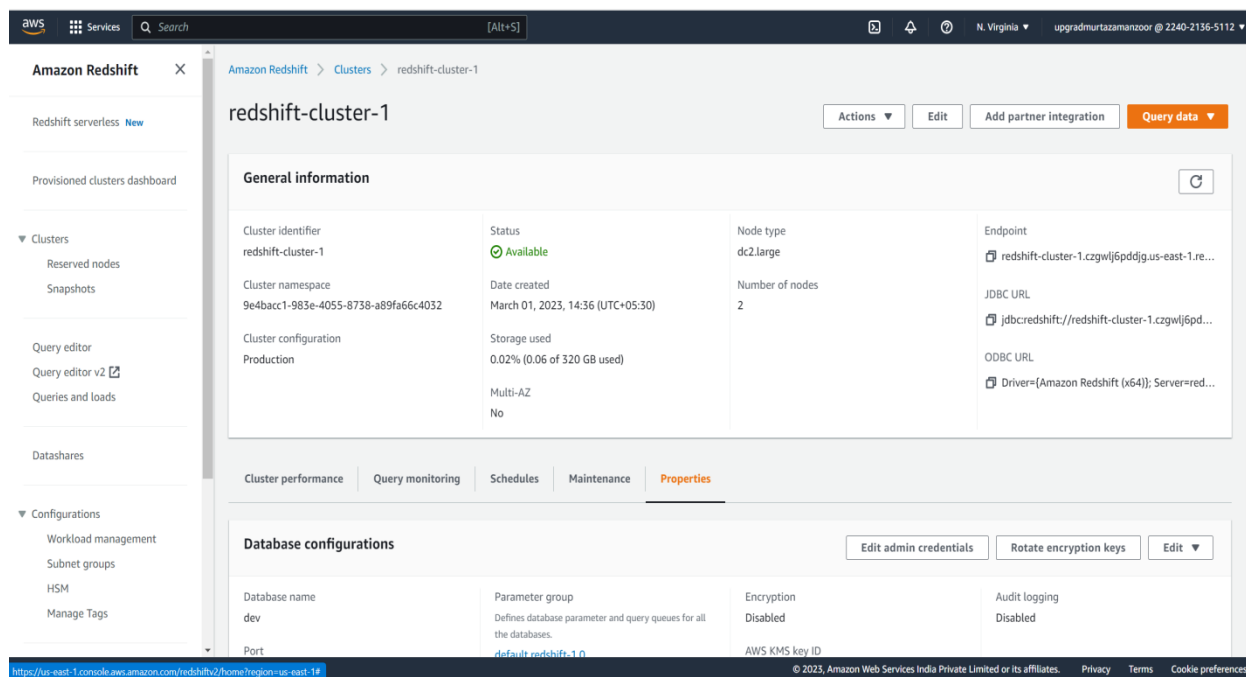
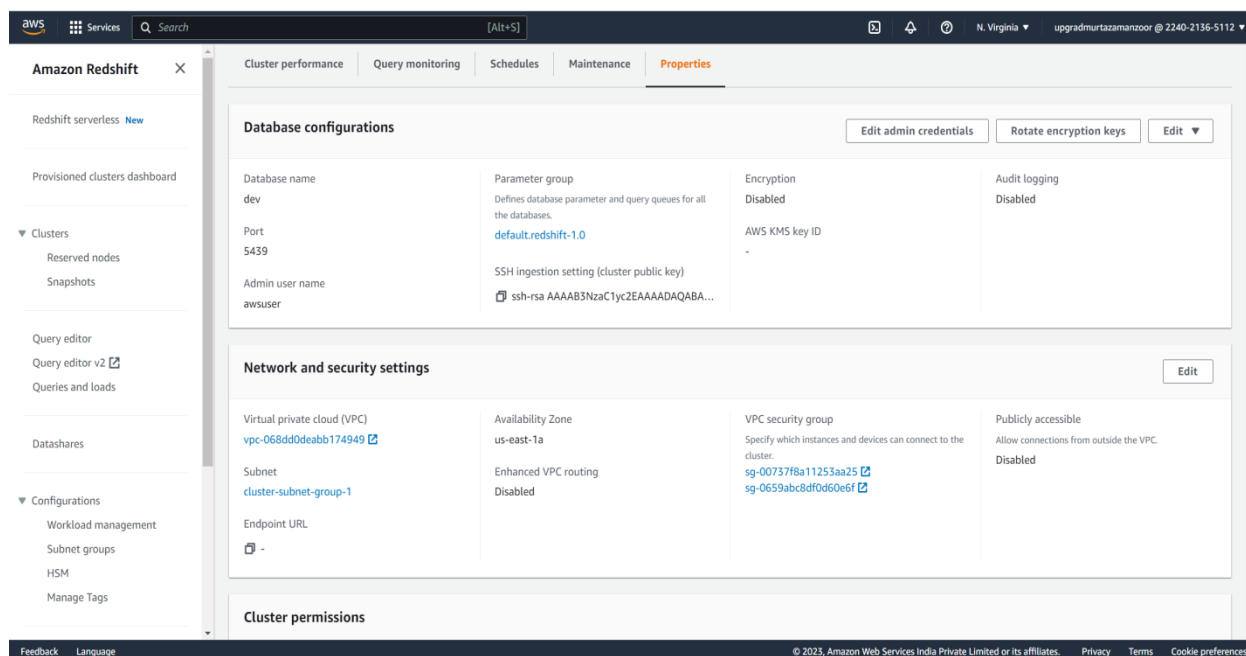


