**Efficient Data Archiving System in cloud**

**Abstract:**

The increasing volume and complexity of data in today's digital age necessitate efficient and scalable data archiving solutions. This paper presents a comprehensive theoretical exploration of data archiving systems, their architecture, and the underlying principles that govern their design and implementation. By analyzing the theoretical foundations and best practices, this paper aims to provide a deeper understanding of data archiving system and its significance in managing vast amounts of data for long-term preservation and accessibility.

**Data Archiving System:**

Data archiving is the practice of identifying data that is no longer active and moving it out of production systems into long-term storage systems. Secure data archiving enables the long-term retention and storage of data. It provides secure locations for storing mission-critical information for use as needed. Once in the archived data management system, the information stays accessible and the system protects its integrity. Data archiving is critical for businesses and organizations that acquire new information regularly yet must retain existing data and remain able to quickly retrieve both types. Trends in government regulations, the law, and corporate policy all skew toward more data, retained longer, and retrieved faster. Data archiving services help companies stay abreast of these trends for lower costs. Archival data is stored so that at any time it can be brought back into service.

For efficient data archiving we used Amazon Glacier for storage purpose so that the data that is no longer active can be sent to Amazon Glacier. Even there are other storage classes, glacier provides extremely low cost archive storage service(Store for as little as $0.01 per gigabyte for month), high-durability storage and so forth. Amazon glacier allow you to retrieve data with 3-5 hours maximum which is better than other methods of other storage classes.

**Amazon S3 vs Amazon Glacier**

Amazon S3 (Simple Storage Service) and Amazon Glacier are two storage services provided by Amazon Web Services (AWS) with distinct characteristics, use cases, and pricing models. Here are the key differences between the pricing of S3 and Glacier:

**- Storage Cost:**

**Amazon S3:**

S3 offers different storage classes, including Standard, Intelligent-Tiering, and One Zone-Infrequent Access (IA), among others. The Standard storage class provides high availability and durability but is typically more expensive than other storage classes. The cost for storage in S3 varies based on the storage class chosen and the AWS region.

**Amazon Glacier:**

Glacier is designed for long-term data archival and cold storage. It offers significantly lower storage costs compared to S3, making it more cost-effective for data that is rarely accessed. Glacier also provides different retrieval options with varying costs, such as Expedited, Standard, and Bulk retrievals, depending on the retrieval time required.

**- Retrieval Cost:**

**Amazon S3:**

Retrieval from S3 is designed for frequently accessed data and is typically fast and reliable. There are no retrieval costs for accessing data stored in the Standard storage class.

**Amazon Glacier:**

Glacier retrieval costs depend on the retrieval option chosen. Expedited retrievals are faster but more expensive, while Standard and Bulk retrievals have lower costs but longer retrieval times. It's essential to consider the retrieval costs when planning to access data stored in Glacier.

**- Data Transfer Costs:**

**Amazon S3:**

Data transfer costs for S3 are incurred when moving data between different AWS regions or when transferring data out of the S3 bucket to the internet or other AWS services.

**Amazon Glacier:**

Glacier data transfer costs apply when moving data between different AWS regions or when transferring data out of Glacier to the internet or other AWS services.

**- Minimum Storage Duration:**

**Amazon S3:**

S3 has no minimum storage duration, and you can store and retrieve data at any time without incurring penalties.

