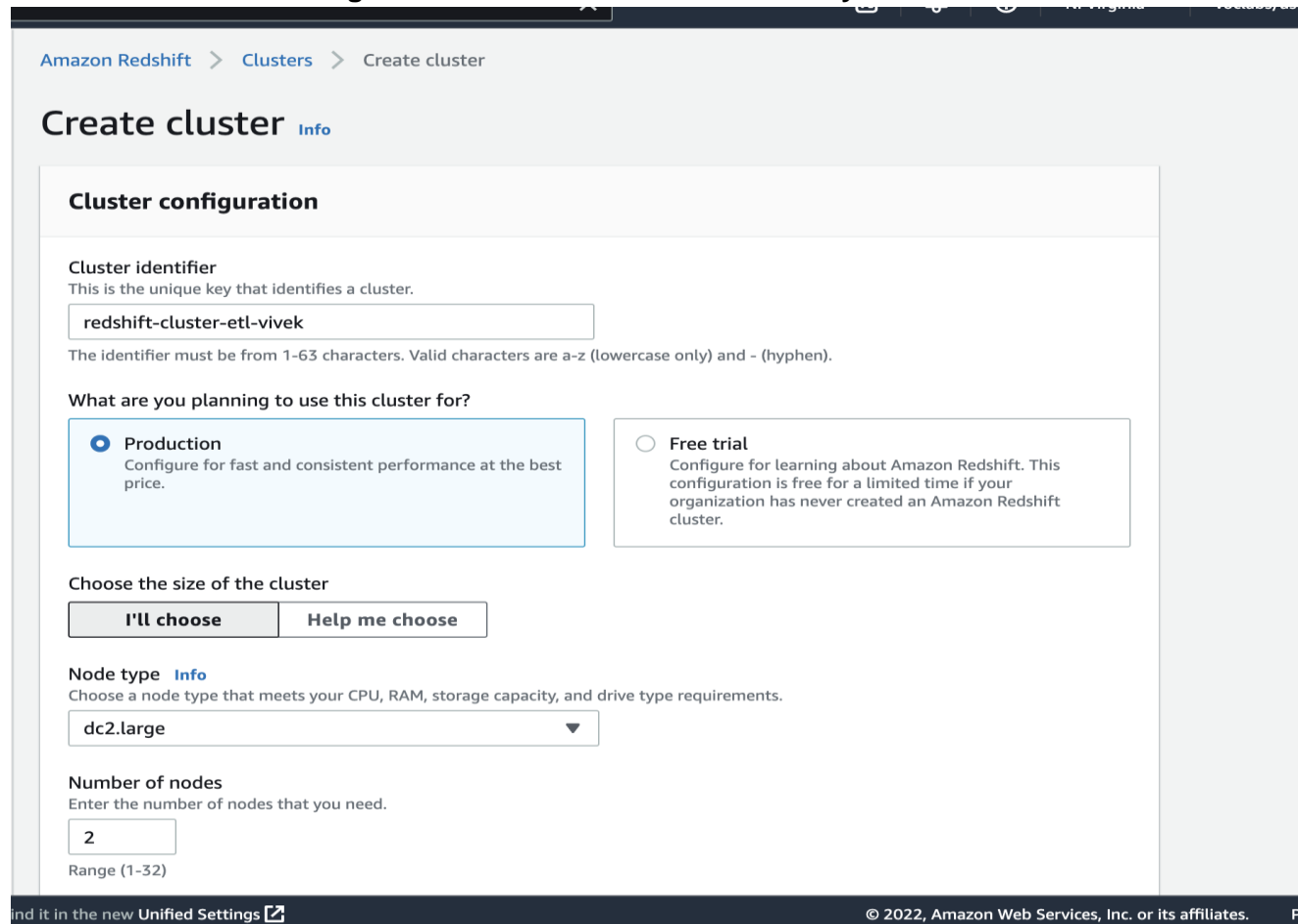


## Creation of a Redshift Cluster

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



Amazon Redshift > Clusters > Create cluster

### Create cluster [Info](#)

#### Cluster configuration

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

redshift-cluster-etl-vivek

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**  
Enter the number of nodes that you need.




2

Range (1-32)

Find it in the new [Unified Settings](#)

© 2022, Amazon Web Services, Inc. or its affiliates. Pr

×



N. Virginia ▼

voclabs/user2238604=vivek @ 3714-3662-9611 ▼

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

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 "@".

☐ Show password

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.

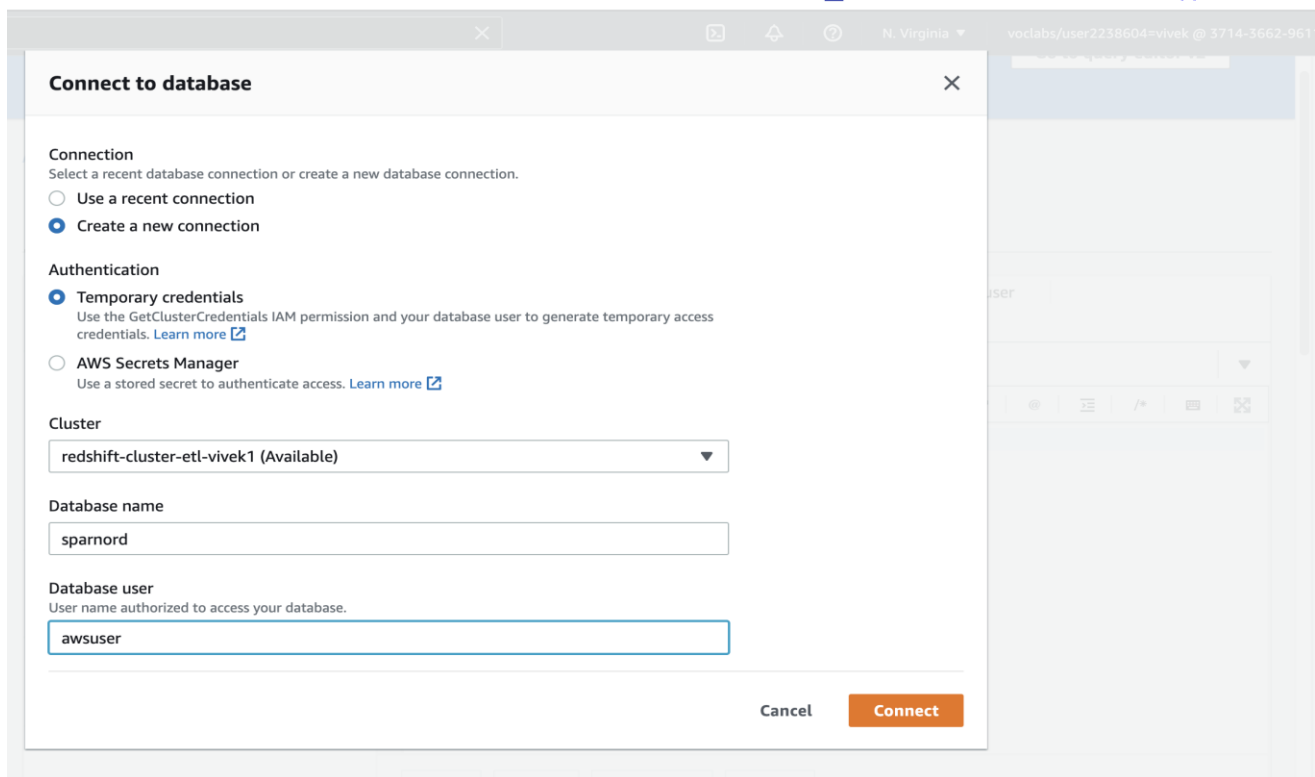
Find it in the new [Unified Settings](#)

