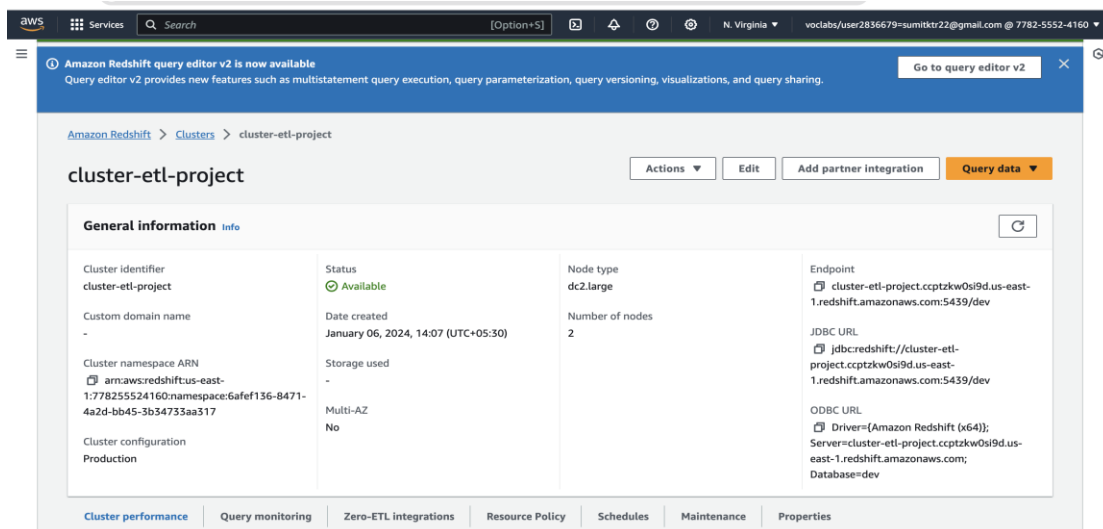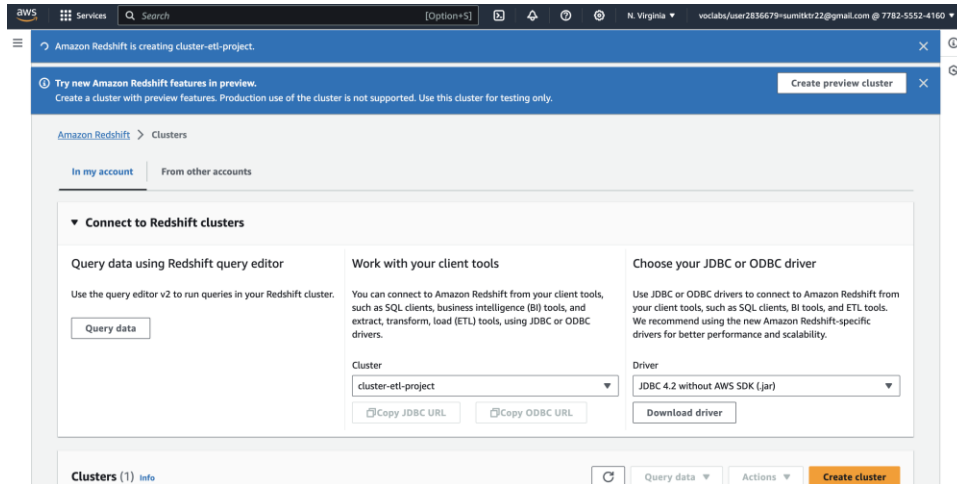# Redshift cluster creation

Screenshots depicting the configuration of the Redshift cluster I've established:

ⓘ **Amazon Redshift query editor v2 is now available**
Query editor v2 provides new features such as multistatement query execution, query parameterization, query versioning, visualizations, and query sharing.

Go to query editor v2 · ✕

Amazon Redshift > Clusters > cluster-etl-project

cluster-etl-project

Actions ▼ · **Edit** · **Add partner integration** · **Query data** ▼

General information Info ⟳

Cluster identifier cluster-etl-project	**Status** ⊘ Available	**Node type** dc2.large	**Endpoint** 🗗 cluster-etl-project.ccptzkw0si9d.us-east-1.redshift.amazonaws.com:5439/dev
Custom domain name -	**Date created** January 06, 2024, 14:07 (UTC+05:30)	**Number of nodes** 2	**JDBC URL** 🗗 jdbc:redshift://cluster-etl-project.ccptzkw0si9d.us-east-1.redshift.amazonaws.com:5439/dev
Cluster namespace ARN 🗗 arn:aws:redshift:us-east-1:778255524160:namespace:6afef136-8471-4a2d-bb45-3b34733aa317	**Storage used** -		**ODBC URL** 🗗 Driver={Amazon Redshift (x64)}; Server=cluster-etl-project.ccptzkw0si9d.us-east-1.redshift.amazonaws.com; Database=dev
Cluster configuration Production	**Multi-AZ** No		

