-----------------------------------------------------------------------------------------------------------------------------------------

11-06-2020: day 8

https://vimeo.com/427986964

NIT$AWs

S3 : **Simple storage service** :

**S3 comes under Object based storage**.

Block based storage: EBS / Instance store: Designed to run OS.. to perform cluster mechanism and to add volume.

object based storage : S3 : Designed to store the object flat files ( flat files means any kind of files pdf, txt,avi,exe…) .. We cannot install/run the applications.

Network based storage: EFS / FSx : Storage over the network.. The resource or the storage is available over network, whatever the server we need to mount we can do it.

S3: Object based storage.

S3 platform global doesn't required region selection. Whatever the bucker we are creating in S3 platform that bucket will be available across the regions. For Eg:- we have IAM it is global , that means whatever the user we are creating that user is available , and that user can access the resources.

Whenever we are creating a bucket, it will ask us to choose the region.

In s3 we use buckets to store the data. (Bucket = Folder with unique namespace)

Bucket is nothing but a folder with unique namespace. Basically whatever the data which we are storing in the bucket, that data can be share with anyone.

**Bucket name limitations:**

--> Bucket names are unique across the globe. [suppose if we created a bucket in Mumbai and it is not available and we are selecting the north Virginia try to create still it will show not available why coz buckets are unique across the globe not the region. They doesn’t required region selection coz they are globe].

--> Min 3 char, Max 63 Char..

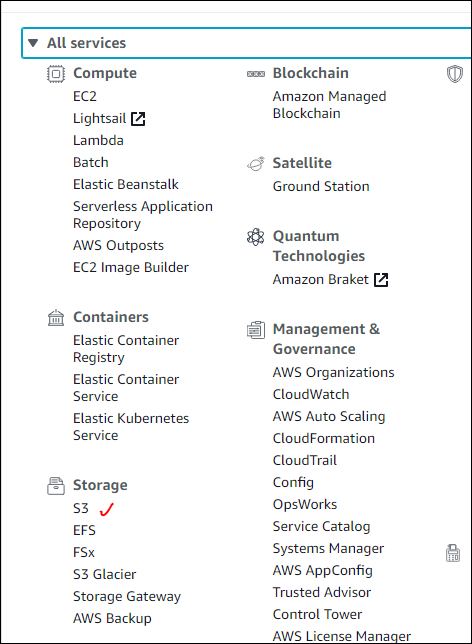
--> Should not start with (.) , end with (.), adjacent( ..) , No special characters

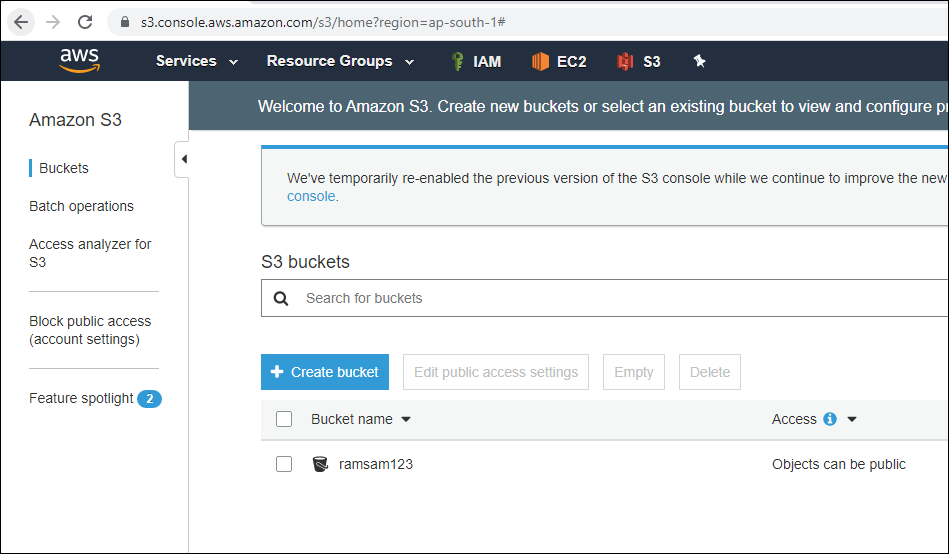
--> No Capital letters

--> Should not resemble IP address format. Eg: - (**192.168.100.1 can’t be created in this format )**

Where do we want to keep all the data? Technically all our data ( physical copy ) is stored in the particular region, it can be operated/accessed from any region server or region resources.

