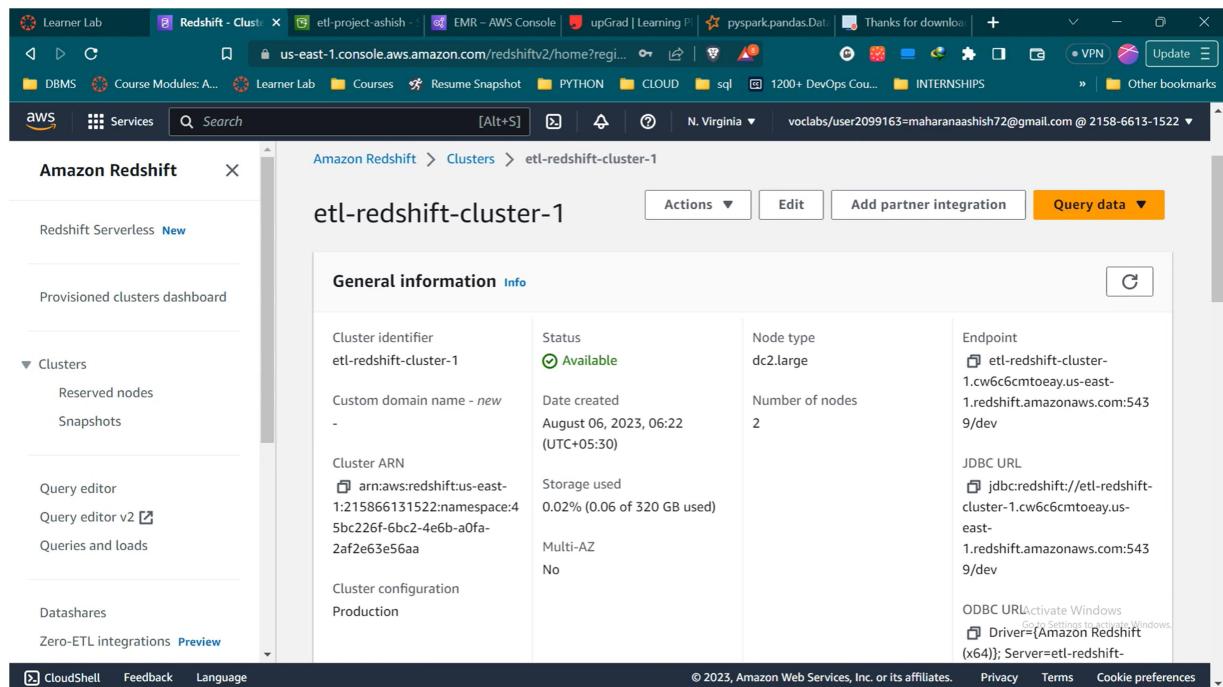
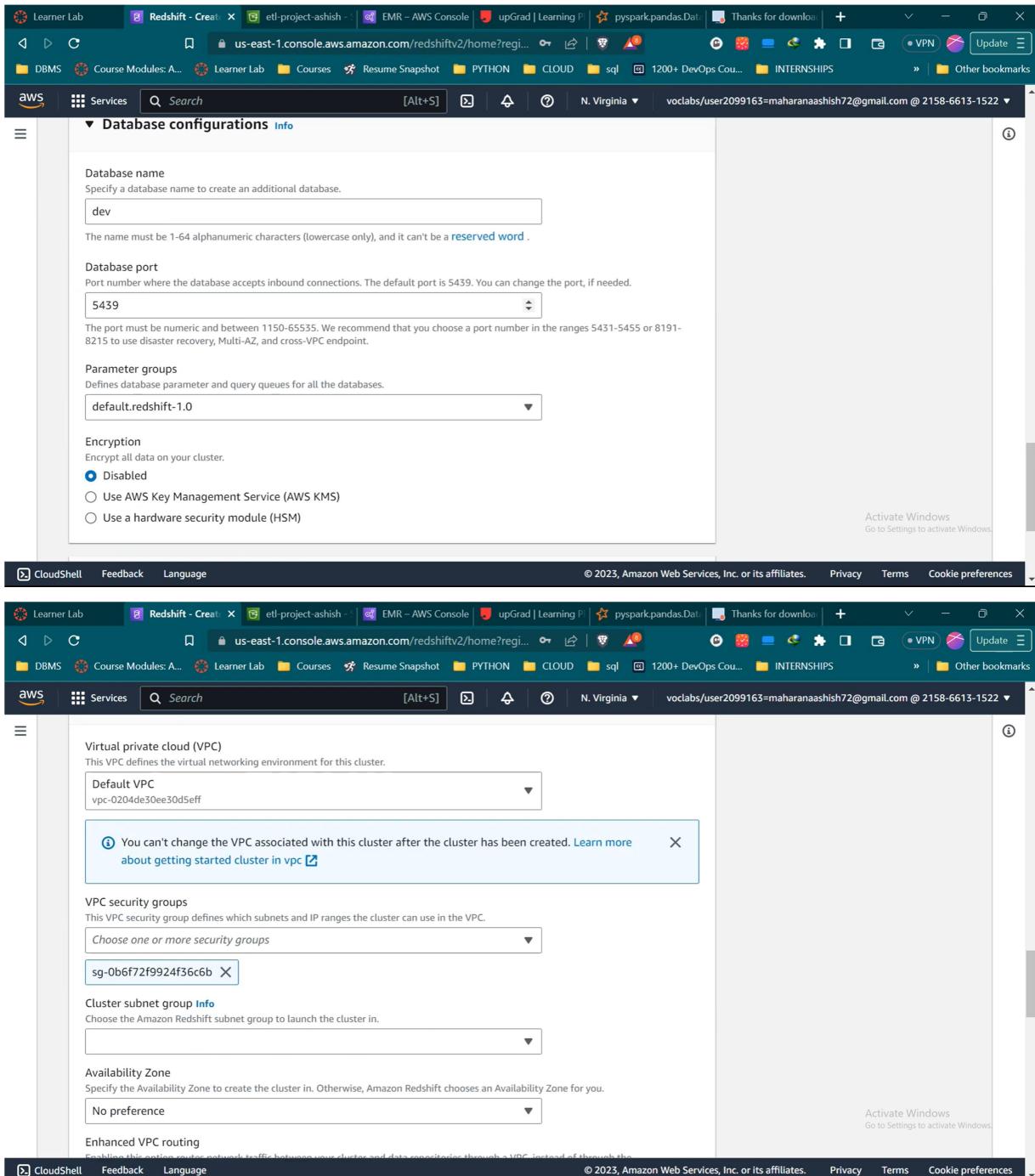


Creation of a Redshift Cluster

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



The screenshot shows the AWS Amazon Redshift console with the cluster details for 'etl-redshift-cluster-1'. The cluster identifier is 'etl-redshift-cluster-1', status is 'Available', node type is 'dc2.large', and there are 2 nodes. The endpoint is 'etl-redshift-cluster-1.cw6c6cmtoeay.us-east-1.redshift.amazonaws.com:5439/dev'. The JDBC URL is 'jdbc:redshift://etl-redshift-cluster-1.cw6c6cmtoeay.us-east-1.redshift.amazonaws.com:5439/dev'. The ODBC URL is 'Driver={Amazon Redshift (x64)}; Server=etl-redshift-1'. The cluster ARN is 'arn:aws:redshift:us-east-1:215866131522:namespace:45bc226f-6bc2-4e6b-a0fa-2af2e63e56aa'. The cluster configuration is 'Production'.



