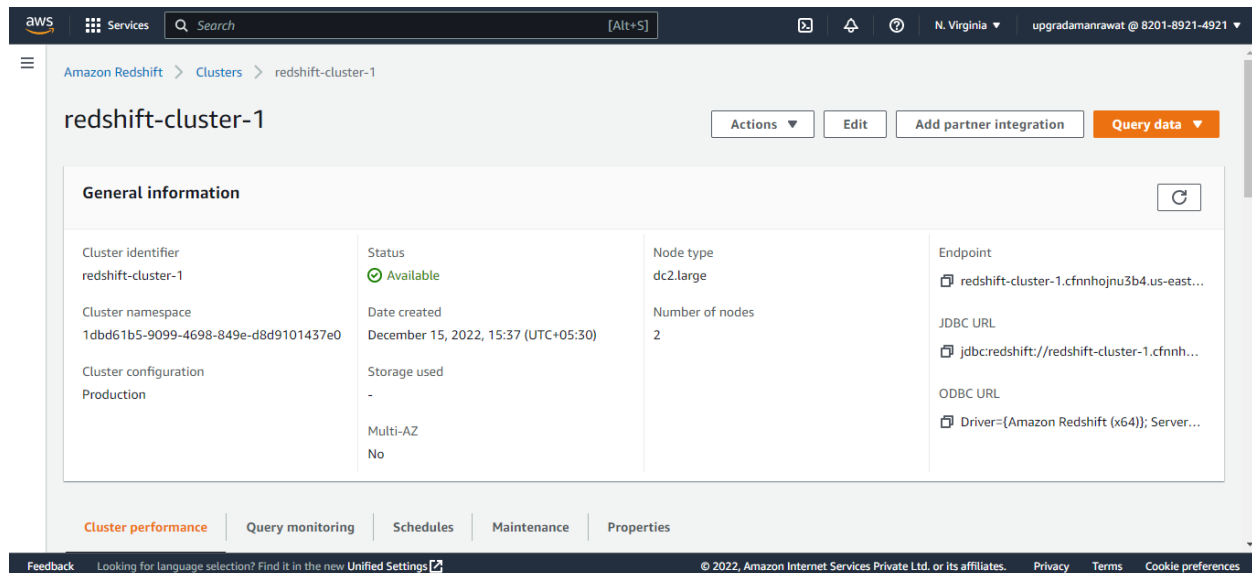


Creation of a Redshift Cluster

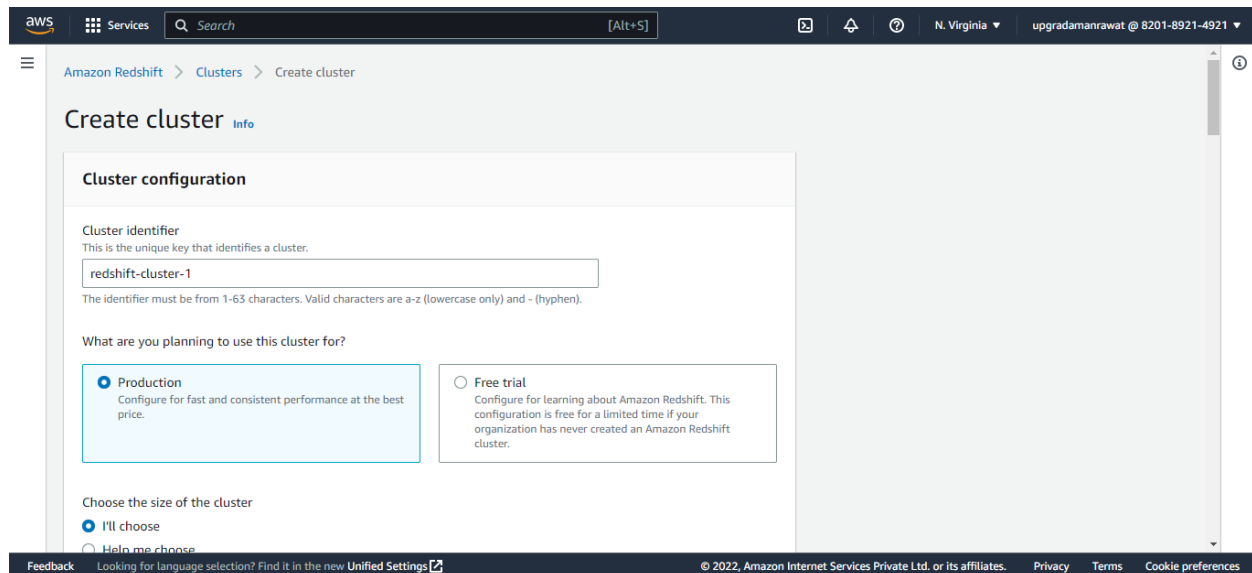
Screenshots of the configuration of the Redshift cluster that you have created:



The screenshot shows the AWS Redshift console for a cluster named 'redshift-cluster-1'. The cluster is in an 'Available' status. The configuration details are as follows:

General information			
Cluster identifier redshift-cluster-1	Status Available	Node type dc2.large	Endpoint redshift-cluster-1.cfnhohnu3b4.us-east...
Cluster namespace 1dbd61b5-9099-4698-849e-d8d9101437e0	Date created December 15, 2022, 15:37 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-1.cfnh...
Cluster configuration Production	Storage used -		ODBC URL Driver={Amazon Redshift (x64)}; Server...
	Multi-AZ No		

Below the general information, there are tabs for Cluster performance, Query monitoring, Schedules, Maintenance, and Properties.



The screenshot shows the 'Create cluster' configuration page in the AWS Redshift console. The cluster identifier is 'redshift-cluster-1'. The user is planning to use the cluster for 'Production'. The configuration is set to 'Free trial'.

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.
redshift-cluster-1
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

☒ **I'll choose**

☐ **Help me choose**

aws

Services

Search

[Alt+S]

N. Virginia

upgradamanrawat @ 8201-8921-4921

Help me choose

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)

Configuration summary

Info

dc2.large | 2 nodes

\$360.00/month

Estimated on-demand compute price

Save more than 60% of your costs by purchasing reserved nodes.

Learn more

320 GB

Total compressed storage

The total storage capacity for the cluster if you deploy the number of nodes that you chose.

aws

Services

Search

[Alt+S]

N. Virginia

upgradamanrawat @ 8201-8921-4921

Amazon Redshift

Redshift serverless

New

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Multi-AZ

No

Cluster performance

Query monitoring

Schedules

Maintenance

Properties

Database configurations

Edit admin credentials

Rotate encryption keys

Edit

Database name

dev

Parameter group

Defines database parameter and query queues for all the databases.

default.redshift-1.0

Port

5439

SSH ingestion setting (cluster public key)

ssh-rsa AAAAB3NzaC1yc2E...

Encryption

Disabled

AWS KMS key ID

-

Audit logging

Disabled

Network and security settings

Edit

aws

Services

Search

[Alt+S]

N. Virginia

upgradamanrawat @ 8201-8921-4921

Associated IAM roles (1)

Info

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.

Set default

Manage IAM roles

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

IAM roles

myredshiftrole

Status

Not applied

Role type

--

Additional configurations

Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

Network and security

Info

Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

aws

Services

Q Search

[Alt+S]

N. Virginia

upgradamanrawat @ 8201-8921-4921

modify these settings now.

Network and security Info

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

Default VPC
vpc-0565c3c933839e5a3

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

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

default
sg-02ead8c9009bce623

Cluster subnet group Info
Choose the Amazon Redshift subnet group to launch the cluster in.

cluster-subnet-group-1

Availability Zone
Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

No preference

Feedback

Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms

Cookie preferences

aws

Services

Q Search

[Alt+S]

N. Virginia

upgradamanrawat @ 8201-8921-4921

Select database Info
To view schemas, select a database.

dev

Select schema Info
To view tables, select a schema.

public

Filter tables

< 1 >

No resources
No resources to display

1 create schema data_atm;

