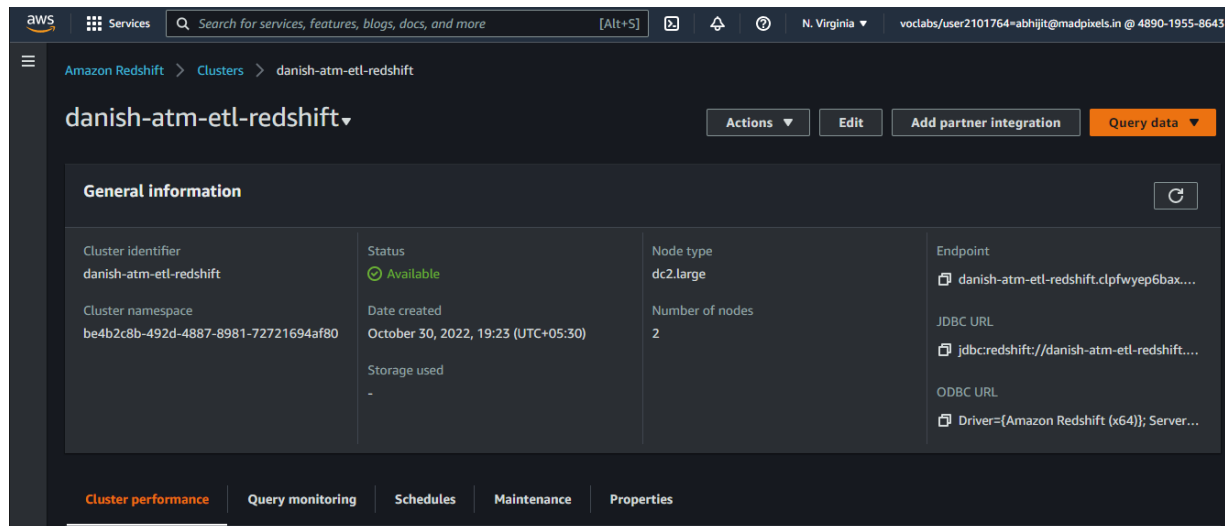


## Creation of a Redshift Cluster

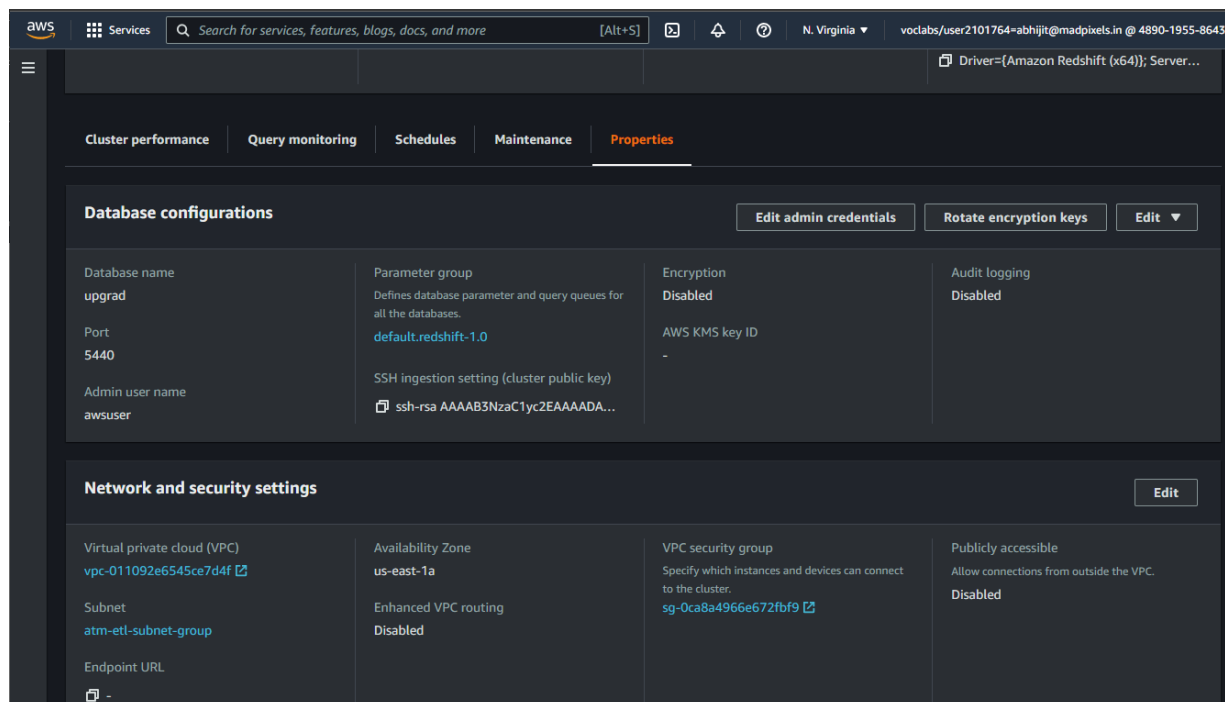
Screenshots of the configuration of the Redshift cluster that you have created:



The screenshot shows the AWS Redshift console interface. The breadcrumb navigation indicates the path: Amazon Redshift > Clusters > danish-atm-etl-redshift. The cluster name 'danish-atm-etl-redshift' is displayed at the top, along with buttons for 'Actions', 'Edit', 'Add partner integration', and 'Query data'. Below this, the 'General information' tab is active, showing a table with the following details:

General information			
Cluster identifier danish-atm-etl-redshift	Status Available	Node type dc2.large	Endpoint danish-atm-etl-redshift.clpwyep6bax...
Cluster namespace be4b2c8b-492d-4887-8981-72721694af80	Date created October 30, 2022, 19:23 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://danish-atm-etl-redshift....
	Storage used -		ODBC URL Driver={Amazon Redshift (x64)}; Server...

