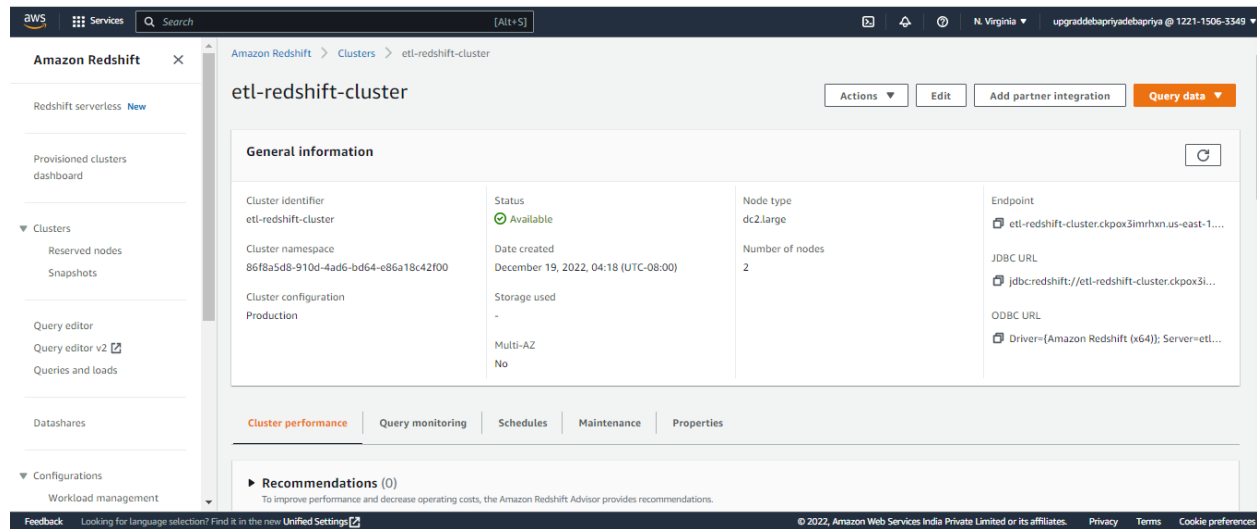


Creation of a Redshift Cluster

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



The screenshot displays the AWS Management Console interface for an Amazon Redshift cluster. The left sidebar shows the navigation menu with options like 'Amazon Redshift', 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', and 'Workload management'. The main content area shows the details for the cluster 'etl-redshift-cluster'. The 'General information' tab is active, displaying a table with the following data:

Cluster identifier	Status	Node type	Endpoint
etl-redshift-cluster	Available	dc2.large	etl-redshift-cluster.cpkox3imrhxn.us-east-1....
Cluster namespace	Date created	Number of nodes	JDBC URL
86f8a5d8-910d-4ad6-bd64-e86a18c42f00	December 19, 2022, 04:18 (UTC-08:00)	2	jdbcredshift://etl-redshift-cluster.cpkox3i...
Cluster configuration	Storage used		ODBC URL
Production	-		Driver=(Amazon Redshift (x64)); Server=etl...
	Multi-AZ		
	No		

Below the table, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'. The 'Recommendations (0)' section is also visible, indicating no recommendations are currently present.

Amazon Redshift

- Redshift serverless *New*
- Provisioned clusters dashboard
- Clusters
 - Reserved nodes
 - Snapshots
- Query editor
 - Query editor v2
 - Queries and loads
- Datashares
- Configurations
 - Workload management

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.

Associated IAM roles (1) [Info](#)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

