AWS S3 (Simple Storage Services) | Steps to transfer the files from EC2 Instance to S3 Bucket

Prepared by: Aravind Panchanathan
Divyansh Nigam
Mayuri Jadhav

Date: March 22, 2024

In this manual, we will cover AWS S3 (Simple Storage Service), a storage service from AWS and in this guide, we will provide detailed instructions on transferring files to S3 through EC2 instance.

1. Introduction to Amazon S3:

Amazon S3, also known as Amazon Simple Storage Service, is a cloud storage service offered by Amazon Web Services (AWS). It's designed for online storage of any type of file - from images and videos to code, documents and scientific data. S3 is known for its scalability, high availability, data durability, and security.

2. Why Use AWS Simple Storage Services (S3):

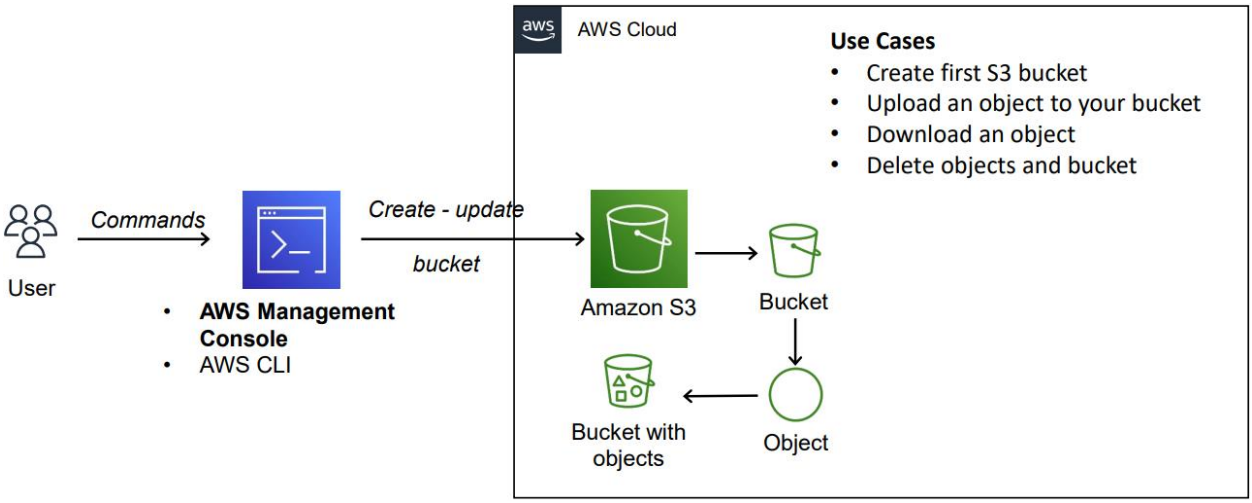
Object storage: S3 uses object storage architecture, which means data is stored in basic units called objects. Each object has a unique key that identifies it within a bucket, a virtual container that groups related objects.

Scalability and Availability: Designed to handle massive amounts of data, S3 can scale to meet your storage needs as they grow. It also offers high availability, meaning your data is stored redundantly across multiple locations to minimize downtime.

Security: S3 offers robust security features to protect your data, including access control lists (ACLs) to define who can access your data and encryption to keep your data confidential.

3. How does AWS Simple Storage Services (S3) work?

- You upload a file to an S3 bucket.
- S3 breaks the file into smaller pieces (objects) and stores them redundantly across its infrastructure.
- You can define access permissions for the bucket and its objects using ACLs.
- Whenever you need to access the file, you can download it from the S3 bucket.



4. Features of Simple Storage Services

- **Scalability:** S3 scales seamlessly to accommodate any amount of data. You only pay for the storage you use, eliminating the need to provision fixed storage capacity.
- **High Availability:** S3 ensures that your data is highly available by storing it redundantly across multiple locations. This minimizes downtime in case of hardware failures.
- **Durability:** S3 offers exceptional data durability, meaning your data is extremely unlikely to be lost. S3 claims 99.999999999% (11 nines) of object durability, ensuring your data remains secure for long periods.
- **Security:** S3 provides robust security features to protect your data. Access control lists (ACLs) allow you to define granular permissions, specifying who can access, upload, download, or delete objects within a bucket. Additionally, S3 supports server-side and client-side encryption to keep your data confidential at rest and in transit.
- **Storage Classes:** S3 offers various storage classes to optimize costs based on your data access needs. Options include Standard for frequently accessed data, Glacier for infrequently accessed data, and Glacier Deep Archive for rarely accessed data, each with corresponding pricing structures.
- **Lifecycle Management:** S3 Lifecycle Management allows you to automate the transition of objects between different storage classes based on user-defined rules. This helps optimize costs by automatically moving less frequently accessed data to cheaper storage tiers.
- **Versioning:** S3 Object Versioning lets you maintain a history of all object versions, allowing you to roll back to previous versions if necessary.

- **Data Analytics Integration:** S3 integrates seamlessly with various AWS data analytics services like Amazon Athena and Amazon EMR, enabling you to perform analytics directly on your S3 data lake.
- **Static Website Hosting:** S3 can be used to host static websites, allowing you to store website content like HTML, CSS, and JavaScript files in S3 buckets.
- **Cloud Endpoints:** S3 supports Cloud Endpoints, which provide private endpoints for accessing S3 resources from within your Amazon Virtual Private Cloud (VPC).

5. What are uses cases of S3?

- **Backup and Disaster Recovery:** S3's durability and scalability make it ideal for storing backups of critical data. You can easily create backups of your databases, applications, and other important information in S3 buckets. In case of a disaster or system failure, you can restore your data quickly and efficiently from S3.
- **Data Archiving:** S3 Glacier and S3 Glacier Deep Archive storage classes offer cost-effective solutions for archiving inactive data that you need to retain for long periods but don't require frequent access.
- **Data Lakes for Analytics:** S3 serves as a central repository for data lakes, storing vast amounts of data from various sources in its raw format. This data can then be easily accessed and analyzed by big data analytics tools like Amazon EMR and Amazon Athena.
- **Web Application Storage:** S3 can store static website content like HTML, CSS, and JavaScript files. You can even leverage S3 Cloud Endpoints for secure access to these resources within your VPC.
- **Media Hosting:** S3 is a great option for storing and distributing images, audio, and video files. Features like content delivery networks (CDNs) can be integrated to deliver content with low latency and high performance.
- **Software Delivery:** S3 can be used to securely store and distribute software downloads for applications or systems.
- **IoT Device Data Storage:** S3 can handle the high volume and variety of data generated by IoT devices, making it a suitable storage solution for sensor data, logs, and other IoT-related information.

