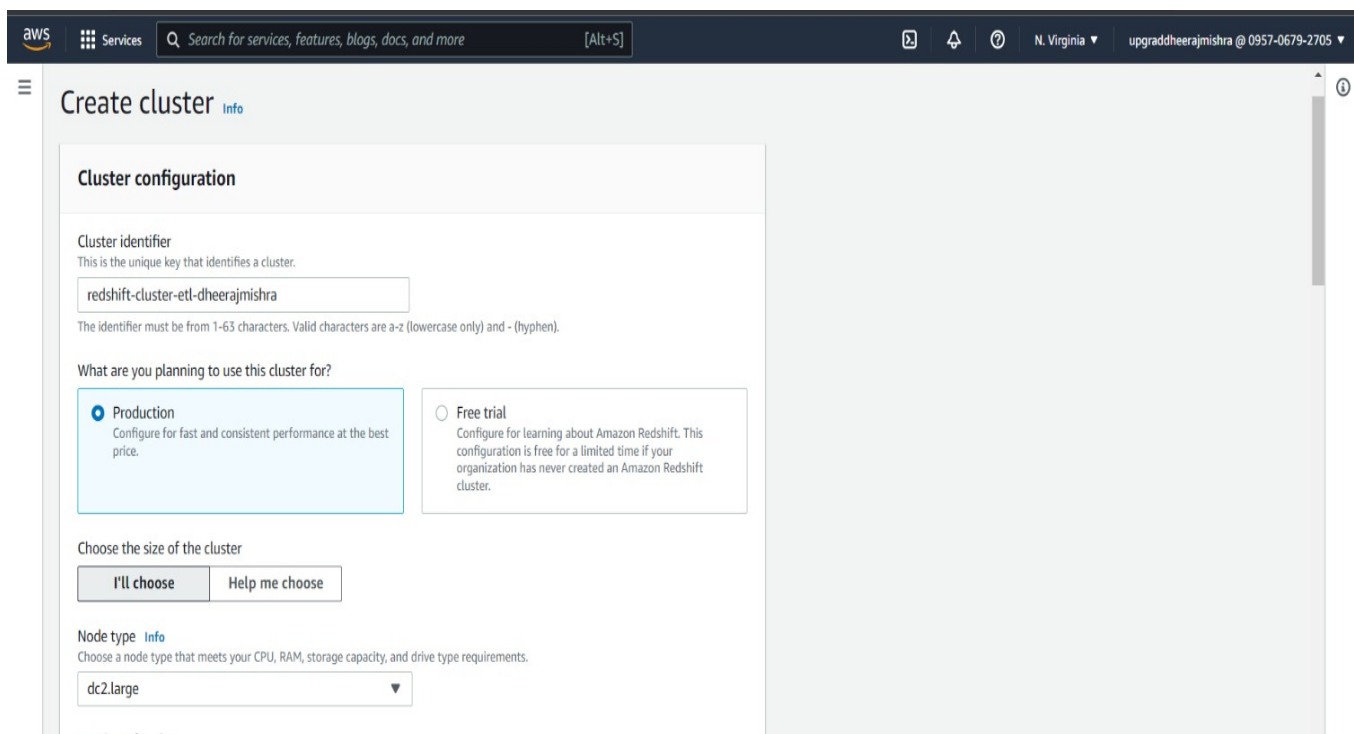


Creation of a Redshift Cluster

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

1. Screenshot of the type of machine used along with number of nodes



Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.
redshift-cluster-etl-dheerajmishra
The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

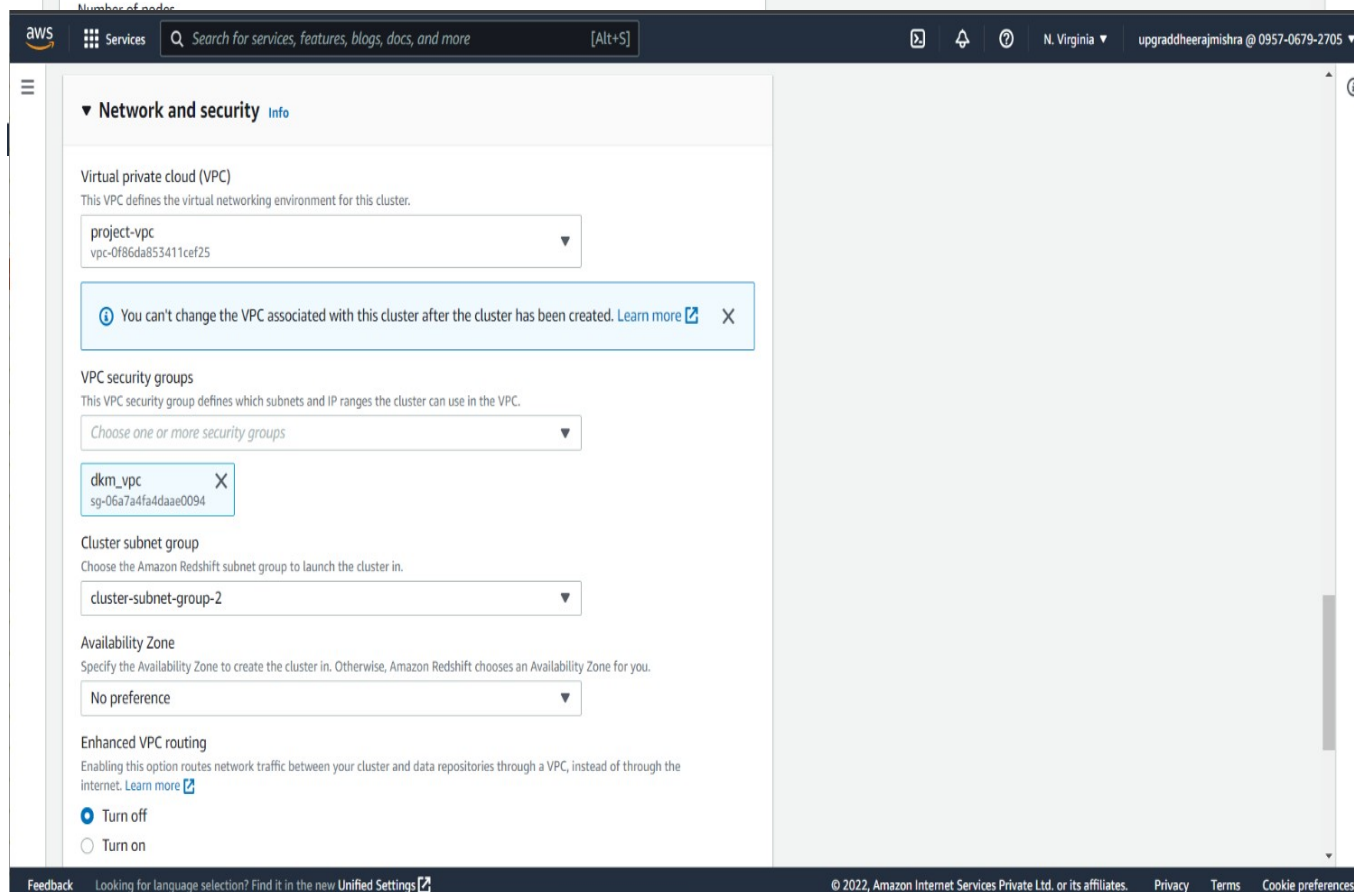
☒ **Production**
Configure for fast and consistent performance at the best price.