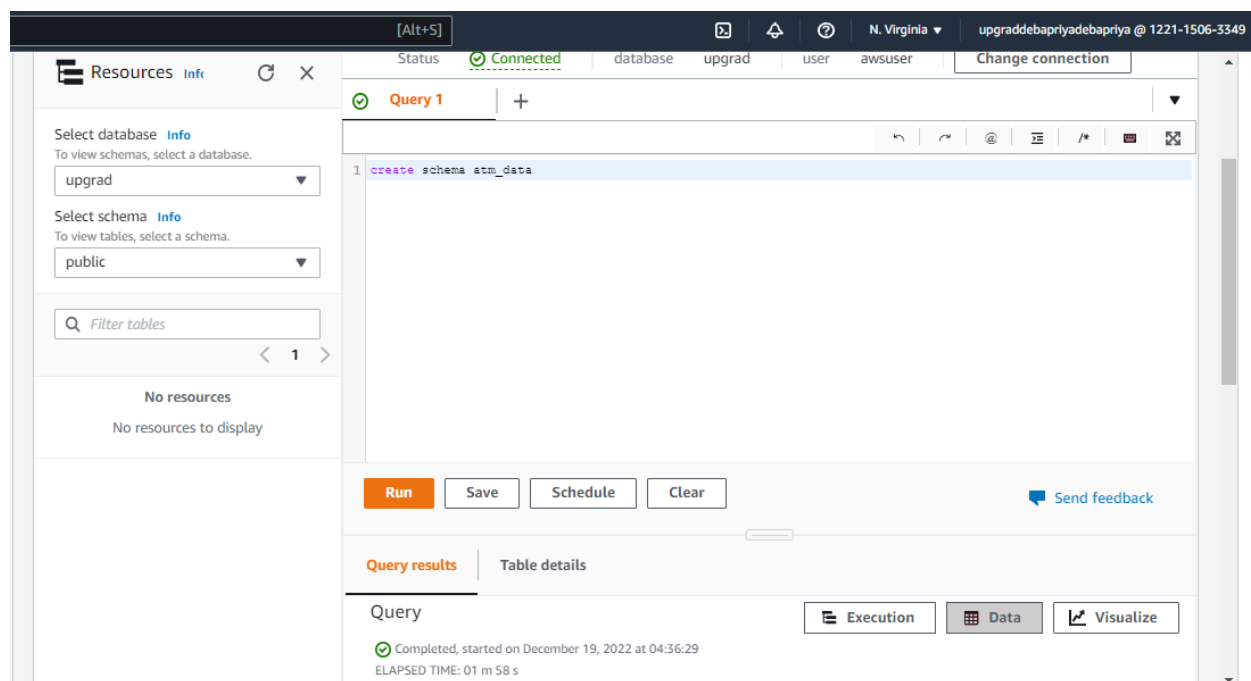
< 1 >

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	Redshift_S3_Full_Access	In-sync	Default

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

Query for creating schema

```
create schema atm_data;
```



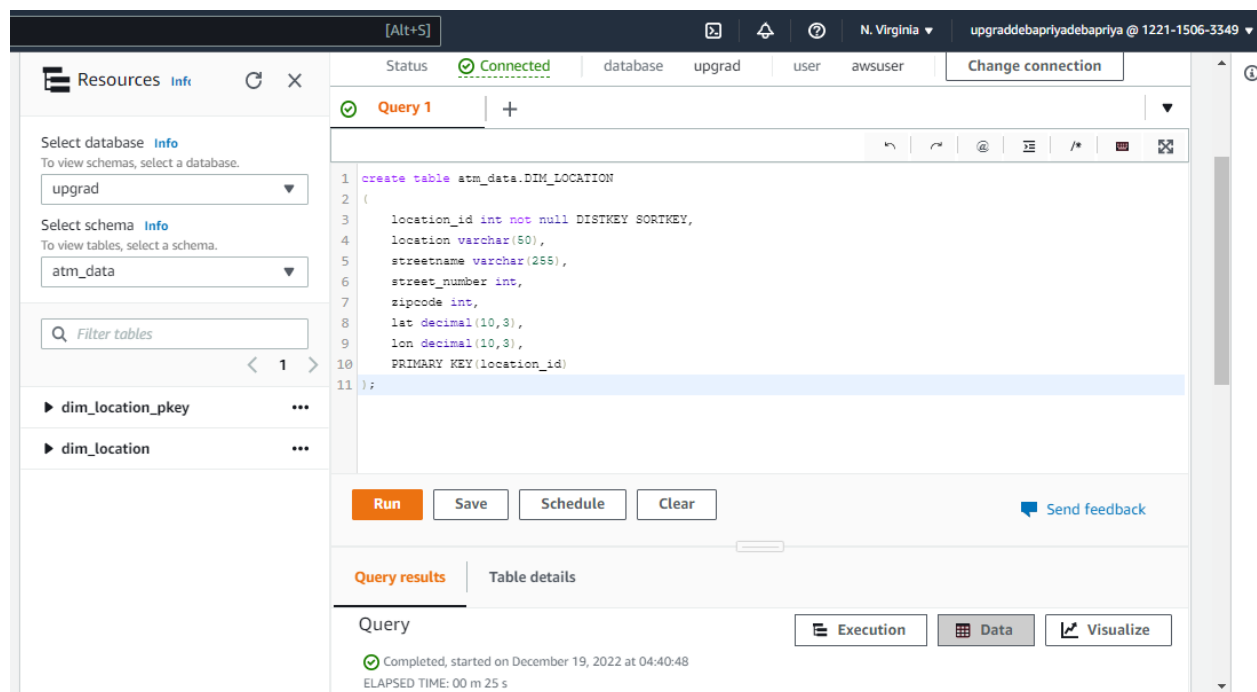
The screenshot displays the AWS Redshift console interface. At the top, the status bar shows 'Connected' and the database 'upgrad' with user 'awsuser'. The left sidebar contains navigation options: 'Resources', 'Info', and a search bar. The main area shows a query editor with the text 'create schema atm_data;'. Below the editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating completion. The execution details show 'Completed, started on December 19, 2022 at 04:36:29' and 'ELAPSED TIME: 01 m 58 s'. The 'Table details' tab is also visible.

Queries 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),  
    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),

