### Creation of a Redshift Cluster

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

A screenshot of a computer

Description automatically generated

### A screenshot of a computer Description automatically generated

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

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

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

**Creating DIM\_LOCATION table schema**

create table etl\_project.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)

);

A screenshot of a computer

Description automatically generated

**Creating DIM\_ATM table schema**

create table etl\_project.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 etl\_project.DIM\_LOCATION(location\_id)

);

A screenshot of a computer

Description automatically generated

**Creating DIM\_DATE table schema**

create table etl\_project.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)

);

A screenshot of a computer

Description automatically generated

**Create DIM\_CARD\_TYPE table schema**

create table etl\_project.DIM\_CARD\_TYPE

(

card\_type\_id int not null DISTKEY SORTKEY,

card\_type varchar(30),

PRIMARY KEY(card\_type\_id)

);

A screenshot of a computer

Description automatically generated

**Creating FACT\_ATM\_TRANS table schema**

create table etl\_project.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),

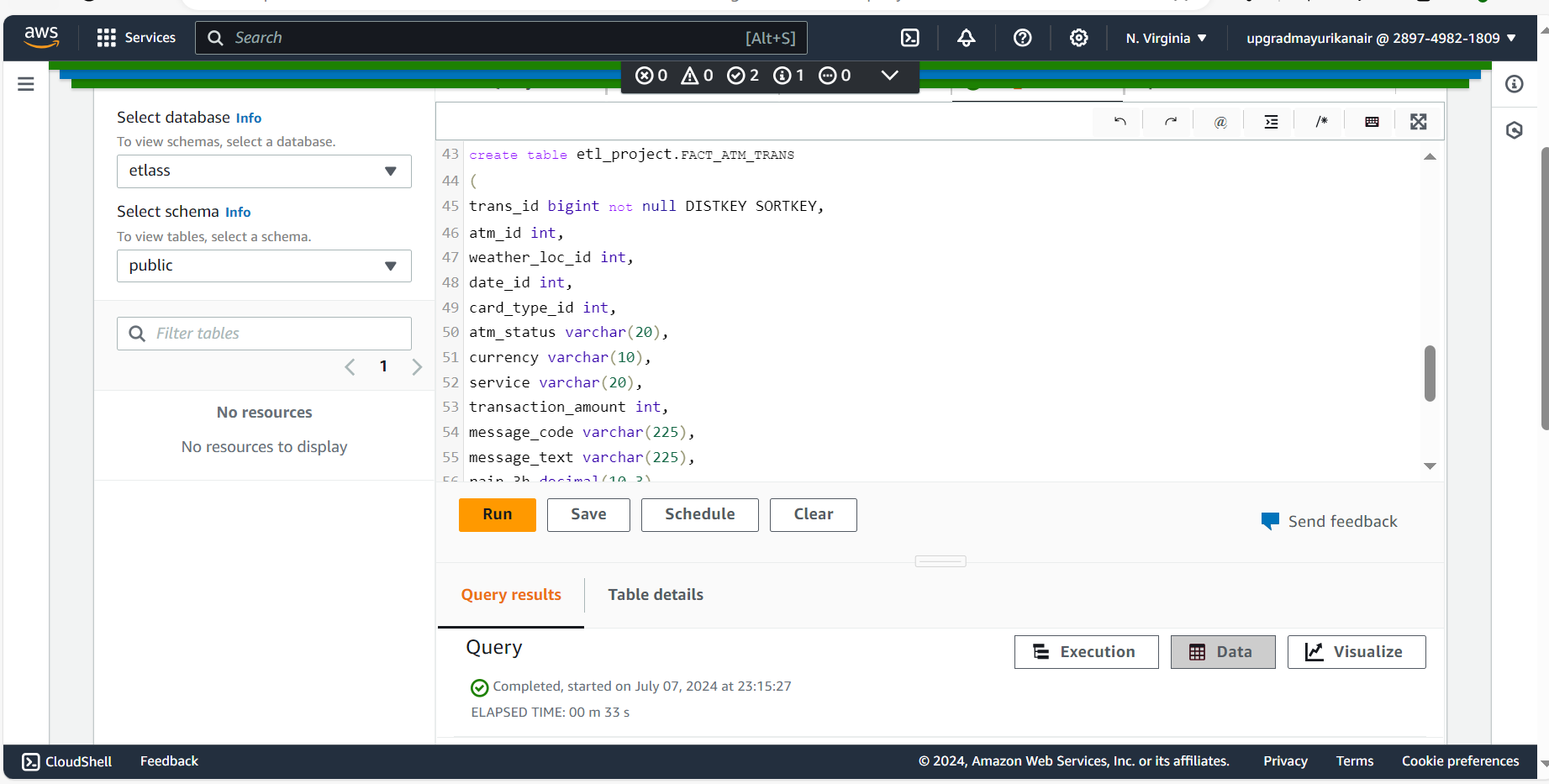
FOREIGN KEY(weather\_loc\_id) references etl\_project.DIM\_LOCATION(location\_id),