© 2022, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

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

Find it in the new **Unified Settings** [↗](#)

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



**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-vivek1 (Available)

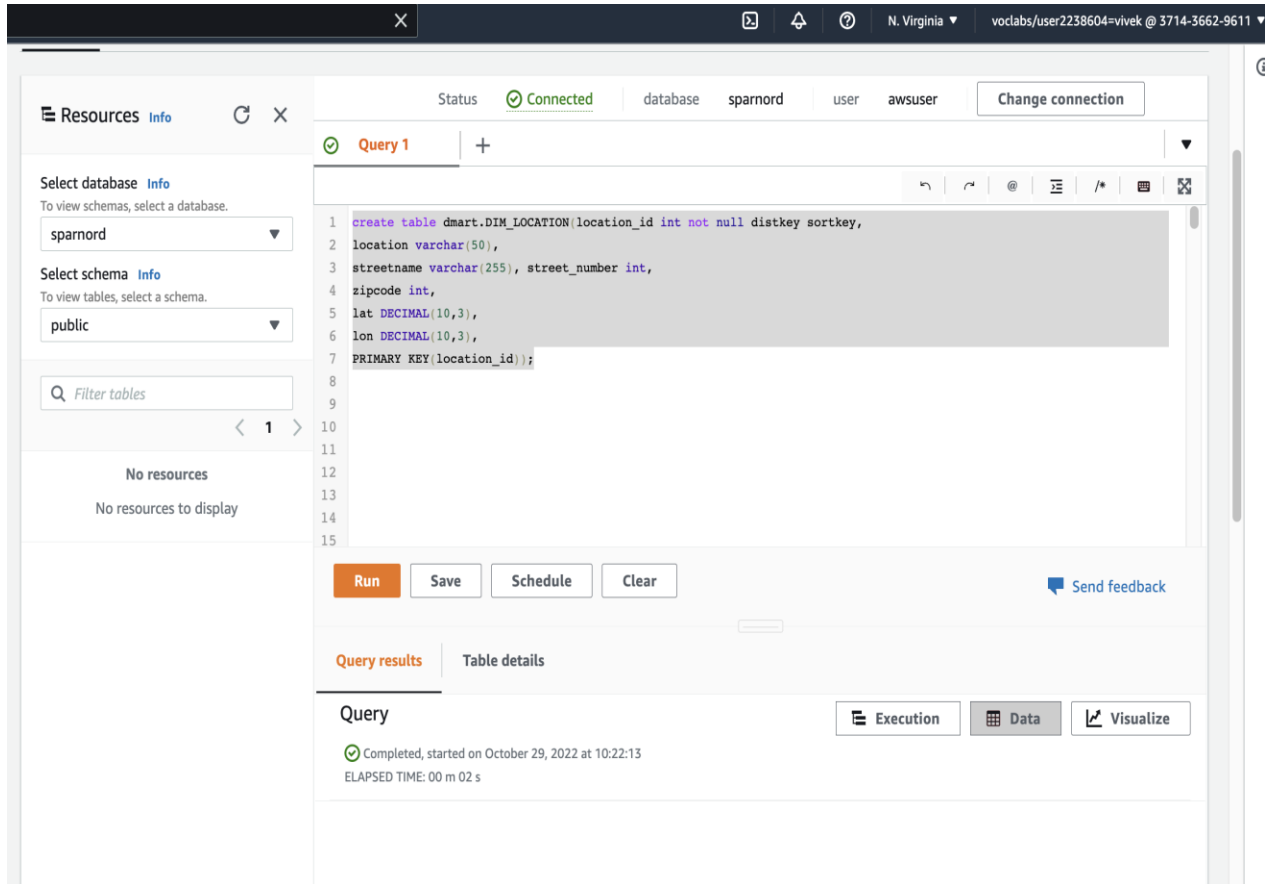
**Database name**  
sparnord

**Database user**  
User name authorized to access your database.  
awsuser

Cancel Connect

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

1. create table dmart.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));