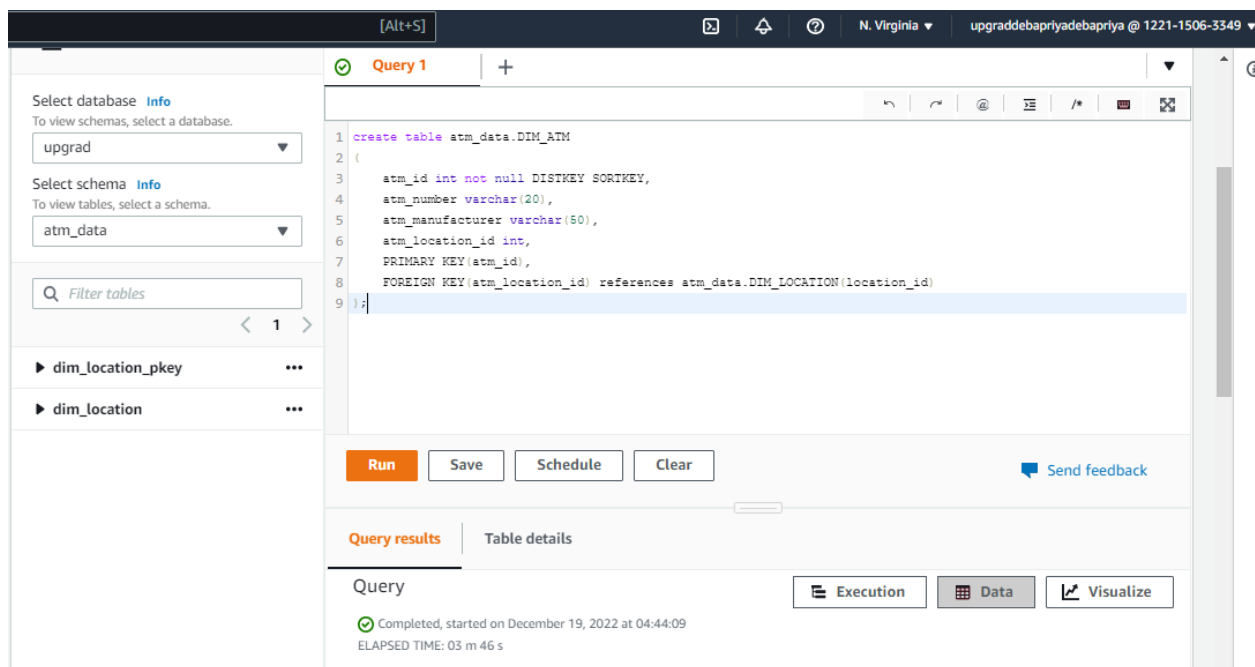
atm_manufacturer varchar(50),

atm_location_id int,

PRIMARY KEY(atm_id),

FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)

);



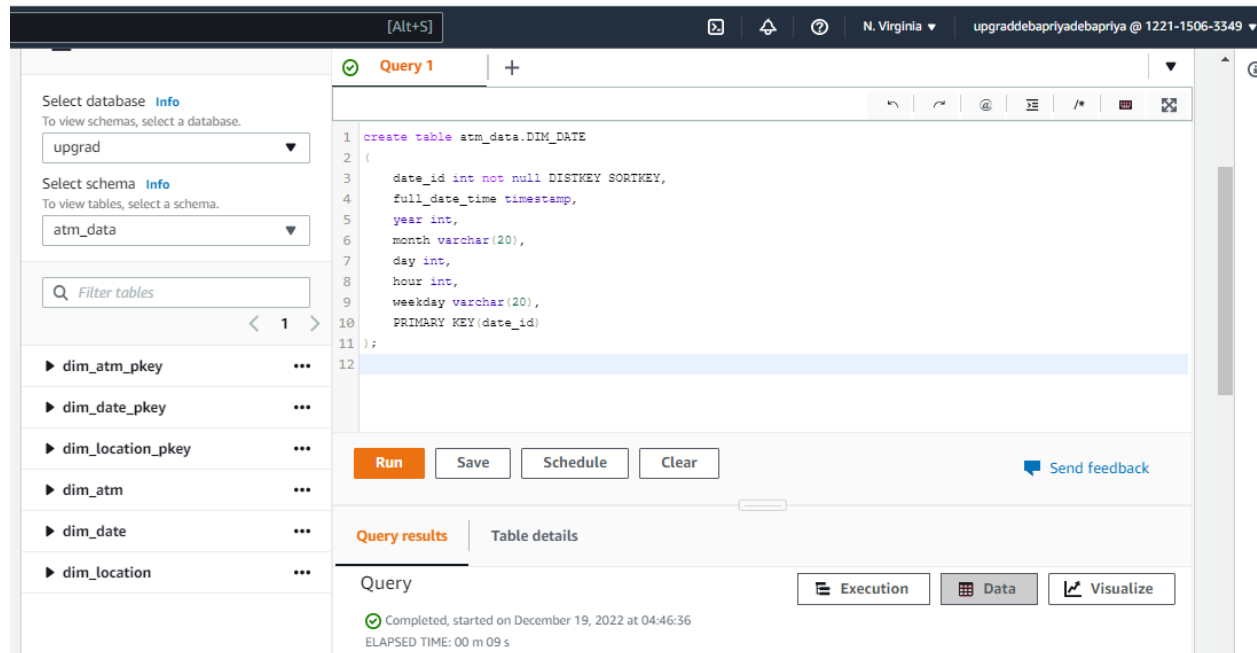
The screenshot shows a SQL query editor interface. On the left, there is a sidebar with a 'Select database' dropdown set to 'upgrad' and a 'Select schema' dropdown set to 'atm_data'. Below these are search filters and a list of tables including 'dim_location_pkey' and 'dim_location'. The main editor area displays a SQL query to create a table named 'atm_data.DIM_ATM'. The query includes columns for 'atm_id' (int, not null, DISTKEY SORTKEY), 'atm_number' (varchar(20)), 'atm_manufacturer' (varchar(50)), and 'atm_location_id' (int). It also defines a primary key on 'atm_id' and a foreign key on 'atm_location_id' that references 'atm_data.DIM_LOCATION(location_id)'. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, the 'Query results' tab is active, showing a status message: 'Completed, started on December 19, 2022 at 04:44:09' and 'ELAPSED TIME: 03 m 46 s'. There are also tabs for 'Execution', 'Data', and 'Visualize'.

```
1 create table atm_data.DIM_ATM
2 (
3     atm_id int not null DISTKEY SORTKEY,
4     atm_number varchar(20),
5     atm_manufacturer varchar(50),
6     atm_location_id int,
7     PRIMARY KEY(atm_id),
8     FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );
```

