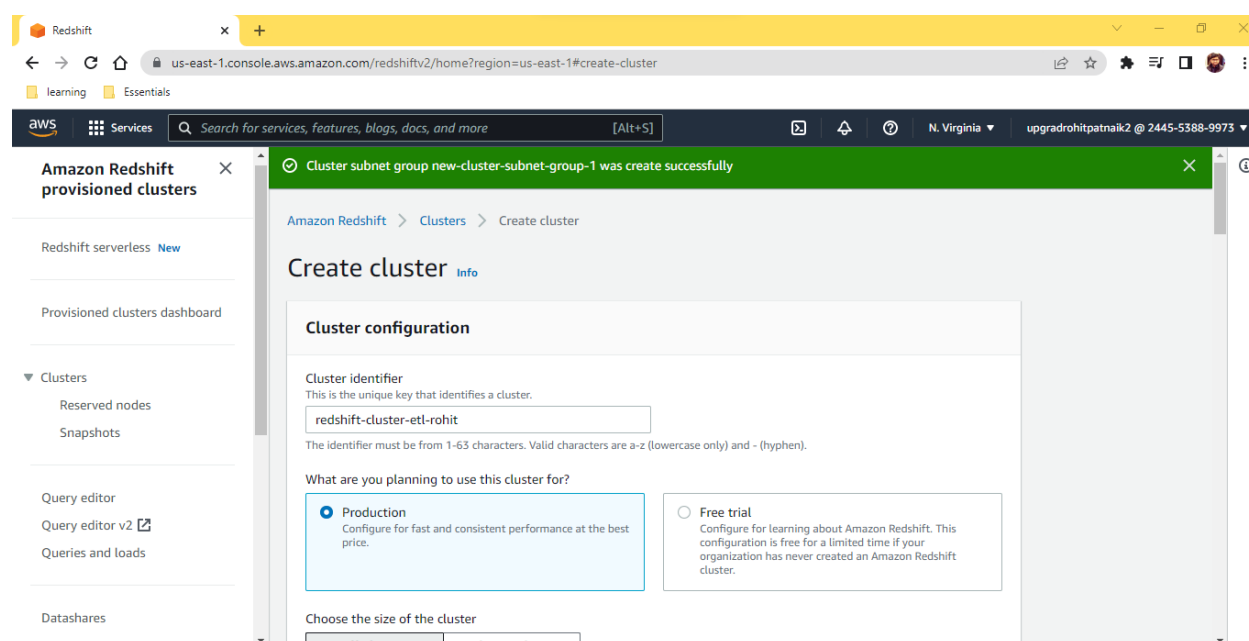


Creation of a Redshift Cluster

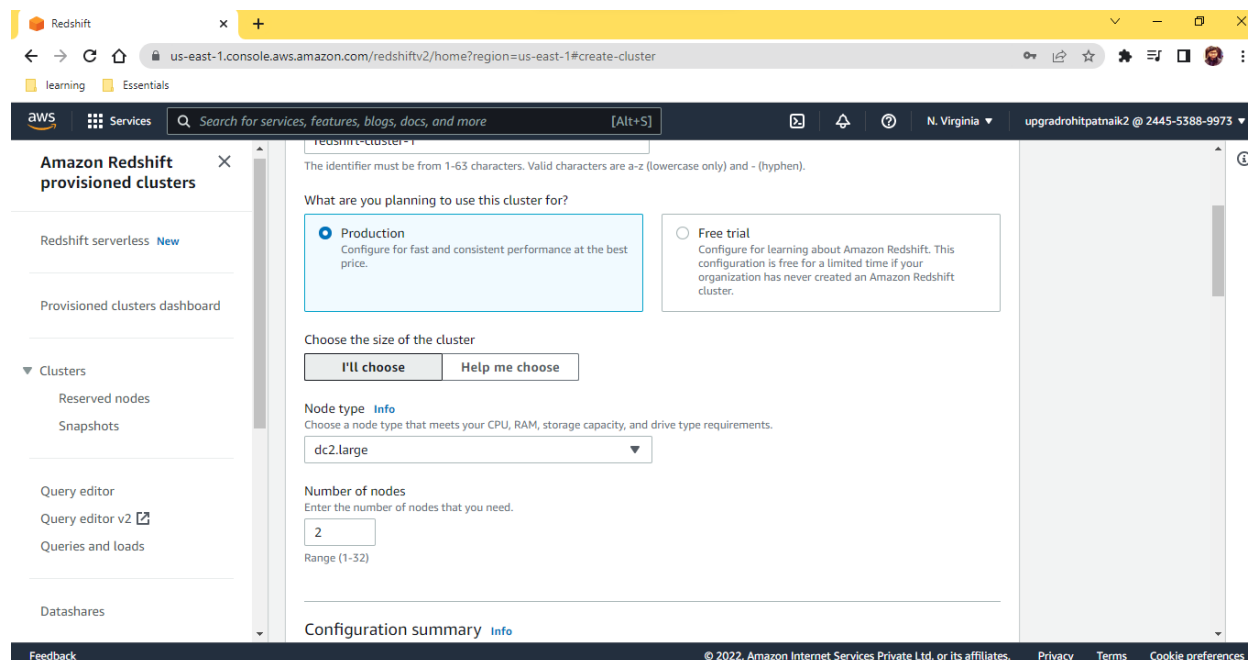
Screenshots of the configuration of the Redshift cluster that you have created:

<Screenshot of the type of machine used along with number of nodes>

Creation of new cluster redshift-cluster-etl-rohit



Change the node type to dc2.large



Redshift

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#create-cluster

learning Essentials

Services Search for services, features, blogs, docs, and more [Alt+S]

N. Virginia upgradrohitpatnaik2 @ 2445-5388-9973

Amazon Redshift provisioned clusters

Redshift serverless **New**

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Feedback

© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences

redshift-cluster-1

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

☒ **Production**
Configure for fast and consistent performance at the best price.