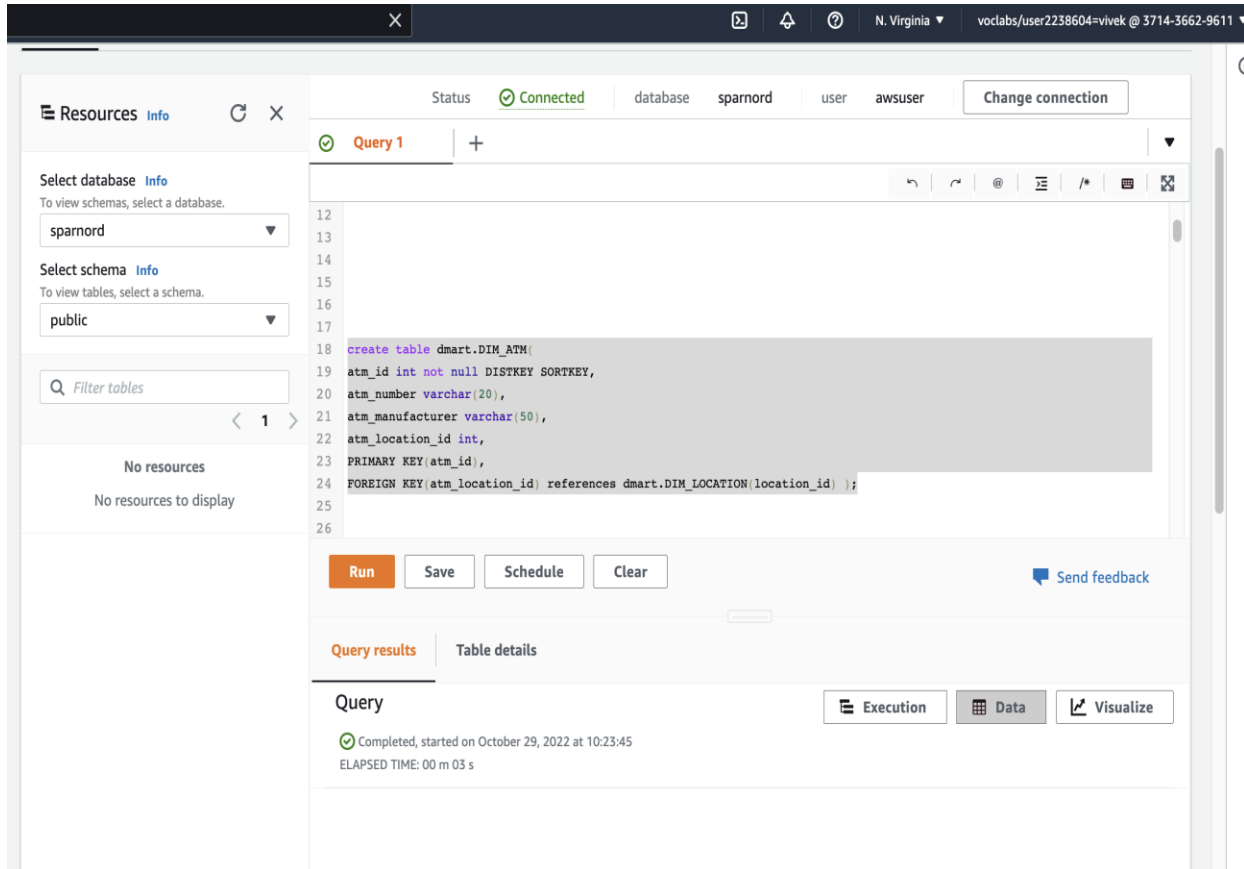


The screenshot shows the upGrad SQL IDE interface. At the top, the status is 'Connected' to the 'sparnord' database. The left sidebar shows the 'Resources' panel with 'Select database' set to 'sparnord' and 'Select schema' set to 'public'. The main editor area displays a SQL query for creating a table named 'dmart.DIM\_LOCATION'. The query is as follows:

```
1 create table dmart.DIM_LOCATION(location_id int not null distkey sortkey,
2 location varchar(50),
3 streetname varchar(255), street_number int,
4 zipcode int,
5 lat DECIMAL(10,3),
6 lon DECIMAL(10,3),
7 PRIMARY KEY(location_id));
8
9
10
11
12
13
14
15
```

Below the query editor, 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 query editor, the 'Query results' tab is active, showing the query execution status: 'Completed, started on October 29, 2022 at 10:22:13' and 'ELAPSED TIME: 00 m 02 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

2. create table dmart.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 dmart.DIM\_LOCATION(location\_id) );



The screenshot shows the AWS Glue console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there are dropdowns for 'Select database' (set to 'sparnord') and 'Select schema' (set to 'public'). A search bar for 'Filter tables' is also present. The main area shows a SQL query in the 'Query 1' tab. The query is as follows:

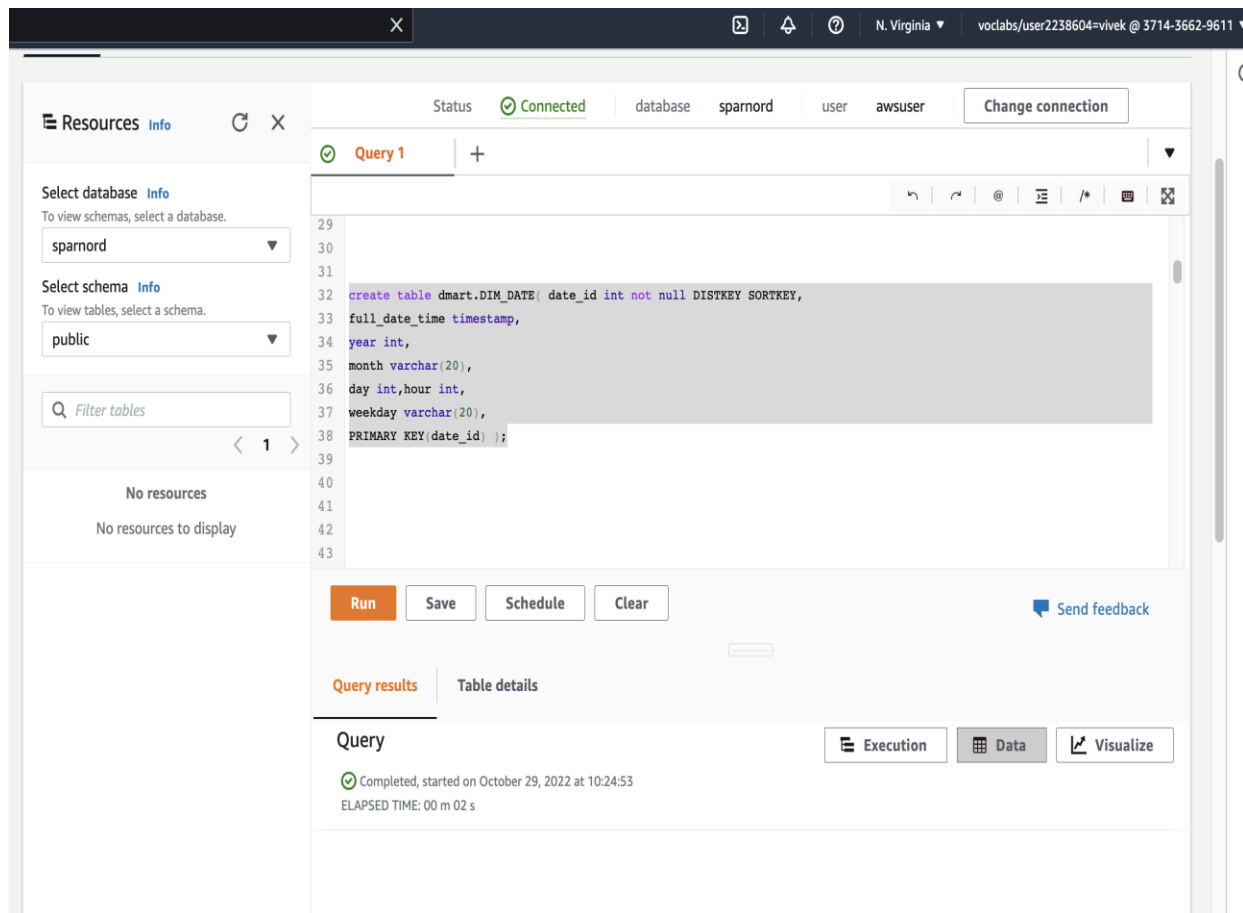
```

12
13
14
15
16
17
18 create table dmart.DIM_ATM(
19   atm_id int not null DISTKEY SORTKEY,
20   atm_number varchar(20),
21   atm_manufacturer varchar(50),
22   atm_location_id int,
23   PRIMARY KEY(atm_id),
24   FOREIGN KEY(atm_location_id) references dmart.DIM_LOCATION(location_id) );
25
26

