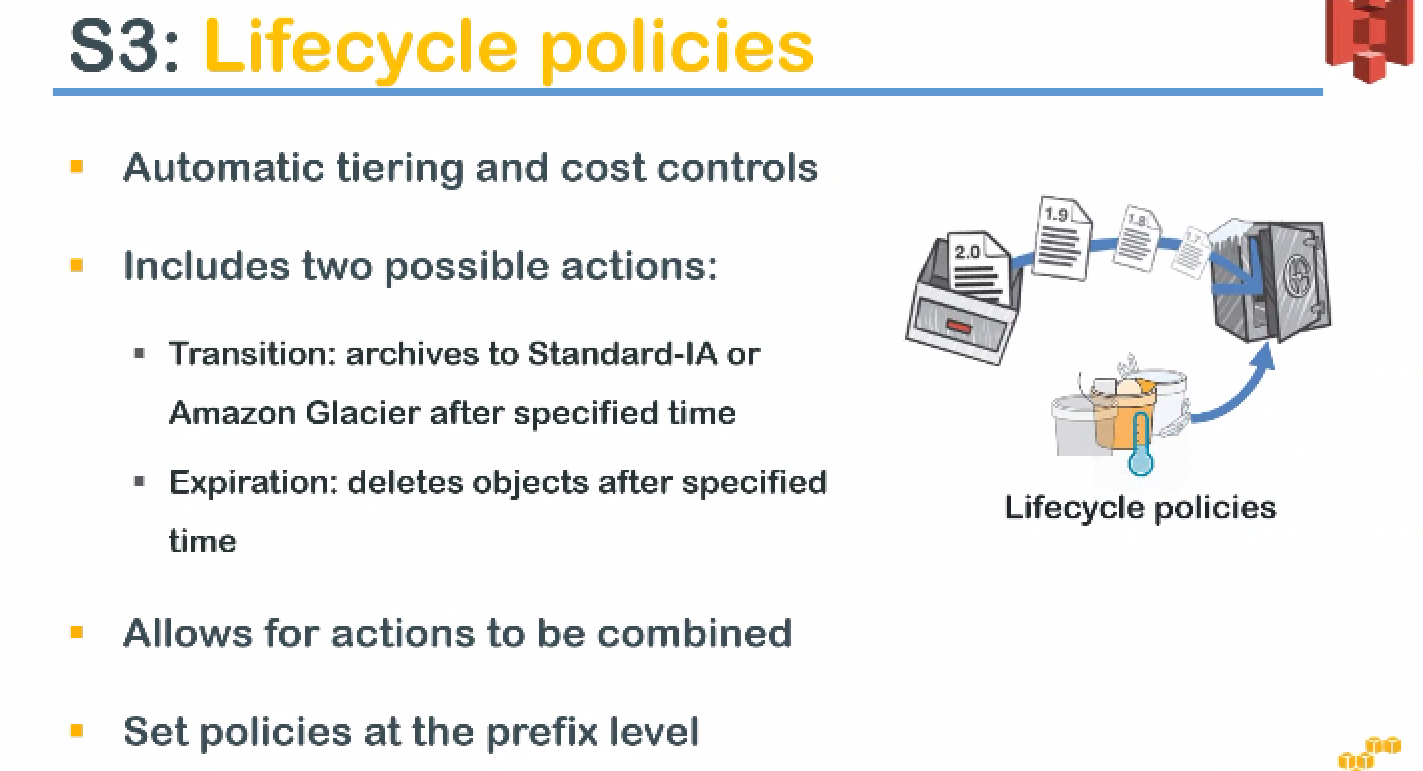
how to create lifecycle policies.

how to use a cross region replication.

**Versioning**:

keeping a track of changes to your files so that you can roll back to any older version or if for a compliance revision., somebody says that it give me a file which was there and this date… Then you can go back and give you them the particular outdated file.

**lifecycle policies**:



* you can have automated clearing and cost control.
* **For example:** in medical industry and insurance industry (very common).., after a certain period of time let us say you have a registered for a policy or visited a doctor.

Doctor visit, during that medical care time, your documents will be requested for access to very frequently but once that activity is completed.. it will be stored only for compliance reasons for a very long time... Say for example seven years.

* During that time, you are not going to access the data frequently. In that case, you can move them to infrequent access storage or a glacier (archival storage) to control your cost and you can do this automatically in Amazon.

So the way that we do that is by doing the transitioning…

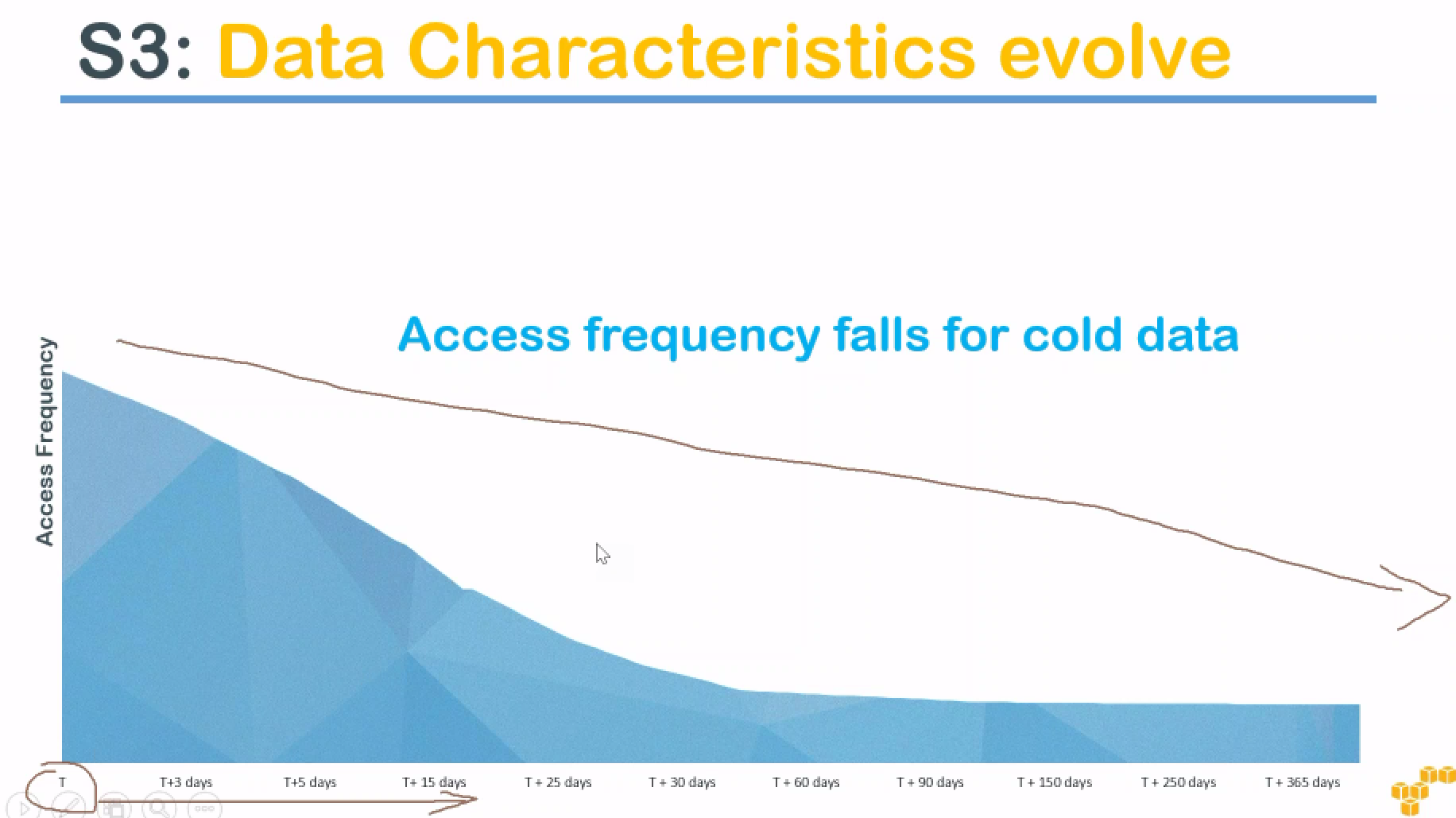
* You can move infrequent access or a glace here depending upon your retrieval mechanisms and also after say seven years or ten years, you don't want that document even for compliance reasons …

let us say it is a telephone bill you are submitting to your company and after the reimbursable process is completed and the tax compliance is completed those stuff copies of the bills are not required Enterprise anymore.

So, they have to be deleted after certain period of time, so you can go ahead and say after 365 days or 370 days delete all the copies of the business.

* We can go change this policy is anytime and they take into effect almost immediately… That is immediately in the sense **the policy minimum duration is like one day .**

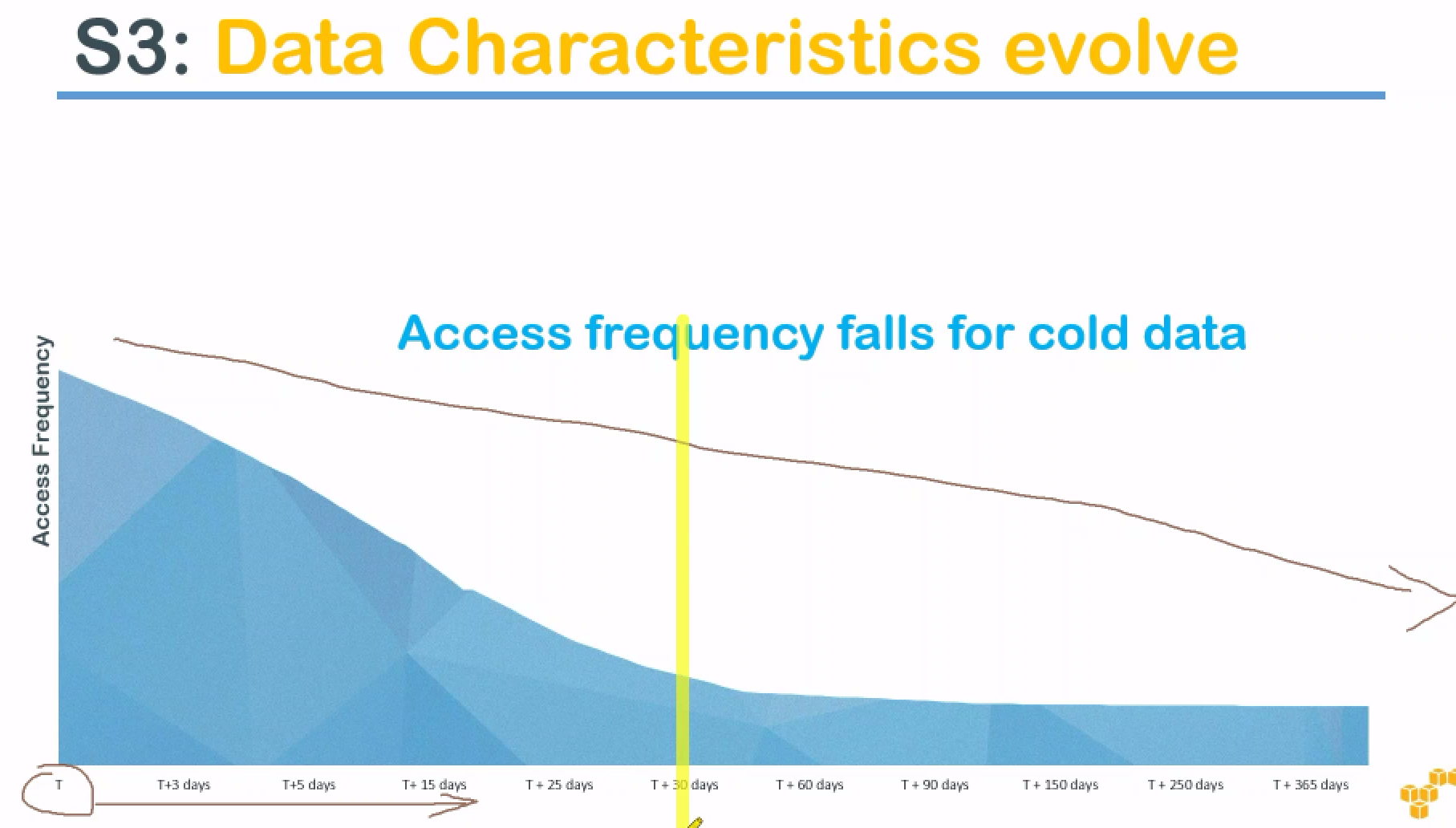
So if you change it within the same day it is going to effect in the midnight cycle..