☐ **Free trial**
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

I'll choose **Help me choose**

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

dc2.large

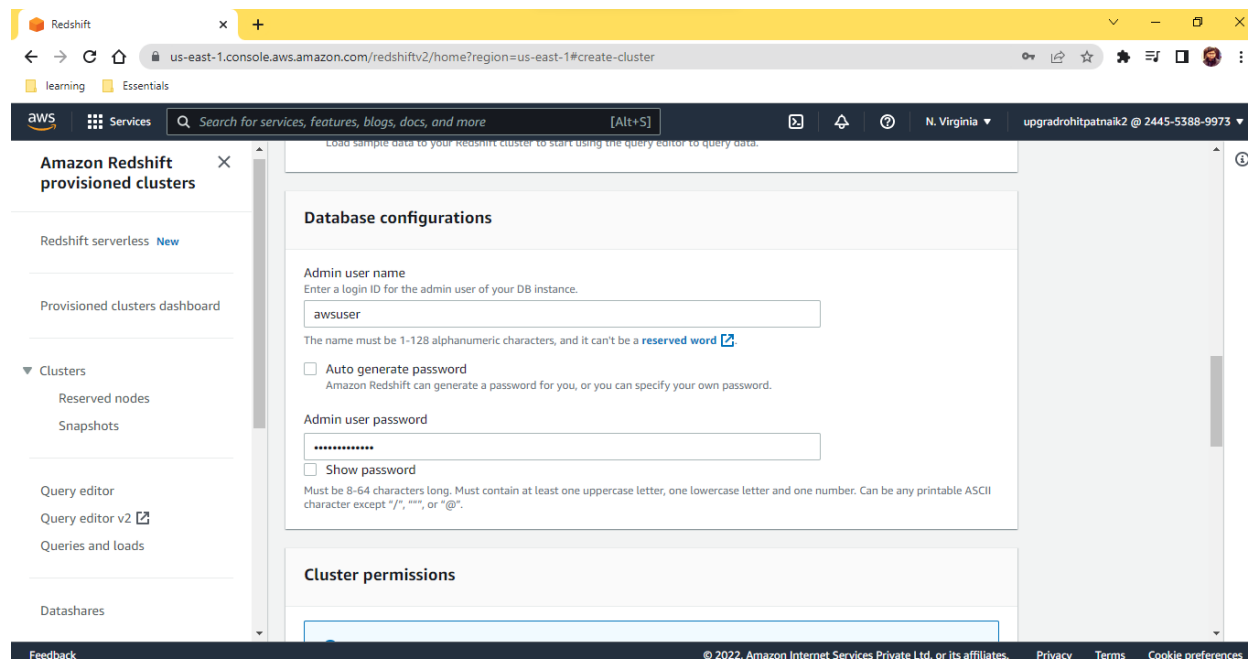
Number of nodes
Enter the number of nodes that you need.

2

Range (1-32)

Configuration summary [Info](#)

Change the database configuration



Redshift

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#create-cluster

learning Essentials

Services Search for services, features, blogs, docs, and more [Alt+S]

N. Virginia upgradrohitpatnaik2 @ 2445-5388-9973

Amazon Redshift provisioned clusters

Redshift serverless **New**

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Feedback

© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences

Load sample data to your redshift cluster to start using the query editor to query data.

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

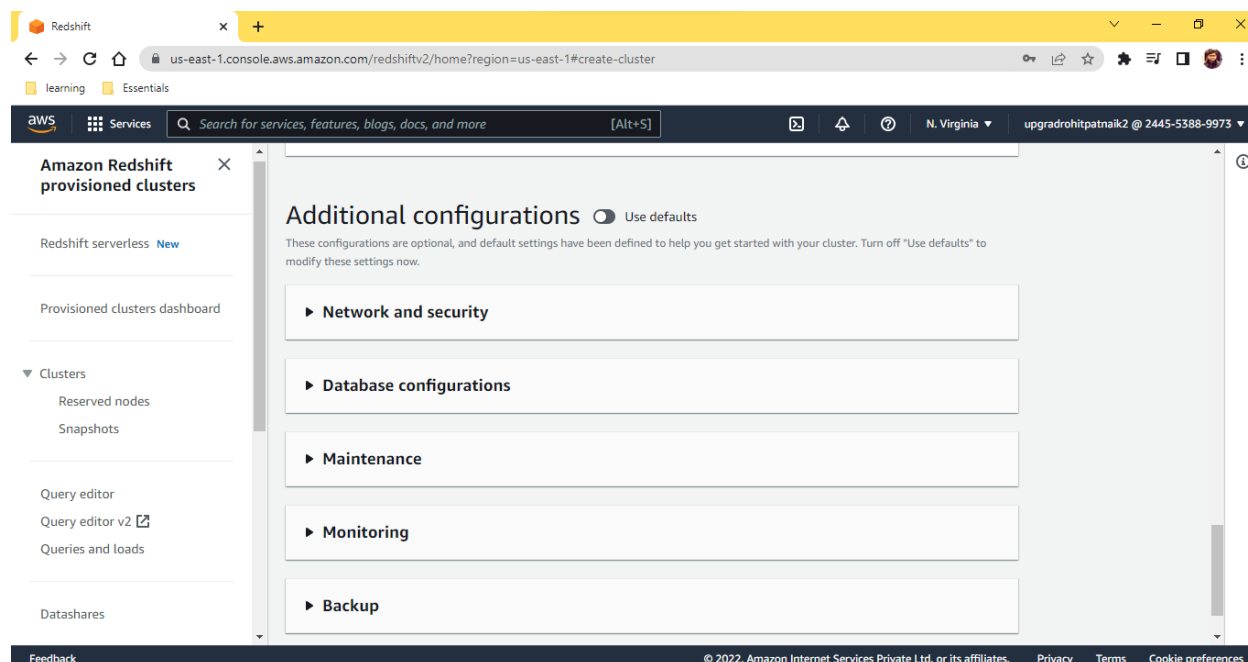
☐ **Auto generate password**
Amazon Redshift can generate a password for you, or you can specify your own password.

Admin user password

☐ **Show password**
Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except "/", "", or "@".

Cluster permissions

Change the additional configuration from default and change in the network and security, database configurations



