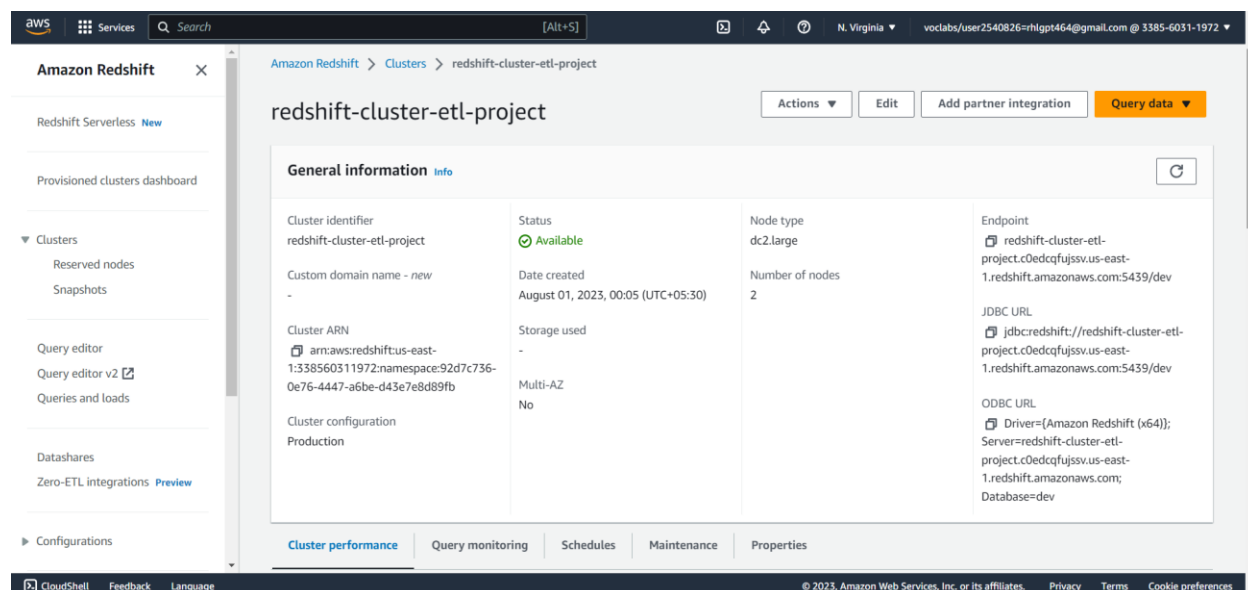
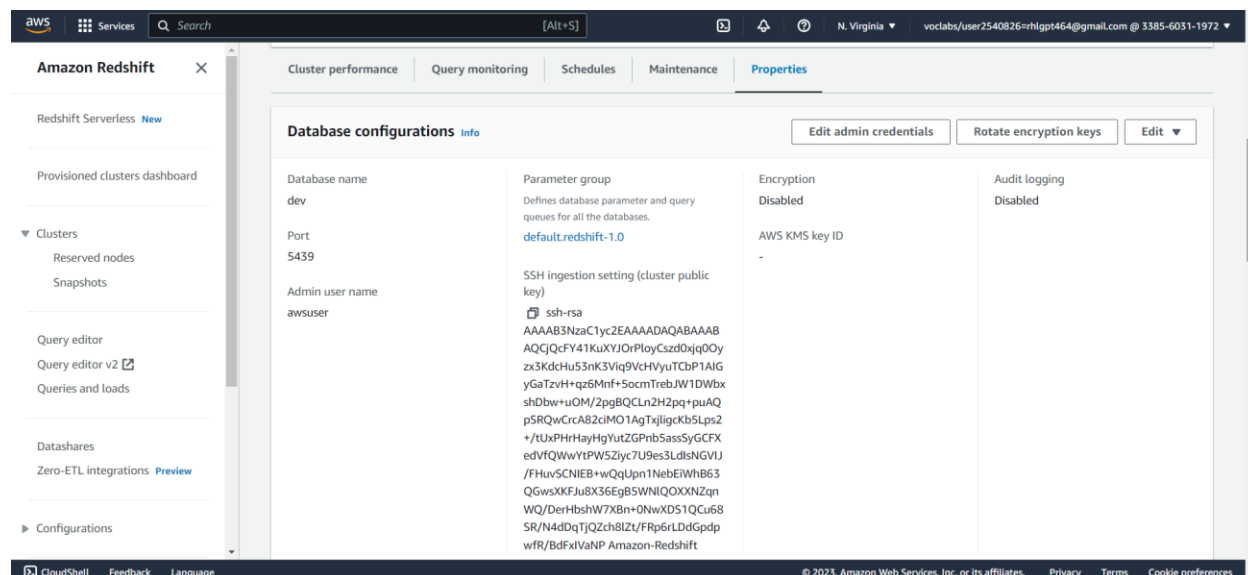


