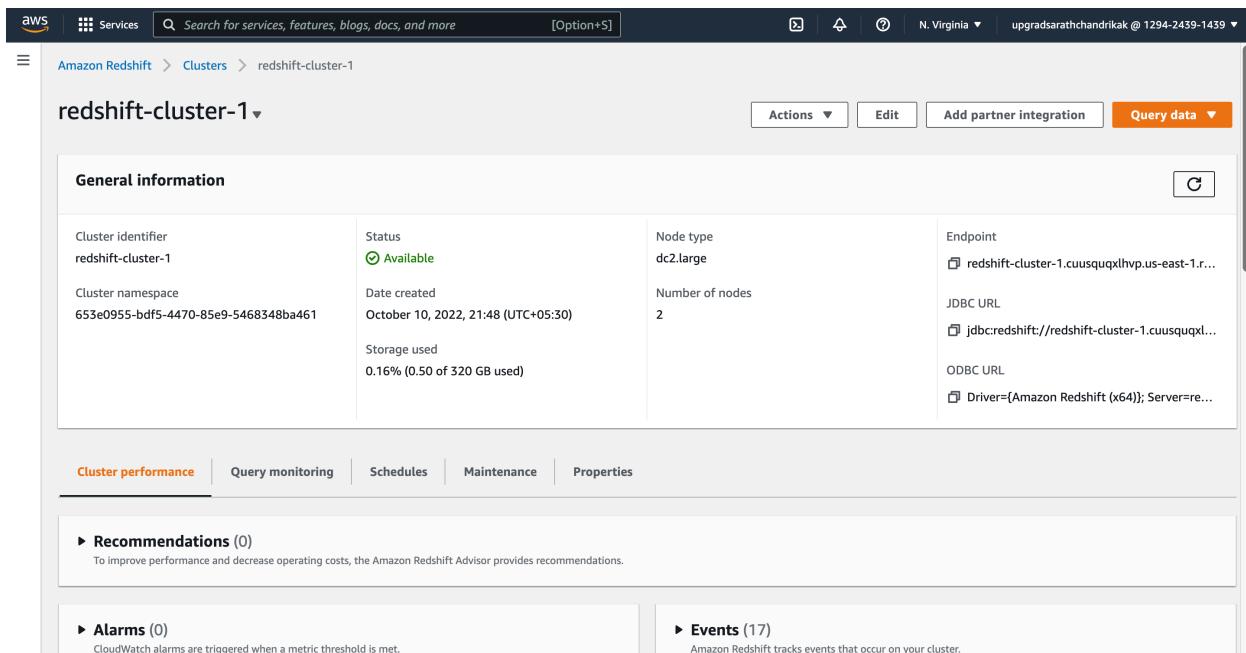


Creation of a Redshift Cluster

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



The screenshot shows the AWS Amazon Redshift Cluster configuration page for a cluster named "redshift-cluster-1".

General information:

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-1	Available	dc2.large	redshift-cluster-1.cuusquqlhv.us-east-1.r...
Cluster namespace	Date created	Number of nodes	JDBC URL
653e0955-bdf5-4470-85e9-5468348ba461	October 10, 2022, 21:48 (UTC+05:30)	2	jdbc:redshift://redshift-cluster-1.cuusquql...
	Storage used		ODBC URL
	0.16% (0.50 of 320 GB used)		Driver={Amazon Redshift (x64)}; Server=re...

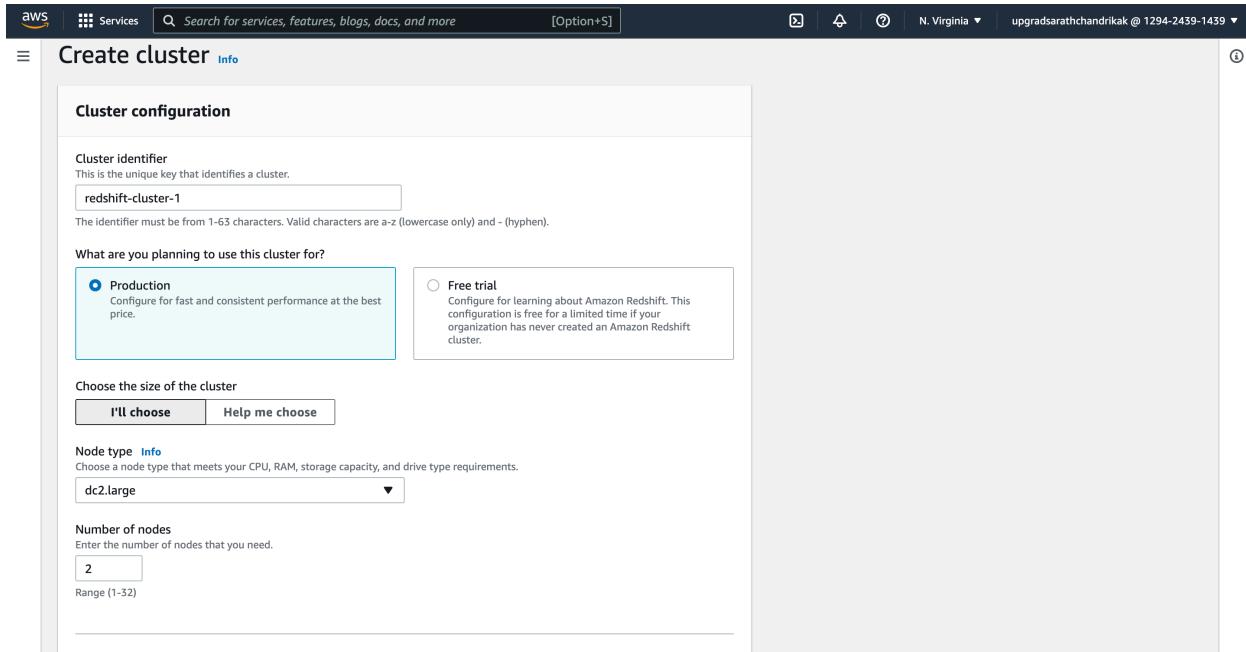
Cluster performance: (Selected tab)

Recommendations (0): To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

Alarms (0): CloudWatch alarms are triggered when a metric threshold is met.

Events (17): Amazon Redshift tracks events that occur on your cluster.

Screenshots of steps and configurations selected to create cluster



Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

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.