6. Key Functionalities of S3?

Object Storage:

- **Fundamental Unit:** Unlike traditional file systems with folders and subfolders, S3 utilizes object storage. Your data resides in objects, which are essentially files with additional metadata attached (think file size, creation date, etc.). Each object has a unique identifier

within a bucket, a virtual container that groups related objects. Imagine buckets as folders and objects as the individual files within those folders.

Management Functionalities:

- **CRUD Operations on Objects:** S3 allows you to perform basic CRUD (Create, Read, Update, Delete) operations on objects within buckets. You can upload files (create objects), download them whenever needed (read), delete unwanted objects, and manage their lifecycle within S3.
- **Listing Buckets and Objects:** For better organization and identification of your stored data, S3 provides functionalities to list all your S3 buckets or view the objects within a specific bucket.

Security Functionalities:

- **Access Control Lists (ACLs):** S3 offers fine-grained access control through ACLs. These lists define who can access (read, write, delete, etc.) your data within a bucket. This ensures restricted access and helps you secure sensitive information.
- **Encryption:** To add an extra layer of security, S3 supports both server-side and client-side encryption. Server-side encryption encrypts your data at rest within S3, while client-side encryption encrypts your data before upload and decrypts it after download.

Data Protection and Availability:

- **Object Versioning:** S3 Object Versioning allows you to maintain a history of all object versions. This acts like a safeguard, enabling you to retrieve previous versions if necessary. For instance, if you accidentally overwrite an important file, you can revert to a previous version.
- **Exceptional Data Durability:** S3 boasts exceptional data durability, meaning your data is extremely unlikely to be lost. S3 claims a staggering 99.999999999% (11 nines) of object durability, ensuring your data remains secure for extended periods.
- **Storage Classes:** S3 provides a variety of storage classes to optimize costs based on how often you access your data. Standard Storage is ideal for frequently accessed data, while Glacier and Glacier Deep Archive are cost-effective options for infrequently accessed data, with varying retrieval times based on the storage class.

Scalability and Integration:

- **Effortless Scaling:** One of S3's strengths is its seamless scaling. S3 can automatically scale up or down to accommodate any amount of data you store. You only pay for the storage you use, eliminating the need to provision fixed storage capacity beforehand.

- **Lifecycle Management:** S3 Lifecycle Management allows you to automate the transition of objects between storage classes based on user-defined rules. This helps you optimize costs by automatically moving less frequently accessed data to cheaper storage tiers.
- **Data Analytics Integration:** S3 integrates with various AWS data analytics services like Amazon Athena and Amazon EMR. This integration allows you to perform data analytics directly on your S3 data lake, eliminating the need to move your data for processing.

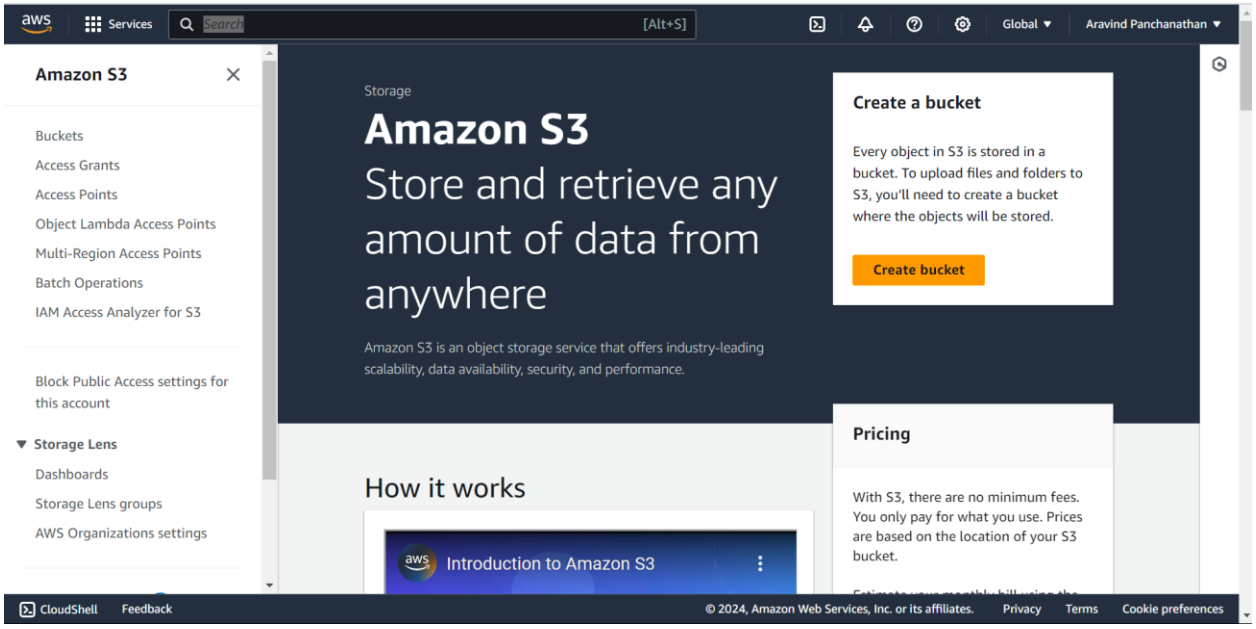
Additional Functionalities:

- **Static Website Hosting:** S3 can be used to host simple websites. You can store website content like HTML, CSS, and JavaScript files in S3 buckets and configure S3 to deliver them as a static website.
- **Cloud Endpoints:** S3 supports Cloud Endpoints, which provide private endpoints for accessing S3 resources from within your Virtual Private Cloud (VPC). This allows you to create secure connections to your S3 buckets from resources within your VPC.