☐ **Free trial**
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.
dc2.large

Number of nodes



Network and security [Info](#)

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.
project-vpc
vpc-0f86da853411cef25

VPC security groups
This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.
Choose one or more security groups
dkm_vpc
sg-06a7a4fa4daae0094

Cluster subnet group
Choose the Amazon Redshift subnet group to launch the cluster in.
cluster-subnet-group-2

Availability Zone
Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.
No preference

Enhanced VPC routing
Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the internet. [Learn more](#)

☒ **Turn off**
☐ **Turn on**

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

N. Virginia

upgraddheerajmishra @ 0957-0679-2705

Database configurations

Database name

Specify a database name to create an additional database.

etl-dheeraj

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word.

Database port

Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.

5555

The port must be numeric (1150-65535).

Parameter groups

Defines database parameter and query queues for all the databases.

default.redshift-1.0

Default parameter group for redshift-1.0

Encryption

Encrypt all data on your cluster.

☒ Disabled

☐ Use AWS Key Management Service (AWS KMS)

☐ Use a hardware security module (HSM)

Maintenance

Feedback


Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms




Cookie preferences



Services


Search for services, features, blogs, docs, and more


[Alt+S]

N. Virginia

upgraddheerajmishra @ 0957-0679-2705





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. [Learn more](#)

Go to query editor v2

Amazon Redshift

>

Clusters

>

redshift-cluster-etl-dheerajmishra

redshift-cluster-etl-dheerajmishra

Actions

Edit

Add partner integration

Query data

General information

Cluster identifier

redshift-cluster-etl-dheerajmishra

Status

Available

Node type

dc2.large

Endpoint

redshift-cluster-etl-dheerajmishra.csk5c26ylco...

Cluster namespace

99b2805a-f0cd-44c6-a5fa-8bbcfcdafb02

Date created

September 24, 2022, 13:24 (UTC+05:30)

Number of nodes

2

JDBC URL

jdbc:redshift://redshift-cluster-etl-dheerajmis...

Storage used

-

ODBC URL

Driver={Amazon Redshift (x64)}; Server=redshi...

Cluster performance

Query monitoring

Schedules

Maintenance

Properties

Feedback

Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms

Cookie preferences

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

sta

Connect to database

×

Connection

Select a recent database connection or create a new database connection.

☐ Use a recent connection

☒ Create a new connection

Authentication

☒ Temporary credentials

Use the GetClusterCredentials IAM permission and your database user to generate temporary access credentials. [Learn more](#)

☐ AWS Secrets Manager

Use a stored secret to authenticate access. [Learn more](#)

Cluster

redshift-cluster-etl-dheerajmishra (Available) ▼

Database name

etl-dheeraj

Database user

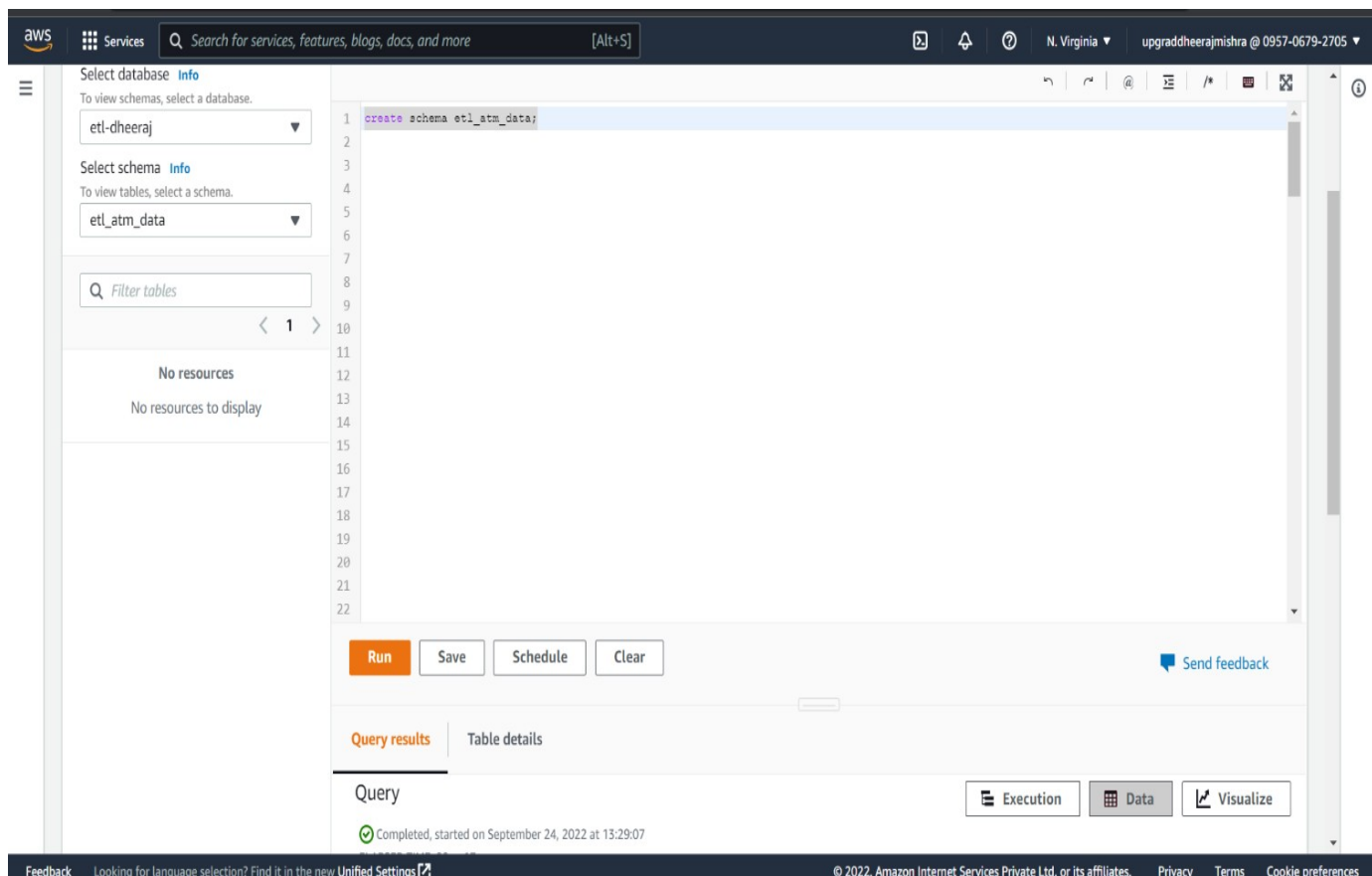
User name authorized to access your database.

awsuser

Cancel

Connect

create schema etl_atm_data;