Now let us see how we are going the create the buckets , so naviagate to the storage 🡪S3 service and do it.

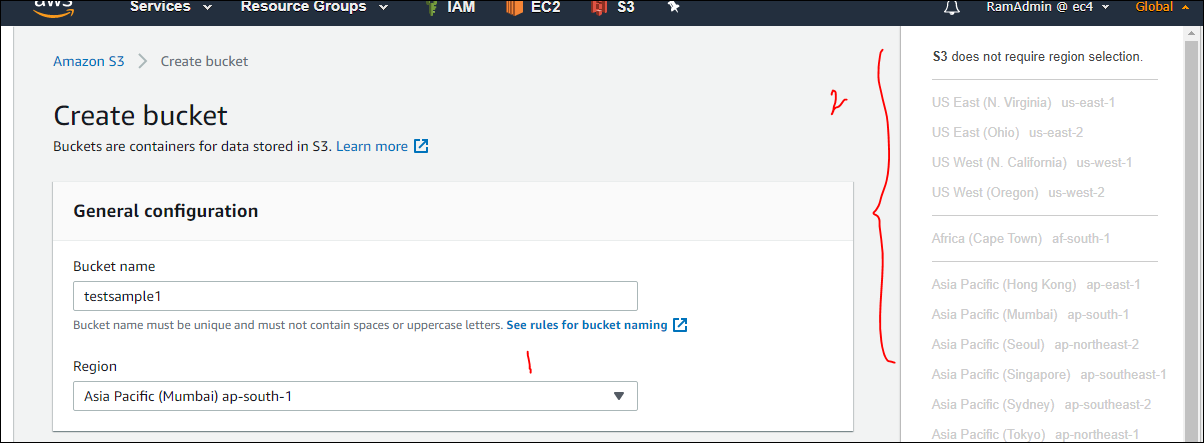




**Rules for Bucket Naming**

The following rules apply for naming S3 buckets:

* 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).
* Bucket names can't begin with xn-- (for buckets created after February 2020).
* Bucket names must be unique within a partition. A partition is a grouping of Regions. AWS currently has three partitions: aws (Standard Regions), aws-cn (China Regions), and aws-us-gov (AWS GovCloud [US] Regions).
* Buckets used with Amazon S3 Transfer Acceleration can't have dots (.) in their names. For more information about transfer acceleration, see [Amazon S3 Transfer Acceleration](https://docs.aws.amazon.com/AmazonS3/latest/dev/transfer-acceleration.html).

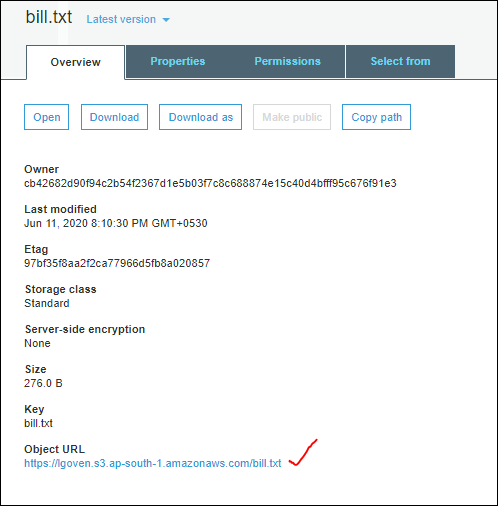


Here in the above image, we can have physical region [1], where exactly the create bucket is stored but it can be accessed globe.