Creation of a Redshift Cluster

Screenshots of the configuration of the Redshift cluster that you have 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-etl-project' cluster details under the 'General information' tab. The cluster status is 'Available'. Key details include: Cluster identifier (redshift-cluster-etl-project), Status (Available), Node type (dc2.large), Endpoint (redshift-cluster-etl-project.c0edcqfujssv.us-east-1.redshift.amazonaws.com:5439/dev), JDBO URL, ODBC URL, Cluster ARN, Date created (August 01, 2023, 00:05 UTC+05:30), Number of nodes (2), Storage used, Multi-AZ (No), and Cluster configuration (Production).



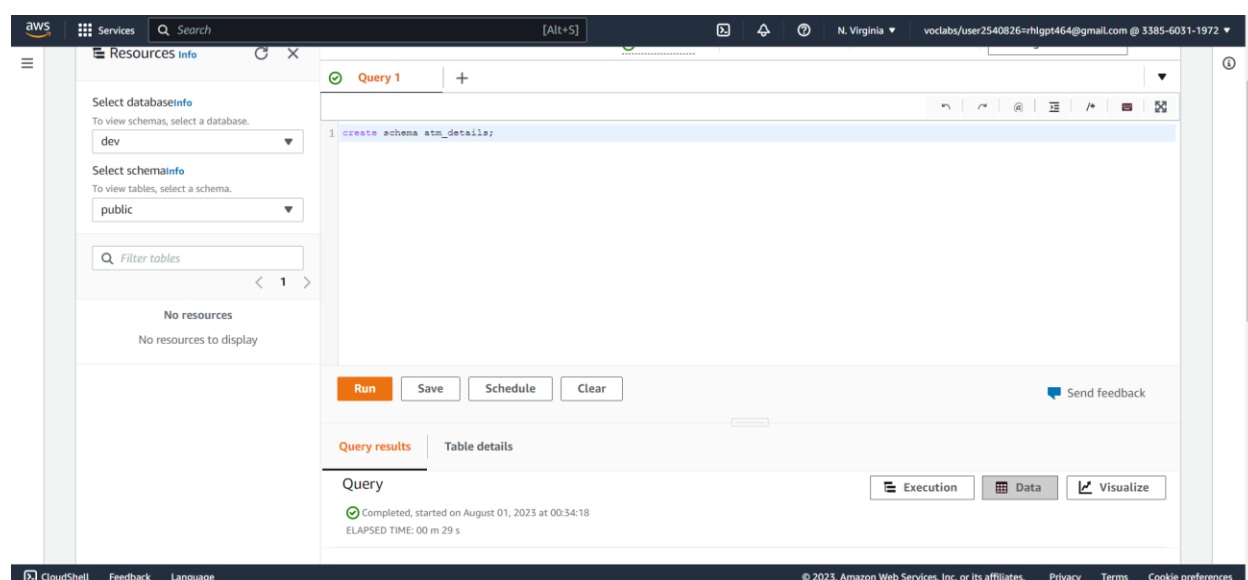
The screenshot shows the Amazon Redshift console interface, specifically the 'Properties' tab for the 'redshift-cluster-etl-project' cluster. The 'Database configurations' section is visible, showing details for the 'dev' database. Key details include: Database name (dev), Port (5439), Admin user name (awsuser), Parameter group (default.redshift-1.0), SSH ingestion setting (cluster public key), Encryption (Disabled), and Audit logging (Disabled).