At the bottom, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'.



The screenshot shows the 'Properties' tab of the AWS Redshift console for the cluster 'danish-atm-etl-redshift'. The 'Database configurations' section is active, showing a table with the following details:

Database configurations			
Database name upgrad	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled AWS KMS key ID -	Audit logging Disabled
Port 5440	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADA...		
Admin user name awsuser			

Below this, the 'Network and security settings' section is active, showing a table with the following details:

Network and security settings			
Virtual private cloud (VPC) vpc-011092e6545ce7d4f	Availability Zone us-east-1a	VPC security group Specify which instances and devices can connect to the cluster. sg-0ca8a4966e672fbf9	Publicly accessible Allow connections from outside the VPC. Disabled
Subnet atm-etl-subnet-group	Enhanced VPC routing Disabled		
Endpoint URL -			

**Services** Search for services, features, blogs, docs, and more [Alt+S] N. Virginia voclabs/user2101764=abhijit@madpixels.in @ 4890-1955-8643

Subnet **atm-etl-subnet-group** Enhanced VPC routing Disabled sg-0ca8a4966e672fbf9

Endpoint URL -

**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.

**Associated IAM roles (1)** Info Set default Manage IAM roles

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.

Search for associated IAM role by name, status, or role type

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

**Granted accounts (0)** Edit Revoke Grant access

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

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

## Creating Schema:

CREATE SCHEMA danish\_atm\_data;

**Amazon Redshift** Query editor

Editor Query history Saved queries Scheduled queries

Status **Connected** database upgrad user awsuser Change connection

Query 1

Select database Info To view schemas, select a database. upgrad

Select schema Info To view tables, select a schema. danish\_atm\_data

Filter tables

No resources No resources to display

Run Save Schedule Clear

Send feedback

**Query results** Table details

Query

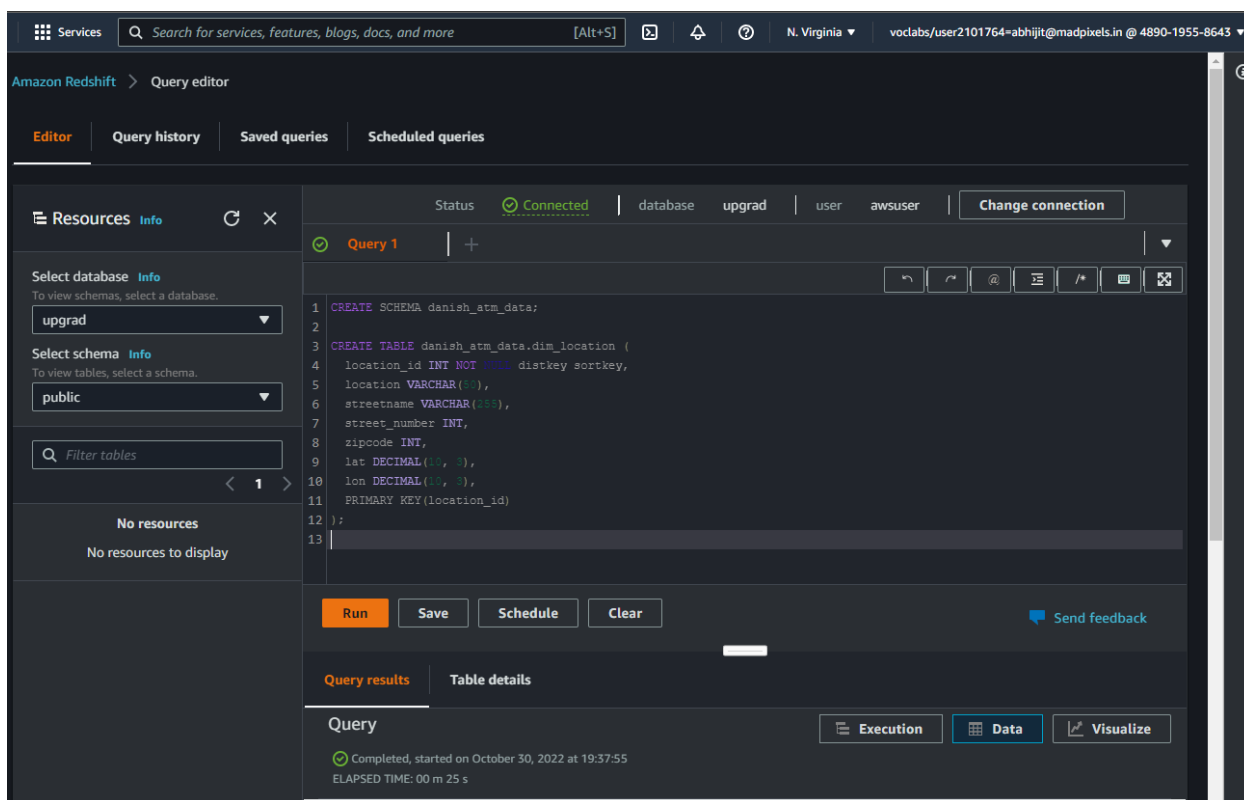
Execution Data Visualize

Completed, started on October 30, 2022 at 19:35:02  
ELAPSED TIME: 00 m 09 s

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

- **Creating location dimension table**

```
CREATE TABLE danish_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)
);
```