FOREIGN KEY(atm\_id) references etl\_project.DIM\_ATM(atm\_id),

FOREIGN KEY(date\_id) references etl\_project.DIM\_DATE(date\_id),

FOREIGN KEY(card\_type\_id) references etl\_project.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**

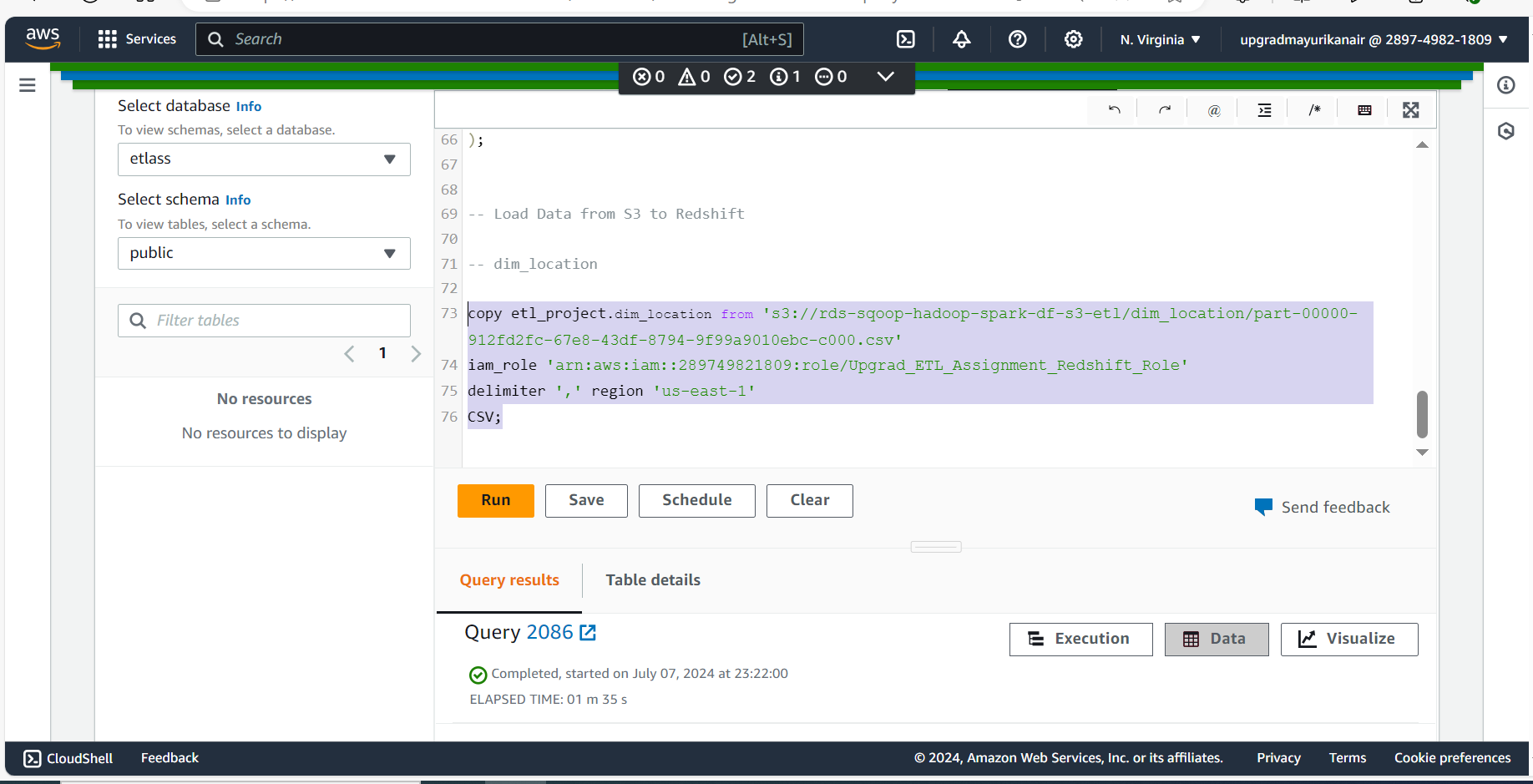
**Populate DIM\_LOCATION**

**copy etl\_project.dim\_location from 's3://rds-sqoop-hadoop-spark-df-s3-etl/dim\_location/part-00000-912fd2fc-67e8-43df-8794-9f99a9010ebc-c000.csv'**