7. Steps to create S3 bucket.

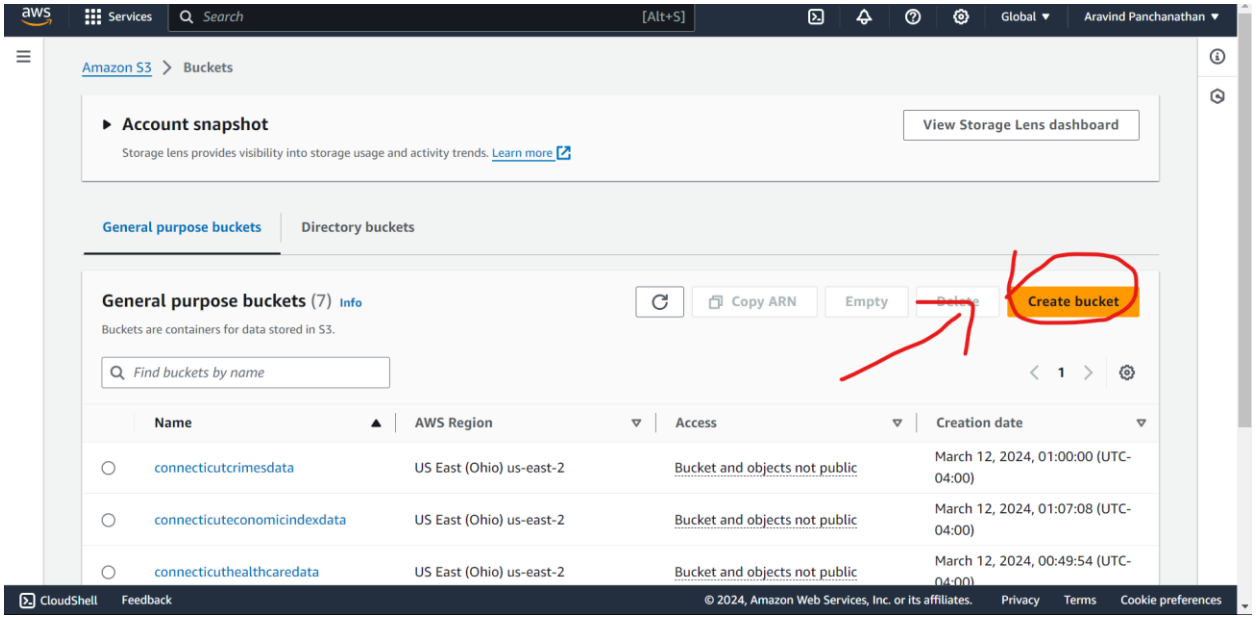
Step 7.1:

Navigate to Amazon S3 (https://s3.console.aws.amazon.com/)



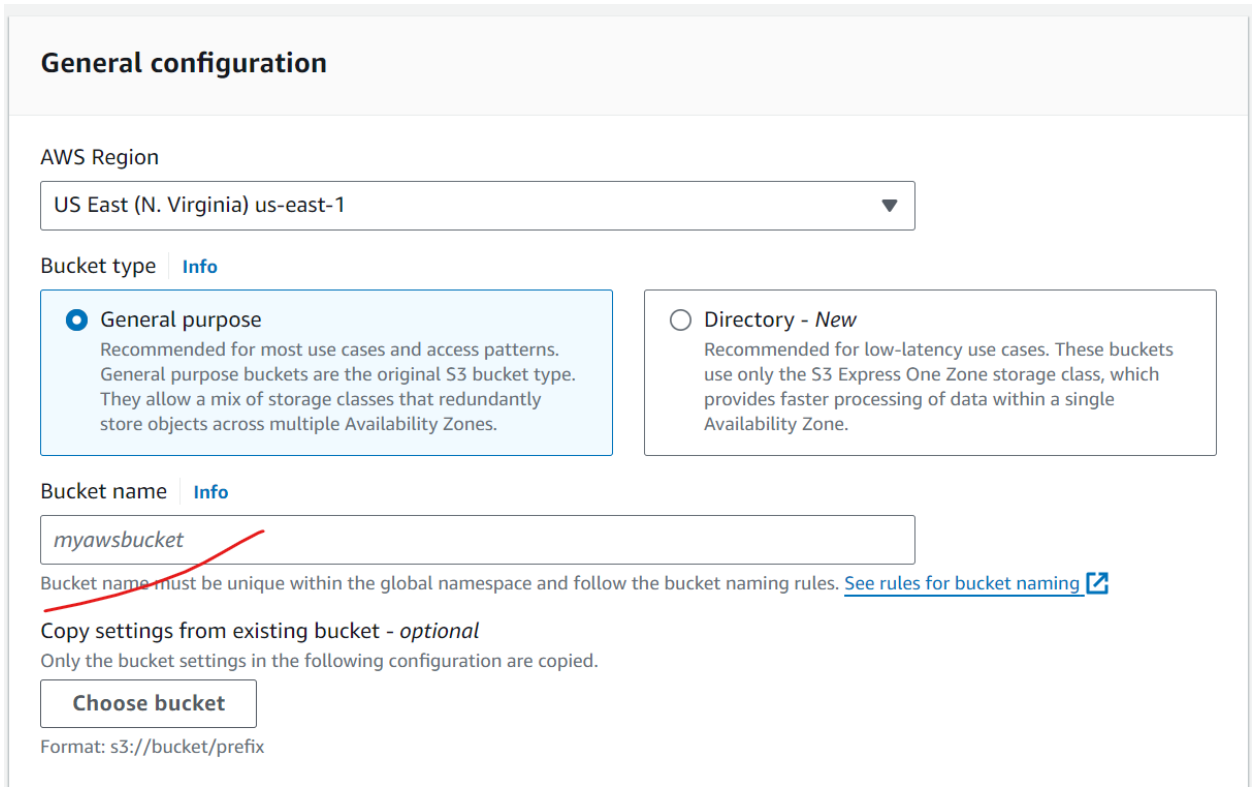
Step 7.2:

Navigate to Buckets and Click on "**Create Bucket**"



Step 7.3:

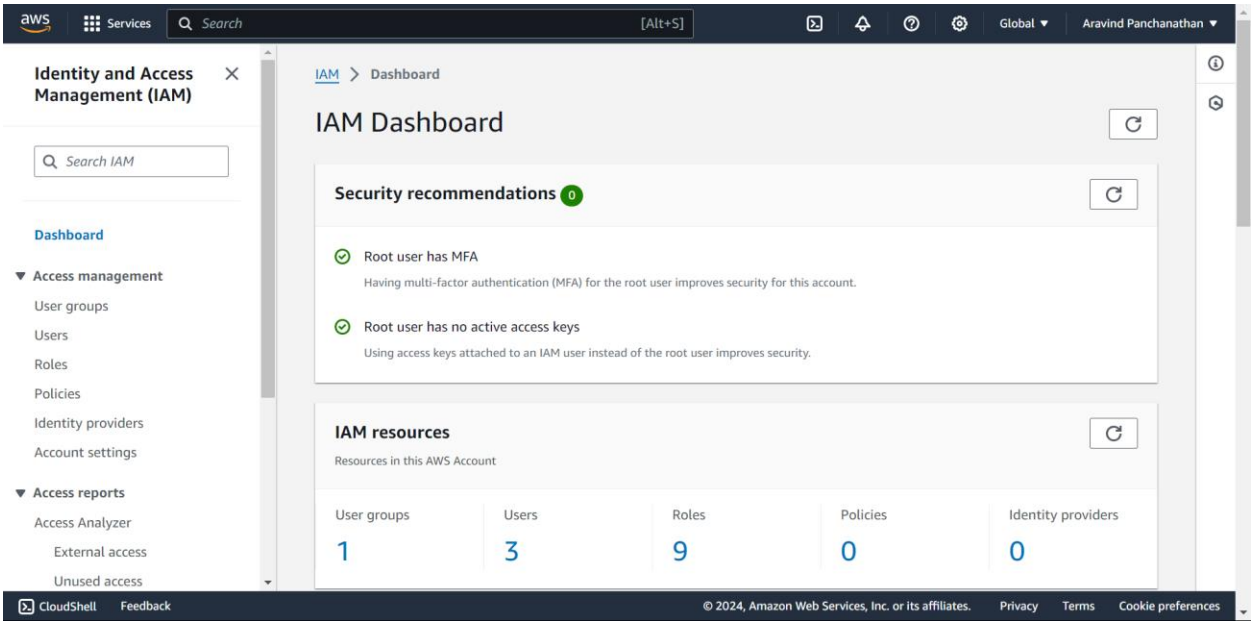
Under General Configuration > Give Bucket a name & follow default options to create a bucket.



8. Configure an IAM Role (EC2-S3 Access):

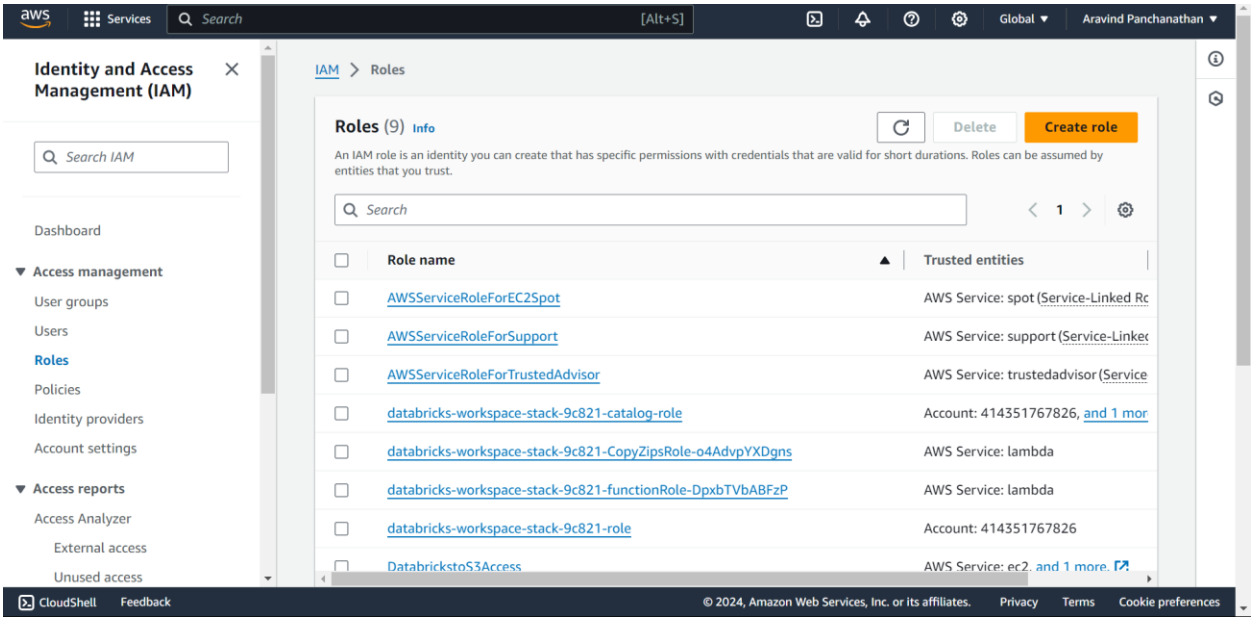
Step 8.1:

- Navigate to AWS IAM (https://us-east-1.console.aws.amazon.com/)



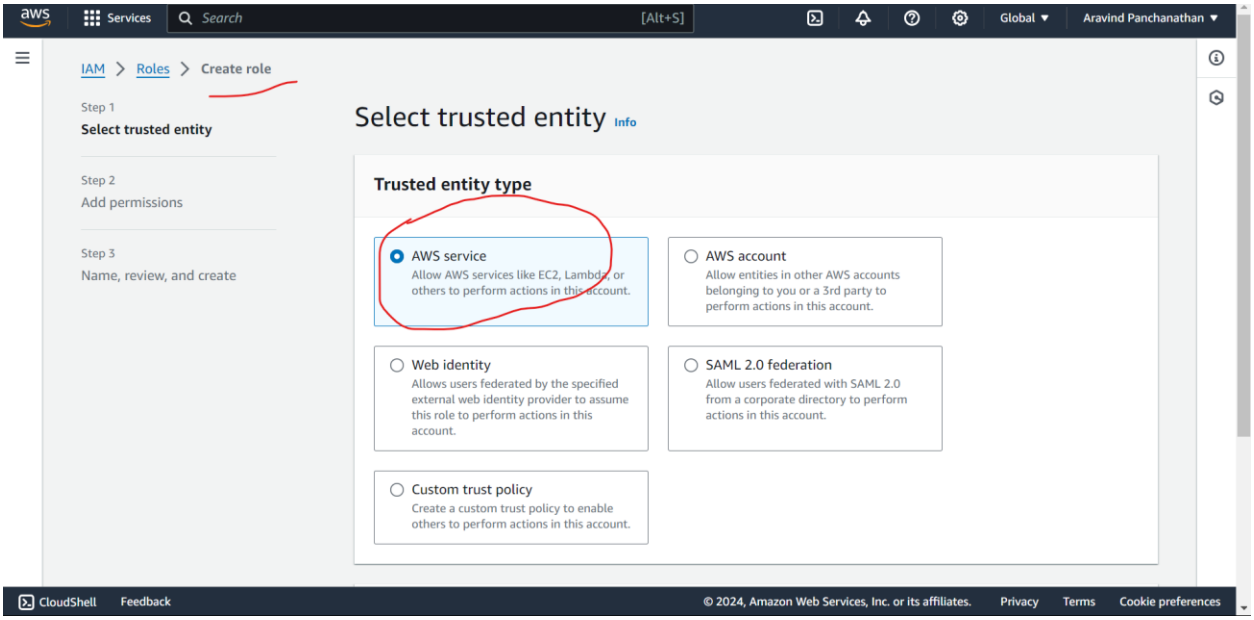
Step 8.2:

- Navigate to Roles



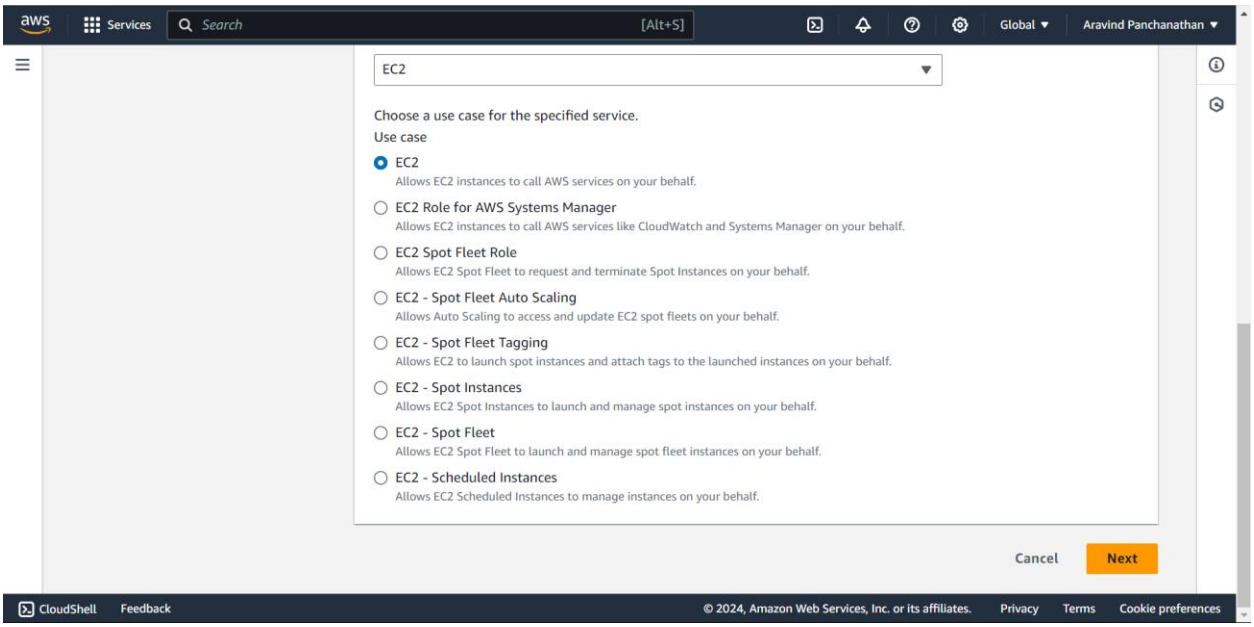
Step 8.3:

Create new role, Select AWS Service.



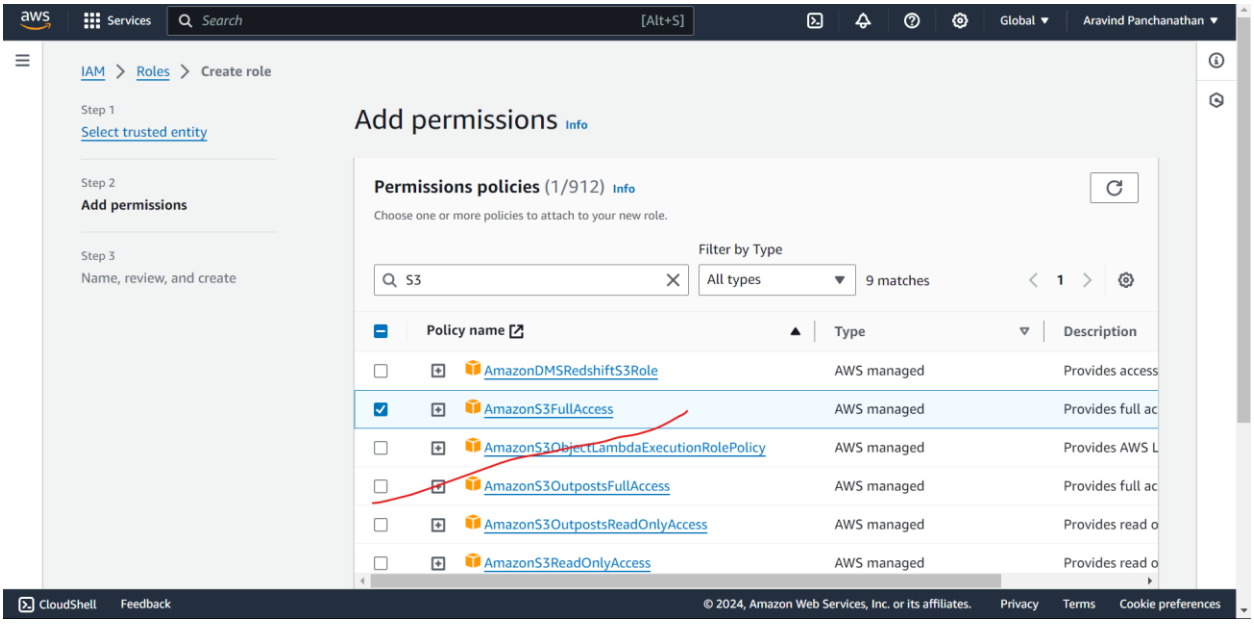
Step 8.4:

Scroll down and select use case to be EC2 & select Next.