Amazon Redshift provisioned clusters

Redshift serverless **New**

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

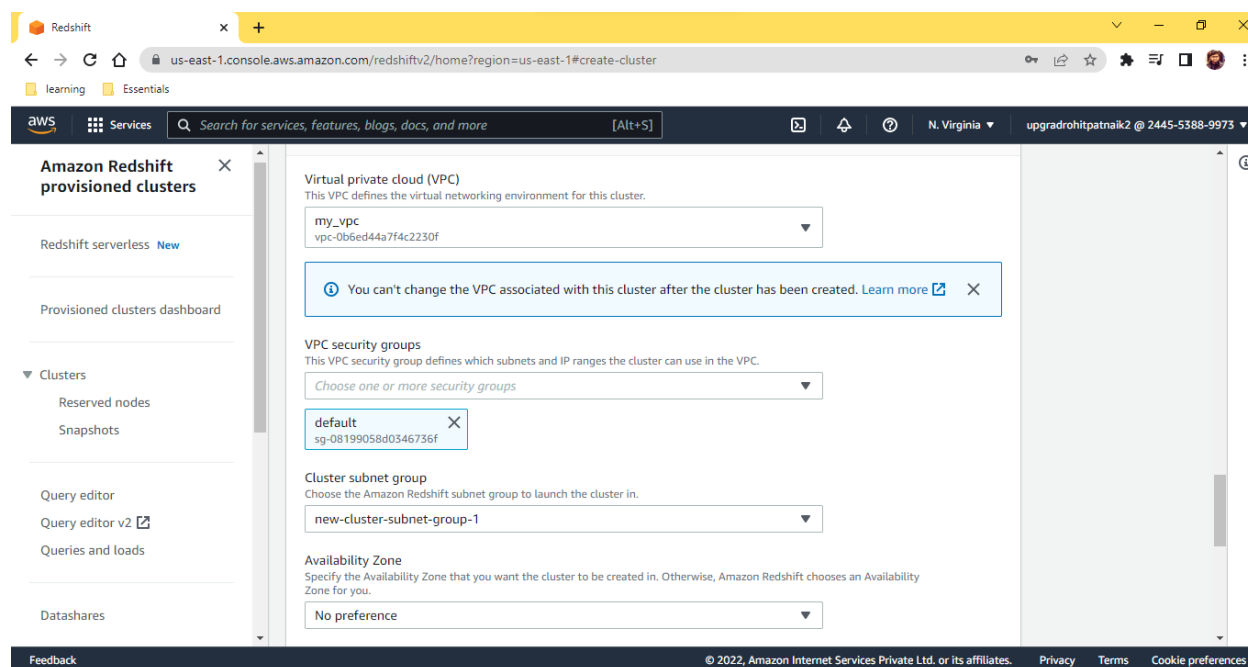
Datashares

Additional configurations ☐ Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

- Network and security
- Database configurations
- Maintenance
- Monitoring
- Backup

Change in the network and security configuration



Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

my_vpc
vpc-0b6ed44a7f4c2230f

You can't change the VPC associated with this cluster after the cluster has been created. [Learn more](#)

VPC security groups

This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.

Choose one or more security groups

default
sg-08199058d0346736f

Cluster subnet group

Choose the Amazon Redshift subnet group to launch the cluster in.

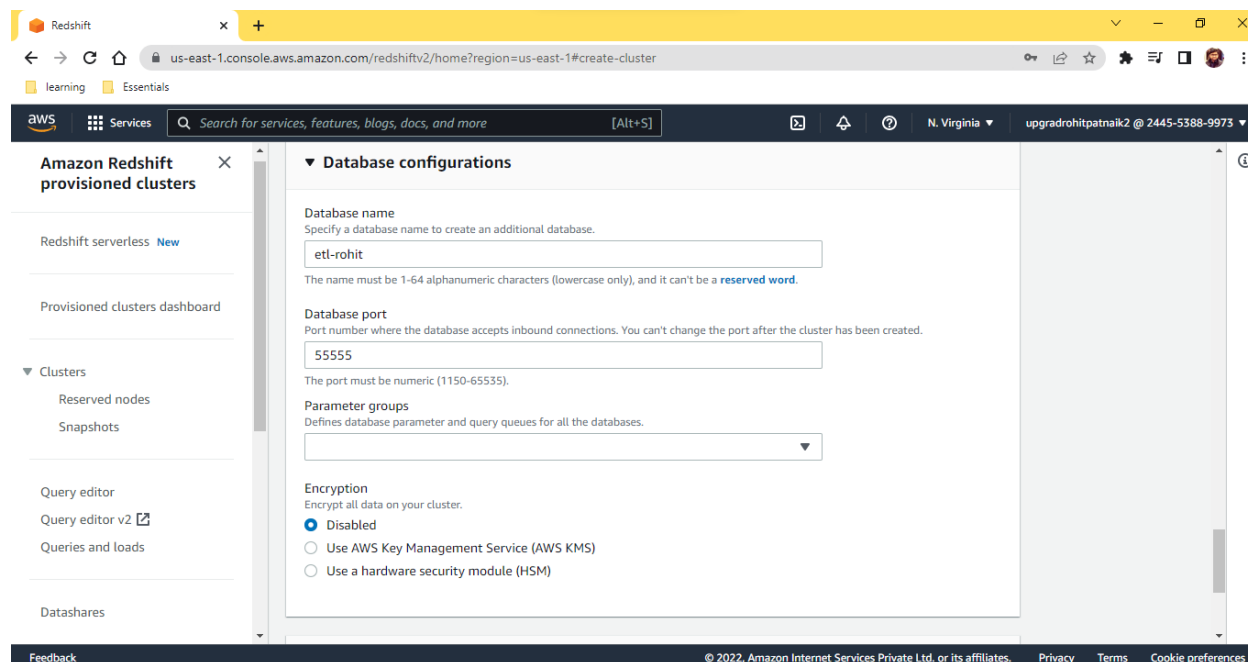
new-cluster-subnet-group-1

Availability Zone

Specify the Availability Zone that you want the cluster to be created in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

No preference

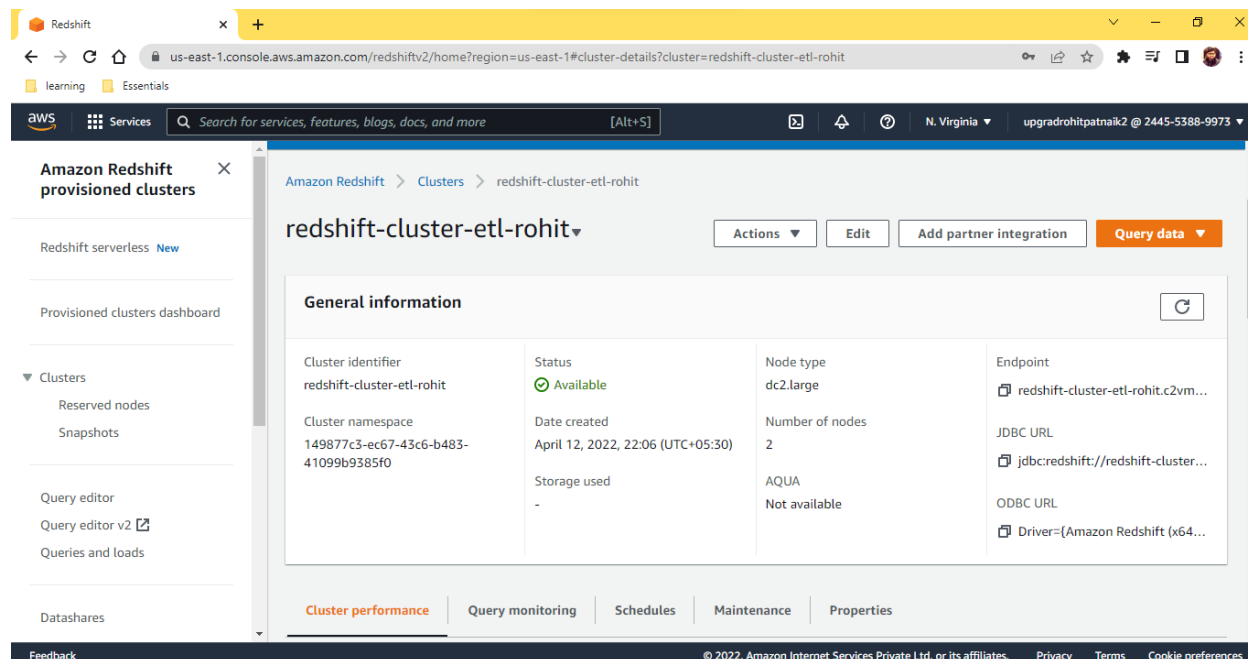
Change in the database configuration



The screenshot shows the 'Database configurations' page in the Amazon Redshift console. The page is titled 'Database configurations' and contains several sections:

- Database name:** A text input field containing 'etl-rohit'. Below it, a note states: 'The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word.'
- Database port:** A text input field containing '5555'. Below it, a note states: 'The port must be numeric (1150-65535).'
- Parameter groups:** A dropdown menu showing a list of parameter groups.
- Encryption:** A section with the heading 'Encrypt all data on your cluster.' and three radio buttons:
 - ☒ Disabled
 - ☐ Use AWS Key Management Service (AWS KMS)
 - ☐ Use a hardware security module (HSM)

Final creation of redshift cluster



The screenshot shows the 'redshift-cluster-etl-rohit' cluster details page in the Amazon Redshift console. The page is titled 'redshift-cluster-etl-rohit' and contains a 'General information' section with a table of cluster details:

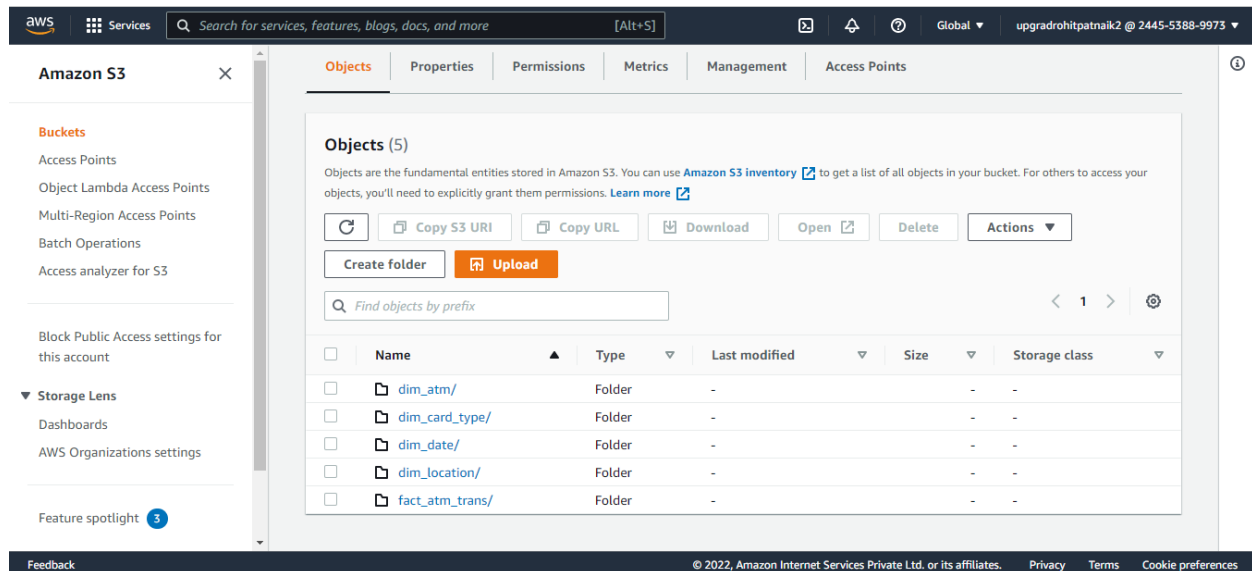
Cluster identifier	Status	Node type	Endpoint
redshift-cluster-etl-rohit	Available	dc2.large	redshift-cluster-etl-rohit.c2vm...
Cluster namespace	Date created	Number of nodes	JDBC URL
149877c3-ec67-43c6-b483-41099b9385f0	April 12, 2022, 22:06 (UTC+05:30)	2	jdbc:redshift://redshift-cluster...
	Storage used	AQUA	ODBC URL
	-	Not available	Driver={Amazon Redshift (x64...

Below the table, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'.

Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

Viewing all the data in the Amazon S3 bucket



Amazon S3

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

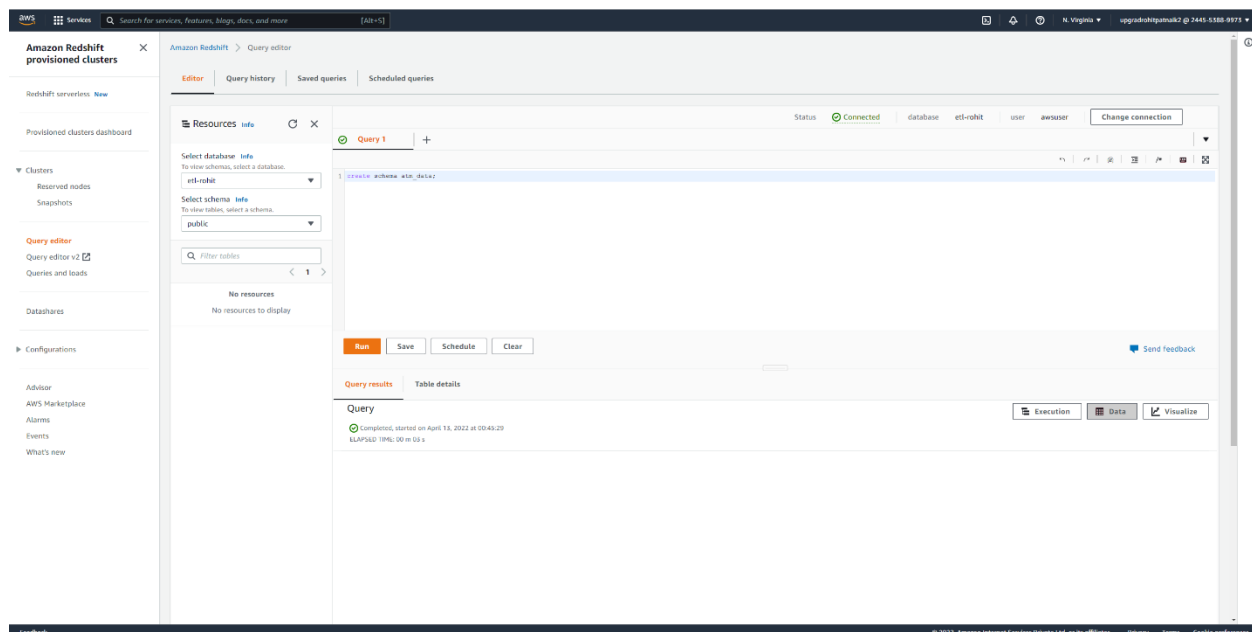
Copy S3 URI Copy URL Download Open Delete Actions