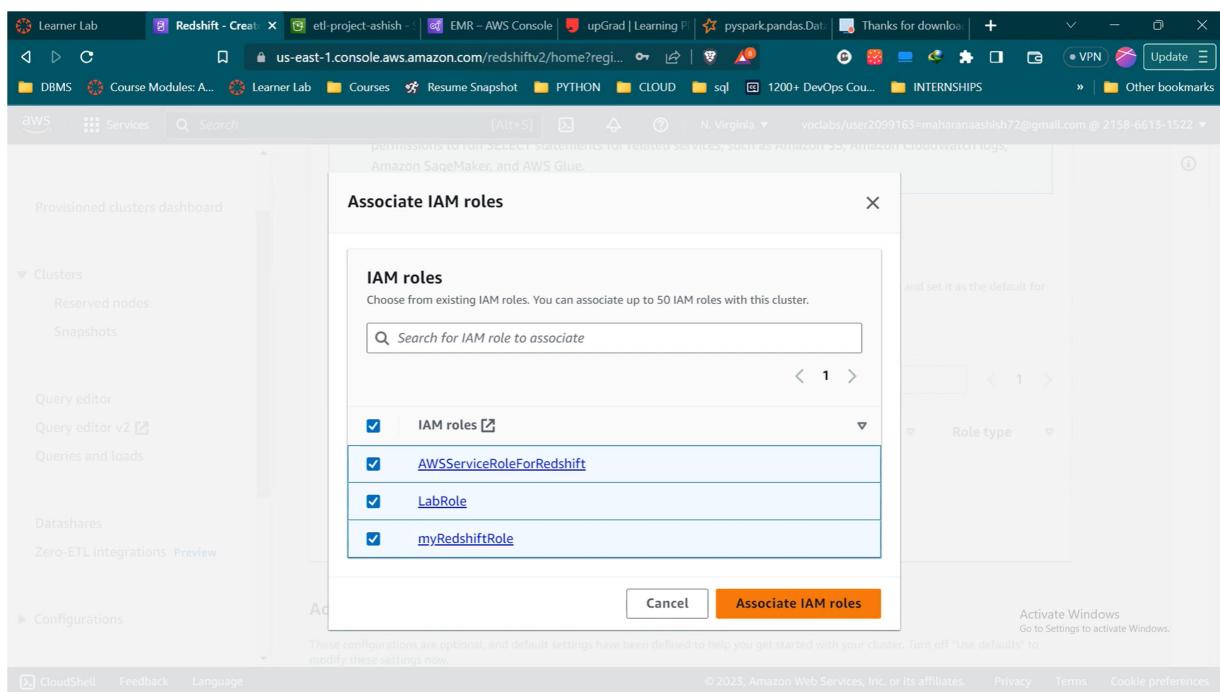
The screenshot shows two stacked screenshots of the AWS Redshift cluster creation interface.

Top Screenshot: Database configurations

- Database name:** dev
- Database port:** 5439
- Parameter groups:** default.redshift-1.0
- Encryption:** Disabled

Bottom Screenshot: Virtual private cloud (VPC)

- Default VPC:** vpc-0204de30ee30d5eff
- Message:** You can't change the VPC associated with this cluster after the cluster has been created. [Learn more about getting started cluster in vpc](#)
- VPC security groups:** sg-0b6f72f9924f36ccb
- Cluster subnet group:** Choose the Amazon Redshift subnet group to launch the cluster in.
- Availability Zone:** No preference
- Enhanced VPC routing:** Enabling this option routes network traffic between your cluster and data concentrators through VPC instead of the web.



The screenshot shows the AWS Redshift console with a modal dialog titled "Associate IAM roles". The dialog lists several IAM roles:

- AWSServiceRoleForRedshift
- LabRole
- myRedshiftRole

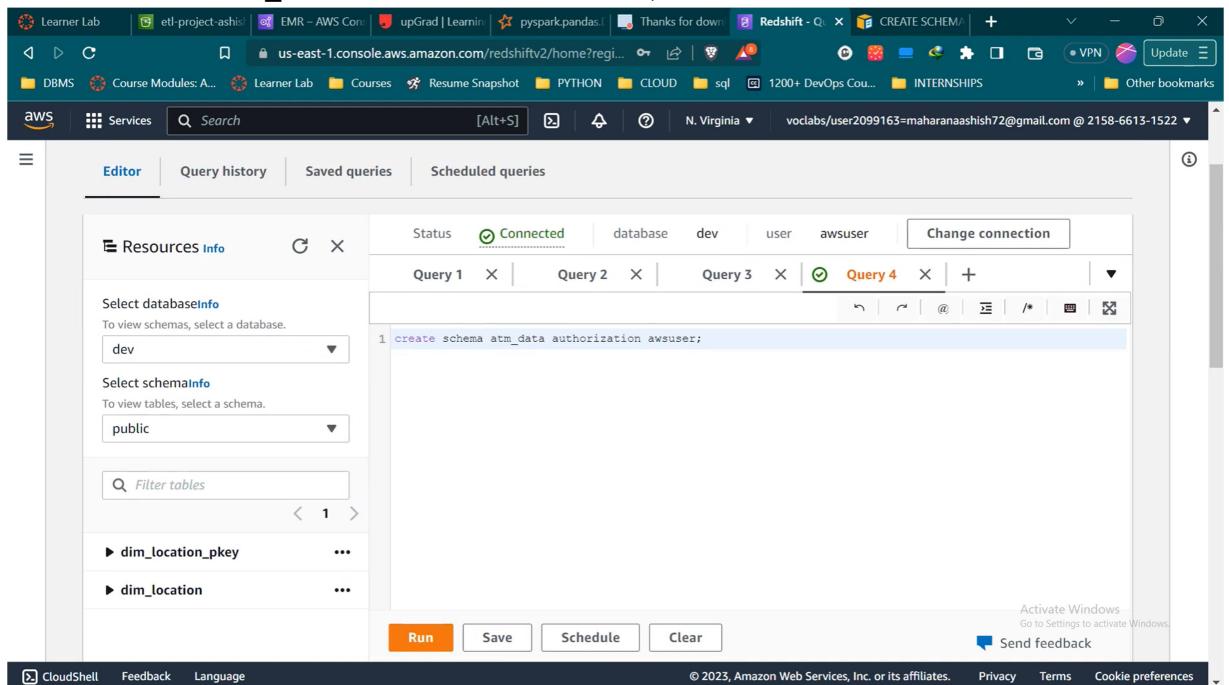
At the bottom right of the dialog are two buttons: "Cancel" and "Associate IAM roles".