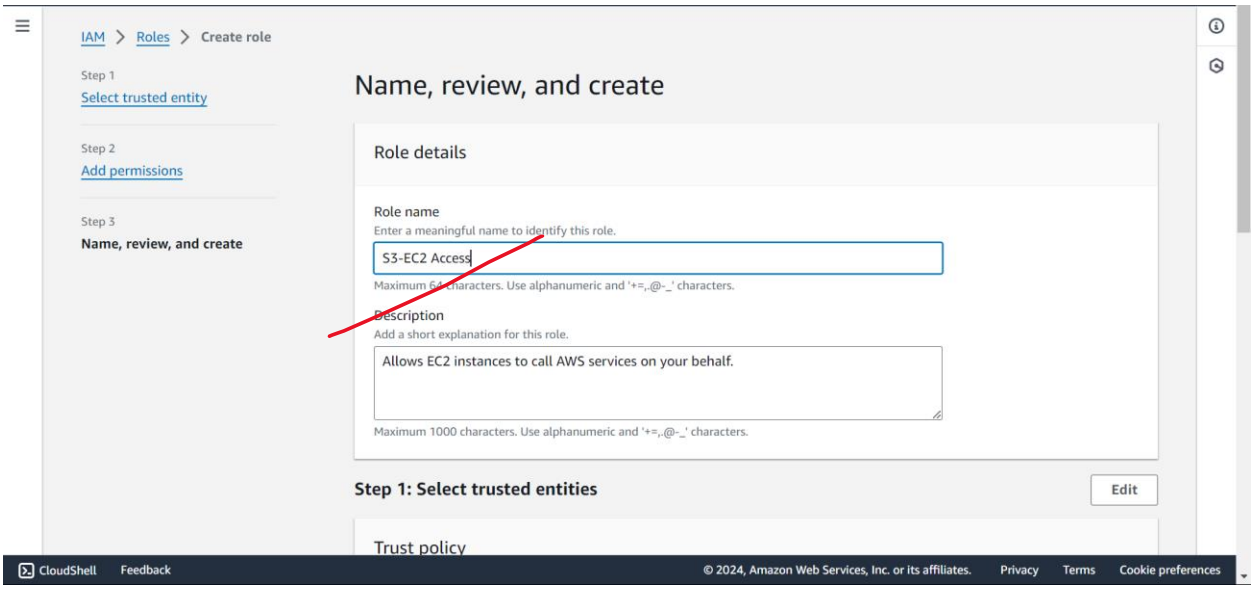
Step 8.5:

Under Permission Policy, Select S3 Full Access



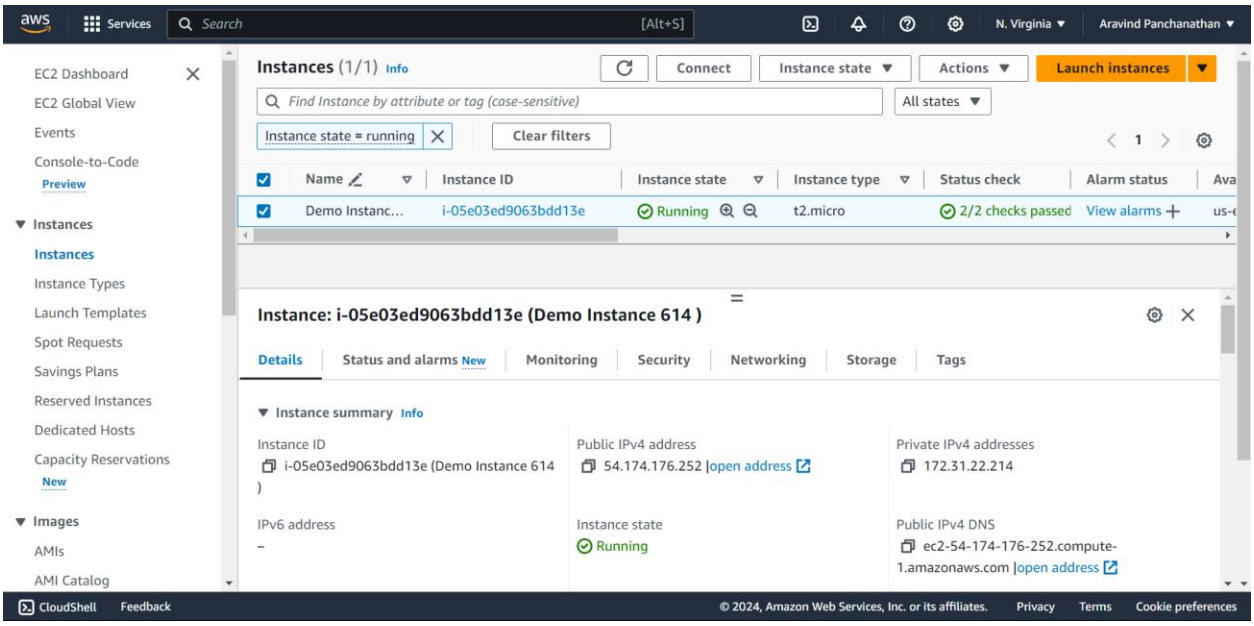
Step 8.6

Give a name for role and select review and create.