* We are going to see that the cold data.. The medical example that I gave you guys is mostly cold data.
* So after the let us say that x axis is the time period ..

T is the time the data was created and the time increases T+3, T+5.. Access frequency is constantly decreasing and it flattens out after some time.

So this is a typical use case is that the industry is seeing... It can apply to most industries enterprises..

* After the initial period of time the data is not frequently accessed. So in those cases you will quite easily move the data into cold storage and reduce your cost and also if possible completely delete them to reduce your job... cost even more.
* So typically if you can see here somewhere along the line if I can put it line here ….after 30 days almost flattens out to very regular usage and you are not using data frequently ….probably once per a day or twice per day for checking backups or checking the versions or things like that ..

So this use case typically applies to almost all industries …And recommend your clients to analysis of the data and see how much frequently they are using it and then recommend them different storage options.

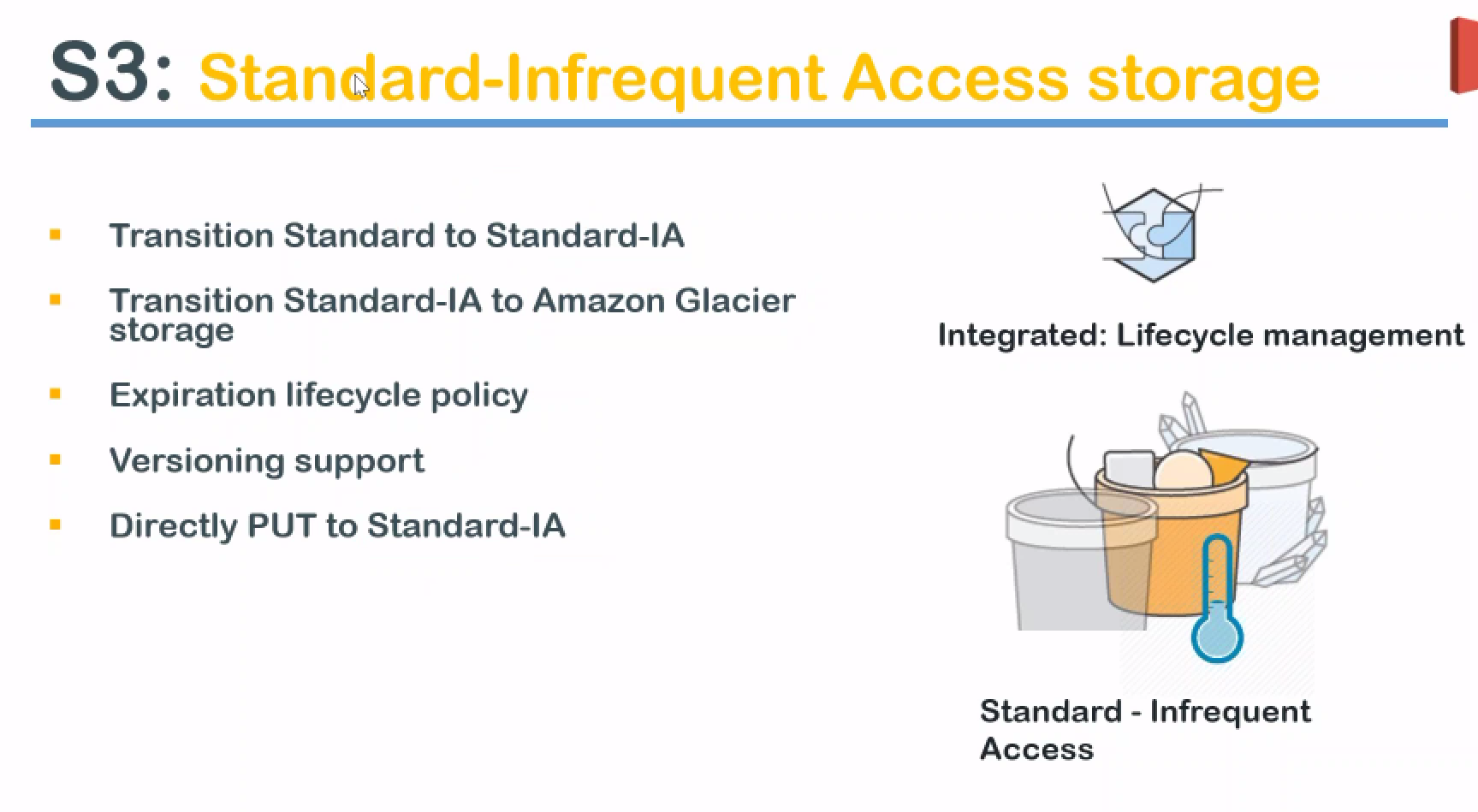
standard infrequent access is one option... the first transitioning there is you can move your data from standard storage class to a standard infrequent access.

Typically, in s3 you have three different storage classes.

1.) default standard

2.) standard ia and then

3.) glacier..



* first transition that you can do is from standard to standard ia and anytime that you want to set it and once IA is set you can move it to Glaciaer… it is not that you have to always come to IA and then move it Glacier… you can also directly move it to Glaciaer .
* But people tend to do it this way because when you're doing it in IA it is still frequently accessible… But if you remember glacier has the retrieval time involved (3to 5 hours of retrieval time) in Galcier..
* There is immediate retrieval for Glacier.. for an additional extra cost . Amazon will retrieve certain amount of data immediately and they say immediately it is there is also a small time period involved from the time you make the request on the day time if data is available to you. But it is not at the 3 to 5 hour time period..
* It allows for versioning support that means that you can move only the latest version or only the older versions... So that your latest versions will be in the standard storage class and only the older versions will be in your infrequent access or glacier class.
* Whichever version, you want to move it automatically you can select that and as I already said expiration policy can be set that means that after 370 days you can delete them…
* **The number of days I am giving you is an arbitrary number it will be always more than the transition period.**

**So, if you say transition for one day you can start deleting it from the second day….**

If you said transition for 30 days you can delete it on the 31st day or expire on 31st today.

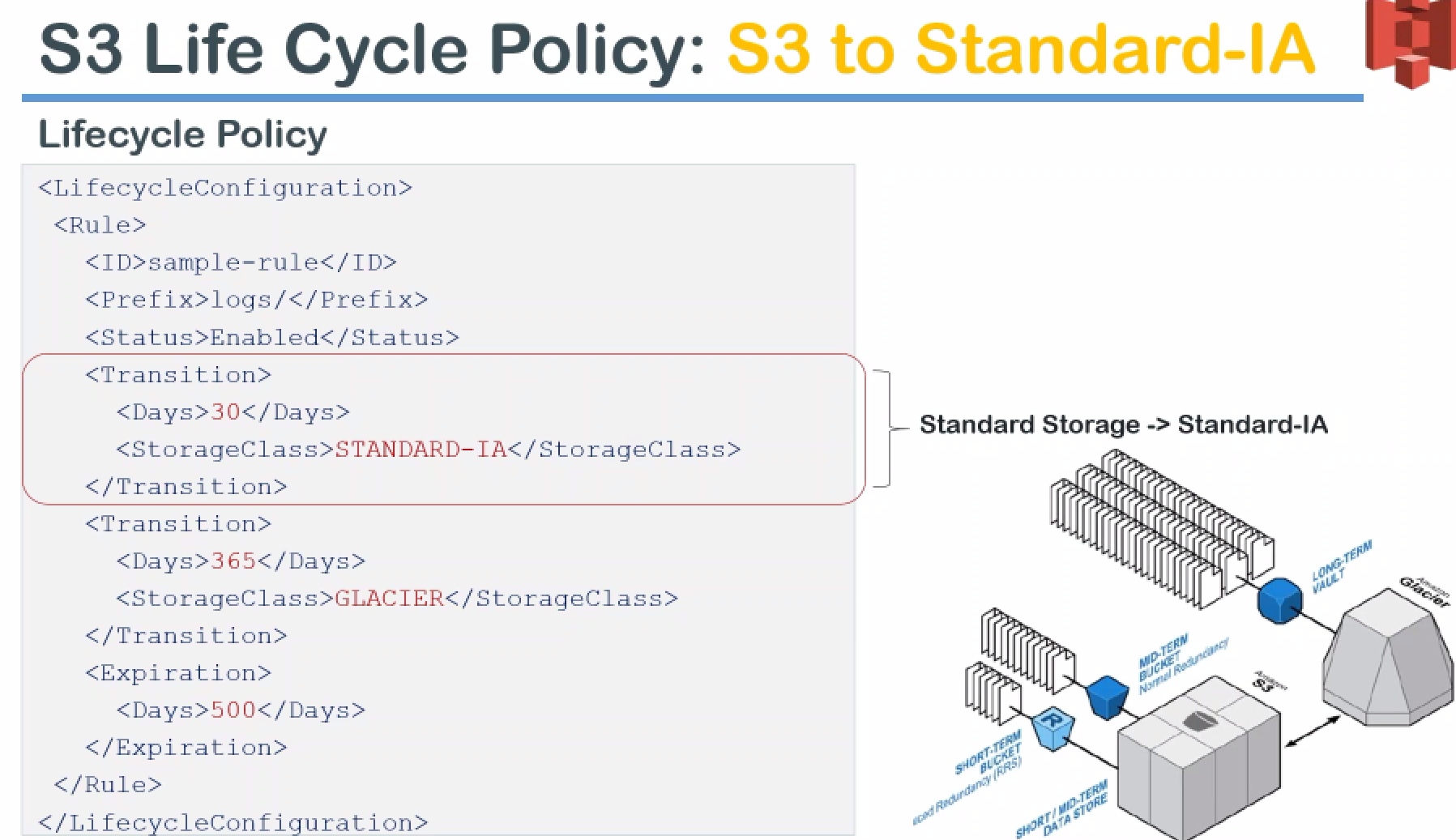
* Also, it is possible to upload directly to standard infrequent Access..

Yesterday we saw that when I have uploaded some object to infrequent or reduce redundancy storage so that is also possible to directly upload your objects to infrequent access ..

* It's not necessary to put it into standard and then you put it into IA..

When you're talking about s3 because it is a huge amount of storage and the cost is cheap ..People tend to keep on adding files and over a period of time. Bucket becomes having a lot of files it is cluttered and it is very difficult to host keep them properly **so set up these policies in the beginning itself or recommend your clients to how these kind of policies so that your buckets are clean and ready efficient or to manage them.**