```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also visible. The 'Query results' tab is selected, showing the query execution status: 'Completed, started on October 29, 2022 at 10:23:45' and 'ELAPSED TIME: 00 m 03 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

**3. create table dmart.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) );**



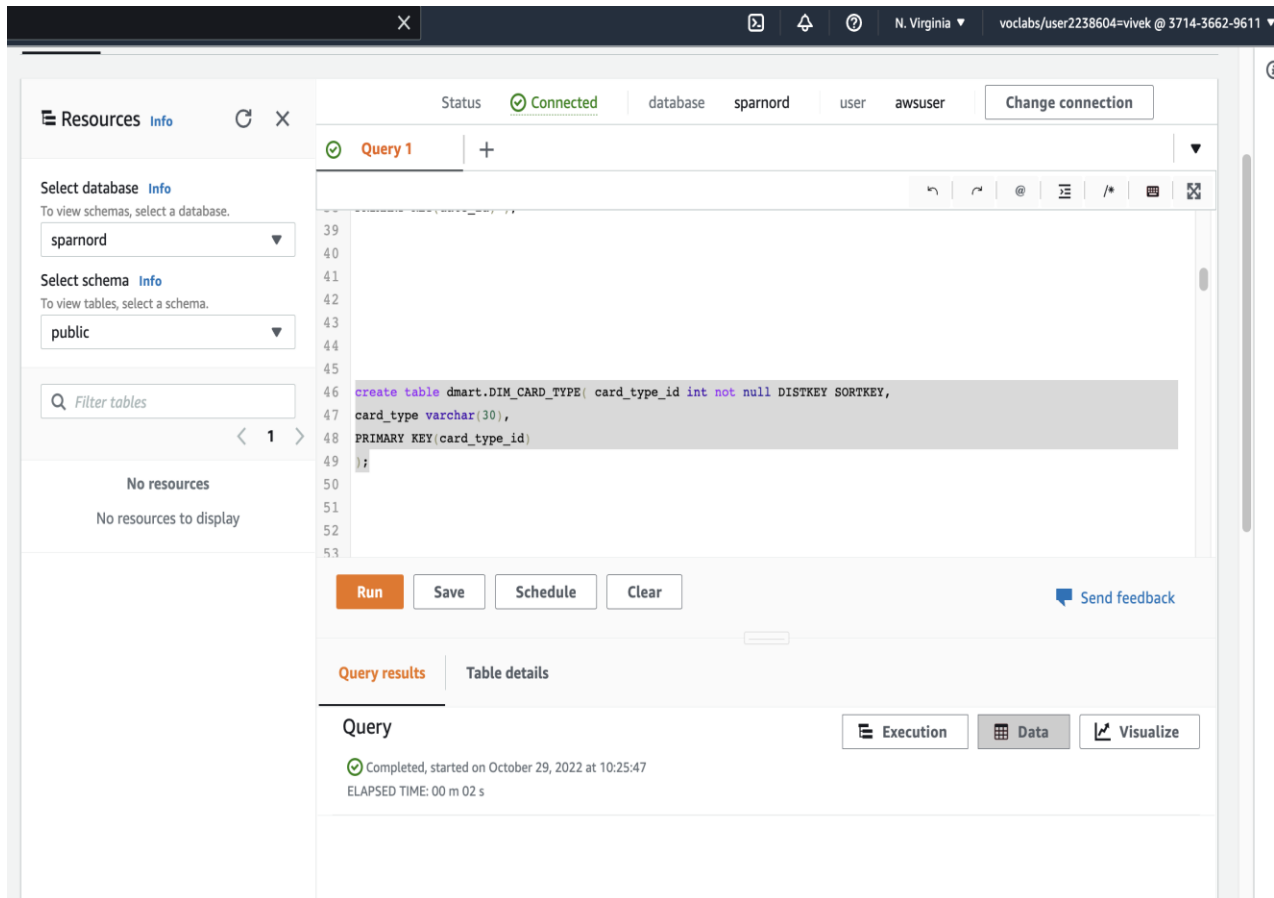
The screenshot shows the upGrad SQL IDE interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there's a 'Select database' dropdown set to 'sparnord' and a 'Select schema' dropdown set to 'public'. A search bar for 'Filter tables' is also present. The main area displays a SQL query in a text editor, with line numbers 29 to 43. The query is:
 

```

    29
    30
    31
    32 create table dmart.DIM_DATE( date_id int not null DISTKEY SORTKEY,
    33 full_date_time timestamp,
    34 year int,
    35 month varchar(20),
    36 day int, hour int,
    37 weekday varchar(20),
    38 PRIMARY KEY(date_id) );
    39
    40
    41
    42
    43
    
```

 Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also visible. The bottom section shows 'Query results' and 'Table details' tabs. The 'Query results' tab is active, showing a status message: 'Completed, started on October 29, 2022 at 10:24:53' and 'ELAPSED TIME: 00 m 02 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

4. create table dmart.DIM\_CARD\_TYPE( card\_type\_id int not null DISTKEY SORTKEY,  
card\_type varchar(30),  
PRIMARY KEY(card\_type\_id)  
);