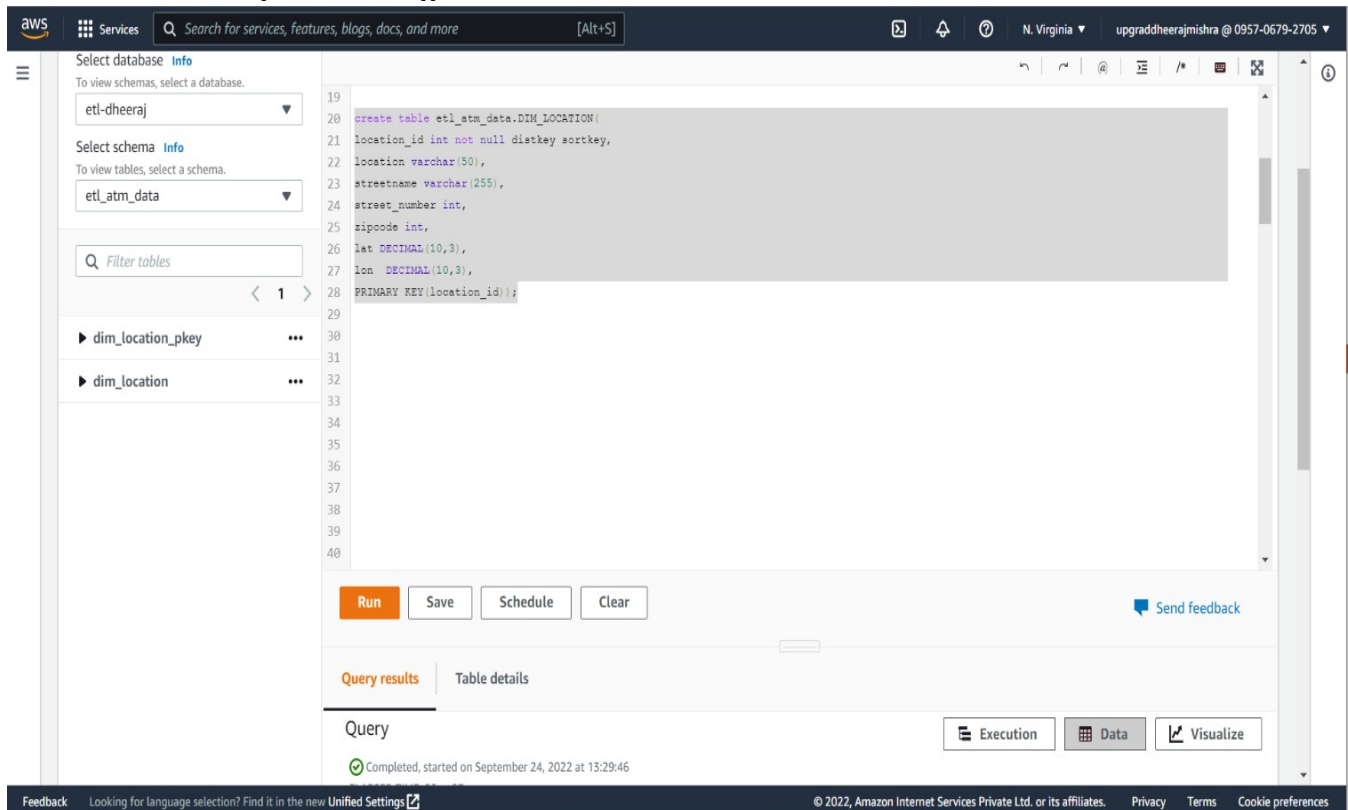


The screenshot shows the AWS Glue console interface. On the left sidebar, under 'Select database', 'etl-dheeraj' is selected. Under 'Select schema', 'etl_atm_data' is selected. The main editor area contains a single line of SQL: `create schema etl_atm_data;`. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. At the bottom, a status bar indicates 'Completed, started on September 24, 2022 at 13:29:07'.

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

1.
create table etl_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));
```



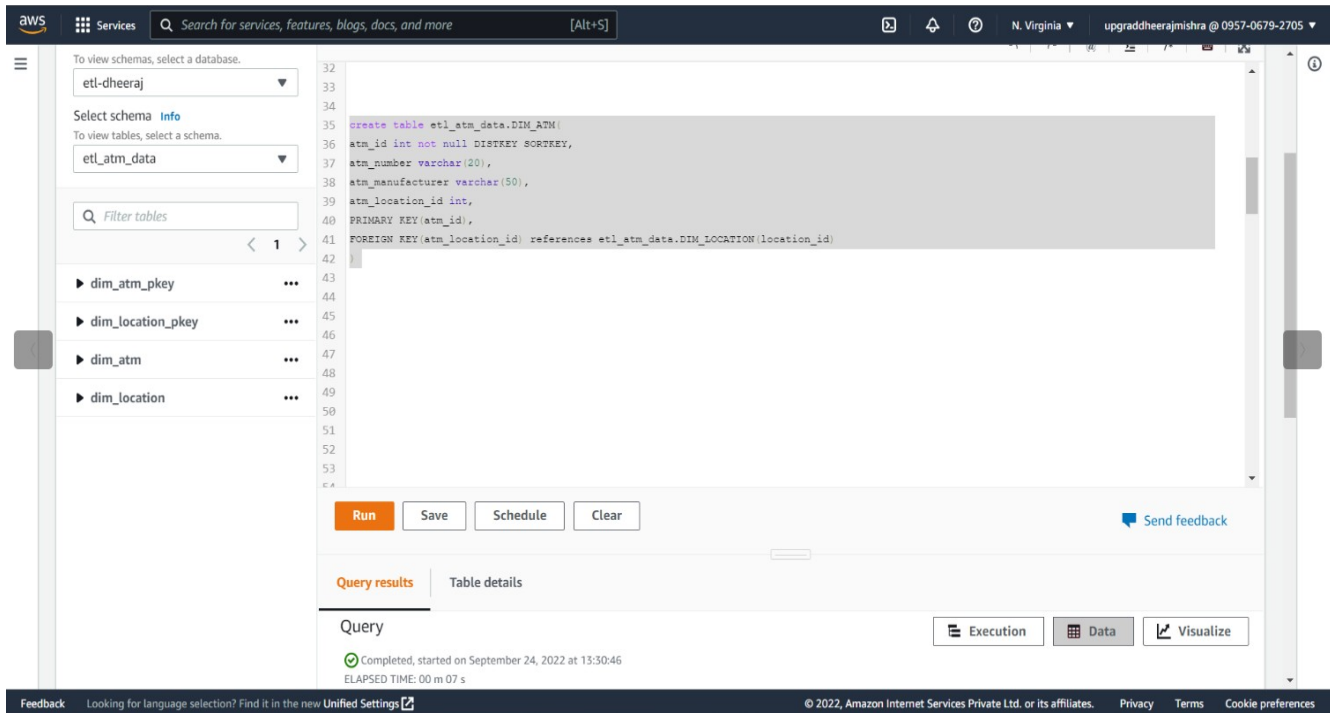
The screenshot shows the AWS Glue console interface. On the left, the 'Select database' dropdown is set to 'etl-dheeraj' and the 'Select schema' dropdown is set to 'etl_atm_data'. Below these, there are filters for tables, showing 'dim_location_pkey' and 'dim_location'. The main area displays a SQL query for creating a table named 'DIM_LOCATION' in the 'etl_atm_data' schema. The query is as follows:

```
19
20 create table etl_atm_data.DIM_LOCATION (
21 location_id int not null distkey sortkey,
22 location varchar(50),
23 streetname varchar(255),
24 street_number int,
25 zipcode int,
26 lat DECIMAL(10,3),
27 lon DECIMAL(10,3),
28 PRIMARY KEY(location_id));
29
30
31
32
33
34
35
36
37
38
39
40
```

Below the query, 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 of 'Completed, started on September 24, 2022 at 13:29:46'. At the bottom, there are links for 'Execution', 'Data', and 'Visualize'.

2.