Create folder Upload

Find objects by prefix

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dim_atm/	Folder	-	-	-
<input type="checkbox"/>	dim_card_type/	Folder	-	-	-
<input type="checkbox"/>	dim_date/	Folder	-	-	-
<input type="checkbox"/>	dim_location/	Folder	-	-	-
<input type="checkbox"/>	fact_atm_trans/	Folder	-	-	-

create schema atm_data;



Amazon Redshift

Query editor

Resources info

Select database: info

Select schema: info

Filter tables

No resources

No resources to display

Query 1

```
1 create schema atm_data;
```

Run Save Schedule Clear

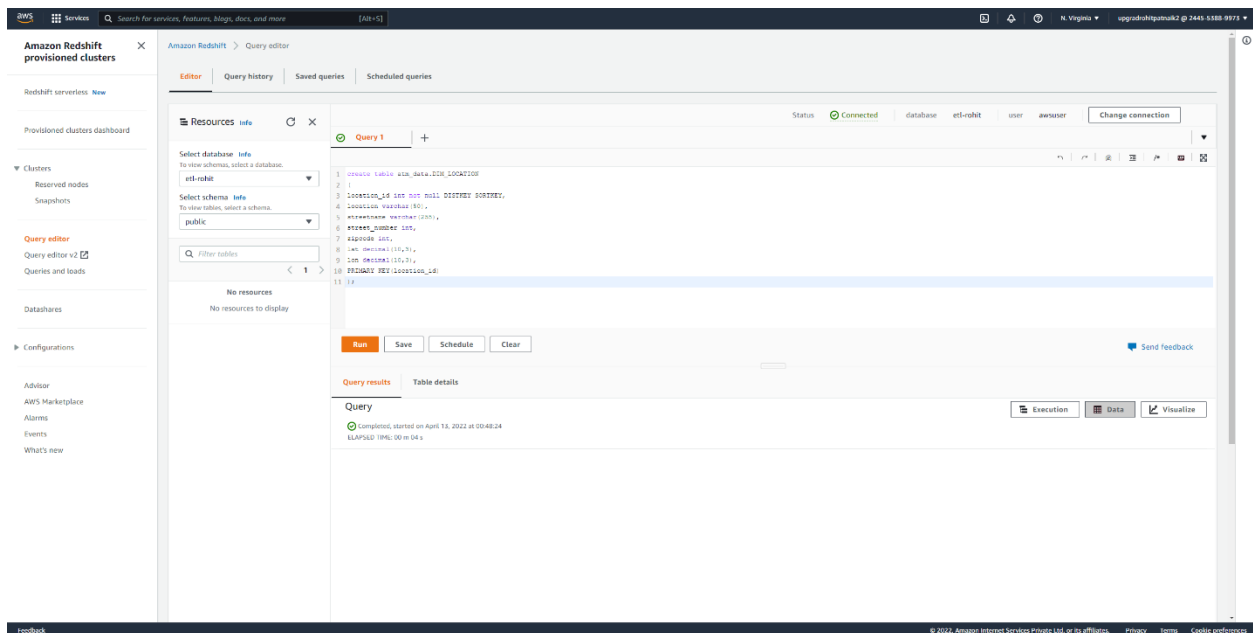
Query results Table details

Query

Completed, started on April 13, 2022 at 00:40:20

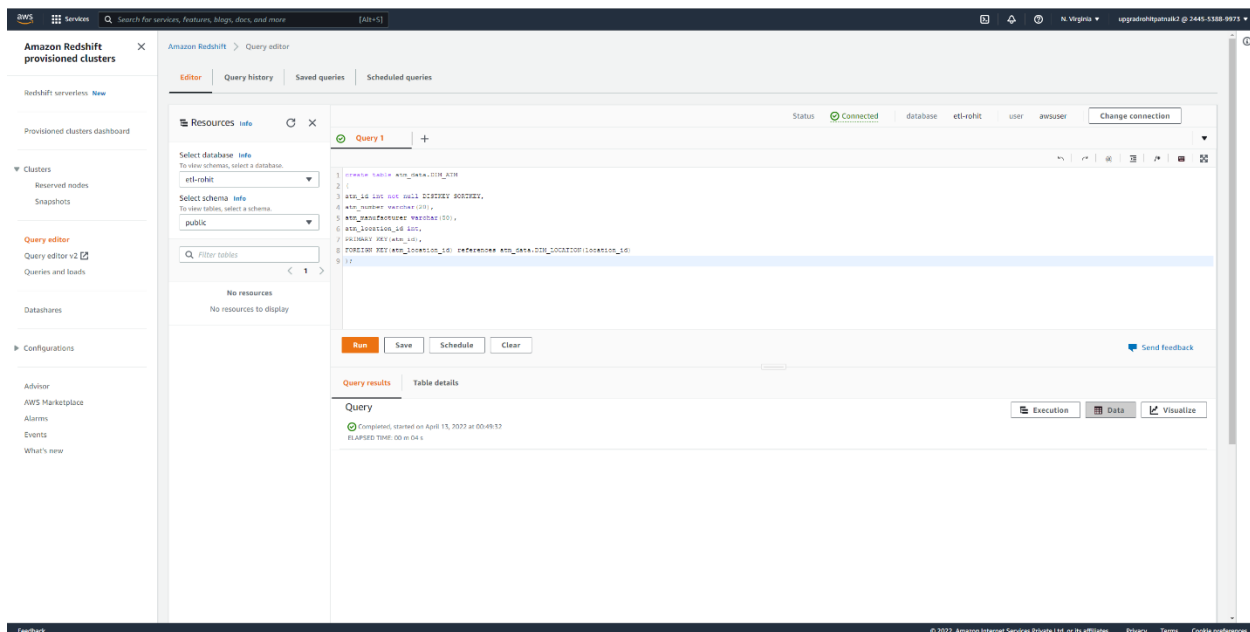
ELAPSED TIME: 00 m 03 s

Create DIM_LOCATION dimension table
create table atm_data.DIM_LOCATION
(
location_id int not null DISTKEY SORTKEY,
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
PRIMARY KEY(location_id)
);