Standard URL : it work for any of the bucket, even if they have dot , if they don’t have dot in the bucket name.

https://s3.ap-south-1.amazonaws.com/**bucketname**/**objectname**

https://s3.ap-south-1.amazonaws.com/**avinashh.website**/A.png

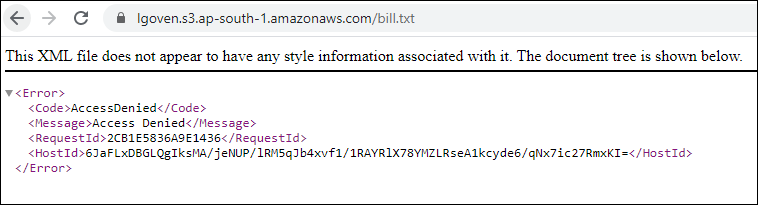


Virtual path : it works for only bucket name which they don’t have any dot in it.

[**https://bucketname.s3.regioncode.amazonaws.com/objectname**](https://bucketname.s3.regioncode.amazonaws.com/objectname) **------- syntax**

https://avinash3.s3.ap-south-1.amazonaws.com/RDS+task.txt

“<https://lgoven.s3.ap-south-1.amazonaws.com/bill.txt> “ [1]



Even if we remove the region code also the below url link will work.

https://bucketname.s3.amazonaws.com/objectname

https://avinash3.s3.amazonaws.com/RDS+task.txt

<https://lgoven.s3.amazonaws.com/bill.txt> [2]

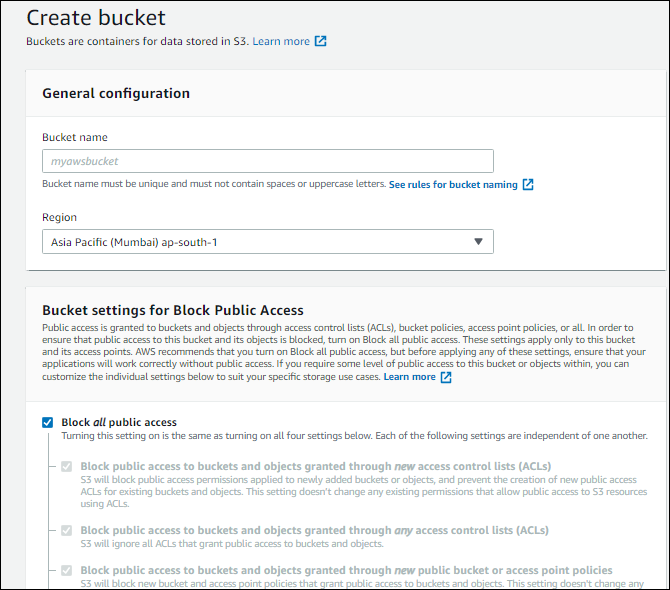
We are not able to access the both urls [1], [2],

to achieve that ==> If we want to share any data from our bucket, firstly enable "**Public access settings**" at bucket level.