Cluster performance · Query monitoring · Zero-ETL integrations · Resource Policy · Schedules · Maintenance · Properties

---

Production

east-1.redshift.amazonaws.com;
Database=dev

Cluster performance · Query monitoring · Zero-ETL integrations · Resource Policy · Schedules · Maintenance · **Properties**

Database configurations Info

Edit admin credentials · **Rotate encryption keys** · **Edit** ▼

Database name dev	**Parameter group** Defines database parameter and query queues for all the databases. default.redshift-1.0	**Encryption** Disabled	**Audit logging** Disabled
Port 5439		**AWS KMS key ID** -	
Admin user name awsuser	**SSH ingestion setting (cluster public key)** 🗗 ssh-rsa		
Admin password ARN 🗗 arn:aws:secretsmanager:us-east-1:778255524160:secret:redshift!cluster-etl-project-awsuser-q7eHGn			

AAAAB3NzaC1yc2EAAAADAQABAAABAQCE
5CCzUfT+TkhV3Sm61eq9K2Yb+L9di2Gul3I
ncQi0E+XV9F4z+qWmLg5Od1ai3RibmjyEH
5zwVVHpjuVH/0KTvwQsGFfRZQamFNUf1a
saYBoTbk+QT7QBi8cwkoU5QBLD0fRWWnf
eUSvOcOFIgVveSMyahzLlIHIOvKBdWSAXW
aIG/GJYs0ROPtp1W6EXbgJFKJ9rLGncpjbEX
IwyUkJOSEqKLhb+dxsbV8Q+jKa8dUgICU1c
TP8AF0eMdcia64HubXtIMXRZ2Nfzn+YMqq
Q7p3RkXUkgQXd/witkdwxFpxqxqJ56yD5M
VI/qeoyvE2fjj+cz8gNAeVoMmkuni1aH
Amazon-Redshift

Network and security settings Info

Edit

|			
Virtual private cloud (VPC) vpc-0a62c993e28b2b981 🗗	**Availability Zone** us-east-1f	**VPC security group** Specify which instances and devices can connect to the cluster. sg-0c9b387f753de6dc8 🗗	**Publicly accessible** Allow connections from outside the VPC. Disabled
Subnet group default	**Enhanced VPC routing** Disabled	**IP address type** -	
Endpoint URL 🗗 -			

