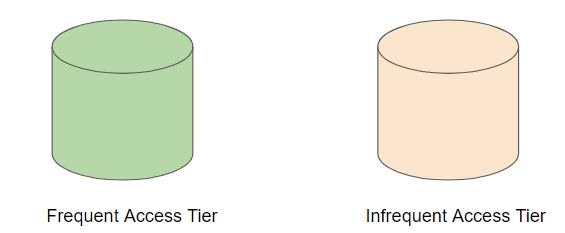
1TB of data stored in Standard IA  = 12.50$

Organization stores terabytes of data in S3.

It will be great if a solution automatically moves infrequent data to Standard IA.

The S3 Intelligent Tiering works by storing data in one of the two access tiers:

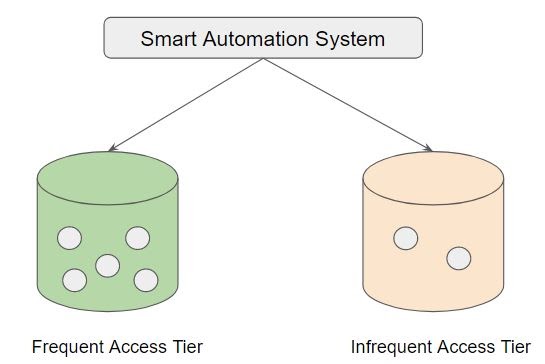
* Frequent Access Tier (Costly)
* Infrequent Access Tier (Much cheaper)



In this tier, the objects are automatically moved to frequent or infrequent access tier based on the access patterns.

Amazon S3 monitors access patterns of the objects in S3 Intelligent-Tiering and moves the ones that have not been accessed for 30 consecutive days to the infrequent access tier.

If an object in the infrequent access tier is accessed, it is automatically moved back to the frequent access tier.



This type of storage class is preferable for long-lived data with access patterns that are unknown or unpredictable.

S3 Intelligent-Tiering like other storage classes is configured at the object level.

1TB of data stored in Standard S3  = 22.88$

1TB of data stored in Standard IA  = 12.50$

1 TB of data stored with Standard-Intelligent = 23$

**S3 Reduced Redundancy Storage (RRS):**

AWS S3 Reduced Redundancy storage enables customers to reduce their costs by storing non-critical, reproducible data at lower levels of redundancy than Amazon S3’s standard storage

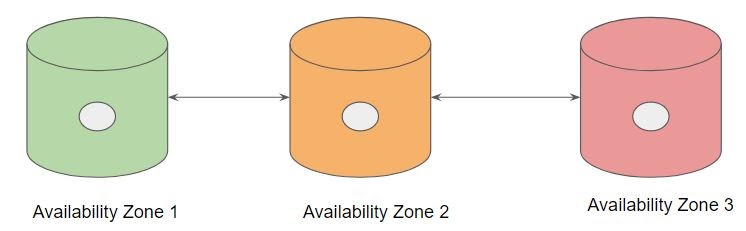
Designed for durability of 99.99% of objects

Designed for 99.99% availability over a given year

**One Zone Infrequent Access (One Zone IA)**

Storage classes like Standard S3, Standard IA stores the data in a minimum of 3 availability zones.

Due to this, the overall cost per of storage is increased with such architecture.



S3 One Zone-IA stores data in a single AZ and costs 20% less than S3 Standard-IA.

It’s a good choice for storing secondary backup copies of on-premises data or easily recreatable data.

Data will be lost in-case of availability zone destruction.



Overview of Pricing comparison between storage classes:

* 1TB of data stored in Standard S3 = 22.88$
* 1TB of data stored in Standard IA = 12.50$
* 1 TB of data stored in One Zone IA = 10$

**Glacier**

* AWS Glacier is meant to be for archiving and for storing long-term backups.
* It may take several hours for the object to get restored.
* 99.999999999% durability of object.
* It is much cheaper than other S3 other storage class ( very low cost )

Example Use Case:-

Backup of Application logs more than 1 year older can be moved to Glacier.

**Glacier Deep Archive**

S3 Glacier Deep Archive is Amazon S3’s lowest-cost storage class and supports long-term retention and digital preservation for data that may be accessed once or twice in a year.

All data stored in S3 Glacier Deep Archive can be restored within 12 hours.

On the contrary, Glacier is ideal for archives where data is regularly retrieved and some of the data may be needed in minutes.