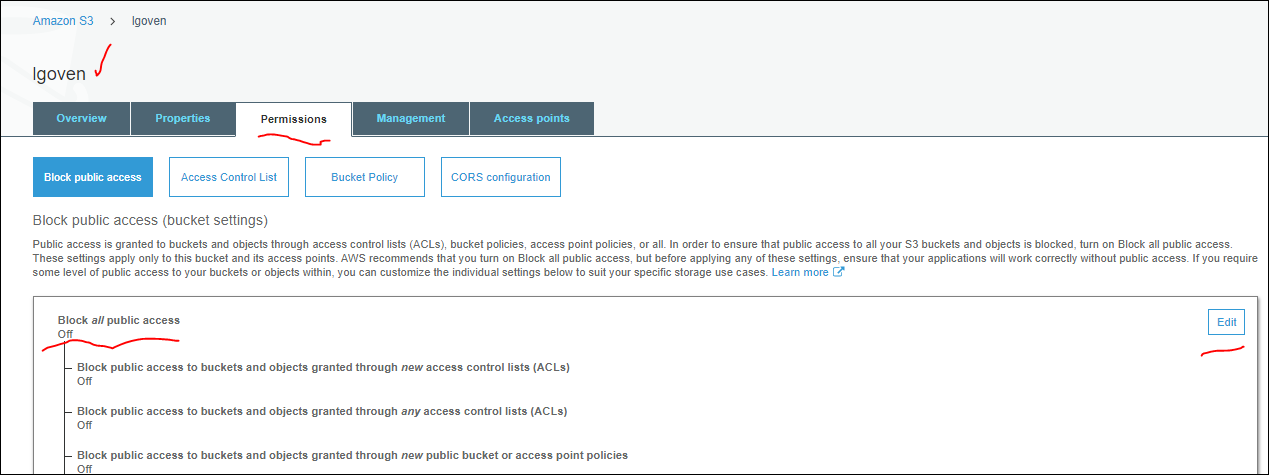
So by default while creating the bucket, we noticed that “Block public access for the bucket got enabled”, so disable that. Even this can be achieved after creation of bucket also, by navigating to the permissions and enable the public access to it in Bucket level.

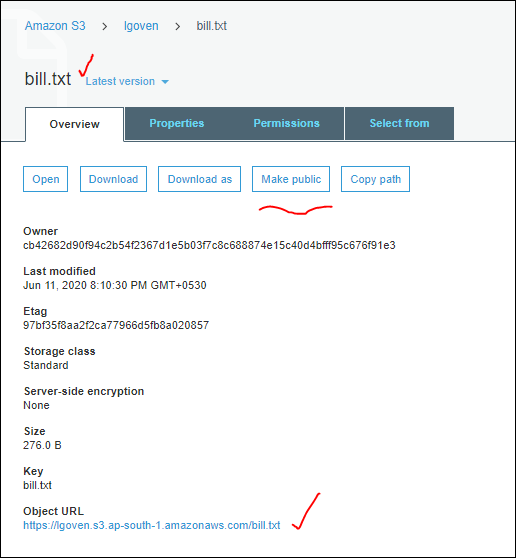
After enabling bucket level Public access, then object level we need to “makes Public” then only that file can be access by public.

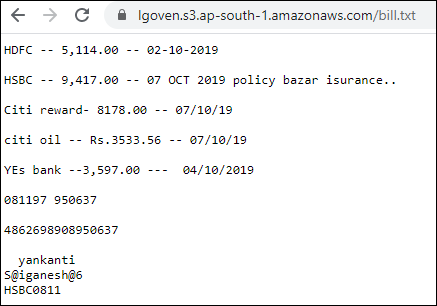
This is how it looks by default :



Once created, we can even edit the permissions for created bucket. By allowing it to public.







This is how we will make create bucket and make it public and object present in it.

* Real time scenario:

In project level where we are going to use it? For eg: we are designing a web site, and that web site contains “N” number of the files and we are keeping all the files in the same server, and that server size is increasing rapidly, in that case we are placing all our data in S3 and we are going to access the data from S3 for the web servers. And we are taking back up periodically for an application, we need to store that back up copy , so we use S3 platform to store those copies.

So we have an application, and it generates logs continuously , we need a place to store those logs, so we need CI [ continuous Integration] and CD [continuous Deployment]. And Dev team has given a .exe file, and we need to place that file and when ever to access , so that time we need S3 platform.

SO finally S3 is storage, we can store any kind of files.

Generally we don’t make any file as public, we are going to have compliance standards, if we are going to make any object publicly then immediately it triggers an alarm. So in real time we don’t use this, it means we don’t make any object public. So on top of this we have another service “**cloud front”, without making bucket public also we can share data.**

For eg: we are designing a wordpress web site, and that web site required some files from back end S3 bucket, so really we don’t need to make each and every object as public which are present the bucket,if we configure the Cloud Front service it will take care.