- Creating date dimension table

create table atm_data.DIM_DATE

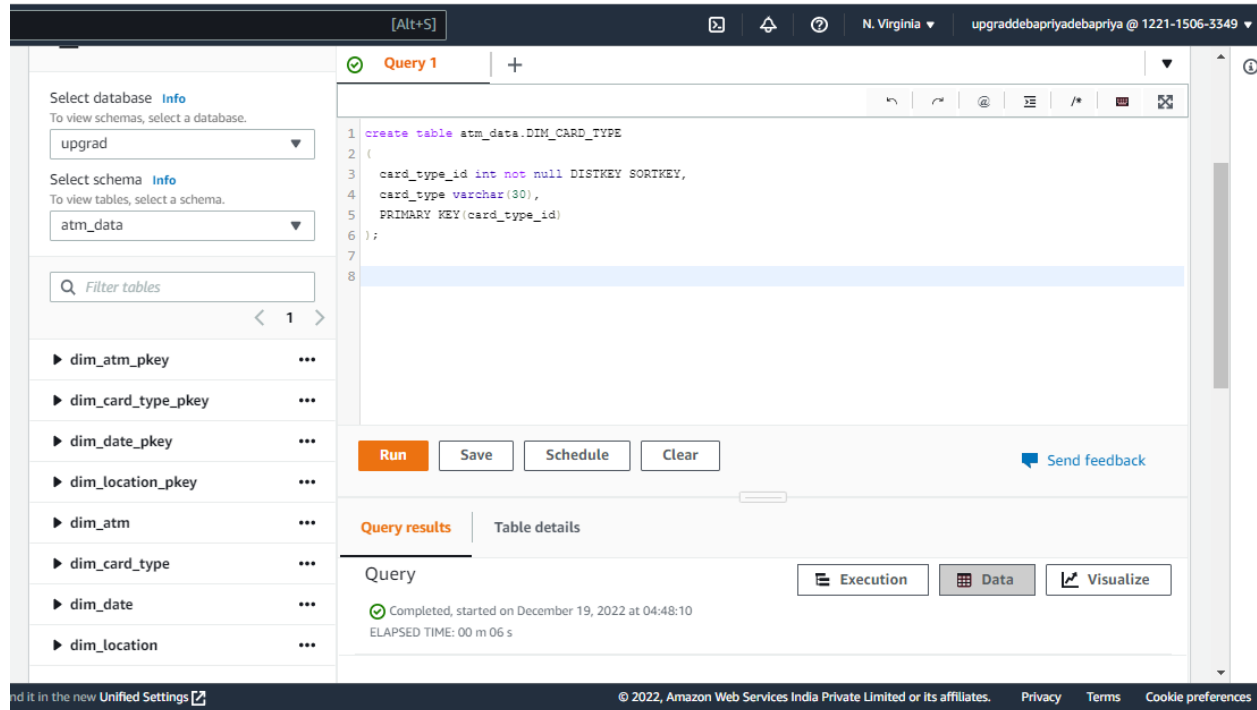
```
(
    date_id int not null DISTKEY SORTKEY,
    full_date_time timestamp,
    year int,
    month varchar(20),
    day int,
    hour int,
    weekday varchar(20),
    PRIMARY KEY(date_id)
);
```