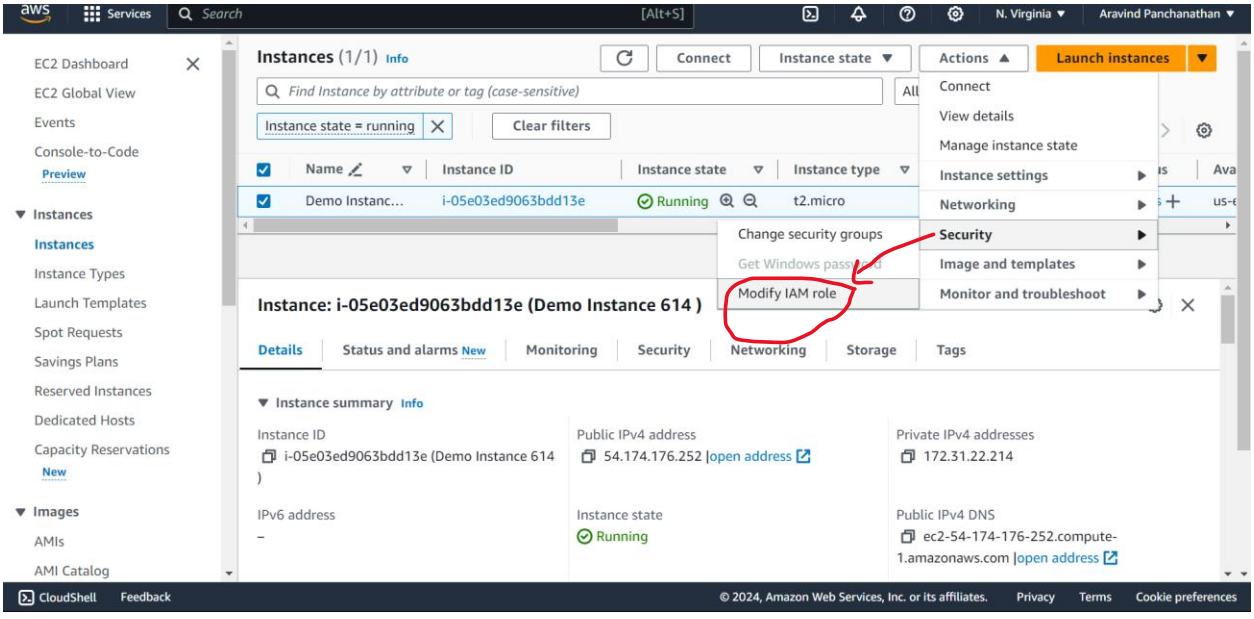


9. Attach IAM role with EC2:

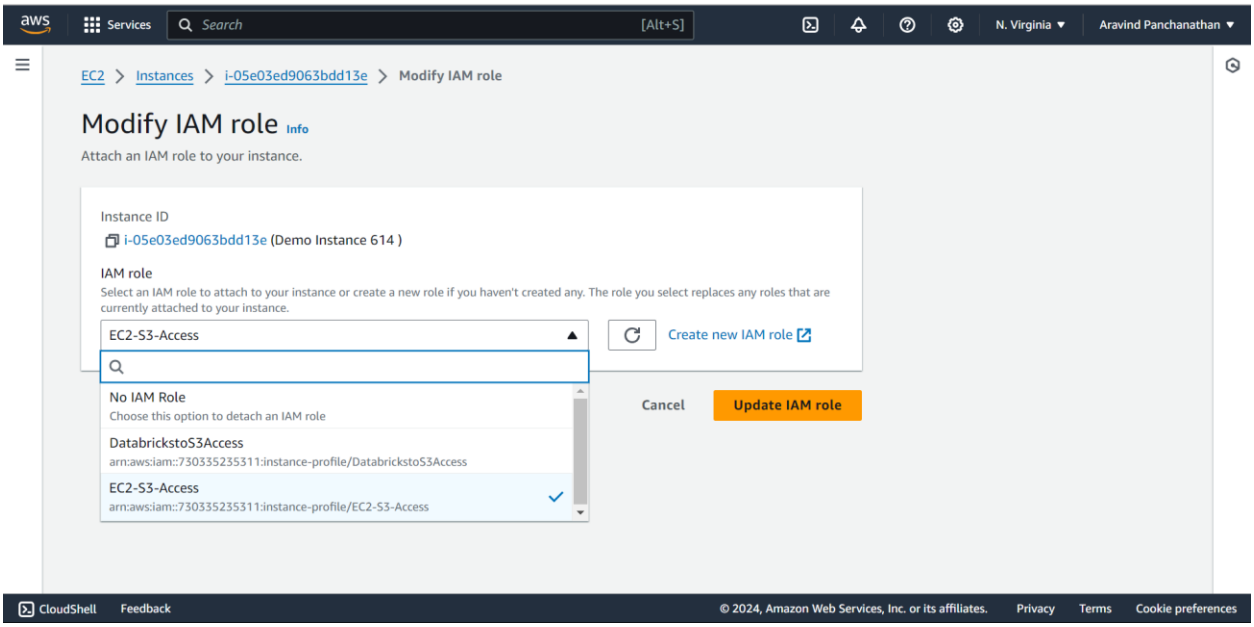
Step 9.1: Navigate to EC2 and choose the Instance



Step 9.2: Click on Actions and Modify IAM Role



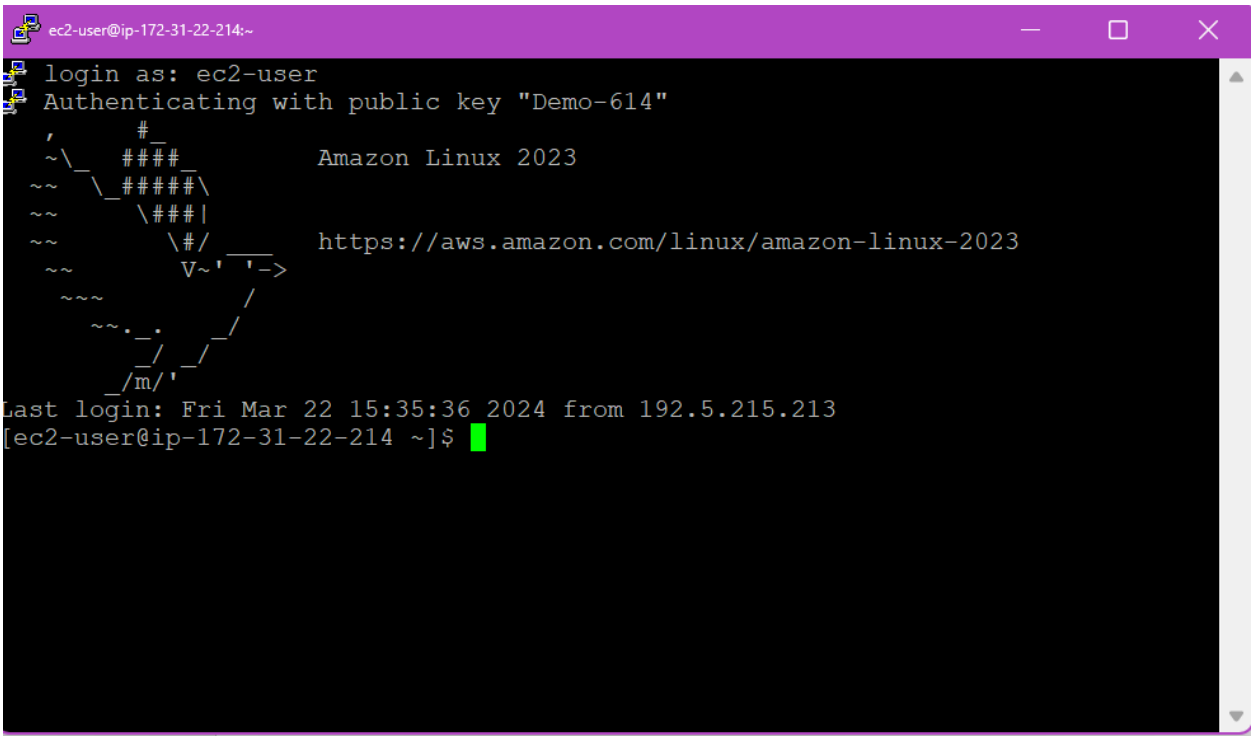
Step 9.3: Choose the appropriate IAM Role