Yes it is difficult to make each and every file [object] as public, we will discuss in coming sessions abt the cloud front , how we can configure it, all those stuff..

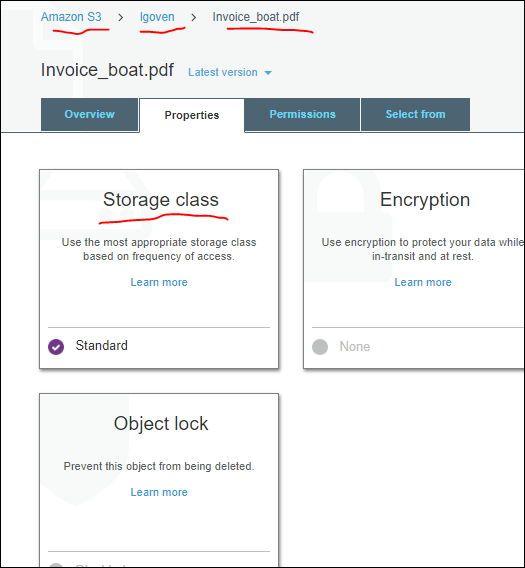
**We have session on CLI , there we can change the permissions, what ever we perform any action with help of GUI, same can be done with CLI.**

We can store any find of file, eg: .py, shortcut file, .pdf… we cant run the program or the open the file in S3, we can only store them. Even if we click on it, it will download the file to local to execute/run it.

It is also possible to make the EC2 files in S3.

What even present in the S3 bucket, they are called as object. Even we can create a folder and we can place all the files [objects] in that folder.

Now we are going to discuss about the “**Properties**”. For an object which is stored in S3 bucket.

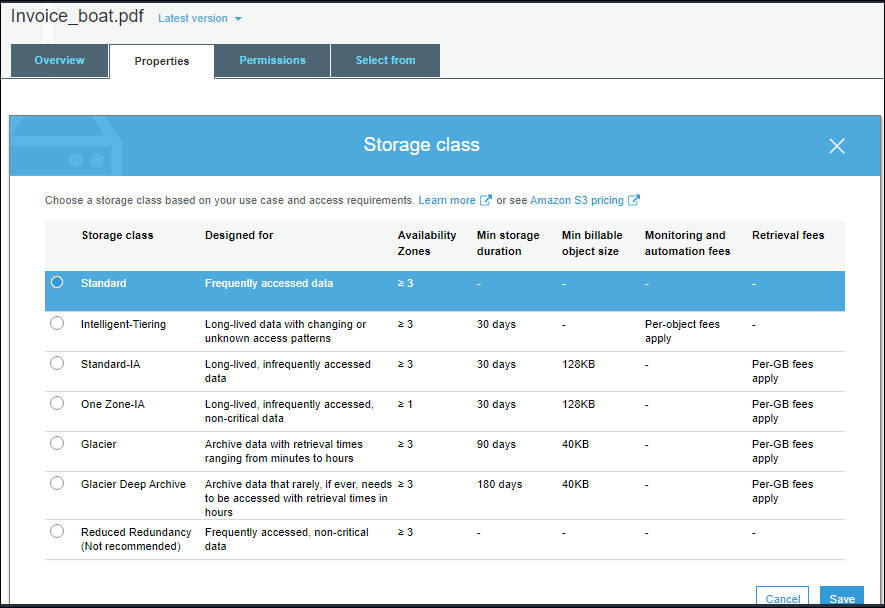


So here AmazonS3 is S3, lgoven is a bucket, invoice\_boat.pdf is an object .

If we expand the storage class property we can see the list of it.