The screenshot shows a SQL query editor interface. On the left, there is a sidebar with a search bar and a list of tables under the 'atm_data' schema. The main area displays a SQL query to create a table named 'DIM_DATE' in the 'atm_data' database. The query includes columns for 'date_id' (int, not null, DISTKEY, SORTKEY), 'full_date_time' (timestamp), 'year' (int), 'month' (varchar(20)), 'day' (int), 'hour' (int), 'weekday' (varchar(20)), and a primary key on 'date_id'. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there is a section for 'Query results' and 'Table details'. The 'Query results' section shows a status message: 'Completed, started on December 19, 2022 at 04:46:36' and 'ELAPSED TIME: 00 m 09 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- 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)
);
```



The screenshot shows the Amazon Redshift console interface. On the left, the 'Select database' dropdown is set to 'upgrad' and the 'Select schema' dropdown is set to 'atm_data'. Below these, a list of tables is displayed, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'dim_atm', 'dim_card_type', 'dim_date', and 'dim_location'. The main area shows a SQL query editor with the following code:

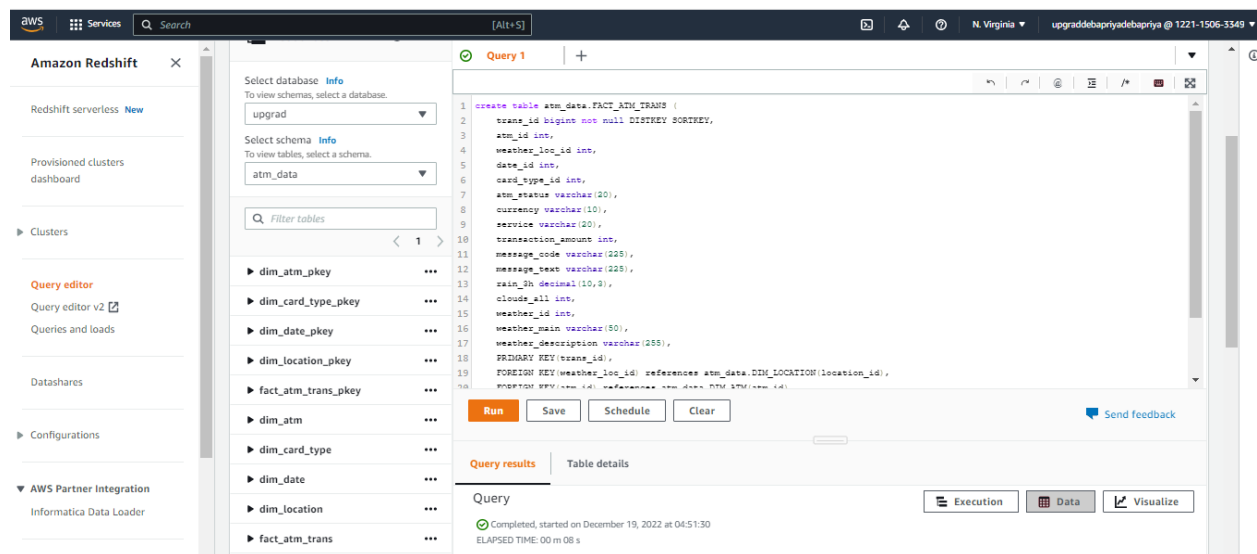
```
1 create table atm_data.DIM_CARD_TYPE
2 (
3     card_type_id int not null DISTKEY SORTKEY,
4     card_type varchar(30),
5     PRIMARY KEY(card_type_id)
6 );
7
8
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of the 'Run' button is a 'Send feedback' link. Below the query editor, the 'Query results' tab is selected, showing the query execution status: 'Completed, started on December 19, 2022 at 04:48:10' and 'ELAPSED TIME: 00 m 06 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Creating atm transactions 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)
);
```



The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Query editor' sidebar is visible, showing the 'Query editor v2' tab. The main area displays the SQL query for creating the 'FACT_ATM_TRANS' table. The query is as follows:

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

Below the query, the 'Run' button is highlighted. The 'Query results' tab is selected, showing the 'Query' details. The query status is 'Completed, started on December 19, 2022 at 04:51:30' and the 'ELAPSED TIME' is '00 m 08 s'.

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

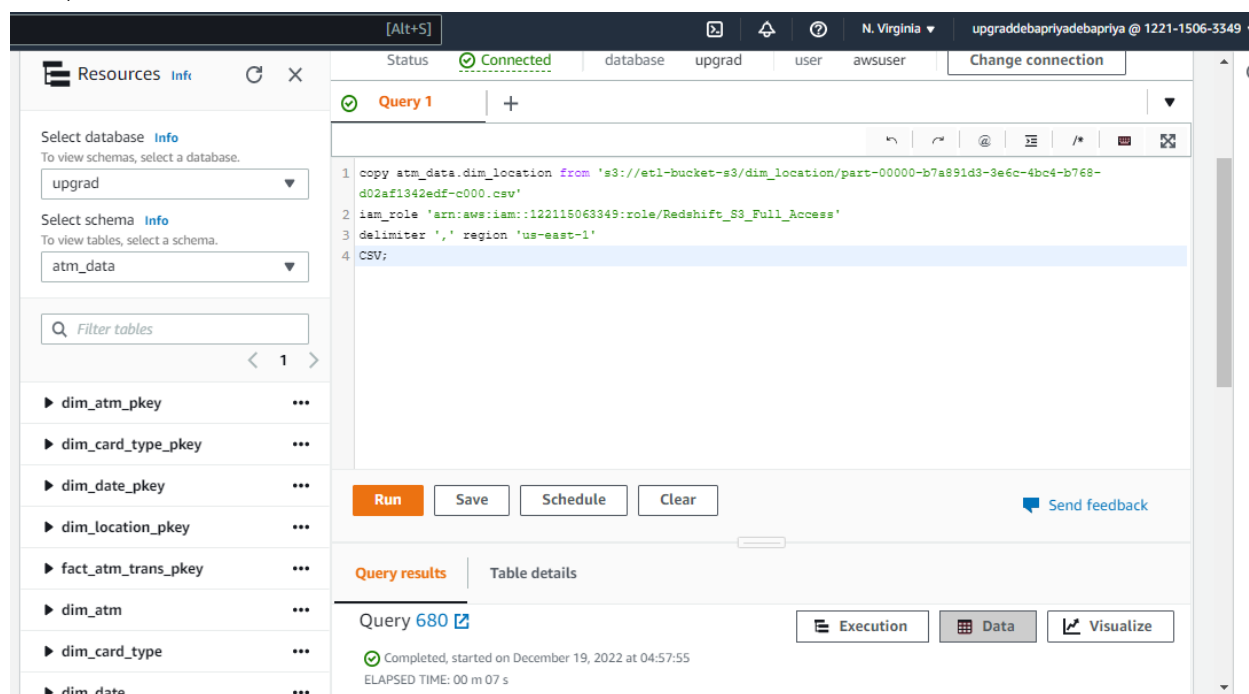
copy atm_data.dim_location from

's3://etl-bucket-s3/dim_location/part-00000-b7a891d3-3e6c-4bc4-b768-d02af1342edf-c000.csv'

iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'

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

CSV;



The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel displays a list of databases and schemas. The 'atm_data' schema is selected, showing 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', and 'dim_date'. The main panel shows a SQL query for 'Query 1' with the following text:

```
1 copy atm_data.dim_location from 's3://etl-bucket-s3/dim_location/part-00000-b7a891d3-3e6c-4bc4-b768-
2 d02af1342edf-c000.csv'
3 iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
4 delimiter ',' region 'us-east-1'
5 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' tab is active, showing 'Query 680' with a status of 'Completed, started on December 19, 2022 at 04:57:55' and an 'ELAPSED TIME: 00 m 07 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to dim_atm table

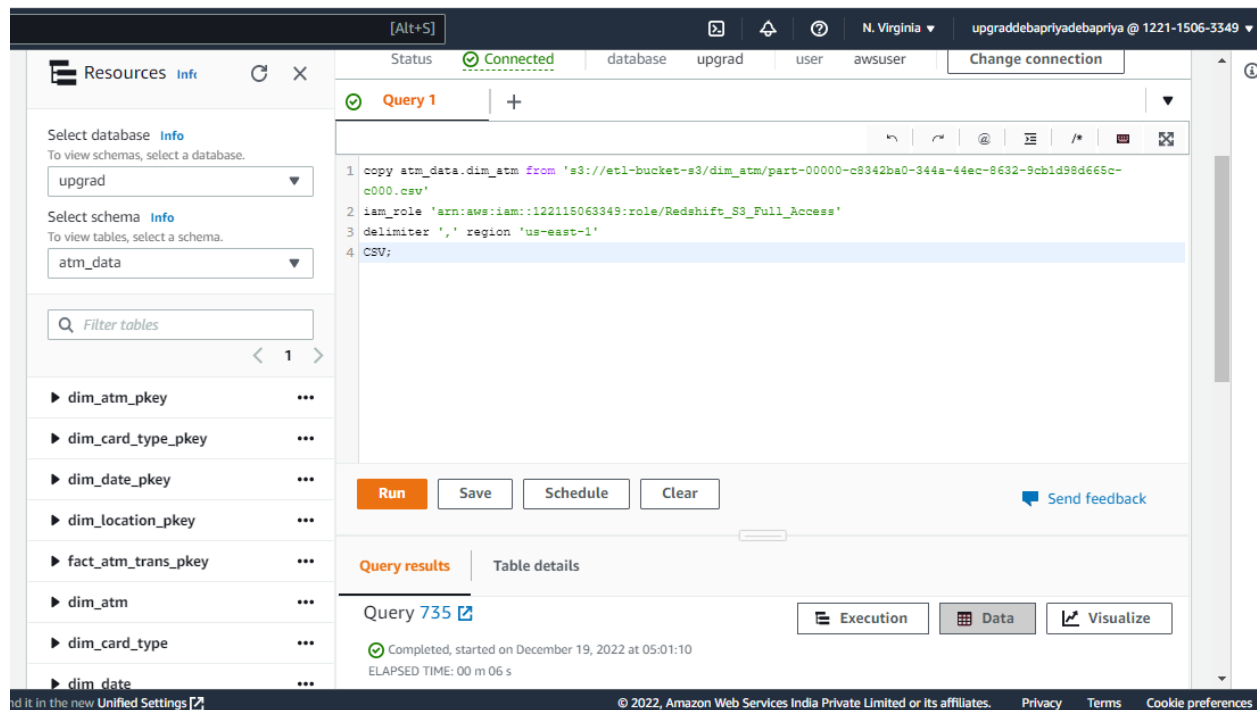
copy atm_data.dim_atm from

's3://etl-bucket-s3/dim_atm/part-00000-c8342ba0-344a-44ec-8632-9cb1d98d665c-c000.csv'

iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'

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

CSV;



