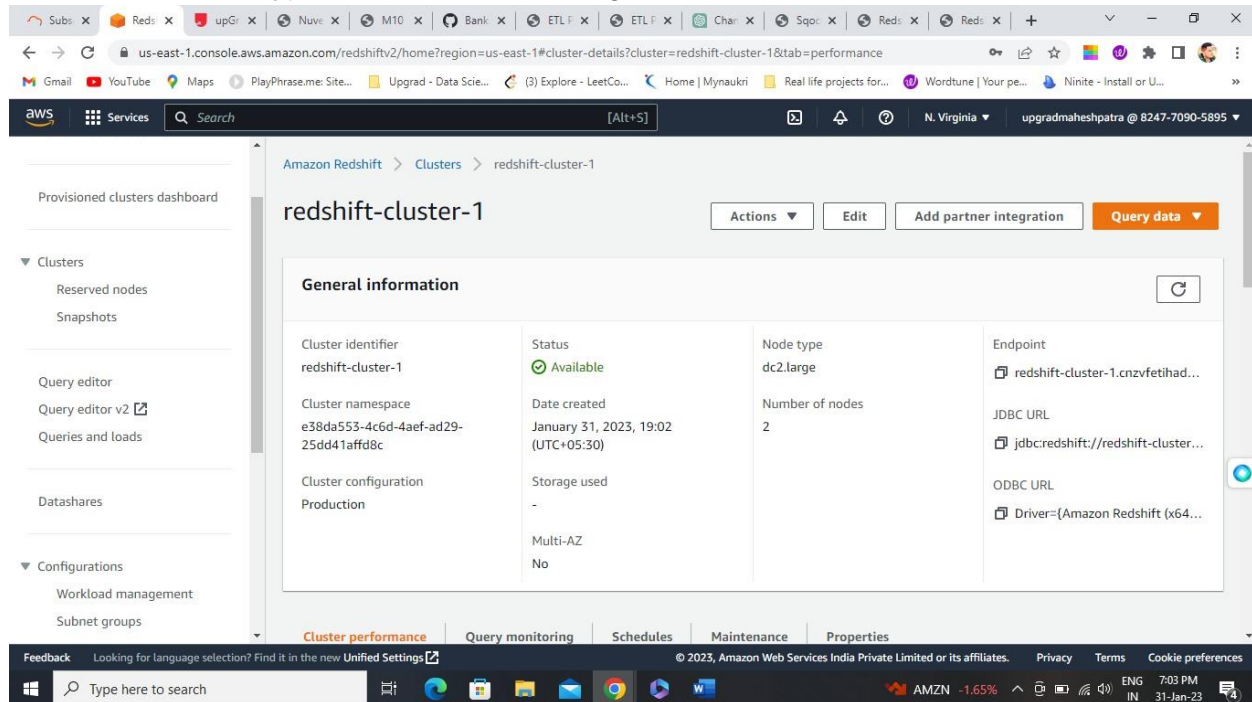


Creation of a Redshift Cluster

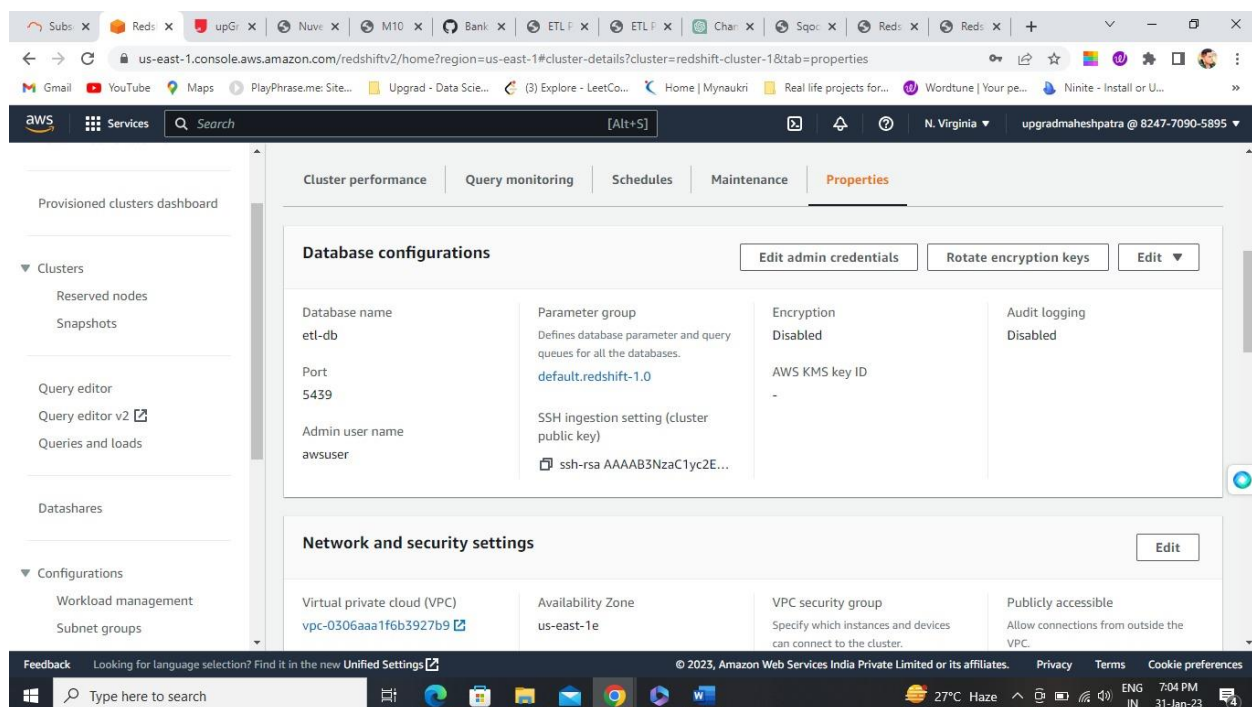
Screenshots of the configuration of the Redshift cluster that you have created:

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



The screenshot shows the AWS Redshift console for a cluster named 'redshift-cluster-1'. The 'General information' tab is selected, displaying the following details:

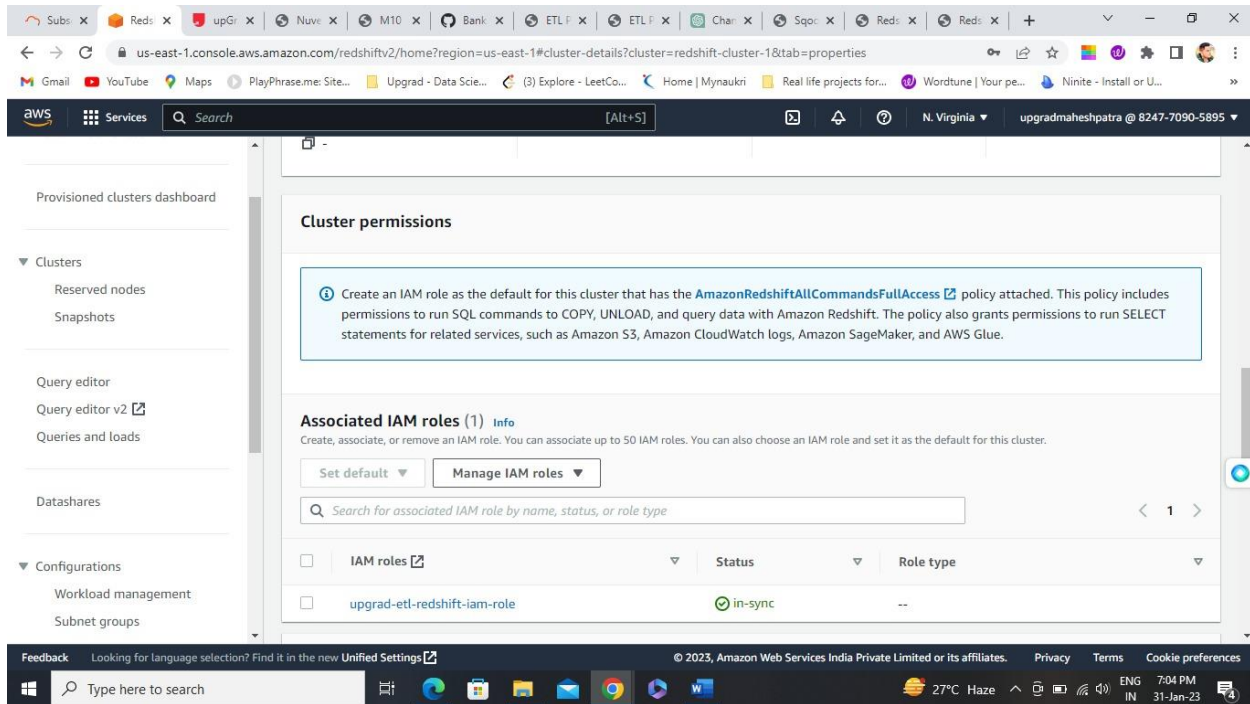
Cluster identifier	Status	Node type	Endpoint
redshift-cluster-1	Available	dc2.large	redshift-cluster-1.cnzfetihad...
Cluster namespace	Date created	Number of nodes	JDBC URL
e38da553-4c6d-4aef-ad29-25dd41affd8c	January 31, 2023, 19:02 (UTC+05:30)	2	jdbc:redshift://redshift-cluster...
Cluster configuration	Storage used		ODBC URL
Production	-		Driver={Amazon Redshift (x64...
	Multi-AZ		
	No		



The screenshot shows the AWS Redshift console for the same cluster, 'redshift-cluster-1', with the 'Properties' tab selected. The 'Database configurations' and 'Network and security settings' are visible.

Database name	Parameter group	Encryption	Audit logging
ett-db	default.redshift-1.0	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5439	ssh-rsa AAAAB3NzaC1yc2E...	-	
Admin user name			
awsuser			

Virtual private cloud (VPC)	Availability Zone	VPC security group	Publicly accessible
vpc-0306aaa1f6b3927b9	us-east-1e	Specify which instances and devices can connect to the cluster.	Allow connections from outside the VPC.



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.