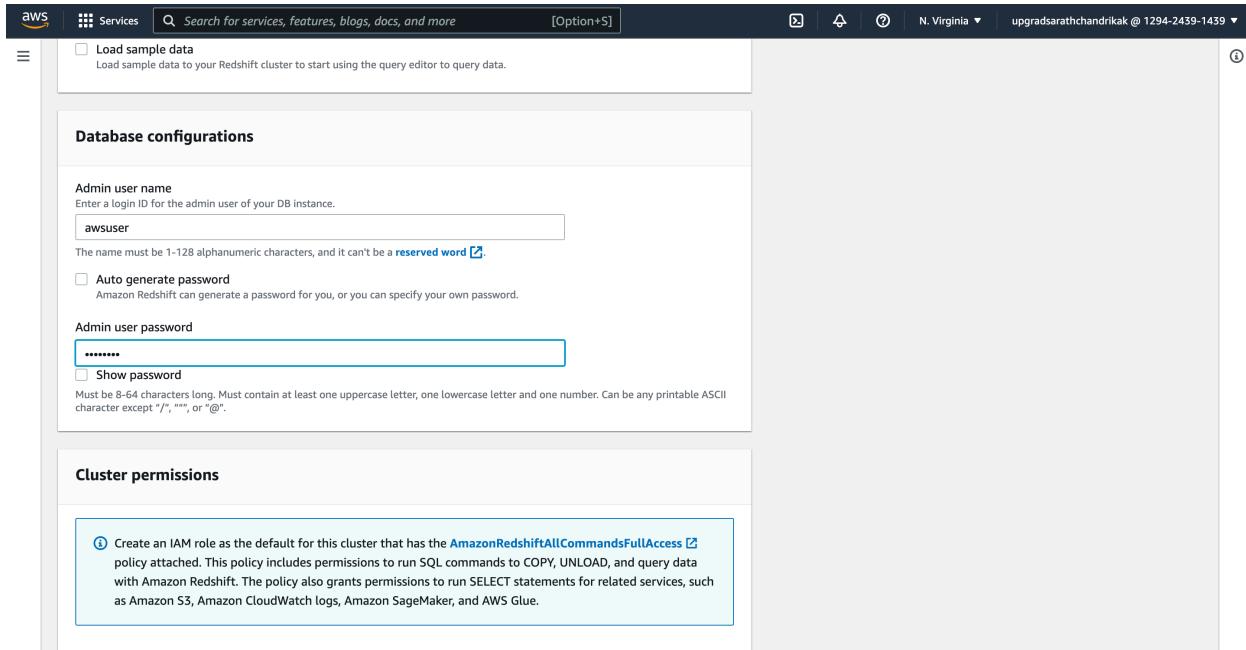
Free trial
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

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

Number of nodes
Enter the number of nodes that you need.

Range (1-32)



Load sample data
Load sample data to your Redshift cluster to start using the query editor to query data.

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

Auto generate password
Amazon Redshift can generate a password for you, or you can specify your own password.

Admin user password

Show password
Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except "/", "\", or "@".

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.

AWS Services Search for services, features, blogs, docs, and more [Option+S] N. Virginia upgradsarathchandrikak @ 1294-2439-1439 ⓘ

Network and security Info

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.

my-vpc-vpc
vpc-09075cc4e7300e4d4

ⓘ You can't change the VPC associated with this cluster after the cluster has been created. [Learn more](#) X

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

my-vpc-security X
sg-0bdb78a5e2e3d5a41

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 [Option+S] N. Virginia upgradsarathchandrikak @ 1294-2439-1439 ⓘ

Database configurations

Database name
Specify a database name to create an additional database.

dev

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.

1678

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

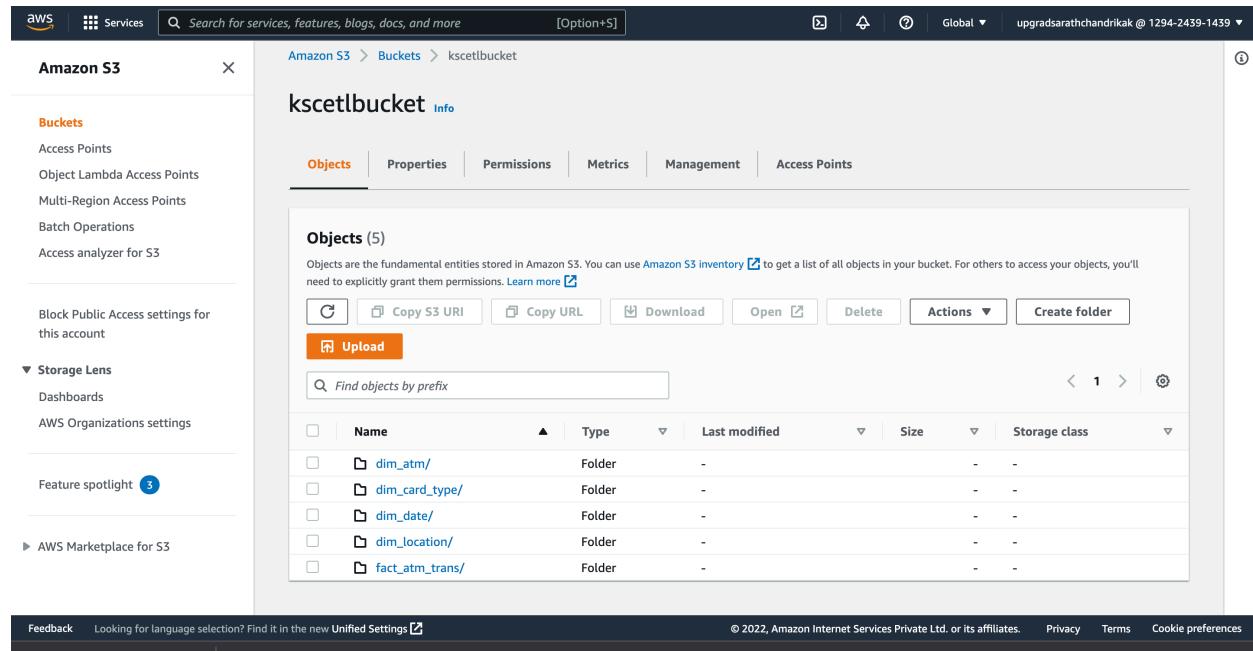
Encryption
Encrypt all data on your cluster.

Disabled Use AWS Key Management Service (AWS KMS) Use a hardware security module (HSM)