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:

- **Creating schema**

```
create schema atm_data authorization awsuser;
```

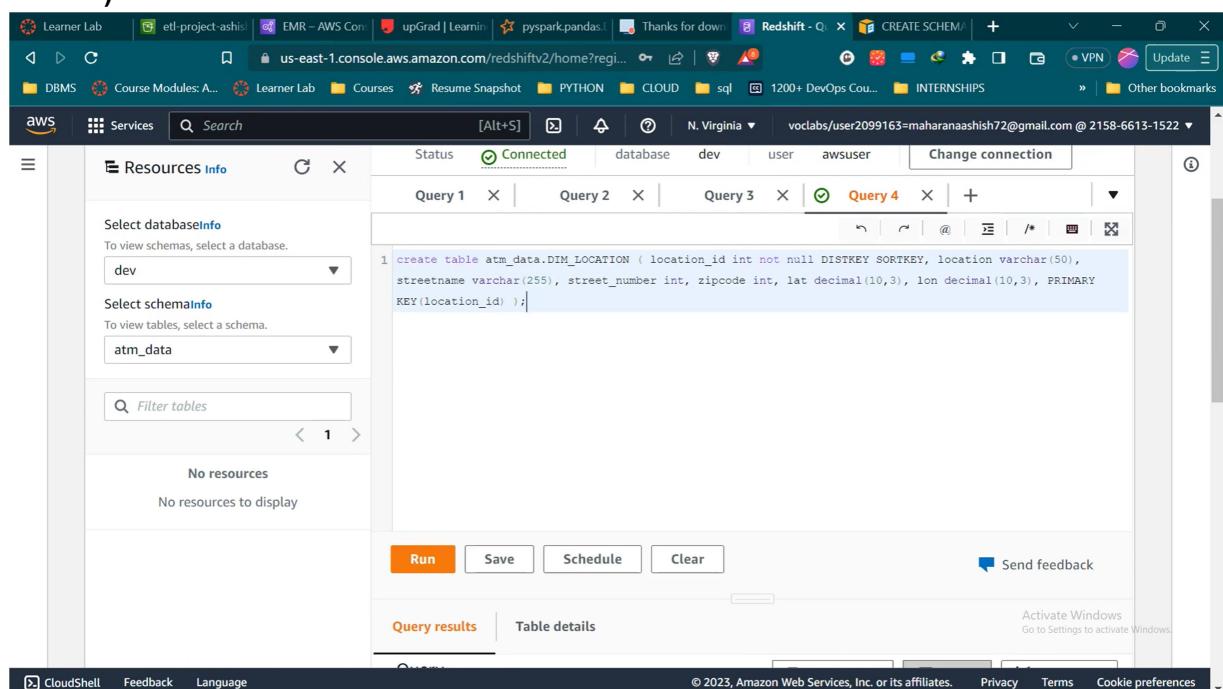


The screenshot shows the AWS Redshift Query Editor interface. The left sidebar displays 'Resources' with dropdown menus for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these are sections for 'Filter tables' and two tables: 'dim_location_pkey' and 'dim_location'. The main area shows four tabs at the top: 'Query 1', 'Query 2', 'Query 3', and 'Query 4' (which is currently active). The query editor window contains the SQL command: '1 create schema atm_data authorization awsuser;'. At the bottom of the editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The status bar at the bottom of the browser window indicates a connection to 'Connected' database 'dev' user 'awsuser'.

<Queries>

- **Creating location dimension table**

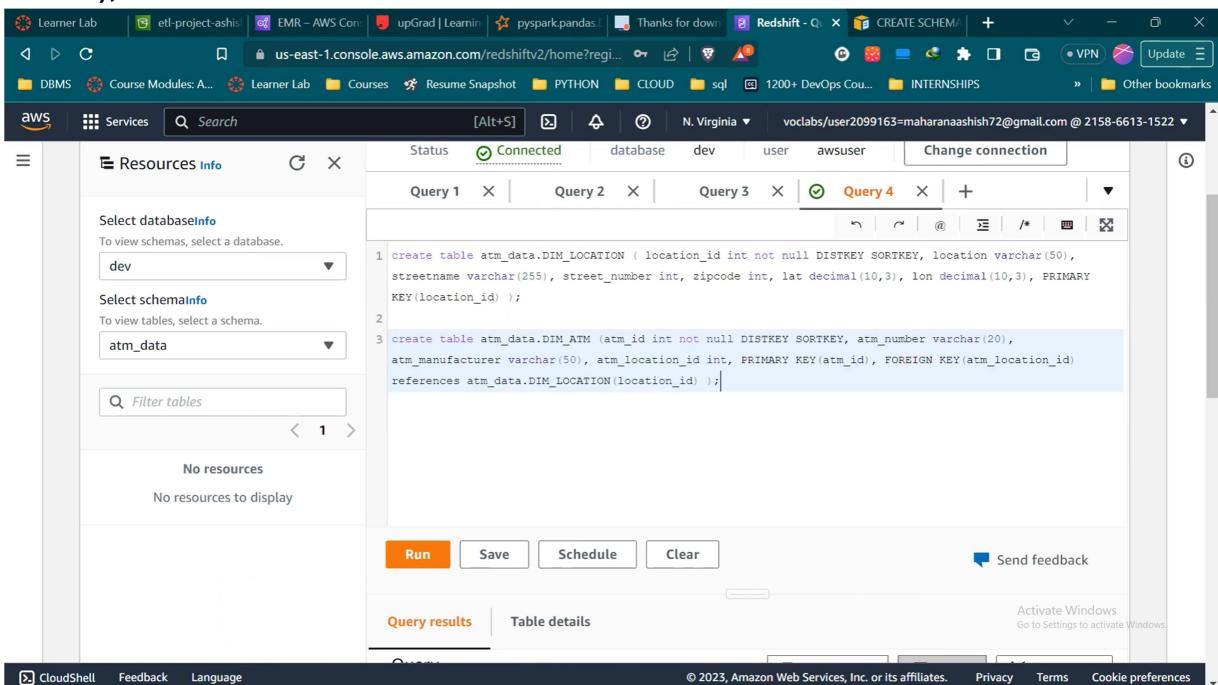
```
create table atm_data.DIM_LOCATION
(
    location_id int not null DISTKEY SORTKEY,
    location varchar(50),
    streetname varchar(255),
    street_number int,
    zipcode int,
    lat decimal(10,3),
    lon decimal(10,3),
    PRIMARY KEY(location_id)
)
```