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

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

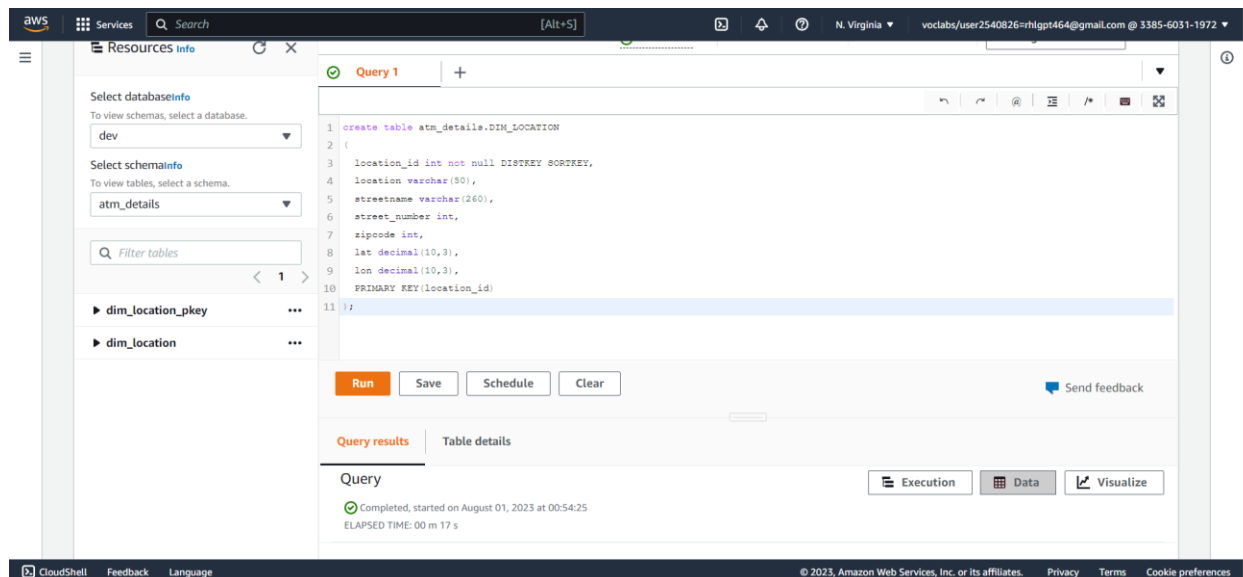
- Query for creating schema:

`create schema atm_details;`



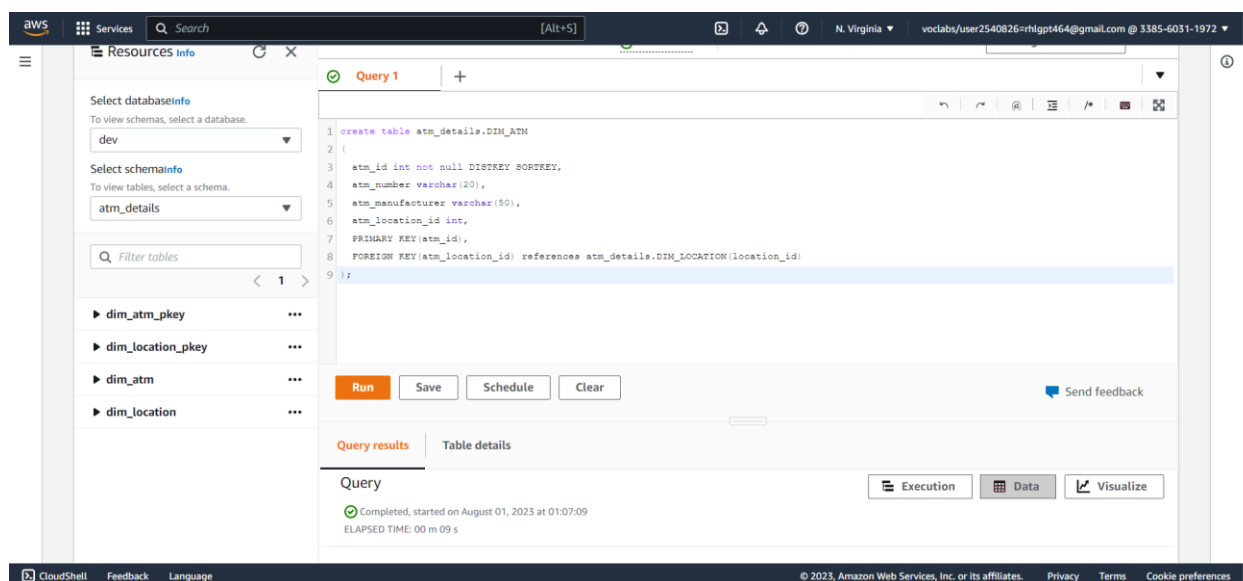
- Query for creating location dimension table:

```
create table atm_details.DIM_LOCATION
(
  location_id int not null DISTKEY SORTKEY,
  location varchar(50),
  streetname varchar(260),
  street_number int,
  zipcode int,
  lat decimal(10,3),
  lon decimal(10,3),
  PRIMARY KEY(location_id)
);
```



