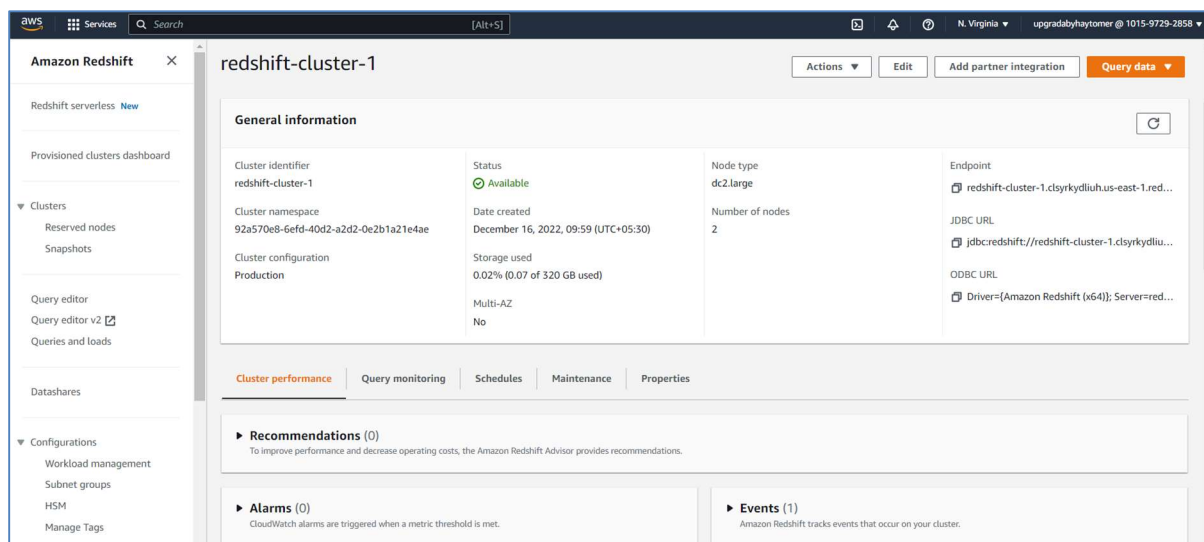
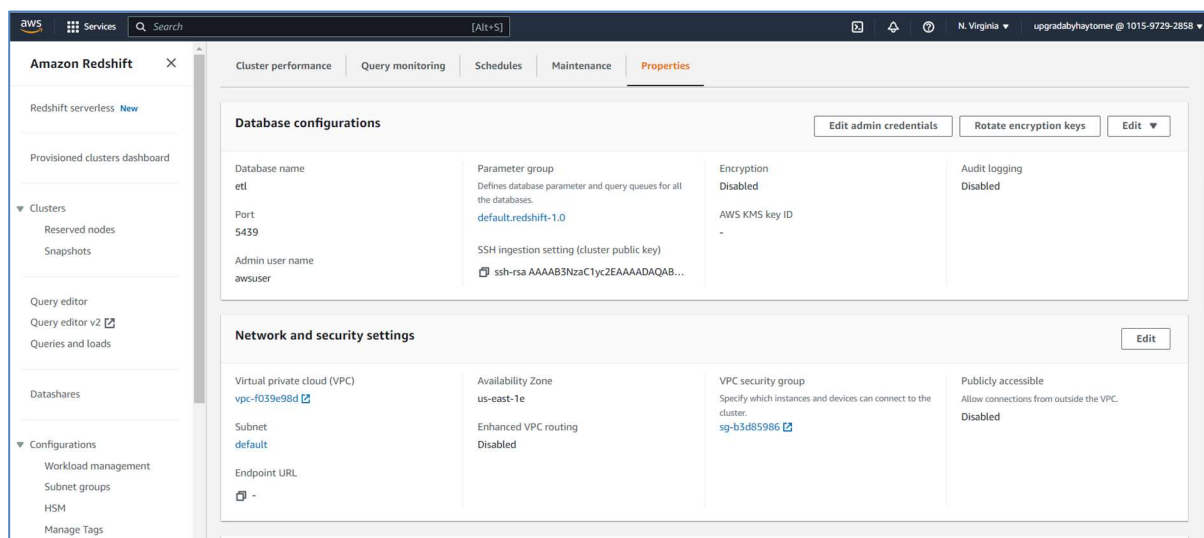