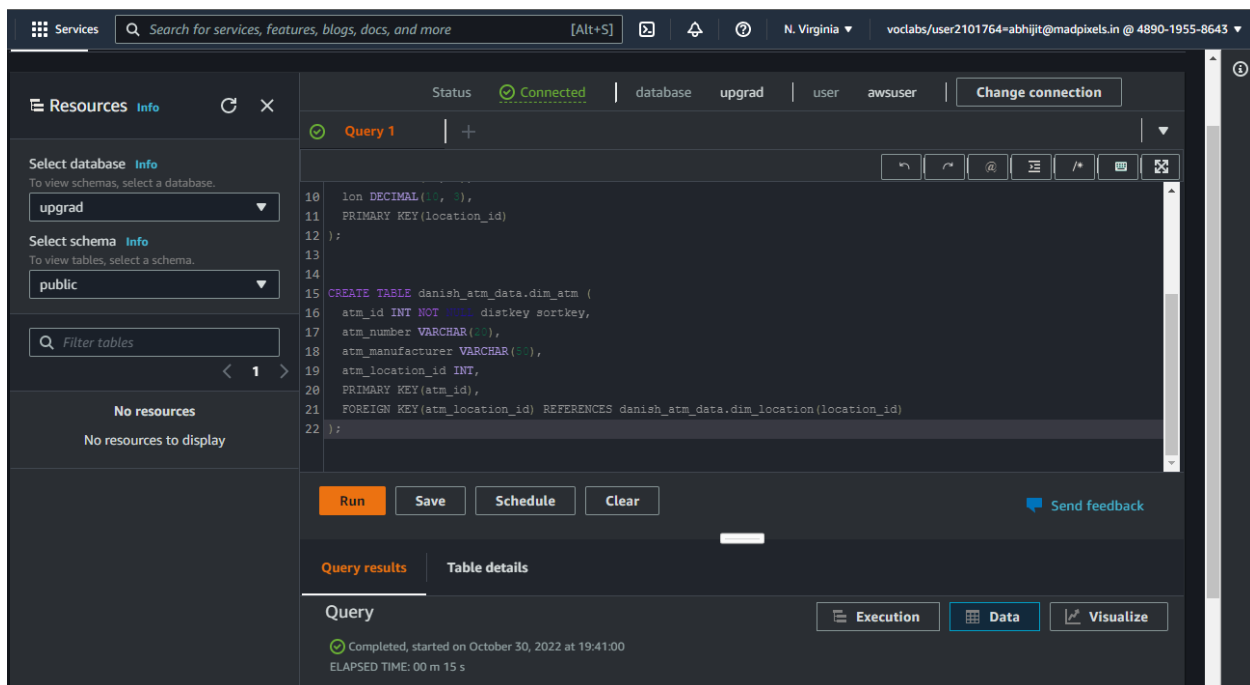


The screenshot displays the Amazon Redshift Query Editor interface. The top navigation bar includes the 'Services' menu, a search bar, and the user's profile information. The main workspace is divided into several sections:

- Left Panel:** Contains 'Resources' and 'Select database' (set to 'upgrad') and 'Select schema' (set to 'public').
- Query Editor:** Shows the SQL query for creating the 'dim\_location' table, with line numbers 1 through 13.
- Buttons:** 'Run', 'Save', 'Schedule', and 'Clear' are visible below the query editor.
- Query Results:** A section at the bottom showing the query status as 'Completed' and the elapsed time as '00 m 25 s'.

## • Creating atm dimension table

```
CREATE TABLE danish_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 danish_atm_data.dim_location(location_id)
);
```



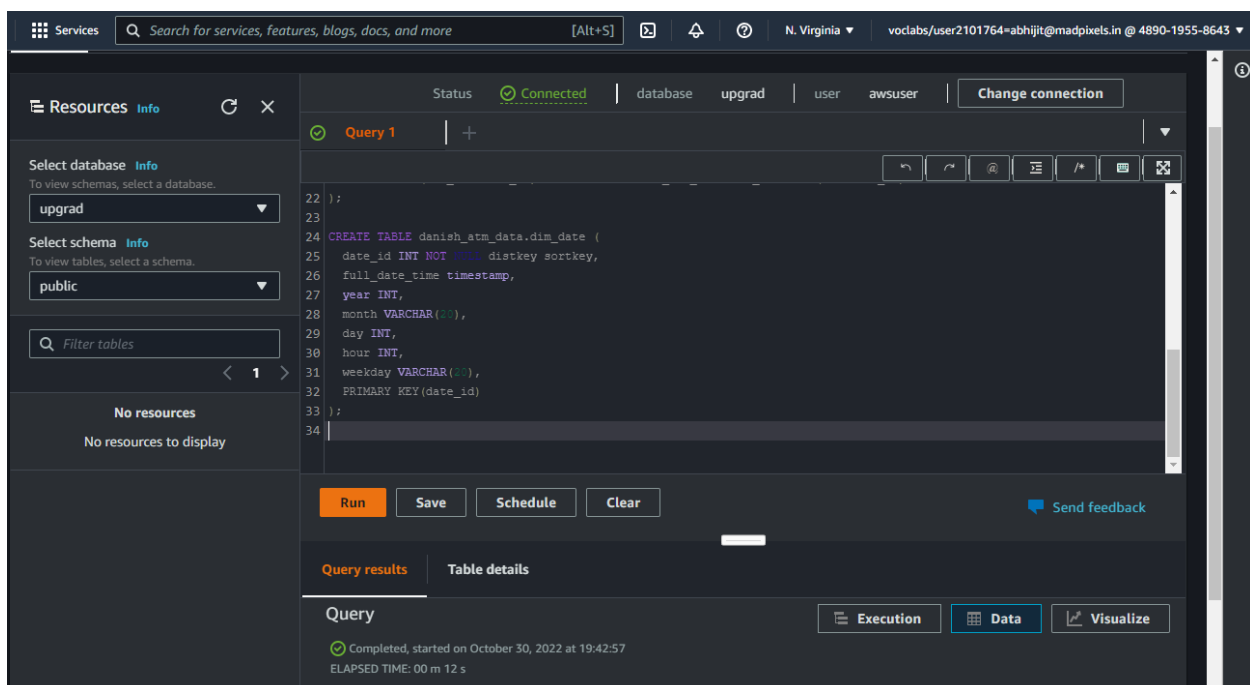
The screenshot shows the upGrad Data Studio interface. On the left, there's a sidebar with 'Resources' and 'Select database' (set to 'upgrad') and 'Select schema' (set to 'public'). The main area displays a SQL query for 'Query 1' with line numbers 10 to 22. The query is:
 

```
10 lon DECIMAL(10, 2),
11 PRIMARY KEY(location_id)
12 );
13
14
15 CREATE TABLE danish_atm_data.dim_atm (
16   atm_id INT NOT NULL distkey sortkey,
17   atm_number VARCHAR(20),
18   atm_manufacturer VARCHAR(50),
19   atm_location_id INT,
20   PRIMARY KEY(atm_id),
21   FOREIGN KEY(atm_location_id) REFERENCES danish_atm_data.dim_location(location_id)
22 );
```