Cluster permissions

ⓘ Create an IAM role as the default for this cluster that has the AmazonRedshiftAllCommandsFullAccess 🗗 policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.
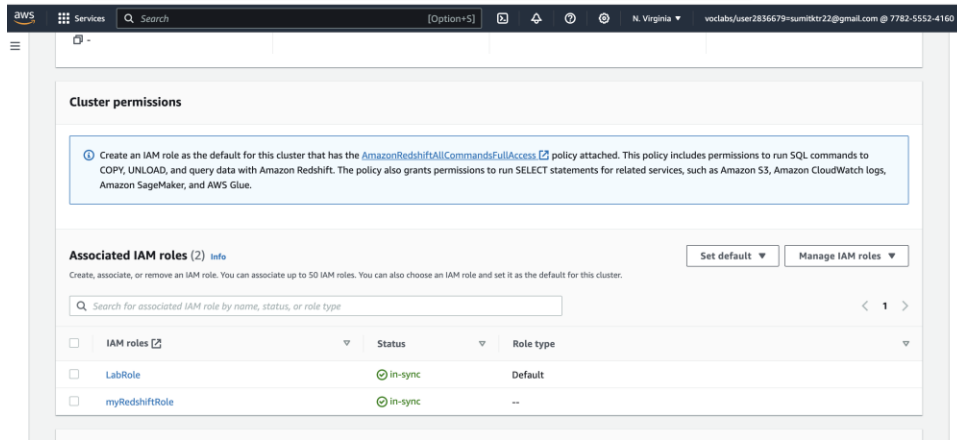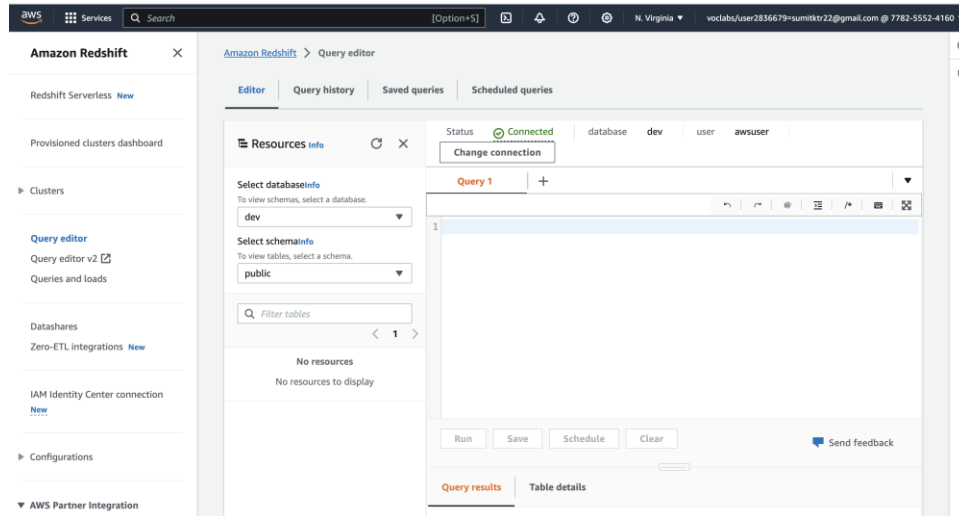
Connection to the database :
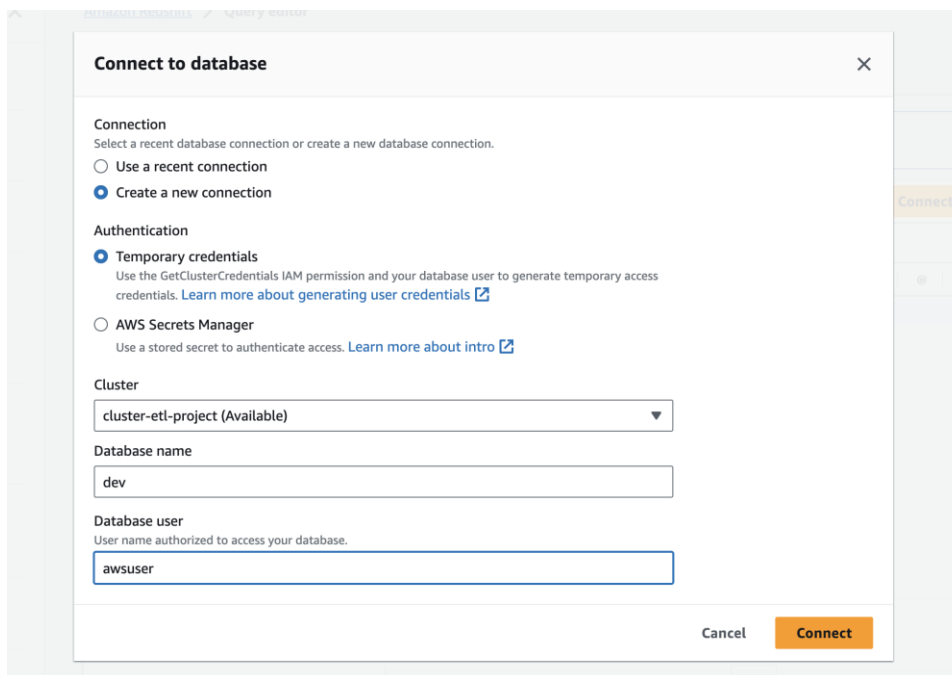
Establishing a database within the Redshift cluster and executing queries to create both dimension and fact tables.

Query for creating schema:
`create schema atm_data;`

Constructing queries for the creation of diverse dimension and fact tables, ensuring proper assignment of primary and foreign keys.