**iam\_role 'arn:aws:iam::289749821809:role/Upgrad\_ETL\_Assignment\_Redshift\_Role'**

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

**CSV;**



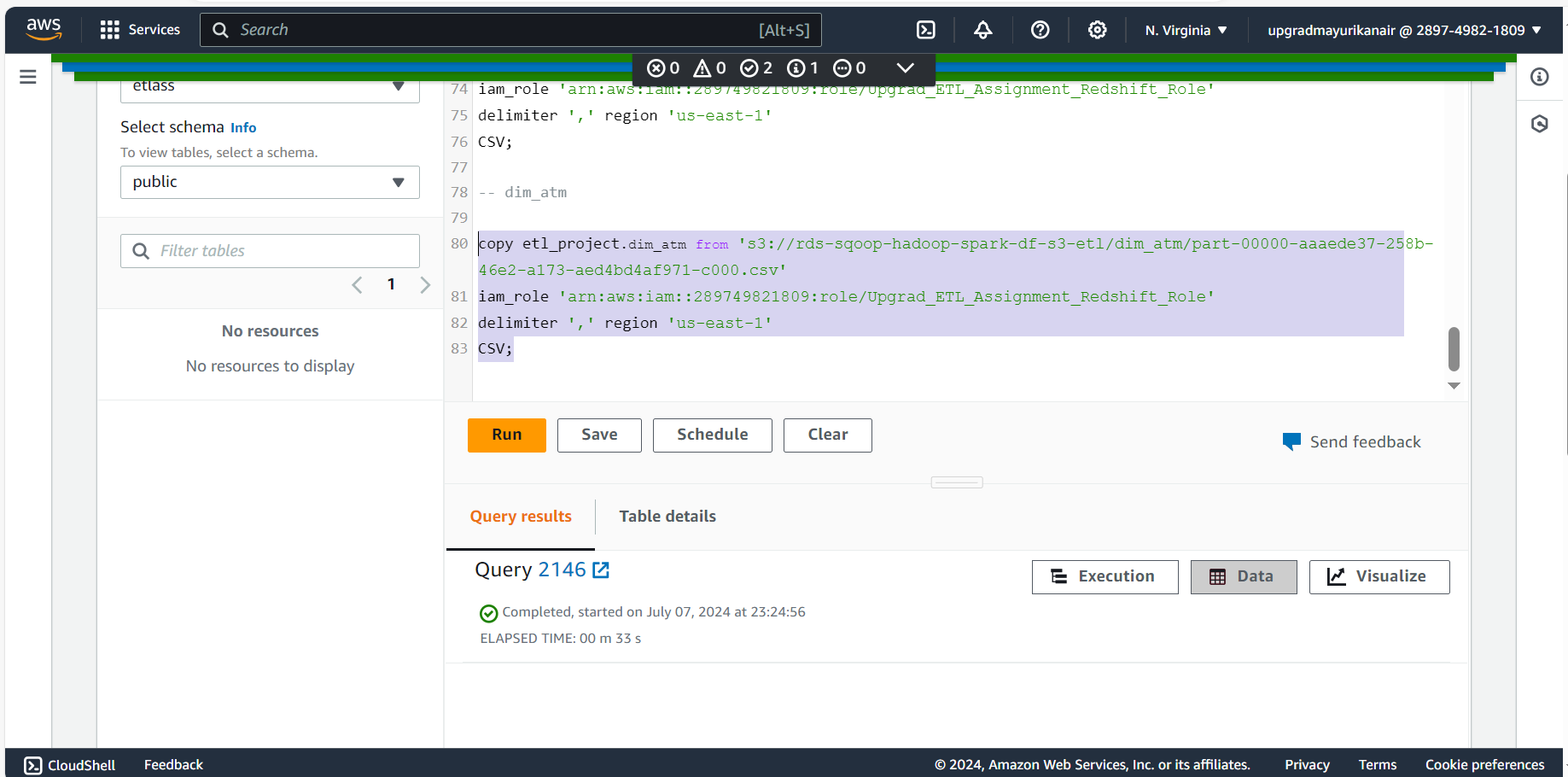
**Populate DIM\_ATM**

**copy etl\_project.dim\_atm from 's3://rds-sqoop-hadoop-spark-df-s3-etl/dim\_atm/part-00000-aaaede37-258b-46e2-a173-aed4bd4af971-c000.csv'**