Pricing Comparison:

1 TB of data stored in Glacier: 14$

1 TB of data stored in Glacier Deep Archive:  10.99$

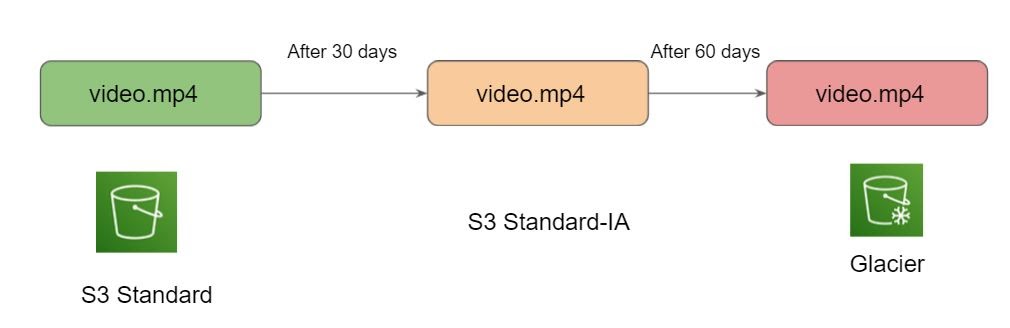
**S3 Lifecycle Policies**

We now understand that there are various S3 Storage classes offered by S3.

We need to make that storage durable + affordable for long term storage.

* We can store 1 months of logs in Amazon S3 Standard.
* Move the logs older than 1 months to S3 Standard IA
* Move the logs older than 6 year to Glacier

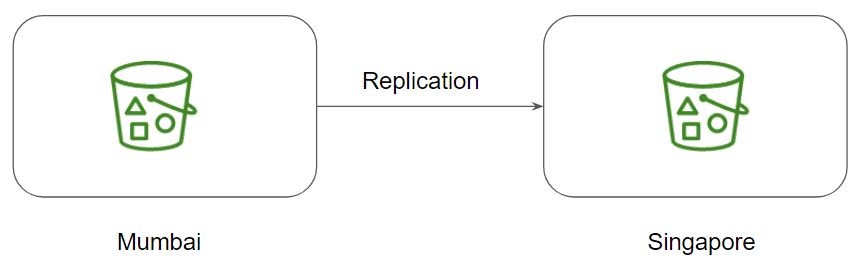
These automatic moving of data to S3 storage classes can be achieved with the help of S3 Lifecycle policies.



**S3 Cross-Region Replication**

Many compliances has a requirement that the data must be replicated across greater distances.

Cross-Region Replication allows data from S3 buckets to be replicated across regions.



Both source and destination buckets must have versioning enabled.

**S3 - Static Website Hosting**

AWS S3 allows us to host static websites.

The static website includes individual webpages which might include static content.

Note:

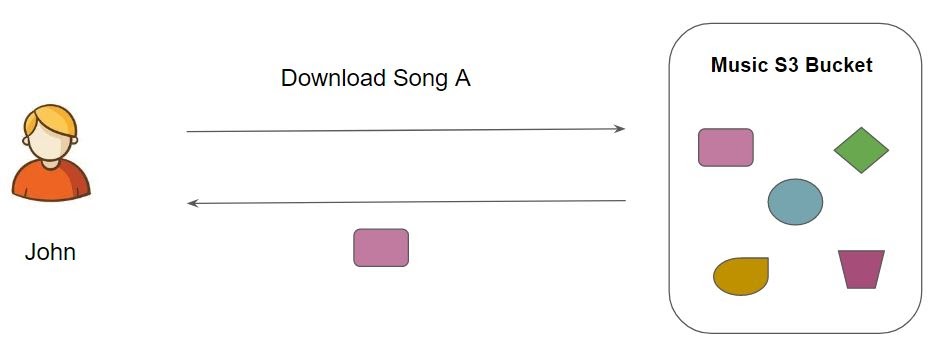
AWS S3 does not support dynamic websites that rely on server-side scripts like PHP, JSP, ASP.NET, and many more.

**S3 - Presigned URLs**

Use-Case of Music Company

Company ABC  is an online Music selling company. Once the user purchases a song, he should be able to download the song. Your company has decided to store all of its song data in S3 due to its highly durable option.

How will you go ahead with this scenario?



Understanding Presigned URL

All objects in S3 are ‘Private’ by default.

However, Object owner can optionally share objects with others by creating a pre-signed URL to grant time-limited permission to download the object.

Achieving the Use Case:-

After a user purchases a song and requests to Download, the application should generate a pre-signed URL that will allow the ‘MP3’ file stored in S3 to be downloaded by the user.