Run Save Schedule Clear

Send feedback

Query results Table details

Query

Execution Data Visualize

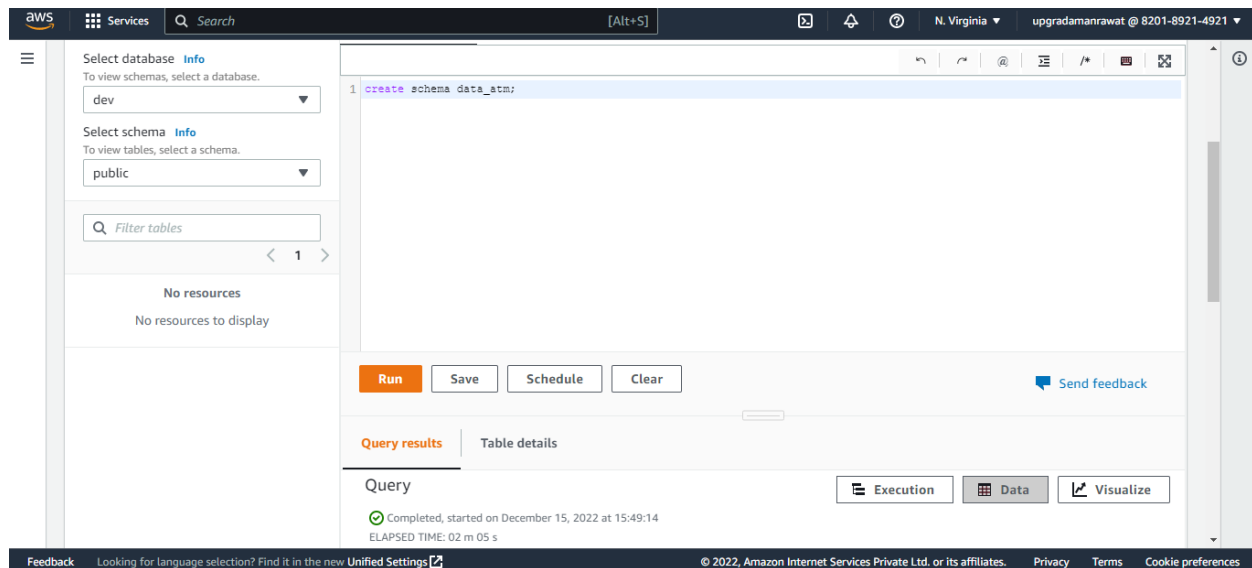
Completed, started on December 15, 2022 at 15:49:14
ELAPSED TIME: 02 m 05 s

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

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

Query for creating schema:

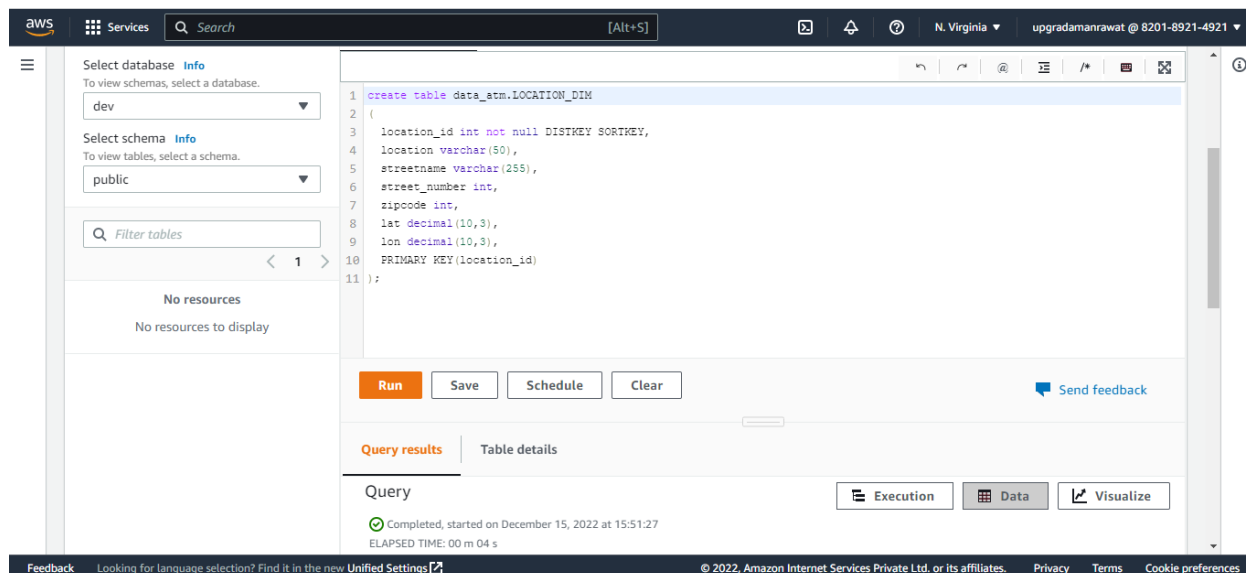
create schema data_atm;