 Below the query editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, the 'Query results' tab is active, showing a status message: 'Completed, started on October 30, 2022 at 19:41:00' and 'ELAPSED TIME: 00 m 15 s'. There are also tabs for 'Execution', 'Data', and 'Visualize'.

## • Creating date dimension table

```
CREATE TABLE danish_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)
);
```



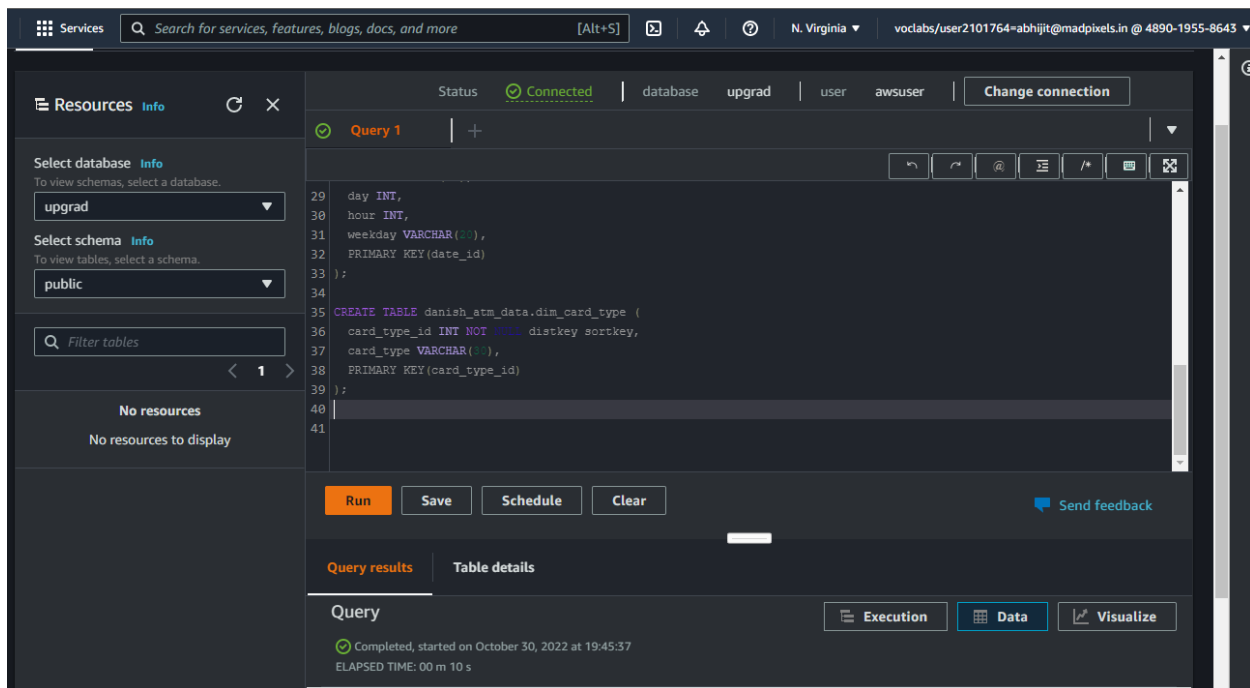
The screenshot displays the upGrad Data Studio interface. On the left sidebar, the 'Resources' panel shows the 'upgrad' database and 'public' schema. The main workspace contains a SQL query editor with the following code:

```
22 );
23
24 CREATE TABLE danish_atm_data.dim_date (
25   date_id INT NOT NULL distkey sortkey,
26   full_date_time timestamp,
27   year INT,
28   month VARCHAR(20),
29   day INT,
30   hour INT,
31   weekday VARCHAR(20),
32   PRIMARY KEY(date_id)
33 );
34
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' tab is active, showing the query execution status: 'Completed, started on October 30, 2022 at 19:42:57' and 'ELAPSED TIME: 00 m 12 s'. The 'Table details' tab is also visible.

## • Creating card type dimension table