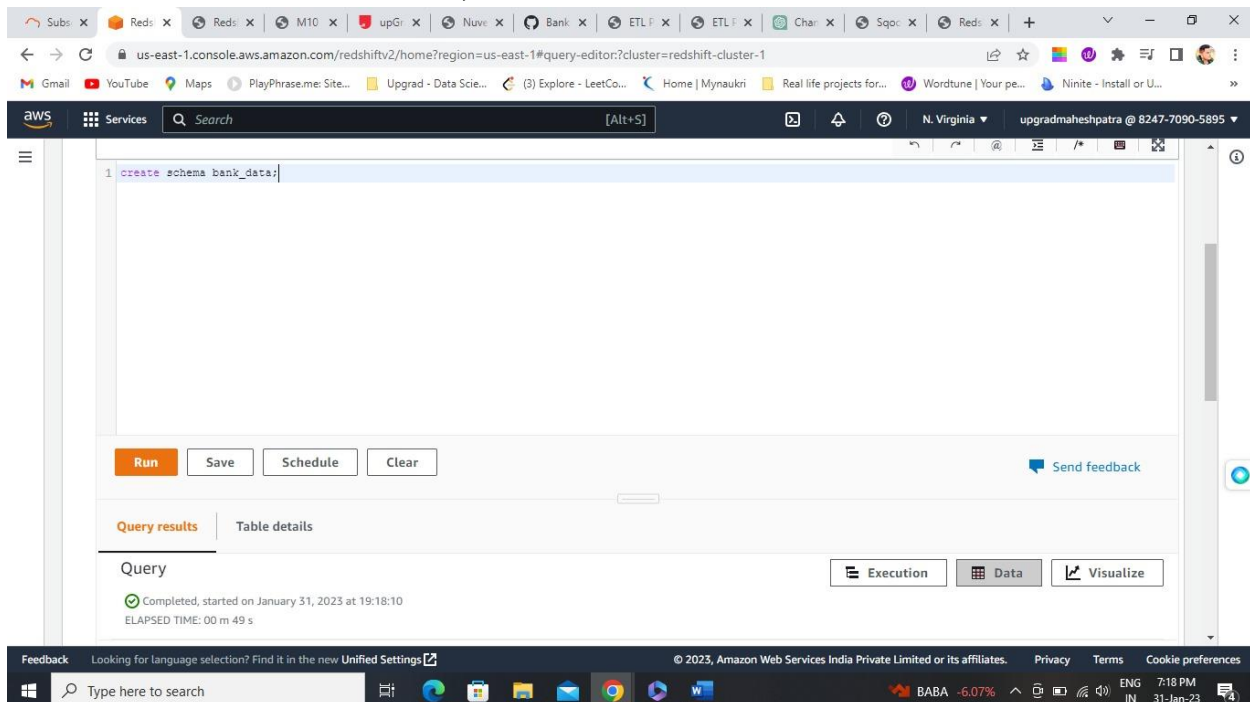
Set default Manage IAM roles

Search for associated IAM role by name, status, or role type

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	upgrad-etl-redshift-iam-role	in-sync	--

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

create schema bank_data;



Query editor

1 create schema bank_data;