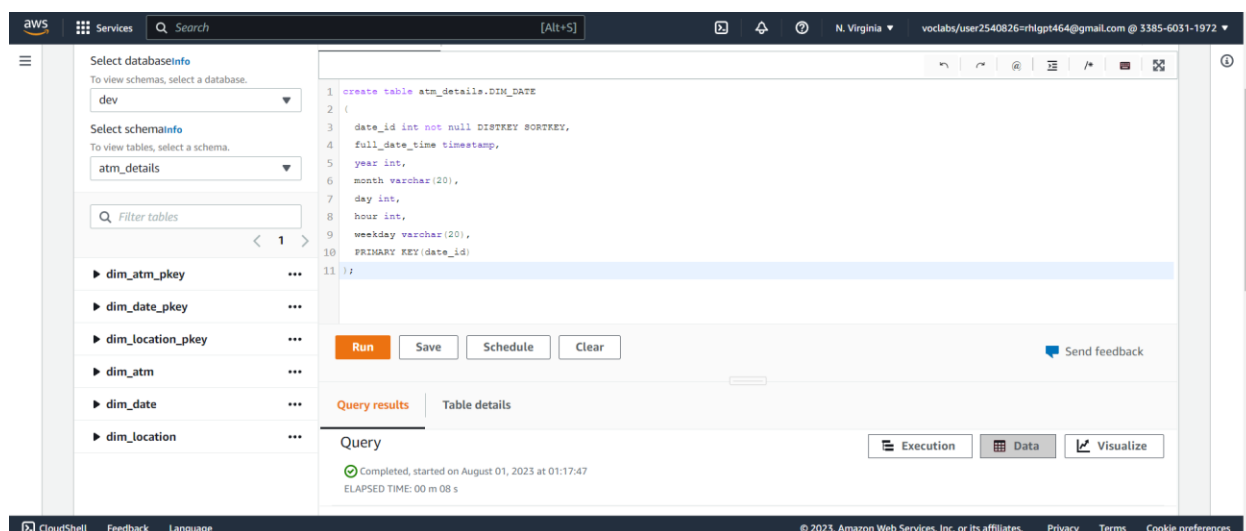
- Query for creating atm dimension table:

```
create table atm_details.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_details.DIM_LOCATION(location_id)
);
```



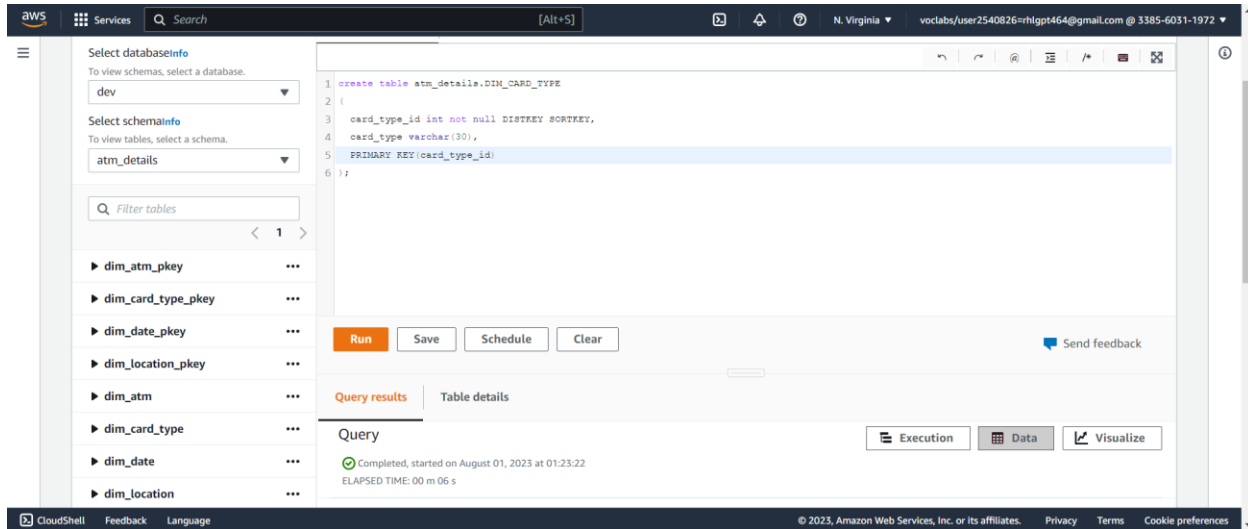
- Query for creating date dimension table:

```
create table atm_details.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)
);
```