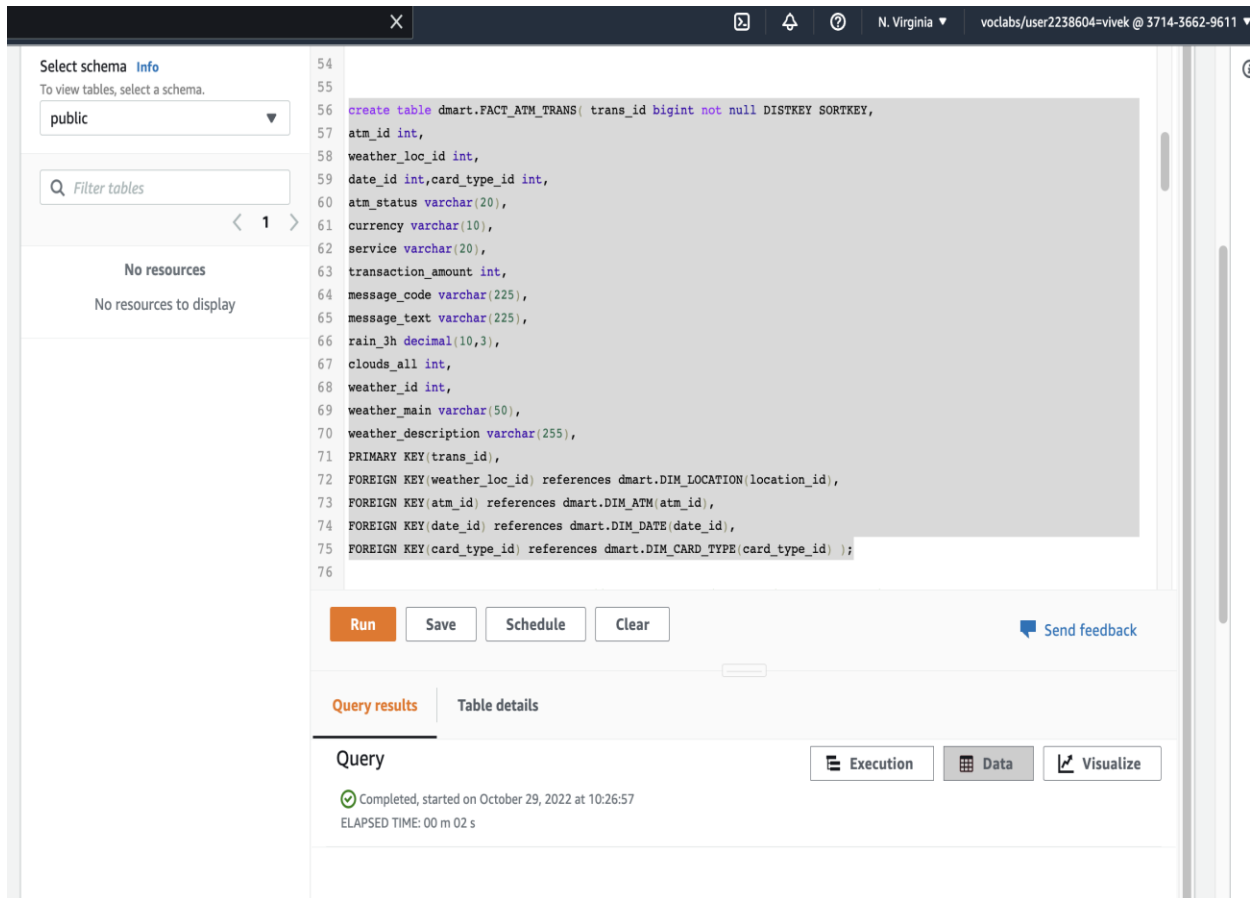
The screenshot shows the upGrad database interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there's a 'Select database' dropdown set to 'sparnord' and a 'Select schema' dropdown set to 'public'. A search bar for tables is also present. The main area shows a SQL query editor with a query named 'Query 1'. The query is:

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

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's a section for 'Query results' and 'Table details'. The 'Query results' section shows the query execution status: 'Completed, started on October 29, 2022 at 10:25:47' and 'ELAPSED TIME: 00 m 02 s'.

5. create table dmart.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 dmart.DIM\_LOCATION(location\_id),  
FOREIGN KEY(atm\_id) references dmart.DIM\_ATM(atm\_id),  
FOREIGN KEY(date\_id) references dmart.DIM\_DATE(date\_id),  
FOREIGN KEY(card\_type\_id) references dmart.DIM\_CARD\_TYPE(card\_type\_id) );





The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Select schema' set to 'public' and a search bar. The main area displays a SQL query to create a table named 'dmart.FACT\_ATM\_TRANS'. The query includes various columns with data types and constraints, and foreign key references to other tables in the 'dmart' schema. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there's a section for 'Query results' and 'Table details', with a status indicating the query is 'Completed'.

```

54
55
56 create table dmart.FACT_ATM_TRANS ( trans_id bigint not null DISTKEY SORTKEY,
57 atm_id int,
58 weather_loc_id int,
59 date_id int, card_type_id int,
60 atm_status varchar(20),
61 currency varchar(10),
62 service varchar(20),
63 transaction_amount int,
64 message_code varchar(225),
65 message_text varchar(225),
66 rain_3h decimal(10,3),
67 clouds_all int,
68 weather_id int,
69 weather_main varchar(50),
70 weather_description varchar(255),
71 PRIMARY KEY(trans_id),
72 FOREIGN KEY(weather_loc_id) references dmart.DIM_LOCATION(location_id),
73 FOREIGN KEY(atm_id) references dmart.DIM_ATM(atm_id),
74 FOREIGN KEY(date_id) references dmart.DIM_DATE(date_id),
75 FOREIGN KEY(card_type_id) references dmart.DIM_CARD_TYPE(card_type_id) );
76