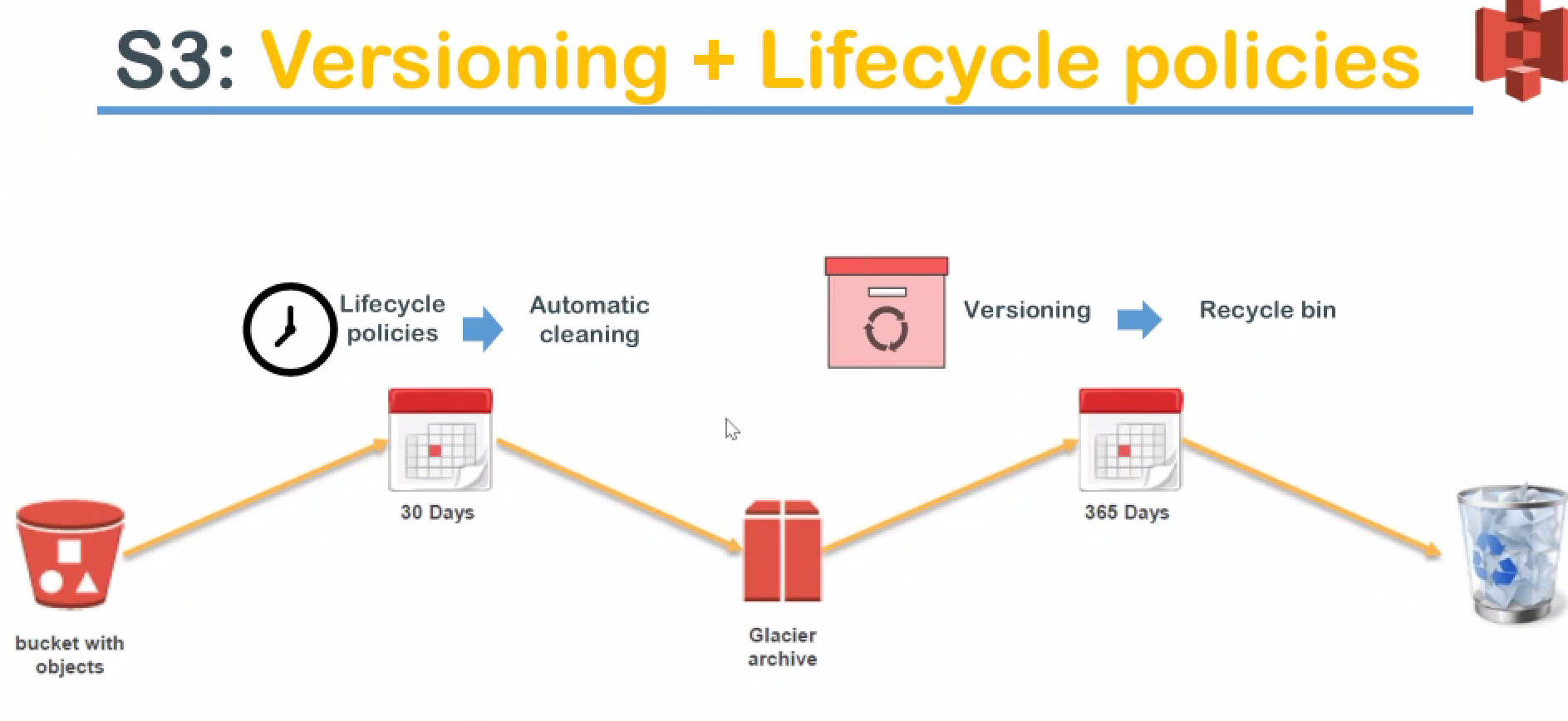
lifecycle policy if you are writing it manually or CLI or a programmatic mode… it policy would look like ..



* Highlighted part is the transition part …It will ask you the number of days, And storage type..

These are the two elements that I am talking about.

* Just below that you can also see the transition after 365 days go to Glacier…
* standard to standard IA🡪Glacier after 365 days -🡪expiration 500 days or one and a half years you automatically the data gets deleted..



**pictorial representation:**

* You have a bucket on the left hand side with lot of objects and after 30 days your life cycle policies will automatically kick in and the cleaning process takes up and it sends it to the glacier. Some data has gone to Glacier.. After 365 days automatically, the glacier life cycle policies will kick in and then they will be sent into your recycle bin or delete it from your account…
* This is another way of to understand what really happens on the background when we are setting up life cycle policies along with versioning..
* Why I mentioned versioning ????is once again you can do this lifecycle policy is only for older files or you can apply it for latest version of files OR you can have both the files affected by the lifecycle policies also …

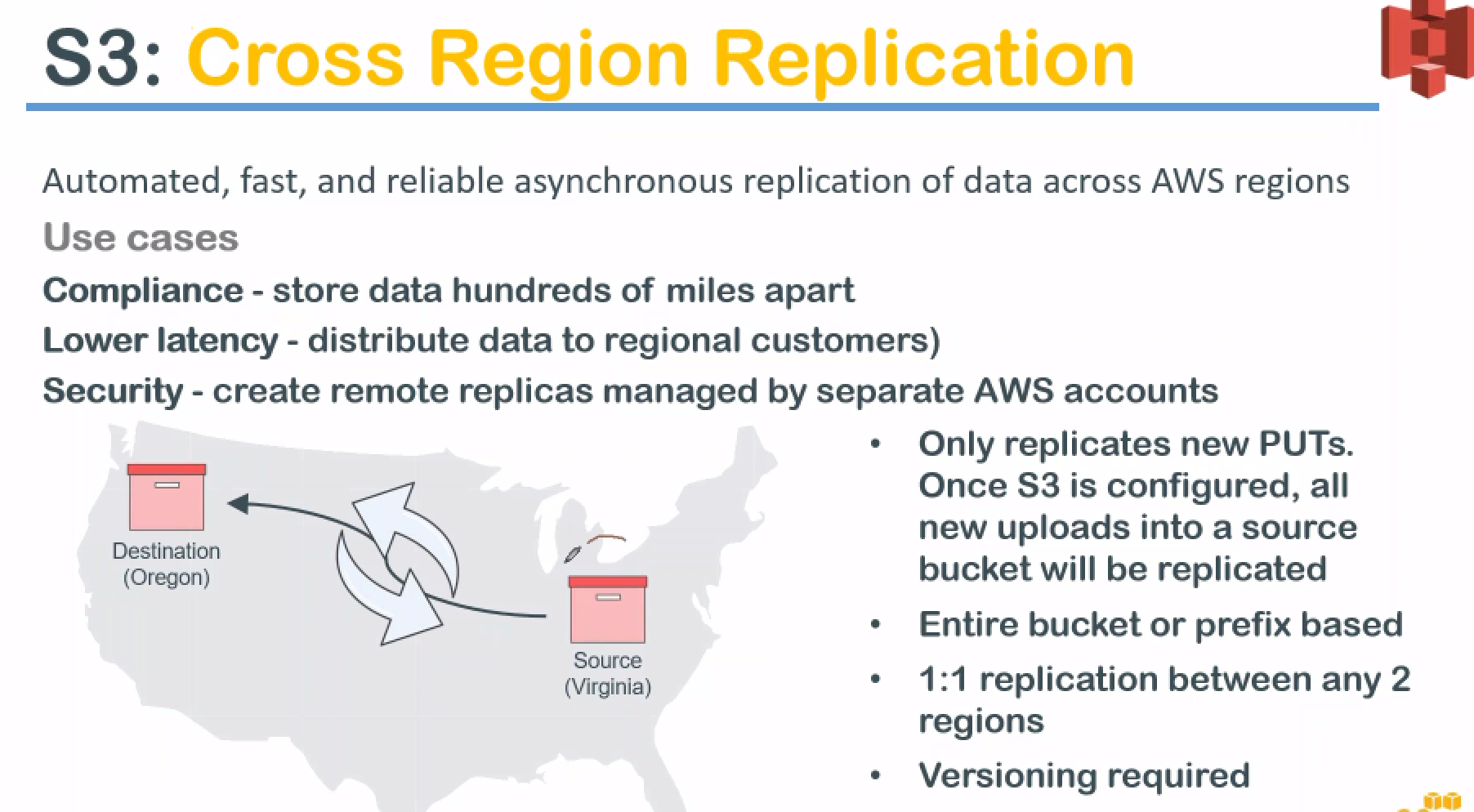
**cross region replication**

* **W**e come across the scenario very often in our professional life things like …your manager comes back comes and says that I want this data to be copied from this server to this Server...

both the servers might be different location. So, we tar or SCP or WinSCP or FTP and then download it and then uploaded again.

These are the different mechanisms that quite often happens in on-premise data center or in the cloud as well.

* Assuming with out AWS, data gets moved from one place to another place but if that is going to happen more regularly and more often Cross regions replication is one fantastic feature that you want to look at.



let us say your source is at Virginia (east coast of the United States) and you want to send all the data to your destination in Oregon region…

* What you do is you go to a s3.., create another bucket in Oregon region ..Say destination bucket.
* Enable versioning on the source and destination and
* Go to the source in the Advanced section of the properties you will have a cross region replication.

It will ask few questions:

* do you want to copy on all data and
* do you have the necessary permissions to copy the data or
* do you want to provide me or create the permissions ….so if you don't have the permissions already configured click on create it will create the necessary permissions to copy to them the source and destination and you enable it finally.

**once you enable it anything added to the source bucket after you enable will get replicated to the other region.** Usually there will be a small delay depending upon the file size it could you get replicated if it is a very small file within one or two minutes people gets replicated but if it is a bigger file it takes some time to get replicated.

**NOTE** **always what only the objects that are added after you enable cross region replication will get replicated**.

Ex: you have thousand objects and then you go ahead and create cross region replication and then expect that those objects on the destination it won't be there but now we have enabled it and of you add another ten objects you can expect those ten objects to be in your destination not the thousand ten objects .

Que: I hear people saying that I enable precision replication though it is not replicating my data which is in my bucket **it doesn't work on already existing objects it works only on the newly added objects…**

**Que:** now I have a bucket so much of data is there now how do I copy all this data there ??

Set up another source locally and then enable cross region replication and you move data from one bucket up to the bucket where you are going to get replicated… so in this case it is an internal move or fresh added to the first bucket and all the data gets duplicated to the destination also …

**another interesting featu**re:

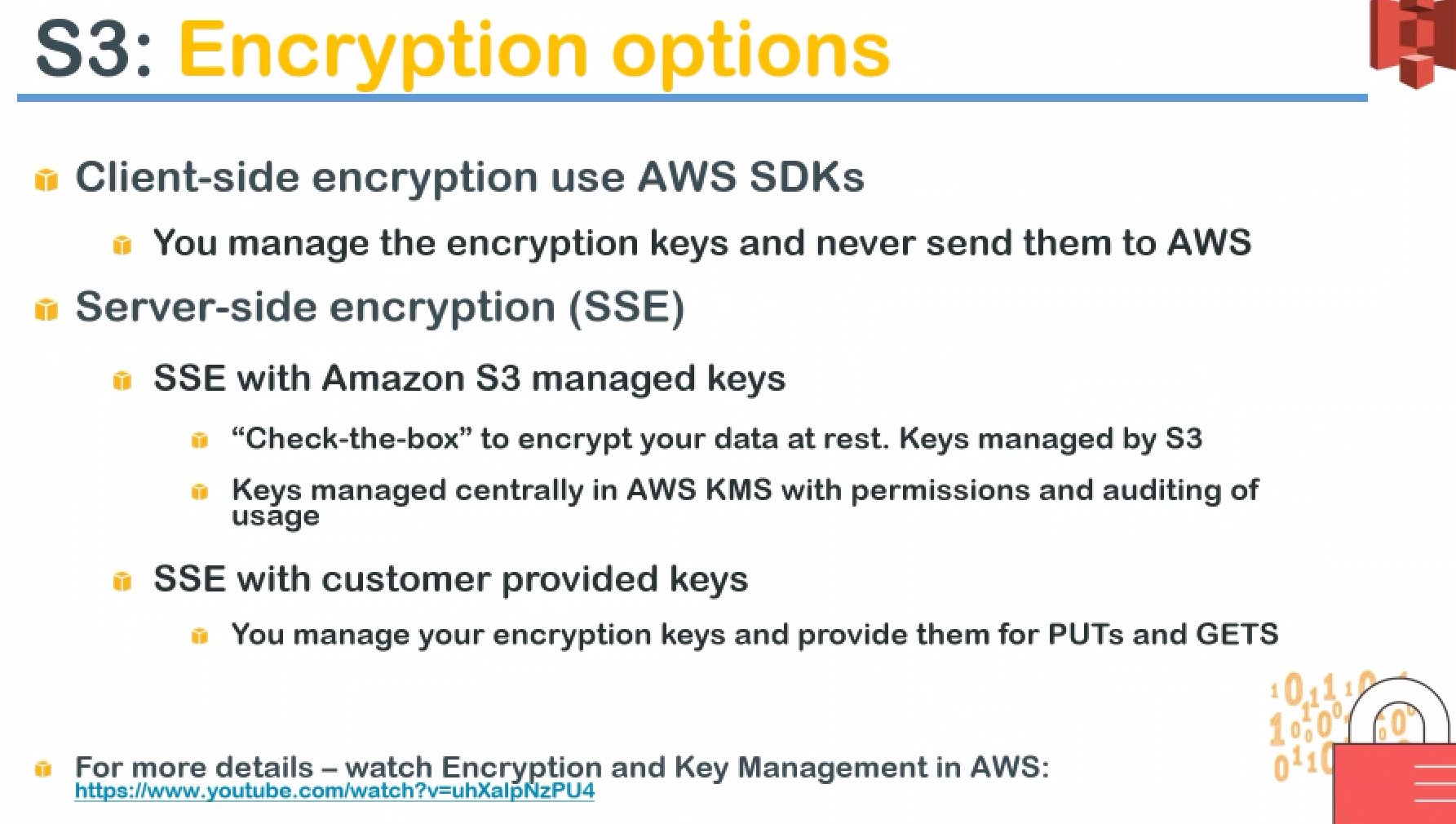
You can do it for an entire bucket or you can use it by prefix based.