**Amazon Glacier:**

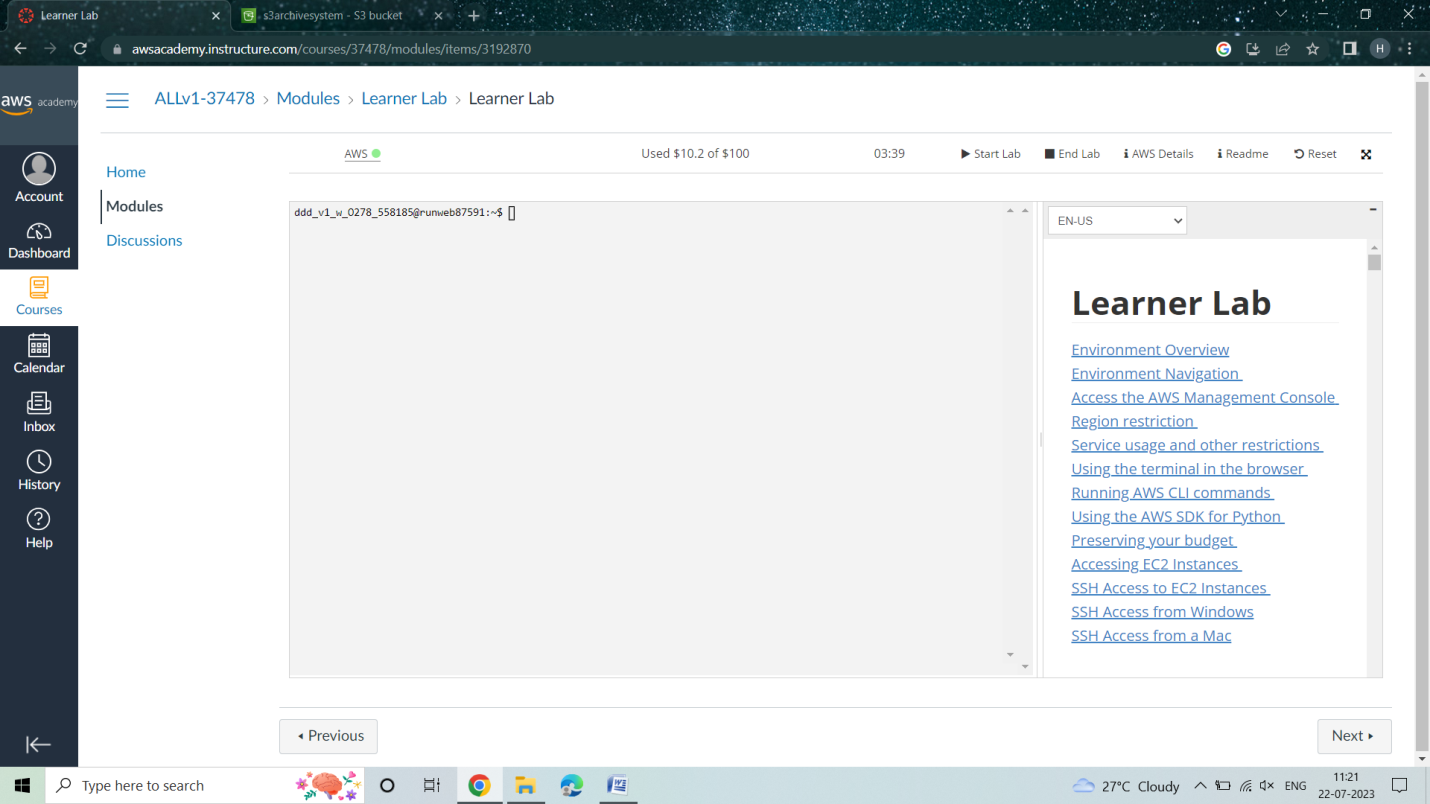
Glacier has a minimum storage duration of 90 days for each archive. Deleting data before the 90-day period may result in additional charges.

In summary, Amazon S3 is ideal for frequently accessed data with relatively higher storage costs, while Amazon Glacier is suitable for long-term data archival and cold storage with lower storage costs but potentially higher retrieval costs depending on the retrieval option chosen. The choice between S3 and Glacier depends on the data access patterns, retention requirements, and budget considerations for your specific use case.

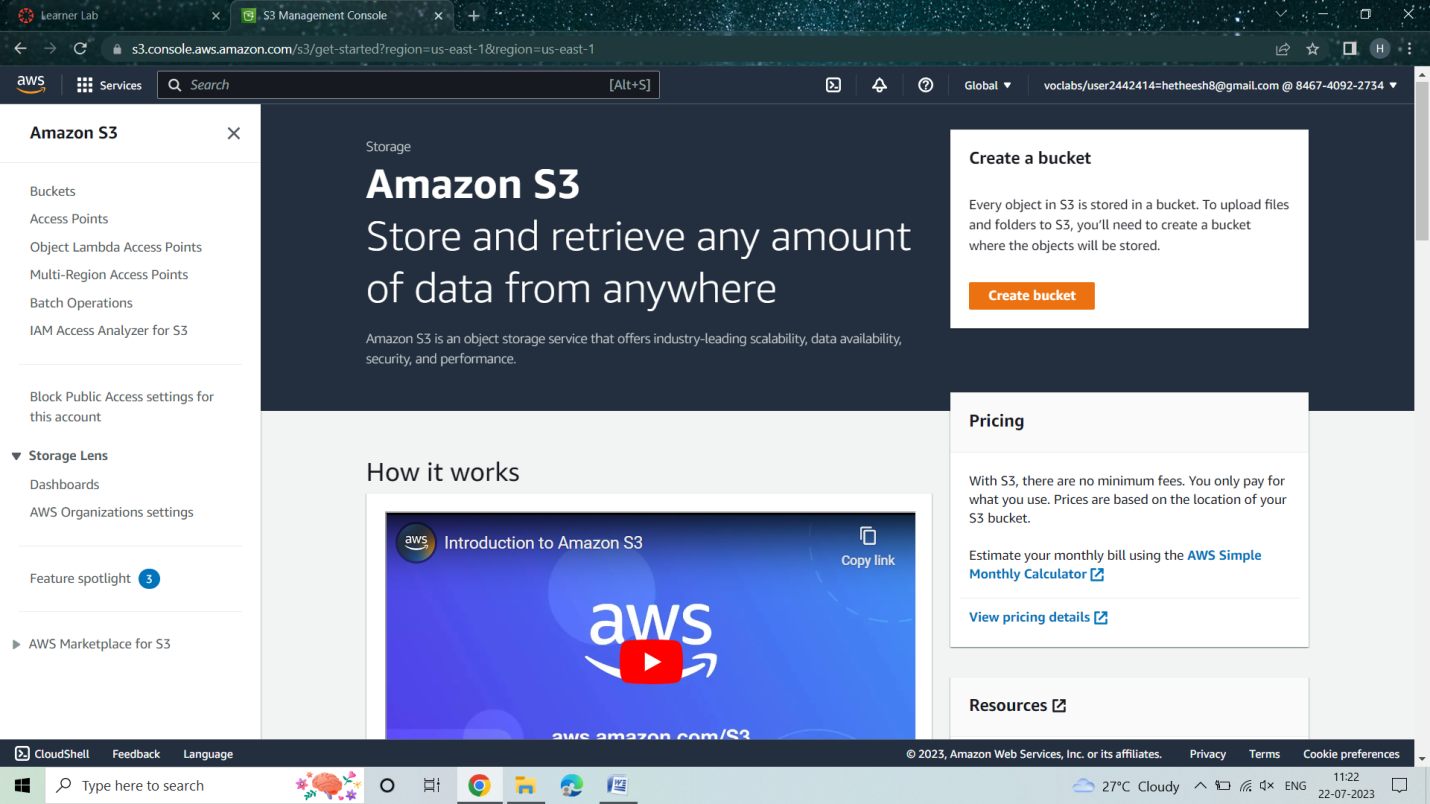
On an average, the approximate price for storing one gigabyte (1 GB) of data in Amazon S3 using the Standard storage class in the US East (N. Virginia) region is around $0.023 per month, where as, the approximate price for storing one gigabyte (1 GB) of data in Amazon Glacier using the Standard storage class in the US East (N. Virginia) region is $0.004 per gigabyte per month. This shows a clear understanding on the pricing in Amazon’s S3 and Glacier.