Maintenance

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

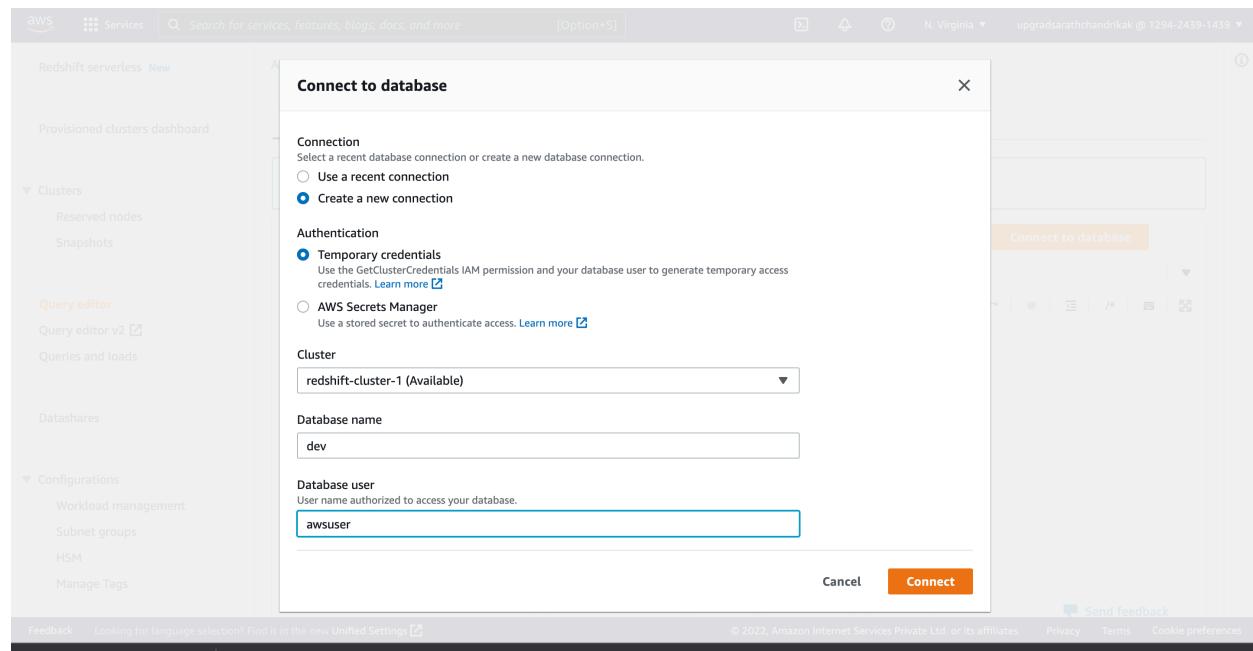
Viewing data from S3 buckets



The screenshot shows the AWS S3 console interface. On the left, the navigation pane includes 'Amazon S3', 'Buckets', 'Access Points', 'Object Lambda Access Points', 'Multi-Region Access Points', 'Batch Operations', 'Access analyzer for S3', 'Block Public Access settings for this account', 'Storage Lens', 'Dashboards', 'AWS Organizations settings', 'Feature spotlight (3)', and 'AWS Marketplace for S3'. The main content area displays the 'kscetlbucket' bucket's details. The 'Objects' tab is selected, showing 5 objects. The table lists the following items:

Name	Type	Last modified	Size	Storage class
dim_atm/	Folder	-	-	-
dim_card_type/	Folder	-	-	-
dim_date/	Folder	-	-	-
dim_location/	Folder	-	-	-
fact_atm_trans/	Folder	-	-	-

Connecting to database

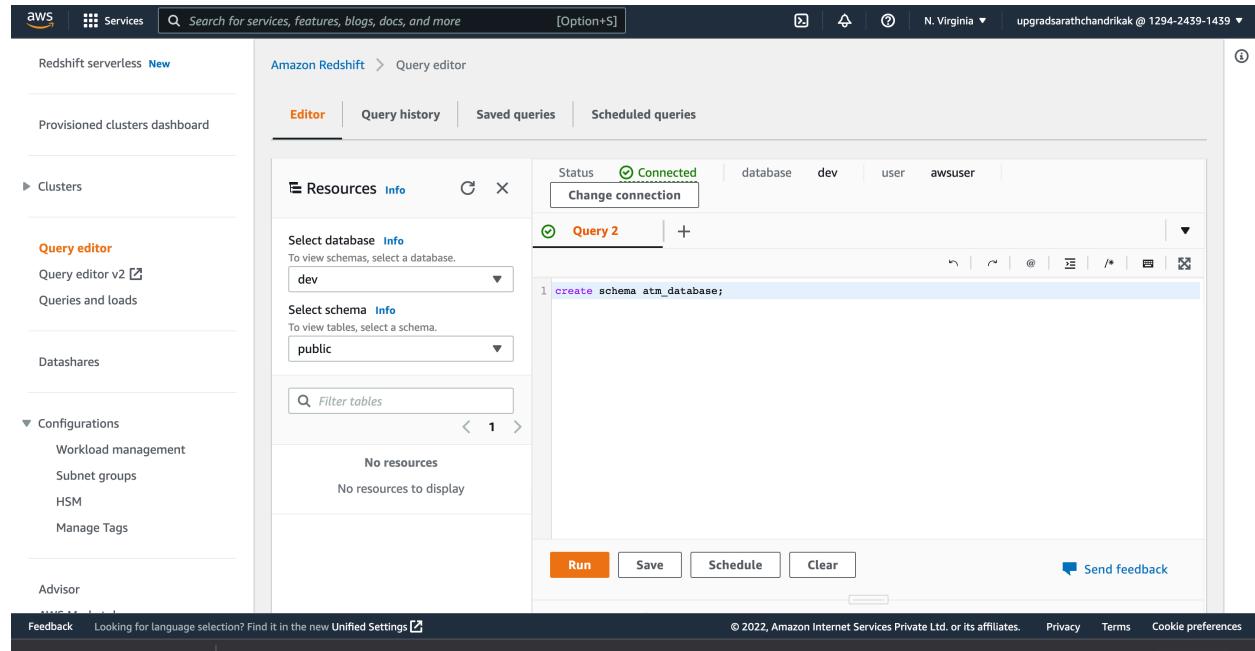


The screenshot shows the 'Connect to database' dialog box over the AWS Redshift serverless console. The left sidebar includes 'Provisioned clusters dashboard', 'Clusters' (with 'Reserved nodes' and 'Snapshots'), 'Query editor' (with 'Query editor v2' and 'Queries and loads'), 'Datashares', 'Configurations' (with 'Workload management', 'Subnet groups', 'HSM', and 'Manage Tags'), and 'Feedback' (with 'Send feedback'). The main dialog has the following fields:

- Connection**: 'Select a recent database connection or create a new database connection.' Options: Use a recent connection (disabled) and Create a new connection.
- Authentication**: Temporary credentials (selected) and AWS Secrets Manager.
- Cluster**: A dropdown menu showing 'redshift-cluster-1 (Available)'.
- Database name**: The input field contains 'dev'.
- Database user**: The input field contains 'awsuser'.
- Buttons**: 'Cancel' and 'Connect'.

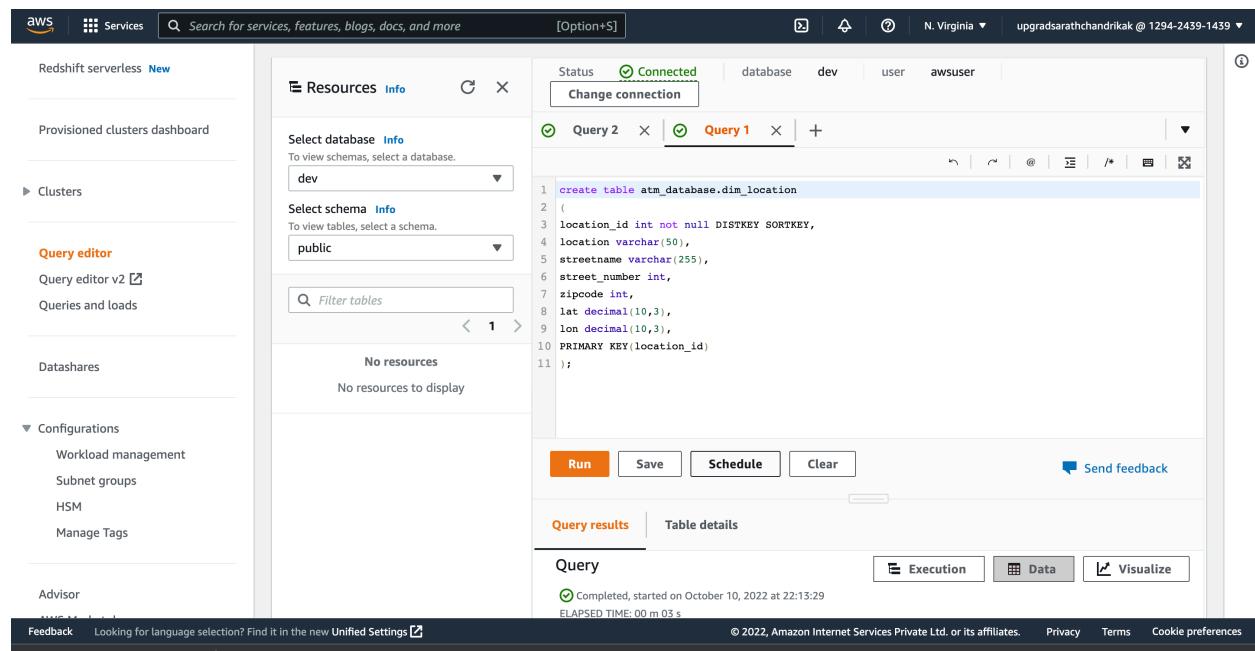
Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

```
create schema atm_database;
```