Ex: let us say you want to move only the certain files which are under images folder or the images directory that also you can do or

you want to do it only for files which are starting with a certain text that also you can do

So, there are multiple filters that can be applied to your account so that you can move certain files and not all the files and in case you want all the files that is also possible so that is about cross region replication.

**encryption**



Amazon allows to do both client-side encryption as well as a server-side encryption.

when you are talking about the client-side encryption what it means is: you can encrypt with your own keys before you upload to Amazon and you can manage the keys yourself and never send them to Amazon.

when you are retrieving it or when you're storing it in Amazon it is all encrypted format. Nobody in them as one will know what is the object or what is the content of that object even if somebody has access with an object without your key they will not be able to see it they will just get an encrypted alphanumeric characters or characters.

This is what client-side encryption using you your own encryption key means…

You will need t=Amazon SDk to do that..so that you can send that encrypted document or in other words you will need a CLI or a command programming language to do this.

**server-side encryption** that is what we saw when we are doing our demos and especially with this part.

* We used Amazon managed keys to encrypt our objects and we just check the box which says encrypt the data at rest and keys are managed by s3 itself.
* if you remember we chose the option which said Amazon s3 master key.. Since it is managed with Amazon while storing or writing it into the s3 bucket it encrypts it… and Whenever you are sending an read request or copy or a get request …,Amazon will decrypt it and send it back to you.
* If you want to have encryption at a transit also then you have to look at SSL or TLS or any other form of encryption or you might have some other encryption or layer by using VPN and tunneling..
* We can manage our own keys using amazon KMS keys…

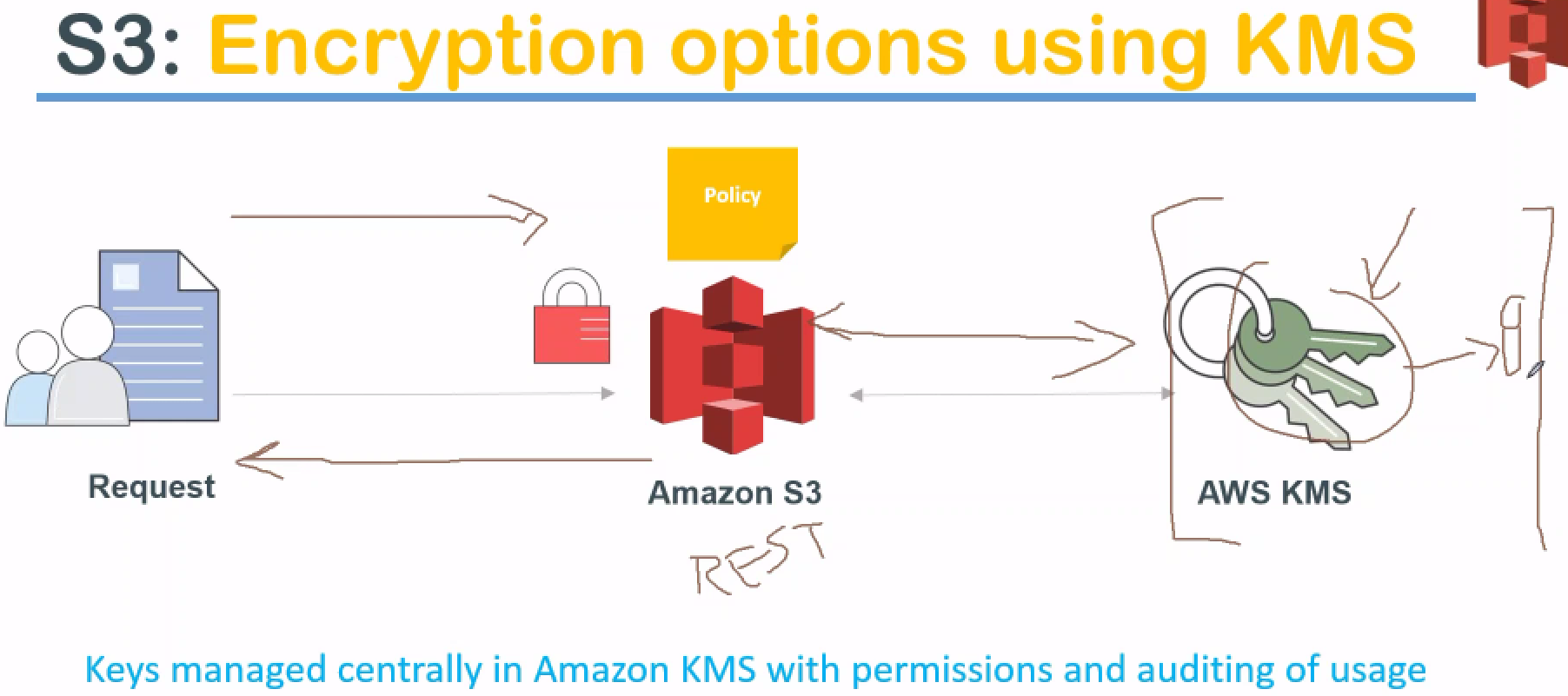
If you have your own 2048 SHA RSA keys you can upload it to Amazon and tell to use my keys to encrypt the data and decrypt the data that is also possible…

* Is there a cost of using encryption using Amazon keys???? Yes.. But a very low marginal cost involved.

It is based on the number of requests you may to the key …Free tier have some requests free..

Cost is involved by using from Amazon keys or using your own keys.

**It is not for the key storage or it is not for the encryption itself …it is just for retrieving those keys and reading or writing of the data to s3 will request a key access from the kms …**

so for that request Amazon charges . 

**how the key encryption happens…**

So if you have a user on the left hand side making the request to Amazon. Data is encrypted so automatically the kms service will be requested for the keys and once the keys comes to s3 the s3 service will automatically decrypt the data and then send it back to the user here.

**so the request goes here and then if the data is encrypted.., Amazon goes internally talks to kms and gets the data and once the data is decrypted it is sent back to the user .**

But whenever there is stored at rest the data will be encrypted using the keys.

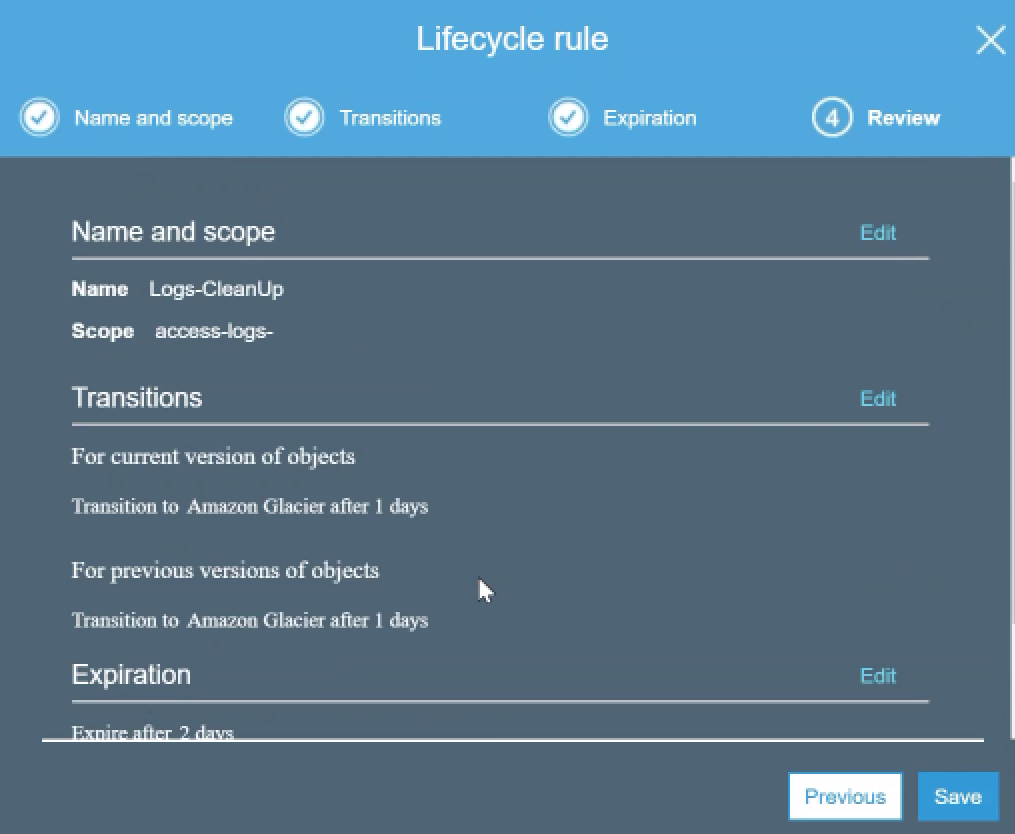
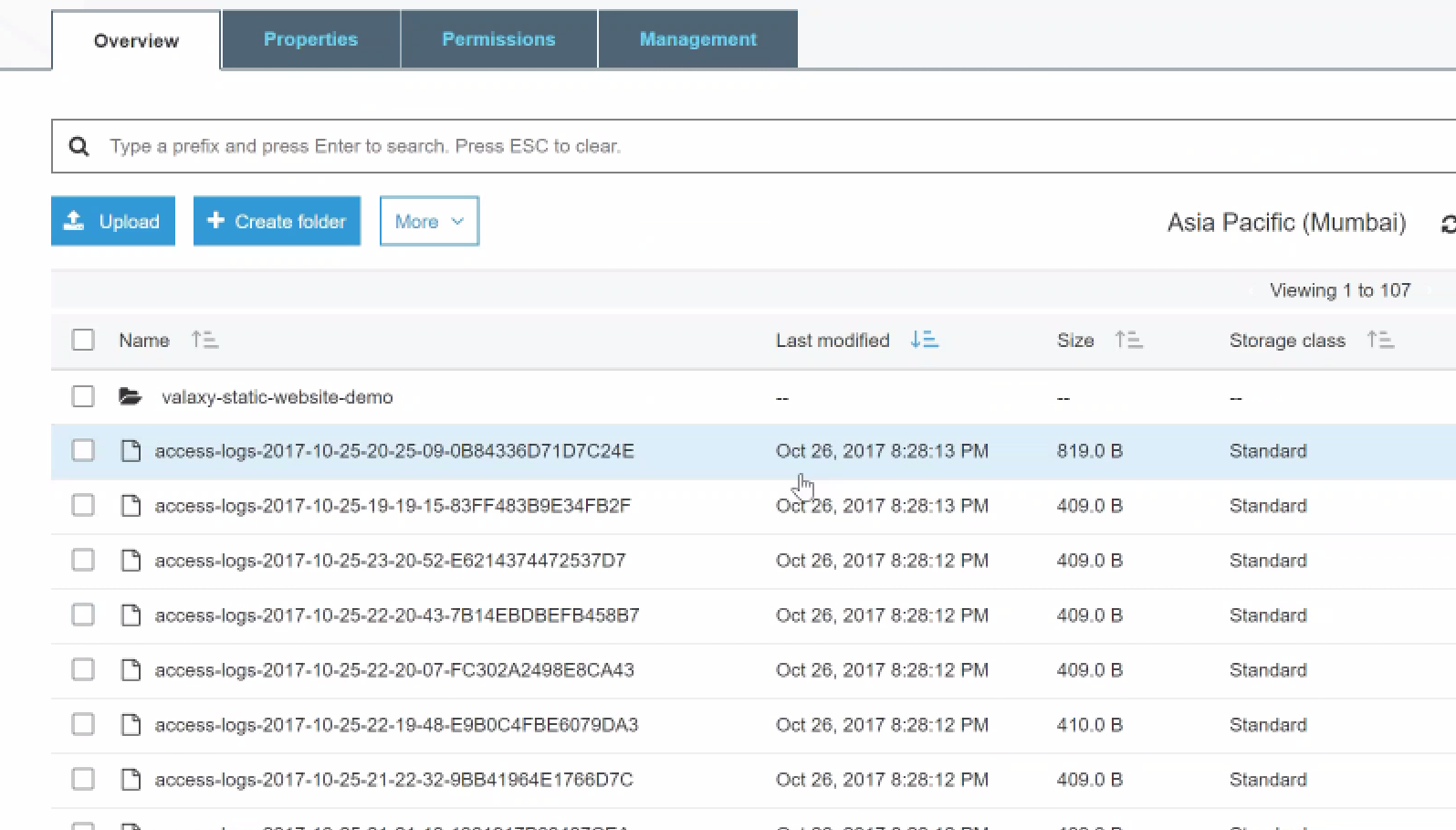
There's there will be one more key here… Amazon will usually encrypt the keys itself using one more key and store it with a master key that is on Amazon part of it that we don't have to worry about it.

but if you are worried about how this key will be stored…

**this key in itself is secured by another master key and all these keys are frequently rotated by Amazon** when I say rotated it means changed more frequently than we can imagine.