```
create table etl_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 etl_atm_data.DIM_LOCATION(location_id)
);
```



The screenshot shows the AWS Glue console interface. On the left, the 'etl-dheeraj' database is selected, and the 'etl_atm_data' schema is chosen. A list of tables is visible, including 'dim_atm_pkey', 'dim_location_pkey', 'dim_atm', and 'dim_location'. The main area displays a SQL query to create a table named 'etl_atm_data.DIM_ATM' with the following schema:

```

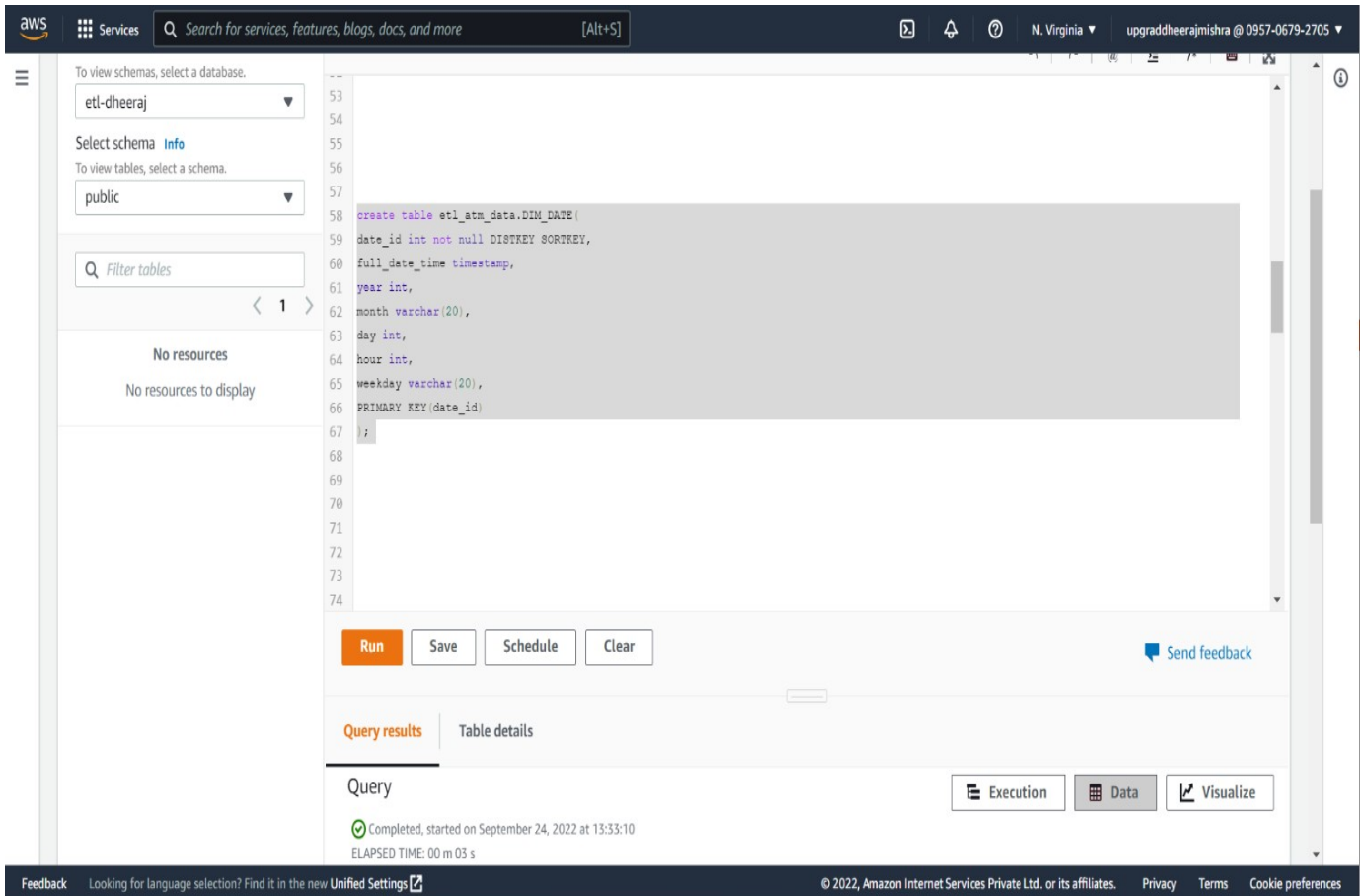
32
33
34
35 create table etl_atm_data.DIM_ATM(
36   atm_id int not null DISTKEY SORTKEY,
37   atm_number varchar(20),
38   atm_manufacturer varchar(50),
39   atm_location_id int,
40   PRIMARY KEY(atm_id),
41   FOREIGN KEY(atm_location_id) references etl_atm_data.DIM_LOCATION(location_id)
42 )
43
44
45
46
47
48
49
50
51
52
53
54

```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is active, showing a status message: 'Completed, started on September 24, 2022 at 13:30:46' and 'ELAPSED TIME: 00 m 07 s'. At the bottom, there are links for 'Feedback', 'Looking for language selection? Find it in the new Unified Settings', and copyright information: '© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences'.

3.
create table etl_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)
);

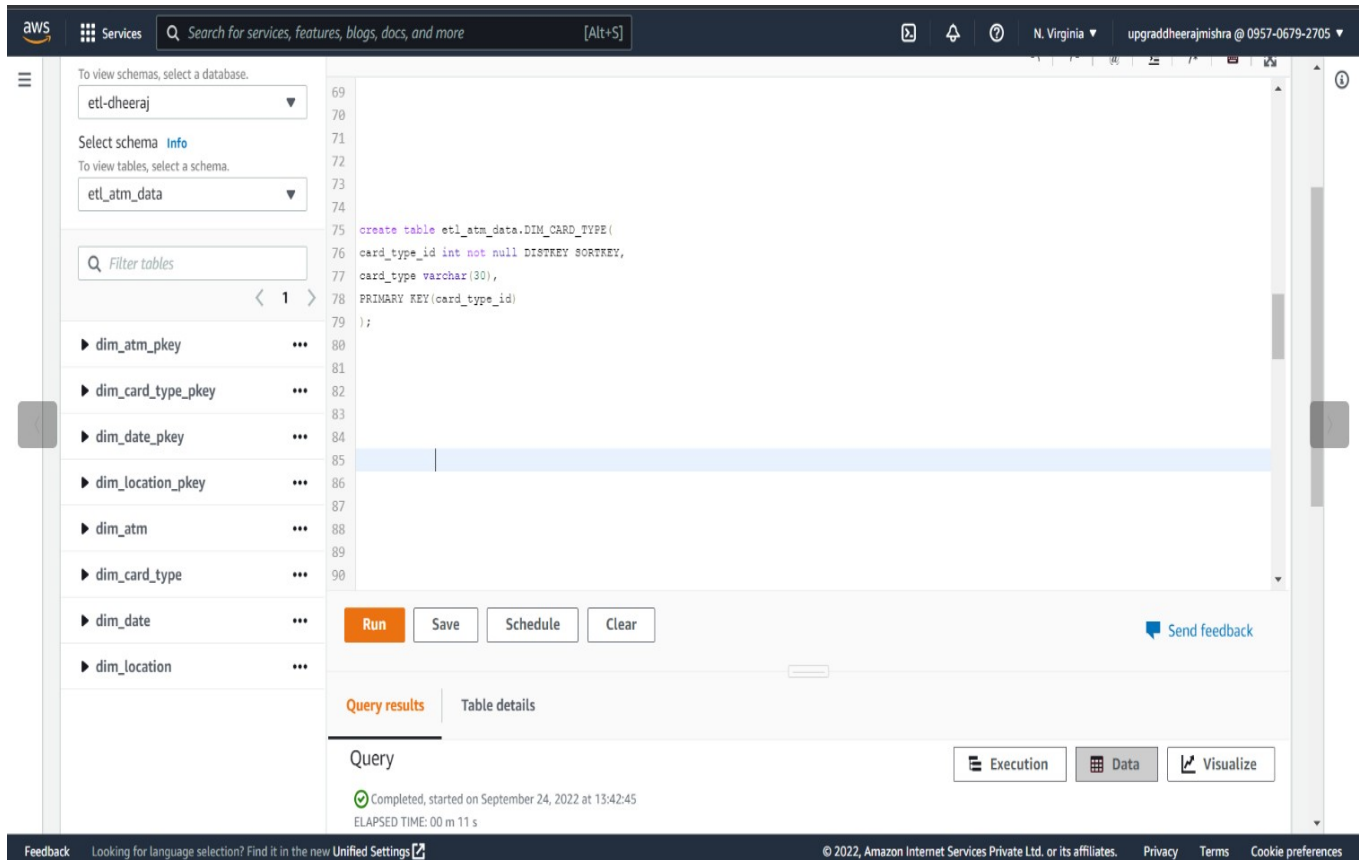


The screenshot shows the AWS Glue console interface. On the left, there's a sidebar with a search bar and a list of databases and schemas. The main area displays a SQL query for creating a table named `etl_atm_data.DIM_DATE`. The query is as follows:

```
create table etl_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)
);
```

Below the query, there are buttons for `Run`, `Save`, `Schedule`, and `Clear`. A `Send feedback` link is also present. The `Query results` tab is selected, showing a message: "Completed, started on September 24, 2022 at 13:33:10 ELAPSED TIME: 00 m 03 s".

4.
create table etl_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 AWS Glue console interface. On the left, a sidebar lists databases and schemas. The main area displays a SQL query for creating a table named `etl_atm_data.DIM_CARD_TYPE`. The query is as follows:

```


69
70
71
72
73
74
75 create table etl_atm_data.DIM_CARD_TYPE(
76   card_type_id int not null DISTKEY SORTKEY,
77   card_type varchar(30),
78   PRIMARY KEY(card_type_id)
79 );
80
81
82
83
84
85
86
87
88
89
90

```

Below the query editor, there are buttons for **Run**, **Save**, **Schedule**, and **Clear**. A **Send feedback** link is also present. The bottom section shows the **Query results** tab, indicating the query is **Completed** and started on September 24, 2022 at 13:42:45. The elapsed time is 00 m 11 s. There are also tabs for **Execution**, **Data**, and **Visualize**.

5.
create table etl_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 etl_atm_data.DIM_LOCATION(location_id),  
FOREIGN KEY(atm_id) references etl_atm_data.DIM_ATM(atm_id),  
FOREIGN KEY(date_id) references etl_atm_data.DIM_DATE(date_id),  
FOREIGN KEY(card_type_id) references etl_atm_data.DIM_CARD_TYPE(card_type_id)  
);
```



Services

Search for services, features, blogs, docs, and more

[Alt+S]

N. Virginia

upgradheerajmishra @ 0957-0679-2705

etl-uneeraj

Select schema

Info

To view tables, select a schema.