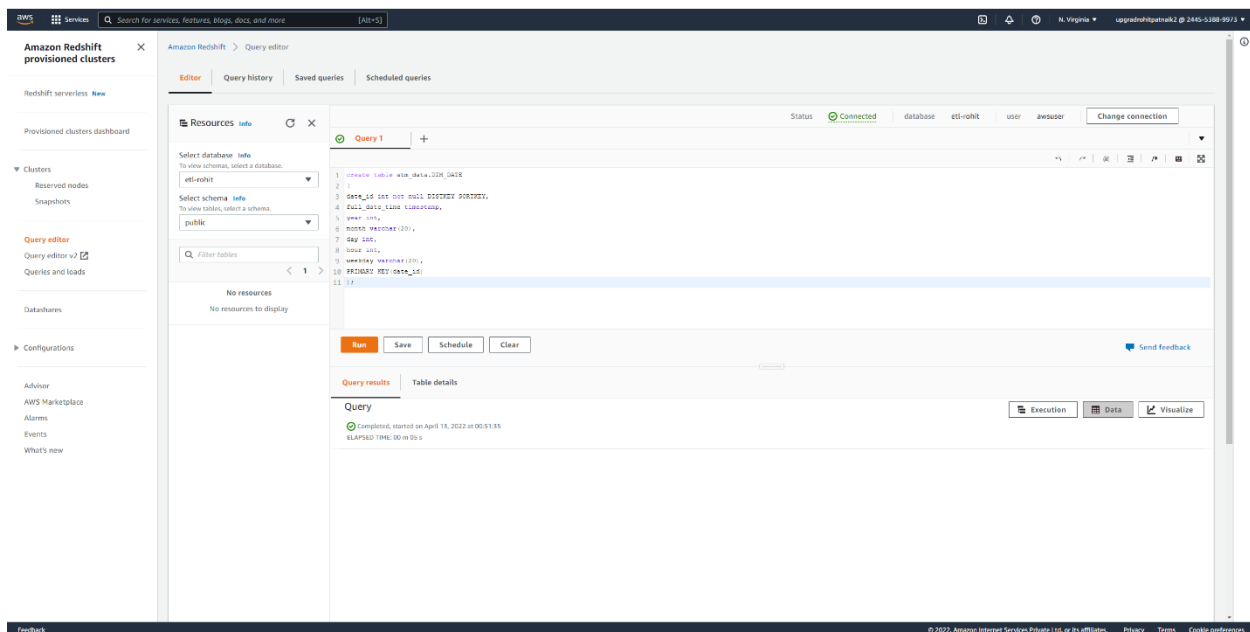
Create DIM_ATM Dimension table

```
create table atm_data.DIM_ATM
(  
  atm_id int not null DISTKEY SORTKEY,  
  atm_number varchar(20),  
  atm_manufacturer varchar(50),  
  atm_location_id int,  
  PRIMARY KEY(atm_id),  
  FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)  
);
```



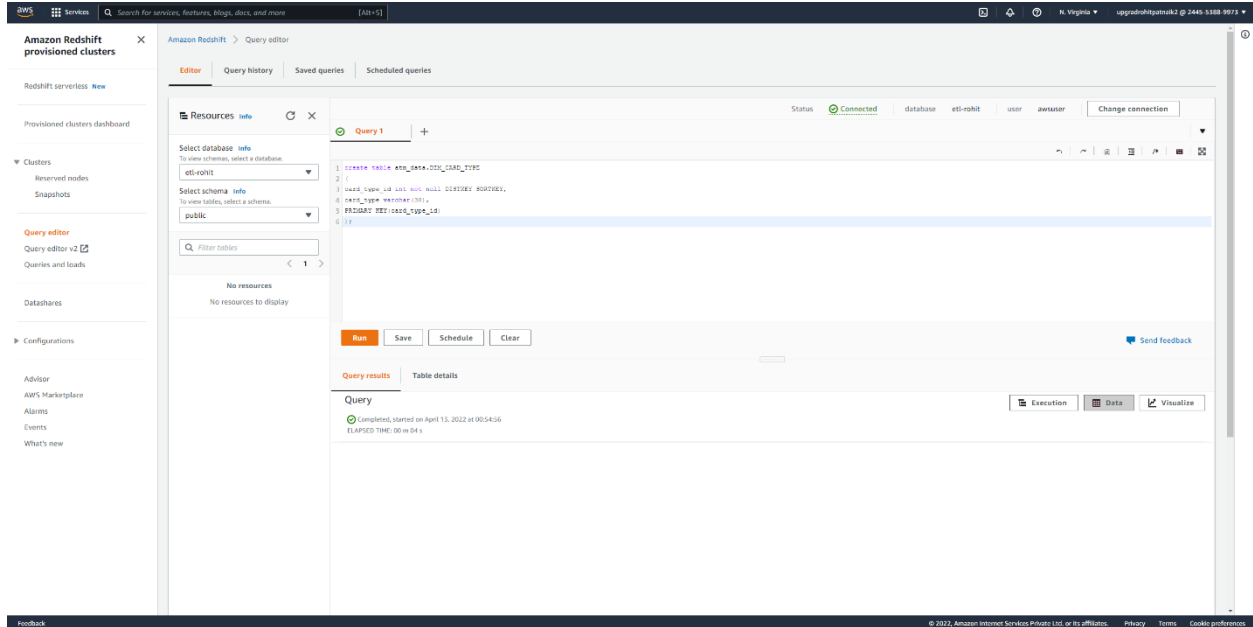
Create DIM_DATE Dimension table

```
create table atm_data.DIM_DATE  
(  
  date_id int not null DISTKEY SORTKEY,  
  full_date_time timestamp,  
  year int,  
  month varchar(20),  
  day int,  
  hour int,  
  weekday varchar(20),  
  PRIMARY KEY(date_id)  
);
```