The screenshot shows the AWS Redshift Query Editor interface. The top navigation bar includes tabs for Learner Lab, etl-project-ashish, EMR - AWS Con..., upGrad | Learnin..., pypspark.pandas..., Thanks for down..., Redshift - Q..., CREATE SCHEMA..., and other items. The main window has a sidebar titled 'Resources Info' with dropdowns for 'Select databaseInfo' (set to dev) and 'Select schemaInfo' (set to atm_data). Below this is a search bar labeled 'Filter tables'. The main area contains four tabs: 'Query 1', 'Query 2', 'Query 3', and 'Query 4'. The 'Query 4' tab is active, showing the SQL code for creating the DIM_LOCATION table. At the bottom of the editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear', along with a 'Send feedback' link. The footer includes links for CloudShell, Feedback, Language, and copyright information: © 2023, Amazon Web Services, Inc. or its affiliates.

- 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)
);
```



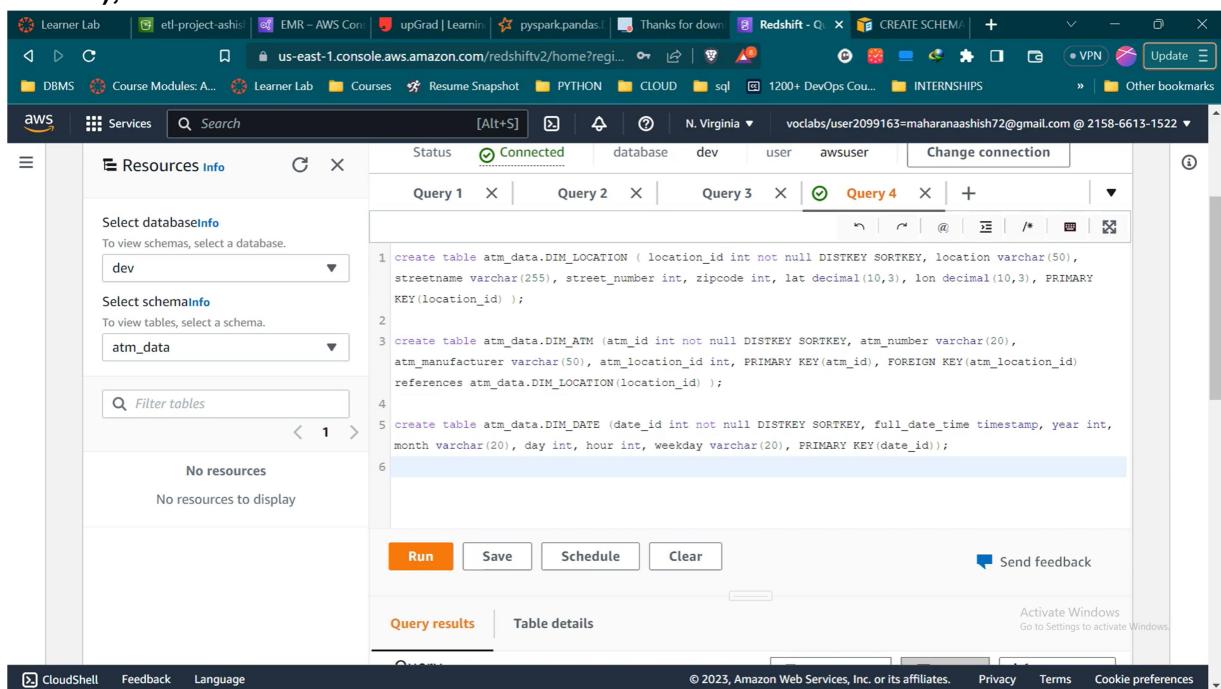
The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar titled 'Resources Info' with dropdown menus for 'Select databaseInfo' (set to 'dev') and 'Select schemaInfo' (set to 'atm_data'). Below these are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, tabs for 'Query results' and 'Table details' are visible. The main area contains four tabs: 'Query 1', 'Query 2', 'Query 3', and 'Query 4'. The 'Query 4' tab is active, displaying the SQL code for creating the DIM_ATM table. The code is as follows:

```
1 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) );
2
3 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) );
```

- 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)
);
```

© Copyright. upGrad Education Pvt. Ltd. All rights reserved



The screenshot shows the AWS Redshift console interface. The top navigation bar includes tabs for Learner Lab, etl-project-ashish, EMR - AWS Com, upGrad | Learning, pypark.pandas, Thanks for down..., Redshift - Q, CREATE SCHEMA, and other options like VPN and Update. Below the navigation bar, there's a toolbar with icons for DBMS, Course Modules, Learner Lab, Courses, Resume Snapshot, PYTHON, CLOUD, sql, 1200+ DevOps Cou..., INTERNSHIPS, and Other bookmarks.