**Example of sending files to S3 from Glacier:**

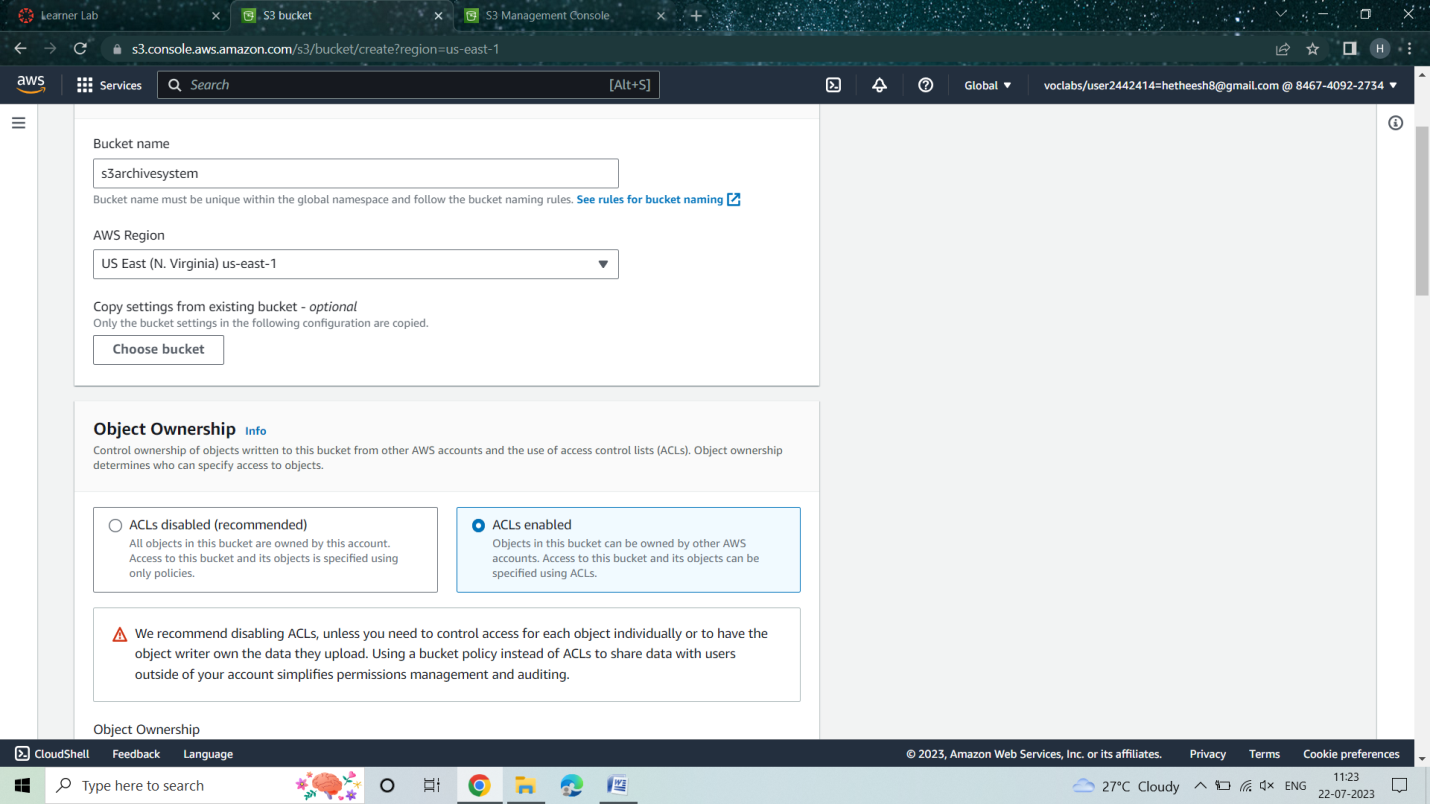
1. Login into AWS Management Console.