**iam\_role 'arn:aws:iam::289749821809:role/Upgrad\_ETL\_Assignment\_Redshift\_Role'**

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

**CSV;**



**Populate DIM\_DATE**

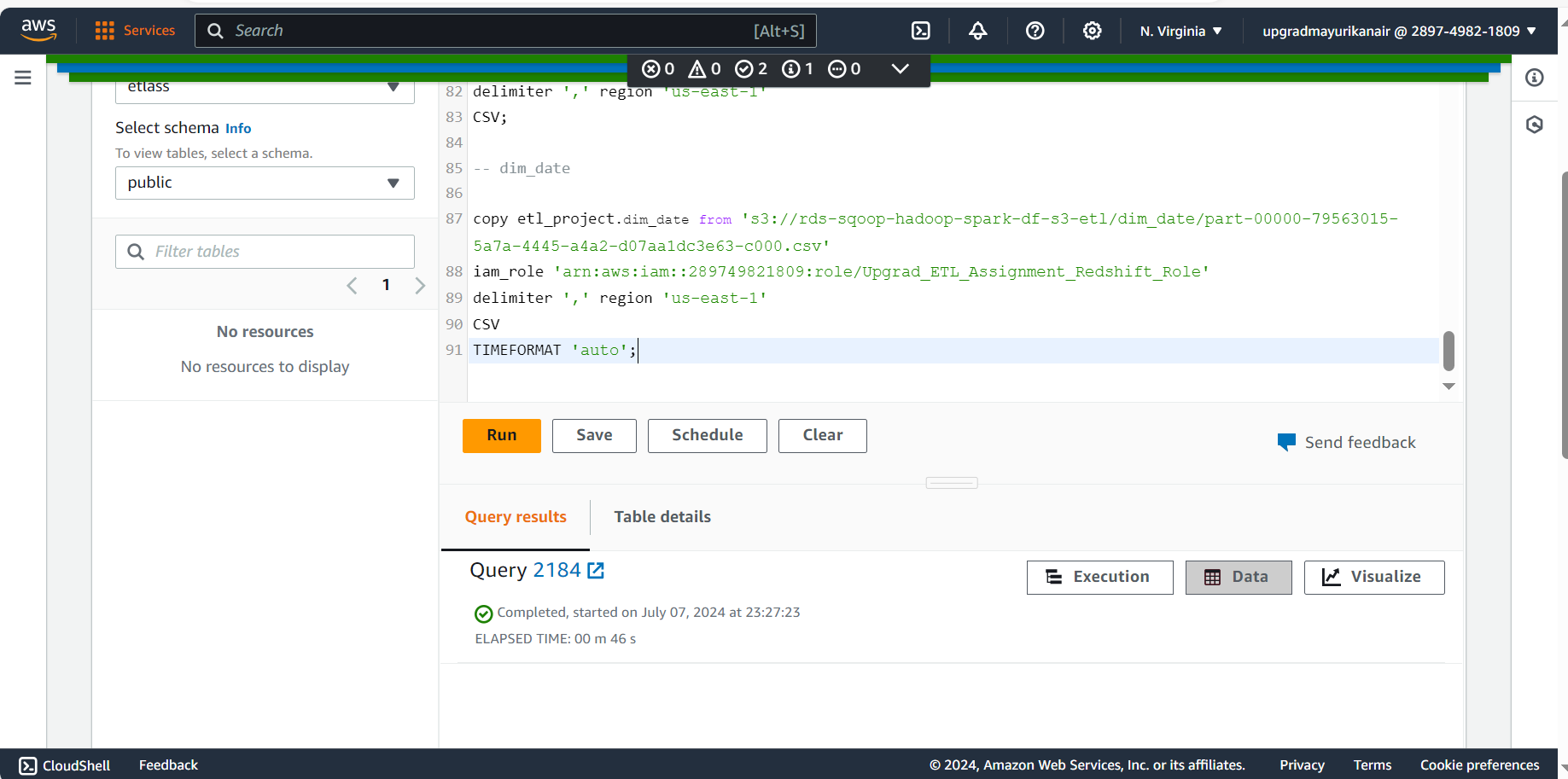
**copy etl\_project.dim\_date from 's3://rds-sqoop-hadoop-spark-df-s3-etl/dim\_date/part-00000-79563015-5a7a-4445-a4a2-d07aa1dc3e63-c000.csv'**