Creating 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)
```
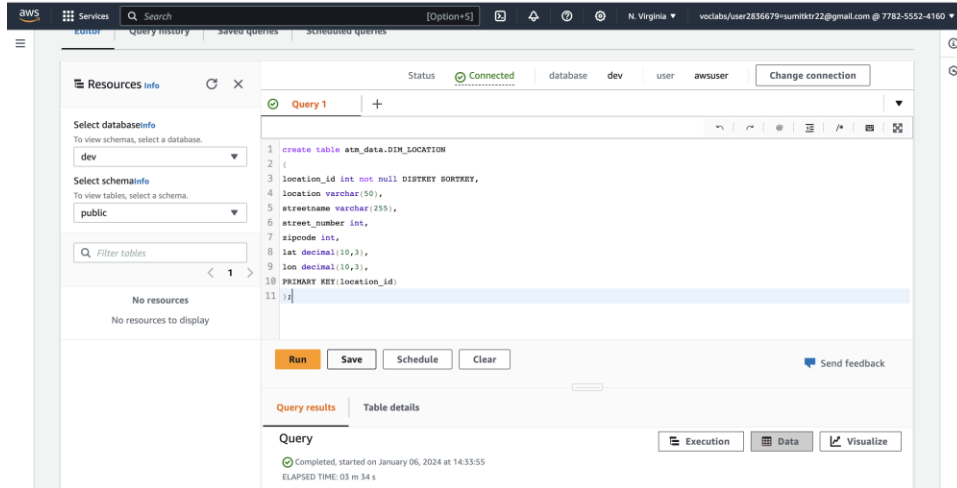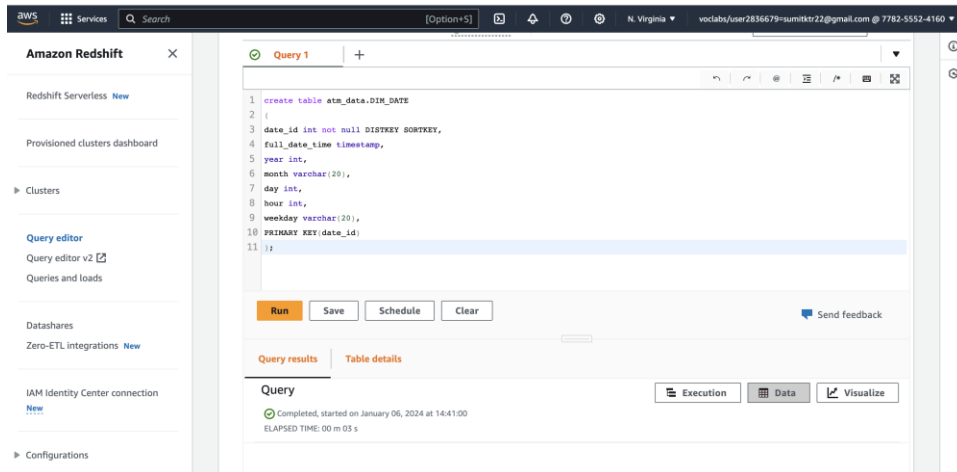
Creating atm dimension table

```
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)
);
```
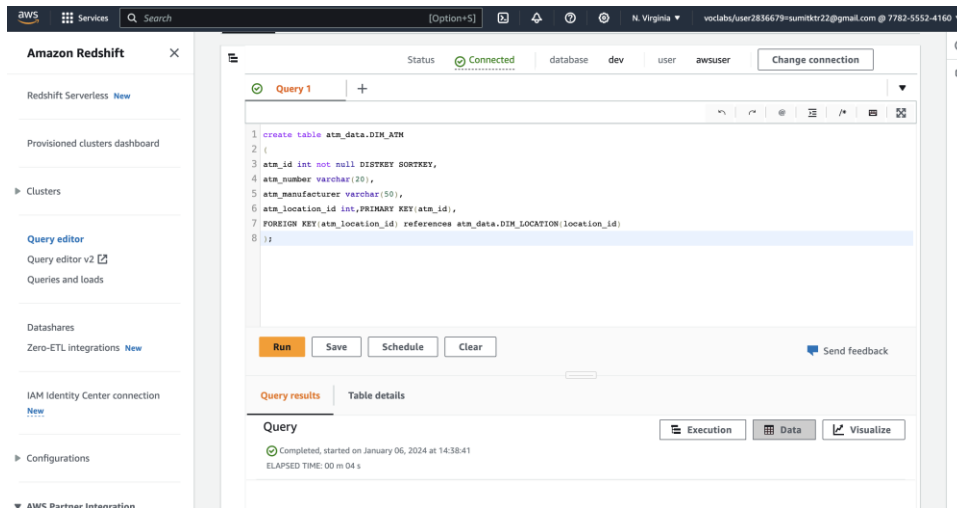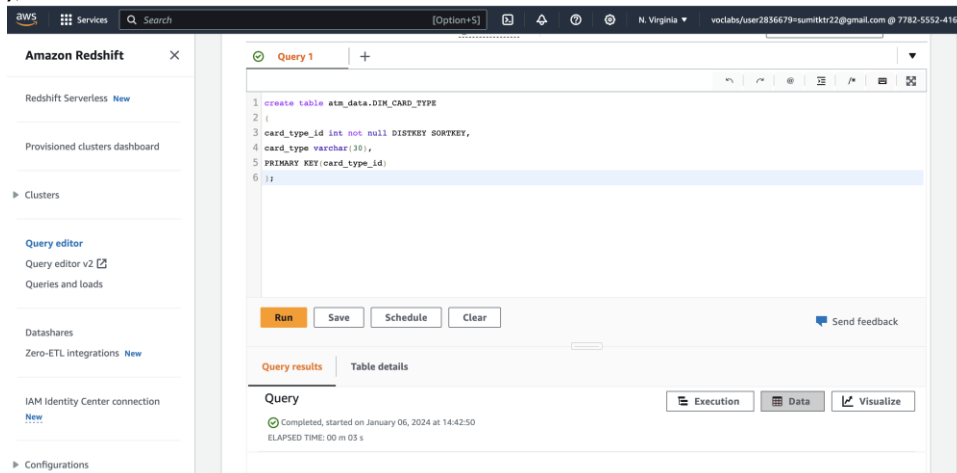
Creating date dimension table