Creation of a Redshift Cluster

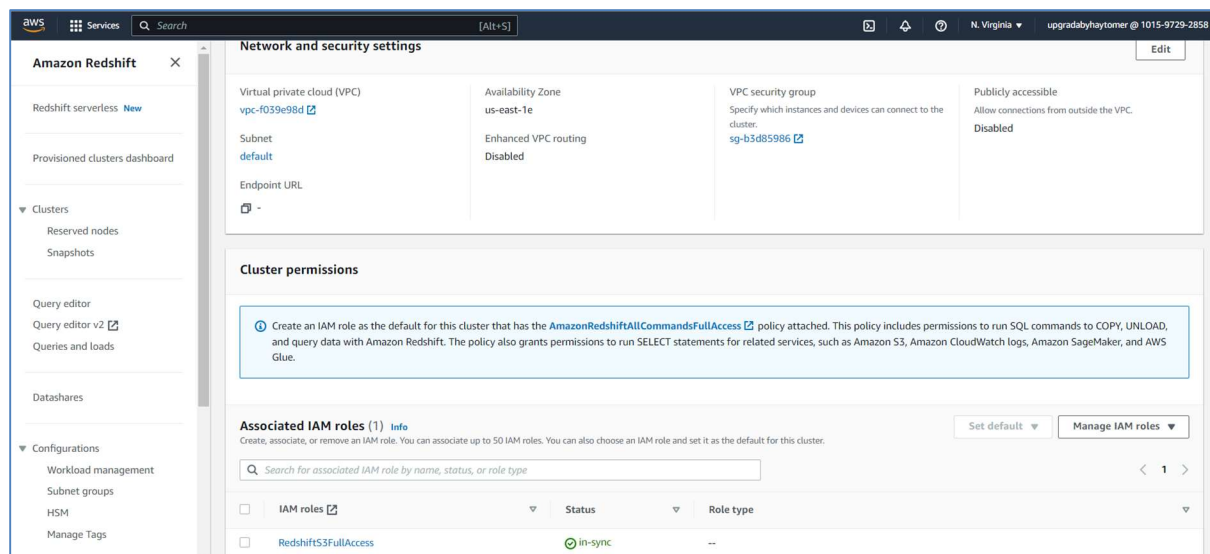
Screenshots of the configuration of the Redshift cluster that we have created:



Screenshot for Cluster Configuration and Network Security Settings:



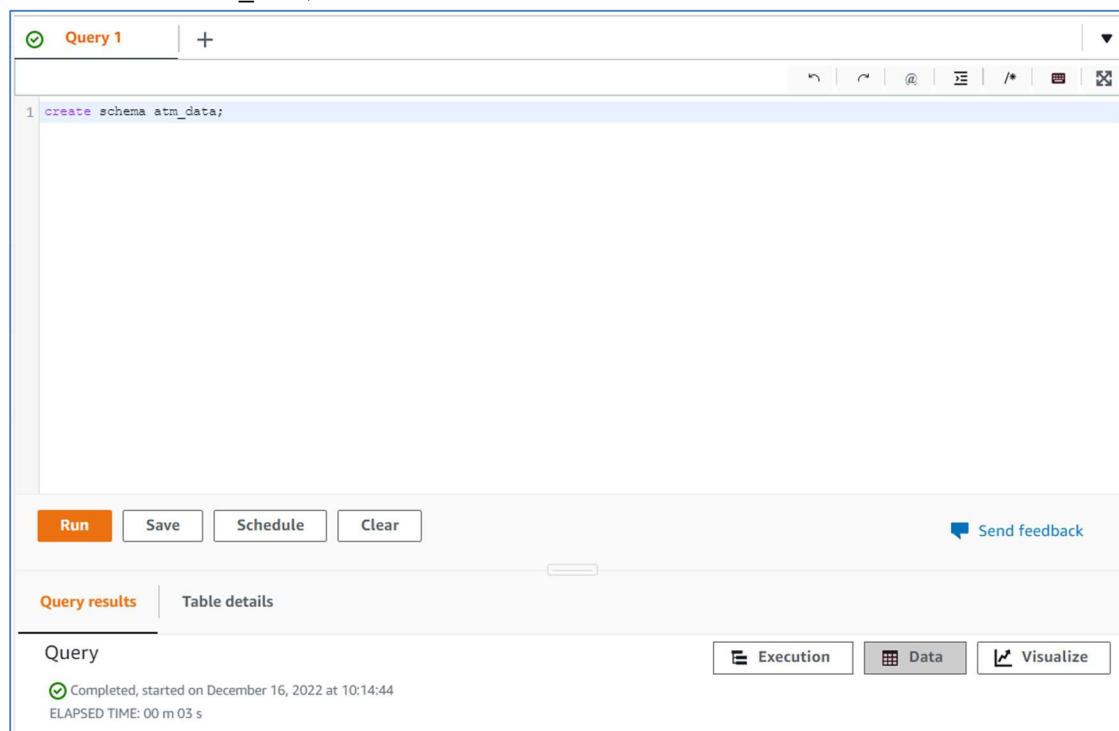
Screenshot of IAM Roles associated with Redshift Cluster for S3 Access



Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables. Below are commands and related screenshots.