S3 service 🡪I have a particular bucket called access logs.

We are going to use this bucket for creating our lifecycle policies. It have 107 objects And let me just sort it by last modified.

Here as you can see we have files which were created on October 26 and if I scroll down all the way you can see here some of them are made on October 25th and some of them made on 24th …

All of them are in standard storage class …

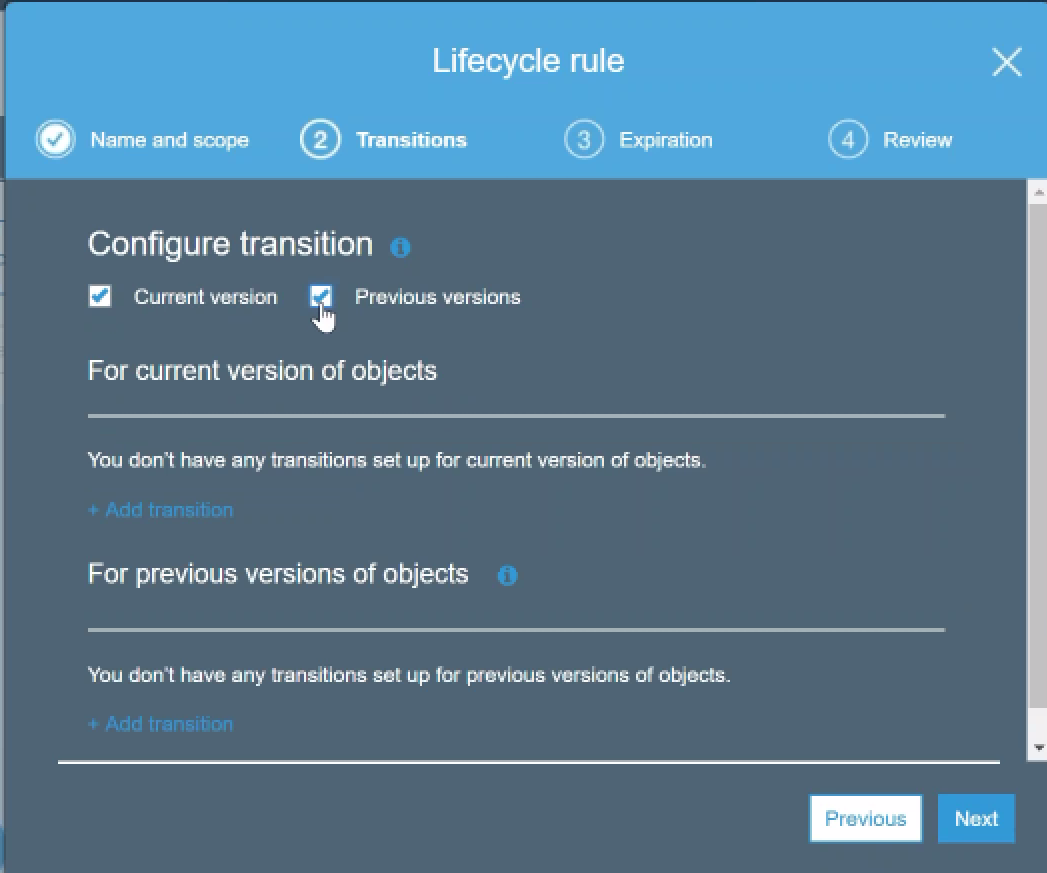
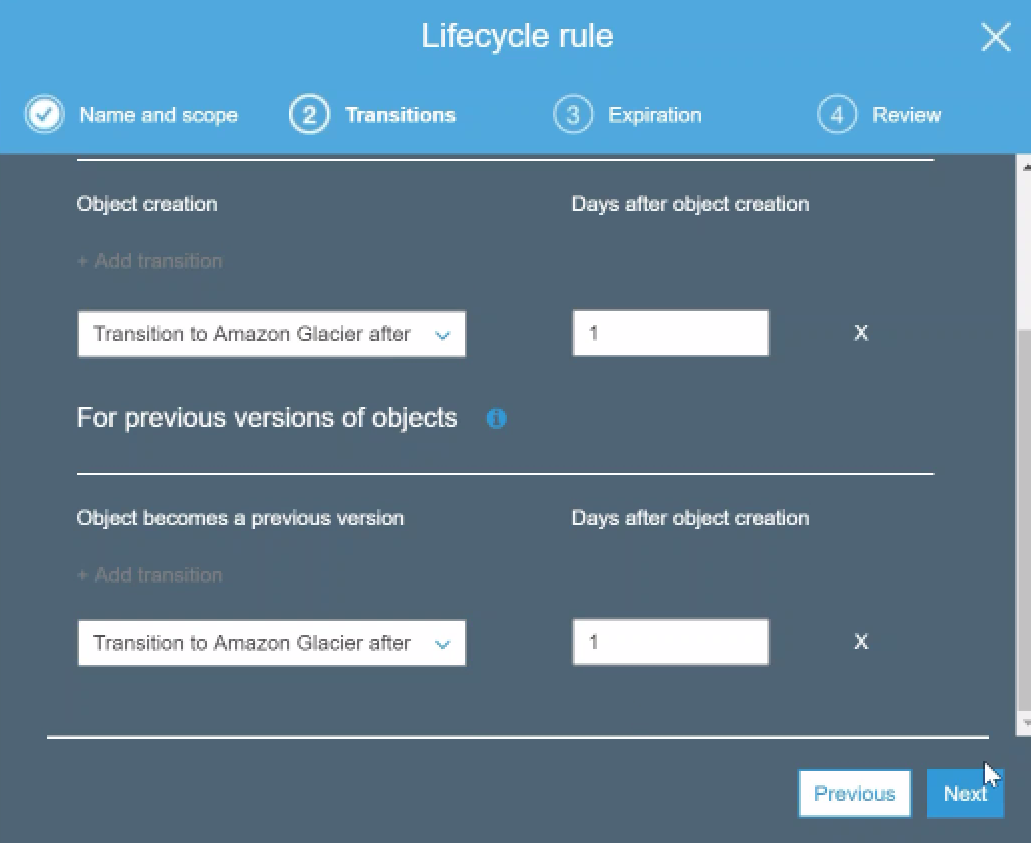
First transition S3 to Reduced Redundancy… But RRS have minimum time period of 30 days…So, we move to Glacier…

GOTO BUCKET🡪 management-🡪lifecycle tab🡪add lifecycle rule.

Name of Rule: logs-cleanup .

Prefix: access-logs

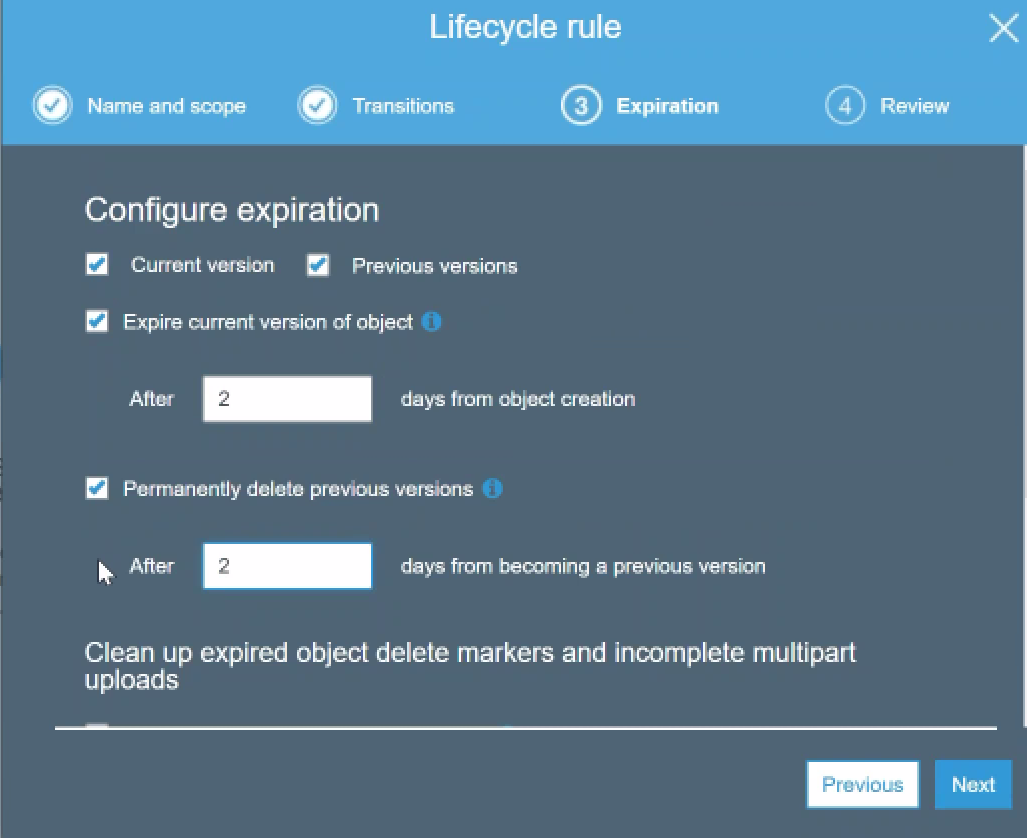
Version: Which versison,you want to take it into effect … choose both my current version as well as my previous versions

 ****

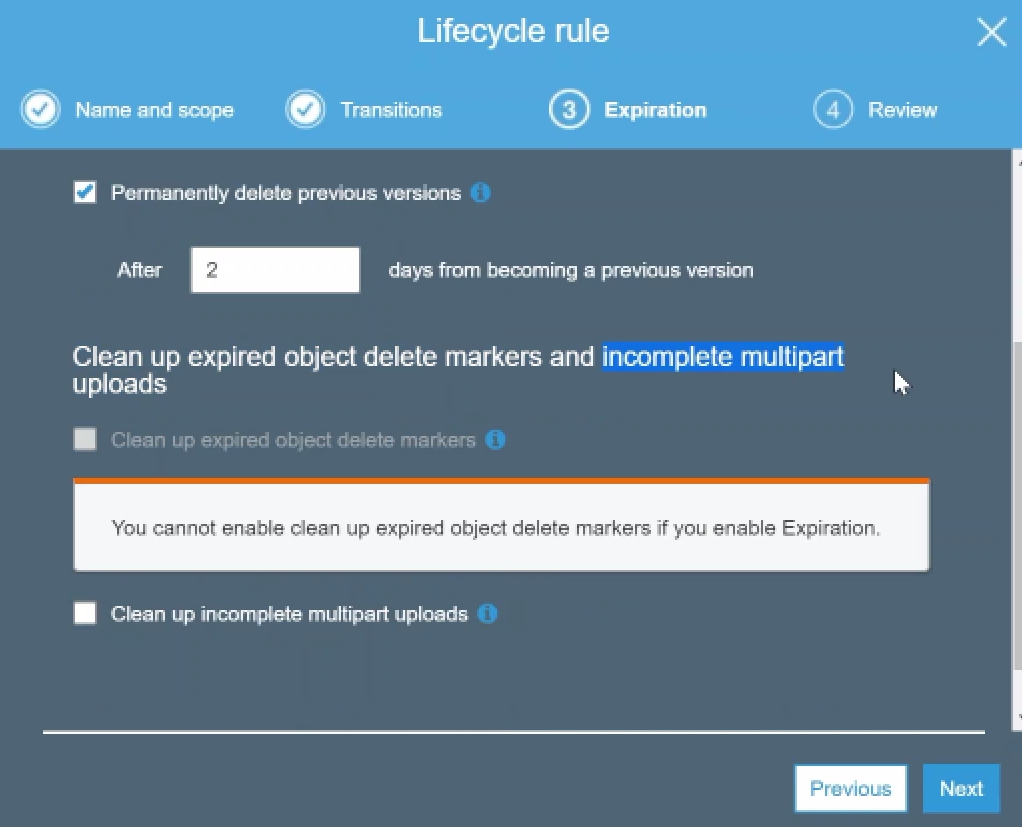
**transition rule:**

here I am going to choose glacier… The reason for that is if we choose a standard ia the minimum has to be 30 days for transitioning to IA storage class.