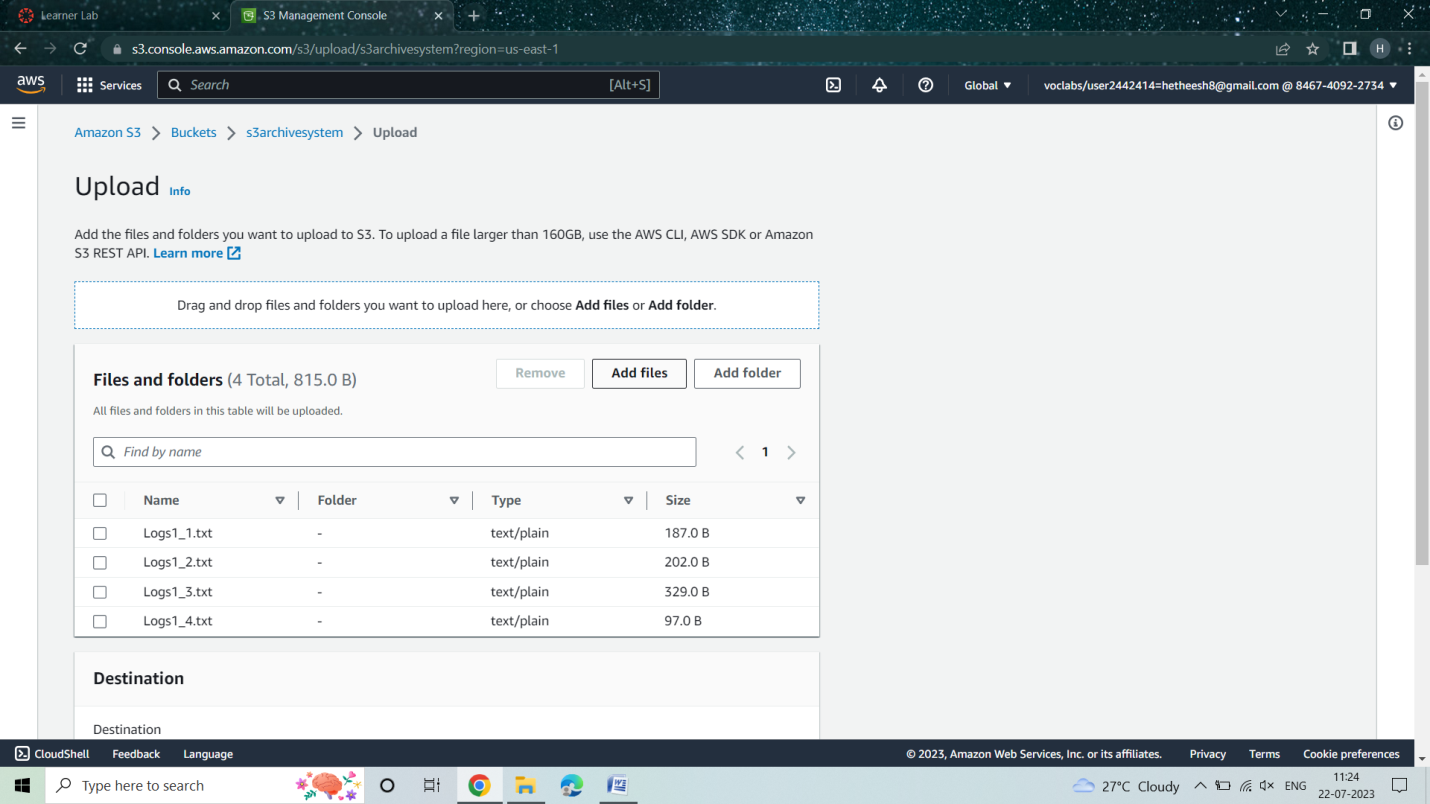
1. In services, select S3.