## Creation of a Redshift Cluster

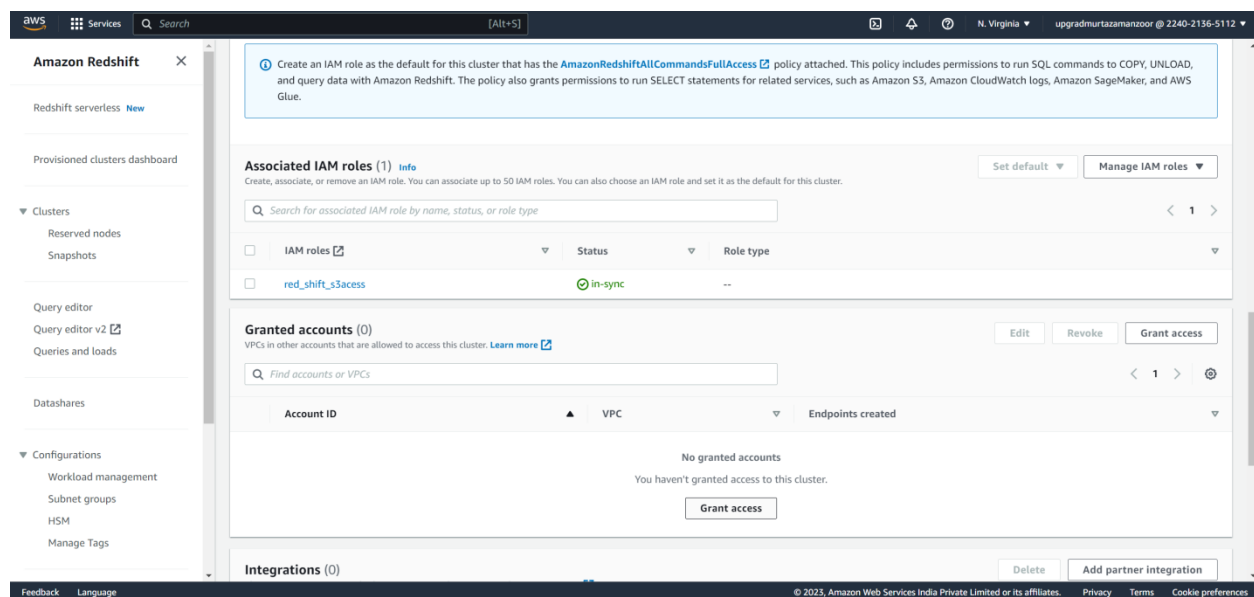
Screenshots of the configuration of the Redshift cluster that has been created:



The screenshot shows the Amazon Redshift console interface. The left sidebar contains navigation options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Datashares', and 'Configurations'. The main content area displays the 'redshift-cluster-1' cluster details under the 'General information' tab. The cluster status is 'Available'. Key details include: Cluster identifier (redshift-cluster-1), Cluster namespace (9e4bacc1-983e-4055-8738-a89fa66c4032), Node type (dc2.large), Number of nodes (2), Date created (March 01, 2023, 14:36 UTC+05:30), Storage used (0.02% of 320 GB), and Multi-AZ (No). The endpoint is redshift-cluster-1.czgwij6pddjg.us-east-1.re... and the JDBC URL is jdbc:redshift://redshift-cluster-1.czgwij6pd... The ODBC URL is Driver={Amazon Redshift (x64)}; Server=red... The bottom of the console shows the URL https://us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1# and the copyright notice for Amazon Web Services India Private Limited.



The screenshot shows the Amazon Redshift console interface, specifically the 'Properties' tab for the 'redshift-cluster-1' cluster. The left sidebar is the same as the previous screenshot. The main content area displays the 'Database configurations' and 'Network and security settings' sections. The 'Database configurations' section includes: Database name (dev), Parameter group (default.redshift-1.0), Encryption (Disabled), Audit logging (Disabled), Port (5439), and Admin user name (awsuser). The 'Network and security settings' section includes: Virtual private cloud (VPC) (vpc-068dd0deabb174949), Subnet (cluster-subnet-group-1), Availability Zone (us-east-1a), Enhanced VPC routing (Disabled), VPC security group (sg-00737f8a11253aa25), and Publicly accessible (Disabled). The bottom of the console shows the URL https://us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1# and the copyright notice for Amazon Web Services India Private Limited.



**Amazon Redshift**

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.

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	red_shift_s3access	in-sync	--

**Granted accounts (0)**

VPCs in other accounts that are allowed to access this cluster. [Learn more](#)

Account ID	VPC	Endpoints created
No granted accounts		

You haven't granted access to this cluster.

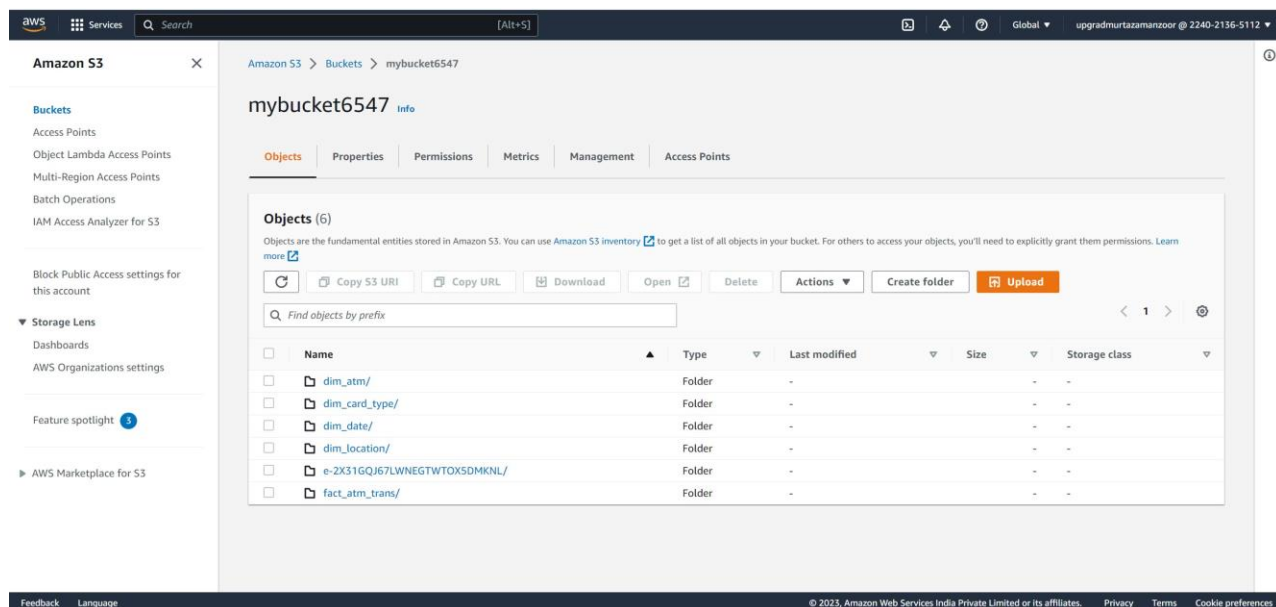
[Grant access](#)