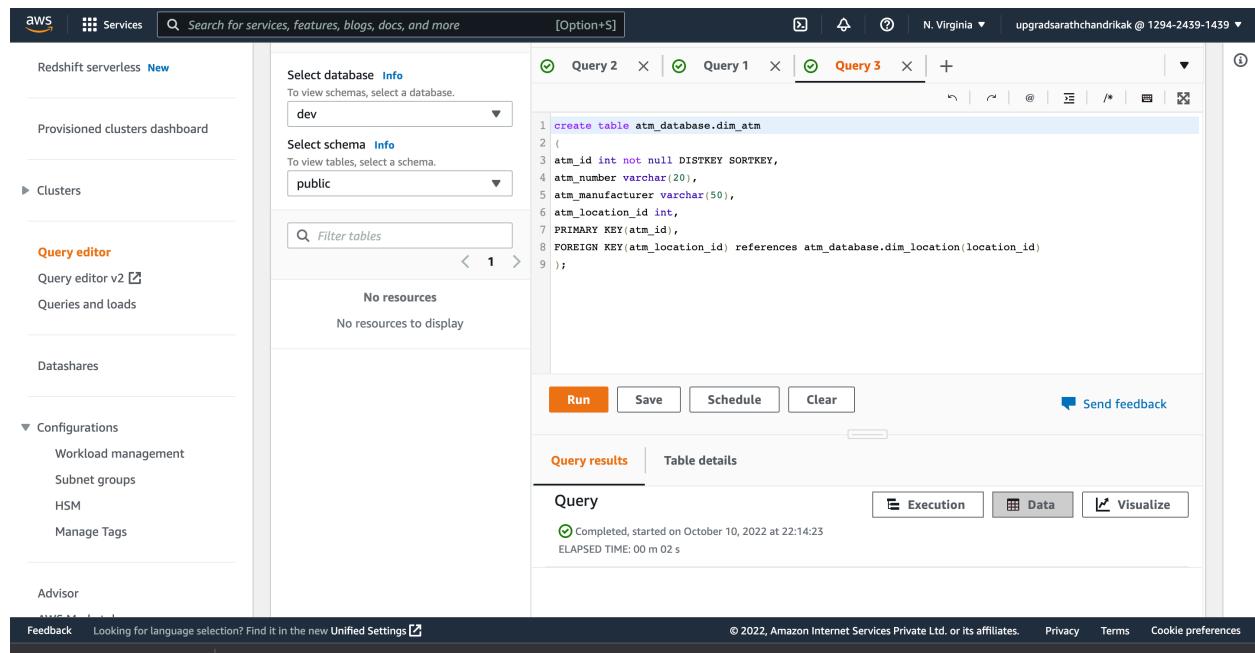
The screenshot shows the Amazon Redshift Query Editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is selected), 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags', and 'Advisor'. The main area has tabs for 'Editor' (selected), 'Query history', 'Saved queries', and 'Scheduled queries'. The 'Editor' tab shows a connection status of 'Connected' to 'database dev user awsuser'. Below that, it says 'Select database' with 'dev' selected and 'Select schema' with 'public' selected. A table list shows 'No resources' with 'No resources to display'. At the bottom, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear', along with a 'Send feedback' link.

```
create table atm_database.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)
);
```



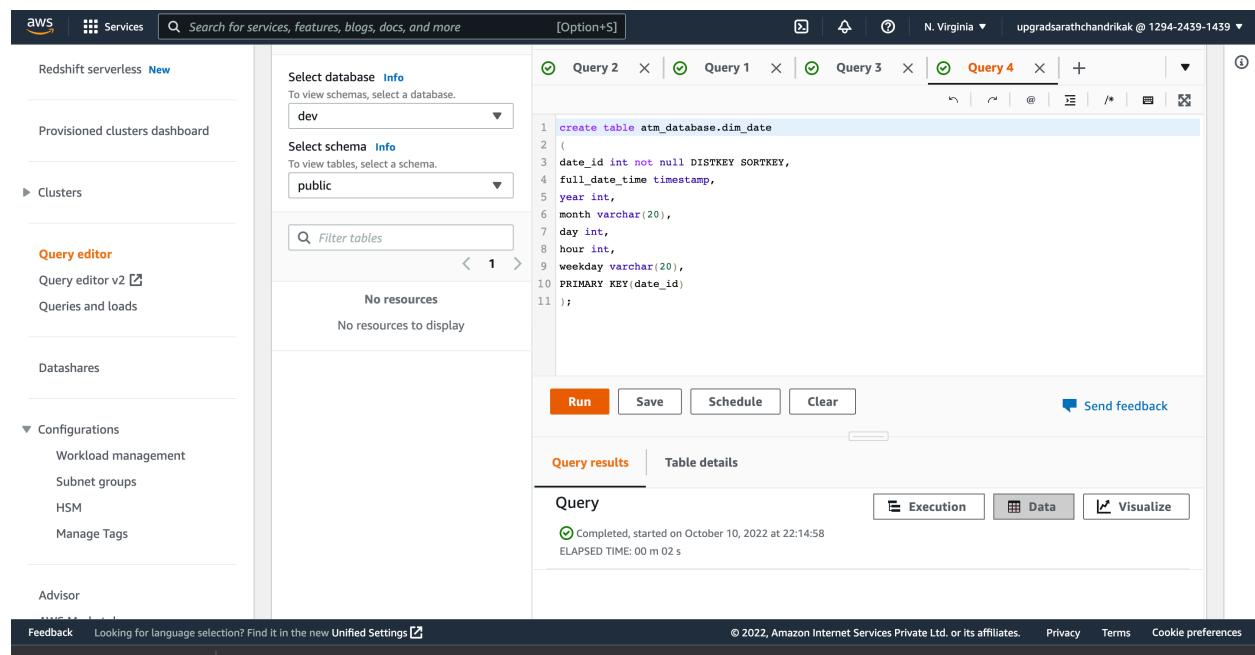
The screenshot shows the AWS Redshift Serverless Query Editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is selected), 'Queries and loads', 'DataShares', 'Configurations', 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags', and 'Advisor'. The main area has tabs for 'Resources' and 'Info'. Under 'Info', it shows 'Status Connected', 'database dev', 'user awsuser', and a 'Change connection' button. There are two query tabs: 'Query 2' (green) and 'Query 1' (orange). The orange tab contains the SQL code provided above. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there are tabs for 'Query results' (selected), 'Table details', and buttons for 'Execution', 'Data', and 'Visualize'. A status message at the bottom says 'Completed, started on October 10, 2022 at 22:13:29 ELAPSED TIME: 00 m 03 s'. The footer includes links for 'Feedback', 'Language selection', 'Unified Settings', '© 2022, Amazon Internet Services Private Ltd. or its affiliates.', 'Privacy', 'Terms', and 'Cookie preferences'.

```
create table atm_database.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_database.dim_location(location_id)
);
```