Run Save Schedule Clear

Query results Table details

Query

Completed, started on January 31, 2023 at 19:18:10
ELAPSED TIME: 00 m 49 s

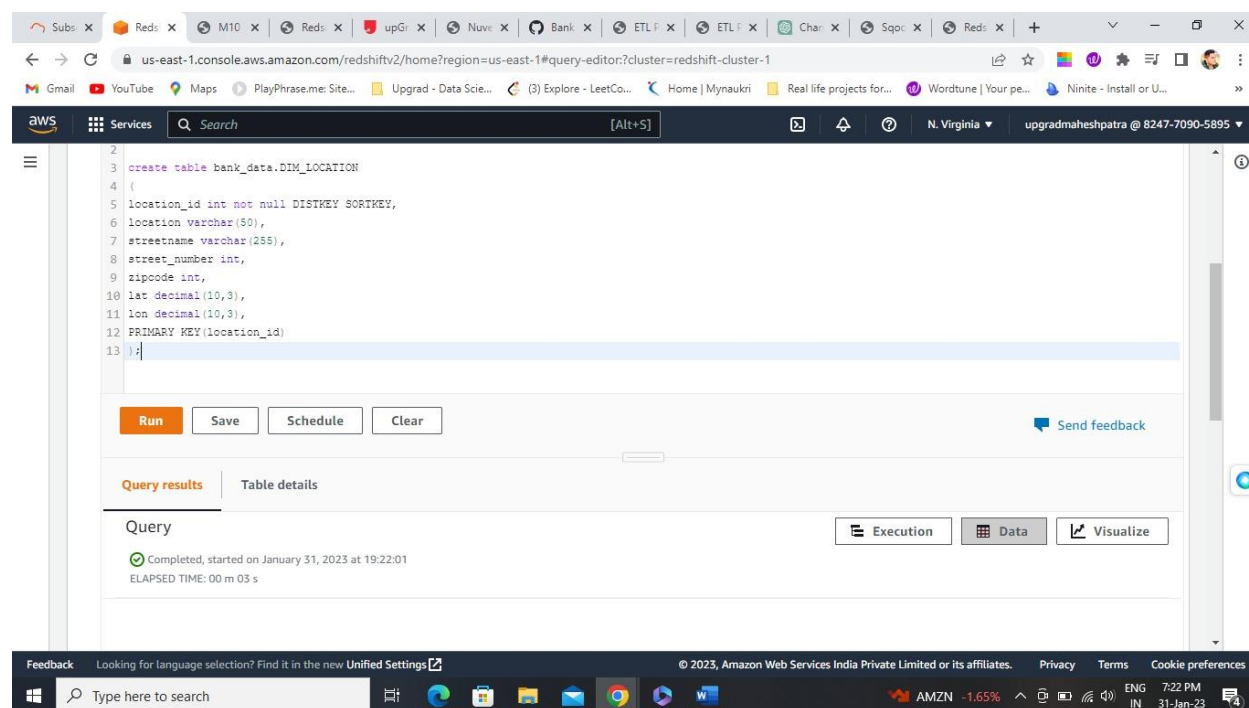
Execution Data Visualize

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

<Queries>

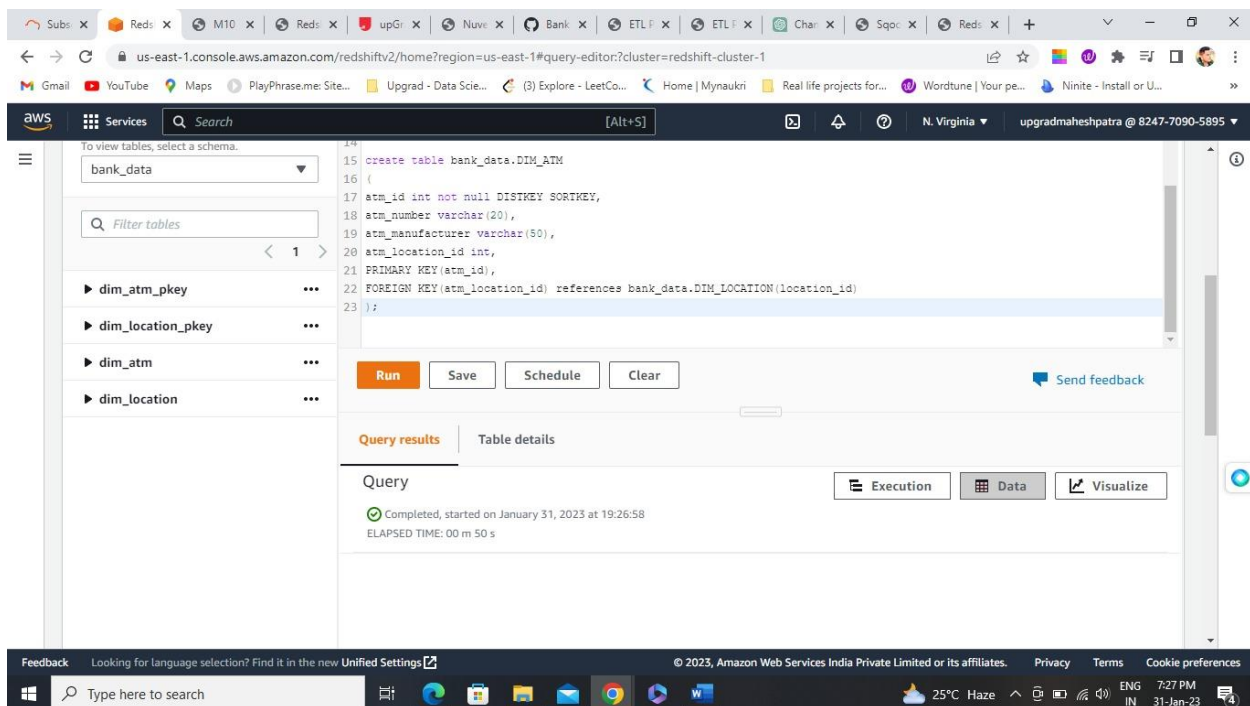
- LOCATION DIMENSION TABLE

```
create table bank_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)
);
```



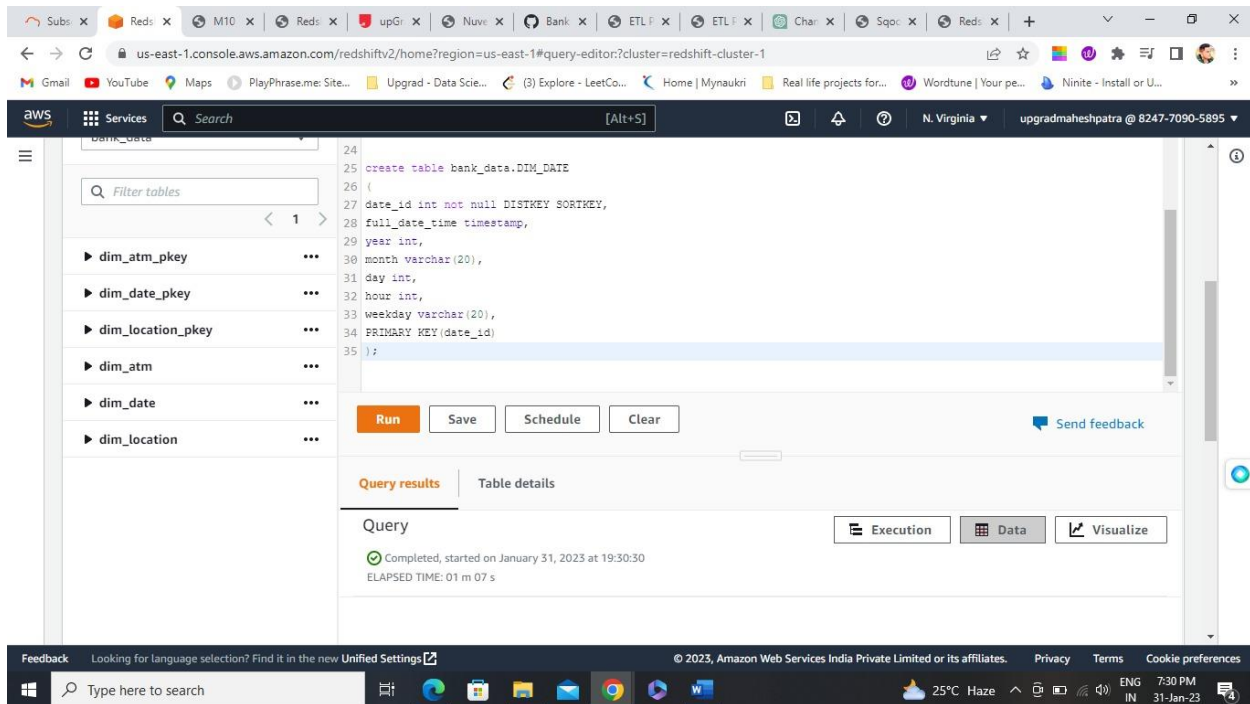
- **ATM DIMENSION TABLE**

```
create table bank_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 bank_data.DIM_LOCATION(location_id)  
);
```



- DATE DIMENSION TABLE

```
create table bank_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)  
);
```