**Integrations (0)**

[Delete](#) [Add partner integration](#)

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

S3 bucket containing Files:



**Amazon S3**

Amazon S3 > Buckets > mybucket6547

**mybucket6547** [Info](#)

[Objects](#) [Properties](#) [Permissions](#) [Metrics](#) [Management](#) [Access Points](#)

**Objects (6)**

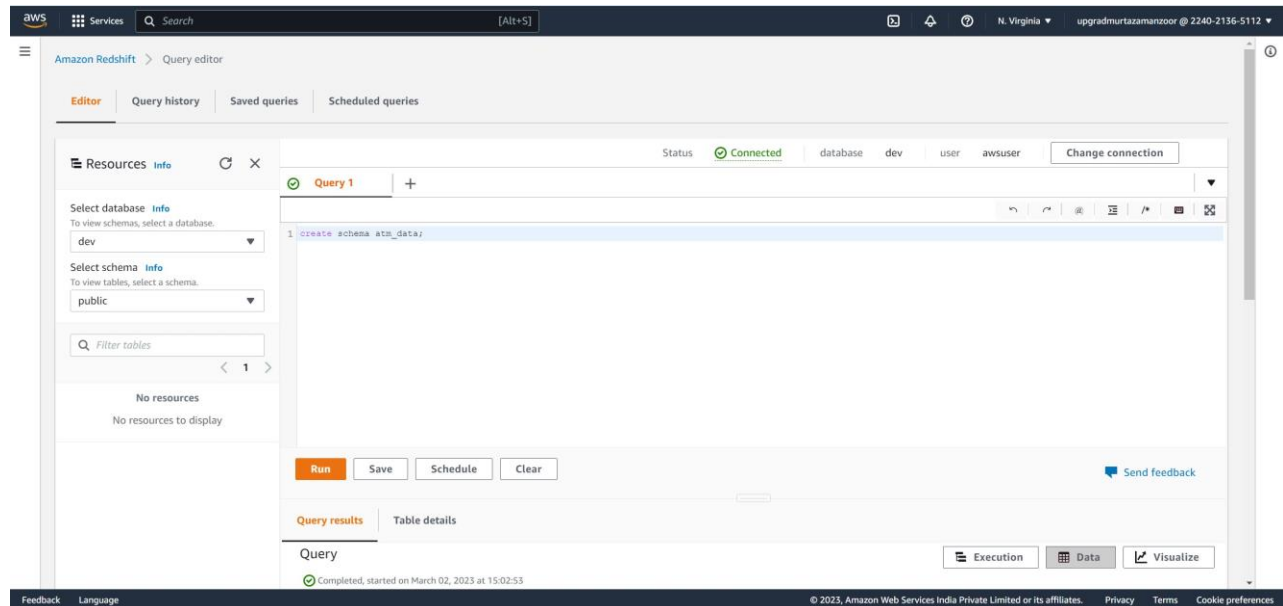
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dim_atm/	Folder	-	-	-
<input type="checkbox"/>	dim_card_type/	Folder	-	-	-
<input type="checkbox"/>	dim_date/	Folder	-	-	-
<input type="checkbox"/>	dim_location/	Folder	-	-	-
<input type="checkbox"/>	e-2X31GQJ67LWNEGWTQXSDMKNL/	Folder	-	-	-
<input type="checkbox"/>	fact_atm_trans/	Folder	-	-	-

Query for creating schema:

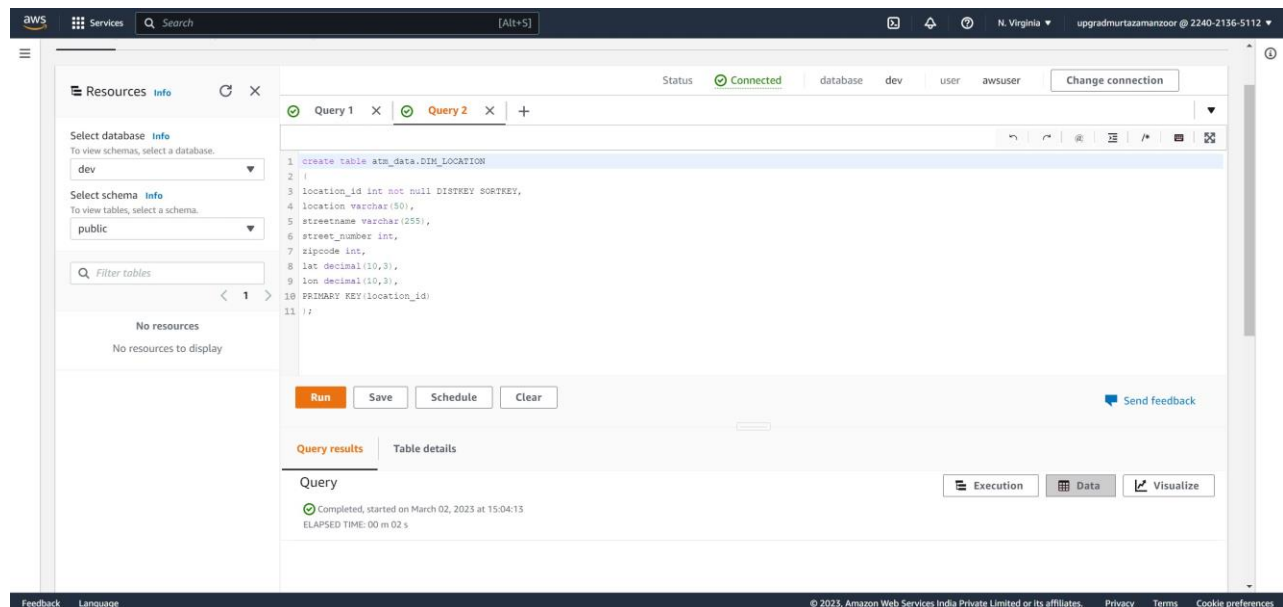
```
create schema atm_data;
```



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

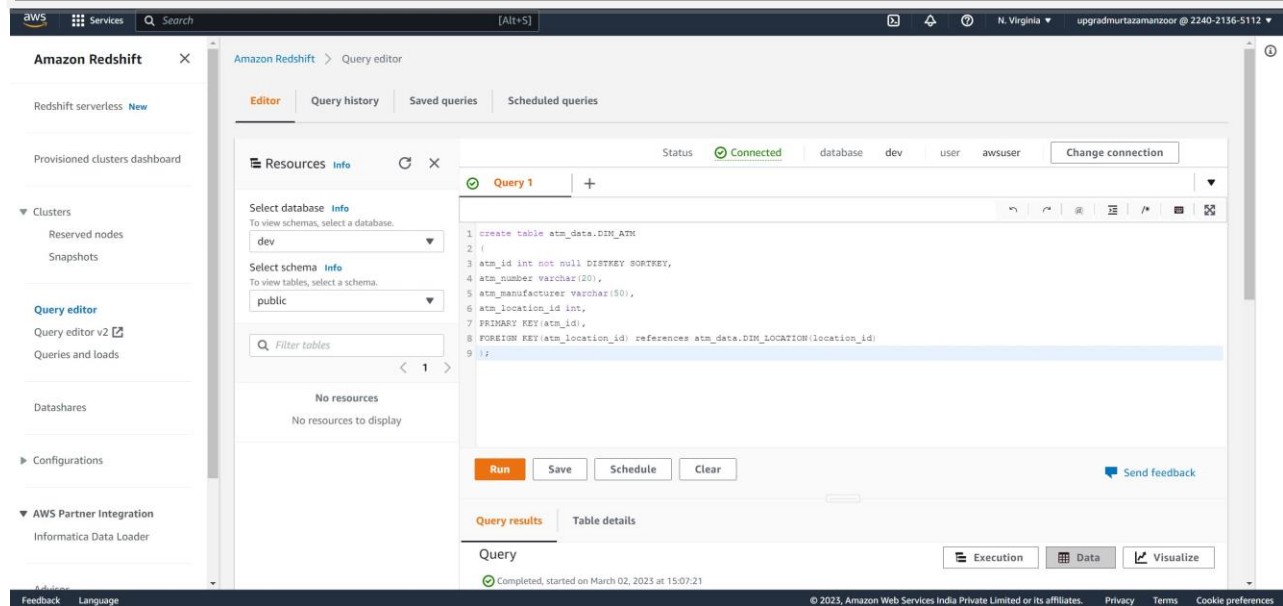
- Creating location dimension table

```
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)
);
```