- **Creating schema atm_data:**

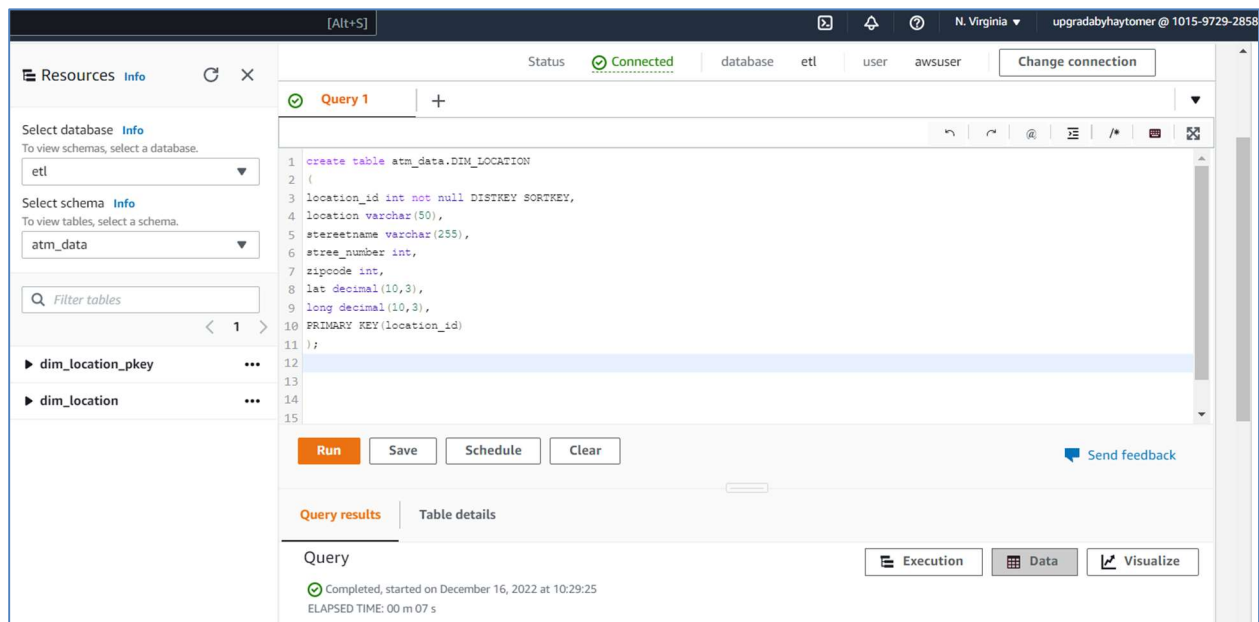
Create schema atm_data;



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

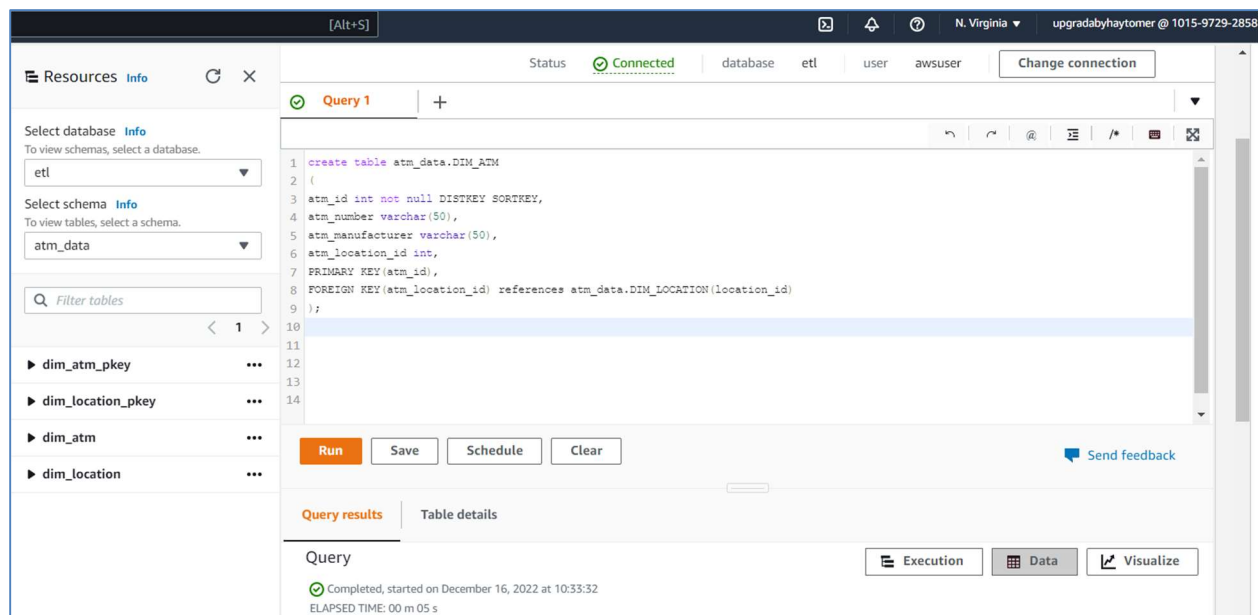
- **Creating Location dimension Table:**

```
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),
long decimal(10,3),
PRIMARY KEY(location_id)
);
```