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

• **Creating location dimension table:**

```
create table data_atm.LOCATION_DIM
(
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 Glue console interface. On the left, there's a sidebar with 'Select database' (dev) and 'Select schema' (public). The main area displays a SQL query to create a table named 'data_atm.LOCATION_DIM'. The query defines columns: 'location_id' (int, not null, DISTKEY SORTKEY), 'location' (varchar(50)), 'streetname' (varchar(255)), 'street_number' (int), 'zipcode' (int), 'lat' (decimal(10,3)), and 'lon' (decimal(10,3)). It also sets 'location_id' as the primary key. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating successful execution on December 15, 2022, at 15:51:27.

• Creating atm dimension table

create table data_atm.ATM_DIM

(

atm_id int not null DISTKEY SORTKEY,

atm_number varchar(20),

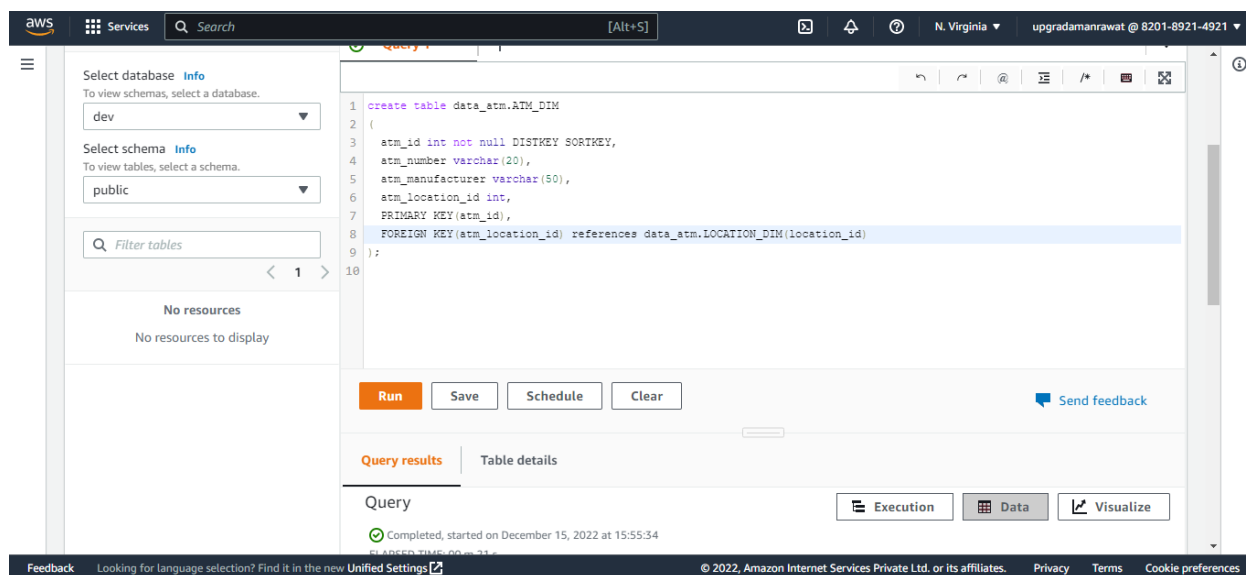
atm_manufacturer varchar(50),

atm_location_id int,

PRIMARY KEY(atm_id),

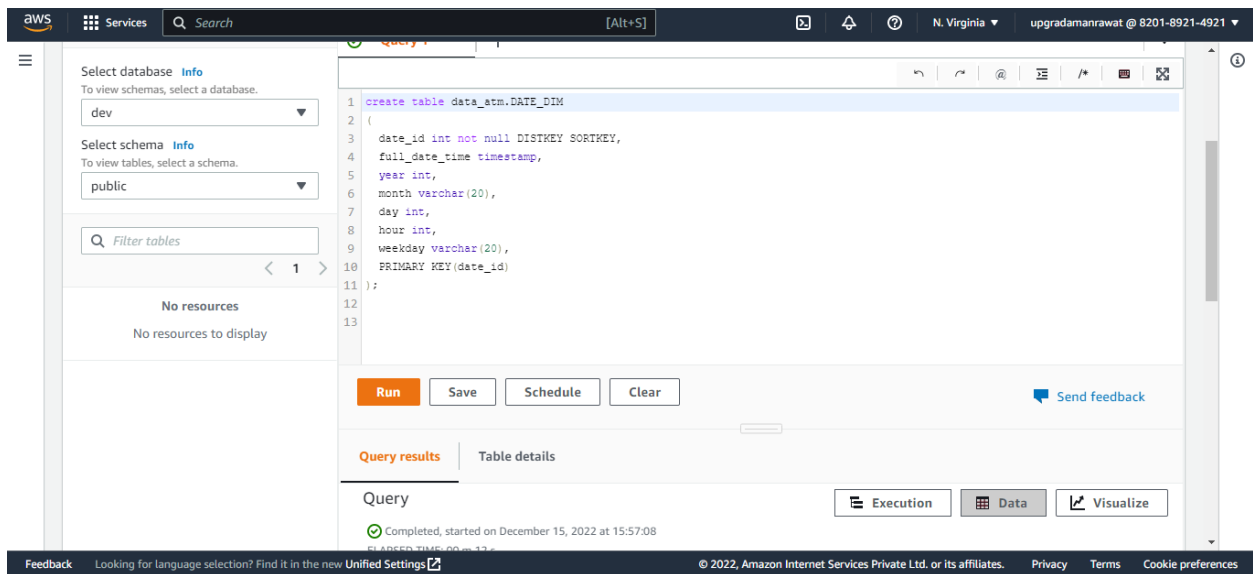
FOREIGN KEY(atm_location_id) references data_atm.LOCATION_DIM (location_id)

);



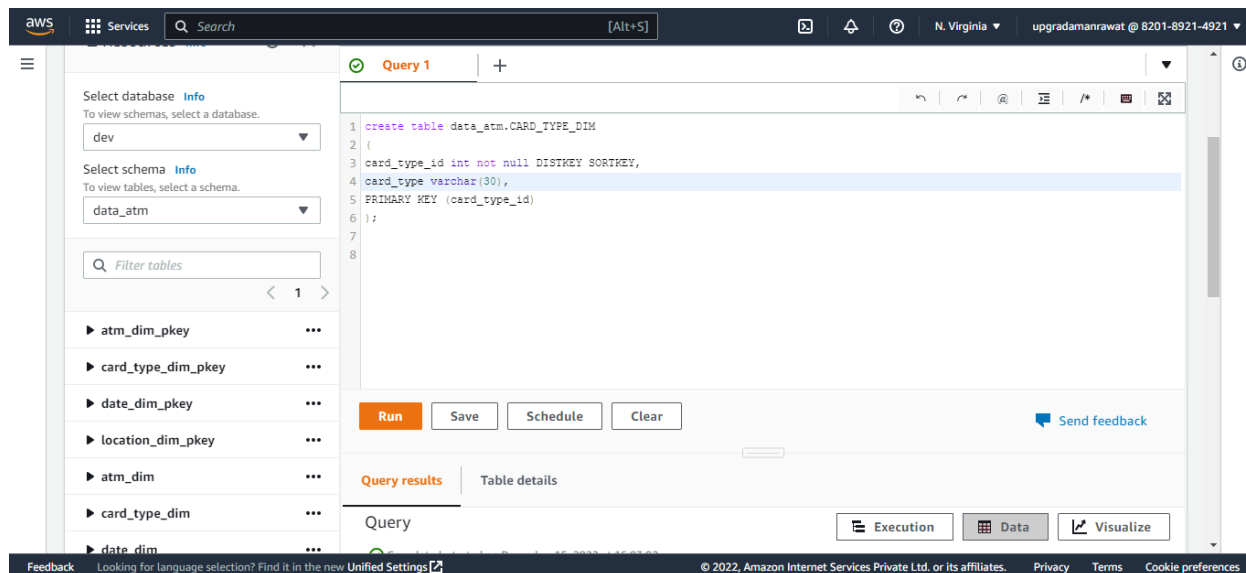
The screenshot shows the AWS Glue console interface. On the left, there's a sidebar with 'Select database' (dev) and 'Select schema' (public). The main area displays a SQL query to create a table named 'data_atm.ATM_DIM'. The query defines columns: 'atm_id' (int, not null, DISTKEY