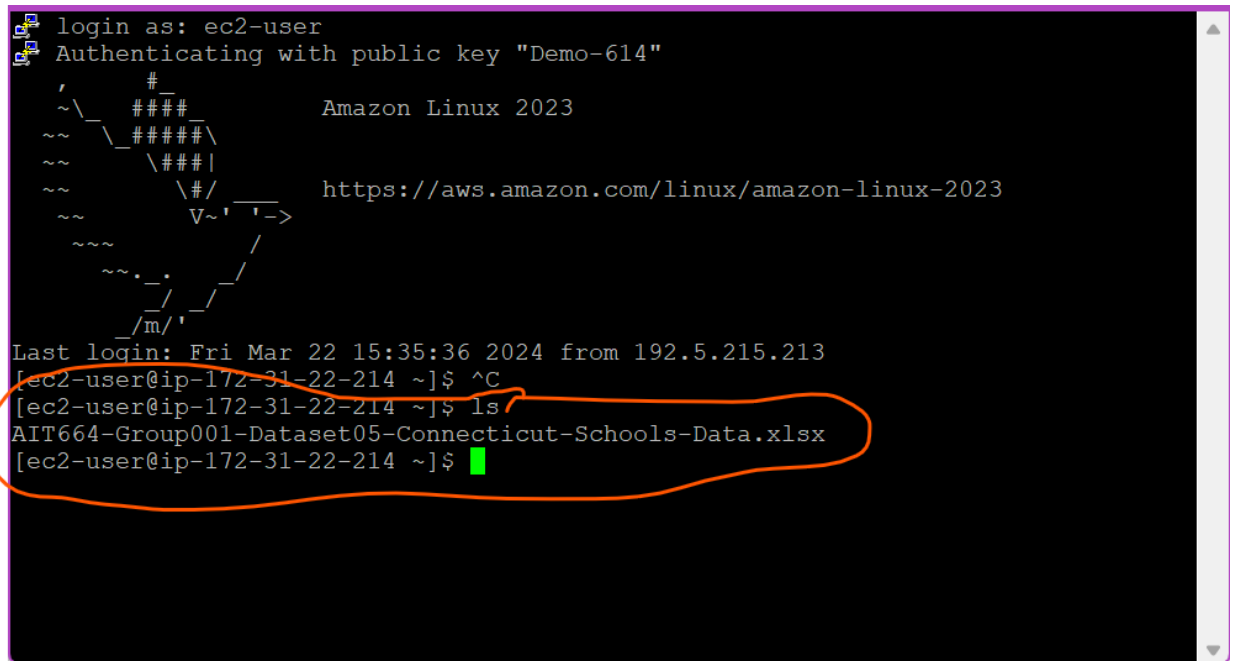
10. Transfer the file from EC2 to S3:

Step 10.1:

Access the EC2 Instance through puTTy & login as ec2-user

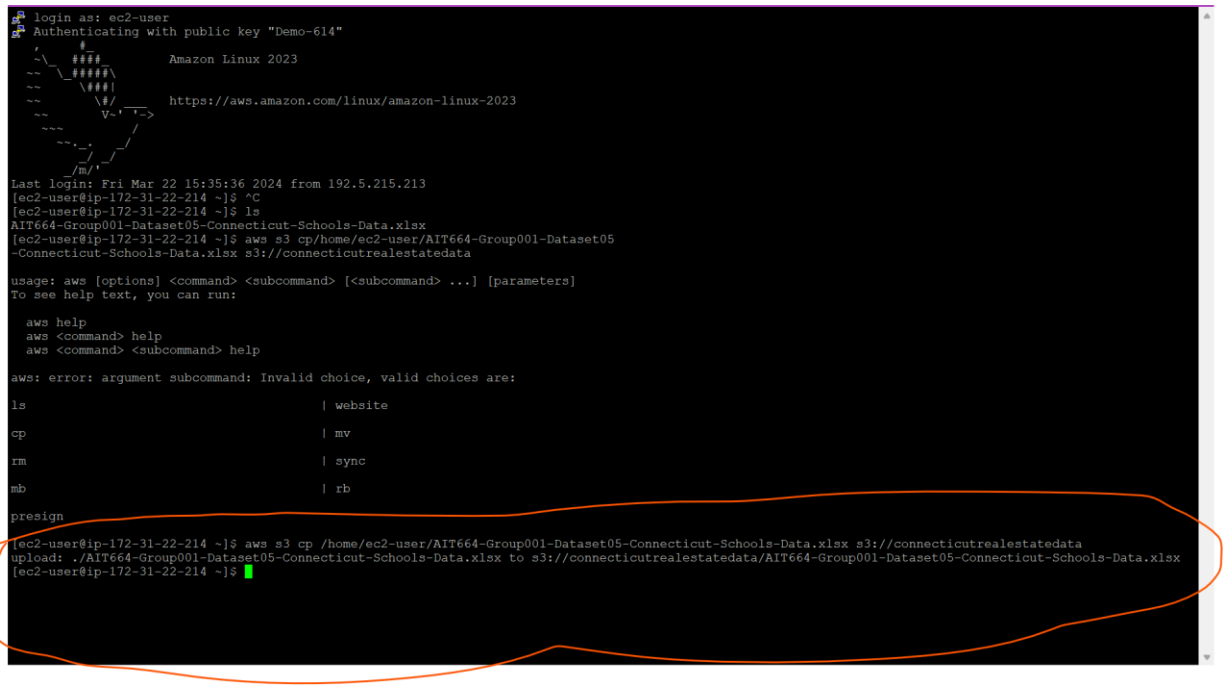


Step 10.2: Type “ls” command on the terminal to access the data file in the system



Step 10.3: Type the command to transfer the file from EC2 Instance to S3 bucket

“aws s3 cp /home/ec2-user/<file-name.csv> s3://<bucket-name>”



Step 10.4: Verify if the file has come inside the bucket by navigating by S3 buckets, and see if the from EC2 file is present inside the bucket