```
CREATE TABLE danish_atm_data.dim_card_type (
  card_type_id INT NOT NULL distkey sortkey,
  card_type VARCHAR(30),
  PRIMARY KEY(card_type_id)
);
```



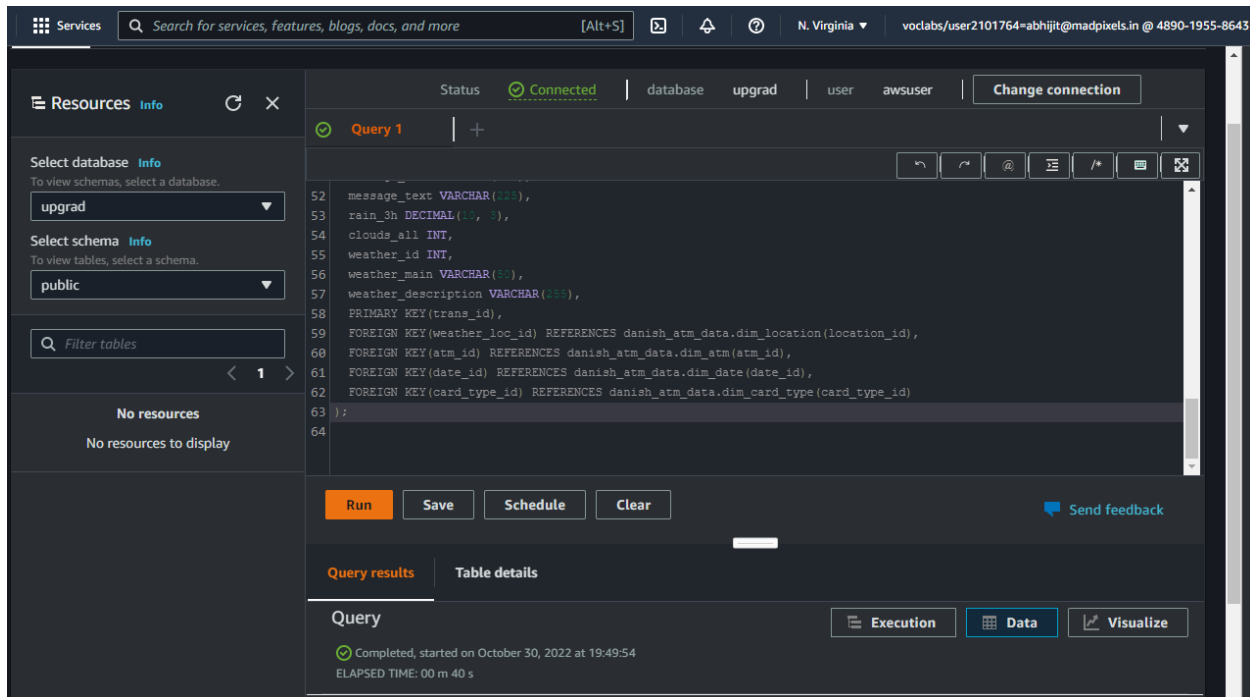
The screenshot shows the Amazon Redshift console interface. On the left, the 'Resources' panel displays the selected database 'upgrad' and schema 'public'. The main area shows a SQL query being executed. The query is as follows:

```
29 day INT,
30 hour INT,
31 weekday VARCHAR(30),
32 PRIMARY KEY(date_id)
33 );
34
35 CREATE TABLE danish_atm_data.dim_card_type (
36   card_type_id INT NOT NULL distkey sortkey,
37   card_type VARCHAR(30),
38   PRIMARY KEY(card_type_id)
39 );
40
41
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' tab is active, showing the query execution status: 'Completed, started on October 30, 2022 at 19:45:37' and 'ELAPSED TIME: 00 m 10 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## • Creating atm transactions fact table

```
CREATE TABLE danish_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 danish_atm_data.dim_location(location_id),
  FOREIGN KEY(atm_id) REFERENCES danish_atm_data.dim_atm(atm_id),
  FOREIGN KEY(date_id) REFERENCES danish_atm_data.dim_date(date_id),
  FOREIGN KEY(card_type_id) REFERENCES danish_atm_data.dim_card_type(card_type_id)
);
```



The screenshot shows the upGrad Data Studio interface. On the left, the 'Resources' panel displays the 'upgrad' database and 'public' schema. The main editor area shows the SQL query for creating the 'fact\_atm\_trans' table. The query is as follows:

```
CREATE TABLE 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 danish_atm_data.dim_location(location_id),
  FOREIGN KEY(atm_id) REFERENCES danish_atm_data.dim_atm(atm_id),
  FOREIGN KEY(date_id) REFERENCES danish_atm_data.dim_date(date_id),
  FOREIGN KEY(card_type_id) REFERENCES danish_atm_data.dim_card_type(card_type_id)
);
```

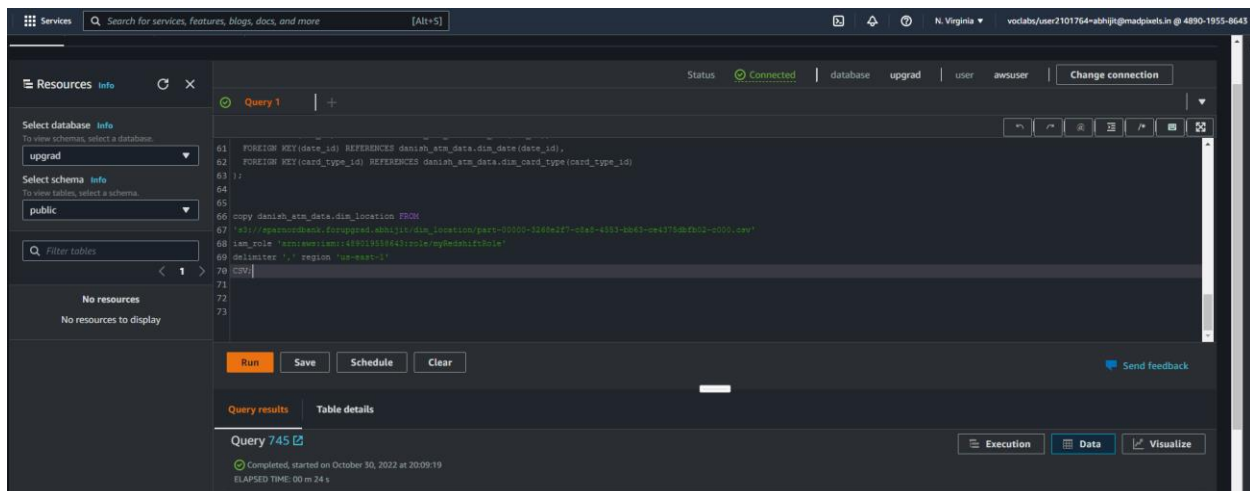
Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The bottom section shows the 'Query results' tab, indicating that the query was completed successfully on October 30, 2022, at 19:49:54, with an elapsed time of 00 m 40 s.