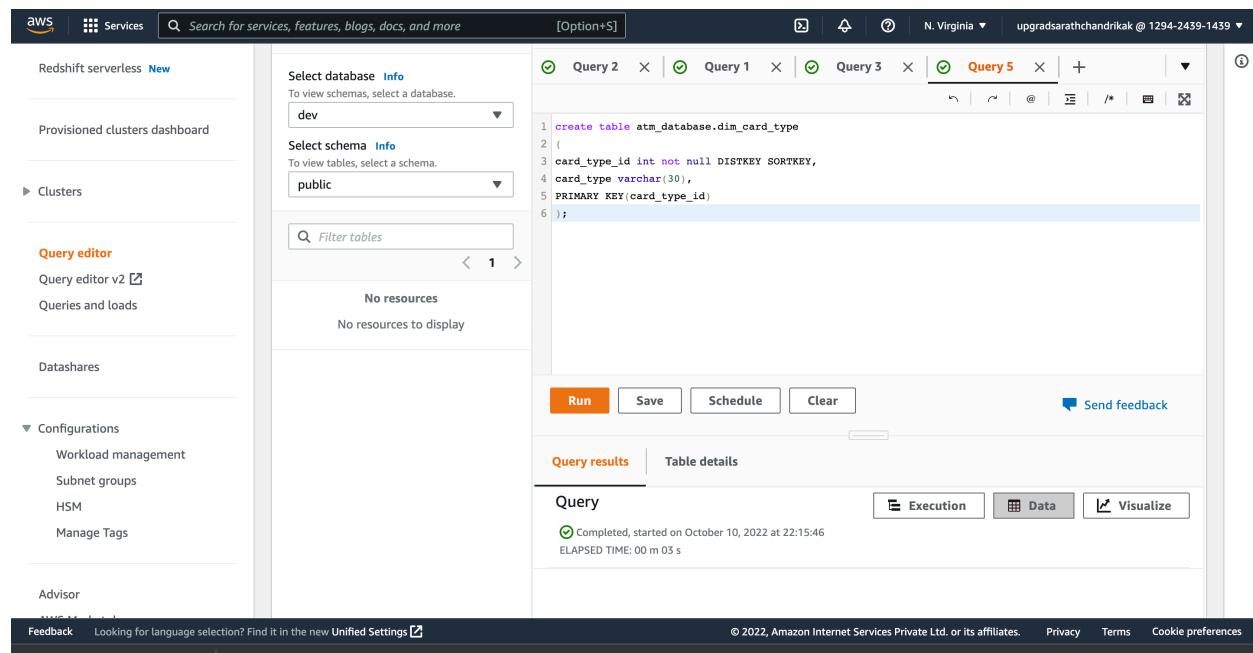
The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is highlighted in orange), 'Queries and loads', 'DataShares', 'Configurations', 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags', and 'Advisor'. The main area has dropdown menus for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). A search bar at the top says 'Search for services, features, blogs, docs, and more'. Below the search bar, there are tabs for 'Query 2', 'Query 1', and 'Query 3' (which is currently active). The query editor contains the SQL code provided above. At the bottom of the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of the editor, there's a 'Send feedback' link. Below the editor, there's a 'Query results' section with tabs for 'Execution', 'Data', and 'Visualize'. It shows a status message: 'Completed, started on October 10, 2022 at 22:14:23' and 'ELAPSED TIME: 00 m 02 s'. At the very bottom of the page, there's a footer with links for 'Feedback', 'Unified Settings', '© 2022, Amazon Internet Services Private Ltd. or its affiliates.', 'Privacy', 'Terms', and 'Cookie preferences'.

```
create table atm_database.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 Redshift Query Editor interface. On the left, the navigation sidebar includes options like Redshift serverless, Provisioned clusters dashboard, Clusters, Query editor (selected), Datashares, Configurations (Workload management, Subnet groups, HSM, Manage Tags), and Advisor. The main area has tabs for Query 2, Query 1, Query 3, and Query 4. Query 4 is active, displaying the SQL code for creating the dim_date table. Below the code are buttons for Run, Save, Schedule, and Clear. At the bottom, there are tabs for Query results and Table details, along with Execution, Data, and Visualize buttons. The status bar at the bottom indicates the query completed on October 10, 2022, at 22:14:58, with an elapsed time of 00 m 02 s.

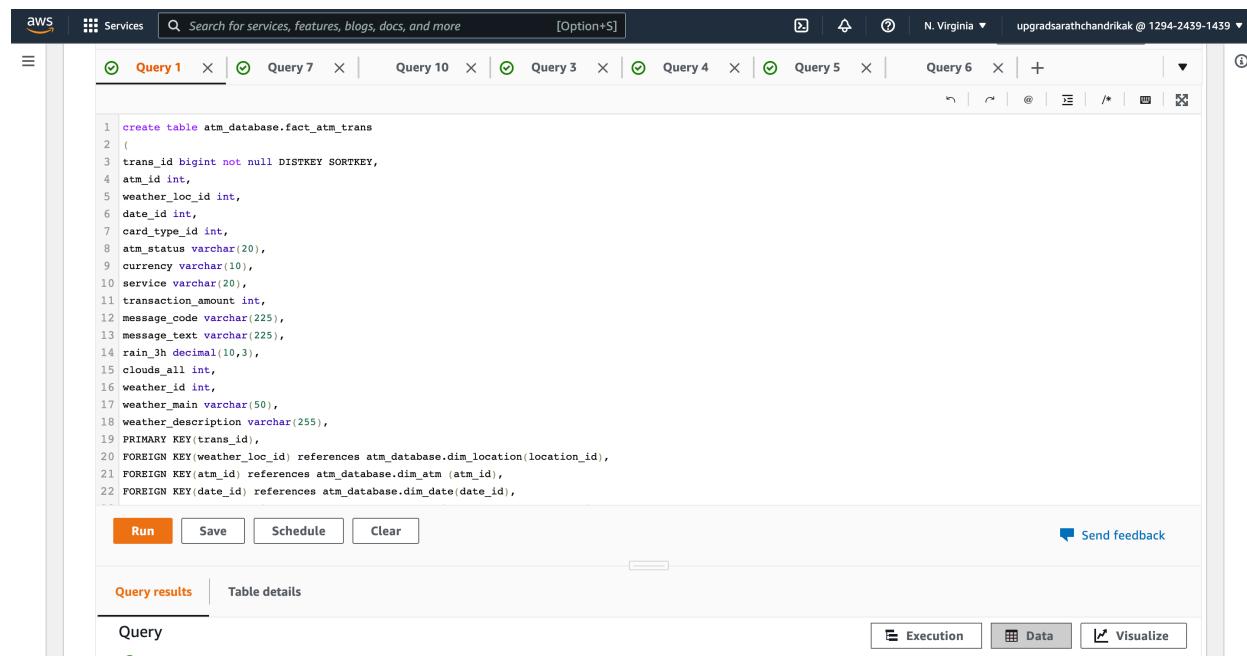
```
create table atm_database.dim_card_type
(
    card_type_id int not null DISTKEY SORTKEY,
    card_type varchar(30),
    PRIMARY KEY(card_type_id)
);
```



The screenshot shows the AWS Redshift serverless Query editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is selected), 'Queries and loads', 'DataShares', 'Configurations' (with sub-options 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags'), and 'Advisor'. The main area has dropdown menus for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these is a 'Filter tables' search bar. A message 'No resources' is displayed. On the right, five tabs are open: 'Query 2', 'Query 1', 'Query 3', 'Query 5' (which is active), and a '+' tab. The 'Query 5' tab contains the SQL code shown at the top of the page. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. At the bottom, there are tabs for 'Query results' (which is selected) and 'Table details', along with buttons for 'Execution', 'Data', and 'Visualize'. A status message indicates the query completed on October 10, 2022, at 22:15:46 with an elapsed time of 00 m 03 s. The footer includes links for 'Feedback', 'Unified Settings', '© 2022, Amazon Internet Services Private Ltd. or its affiliates.', 'Privacy', 'Terms', and 'Cookie preferences'.