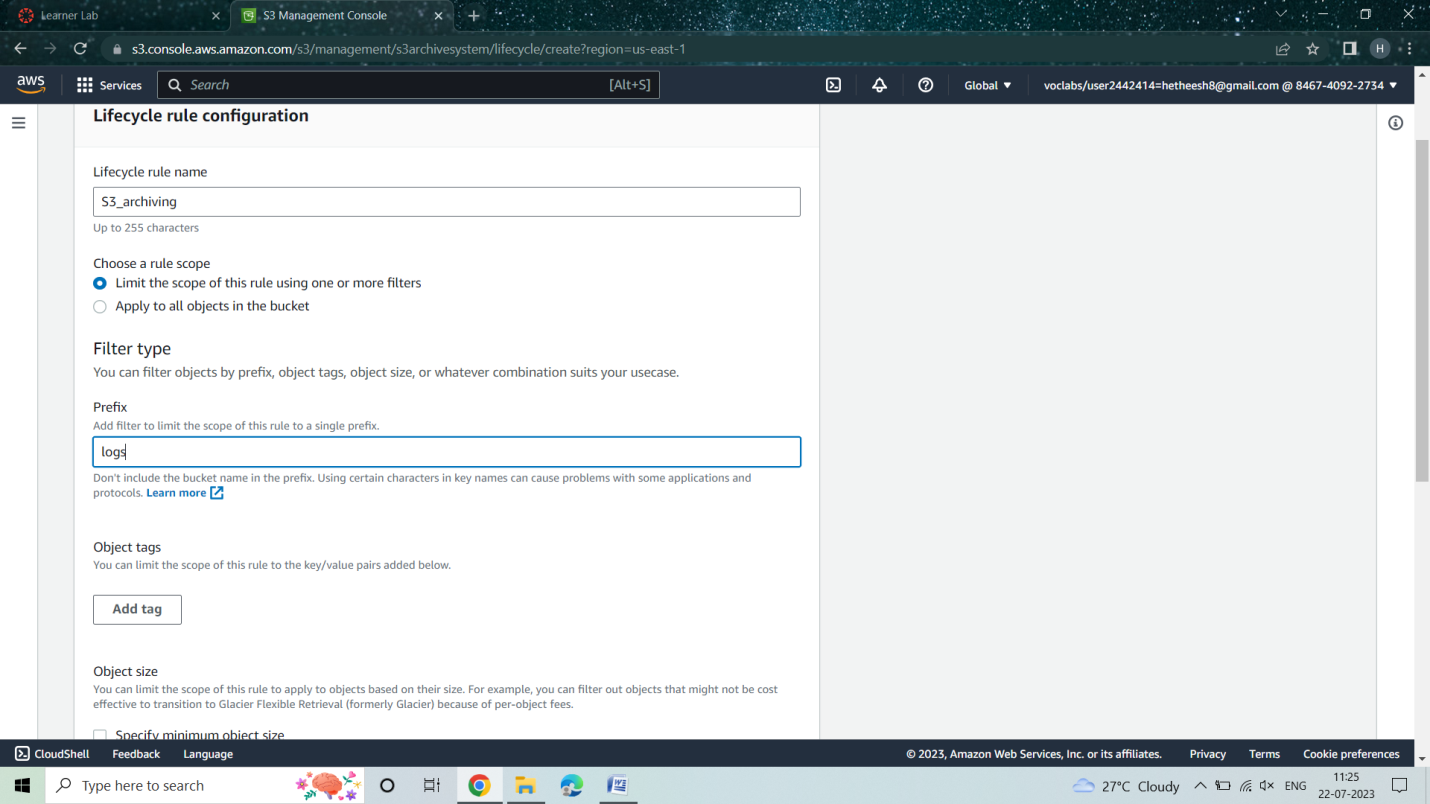
1. Click on “Create Bucket”, name the bucket(say s3archivesystem) set ACLs enabled and allow public access and click on create.