The main area has a sidebar titled "Resources Info" with dropdowns for "Select databaseInfo" (set to dev) and "Select schemaInfo" (set to atm_data). A "Filter tables" input field is also present. The main content area displays three SQL queries:

```

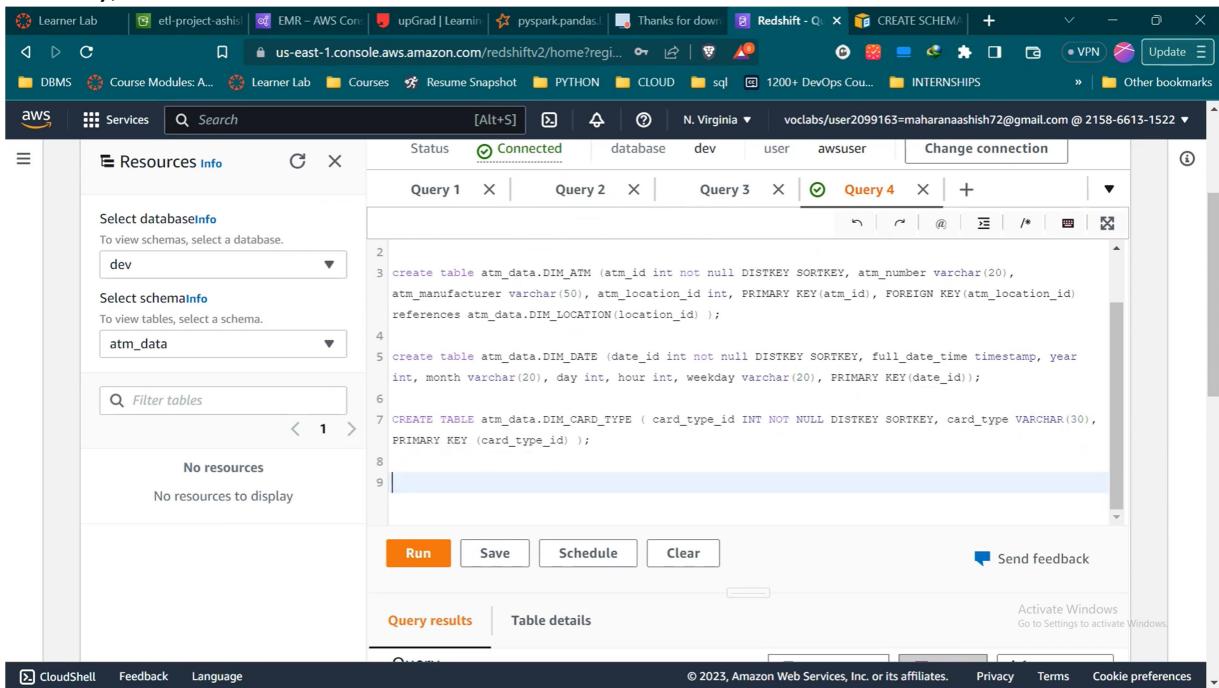
1 create table atm_data.DIM_LOCATION ( location_id int not null DISTKEY SORTKEY, location varchar(50),
streetname varchar(255), street_number int, zipcode int, lat decimal(10,3), lon decimal(10,3), PRIMARY
KEY(location_id) );
2
3 create table atm_data.DIM_ATM (atm_id int not null DISTKEY SORTKEY, atm_number 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) );
4
5 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));
6

```

Below the queries are buttons for "Run", "Save", "Schedule", and "Clear". There's also a "Send feedback" link and a note about activating Windows. At the bottom, there are links for CloudShell, Feedback, Language, and a footer with copyright information: "© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".

- 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)
);
```

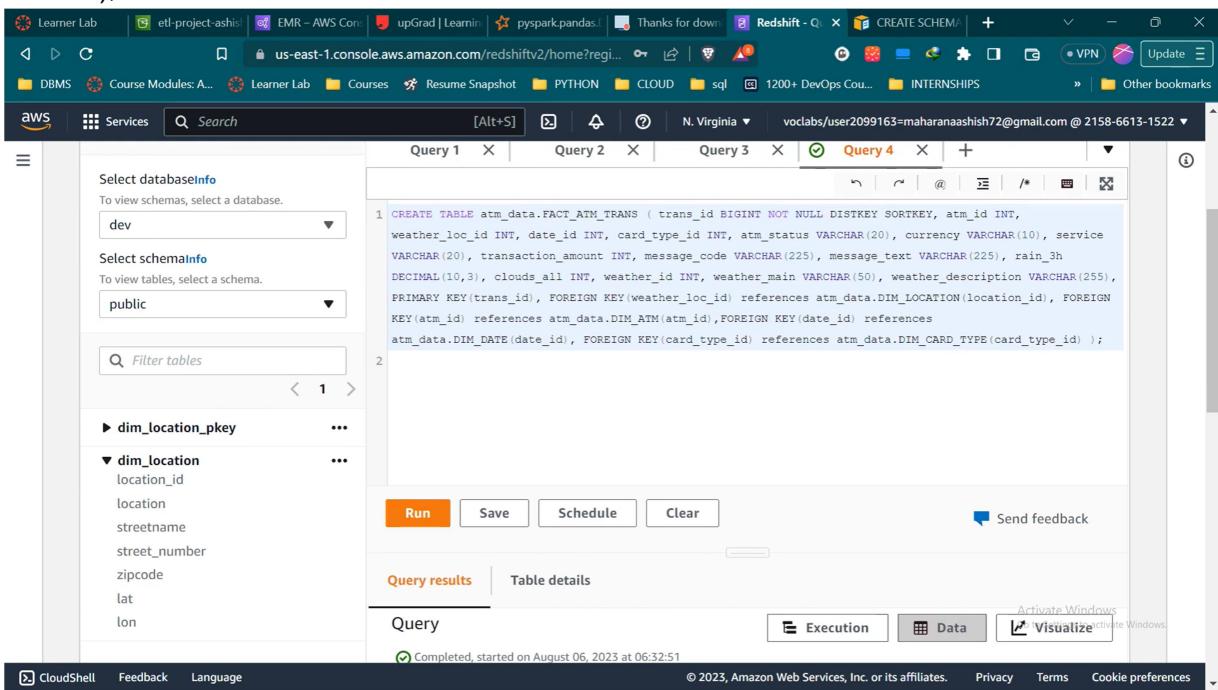


The screenshot shows the AWS Redshift console interface. The top navigation bar includes tabs like Learner Lab, etl-project-ashish, EMR - AWS Con, upGrad | Learni..., pyspark.pandas, Thanks for down..., Redshift - Q, CREATE SCHEMA, VPN, Update, DBMS, Course Modules: A..., Learner Lab, Courses, Resume Snapshot, PYTHON, CLOUD, sql, 1200+ DevOps Cou..., INTERNSHIPS, Other bookmarks. Below the navigation is a search bar and a connection status bar showing 'Connected' to 'dev' database as 'awsuser'. There are four tabs open in the query editor: Query 1, Query 2, Query 3, and Query 4 (the active tab). The code in Query 4 is the provided SQL statement for creating the DIM_CARD_TYPE table. At the bottom of the screen, there are buttons for Run, Save, Schedule, Clear, and Send feedback, along with links for CloudShell, Feedback, Language, and cookie preferences.