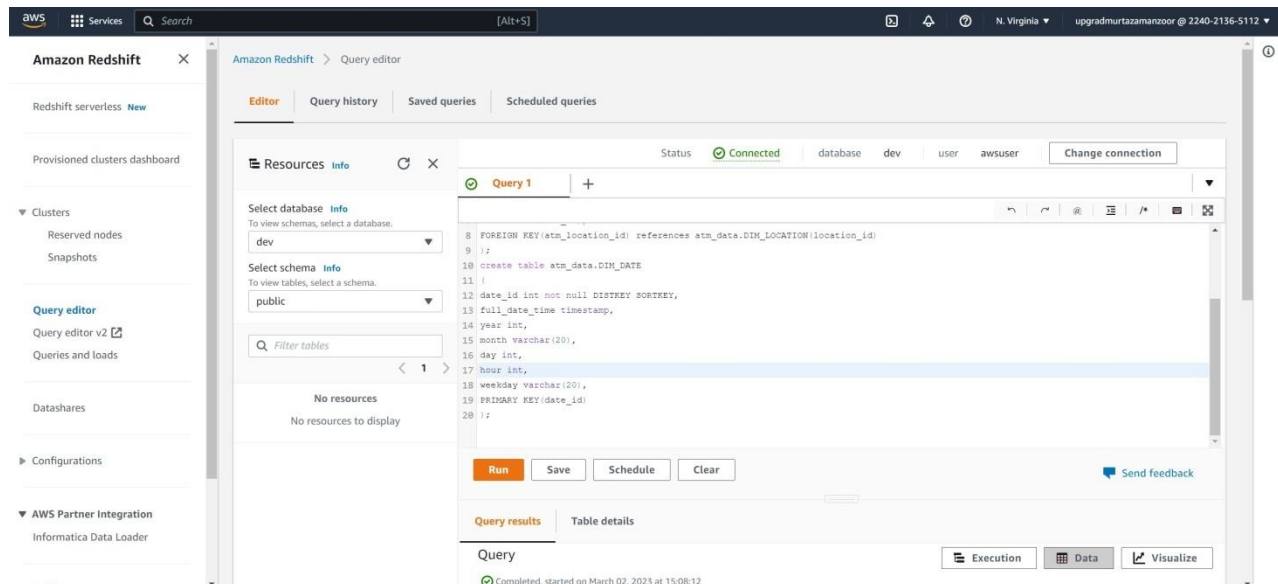
The screenshot displays the Amazon Redshift Query Editor interface. The left sidebar shows the navigation menu with options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', and 'AWS Partner Integration'. The main area is titled 'Amazon Redshift > Query editor' and includes tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. The 'Editor' tab is active, showing a SQL query to create a table named 'DIM\_ATM' in the 'atm\_data' database. The query is as follows:

```
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 ;}
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' section at the bottom shows the query status as 'Completed, started on March 02, 2023 at 15:07:21'.

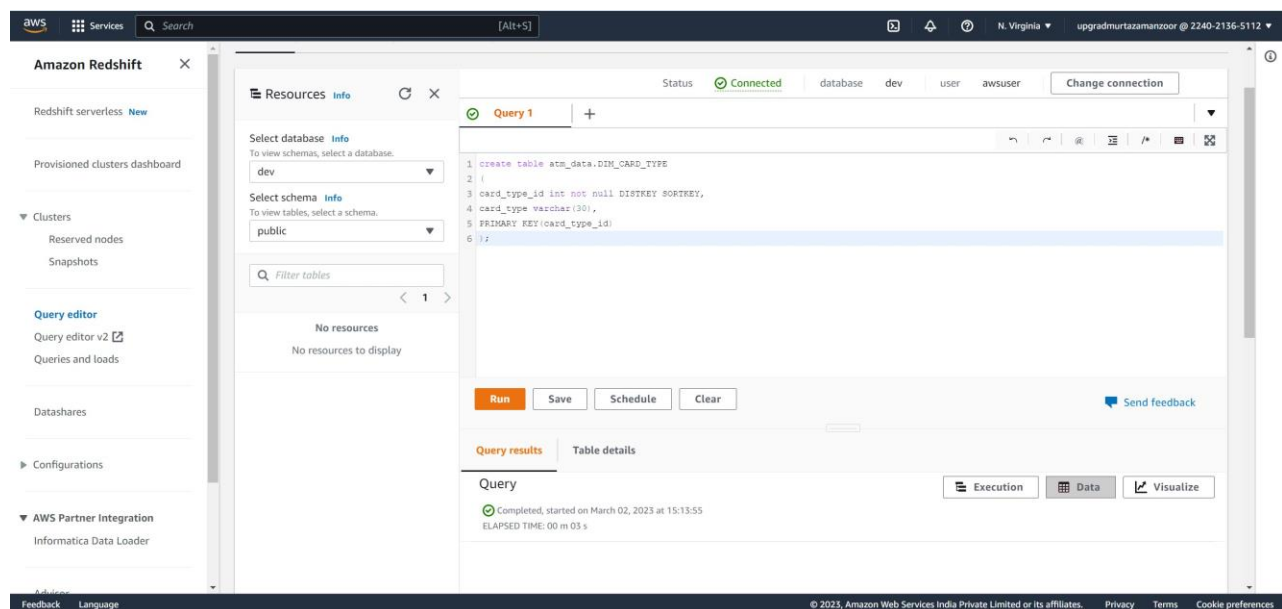
- Creating date dimension table

```
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)
);
```



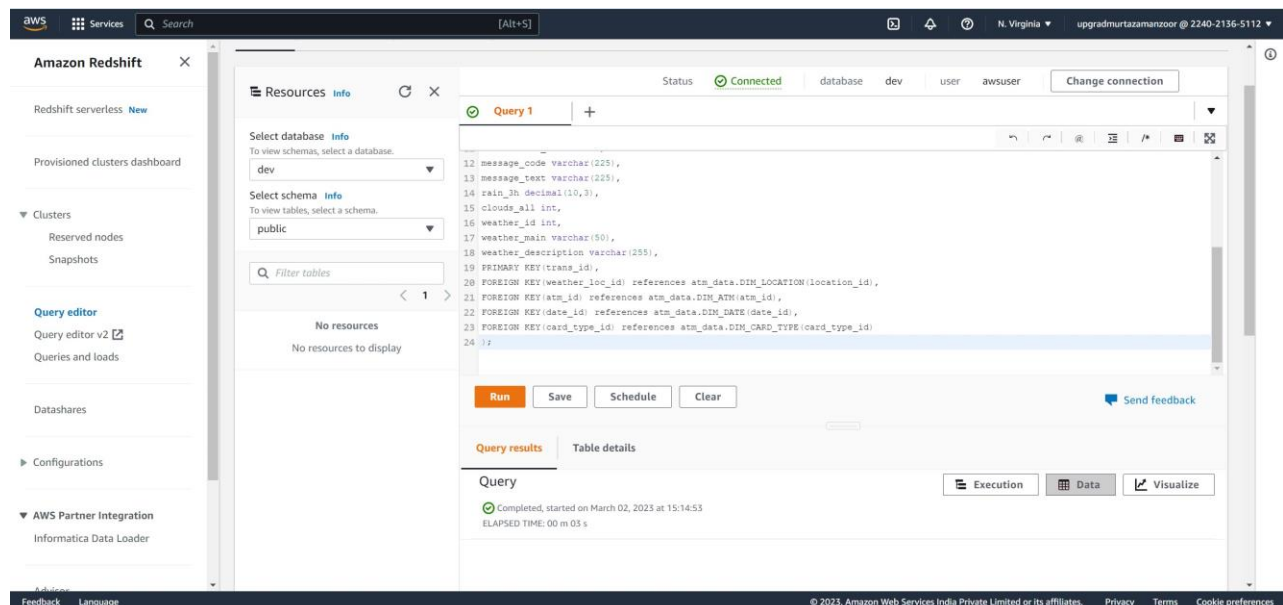
- Creating card type dimension table