## 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 danish\_atm\_data.dim\_location FROM  
 's3://sparnordbank.forupgrad.abhijit/dim\_location/part-00000-3268e2f7-c8a8-4553-bb63-ce4375dbfb02-c000.csv'  
 iam\_role 'ARN:AWS:IAM::489019558643:role/myredshiftrole'  
 delimiter ',' region 'us-east-1' CSV;





## • Copying the data to dim\_atm table

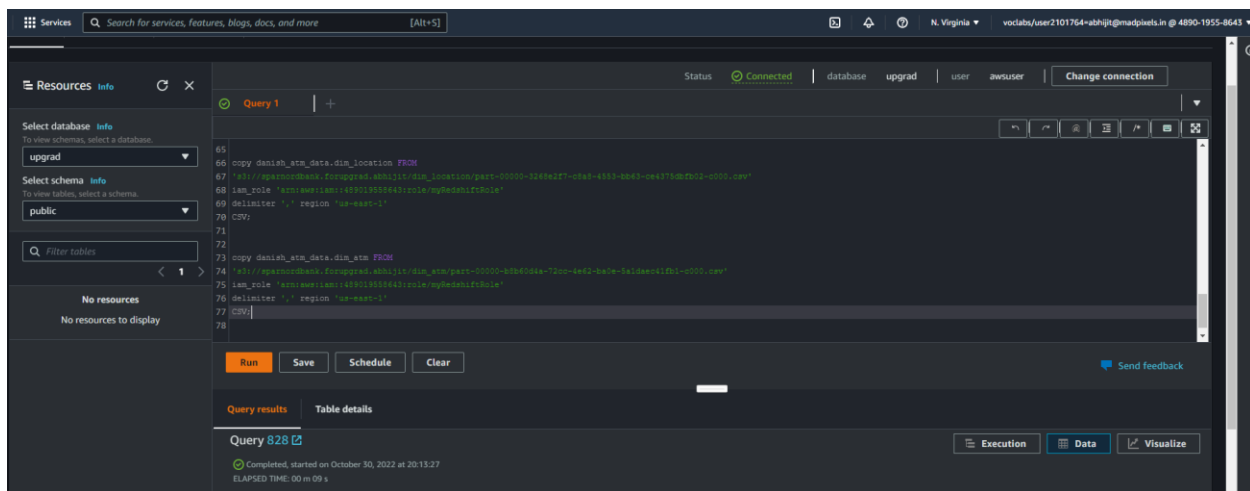
copy danish\_atm\_data.dim\_atm FROM

's3://sparnordbank.forupgrad.abhijit/dim\_atm/part-00000-b8b60d4a-72cc-4e62-ba0e-5a1daec41fb1-c000.csv'

iam\_role 'arn:aws:iam::489019558643:role/myRedshiftRole'

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

CSV;



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. The main area displays a SQL query for 'Query 1'. The query is as follows:

```

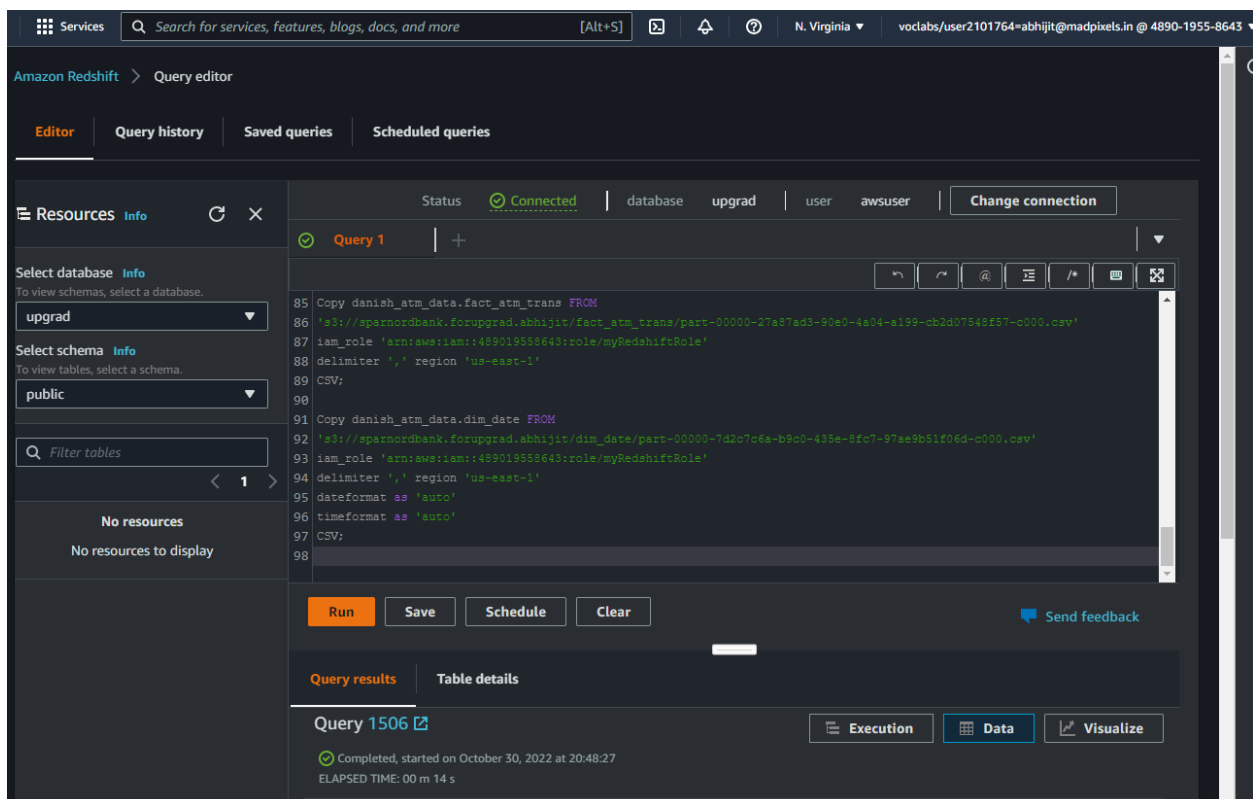
65 copy danish_atm_data.dim_location FROM
66 's3://sparnordbank.forupgrad.abhijit/dim_location/part-00000-b8b60d4a-72cc-4e62-ba0e-5a1daec41fb1-c000.csv'
67 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
68 delimiter ',' region 'us-east-1'
69 CSV;
70
71
72
73 copy danish_atm_data.dim_atm FROM
74 's3://sparnordbank.forupgrad.abhijit/dim_atm/part-00000-b8b60d4a-72cc-4e62-ba0e-5a1daec41fb1-c000.csv'
75 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
76 delimiter ',' region 'us-east-1'
77 CSV;
78