We will do the same thing for Amazon previous version of the objects also… We have to do for current version and as well as previous version. set it to 1 then click on next.

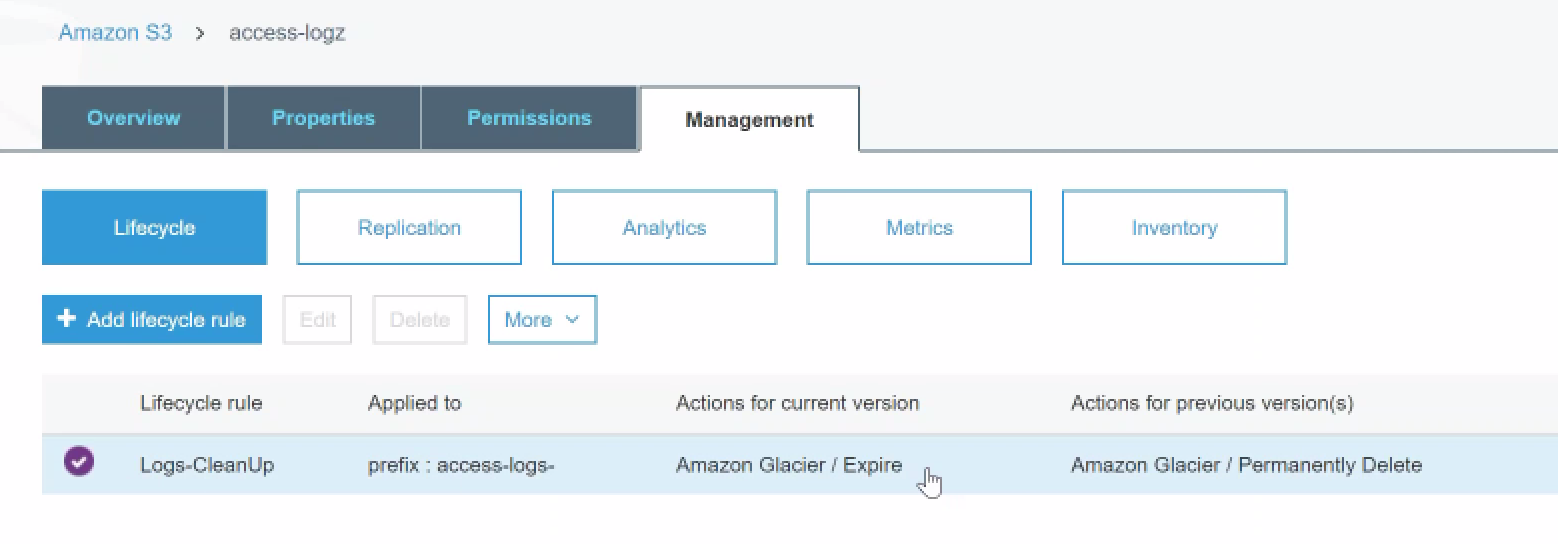
**expiration** after how many days I wanted to set expiration what I am going to do is set it for both and I am going to set the expiration for two days.

So when we come back on Monday we should see all this access - logs set up that is whichever is marked for expiration tomorrow will be deleted or by Saturday or Sunday so when you come back on Monday or things after that we should not be able to see some of the files that are marked for expiration.

**Another interesting thing:**

Clean up objects and incomplete a multi-part uploads:

What happens if the upload is disconnected and will there be a lot of objects so if I go ahead and click on this it will automatically clean up my incomplete multi-part uploads after seven days… I can change it to saying after the first day also you can clean it up.

It will give me a summary : log name, prefix, which is going to delete and it is going to transition to place here after one day and it is going to expire them after two days and permanently delete to them expiry happens after 2 days and deletion happens after 3 days.

Here, we can see rule for access\_logs…We can create other rule for Site\_logs also.

We have set up our lifecycle policies now, and min. duration is about 1 day.. So tomorrow we will come back and see what happens to our files in the those buckets and we possibly see them for expiration also on Monday or Tuesday or somewhere in the future.

**Cross region replication**

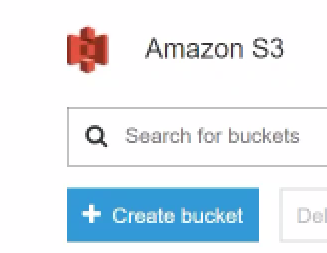
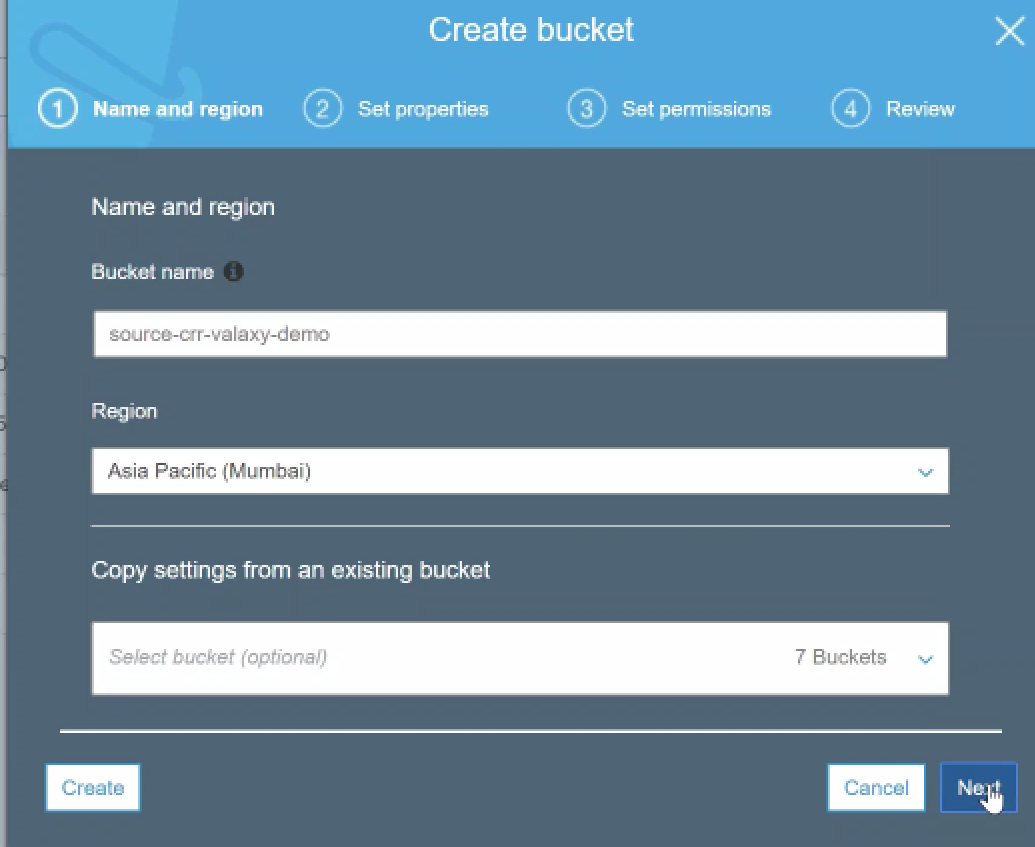
**Source bucket Destination bucket Persmissions**

You can use any of the existing buckets for a source and destination but to make it a simple and clean…

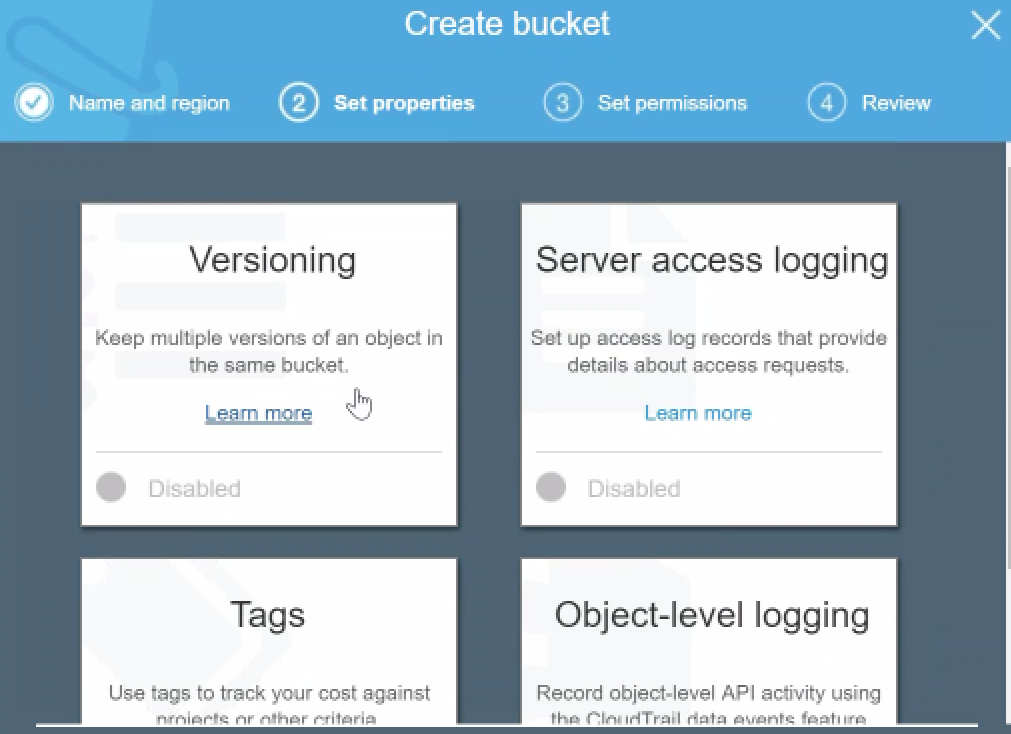
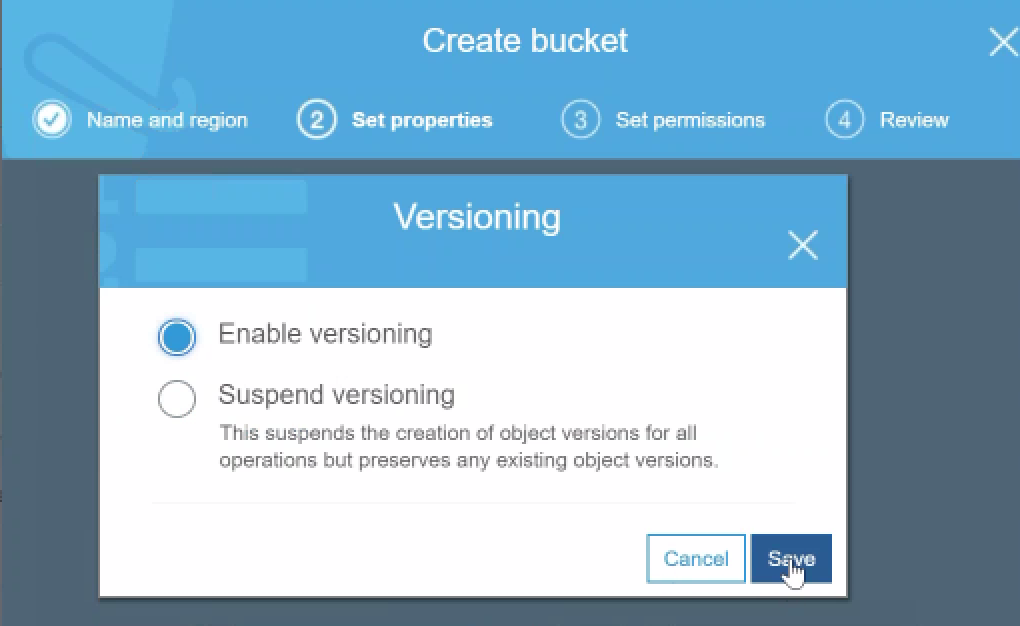
what we will do is we will set up for two different buckets for source and destination

1. let us create the source bucket.