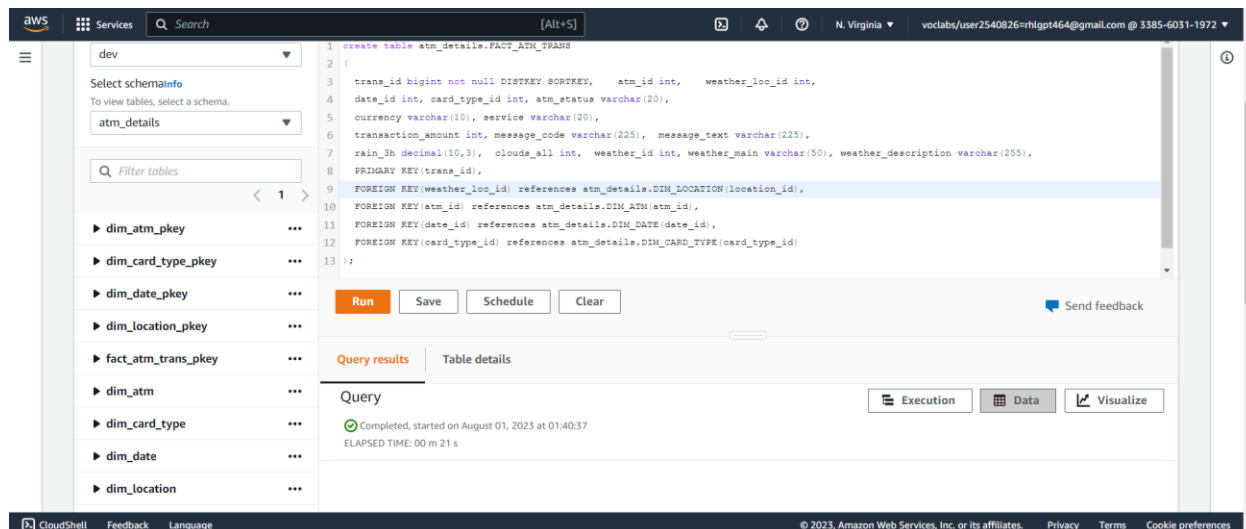
- Query for creating card type dimension table:

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



- Query for creating atm transactions fact table:

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



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

- Query to copy the data to location dimension table:

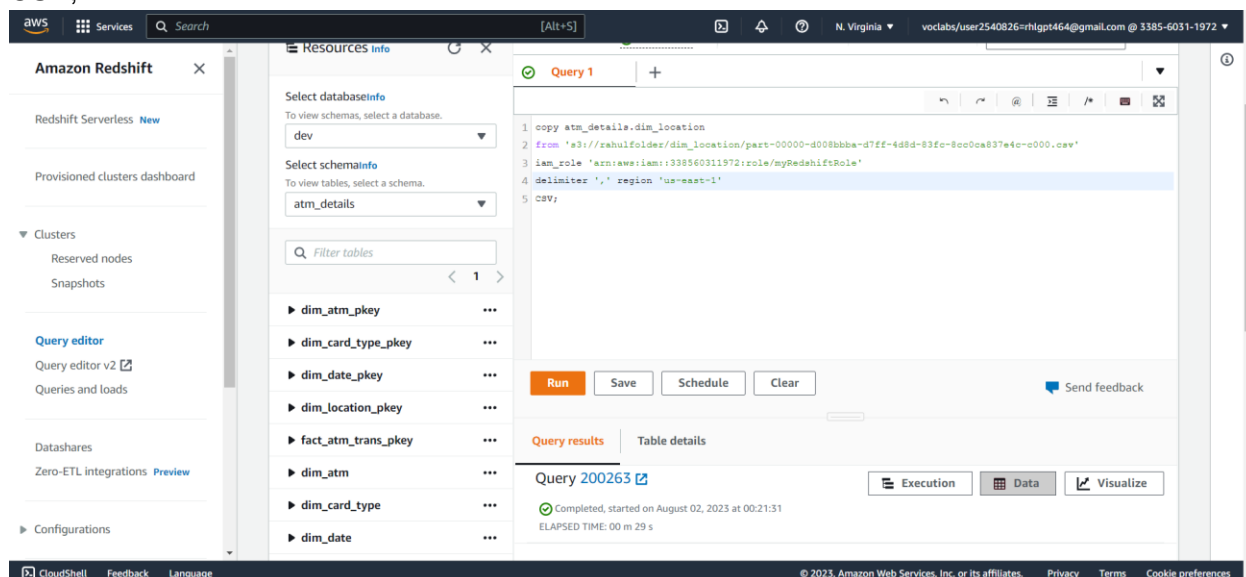
copy atm_details.dim_location

from 's3://rahulfolder/dim_location/part-00000-d008bbba-d7ff-4d8d-83fc-8cc0ca837e4c-c000.csv'

iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'