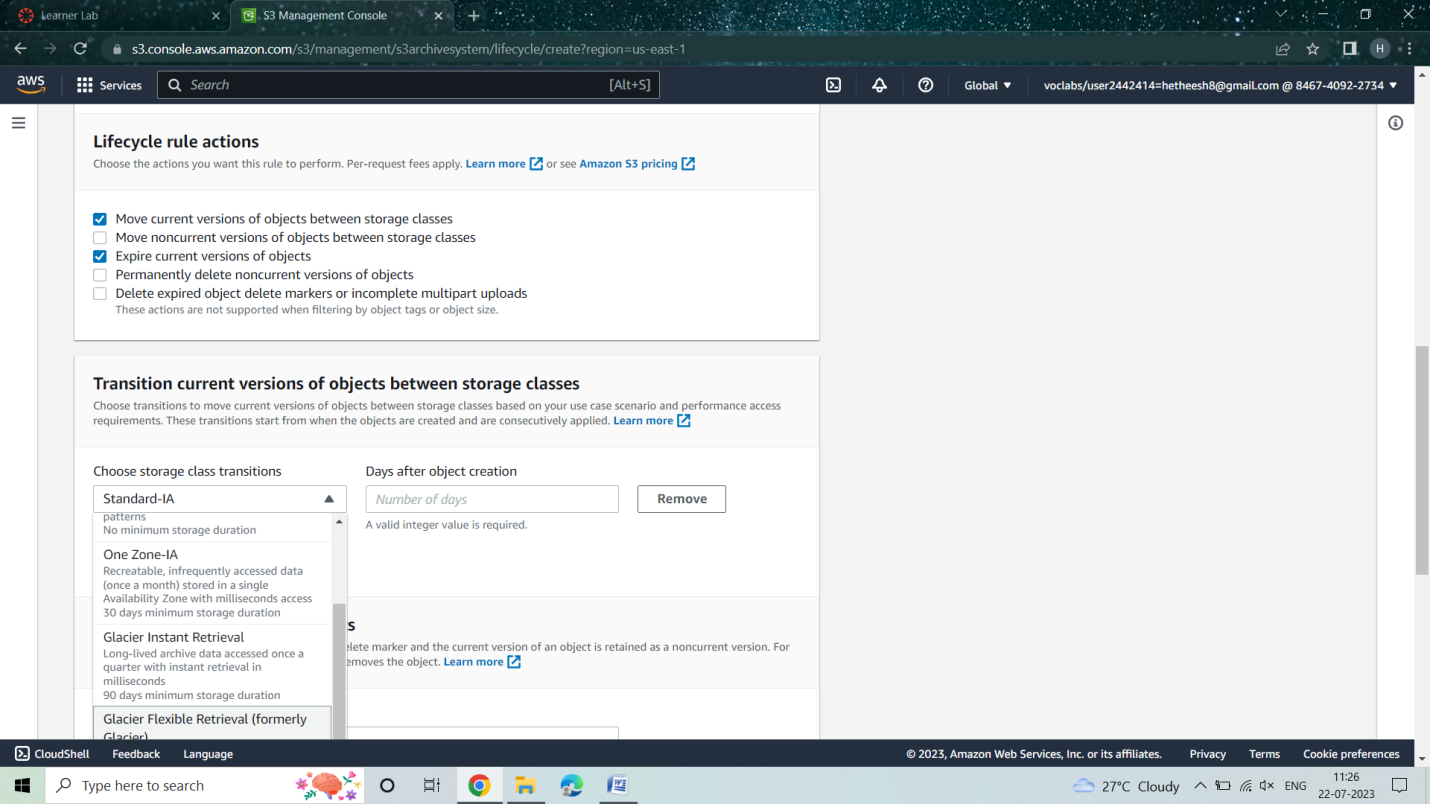
1. Now select the created bucket and upload the files to be moved to glacier after a certain time.