```
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

```
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. The left sidebar contains navigation options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Datashares', 'Configurations', and 'AWS Partner Integration'. The main area displays the 'Query 1' editor with the SQL code for creating the 'FACT\_ATM\_TRANS' table. The 'Resources' panel on the left shows the selected database 'dev' and schema 'public'. The 'Query results' panel at the bottom indicates that the query was completed successfully on March 02, 2023, at 15:14:53, with an elapsed time of 00 m 03 s.

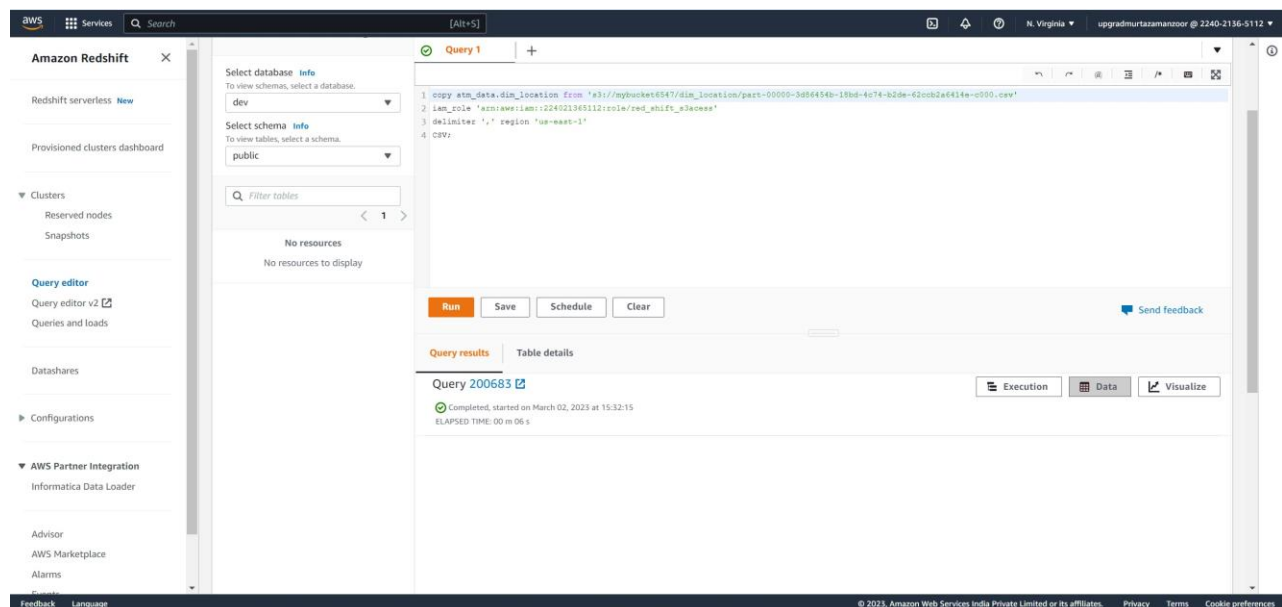


## Loading data into a Redshift cluster from Amazon S3 bucket

### 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://mybucket6547/dim_location/part-00000-3d86454b-18bd-4c74-b2de-62ccb2a6414e-c000.csv'  
iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'  
delimiter ',' region 'us-east-1'  
CSV;
```