**Storage class** is very important for certification exam also for associate, professional.

We have different S3 storage classes. As listed below



**S3 Standard** : Frequently accessed data : Data spread across >=3 AZs.. : Data available immediately

Availability : 99.99 %

Durability : 99.999999999 %

It is designed for Frequently accessed data, for eg : if we take google drive, we get 100gb data storeage, out of this 10 gb / 20 gb we used also but end of the month we get bill for 100 gb. But in S3 how we storeage for that we get bill no pre fee is applied.

For eg:

20 gb data daily using

20 monthy once we are using

10 gb data is not imp data / even if data is deleted we can easily reproduce it.

20 gb data for compliance purpose we need data , we really not accessing data , and whenever we required data we have some duration to and some flexible to restore the data .

So all these all are different storage classes as mentioned in the above example. Basically all these things are called as access patterns. How frequency we are accessing them, when we are accessing, how much data we are access is called access pattern.

So depend on this we need to choose the storage class,

First storage classe access pattern is : Frequently accessed data:

We have a bucket with in this bucket we have some data [object]which is frequently accessed.now that is **S3 Standard.**

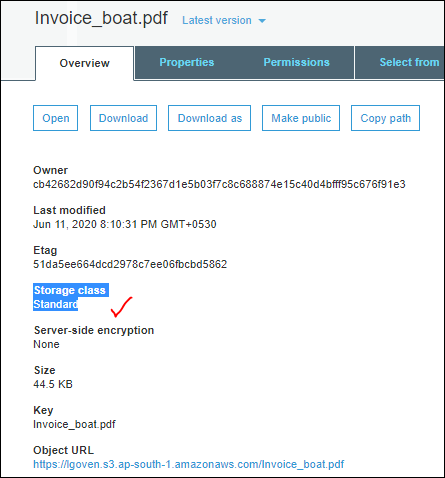
So for even storage class we have Availability: 99.99 % [entire data ]

Durability: 99.999999999 % [means do we have any changes to lose the data].

So it we select S3 standard as storage class, and we are storing 100 million files in a year, there may be one chance to lose one single file [object] out of it.that is called durability.

We may have a double for that 0.1%, we have another options we can discuss them in next session.

So here amazon is given this Availability and durability for us by placing the data across 3 AZ s,



Backend amazon is maintaining 3 copies of the file which we stored in the bucket.

**Can we able to view those copies of the file? NO, we can’t see them**

**So this is Amazon backend mechanism.**

Second one is **S3 Standard Infrequently Access**. [S3 Standard-IA].

S3 Standard-IA : Less Freq accessed data : Data spread across >=3 AZs.. : Data available immediately

Availability : 99.9 %

Durability : 99.999999999 %

It is designed for less frequently accessed data. here also data will spread across 3 Availability Zones [AZs]

Third one is **S3 One Zone Infrequently Access**

Both **Standard Infrequently Access and One Zone Infrequently Access are designed with same mechanism.**

But for one zone data will be spread only in one AZ.

One Zone-IA : Less Freq accessed data : Data spread across 1 AZs.. : Data available immediately

Availability : 99.5 %

Durability : 99.999999999 %

So for all three storage classes data is availability immediately. No need to wait to retrieve the data.

So based on our access pattern we need to choose one of this.

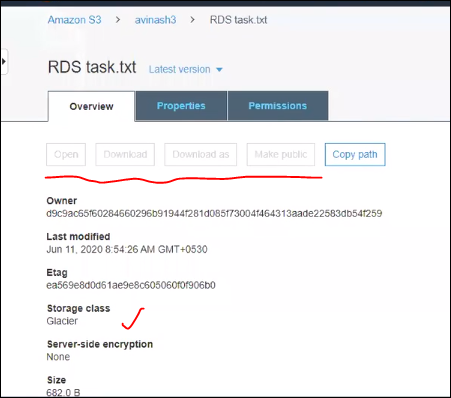
Next is **S3 Glacier: it is designed for long time archrivals :**

S3 Glacier : Long time archievals : Data spread across >=3 AZs.. : Data not available immediately.. We need to restore the view the data.. Restoration : Minutes to Hours..