```

Run Save Schedule Clear

Send feedback

Query results Table details

Query

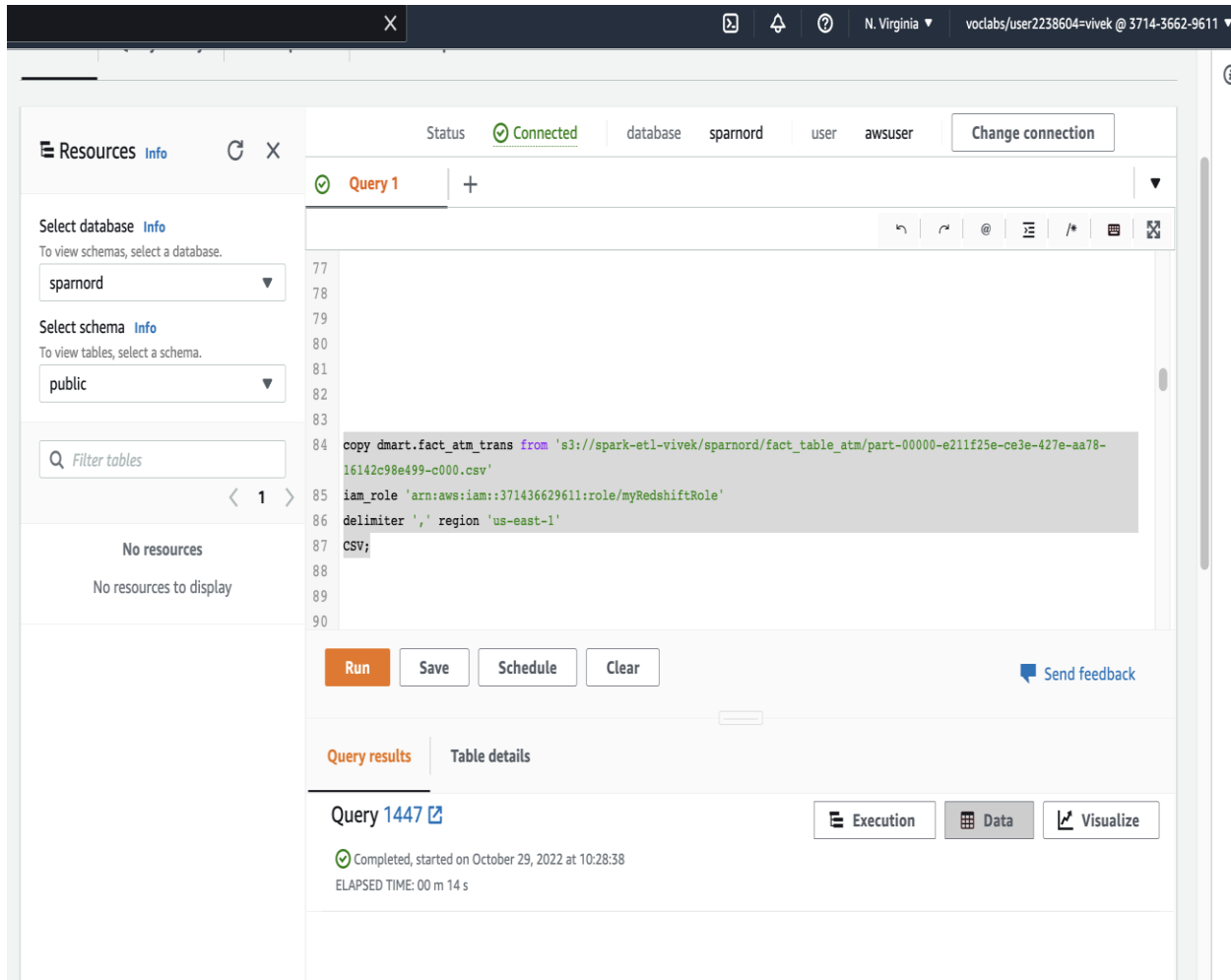
Completed, started on October 29, 2022 at 10:26:57  
ELAPSED TIME: 00 m 02 s

Execution Data 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:

1. copy dmart.fact\_atm\_trans from 's3://spark-etl-vivek/sparnord/fact\_table\_atm/part-00000-e211f25e-ce3e-427e-aa78-16142c98e499-c000.csv'  
iam\_role 'arn:aws:iam::371436629611:role/myRedshiftRole'  
delimiter ',' region 'us-east-1'  
CSV;