```
2
3 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) );
4
5 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));
6
7 CREATE TABLE atm_data.DIM_CARD_TYPE ( card_type_id INT NOT NULL DISTKEY SORTKEY, card_type VARCHAR(30),
PRIMARY KEY (card_type_id) );
8
9
```

- **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 AWS Redshift console interface. The top navigation bar includes tabs like Learner Lab, etl-project-asinh, EMR - AWS Com, upGrad | Learnin, pypark.pandas, Thanks for down..., Redshift - Q, CREATE SCHEMA, and others. Below the navigation is a toolbar with icons for DBMS, Course Modules, Learner Lab, Courses, Resume Snapshot, PYTHON, CLOUD, sql, 1200+ DevOps Cou..., INTERNSHIPS, and Other bookmarks.

The main area has a sidebar titled "Select databaseInfo" with dropdowns for "dev" and "public". It also has a "Filter tables" search bar and a table list for "dim_location_pkey" and "dim_location".

The central workspace contains four tabs: Query 1, Query 2, Query 3, and Query 4. Query 4 is active and displays the SQL code for creating the FACT_ATM_TRANS table. The code is identical to the one provided above.

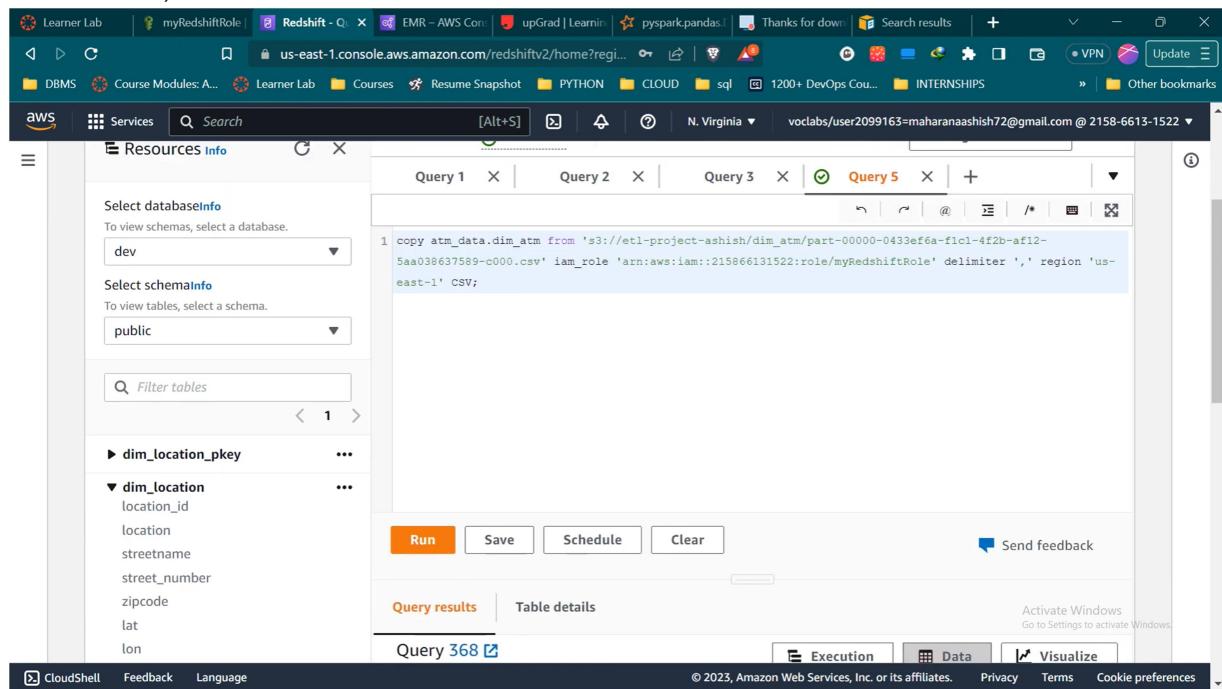
Below the queries are buttons for Run, Save, Schedule, Clear, and Send feedback. At the bottom, there are tabs for Query results and Table details, along with Execution, Data, and Visualize buttons. A status bar at the bottom indicates the query was completed on August 06, 2023, at 06:32:51.

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_atm table

```
copy atm_data.dim_atm from 's3://etl-project-ashish/dim_atm/part-00000-0433efa-f1c1 - 4f2b-af12-5aa038637589-c000.csv'
iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



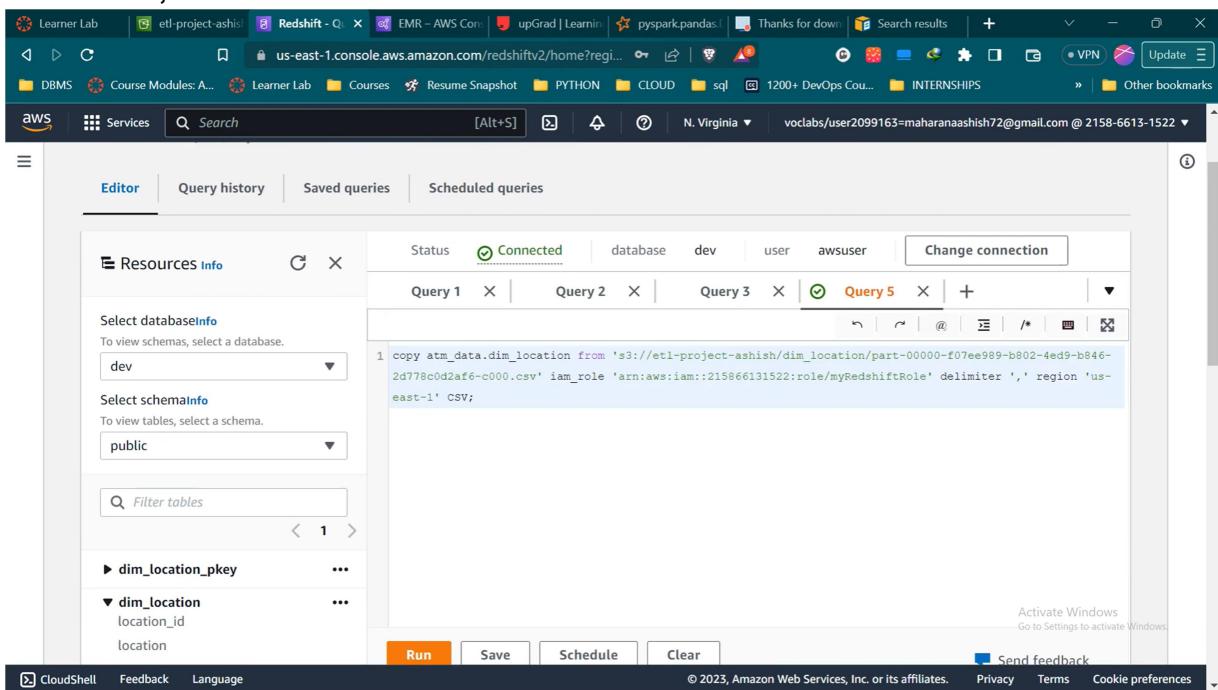
The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar titled 'Resources Info' with dropdowns for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these are sections for 'dim_location_pkey' and 'dim_location'. In the main area, five tabs are open: 'Query 1', 'Query 2', 'Query 3', 'Query 5' (which is active), and 'Query 4'. The 'Query 5' tab contains the following SQL command:

```
copy atm_data.dim_atm from 's3://etl-project-ashish/dim_atm/part-00000-0433efa-f1c1 - 5aa038637589-c000.csv' iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;
```

Below the query tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom of the screen, there are tabs for 'Query results' and 'Table details', along with buttons for 'Execution', 'Data', and 'Visualize'. The status bar at the bottom indicates 'Query 368'.