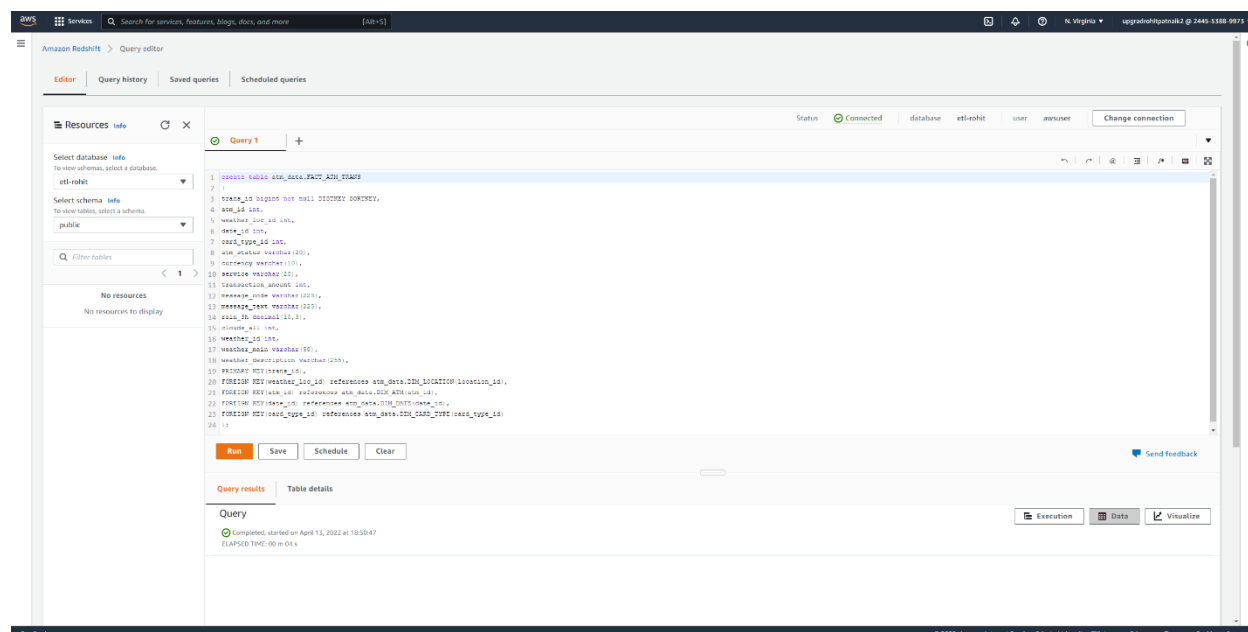
Create DIM_CARD_TYPE Dimension table

```
create table atm_data.DIM_CARD_TYPE  
(  
  card_type_id int not null DISTKEY SORTKEY,  
  card_type varchar(30),  
  PRIMARY KEY(card_type_id)  
);
```



Create FACT_ATM_TRANS Fact table

```
create table atm_data.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount int,
message_code varchar(225),
message_text varchar(225),
rain_3h decimal(10,3),
clouds_all int,
weather_id int,
weather_main varchar(50),
weather_description varchar(255),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
);
```



Loading data into a Redshift cluster from Amazon S3 bucket

Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

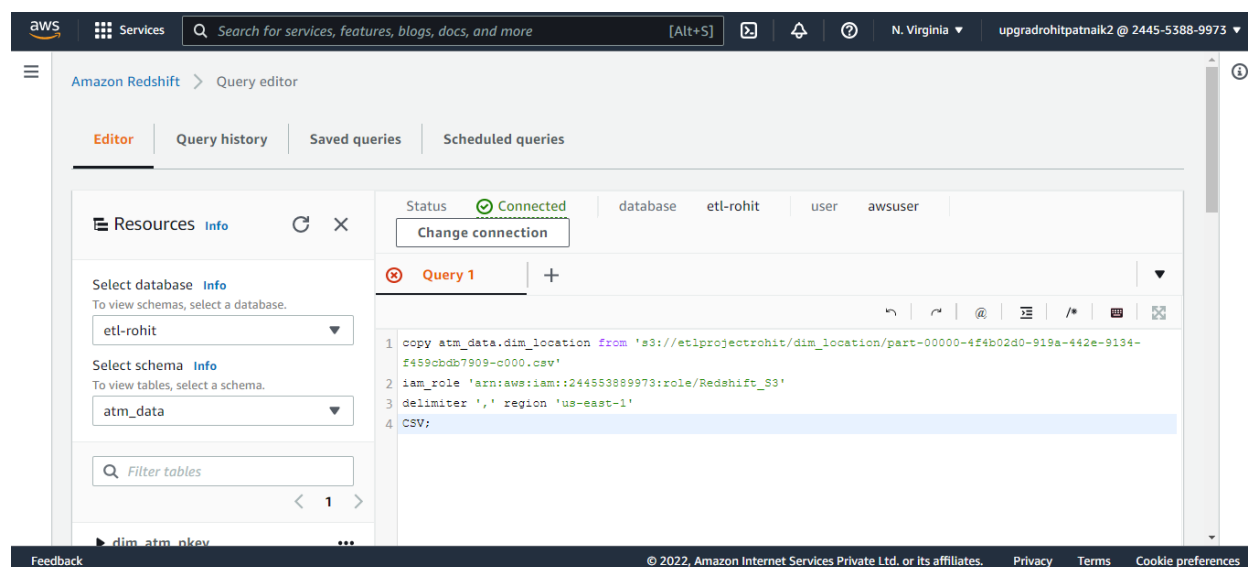
Copying the data to dim_location table

copy atm_data.dim_location from 's3://etlprojectrohit/dim_location/part-00000-4f4b02d0-919a-442e-9134-f459cbdb7909-c000.csv'

iam_role 'arn:aws:iam::244553889973:role/Redshift_S3'

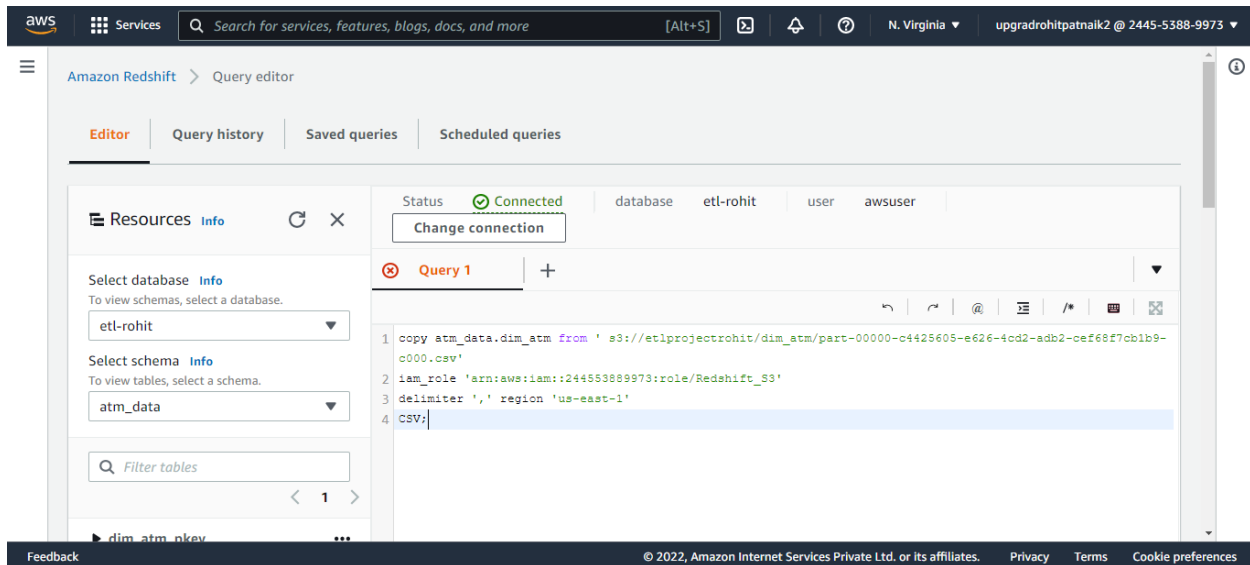
delimiter ',' region 'us-east-1'