**iam\_role 'arn:aws:iam::289749821809:role/Upgrad\_ETL\_Assignment\_Redshift\_Role'**

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

**CSV**

**TIMEFORMAT 'auto';**



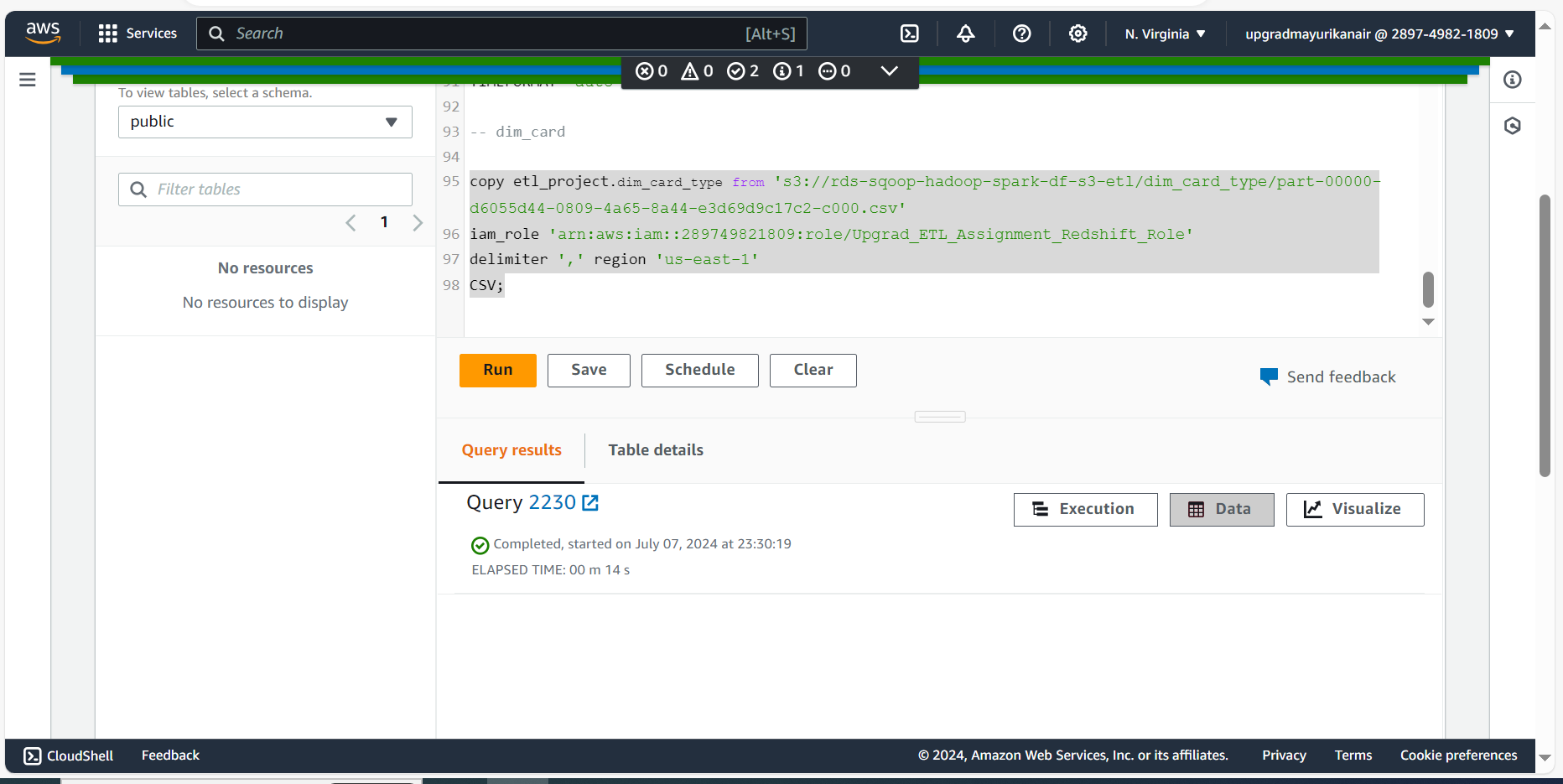
**Populate DIM\_CARD\_TYPE**

**copy etl\_project.dim\_card\_type from 's3://rds-sqoop-hadoop-spark-df-s3-etl/dim\_card\_type/part-00000-d6055d44-0809-4a65-8a44-e3d69d9c17c2-c000.csv'**