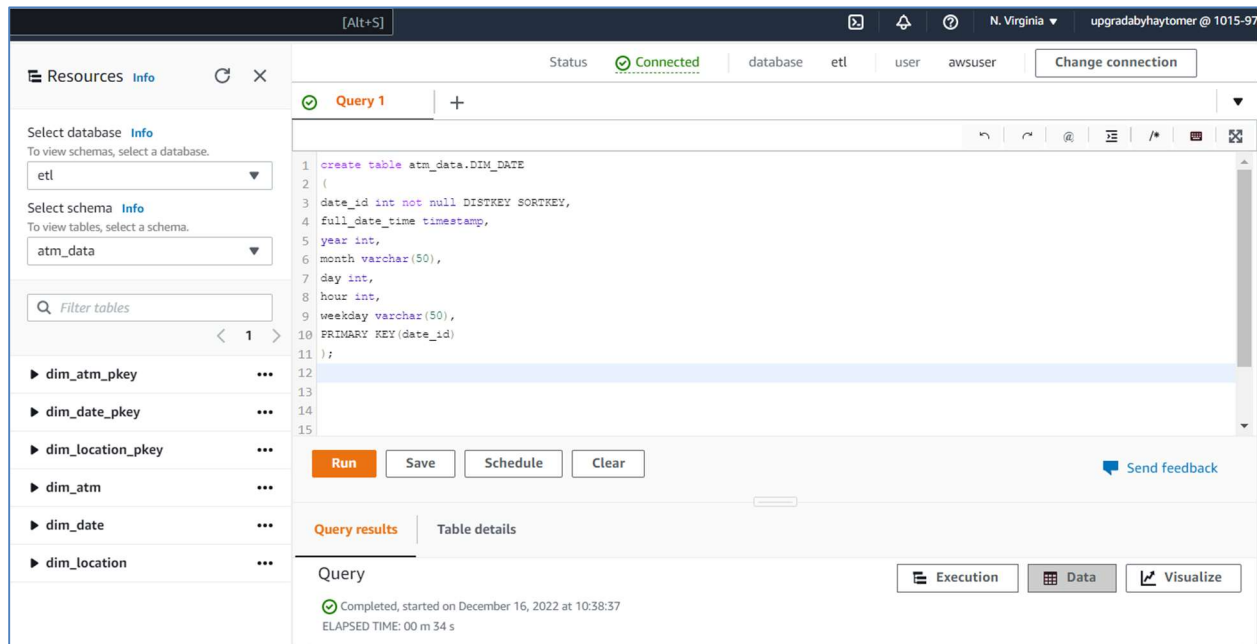
- **Creating ATM Dimension Table:**

```
create table atm_data.DIM_ATM
(
atm_id int not null DISTKEY SORTKEY,
atm_number varchar(50),
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(50),
  day int,
  hour int,
  weekday varchar(50),
  PRIMARY KEY(date_id)
);
```



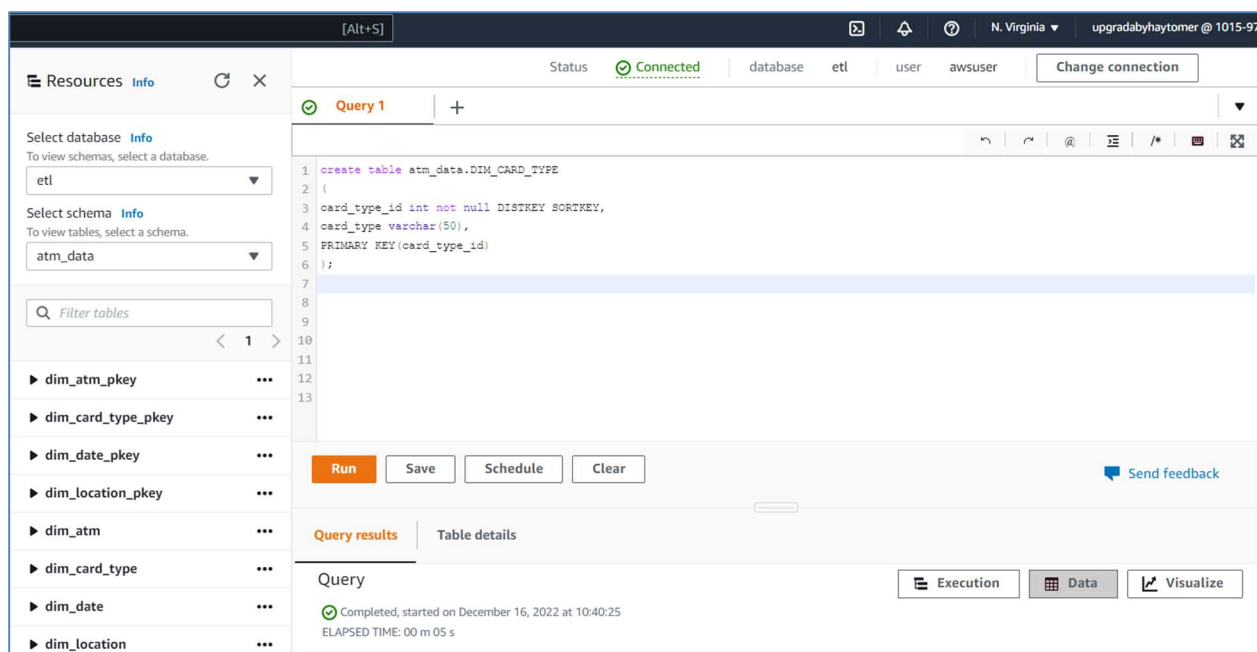
The screenshot shows the upGrad SQL IDE interface. On the left, the 'Resources' panel displays the database 'etl' and schema 'atm_data'. The main editor area contains the following SQL query:

```
1 create table atm_data.DIM_DATE
2 (
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int,
6   month varchar(50),
7   day int,
8   hour int,
9   weekday varchar(50),
10  PRIMARY KEY (date_id)
11 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing the query execution status: 'Completed, started on December 16, 2022 at 10:38:37' and 'ELAPSED TIME: 00 m 34 s'.

- **Creating Card Type Dimension Table:**