SORTKEY), 'atm_number' (varchar(20)), 'atm_manufacturer' (varchar(50)), and 'atm_location_id' (int). It sets 'atm_id' as the primary key and includes a foreign key constraint: 'FOREIGN KEY(atm_location_id) references data_atm.LOCATION_DIM(location_id)'. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating successful execution on December 15, 2022, at 15:55:34.

- Creating date dimension table
- ```
create table data_atm.DATE_DIM
(
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)
);
```



- Creating card-type dimension table

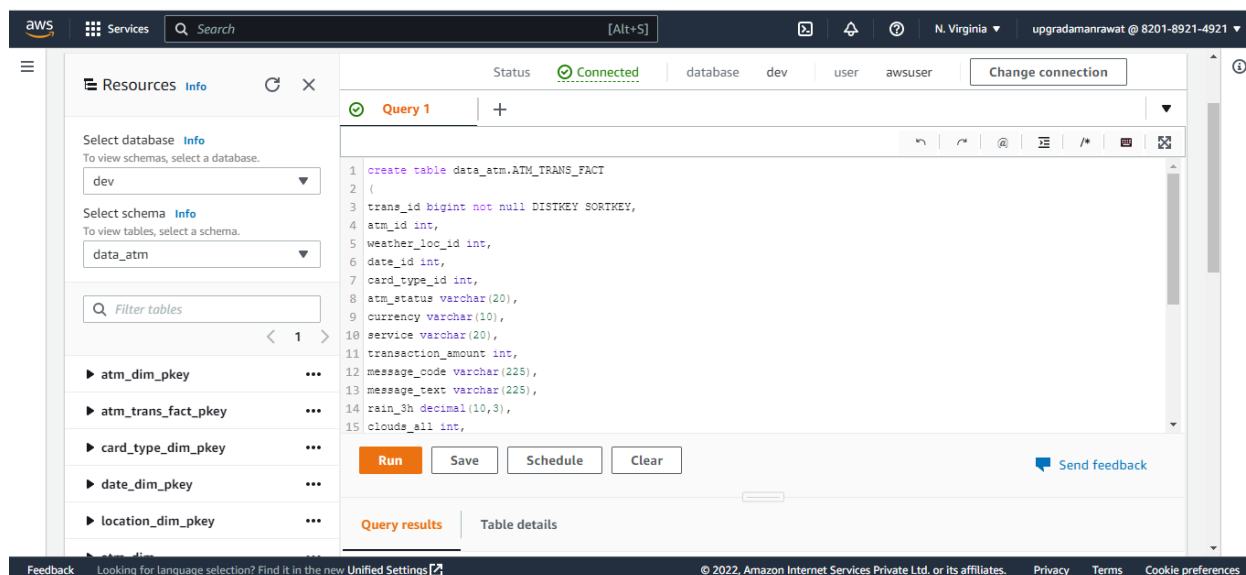
```
create table data_atm.CARD_TYPE_DIM
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);
```



## • Creating atm transactions fact table