```

create table atm_database.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_database.dim_location(location_id),
  FOREIGN KEY(atm_id) references atm_database.dim_atm (atm_id),
  FOREIGN KEY(date_id) references atm_database.dim_date(date_id),
  FOREIGN KEY(card_type_id) references atm_database.dim_card_type(card_type_id)
);
  
```



The screenshot shows the AWS Lambda SQL interface. At the top, there are tabs for 'Services' and a search bar. On the right, it shows the user 'upgradsarathchandrikak @ 1294-2439-1439'. Below the tabs, there are six open queries labeled 'Query 1' through 'Query 6'. Query 1 is the one we just created. The code for Query 1 is displayed in the main editor area:

```

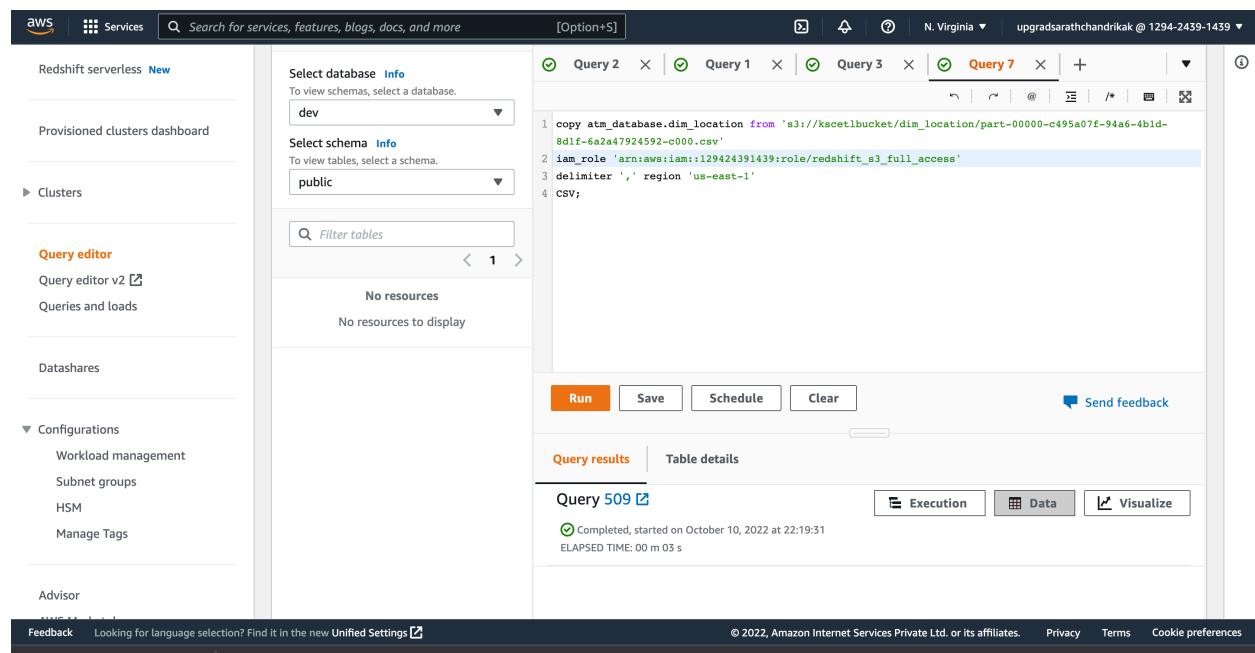
1 create table atm_database.fact_atm_trans
2 (
3   trans_id bigint not null DISTKEY SORTKEY,
4   atm_id int,
5   weather_loc_id int,
6   date_id int,
7   card_type_id int,
8   atm_status varchar(20),
9   currency varchar(10),
10  service varchar(20),
11  transaction_amount int,
12  message_code varchar(225),
13  message_text varchar(225),
14  rain_3h decimal(10,3),
15  clouds_all int,
16  weather_id int,
17  weather_main varchar(50),
18  weather_description varchar(255),
19  PRIMARY KEY(trans_id),
20  FOREIGN KEY(weather_loc_id) references atm_database.dim_location(location_id),
21  FOREIGN KEY(atm_id) references atm_database.dim_atm (atm_id),
22  FOREIGN KEY(date_id) references atm_database.dim_date(date_id),
  
```

Below the code, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of the code area, there is a 'Send feedback' button. At the bottom, there are tabs for 'Query results' and 'Table details', and a 'Query' input field. On the far right, there are buttons for 'Execution', 'Data', and 'Visualize'.

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

```
copy atm_database.dim_location from 's3://kscetlbucket/dim_location/part-00000-c495a07f-94a6-4b1d-8d1f-6a2a47924592-c000.csv'  
iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```