```
create table atm_data.DIM_CARD_TYPE
(
  card_type_id int not null DISTKEY SORTKEY,
  card_type varchar(50),
  PRIMARY KEY(card_type_id)
);
```



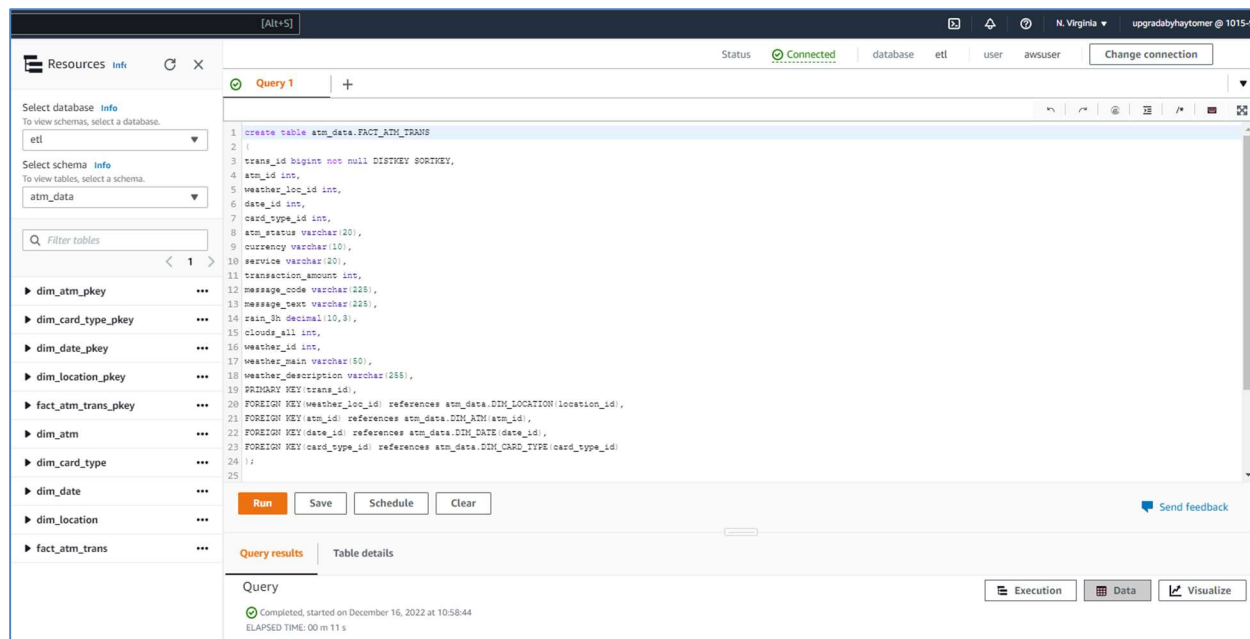
The screenshot shows the upGrad SQL IDE interface. On the left, the 'Resources' panel displays the database 'etl' and schema 'atm_data'. The main editor area contains the following SQL query:

```
1 create table atm_data.DIM_CARD_TYPE
2 (
3   card_type_id int not null DISTKEY SORTKEY,
4   card_type varchar(50),
5   PRIMARY KEY (card_type_id)
6 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing the query execution status: 'Completed, started on December 16, 2022 at 10:40:25' and 'ELAPSED TIME: 00 m 05 s'.

- **Creating Fact ATM Trans 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)
);
```



The screenshot shows a database query editor interface. On the left, there's a sidebar with a tree view showing the database structure: 'etl' database, 'atm_data' schema, and various tables like 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main area displays a SQL query for creating the 'fact_atm_trans' table. The query is as follows:

```
1 create table atm_data.FACT_ATM_TRANS
2 (
3 trans_id bigint not null DISTKEY SORTKEY,
4 atm_id int,
5 weather_loc_id int,
6 date_id int,
7 card_type_id int,
8 atm_status varchar(20),
9 currency varchar(10),
10 service varchar(20),
11 transaction_amount int,
12 message_code varchar(225),
13 message_text varchar(225),
14 rain_3h decimal(10,3),
15 clouds_all int,
16 weather_id int,
17 weather_main varchar(50),
18 weather_description varchar(255),
19 PRIMARY KEY(trans_id),
20 FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
21 FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
22 FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
23 FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
24 );
25
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. At the bottom, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows the query execution status: 'Completed, started on December 16, 2022 at 10:58:44' and 'ELAPSED TIME: 00 m 11 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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