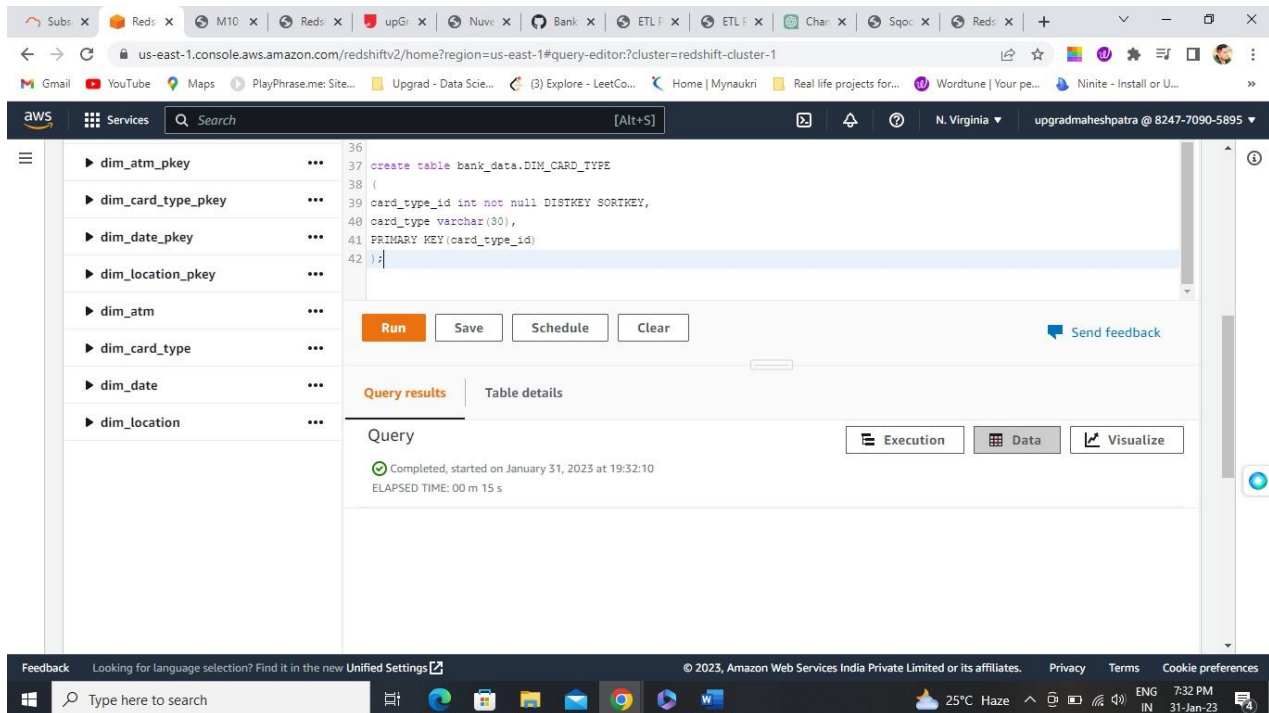
The screenshot displays the AWS Redshift console interface. On the left, a sidebar lists various tables including 'dim_atm_pkey', 'dim_date_pkey', 'dim_location_pkey', 'dim_atm', 'dim_date', and 'dim_location'. The main area shows a SQL query editor with the following code:

```
24  
25 create table bank_data.DIM_DATE  
26 (  
27   date_id int not null DISTKEY SORTKEY,  
28   full_date_time timestamp,  
29   year int,  
30   month varchar(20),  
31   day int,  
32   hour int,  
33   weekday varchar(20),  
34   PRIMARY KEY(date_id)  
35 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing a status message: 'Completed, started on January 31, 2023 at 19:30:30' and 'ELAPSED TIME: 01 m 07 s'. The bottom of the screen shows the Windows taskbar with the date '31-Jan-23' and time '7:30 PM'.

- CARD TYPE DIMENSION TABLE

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



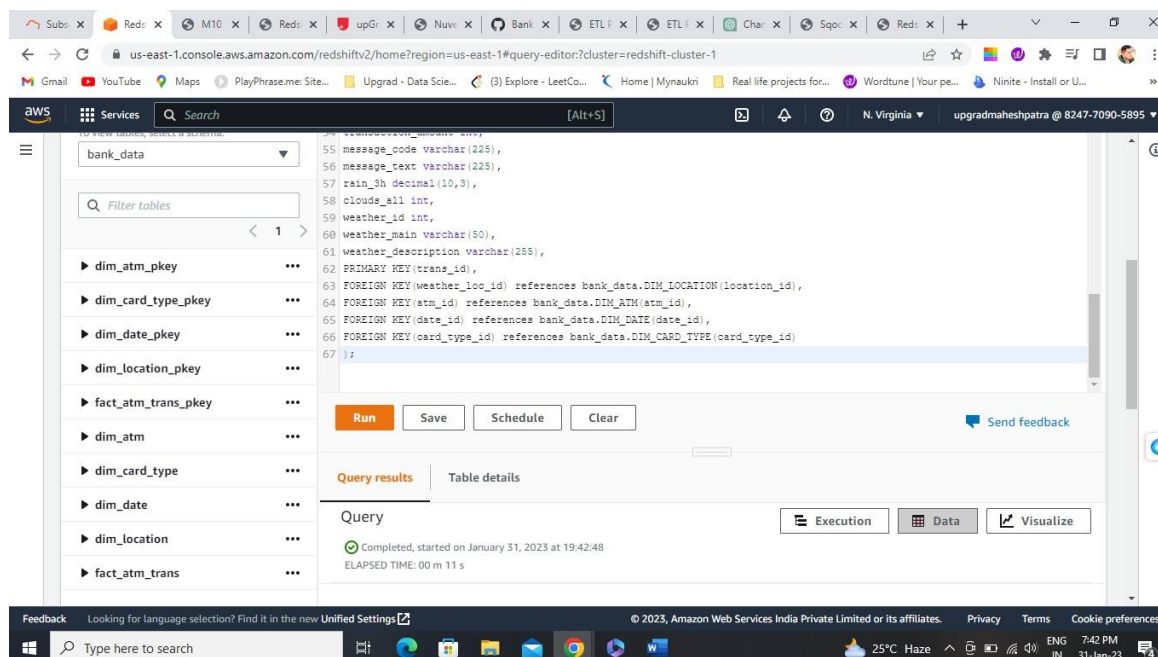
The screenshot shows the AWS Redshift console interface. On the left, a sidebar lists various dimension tables: dim_atm_pkey, dim_card_type_pkey, dim_date_pkey, dim_location_pkey, dim_atm, dim_card_type, dim_date, and dim_location. The main panel displays a SQL query editor with the following code:

```
36  
37 create table bank_data.DIM_CARD_TYPE  
38 (  
39   card_type_id int not null DISTKEY SORTKEY,  
40   card_type varchar(30),  
41   PRIMARY KEY (card_type_id)  
42 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is selected, showing a status message: 'Query Completed, started on January 31, 2023 at 19:32:10 ELAPSED TIME: 00 m 15 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'. The bottom of the screen shows the Windows taskbar with the search bar and various application icons.