etl_atm_data

▼

Filter tables

Q

< 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

...

```

91 create table etl_atm_data.FACT_ATM_TRANS (
92 trans_id bigint not null DISTKEY SORTKEY,
93 atm_id int,
94 weather_loc_id int,
95 date_id int,
96 card_type_id int,
97 atm_status varchar(20),
98 currency varchar(10),
99 service varchar(20),
100 transaction_amount int,
101 message_code varchar(225),
102 message_text varchar(225),
103 rain_3h decimal(10,3),
104 clouds_all int,
105 weather_id int,
106 weather_main varchar(50),
107 weather_description varchar(255),
108 PRIMARY KEY(trans_id),
109 FOREIGN KEY(weather_loc_id) references etl_atm_data.DIM_LOCATION(location_id),
110 FOREIGN KEY(atm_id) references etl_atm_data.DIM_ATM(atm_id),
111 FOREIGN KEY(date_id) references etl_atm_data.DIM_DATE(date_id),
112 FOREIGN KEY(card_type_id) references etl_atm_data.DIM_CARD_TYPE(card_type_id)
113 );

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

Visualize

Completed, started on September 24, 2022 at 13:44:00

ELAPSED TIME: 00 m 05 s

Feedback

Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms

Cookie preferences

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

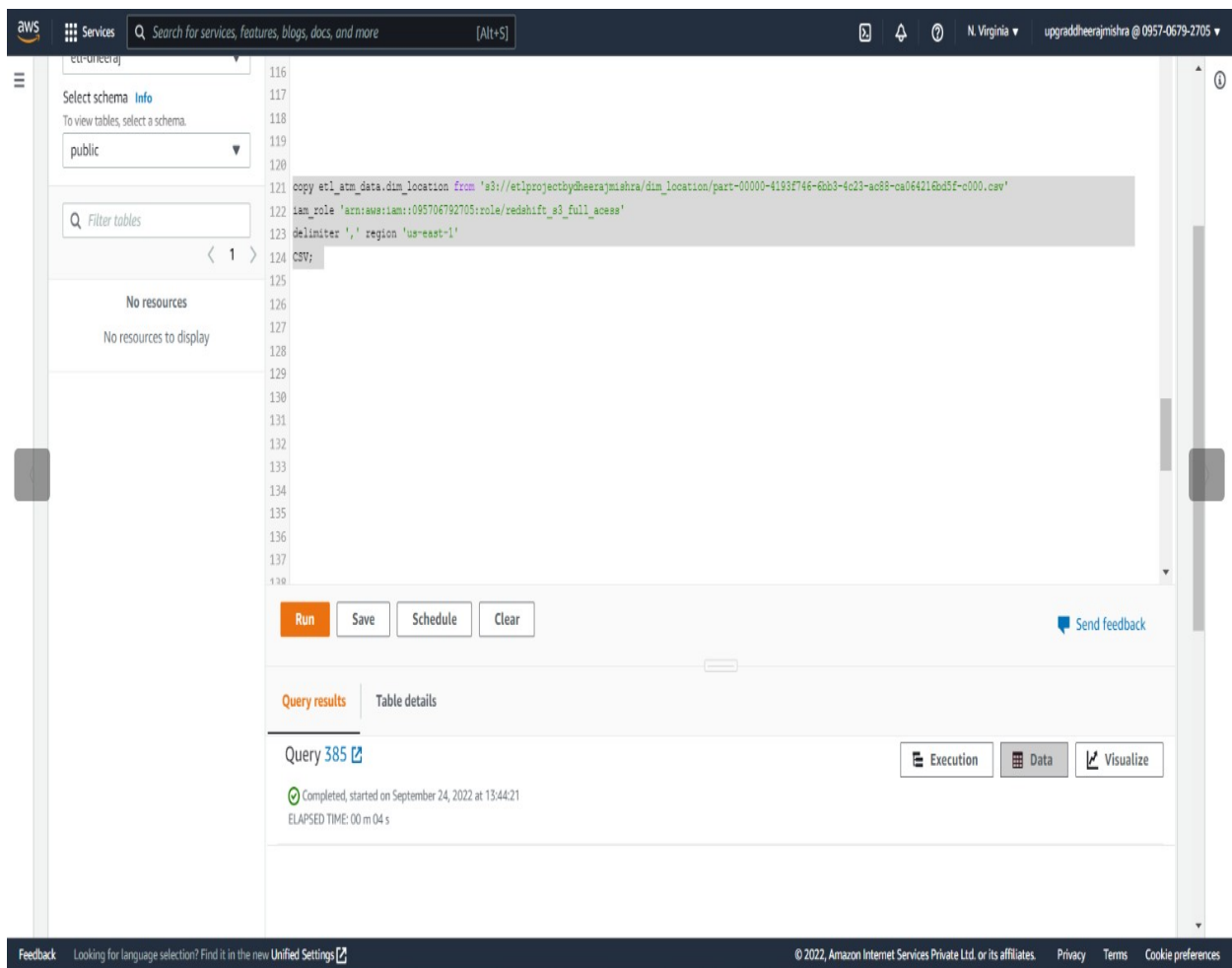
1.

copy etl_atm_data.dim_location from 's3://etlprojectbydheerajmishra/dim_location/part-00000-dab6b198-a75d-4dbd-8629-671d79c580ef-c000.csv'

iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'

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

CSV;



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Select schema' (public) and 'Filter tables'. The main area displays a SQL query being executed. The query is as follows:

```
copy etl_atm_data.dim_location from 's3://etlprojectbydheerajmishra/dim_location/part-00000-4193f746-6bb3-4c23-ac88-ca064216bd5f-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
CSV;
```

Below the query, 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 buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows the query ID 'Query 385' and its execution status: 'Completed, started on September 24, 2022 at 13:44:21' with an 'ELAPSED TIME: 00 m 04 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

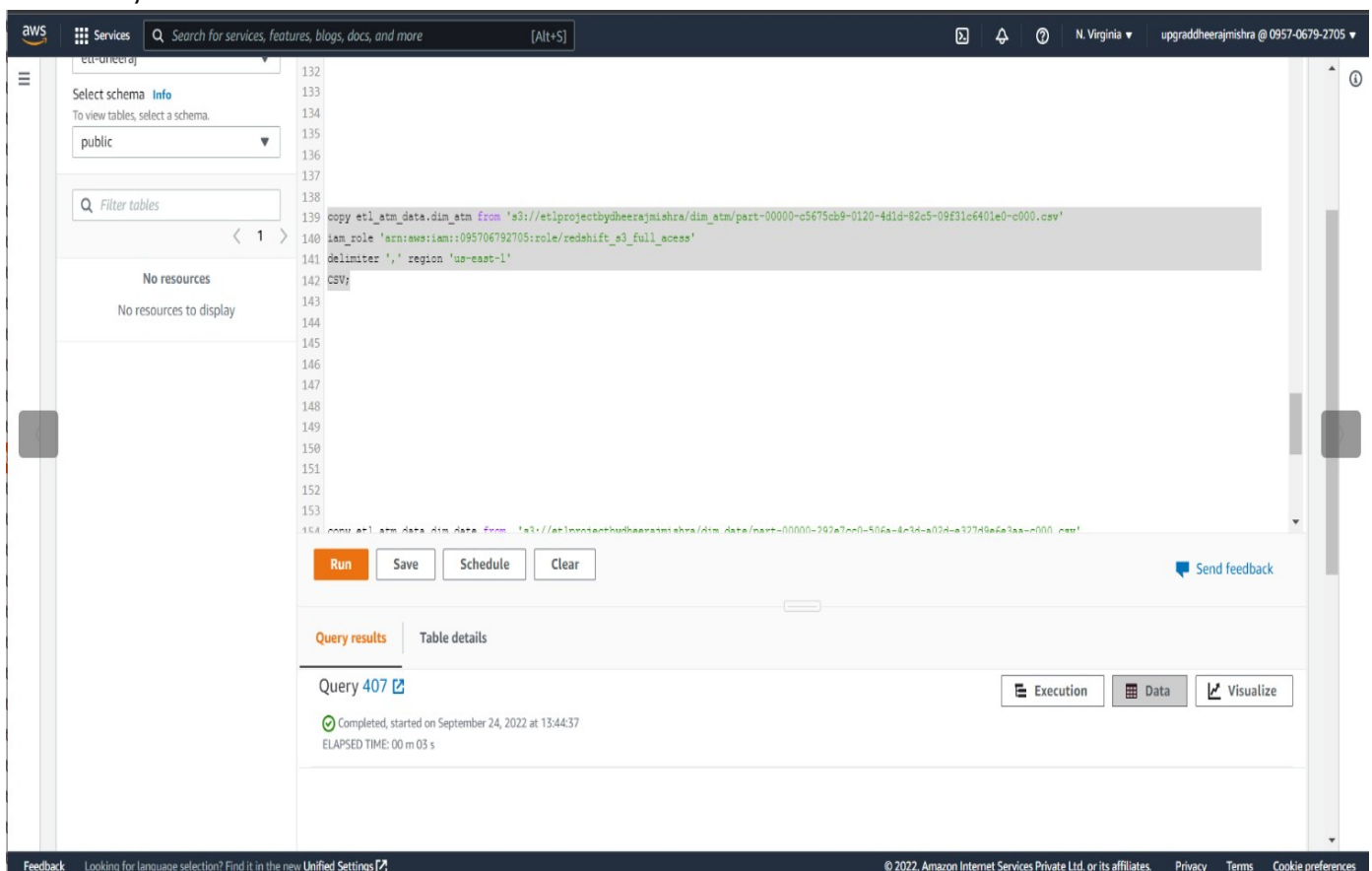
2.

copy etl_atm_data.dim_atm from 's3://etlprojectbydheerajmishra/dim_atm/part-00000-c9dc5033-485a-462c-a434-27aa8851297b-c000.csv'

iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_aces'

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

CSV;



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with a search bar and a list of schemas. The main area displays a SQL query being executed. The query is as follows:

```
copy etl_atm_data.dim_atm from 's3://etlprojectbydheerajmishra/dim_atm/part-00000-c9dc5033-485a-462c-a434-27aa8851297b-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_aces'
delimiter ',' region 'us-east-1'
CSV;
```

Below the query, 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 buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing the status of the query: 'Query 407' is completed, started on September 24, 2022, at 13:44:37, and the elapsed time is 00 m 03 s. There are also buttons for 'Execution', 'Data', and 'Visualize'.

3.

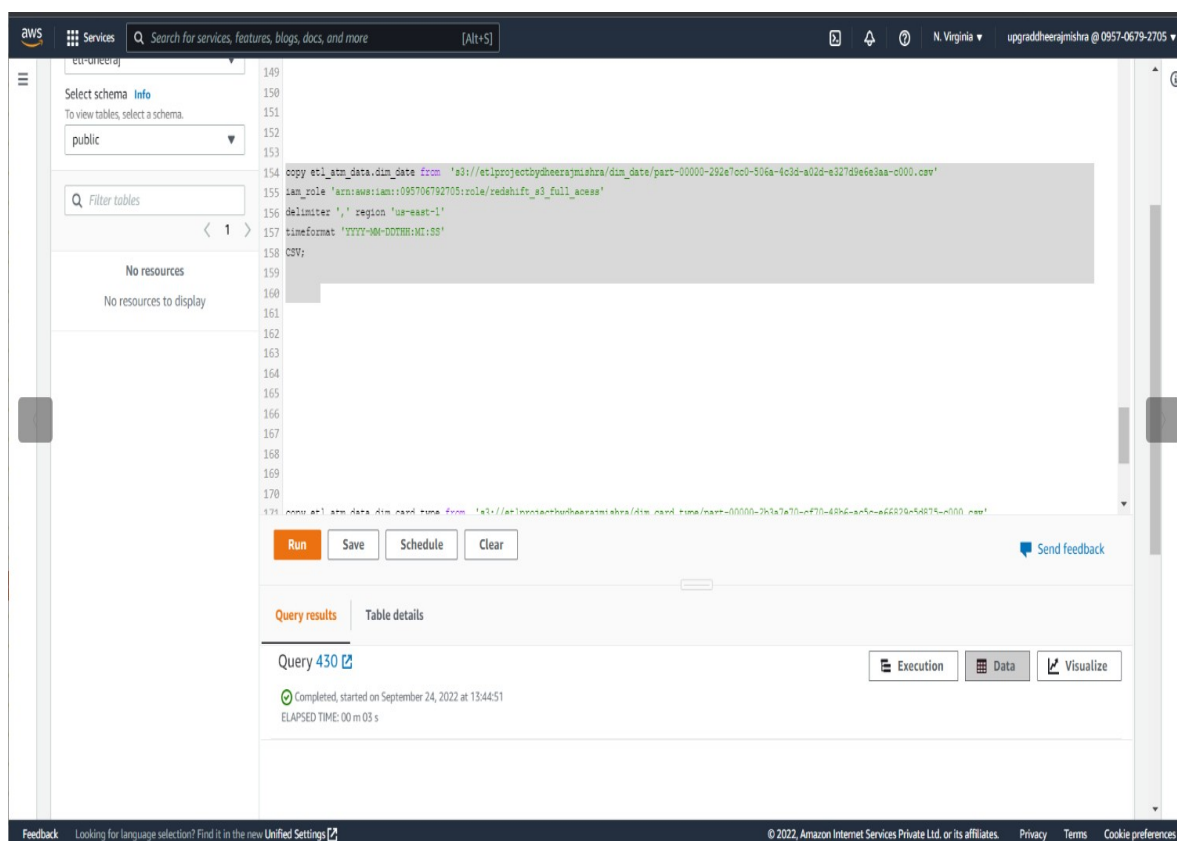
copy etl_atm_data.dim_date from 's3://etlprojectbydheerajmishra/dim_date/part-00000-02bf4ea0-0dad-44ce-b28a-1d76ef2b8aa6-c000.csv'

iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'

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

timeformat 'YYYY-MM-DDTHH:MI:SS'

CSV;



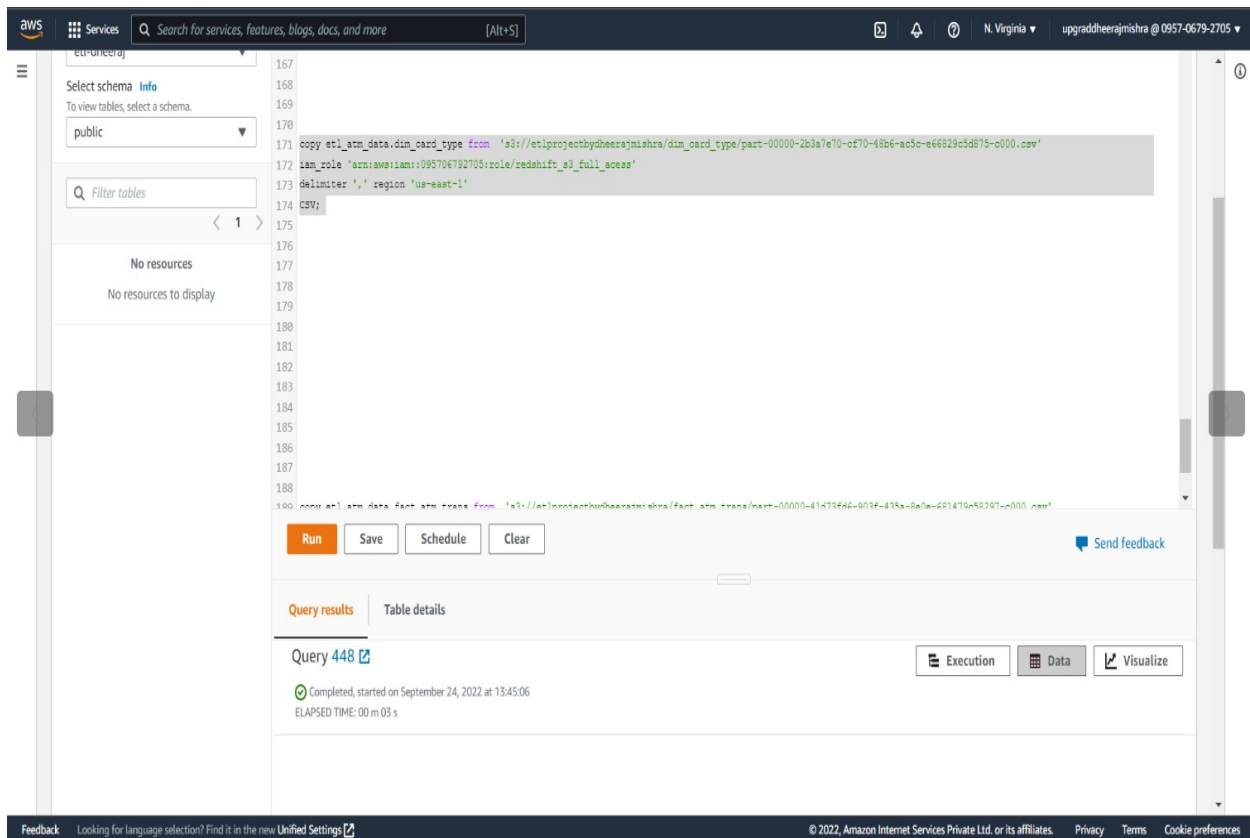
4.

copy etl_atm_data.dim_card_type from
's3://etlprojectbydheerajmishra/dim_card_type/part-00000-438a36c1-9e05-4faa-9839-6625e2e0be3a-c000.csv'

iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'

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

CSV;



The screenshot shows the AWS Glue console interface. On the left, there is a sidebar with a search bar and a list of tables. The main area displays a query editor with the following SQL code:

```
copy etl_atm_data.dim_card_type from 's3://etlprojectbydheerajmishra/dim_card_type/part-00000-438a36c1-9e05-4faa-9839-6625e2e0be3a-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
CSV;
```

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, there is a section for 'Query results' and 'Table details'. The 'Query results' tab is active, showing the following information:

- Query 448
- Completed, started on September 24, 2022 at 13:45:06
- ELAPSED TIME: 00 m 03 s

At the bottom of the console, there is a footer with the text: '© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences'.

5.

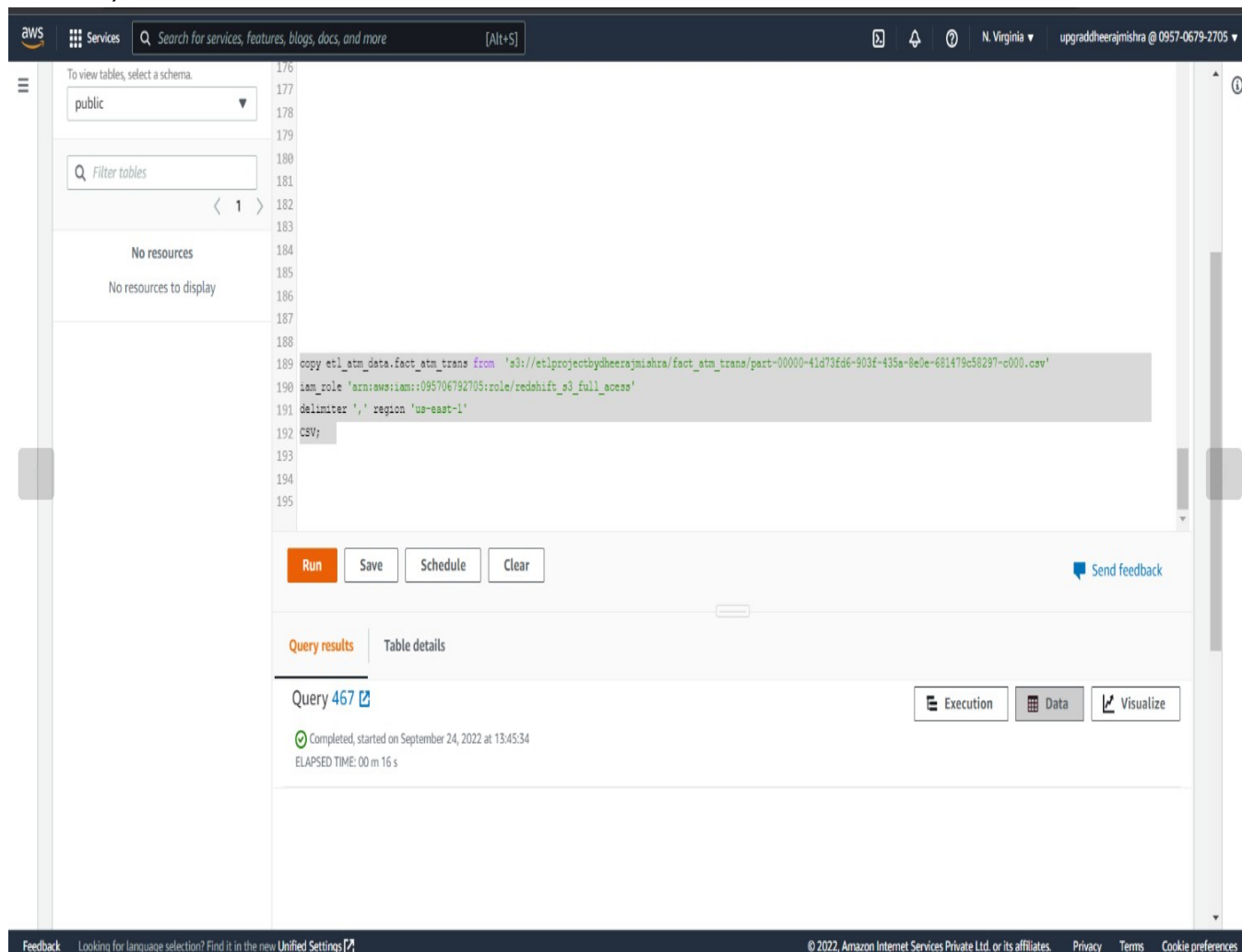
copy etl_atm_data.fact_atm_trans from

's3://etlprojectbydheerajmishra/fact_atm_trans/part-00000-5cbe2097-b629-4447-8fee-0f427ec95a5f-c000.csv'

iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'

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

CSV;



The screenshot shows the AWS Redshift console interface. At the top, there's a navigation bar with the AWS logo, 'Services' link, a search bar, and user information. The main area is divided into a left sidebar and a main content area. The sidebar shows a 'public' schema selected and a 'Filter tables' search bar. The main content area displays a SQL query in a text editor. The query is as follows:

```
copy etl_atm_data.fact_atm_trans from 's3://etlprojectbydheerajmishra/fact_atm_trans/part-00000-5cbe2097-b629-4447-8fee-0f427ec95a5f-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. 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 'Query 467' with a status of 'Completed, started on September 24, 2022 at 13:45:34' and an 'ELAPSED TIME: 00 m 16 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'. At the bottom of the console, there's a footer with a 'Feedback' link, a language selection notice, and copyright information for Amazon Internet Services Private Ltd. or its affiliates, along with links for 'Privacy', 'Terms', and 'Cookie preferences'.