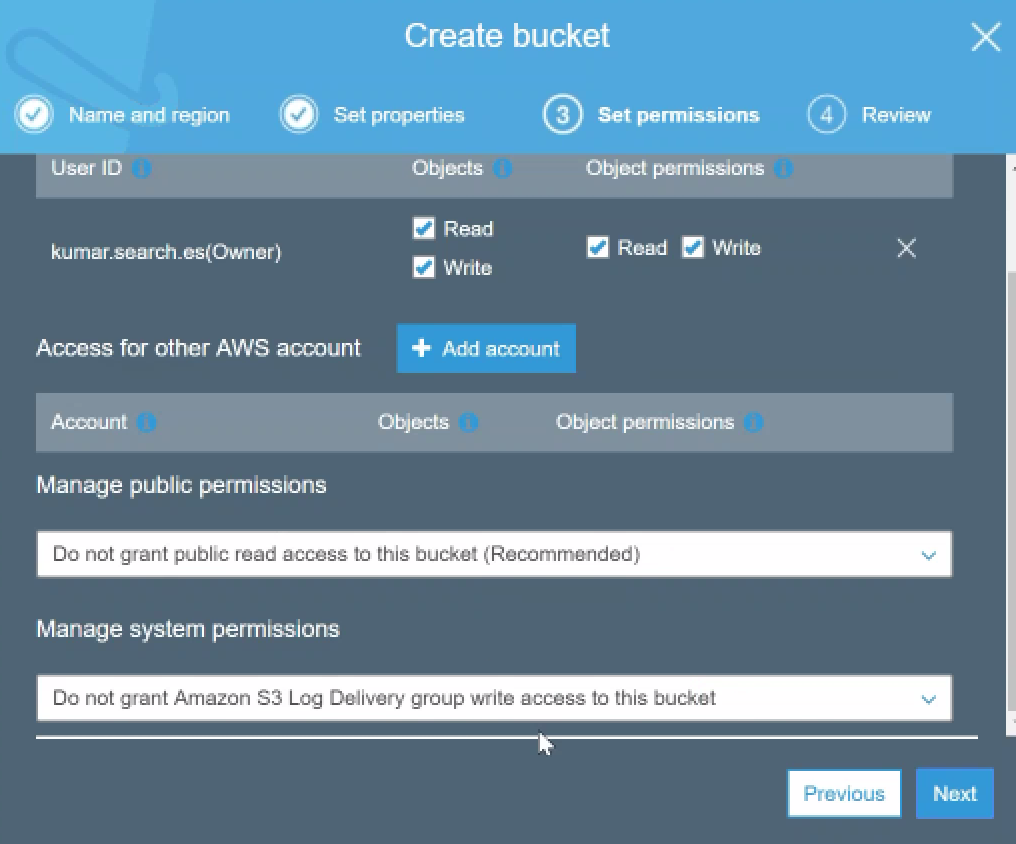
Bucket name, region(Next)

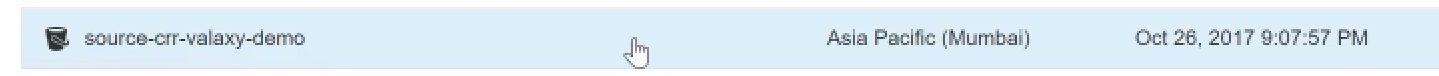
 

1. **NOTE** : for across region replication to work you need to have versioning enabled.

1. You don't need to change any of those things here… just click on create the bucket.





In Mumbai, it got created…

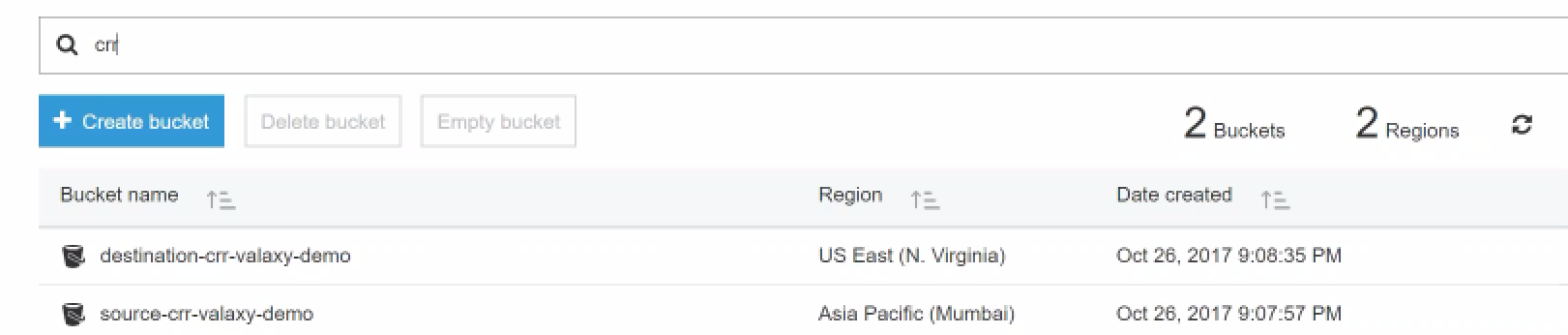
1. let us create our distinction bucket.

Bucketname: destination-CRR

Region: Virginia

We are going to move data from India to the US automatically now…

1. enable versioning

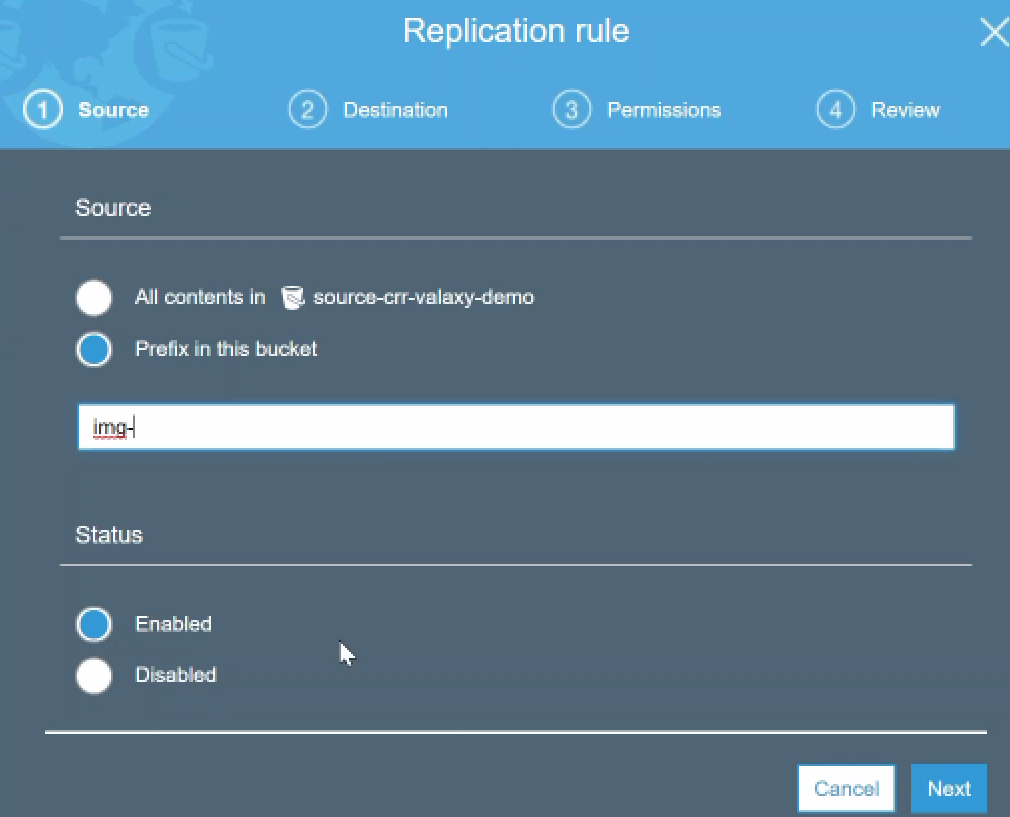
we have got two buckets available for us and let me just filter it, based on CRR.

we have a source bucket and we also have a destination bucket …

1.)source bucket-🡪 management 🡪 replication-🡪add a rule.

It has options:

* So, I want all contents in this object or
* a certain prefix -🡪 Select prefix -🡪images IMG.
* Status—Enabled

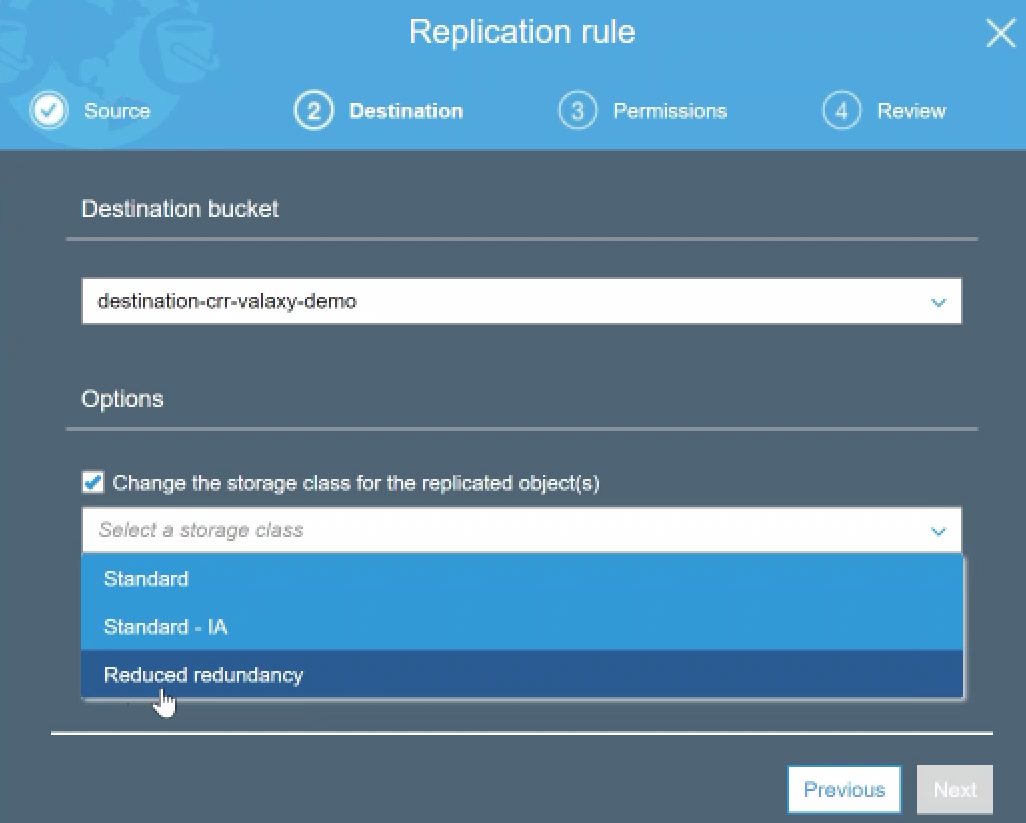


1. Next…. it will ask me which is my

* destination bucket…>destination CRR.
* do you want to change the storage class for the replicated object??

Let’s say we want to use it as a backup mechanism. Sometimes people use this for a backup and recovery mechanism.

They take a snapshot or a database backup and put it into one Region’s bucket and then use CRR to copy it across the globe and move it to another place across the globe.