- ATM TRANSACTIONS FACT TABLE

```
create table bank_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 bank_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references bank_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references bank_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references bank_data.DIM_CARD_TYPE(card_type_id)
);
```



The screenshot shows the AWS Redshift console interface. The SQL query for creating the `FACT_ATM_TRANS` table is entered in the editor. The query includes a primary key on `trans_id` and foreign keys referencing `DIM_LOCATION`, `DIM_ATM`, `DIM_DATE`, and `DIM_CARD_TYPE`. The 'Query results' tab indicates the query was completed successfully on January 31, 2023, at 19:42:48, with an elapsed time of 00 m 11 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

<Queries>

- Copying the data to dim_atm table

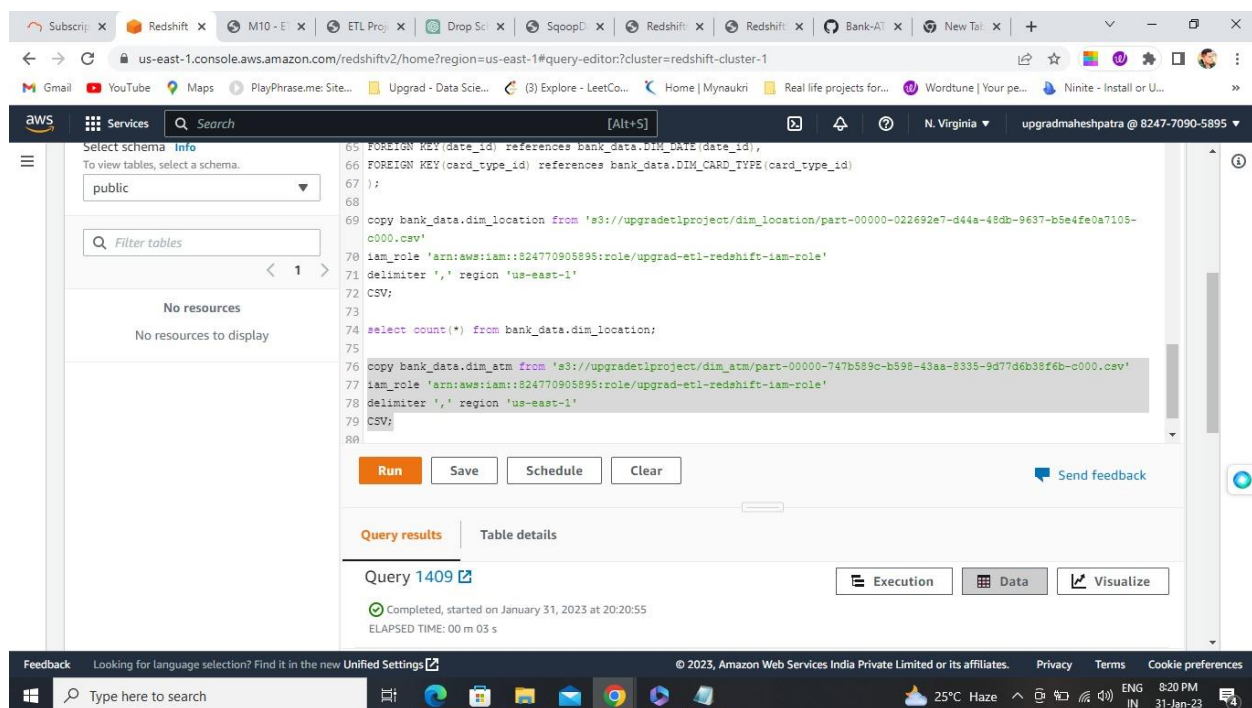
select count(*) from bank_data.dim_location;

copy bank_data.dim_atm from 's3://upgradetlproject/dim_atm/part-00000-747b589c-b598-43aa-8335-9d77d6b38f6b-c000.csv'

iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'

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

CSV;



The screenshot shows the AWS Redshift console interface. The left sidebar displays the 'Select schema' dropdown set to 'public' and a 'Filter tables' search bar. The main area contains a SQL editor with the following queries:

```

65 FOREIGN KEY(date_id) references bank_data.DIM_DATE(date_id),
66 FOREIGN KEY(card_type_id) references bank_data.DIM_CARD_TYPE(card_type_id)
67 );
68
69 copy bank_data.dim_location from 's3://upgradetlproject/dim_location/part-00000-022692e7-d44a-48db-9637-b5e4fe0a7105-
70 c000.csv'
71 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
72 delimiter ',' region 'us-east-1'
73 CSV;
74
75 select count(*) from bank_data.dim_location;
76
77 copy bank_data.dim_atm from 's3://upgradetlproject/dim_atm/part-00000-747b589c-b598-43aa-8335-9d77d6b38f6b-c000.csv'
78 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
79 delimiter ',' region 'us-east-1'
80 CSV;
81

```

Below the 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 the following information:

- Query 1409
- Completed, started on January 31, 2023 at 20:20:55
- ELAPSED TIME: 00 m 03 s

The bottom of the screenshot shows the Windows taskbar with the search bar and system tray.

- Copying the data to dim_date table

```
select count(*) from bank_data.dim_atm;
```