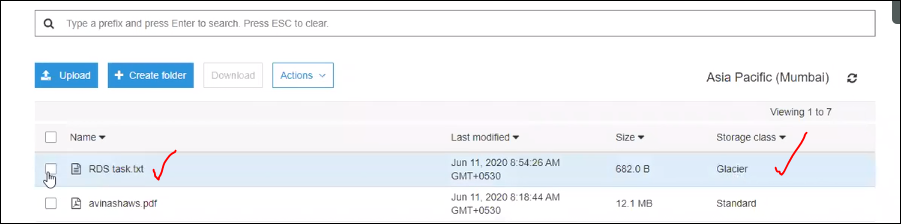
Availability : 99.99 %

Durabiluty : 99.999999999 %

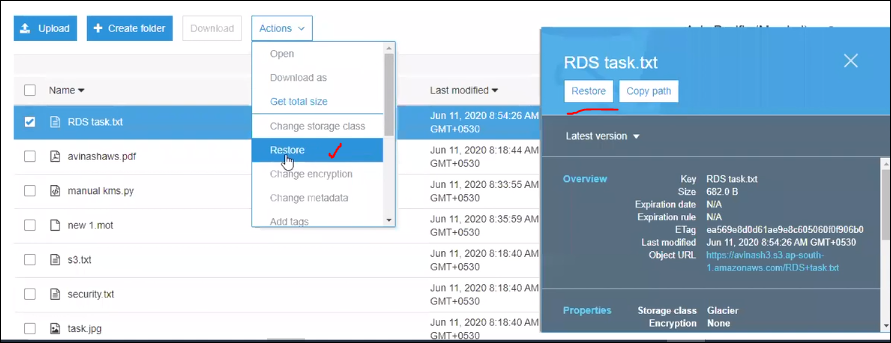
But here data not Available immediately we need to restore the data. Advantage is less cost.



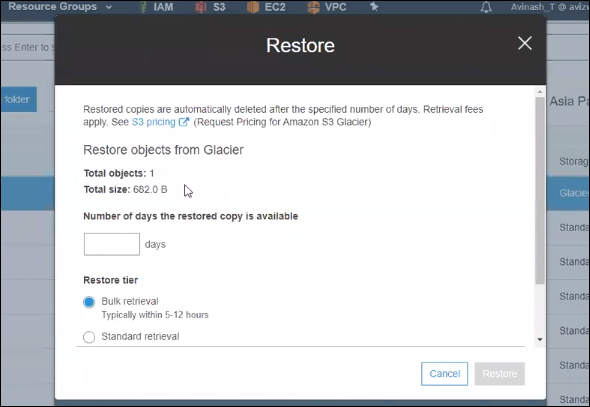
In the above image we can see that storage class is Glacier, where we can immedialtey restore the data where all the options are got disabled.if we need to view the data in the file , we need to restore it as shown below



Select the object which has storage class as Glacier.



Perform the action Restore by clicking on it



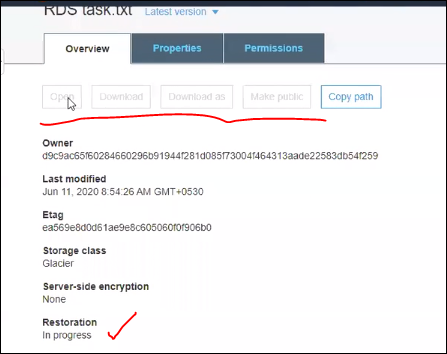
In the above image, we can see the restore tier as :

Bulk retrieval : Typically within 5-12 hours

Standard retrieval : Typically within 3 - 5 hours

Expedited retrieval : Typically within 1 - 5 minutes when retrieving less than 250MB

After selection of the Restore tier, we can see as below image that restore status is in progress. once the restore is completed, then only we can able to access the buttons. So this the resoron that data is not available immediately .