```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. At the bottom, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 828' with a status of 'Completed, started on October 30, 2022 at 20:13:27' and 'ELAPSED TIME: 00 m 09 s'.

## • Copying the data to dim\_date table

```
Copy danish_atm_data.dim_date FROM
's3://sparnordbank.forupgrad.abhijit/dim_date/part-00000-7d2c7c6a-b9c0-435e-8fc7-
97ae9b51f06d-c000.csv'
iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
dateformat as 'auto'
timeformat as 'auto'
CSV;
```



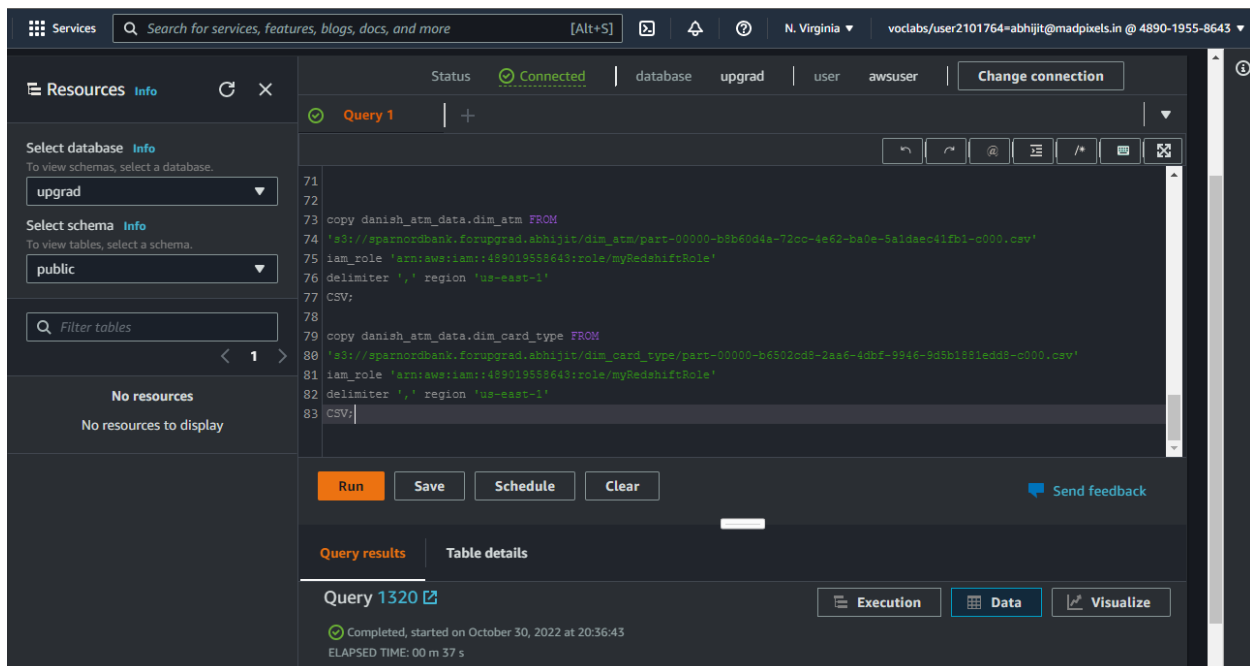
The screenshot shows the Amazon Redshift Query Editor interface. The top navigation bar includes 'Services', a search bar, and user information. The main interface has tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. The 'Editor' tab is active, showing a SQL query in a text area. The query is as follows:

```
85 Copy danish_atm_data.fact_atm_trans FROM
86 's3://sparnordbank.forupgrad.abhijit/fact_atm_trans/part-00000-27a87ad3-90e0-4a04-a199-cb2d07548f57-c000.csv'
87 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
88 delimiter ',' region 'us-east-1'
89 CSV;
90
91 Copy danish_atm_data.dim_date FROM
92 's3://sparnordbank.forupgrad.abhijit/dim_date/part-00000-7d2c7c6a-b9c0-435e-8fc7-97ae9b51f06d-c000.csv'
93 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
94 delimiter ',' region 'us-east-1'
95 dateformat as 'auto'
96 timeformat as 'auto'
97 CSV;
98
```