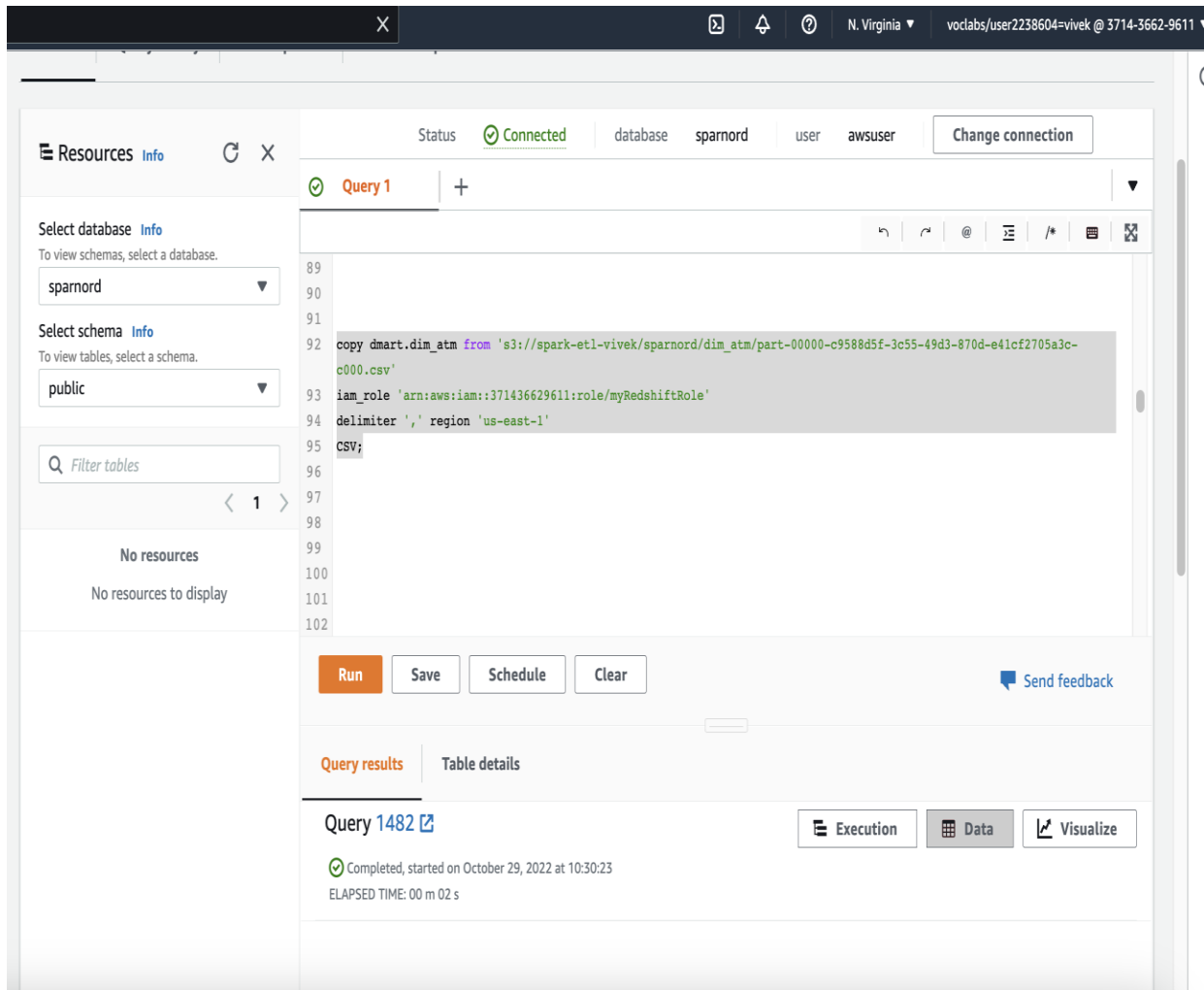


The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. The main area displays a SQL query being executed. The query is:

```
copy dmart.fact_atm_trans from 's3://spark-etl-vivek/sparnord/fact_table_atm/part-00000-e211f25e-ce3e-427e-aa78-16142c98e499-c000.csv'
iam_role 'arn:aws:iam::371436629611:role/myRedshiftRole'
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 selected, showing the query ID 'Query 1447' and its execution status: 'Completed, started on October 29, 2022 at 10:28:38' with an 'ELAPSED TIME: 00 m 14 s'.

2. copy dmart.dim\_atm from 's3://spark-etl-vivek/sparnord/dim\_atm/part-00000-c9588d5f-3c55-49d3-870d-e41cf2705a3c-c000.csv'  
iam\_role 'arn:aws:iam::371436629611:role/myRedshiftRole'  
delimiter ',' region 'us-east-1'  
CSV;

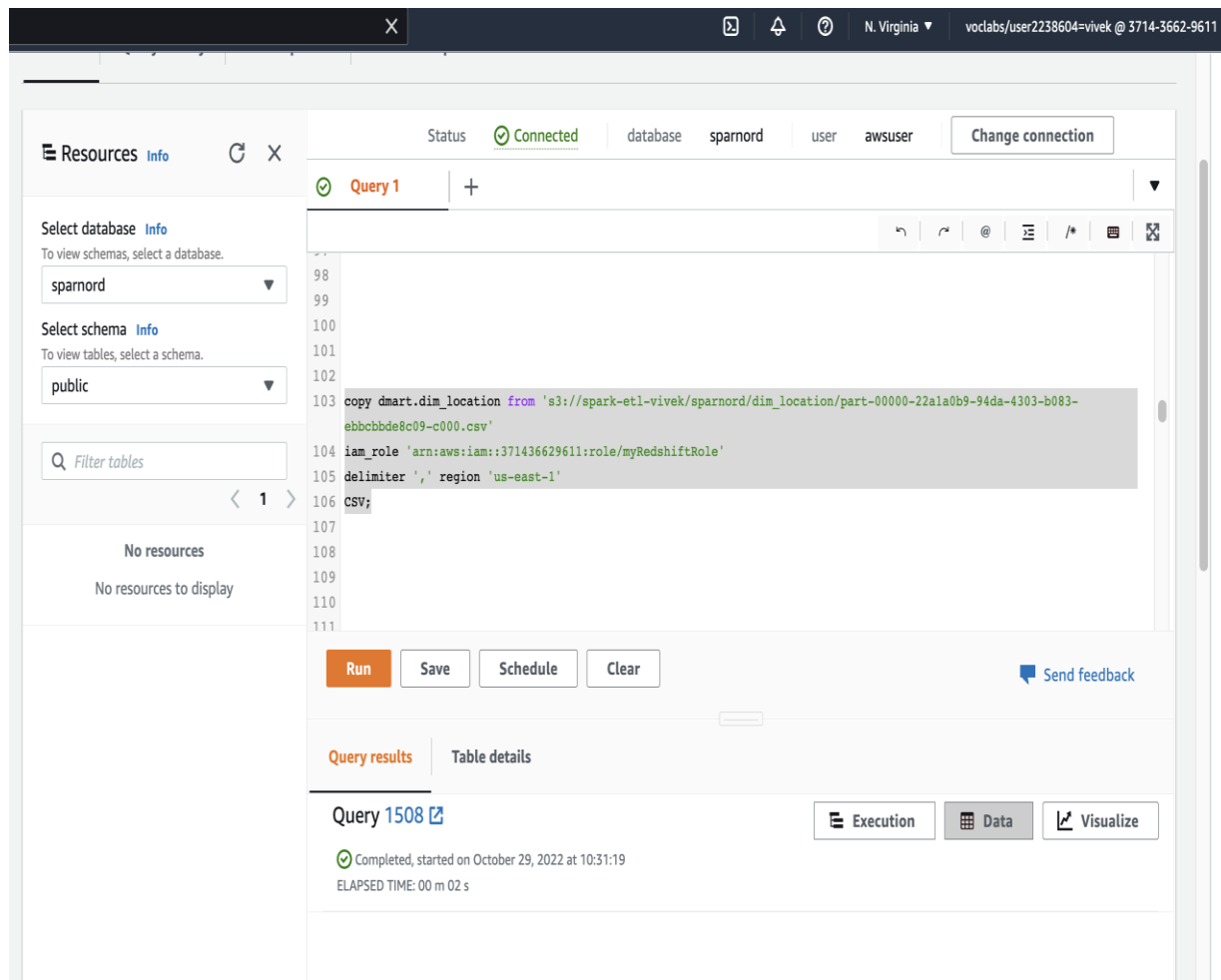


The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Select database' (sparnord) and 'Select schema' (public). The main area shows a SQL query being executed. The query is:

```
copy dmart.dim_atm from 's3://spark-etl-vivek/sparnord/dim_atm/part-00000-c9588d5f-3c55-49d3-870d-e41cf2705a3c-c000.csv'
iam_role 'arn:aws:iam::371436629611:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is selected, showing 'Query 1482' with a status of 'Completed, started on October 29, 2022 at 10:30:23' and 'ELAPSED TIME: 00 m 02 s'.