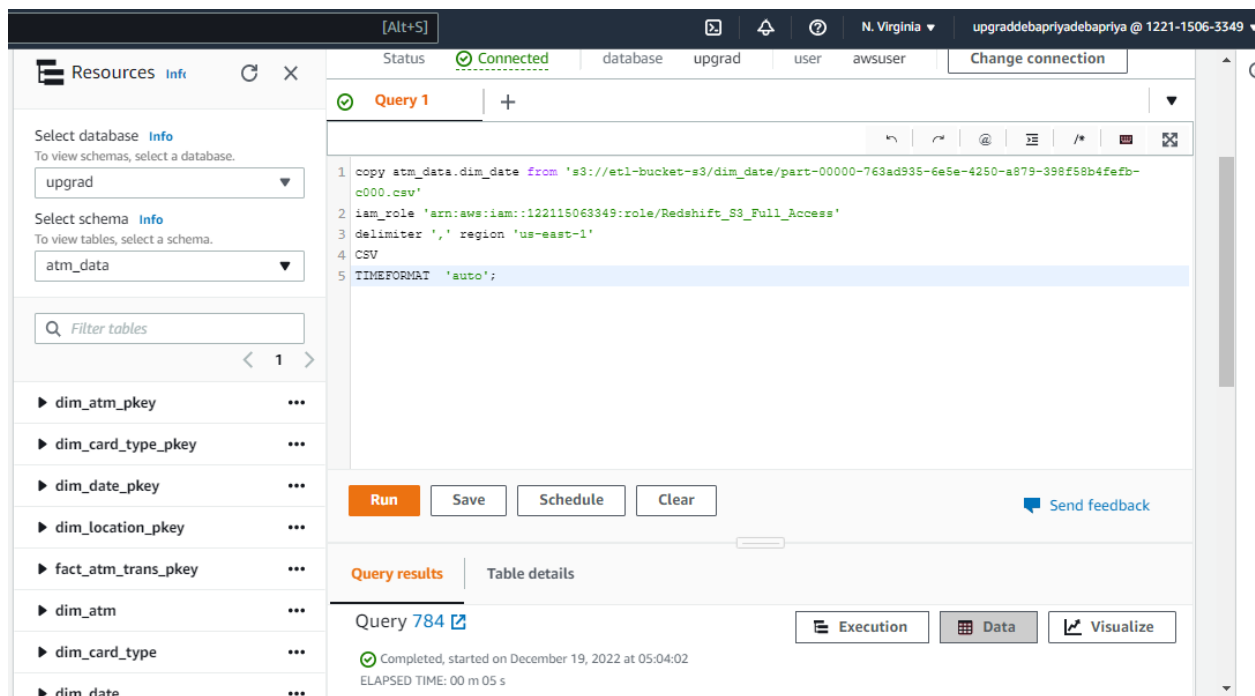
The screenshot displays the Amazon Redshift console interface. On the left, the 'Resources' sidebar shows the database 'upgrad' and schema 'atm_data'. The main panel shows a SQL query being executed:

```
1 copy atm_data.dim_atm from 's3://etl-bucket-s3/dim_atm/part-00000-c8342ba0-344a-44ec-8632-9cb1d98d665c-c000.csv'
2 iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab shows the query status as 'Completed' with an elapsed time of 00 m 06 s. The footer of the console displays the copyright notice: '© 2022, Amazon Web Services India Private Limited or its affiliates. Privacy Terms Cookie preferences'.

- Copying the data to dim_date table

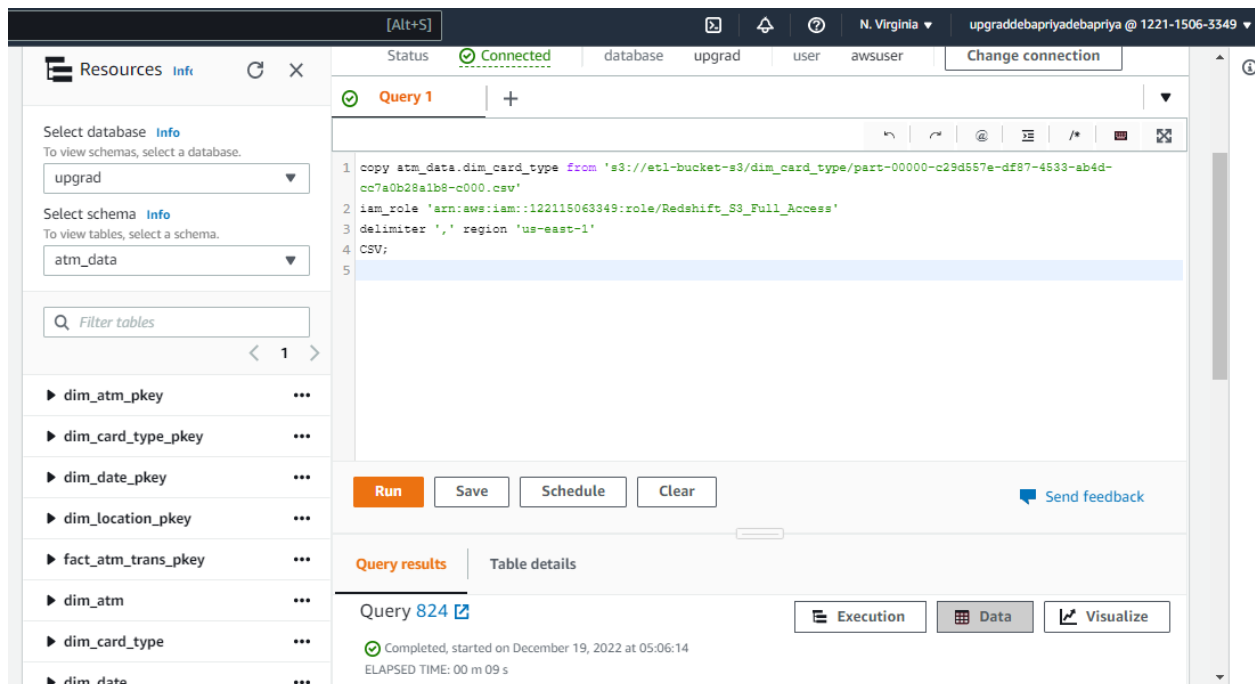
```
copy atm_data.dim_date from
's3://etl-bucket-s3/dim_date/part-00000-763ad935-6e5e-4250-a879-398f58b4fefb-c000.csv'
iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