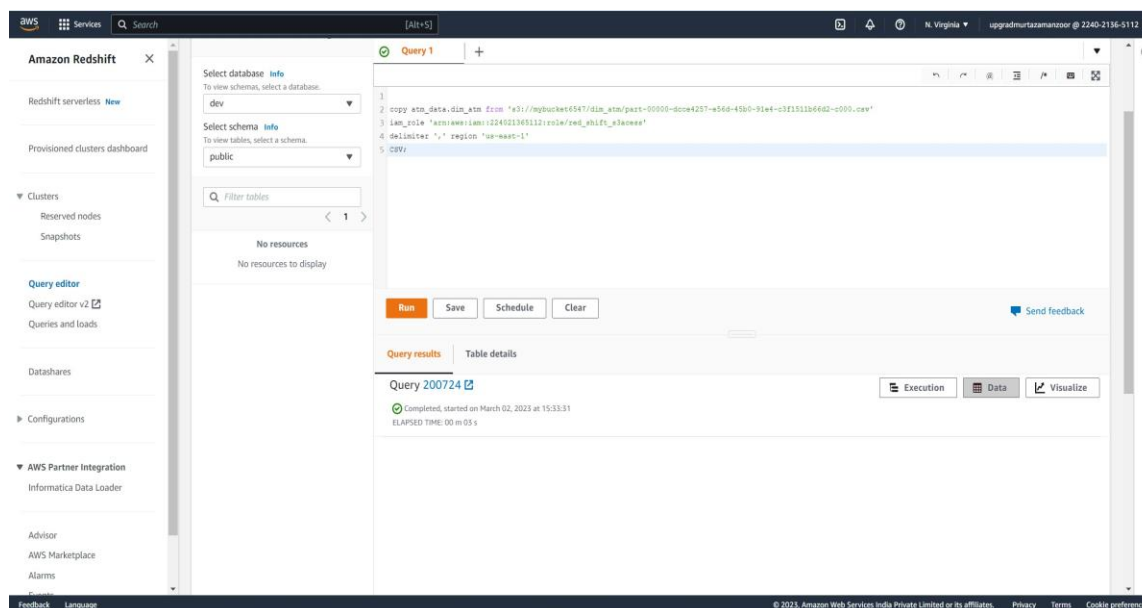
The screenshot displays the AWS Redshift Query Editor interface. On the left, the navigation pane shows the 'Amazon Redshift' section with options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Databases', 'Configurations', and 'AWS Partner Integration'. The main area is divided into three panes. The left pane shows the 'dev' database and 'public' schema. The middle pane contains the SQL query: 

```
1 copy atm_data.dim_location from 's3://mybucket6547/dim_location/part-00000-3d86454b-18bd-4c74-b2de-62ccb2a6414e-c000.csv'  
2 iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

 The right pane shows the 'Query results' tab for 'Query 200683', indicating the query is 'Completed, started on March 02, 2023 at 15:32:15' with an 'ELAPSED TIME: 00 m 06 s'. The bottom of the interface includes a footer with 'Feedback', 'Language', and copyright information for Amazon Web Services India Private Limited.

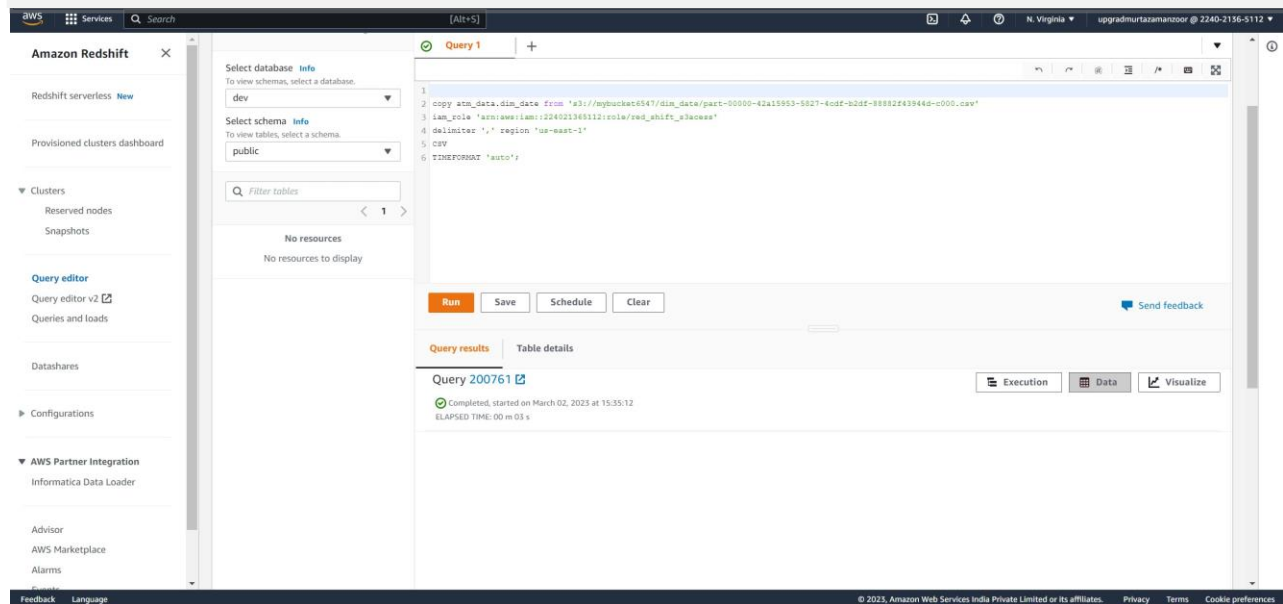
- Copying the data to dim\_atm table

```
copy atm_data.dim_atm from 's3://mybucket6547/dim_atm/part-00000-dcce4257-e56d-45b0-91e4-c3f1511b66d2-c000.csv'
iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
delimiter ',' region 'us-east-1'
CSV;
```



- Copying the data to dim\_date table

```
copy atm_data.dim_date from 's3://mybucket6547/dim_date/part-00000-42a15953-5827-4cdf-b2df-88882f43944d-c000.csv'
iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



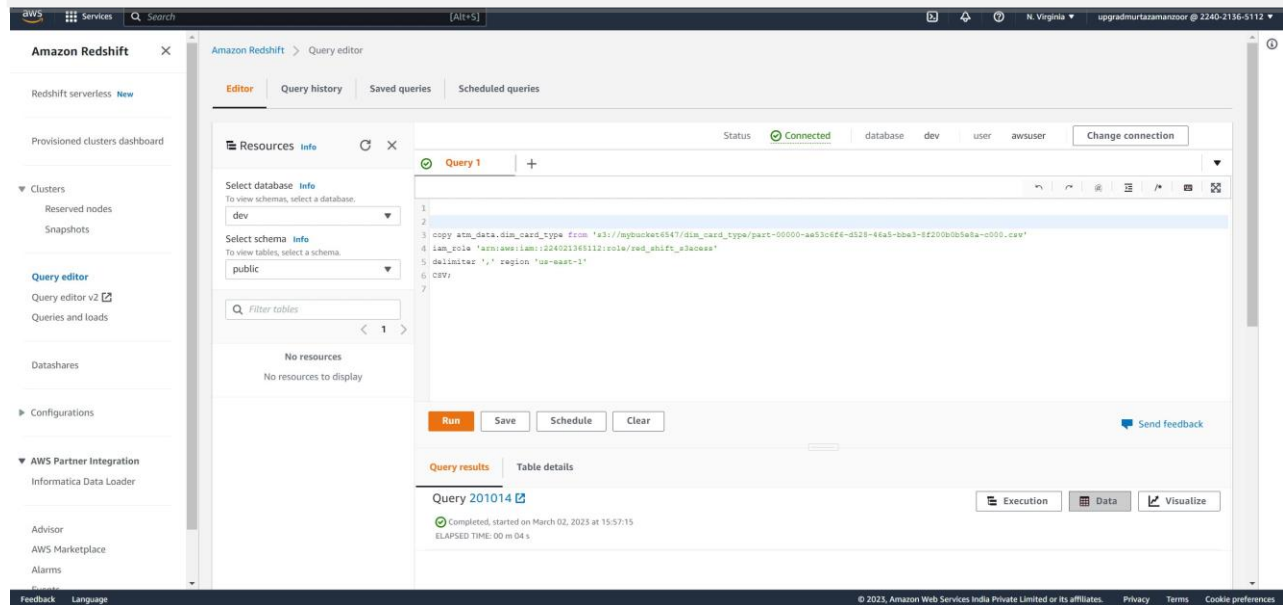
The screenshot shows the Amazon Redshift Query Editor interface. On the left, there is a sidebar with navigation options: Amazon Redshift, Redshift serverless, Provisioned clusters dashboard, Clusters (Reserved nodes, Snapshots), Query editor (Query editor v2, Queries and loads), Databases, Configurations, and AWS Partner Integration (Informatica Data Loader, Advisor, AWS Marketplace, Alarms). The main area is divided into two panes. The left pane shows the 'Select database' dropdown set to 'dev' and the 'Select schema' dropdown set to 'public'. Below these, there is a search bar for tables and a message 'No resources' with 'No resources to display'. The right pane shows a SQL query being executed. The query is:
 

```
1 copy atm_data.dim_date from 's3://mybucket6547/dim_date/part-00000-42a15953-5827-4cdf-b2df-88882f43944d-c000.csv'
2 iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
```