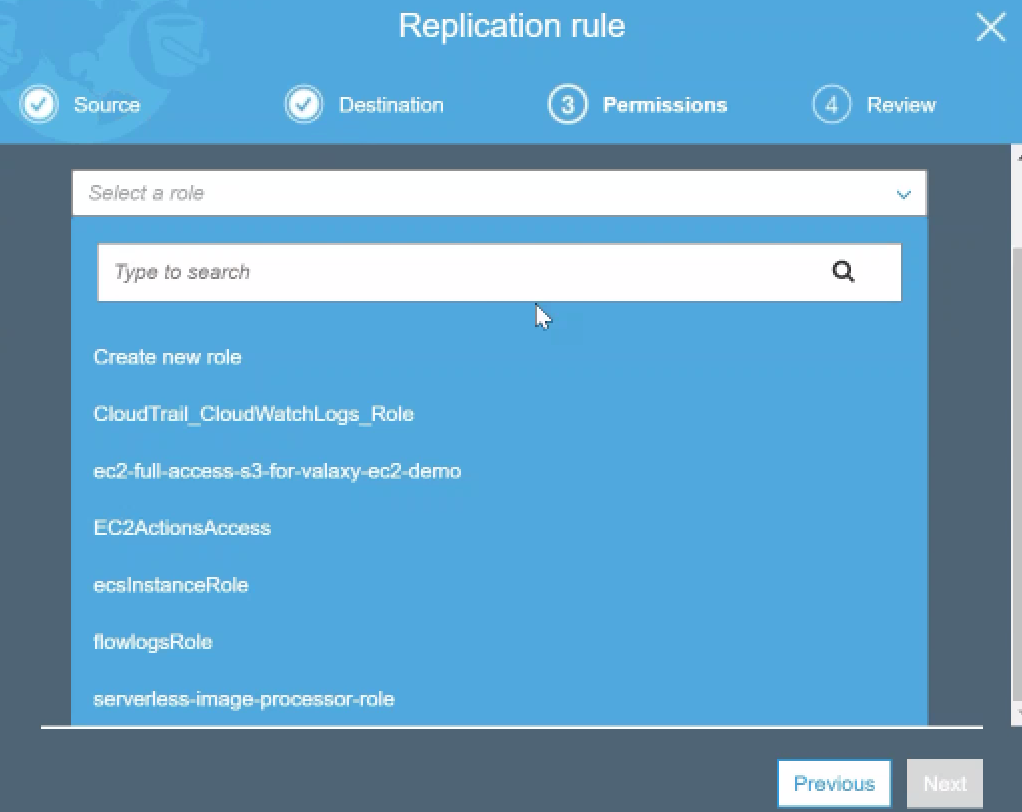
**Example:** your Prod. Databases is in Virginia, you can move the database snapshot to Singapore for disaster recovery purpose and

Since it is a backup you are not going to use it regularly you can put it into a different storage class for example Reduced Redundancy or standard IA…

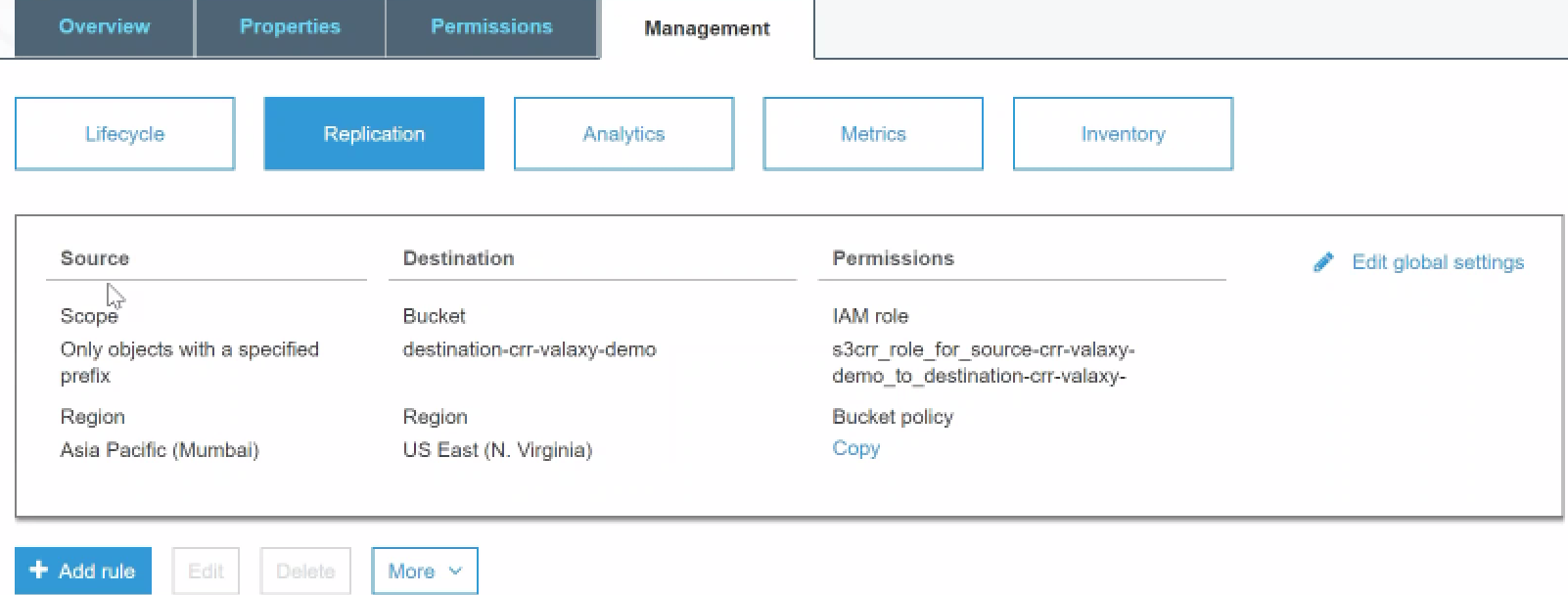
( Lets select reduce redundancy..)

1. **permissions** :  
   this is my new bucket and I'm not set up anything here.

Click on existing roles that is in my account but I don't want to use any of them because I don't know what it does so I'm going to do click on create a new role …

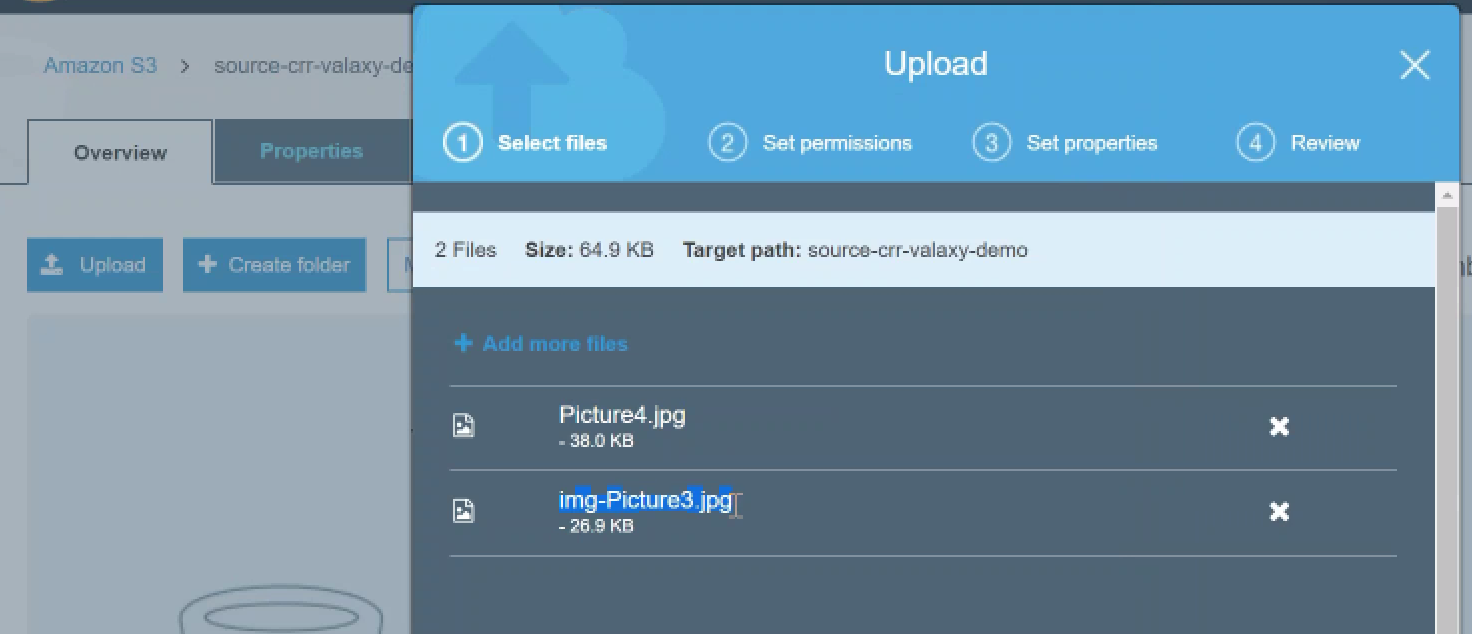
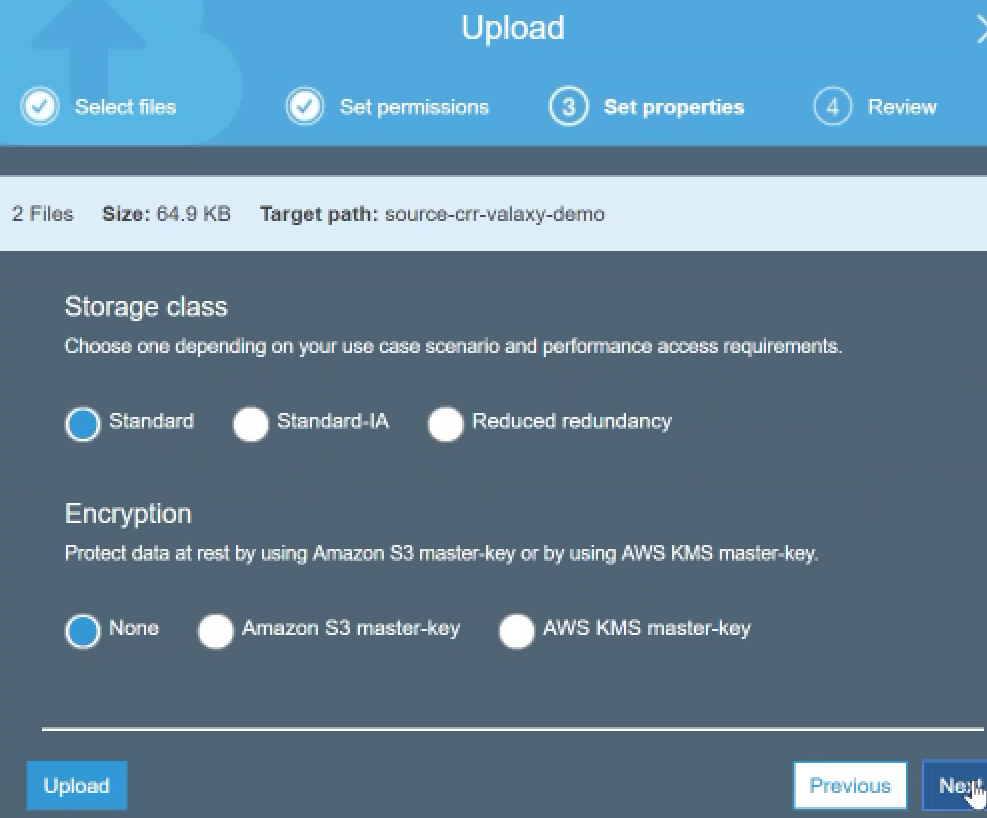


Amazon automatically will set up the necessary permissions between the source and decision buckets to copy data and put it in the distinction….

You can see here the IAM role got created… the source , the destination and the necessary permissions created 

Both source and dest buckets are empty…

I am going to upload a few objects particularly one with the prefix of IMG and another one without the prefix of IMG..

Only IMG file should replicate to destination…

Permissions-🡪No change

Choose storage class: standard here

🡪let us see what happens to my storage class in destination.

time on my screen is 914 p.m. so I'm just going to give it a couple of minutes for the replication to take into effect.

remember that this is not an instant copy or a real time copy so there are a lot of moving pieces on the background so don't try to configure this for real-time access or simultaneous the backup things like that you can use it for a backup .., but not for the real-time use cases…