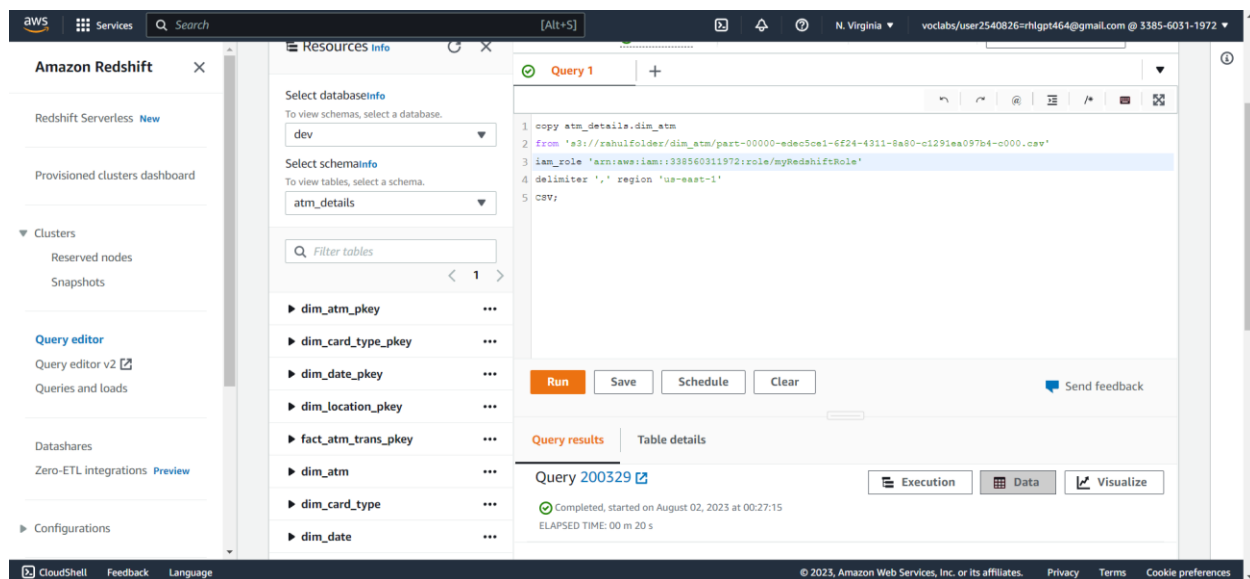
delimiter ',' region 'us-east-1'

CSV;



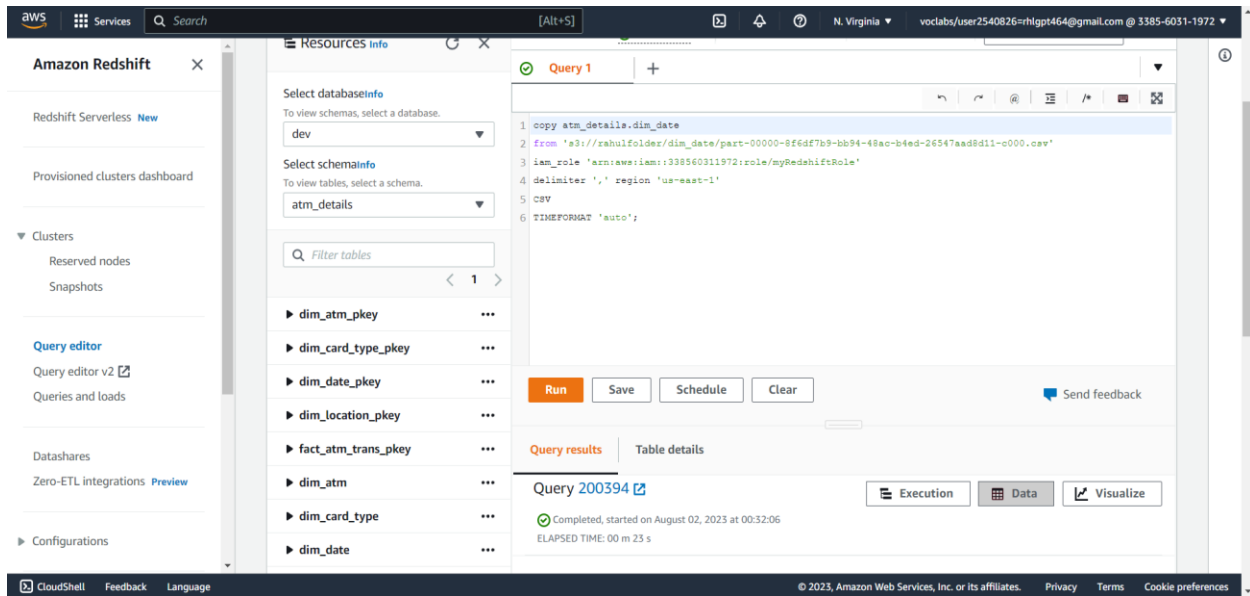
- Query to copy the data to atm table:

```
copy atm_details.dim_atm
from 's3://rahulfolder/dim_atm/part-00000-edec5ce1-6f24-4311-8a80-c1291ea097b4-c000.csv'
iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



- Query to copy the data to date table:

```
copy atm_details.dim_date
from 's3://rahulfolder/dim_date/part-00000-8f6df7b9-bb94-48ac-b4ed-26547aad8d11-c000.csv'
iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



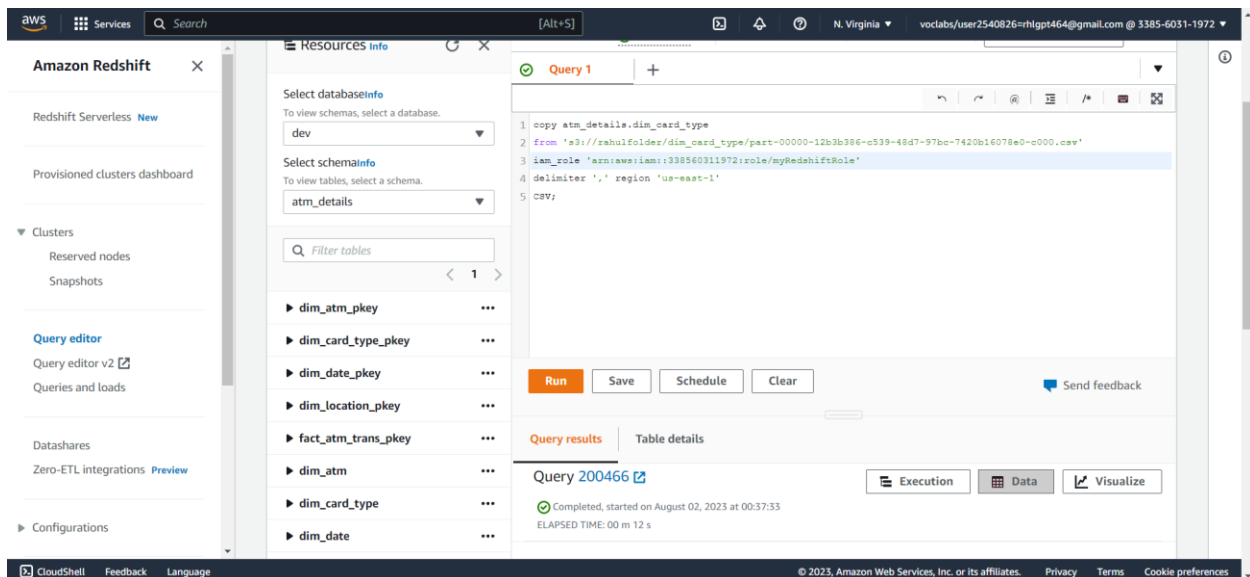
The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' pane is open, showing the 'dev' database and 'atm_details' schema. The 'Query editor' pane is active, displaying a SQL query:

```
1 copy atm_details.dim_date
2 from 's3://rahulfolder/dim_date/part-00000-8f6df7b9-bb94-48ac-b4ed-26547aad8d11-c000.csv'
3 iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV
6 TIMEFORMAT 'auto';
```

Below the query, the 'Run' button is highlighted. The 'Query results' pane shows the query status: 'Query 200394' is completed, started on August 02, 2023 at 00:32:06, and elapsed time is 00 m 23 s.