3. copy dmart.dim\_location from 's3://spark-etl-vivek/sparnord/dim\_location/part-00000-22a1a0b9-94da-4303-b083-ebbcbbde8c09-c000.csv'  
iam\_role 'arn:aws:iam::371436629611:role/myRedshiftRole'  
delimiter ',' region 'us-east-1'  
CSV;

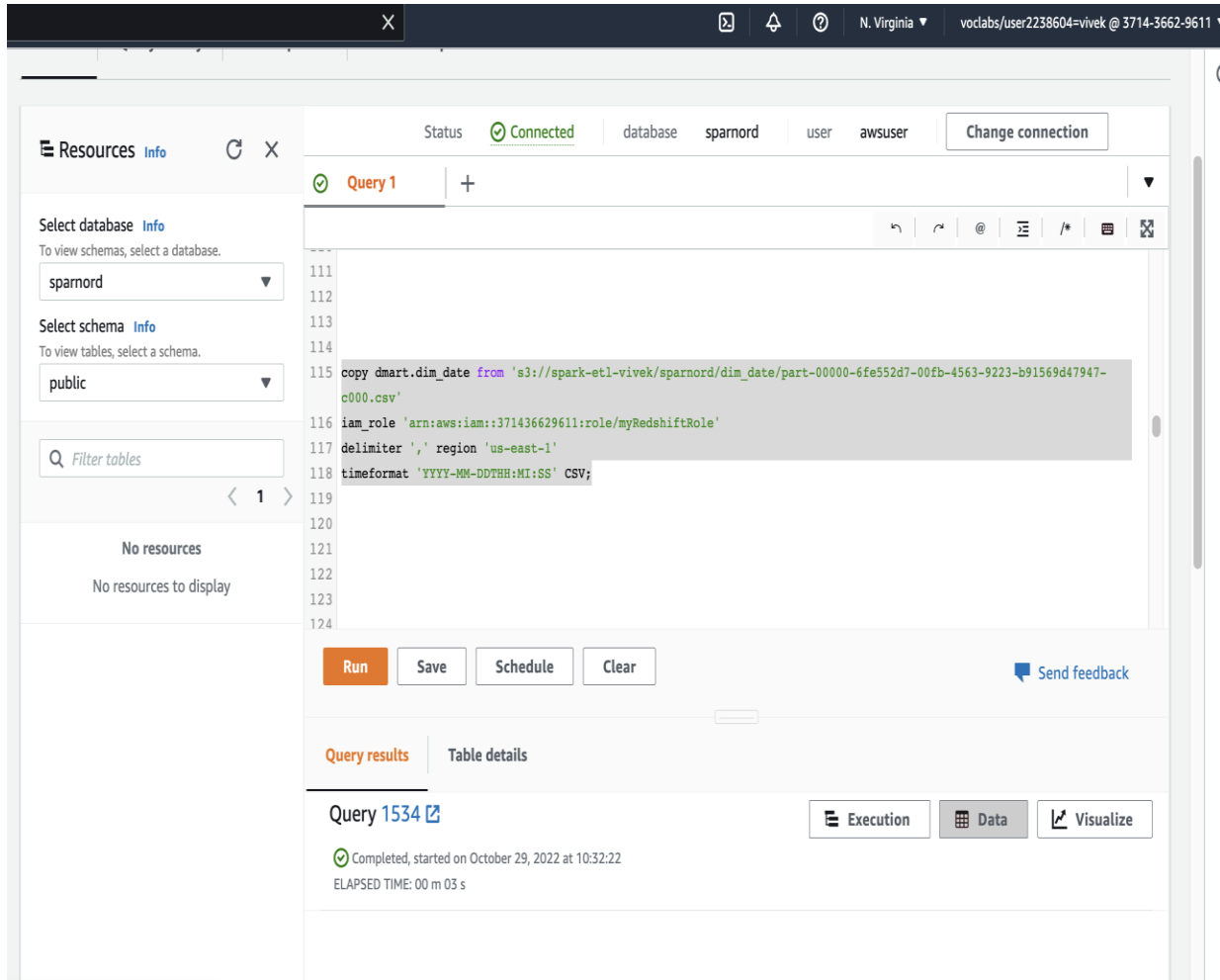


The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Select database' (sparnord) and 'Select schema' (public). The main area displays a SQL query being executed. The query is:

```
copy dmart.dim_location from 's3://spark-etl-vivek/sparnord/dim_location/part-00000-22a1a0b9-94da-4303-b083-ebbcbbde8c09-c000.csv'
iam_role 'arn:aws:iam::371436629611:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The status bar at the bottom indicates 'Query 1508' is 'Completed, started on October 29, 2022 at 10:31:19' with an 'ELAPSED TIME: 00 m 02 s'.

**4. copy dmart.dim\_date from 's3://spark-etl-vivek/sparnord/dim\_date/part-00000-6fe552d7-00fb-4563-9223-b91569d47947-c000.csv'**  
**iam\_role 'arn:aws:iam::371436629611: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 with 'Resources' and 'Select database' (sparnord) and 'Select schema' (public). The main area displays a SQL query being executed. The query is:

```
copy dmart.dim_date from 's3://spark-etl-vivek/sparnord/dim_date/part-00000-6fe552d7-00fb-4563-9223-b91569d47947-c000.csv'
iam_role 'arn:aws:iam::371436629611:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS' CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The status bar at the bottom indicates 'Query 1534' is 'Completed, started on October 29, 2022 at 10:32:22' with an 'ELAPSED TIME: 00 m 03 s'.

5. copy dmart.dim\_card\_type from 's3://spark-etl-vivek/sparnord/dim\_card/part-00000-69e844a0-b253-482c-92a9-8934bd70008c-c000.csv'  
iam\_role 'arn:aws:iam::371436629611:role/myRedshiftRole'  
delimiter ',' region 'us-east-1'  
CSV;

Resources

Info

Select database

Info

To view schemas, select a database.

sparnord

Select schema

Info

To view tables, select a schema.

public

Filter tables

No resources

No resources to display

Status

Connected

database

sparnord

user

awsuser

Change connection

Query 1

+

copy dmart.dim\_card\_type from 's3://spark-etl-vivek/sparnord/dim\_card/part-00000-69e844a0-b253-482c-92a9-8934bd70008c-c000.csv'

iam\_role 'arn:aws:iam::371436629611:role/myRedshiftRole'

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

CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 1560

Execution

Data

Visualize

Completed, started on October 29, 2022 at 10:33:24

ELAPSED TIME: 00 m 03 s