**iam\_role 'arn:aws:iam::289749821809:role/Upgrad\_ETL\_Assignment\_Redshift\_Role'**

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

**CSV;**



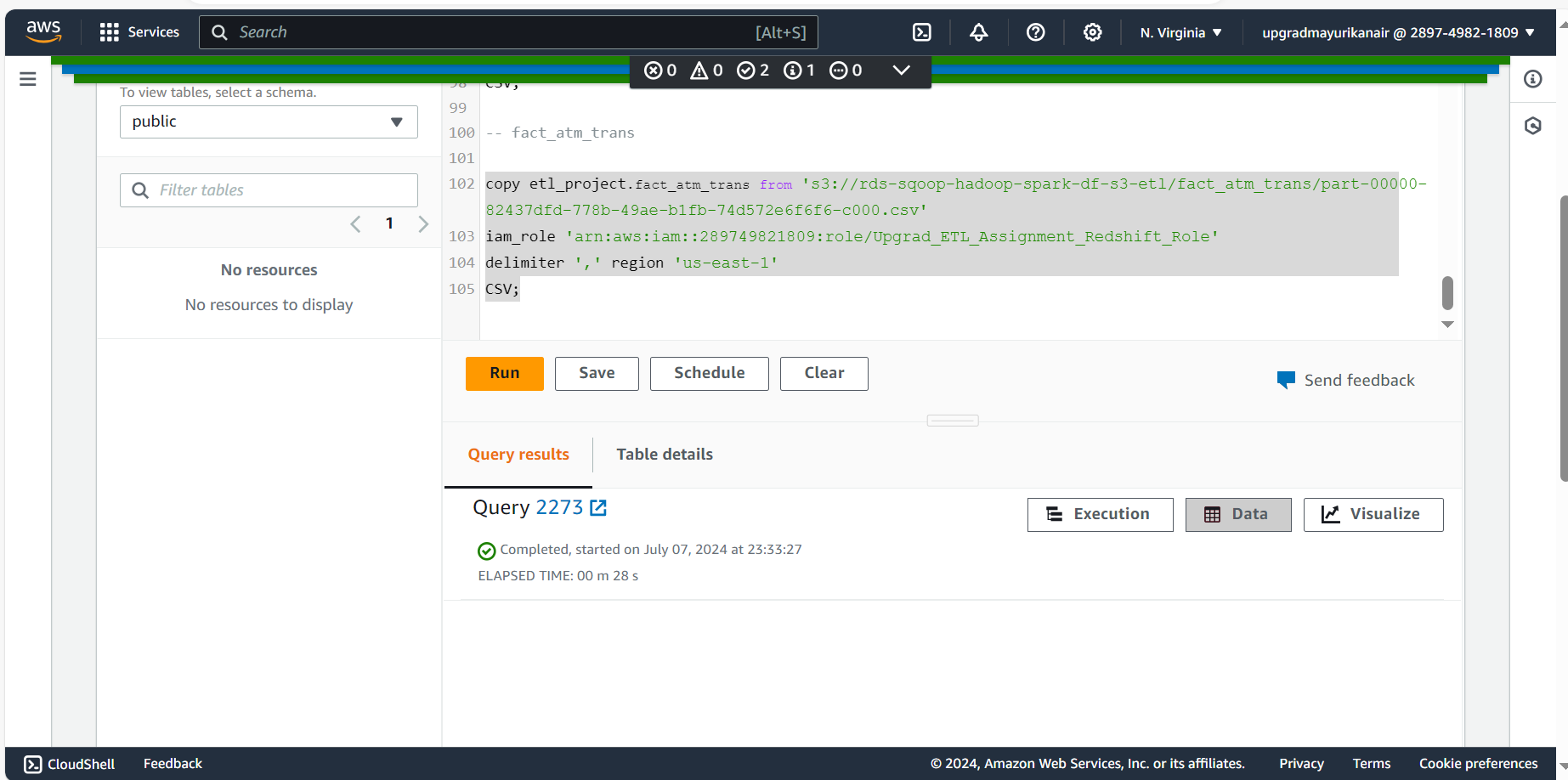
**Populate fact\_atm\_trans**

**copy etl\_project.fact\_atm\_trans from 's3://rds-sqoop-hadoop-spark-df-s3-etl/fact\_atm\_trans/part-00000-82437dfd-778b-49ae-b1fb-74d572e6f6f6-c000.csv'**

**iam\_role 'arn:aws:iam::289749821809:role/Upgrad\_ETL\_Assignment\_Redshift\_Role'**

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

**CSV;**



**OVERALL TABLES**