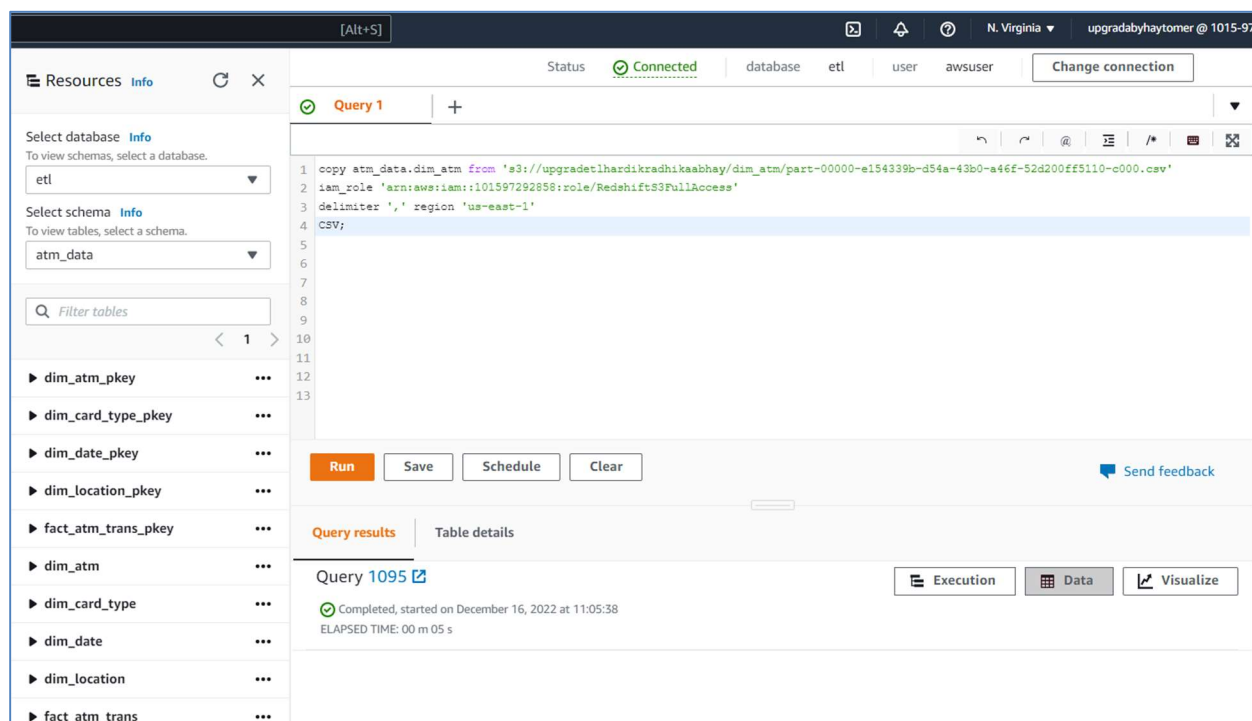
- Copying the data to dim_atm from S3 location:

```
copy atm_data.dim_atm from 's3://upgradetlhardikradhikaabhay/dim_atm/part-00000-e154339b-d54a-43b0-a46f-52d200ff5110-c000.csv'
```

```
iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
```

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

```
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Under 'Resources', there's a 'Select database' dropdown set to 'etl' and a 'Select schema' dropdown set to 'atm_data'. Below these are search filters and a list of tables including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main area shows a SQL query being executed. The query is:
 1 copy atm_data.dim_atm from 's3://upgradetlhardikradhikaabhay/dim_atm/part-00000-e154339b-d54a-43b0-a46f-52d200ff5110-c000.csv'
 2 iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
 3 delimiter ',' region 'us-east-1'
 4 CSV;
 5
 6
 7
 8
 9
 10
 11
 12
 13
 Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 1095' with a status of 'Completed, started on December 16, 2022 at 11:05:38' and 'ELAPSED TIME: 00 m 05 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

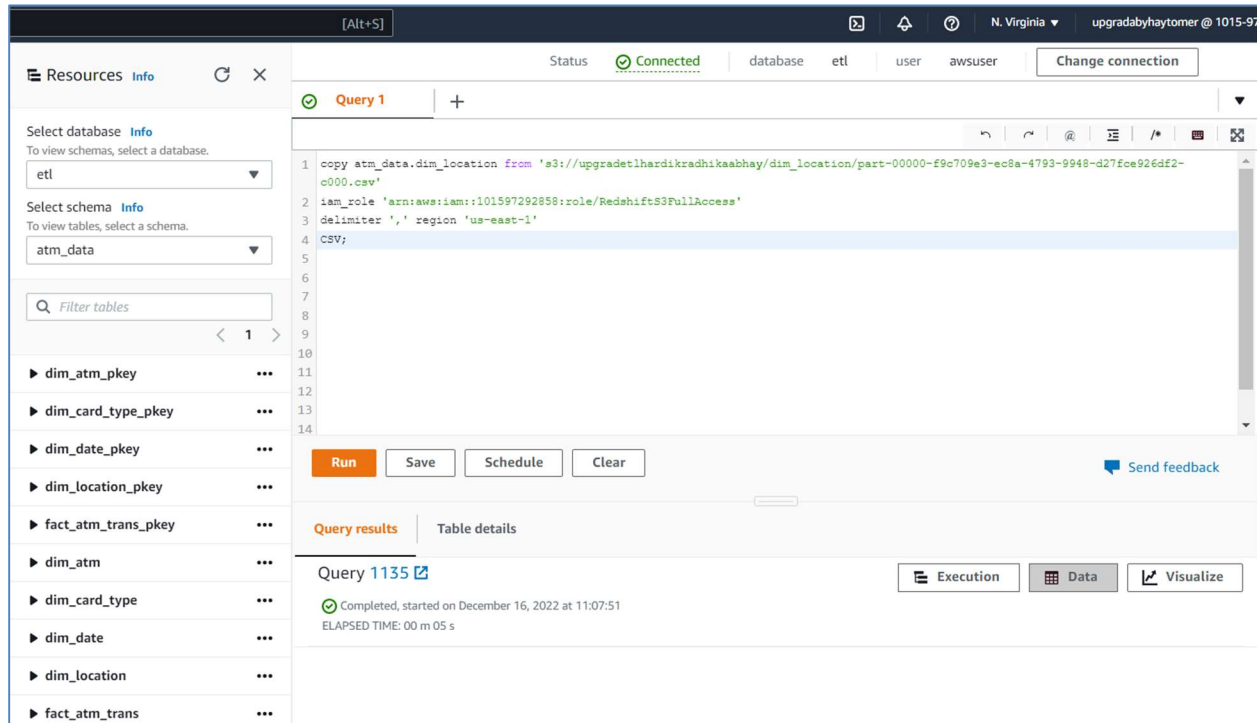
- Copying the data to dim_location from S3 location:

```
copy atm_data.dim_location from 's3://upgradetlhardikradhikaabhay/dim_location/part-00000-f9c709e3-ec8a-4793-9948-d27fce926df2-c000.csv'
```

```
iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
```

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

```
CSV;
```



The screenshot shows the Amazon Redshift console interface. On the left, the 'Resources' sidebar is visible, showing the database 'etl' and schema 'atm_data'. The main area displays a SQL query for 'Query 1' with the following code:

```
1 copy atm_data.dim_location from 's3://upgradetlhardikradhikaabhay/dim_location/part-00000-f9c709e3-e08a-4793-9948-d27fce926df2-c000.csv'
2 iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 1135' with a status of 'Completed, started on December 16, 2022 at 11:07:51' and an 'ELAPSED TIME: 00 m 05 s'. The 'Table details' tab is also visible.

- **Copying the data to dim_date from S3 location:**

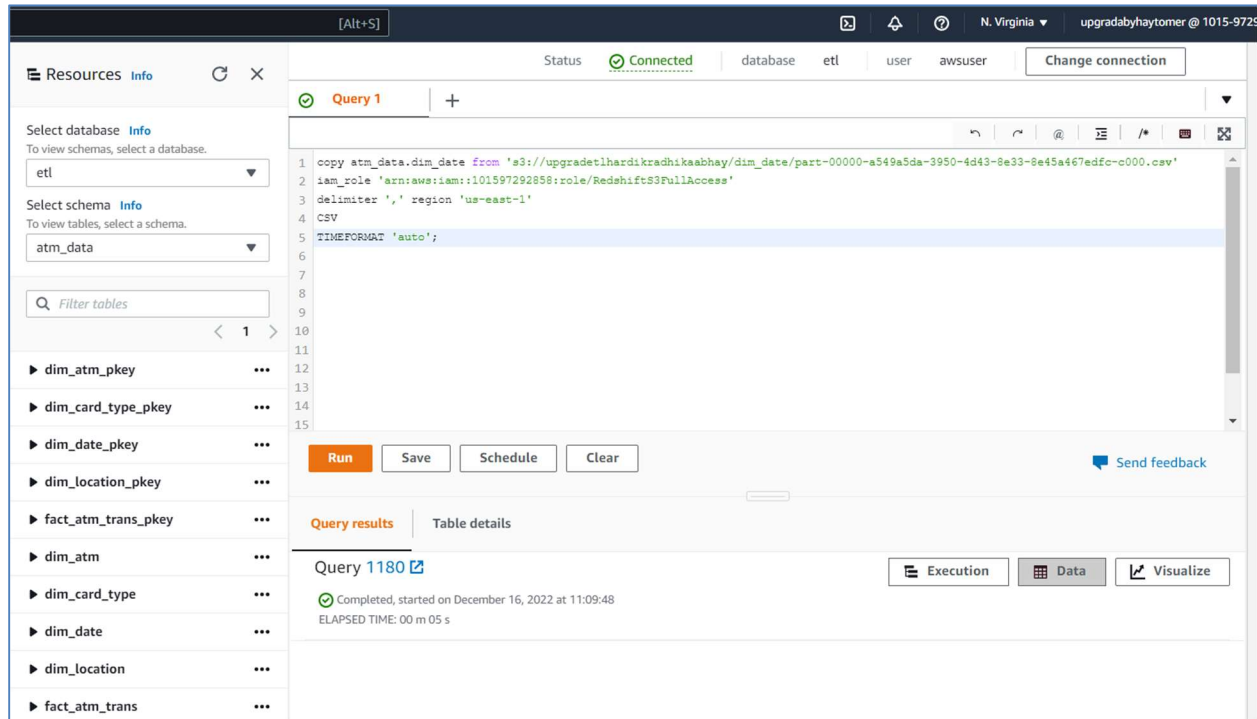
copy atm_data.dim_date from 's3://upgradetlhardikradhikaabhay/dim_date/part-00000-a549a5da-3950-4d43-8e33-8e45a467edfc-c000.csv'

iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'