- Copying the data to dim_location table'

```
copy atm_data.dim_location from 's3://etl-project-ashish/dim_location/part-00000-f07ee989-b902-4ed9-b846-2d778c0d2af6-c000.csv'
iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



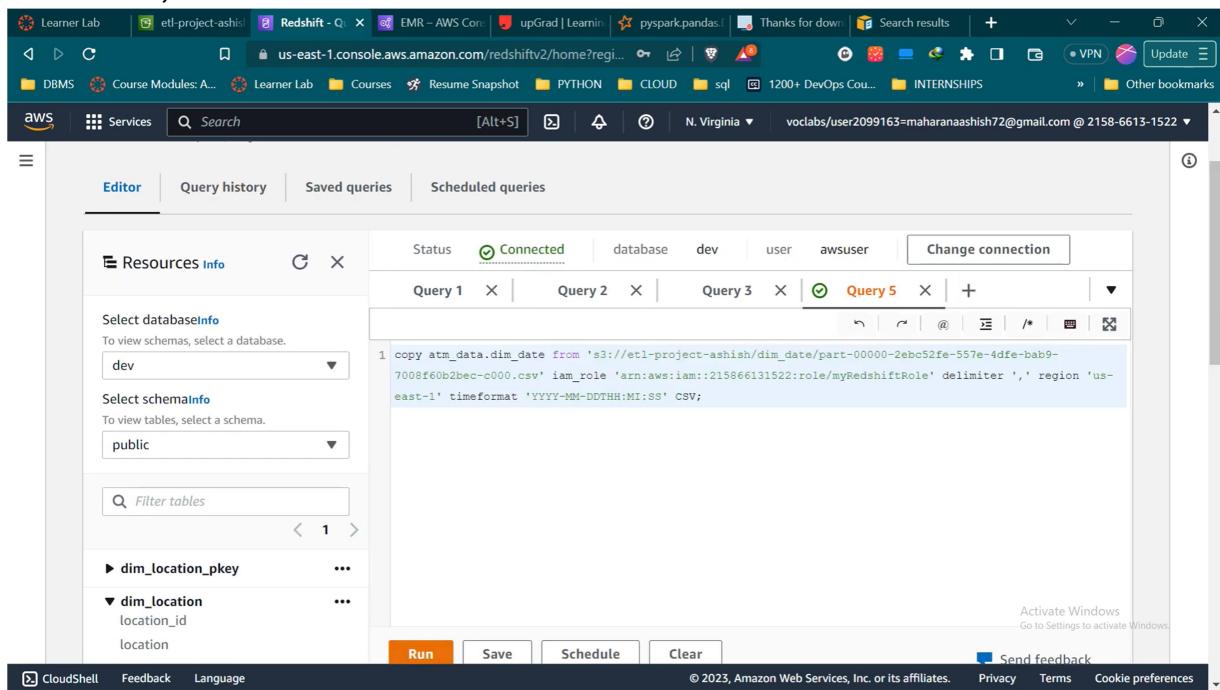
The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar titled 'Resources' with dropdown menus for 'Info' (selected), 'Database', and 'Schema'. Under 'Database', 'dev' is selected. Under 'Schema', 'public' is selected. A search bar at the bottom of this sidebar says 'Filter tables'. Below the sidebar, two tables are listed: 'dim_location_pkey' and 'dim_location'. The 'dim_location' table has three columns: 'location_id', 'location', and an ellipsis (...). On the right, the main area is titled 'Editor' and shows five tabs: 'Query history', 'Saved queries', 'Scheduled queries', 'Query 1', 'Query 2', 'Query 3', and 'Query 5' (which is currently active). The query content is as follows:

```
1 copy atm_data.dim_location from 's3://etl-project-ashish/dim_location/part-00000-f07ee989-b902-4ed9-b846-2d778c0d2af6-c000.csv' iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;
```

Below the query text are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom of the editor, there are links for 'CloudShell', 'Feedback', 'Language', and copyright information: '© 2023, Amazon Web Services, Inc. or its affiliates.' followed by 'Privacy', 'Terms', and 'Cookie preferences'.

- Copying the data to dim_date_table

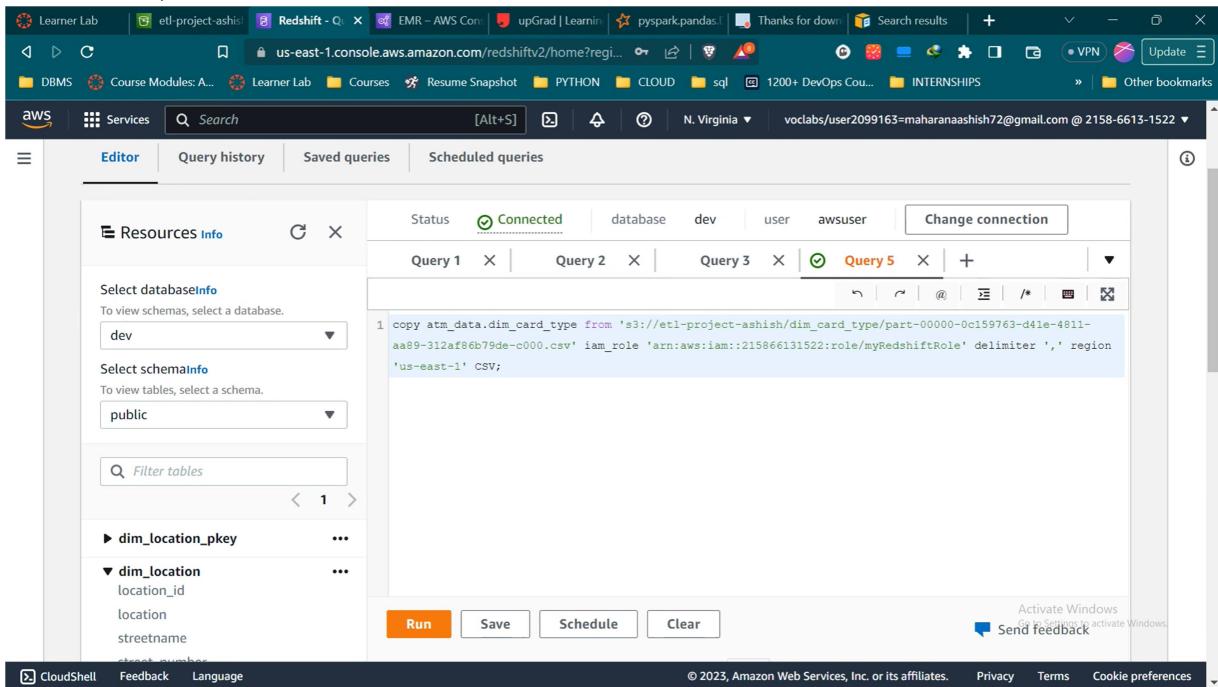
```
copy atm_data.dim_date from 's3://etl-project-ashish/dim_date/part-00000-2ebc52fe-557e-4dfe-bab9-7008f60b2bec-c000.csv'
iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar titled 'Resources Info' with dropdown menus for 'Select databaseInfo' (set to 'dev') and 'Select schemaInfo' (set to 'public'). Below these are sections for 'Filter tables' and two tables listed: 'dim_location_pkey' and 'dim_location'. The 'dim_location' table has rows for 'location_id' and 'location'. On the right, the main area is the 'Editor' tab of the query history. It shows five tabs: 'Query 1', 'Query 2', 'Query 3', 'Query 5' (which contains the copied SQL command), and a new tab. Below the tabs is a toolbar with buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom of the editor, there are links for 'CloudShell', 'Feedback', and 'Language', along with copyright information for Amazon Web Services.