```
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)
);
```


Creating 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)
);
```

Creating atm transactions fact table

```sql
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
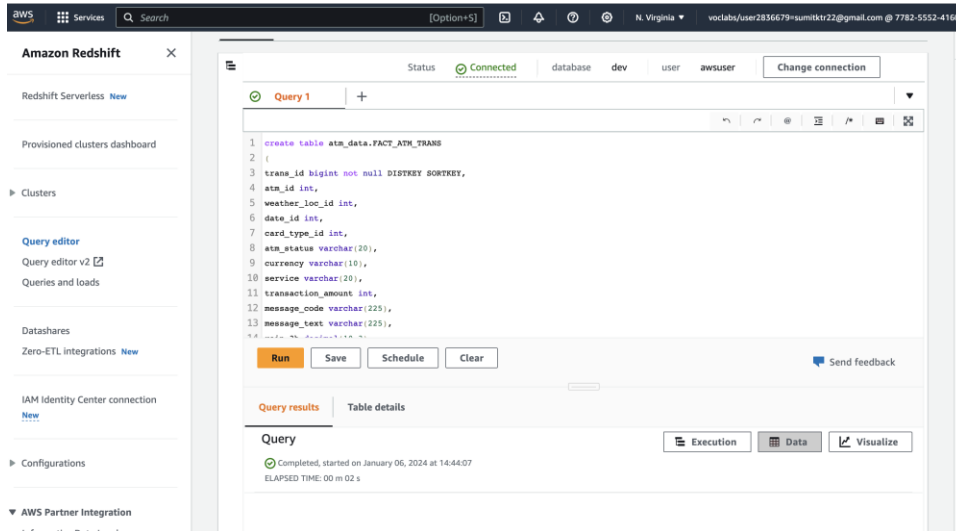
Copying the data to dim_atm table

```
copy atm_data.dim_location from 's3://upgrad-bucket-s3/dim_location/part-00000-
81b80018-e175-4d8e-8d0c-525d4db67034-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
CSV;
```
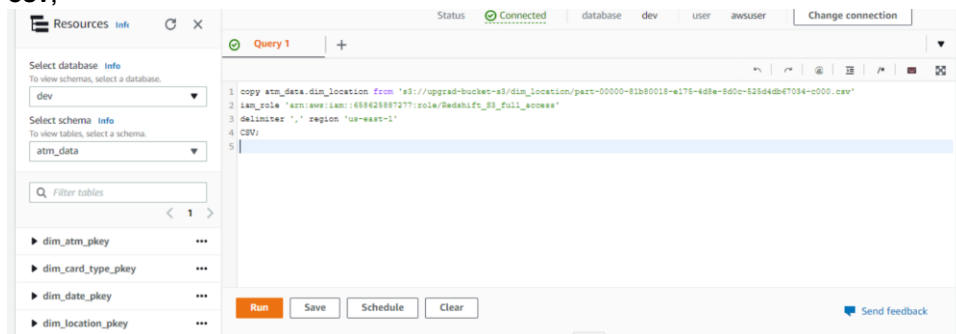

Copying the data to dim_date table