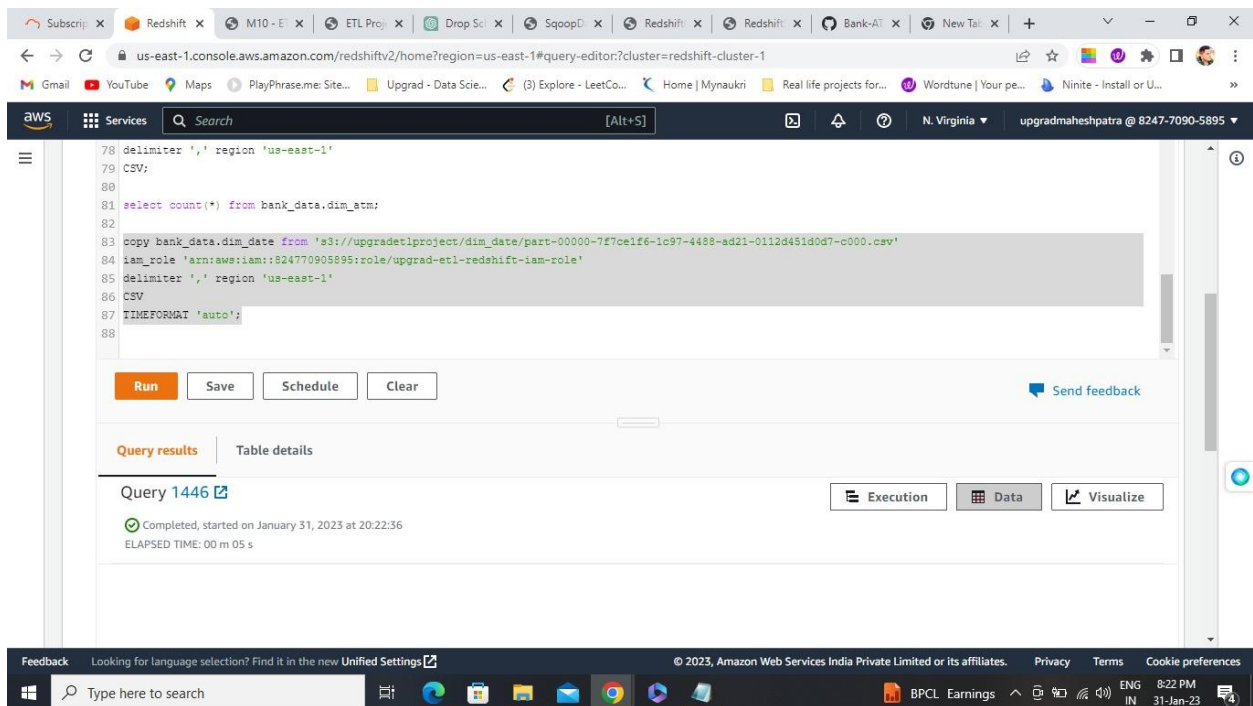
```
copy bank_data.dim_date from 's3://upgradetlproject/dim_date/part-00000-7f7ce1f6-1c97-4488-ad21-0112d451d0d7-c000.csv'
```

```
iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
```

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

```
CSV
```

```
TIMEFORMAT 'auto';
```



The screenshot displays the AWS Redshift console interface. At the top, there's a navigation bar with the AWS logo and search bar. Below it, the 'Query editor' tab is active, showing a SQL script. The script includes a count query for 'bank_data.dim_atm' and a 'copy' command to load data from an S3 bucket into 'bank_data.dim_date'. The 'copy' command specifies an IAM role, delimiter, region, CSV format, and TIMEFORMAT. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button has been clicked, and the 'Query results' tab is selected, showing 'Query 1446' as 'Completed' with a status icon. The execution details indicate it started on January 31, 2023, at 20:22:36 and took 00 m 05 s to complete. The bottom of the screen shows the Windows taskbar with various application icons and the system clock.

- Copying the data to dim_card_type table

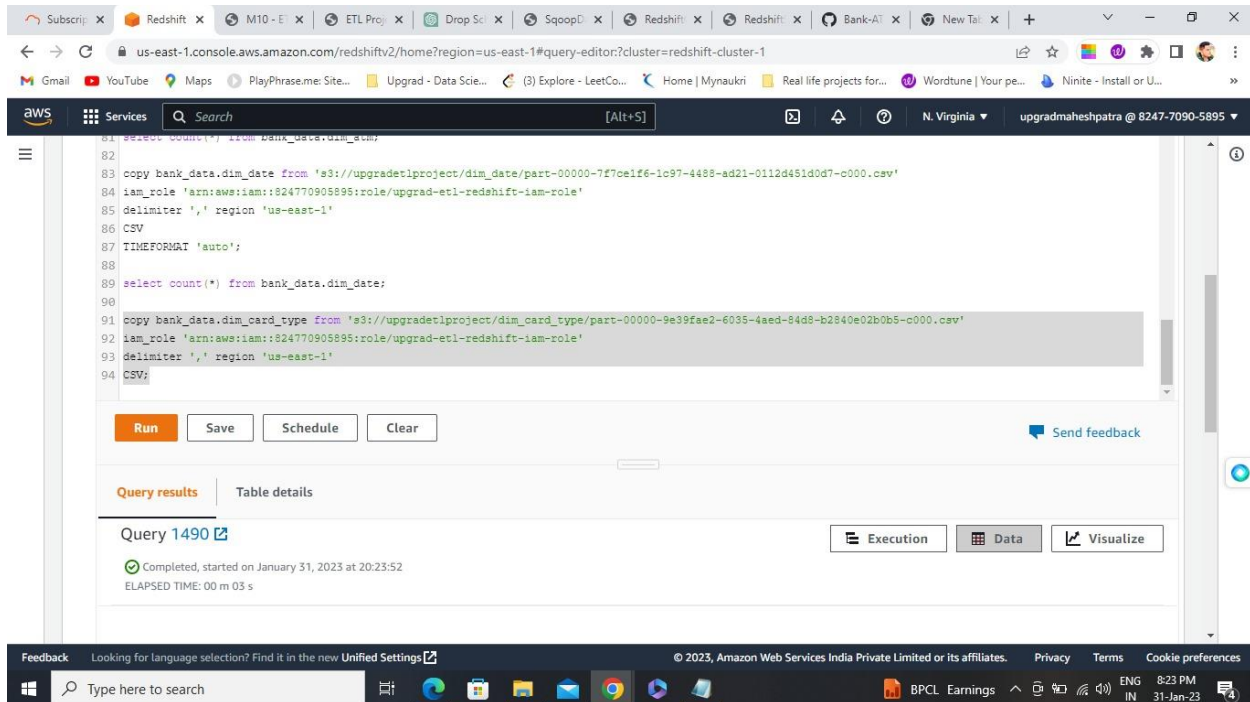
select count(*) from bank_data.dim_date;

copy bank_data.dim_card_type from 's3://upgradetlproject/dim_card_type/part-00000-9e39fae2-6035-4aed-84d8-b2840e02b0b5-c000.csv'

iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'