CSV;



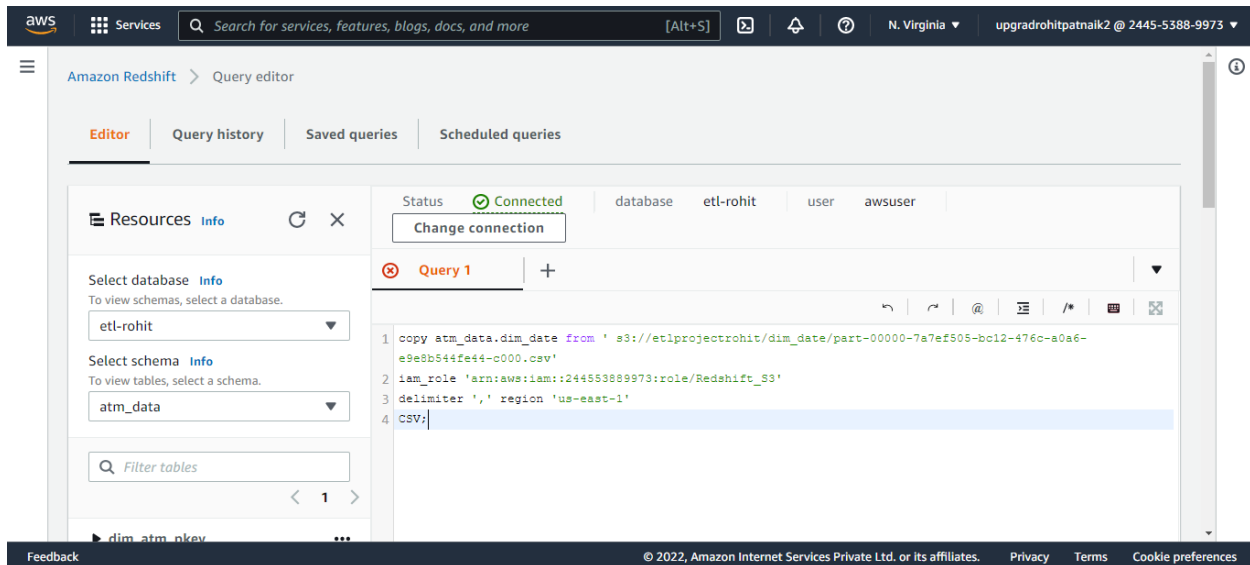
Copying the data to dim_atm table

```
copy atm_data.dim_atm from ' s3://etlprojectrohit/dim_atm/part-00000-c4425605-e626-4cd2-  
adb2-cef68f7cb1b9-c000.csv'  
iam_role 'arn:aws:iam::244553889973:role/Redshift_S3'  
delimiter ',' region 'us-east-1'  
CSV;
```



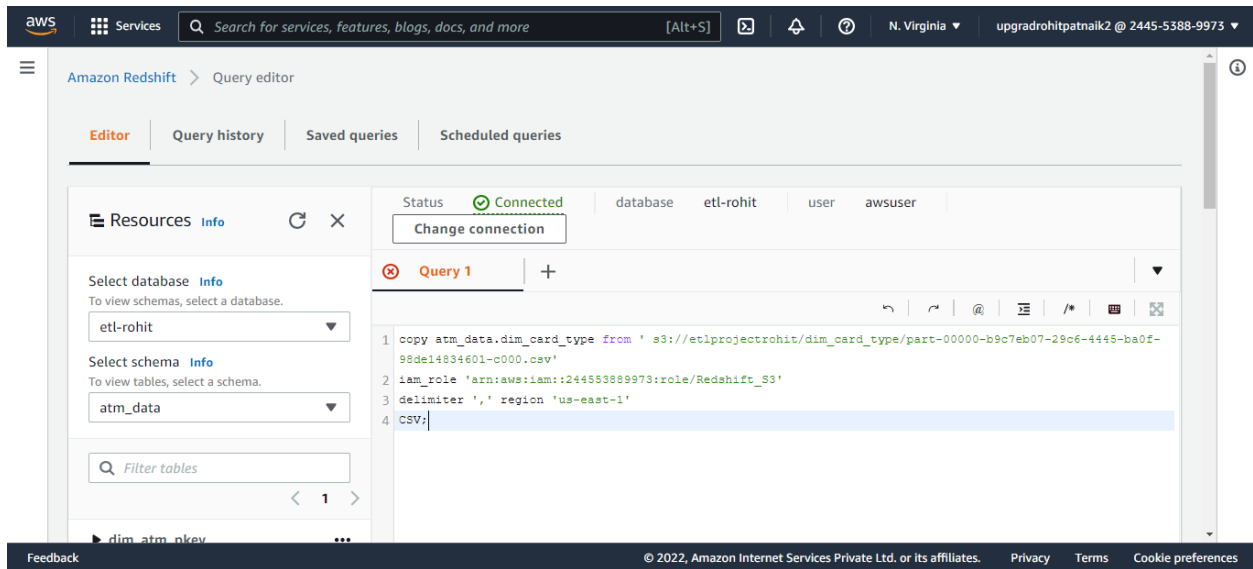
Copying the data to dim_date table

```
copy atm_data.dim_date from ' s3://etlprojectrohit/dim_date/part-00000-7a7ef505-bc12-476c-
a0a6-e9e8b544fe44-c000.csv'
iam_role 'arn:aws:iam::244553889973:role/Redshift_S3'
delimiter ',' region 'us-east-1'
CSV;
```



Copying the data to dim_card_type table

```
copy atm_data.dim_card_type from ' s3://etlprojectrohit/dim_card_type/part-00000-b9c7eb07-29c6-4445-ba0f-98de14834601-c000.csv'  
iam_role 'arn:aws:iam::244553889973:role/Redshift_S3'  
delimiter ',' region 'us-east-1'  
CSV;
```



Copying the data to fact_atm_trans table

```
copy atm_data.fact_atm_trans from ' s3://etlprojectrohit/fact_atm_trans/part-00000-978dd709-2ef2-4145-8ab5-9981558a8c60-c000.csv'  
iam_role 'arn:aws:iam::244553889973:role/Redshift_S3'  
delimiter ',' region 'us-east-1'  
CSV;
```