Below the query editor, 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, the 'Query results' tab is selected, showing the query ID 'Query 1506'. The status indicates 'Completed, started on October 30, 2022 at 20:48:27' with an 'ELAPSED TIME: 00 m 14 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to dim\_card\_type table

```
copy danish_atm_data.dim_card_type FROM
'S3://sparnordbank.forupgrad.abhijit/dim_card_type/part-00000-b6502cd8-2aa6-4dbf-9946-
9d5b1881edd8-c000.csv'
iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole' delimiter ',' region 'us-east-1'
CSV;
```



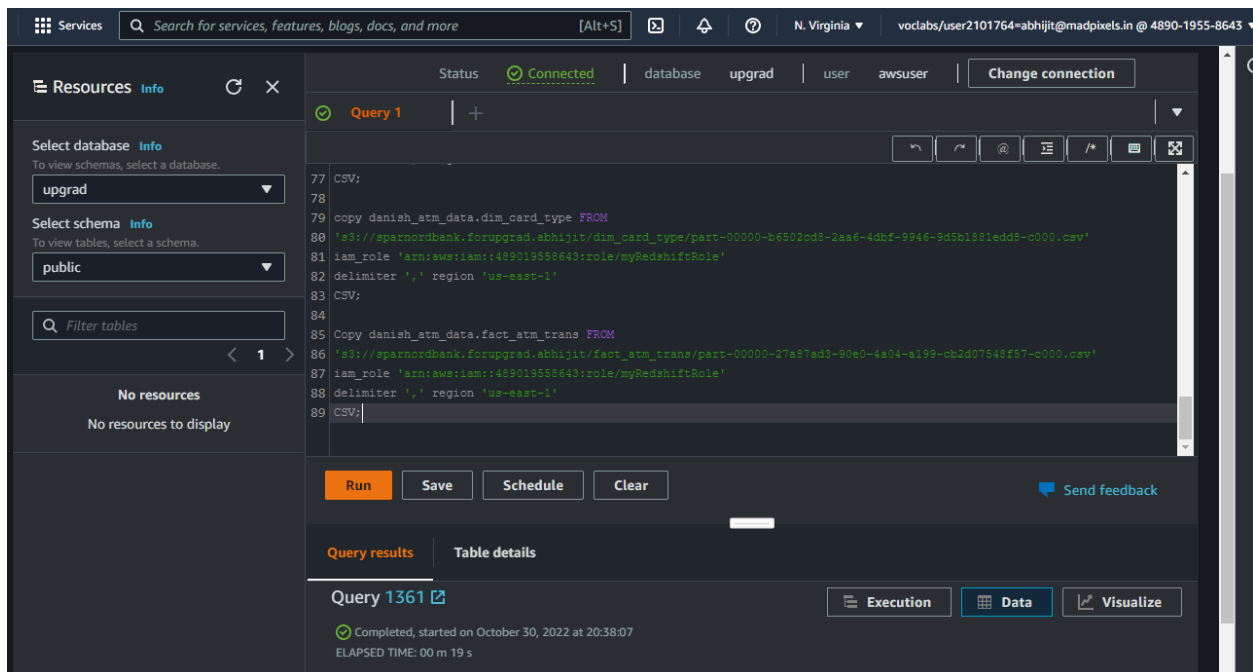
The screenshot displays the AWS Redshift console interface. On the left, the 'Resources' panel shows the selected database 'upgrad' and schema 'public'. The main area shows a SQL query being executed, labeled 'Query 1'. The query text is as follows:

```
71
72
73 copy danish_atm_data.dim_atm FROM
74 's3://sparnordbank.forupgrad.abhijit/dim_atm/part-00000-b6502cd8-2aa6-4dbf-9946-9d5b1881edd8-c000.csv'
75 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
76 delimiter ',' region 'us-east-1'
77 CSV;
78
79 copy danish_atm_data.dim_card_type FROM
80 's3://sparnordbank.forupgrad.abhijit/dim_card_type/part-00000-b6502cd8-2aa6-4dbf-9946-9d5b1881edd8-c000.csv'
81 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
82 delimiter ',' region 'us-east-1'
83 CSV;
```

Below the query text, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is selected, showing 'Query 1320'. The status indicates 'Completed, started on October 30, 2022 at 20:36:43' with an 'ELAPSED TIME: 00 m 37 s'.

- Copying the data to fact\_atm\_trans table

Copy danish\_atm\_data.fact\_atm\_trans FROM  
 'S3://sparnordbank.forupgrad.abhijit/fact\_atm\_trans/part-00000-27a87ad3-90e0-4a04-a199-  
 cb2d07548f57-c000.csv'  
 iam\_role 'arn:aws:iam::489019558643:role/myRedshiftRole' delimiter ',' region 'us-east-1'  
 CSV;



The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. The main area displays a SQL query being executed. The query is as follows:

```

77 CSV;
78
79 copy danish_atm_data.dim_card_type FROM
80 's3://sparnordbank.forupgrad.abhijit/dim_card_type/part-00000-b6902cd8-2aa6-4dbf-9946-9d5b1881edd8-c000.csv'
81 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
82 delimiter ',' region 'us-east-1'
83 CSV;
84
85 Copy danish_atm_data.fact_atm_trans FROM
86 's3://sparnordbank.forupgrad.abhijit/fact_atm_trans/part-00000-27a87ad3-90e0-4a04-a199-cb2d07548f57-c000.csv'
87 iam_role 'arn:aws:iam::489019558643:role/myRedshiftRole'
88 delimiter ',' region 'us-east-1'
89 CSV;
  
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link. Below the buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 1361' with a status of 'Completed, started on October 30, 2022 at 20:38:07' and an 'ELAPSED TIME: 00 m 19 s'.