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

CSV

TIMEFORMAT 'auto';



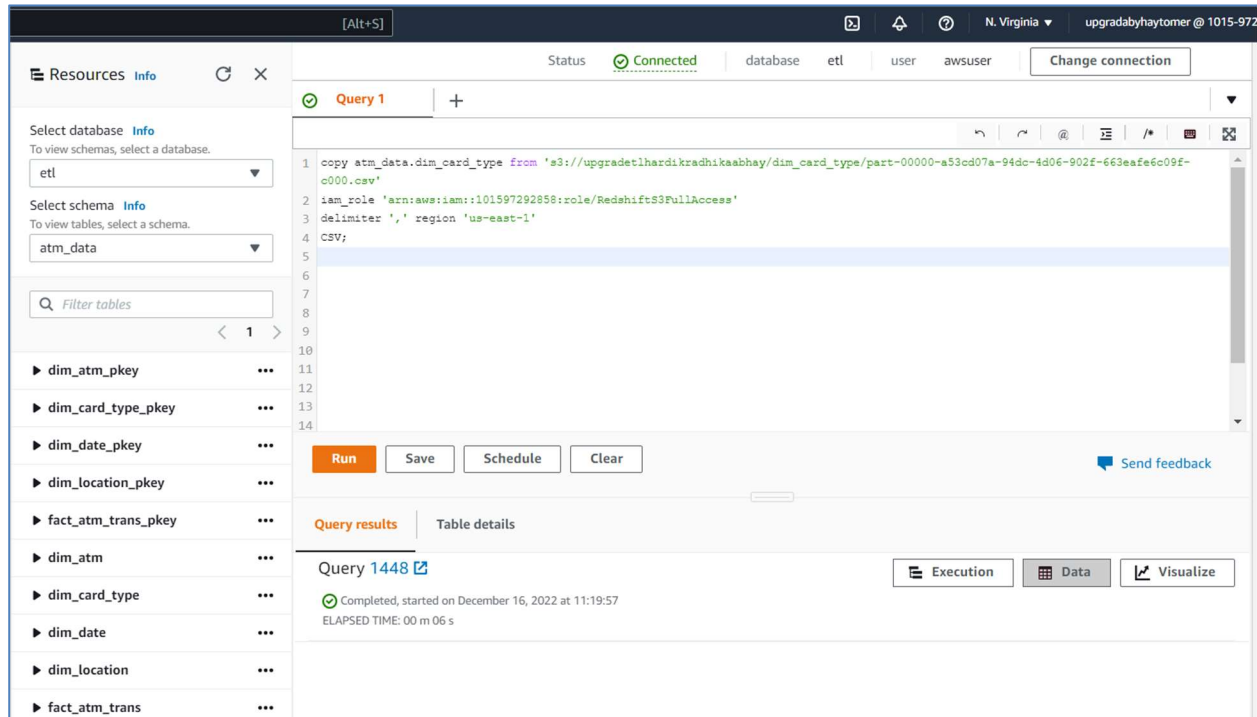
The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel displays the database 'etl' and schema 'atm_data'. The main area shows a SQL query for 'Query 1':

```
1 copy atm_data.dim_date from 's3://upgradetlhardikradhikaabhay/dim_date/part-00000-a549a5da-3950-4d43-8e33-8e45a467edfc-c000.csv'
2 iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab shows the query completed successfully on December 16, 2022, at 11:09:48, with an elapsed time of 00 m 05 s.

- **Copying the data to dim_card_type from S3 location:**

```
copy atm_data.dim_card_type from 's3://upgradetlhardikradhikaabhay/dim_card_type/part-00000-
a53cd07a-94dc-4d06-902f-663eafe6c09f-c000.csv'
iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
delimiter ',' region 'us-east-1'
CSV;
```



- **Copying the data to fact_atm_trans table from S3 location:**

```
copy atm_data.fact_atm_trans from 's3://upgradetlhardikradhikaabhay/fact_atm_trans/part-00000-
ea5f0521-d9cd-44af-b159-4d6fd35a29f3-c000.csv'
iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
delimiter ',' region 'us-east-1'
CSV;
```

Resources Info

Select database Info

To view schemas, select a database.

etl

Select schema Info

To view tables, select a schema.

atm_data

Filter tables

1

dim_atm_pkey

dim_card_type_pkey

dim_date_pkey

dim_location_pkey

fact_atm_trans_pkey

dim_atm

dim_card_type

dim_date

dim_location

fact_atm_trans

Status Connected

database etl user awsuser

Change connection

Query 1

```

1 copy atm_data.fact_atm_trans from 's3://upgradetlhardikradhikaabhay/fact_atm_trans/part-00000-ea5f0521-d9cd-44af-b159-4d6fd35a29f3-c000.csv'
2 iam_role 'arn:aws:iam::101597292858:role/RedshiftS3FullAccess'
3 delimiter ',' region 'us-east-1'
4 CSV;

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 1487

Execution

Data

Visualize

Completed, started on December 16, 2022 at 11:21:24

ELAPSED TIME: 00 m 17 s