**S3 Glacier Deep Archieve**: Long time archievals : Data spread across >=3 AZs.. : Data not available immediately.. We need to restore the view the data.. Restoration : Hours..

Availability : 99.99 %

Durability : 99.999999999 %

These are the different storage classes we have in S3 platform.

**🡪Let us analyze where we can use these storage classes for real time scenario.**

Hospital application, for it we have designed an app, for that scenario for us is:

Doctors are going to have their own tab and verify all the patients records and from mrg to eveng 5-6 doctors visit the patients and verify his details and basically he is a in house patient, data is from mrg to evng whatever data we are performing on him like scan , treatment, like that.. so even thing we are capturing and storing in S3,and by using that application we are retrieving the data from end of the day.

How they are retrieving the data, by entering the patient id, they are retrieving data by using app. So what is our access pattern is that , they are frequently accessing data, we need to choose data , so for that we use standard storage class. So this first scenario.

Second scenario is :

Suppose patient got discharged from critical illness, so hospital has given an option for him that whenever he is facing any difficulty , he can visit the hospital immediate without any pre appointment, so in tha case , does he visit the hospital daliy bases ? he may visit , he may not visit, but when he visit the hospital we should not make him for sometimeto get the records and data related to him, but we are not accessing on daily bases. In this case we can choose Standards –IA or one zone –IA.

Next scenario:

We have one OP patient , he just visited the hospital and its not a critical ill, if he need to visit the hospital, he has to take prior appointment that to 24 hr before. so data is not going to visit daily basics, so here no need to have frequently access. And data should not be Availability immediately also, as he is taking appointment ,if appointment is confirmed then we can initialize the restoration, so here we are getting 24 hr time to restore the data, so we can choose “ Glacier or Glacier Deep Archive” . so here if the restoration is minutes to hours we can select Glacier storage class or if the restoration is hours bases then we can Glacier Deep Archive.

RRS : Redused redundancy storage : Not Recommended by AWS : Non Imp data , Easily reproducible data..

Availability : 99.99 %

Durability : 99.99 %

Amazon is going to roll out RRS, so in future it doesn’t takes place.

Suppose we have some data , we really don’t know which type of access pattern, we have another storage class also i.e., **Intelligent tier** : If we are unsure about the access patterns, we can choose this..

Based on the storage class also price going to vary :

Standard is regular one.

Standard –IA cost less.

One zone –IA cost very very less.

S3 glacier charge is very less for 1 TB the charge 1 $.

RRS cost more than standard, but not recommended by Amazon.

**Intelligent tier this is more costly, coz Amazon is maintaining all storage classes everything for us.**

**FREE TIER : S3-Standard Only**

**--> 5 GB Standard storage / Month**

**--> 2000 PUT (upload)**

**--> 20,000 GET (download)**

We can automatically transit the object from one storage class to another storage class, Back up option for data , achieving 0.1% for storage ,

* Which mechanism is suitable for project , who can decide this?

Solution Architect will decide this, but not only him, but for example , if we got a change to directly interact with the client once he give u the use case how he is going to design it and the access pattern as he is going to explain, and we know the storage classes, and we need to suggest him.

* Which is mostly used in all projects?

Standard is the common one used. It went with most of the organization, and we have the to analyze the access pattern based on that we need to suggest the storage class.

* Combinations can we use like for one use case one storage pattern and for other use case other storage class?

We are going to configure the transit from one storage class to another storage class, this is discussed in 12-06-20.

* Is there any performance issue between storage classes?

Performance doesn’t depend on that, but regarding to performance we are going to discuss it on common classes, that depends on the Naming formats we are giving.

Basically performance is not depends on storage classes.

Performance depends on S3 platform, S3 bucket and folder and objects.

Basically IO operations are for block storage not for object storage. For object based storage we called as performance, i.e., put & get operations, how to improve it, in coming sessions.

---------XXXXXXXXXXXXXXXXXXXXXXX--------------------