1. Now go to “Management” in the bucket and click on “Add Lifecycle rule”.

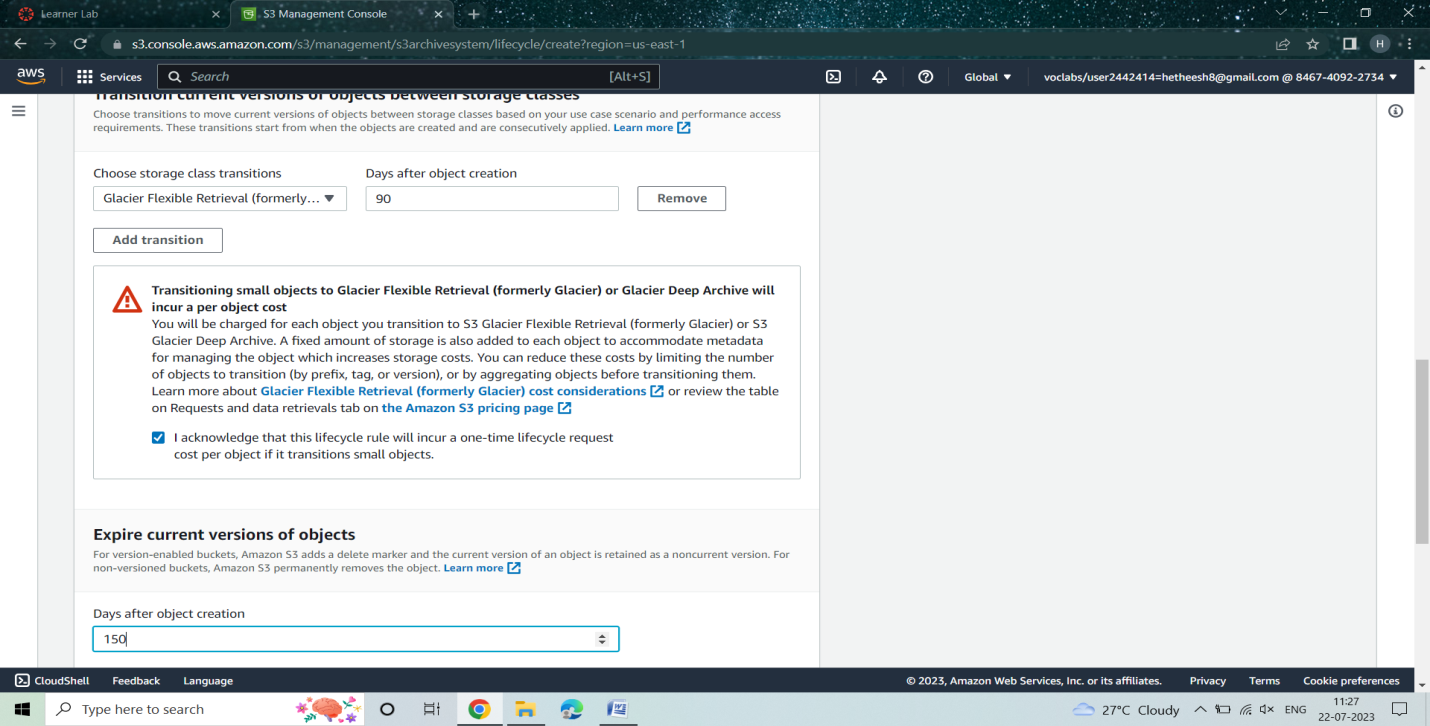


1. In lifecycle rule action select the check boxes as shown in below Figure and in “Transition current versions of objects between storage class”, choose “Glacier Flexible Retrieval(formerly Glacier)” as storage class.

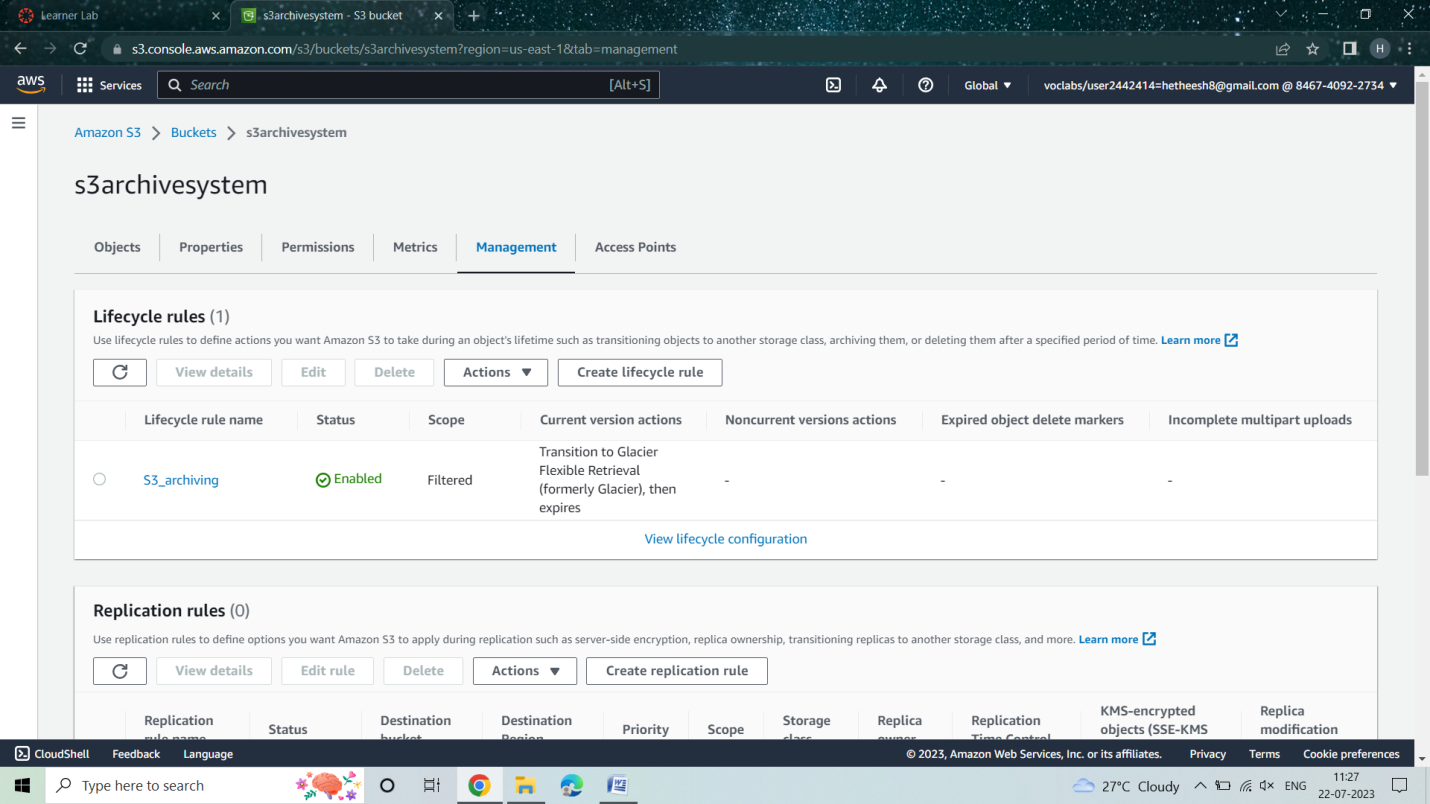


1. Specify the “Days after object creation” in both columns i.e for Moving to Glacier and also days after which the logs should be expired.