- Copying the data to dim_card_type table

```
copy atm_data.dim_card_type from 's3://etl-project-ashish/dim_card_type/part-00000-0c159763-d41e-4811-aa89-312af86b79de-c000.csv'
iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



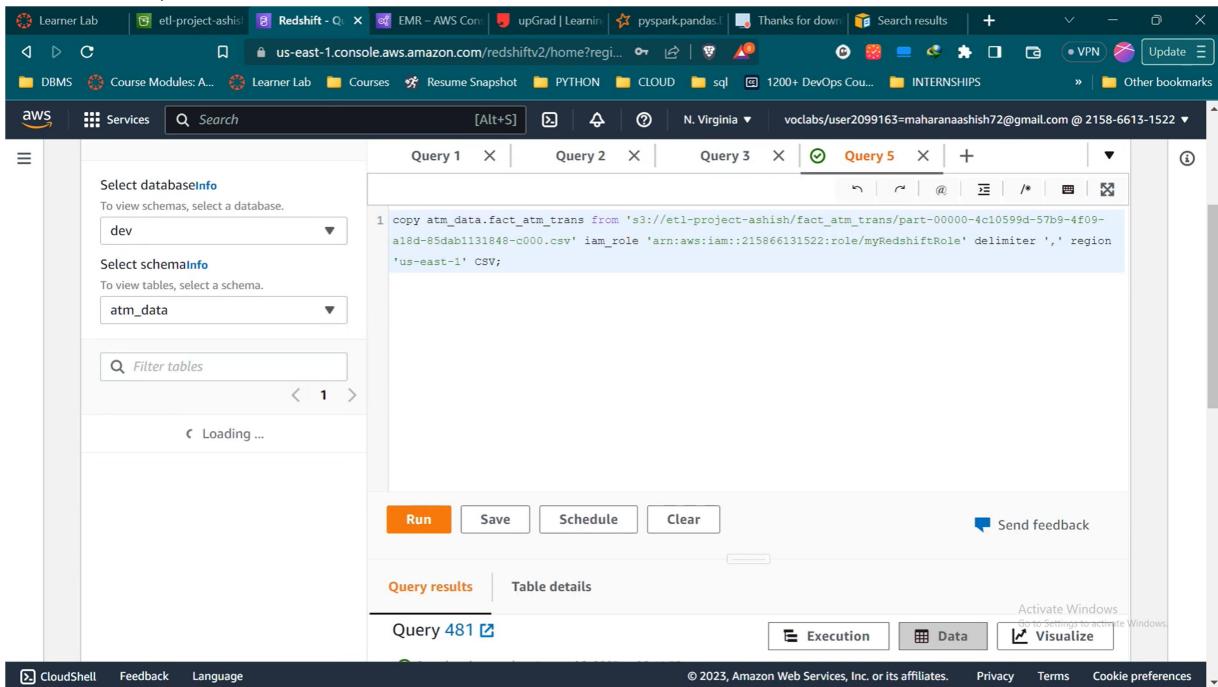
The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar titled 'Resources Info' with dropdowns for 'Select databaseInfo' (set to 'dev') and 'Select schemaInfo' (set to 'public'). Below these are sections for 'Filter tables' and two tables listed: 'dim_location_pkey' and 'dim_location'. The 'dim_location' table has three columns: 'location_id', 'location', and 'streetname'. On the right, the main area shows a query editor with five tabs: 'Query 1', 'Query 2', 'Query 3', 'Query 4', and 'Query 5' (which is currently active). The code in the editor is:

```
copy atm_data.dim_card_type from 's3://etl-project-ashish/dim_card_type/part-00000-0c159763-d41e-4811-aa89-312af86b79de-c000.csv' iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;
```

Below the code are four buttons: 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom right of the editor, there are links for 'Activate Windows', 'Send feedback', and 'CloudShell', along with standard footer links for 'Feedback', 'Language', 'Privacy', 'Terms', and 'Cookie preferences'.

- Copying the data to fact_atm_trans table

```
copy atm_data.fact_atm_trans from 's3://etl-project-ashish/fact_atm_trans/part-00000-4c10599d-57b9-4f09-a18d-85dab1131848-c000.csv'  
iam_role 'arn:aws:iam::215866131522:role/myRedshiftRole'  
delimiter ',' region 'us-east-1'  
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with dropdown menus for 'Select databaseInfo' (set to 'dev') and 'Select schemaInfo' (set to 'atm_data'). Below these are buttons for 'CloudShell', 'Feedback', and 'Language'. The main area contains five tabs: 'Query 1', 'Query 2', 'Query 3', 'Query 5' (which is active), and 'Query 481'. The 'Query 5' tab contains the COPY command shown in the code block above. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of the tabs is a 'Send feedback' button. At the bottom, there are tabs for 'Query results' and 'Table details', and buttons for 'Execution', 'Data', and 'Visualize'. The status bar at the bottom includes links for 'Activate Windows', 'CloudWatch Metrics', 'CloudWatch Metrics Insights', and copyright information: '© 2023, Amazon Web Services, Inc. or its affiliates.'