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

CSV;



The screenshot shows the AWS Redshift console interface. The top navigation bar includes the AWS logo, a search bar, and the user's profile. The main content area displays a SQL query editor with the following code:

```

81 select count(*) from bank_data.dim_date;
82
83 copy bank_data.dim_date from 's3://upgradetlproject/dim_date/part-00000-7f70e1f6-1c97-4488-ad21-0112d451d0d7-c000.csv'
84 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
85 delimiter ',' region 'us-east-1'
86 CSV
87 TIMEFORMAT 'auto';
88
89 select count(*) from bank_data.dim_date;
90
91 copy bank_data.dim_card_type from 's3://upgradetlproject/dim_card_type/part-00000-9e39fae2-6035-4aed-84d8-b2840e02b0b5-c000.csv'
92 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
93 delimiter ',' region 'us-east-1'
94 CSV;

```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link. Below the buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing the query ID 'Query 1490' and its execution status: 'Completed, started on January 31, 2023 at 20:23:52' with an 'ELAPSED TIME: 00 m 03 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'. The bottom of the screen shows the Windows taskbar with various application icons and the system clock indicating 8:23 PM on 31-Jan-23.

- Copying the data to fact_atm_trans table

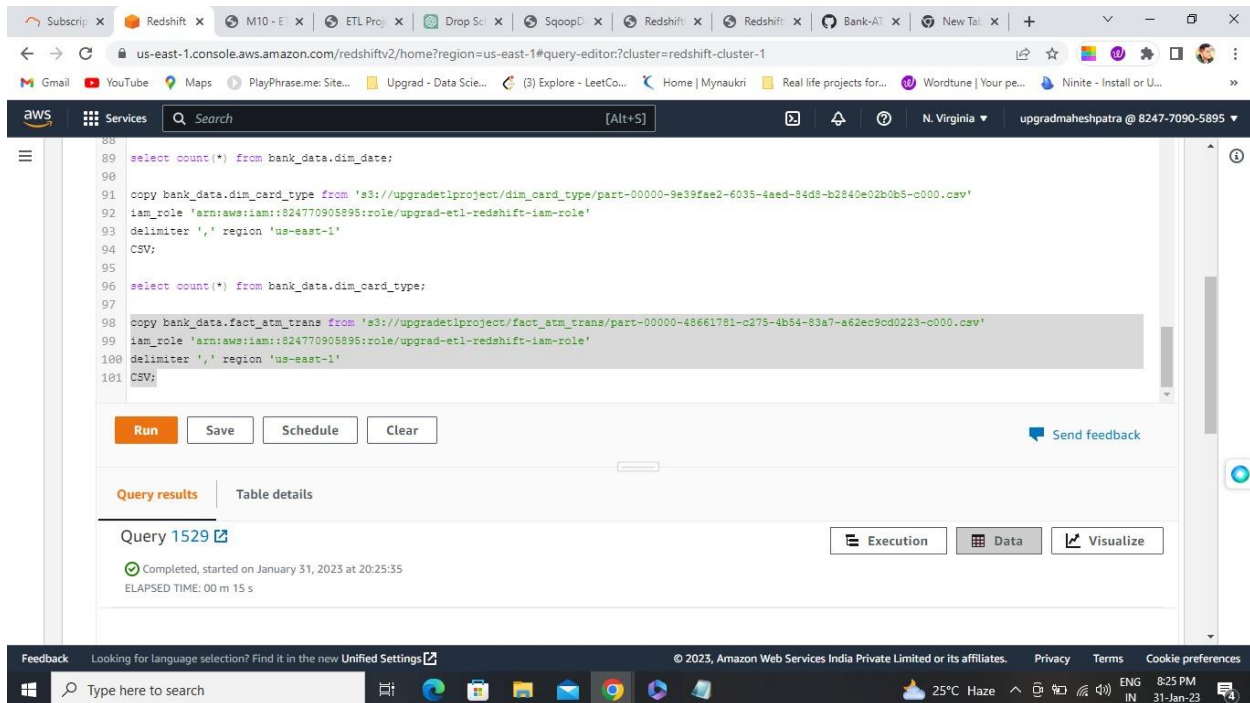
```
select count(*) from bank_data.dim_card_type;
```

```
copy bank_data.fact_atm_trans from 's3://upgradetlproject/fact_atm_trans/part-00000-48661781-c275-4b54-83a7-a62ec9cd0223-c000.csv'
```

```
iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
```

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

```
CSV;
```



The screenshot shows the AWS Redshift Query Editor interface. The SQL editor contains the following queries:

```

88 select count(*) from bank_data.dim_date;
89
90
91 copy bank_data.dim_card_type from 's3://upgradetlproject/dim_card_type/part-00000-9a39fae2-6035-4aed-84d8-b2840e02b0b5-c000.csv'
92 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
93 delimiter ',' region 'us-east-1'
94 CSV;
95
96 select count(*) from bank_data.dim_card_type;
97
98 copy bank_data.fact_atm_trans from 's3://upgradetlproject/fact_atm_trans/part-00000-48661781-c275-4b54-83a7-a62ec9cd0223-c000.csv'
99 iam_role 'arn:aws:iam::824770905895:role/upgrad-etl-redshift-iam-role'
100 delimiter ',' region 'us-east-1'
101 CSV;
  
```

Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link.

Below the buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing the following information:

- Query 1529
- Completed, started on January 31, 2023 at 20:25:35
- ELAPSED TIME: 00 m 15 s

At the bottom of the interface, there are buttons for 'Execution', 'Data', and 'Visualize'. The 'Data' button is highlighted.

The bottom of the screenshot shows the Windows taskbar with the search bar, taskbar icons, and system tray information (25°C Haze, 8:25 PM, 31-Jan-23).