The screenshot shows the upGrad Redshift console interface. On the left, there's a sidebar with 'Resources' 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', and 'dim_date'. The main area displays a SQL query in a text editor, which is the same query provided in the text block above. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 784' with a status of 'Completed, started on December 19, 2022 at 05:04:02' and an 'ELAPSED TIME: 00 m 05 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to dim_card_type table

copy atm_data.dim_card_type from
 's3://etl-bucket-s3/dim_card_type/part-00000-c29d557e-df87-4533-ab4d-cc7a0b28a1b8-c000.csv'
 v'
 iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
 delimiter ',' region 'us-east-1'
 CSV;



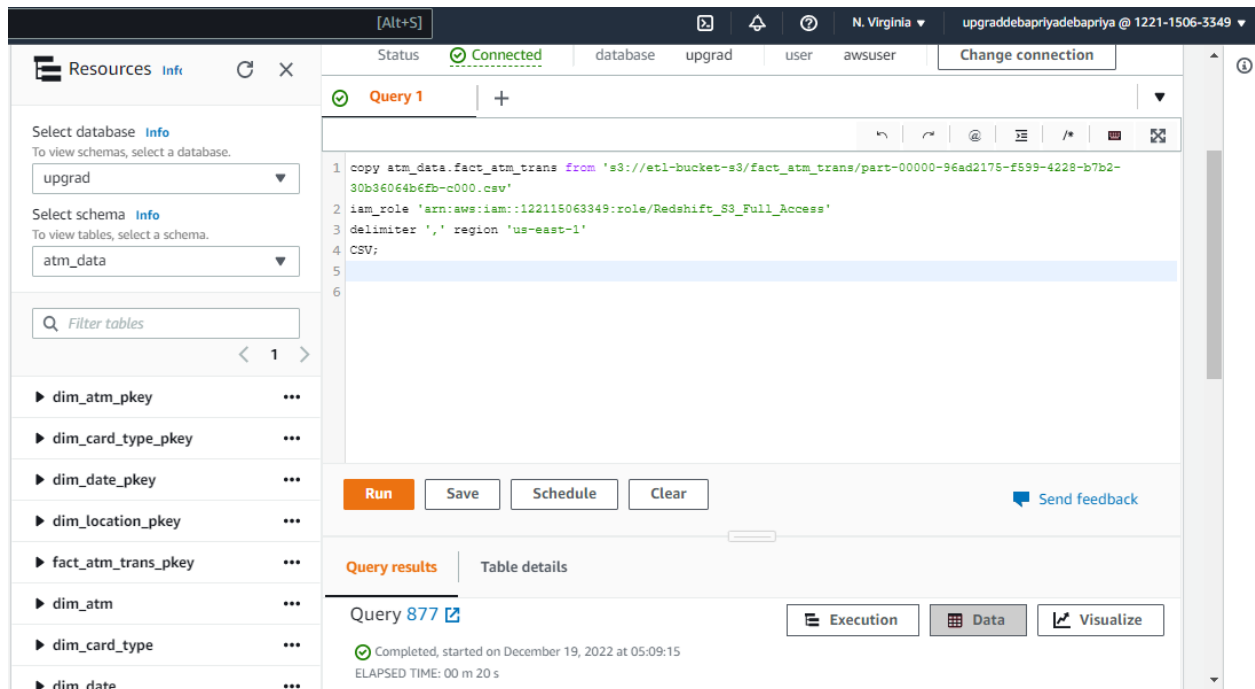
The screenshot shows the upGrad Redshift console interface. On the left, the 'Resources' panel displays a list of tables under the 'atm_data' schema, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', and 'dim_date'. The main panel shows a SQL query being executed:

```
1 copy atm_data.dim_card_type from 's3://etl-bucket-s3/dim_card_type/part-00000-c29d557e-df87-4533-ab4d-cc7a0b28a1b8-c000.csv'
2 iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
3 delimiter ',' region 'us-east-1'
4 CSV;
5
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is selected, showing the query ID 'Query 824' and its execution status: 'Completed, started on December 19, 2022 at 05:06:14' with an 'ELAPSED TIME: 00 m 09 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to fact_atm_trans table

copy atm_data.fact_atm_trans from
's3://etl-bucket-s3/fact_atm_trans/part-00000-96ad2175-f599-4228-b7b2-30b36064b6fb-c000.csv'
v'
iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'
delimiter ',' region 'us-east-1'
CSV;



The screenshot shows the upGrad 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 'upgrad' 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', and 'dim_date'. The main area shows a SQL query being executed. The query is:
`1 copy atm_data.fact_atm_trans from 's3://etl-bucket-s3/fact_atm_trans/part-00000-96ad2175-f599-4228-b7b2-30b36064b6fb-c000.csv'`
`2 iam_role 'arn:aws:iam::122115063349:role/Redshift_S3_Full_Access'`
`3 delimiter ',' region 'us-east-1'`
`4 CSV;`
`5`
`6`
Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of the buttons is a 'Send feedback' link. Below the query editor, there's a 'Query results' tab and a 'Table details' tab. The 'Query results' tab is active, showing 'Query 877' with a status of 'Completed, started on December 19, 2022 at 05:09:15' and 'ELAPSED TIME: 00 m 20 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.