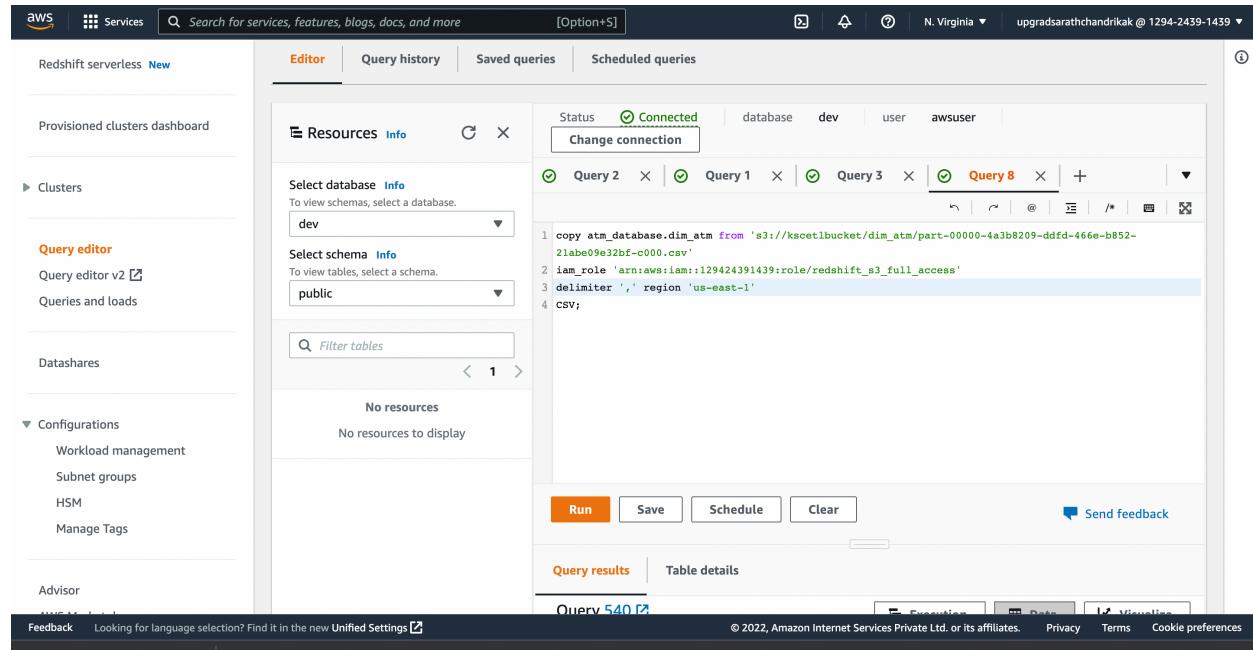


The screenshot shows the AWS Management Console interface for a Redshift cluster. On the left, the navigation pane includes options like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor v2' (selected), 'Queries and loads', 'DataShares', 'Configurations' (with 'Workload management', 'Subnet groups', 'HSM', and 'Manage Tags'), and 'Advisor'. The main area is the 'Query editor' where a query has been run. The query text is:

```
1 copy atm_database.dim_location from 's3://kscetlbucket/dim_location/part-00000-c495a07f-94a6-4b1d-8d1f-6a2a47924592-c000.csv'  
2 iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'  
3 delimiter ',' region 'us-east-1'  
4 CSV;
```

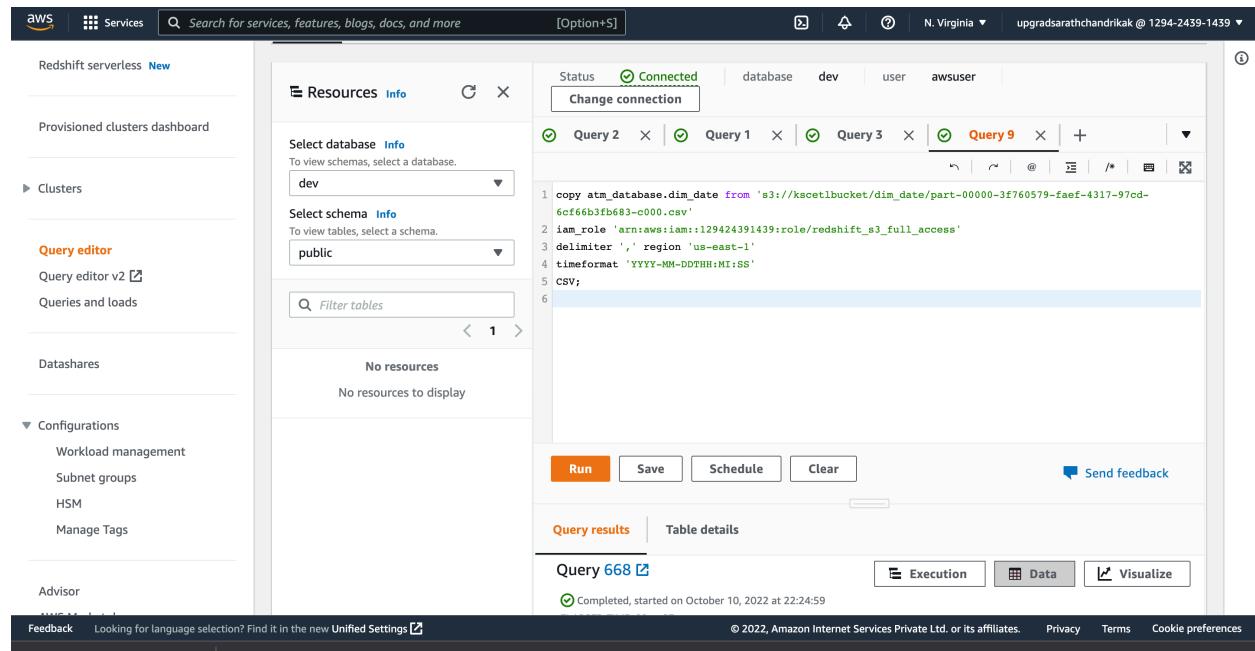
The 'Run' button is highlighted in orange. Below the query results, it says 'Completed, started on October 10, 2022 at 22:19:31' and 'ELAPSED TIME: 00 m 03 s'. There are tabs for 'Query results' and 'Table details', and buttons for 'Execution', 'Data', and 'Visualize'.

```
copy atm_database.dim_atm from 's3://kscetlbucket/dim_atm/part-00000-4a3b8209-ddfd-466e-b852-21abe09e32bf-c000.csv'  
iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```



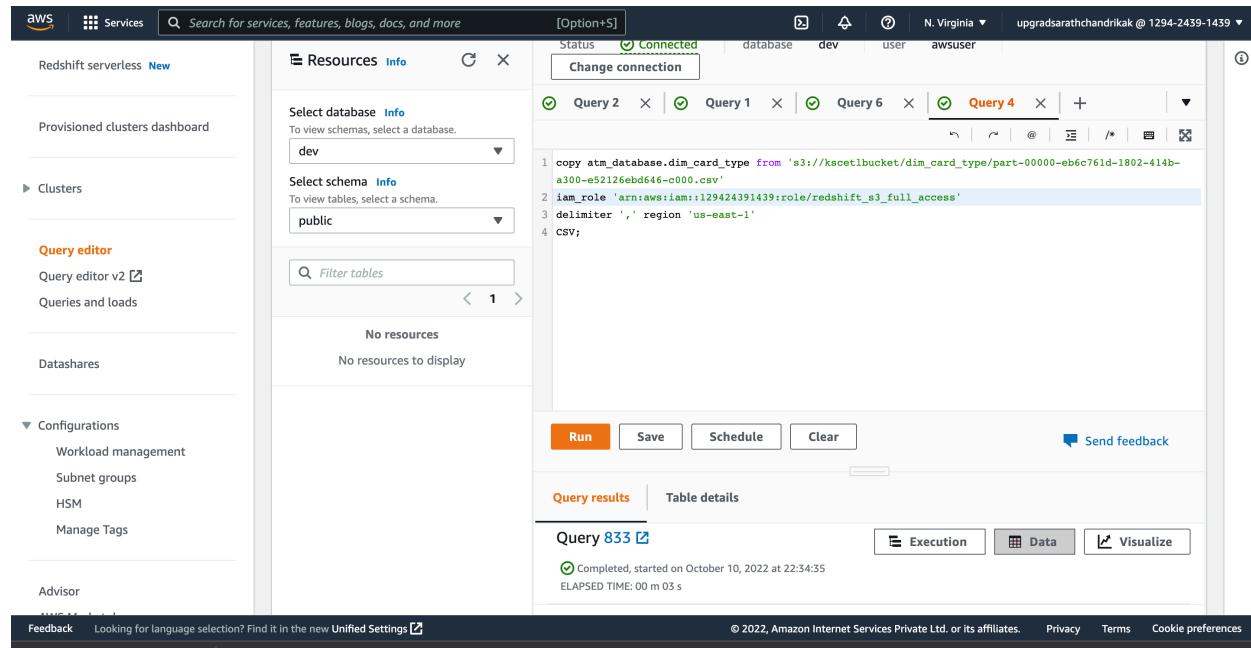
The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar with navigation links like 'Provisioned clusters dashboard', 'Clusters', 'Query editor v2', 'Queries and loads', 'Databricks', 'Configurations' (with 'Workload management', 'Subnet groups', 'HSM', and 'Manage Tags'), and 'Advisor'. The main area is titled 'Editor' and shows a 'Resources' section with dropdown menus for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these are tabs for 'Filter tables' and 'No resources'. A status bar at the top indicates 'Status Connected' and 'database dev user awsuser'. In the center, there are four tabs labeled 'Query 1', 'Query 2', 'Query 3', and 'Query 8'. The 'Query 8' tab is active, displaying the SQL code provided in the question. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there are tabs for 'Query results', 'Table details', and 'Visualizations', along with a 'Feedback' link and copyright information.