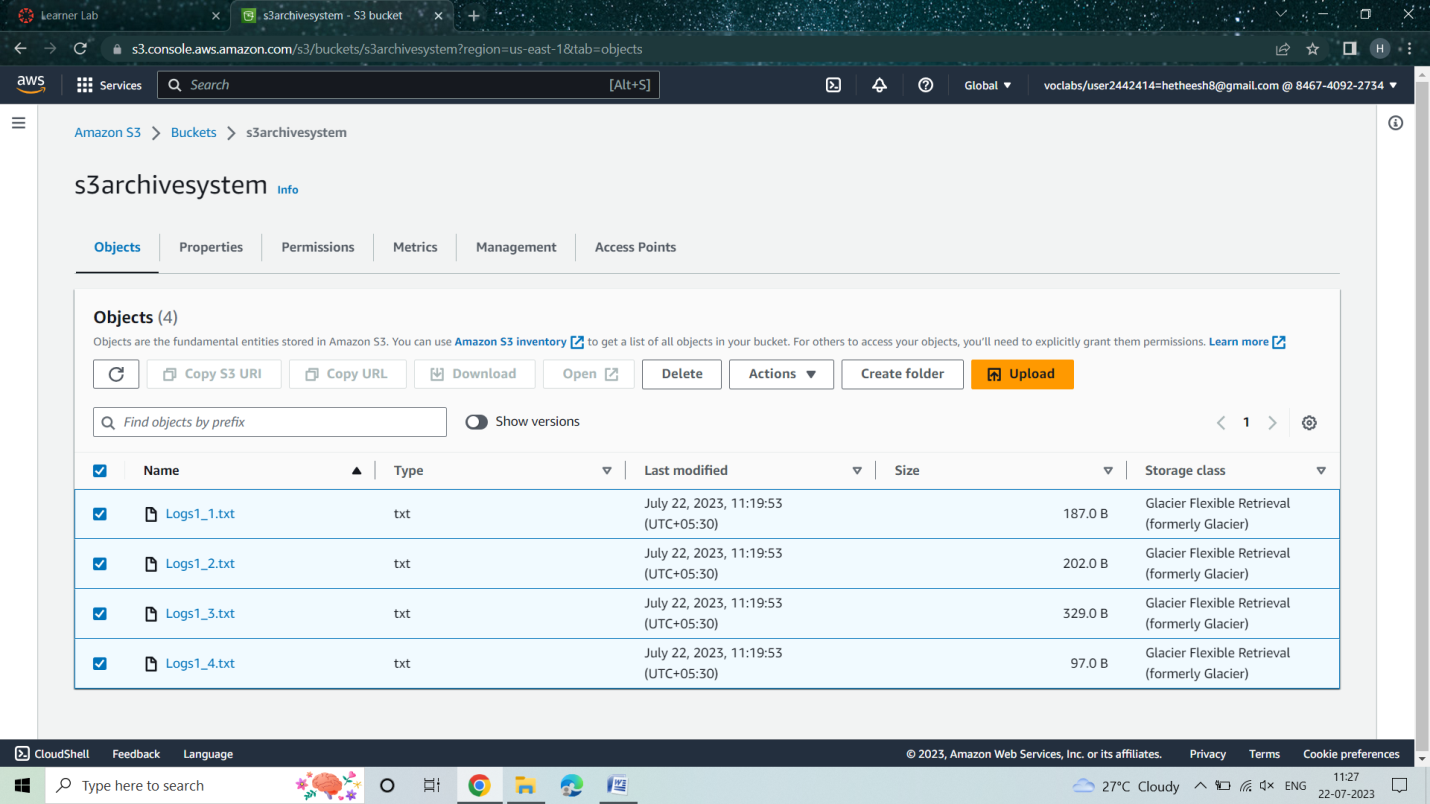
**Note:** Minimum days for moving into Glacier Flexible Retrieval(formerly Glacier) is 90 days.

****

1. A Lifecycle rule will be created as following.



1. We can also find that the files in the S3 bucket will have the storage class as “Glacier Flexible Retrieval(formerly Glacier)”.



**RESULT:**

After 90 days the logs will automatically moved to Glacier and get expired after 150 days. Before the logs get expired they can restored to respective S3 bucket whenever needed. This makes the process of storing files in S3 both cost-effective and as a easy way to retrieve data.