```
copy atm_data.dim_atm from 's3://upgrad-bucket-s3/dim_atm/part-00000-86ed9a65-6e1e-
```
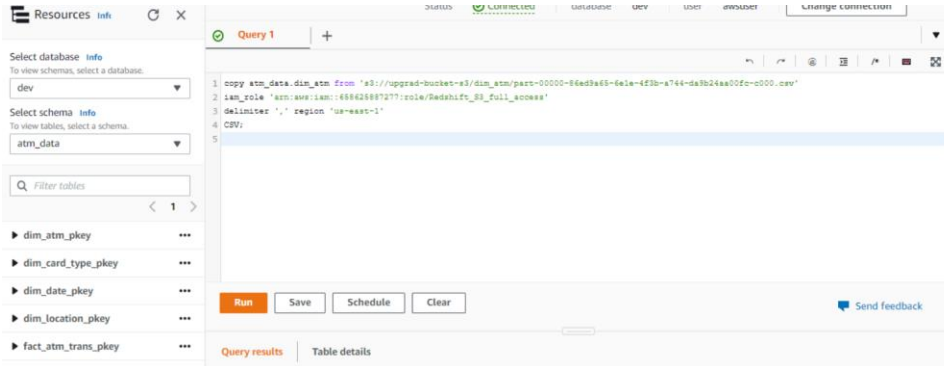
4f3b-a744-da9b24aa00fc-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
CSV;

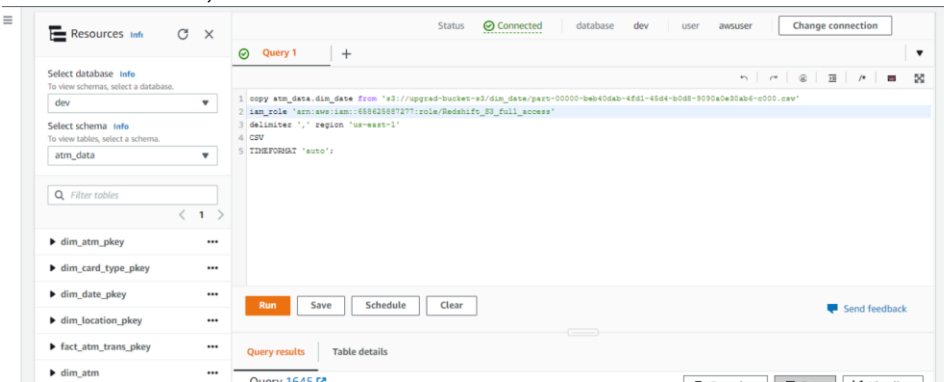copy atm_data.dim_date from 's3://upgrad-bucket-s3/dim_date/part-00000-beb40dab-4fd1-45d4-b0d8-9090a0e30ab6-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
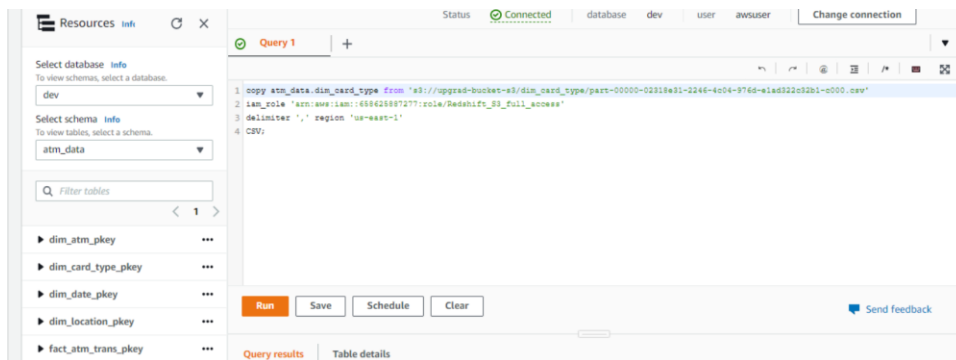CSV
TIMEFORMAT 'auto';

Copying the data to dim_card_type table

copy atm_data.dim_card_type from 's3://upgrad-bucket-s3/dim_card_type/part-00000-02318e31-2246-4c04-976d-e1ad322c32b1-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
CSV;

Copying the data to fact_atm_trans table

copy atm_data.fact_atm_trans from 's3://upgrad-bucket-s3/fact_atm_trans/part-00000-fb72e768-f589-4caa-bd26-1bc6fcbbdb8a-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
CSV;