```
copy atm_database.dim_date from 's3://kscetlbucket/dim_date/part-00000-3f760579-faef-
4317-97cd-6cf66b3fb683-c000.csv'
iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



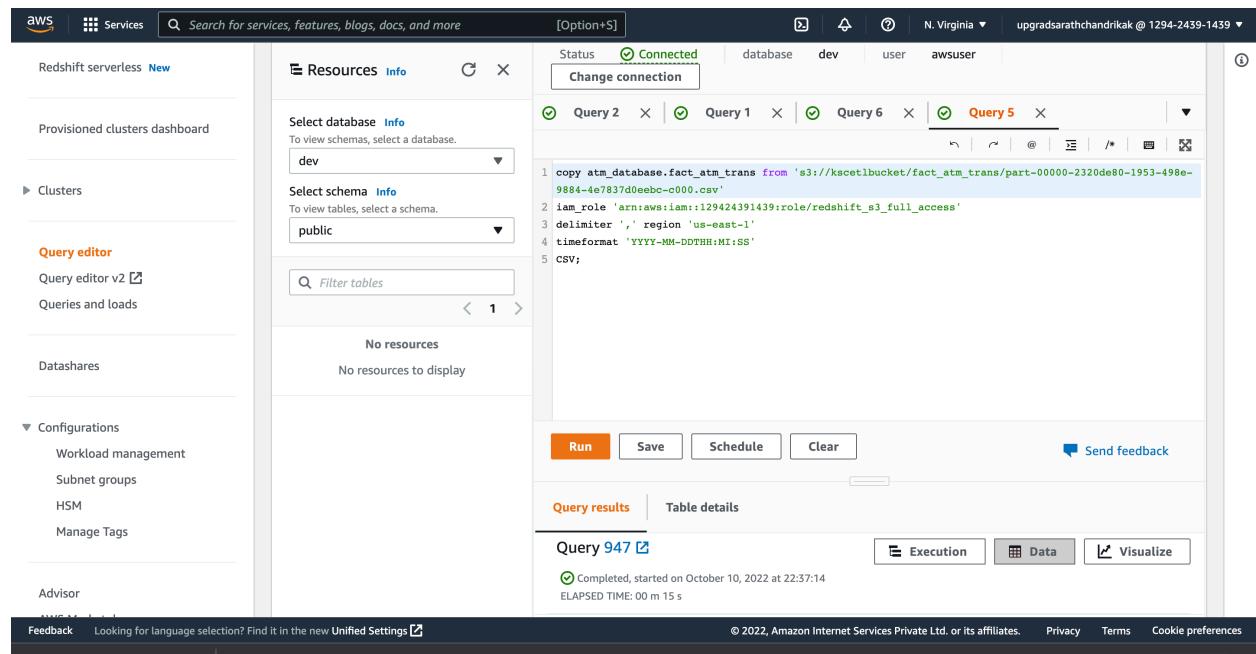
The screenshot shows the AWS Redshift Serverless Query Editor. On the left, there's a sidebar with navigation links like 'Redshift serverless New', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is selected), 'Configurations', 'Advisor', and 'Feedback'. The main area has tabs for 'Resources' and 'Info'. Under 'Info', 'Select database' is set to 'dev' and 'Select schema' is set to 'public'. A search bar 'Filter tables' is present. Below these, it says 'No resources' and 'No resources to display'. To the right, the query editor interface is shown with a status bar indicating 'Status Connected' and 'database dev user awouser'. There are tabs for 'Query 1' through 'Query 9'. The 'Query 9' tab is active, displaying the copied query. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' button is also visible. At the bottom, there are tabs for 'Query results' (which is selected) and 'Table details', along with buttons for 'Execution', 'Data', and 'Visualize'. The status bar at the bottom right shows 'Completed, started on October 10, 2022 at 22:24:59'.

```
copy atm_database.dim_card_type from 's3://kscetlbucket/dim_card_type/part-00000-eb6c761d-1802-414b-a300-e52126ebd646-c000.csv'
iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (selected), 'Datashares', 'Configurations', 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags', and 'Advisor'. The main area has tabs for 'Resources' (selected) and 'Info'. Under 'Resources', it says 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). Below these are 'Filter tables' and a 'No resources' message. The right side shows a query editor with four tabs: 'Query 2', 'Query 1', 'Query 6', and 'Query 4' (selected). The code in 'Query 4' is the one shown at the top of the page. Below the tabs are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there's a 'Query results' section showing 'Completed' status, the start time 'October 10, 2022 at 22:34:35', and an 'ELAPSED TIME: 00 m 03 s'. There are also tabs for 'Table details', 'Execution', 'Data', and 'Visualize'.

```
copy atm_database.fact_atm_trans from 's3://kscetlbucket/fact_atm_trans/part-00000-2320de80-1953-498e-9884-4e7837d0eebc-c000.csv'
iam_role 'arn:aws:iam::129424391439:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



The screenshot shows the AWS Redshift Query Editor interface. On the left, there's a sidebar with navigation links like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (which is highlighted in orange), 'Queries and loads', 'Dashboards', 'Configurations' (with sub-options 'Workload management', 'Subnet groups', 'HSM', 'Manage Tags'), and 'Advisor'. The main area has tabs for 'Resources' (selected) and 'Info'. Under 'Info', it shows 'Status: Connected' and dropdowns for 'database' (set to 'dev') and 'user' (set to 'awsuser'). Below these are dropdowns for 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). A search bar 'Filter tables' is also present. The central part of the screen displays a query editor with six tabs: 'Query 2', 'Query 1', 'Query 6', 'Query 5' (which is currently active and contains the COPY command shown in the code block above), and two others. Below the tabs is a code editor window showing the COPY command. At the bottom of the editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear', along with a 'Send feedback' link. Further down are tabs for 'Query results' and 'Table details', and a section for 'Query 947' which shows it completed successfully. The footer includes links for 'Feedback', 'Unified Settings', 'Privacy', 'Terms', and 'Cookie preferences'.