 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 'Query 200761' with a status of 'Completed, started on March 02, 2023 at 15:35:12' and 'ELAPSED TIME: 00 m 03 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://mybucket6547/dim_card_type/part-00000-ae53c6f6-d528-46a5-bbe3-8f200b0b5e8a-c000.csv'
iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
delimiter ',' region 'us-east-1'
CSV;
```



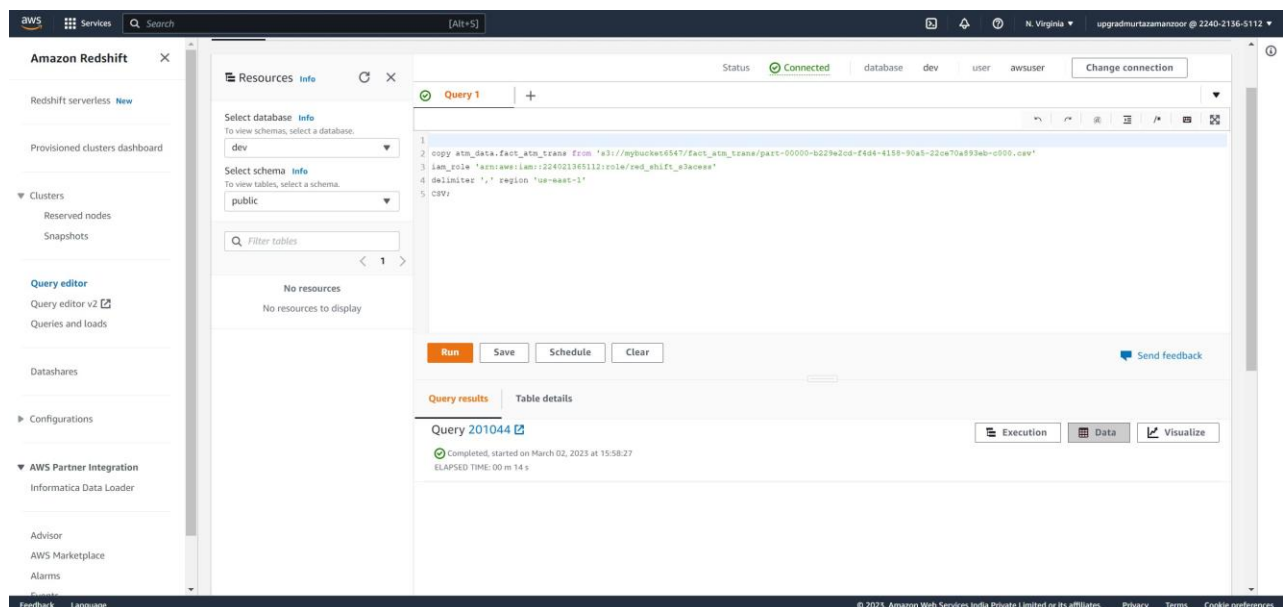
The screenshot shows the Amazon Redshift Query Editor interface. The query being executed is:

```
1
2
3 copy atm_data.dim_card_type from 's3://mybucket6547/dim_card_type/part-00000-ae53c6f6-d528-46a5-bbe3-8f200b0b5e8a-c000.csv'
4 iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
5 delimiter ',' region 'us-east-1'
6 CSV;
7
```

The query has been executed successfully, as indicated by the status "Completed, started on March 02, 2023 at 15:57:15" and "ELAPSED TIME: 00 m 04 s".

- Copying the data to fact\_atm\_trans table

```
copy atm_data.fact_atm_trans from 's3://mybucket6547/fact_atm_trans/part-00000-
b229e2cd-f4d4-4158-90a5-22ce70a893eb-c000.csv'
iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot displays the Amazon Redshift Query Editor interface. On the left, the navigation pane shows the 'Query editor' section. The main area is divided into three panes: 'Resources' (showing 'dev' database and 'public' schema), 'Query 1' (containing the SQL query), and 'Query results' (showing the execution status).

The SQL query in the 'Query 1' pane is:

```
1 copy atm_data.fact_atm_trans from 's3://mybucket6547/fact_atm_trans/part-00000-b229e2cd-f4d4-4158-90a5-22ce70a893eb-c000.csv'
2
3 iam_role 'arn:aws:iam::224021365112:role/red_shift_s3access'
4 delimiter ',' region 'us-east-1'
5 CSV;
```

The 'Query results' pane shows that the query was 'Completed, started on March 02, 2023 at 15:58:27' with an 'ELAPSED TIME: 00 m 14 s'. Below the results, there are tabs for 'Execution', 'Data', and 'Visualize'.