- Query to copy the data to card_type table:

```
copy atm_details.dim_card_type
from 's3://rahulfolder/dim_card_type/part-00000-12b3b386-c539-48d7-97bc-7420b16078e0-c000.csv'
iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' pane is open, showing the 'dev' database and 'atm_details' schema. The 'Query editor' pane is active, displaying a SQL query:

```
1 copy atm_details.dim_card_type
2 from 's3://rahulfolder/dim_card_type/part-00000-12b3b386-c539-48d7-97bc-7420b16078e0-c000.csv'
3 iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
```

Below the query, the 'Run' button is highlighted. The 'Query results' pane shows the query status: 'Query 200466' is completed, started on August 02, 2023 at 00:37:33, and elapsed time is 00 m 12 s.

- Query to copy the data to fact_atm_trans table:

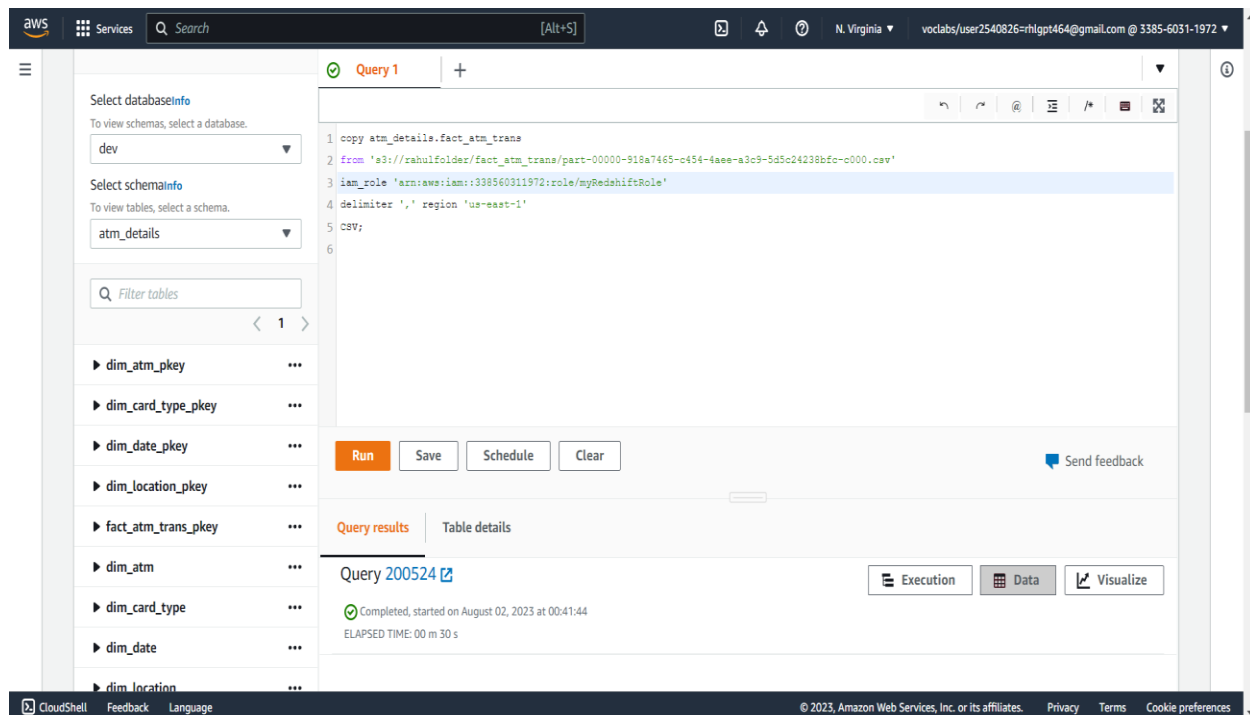
```
copy atm_details.fact_atm_trans
```

```
from 's3://rahulfolder/fact_atm_trans/part-00000-918a7465-c454-4aee-a3c9-5d5c24238bfc-c000.csv'
```

```
iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
```

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

```
CSV;
```



The screenshot displays the AWS Redshift console interface. On the left sidebar, under 'Select database info', the database 'dev' is selected. Under 'Select schema info', the schema 'atm_details' is selected. A list of tables is shown, 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', and 'dim_location'. The main panel shows a SQL query labeled 'Query 1' with the following text:

```
1 copy atm_details.fact_atm_trans
2 from 's3://rahulfolder/fact_atm_trans/part-00000-918a7465-c454-4aee-a3c9-5d5c24238bfc-c000.csv'
3 iam_role 'arn:aws:iam::338560311972:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
6
```

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 the query ID 'Query 200524' and its execution status: 'Completed, started on August 02, 2023 at 00:41:44' with an 'ELAPSED TIME: 00 m 30 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.