```
create table data_atm.ATM_TRANS_FACT
(
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 data_atm.LOCATION_DIM (location_id),
FOREIGN KEY(atm_id) references data_atm.ATM_DIM(atm_id),
FOREIGN KEY(date_id) references data_atm.DATE_DIM(date_id),
FOREIGN KEY(card_type_id) references data_atm.CARD_TYPE_DIM(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

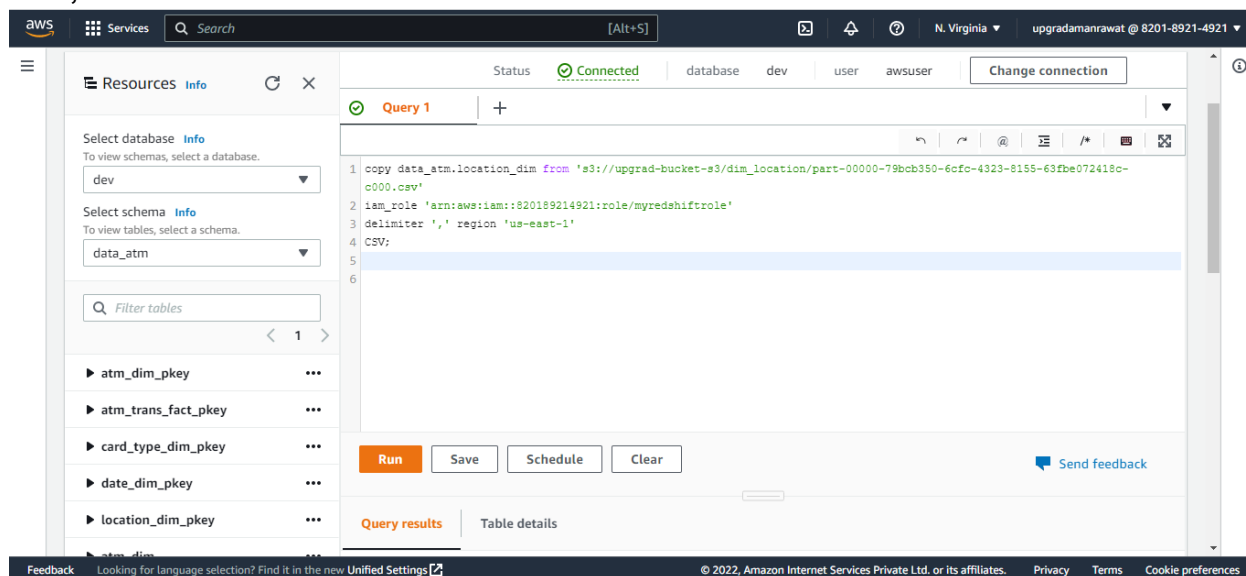
- Copying the data to location\_dim table

copy data\_atm.location\_dim from 's3://upgrad-bucket-s3/dim\_location/part-00000-79bcb350-6cfc-4323-8155-63fbe072418c-c000.csv'

iam\_role 'arn:aws:iam::820189214921:role/myredshiftrole'

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

CSV;



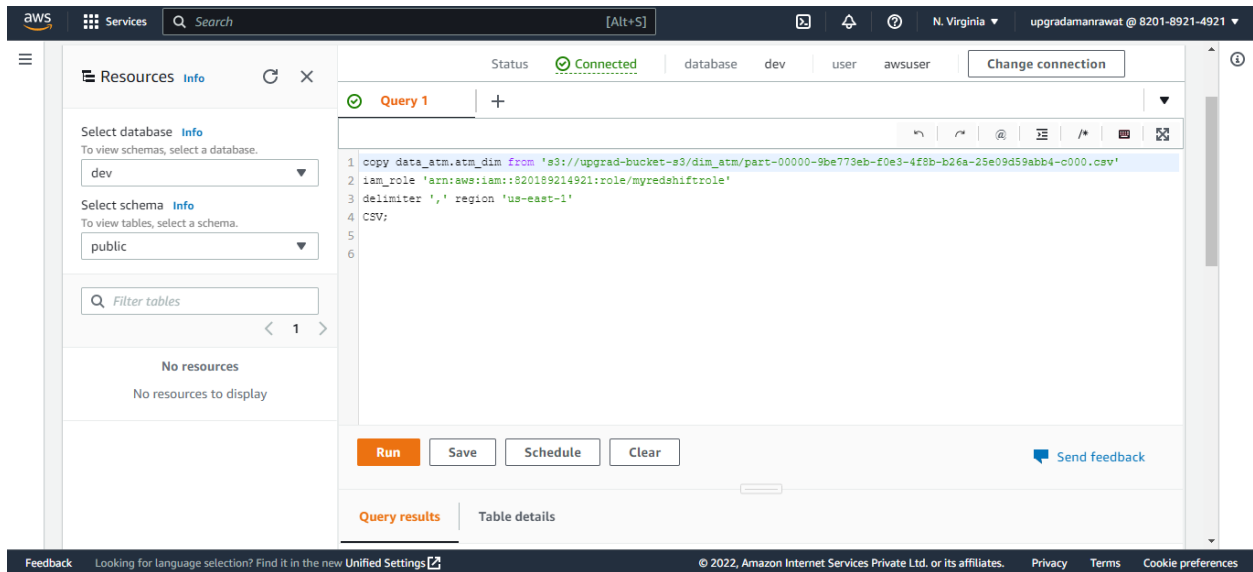
- Copying the data to atm\_dim table

**copy data\_atm.atm\_dim from 's3://upgrad-bucket-s3/dim\_atm/part-00000-9be773eb-f0e3-4f8b-b26a-25e09d59abb4-c000.csv'**

**iam\_role 'arn:aws:iam::820189214921:role/myredshiftrole'**

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

**CSV;**



- Copying the data to date\_dim table

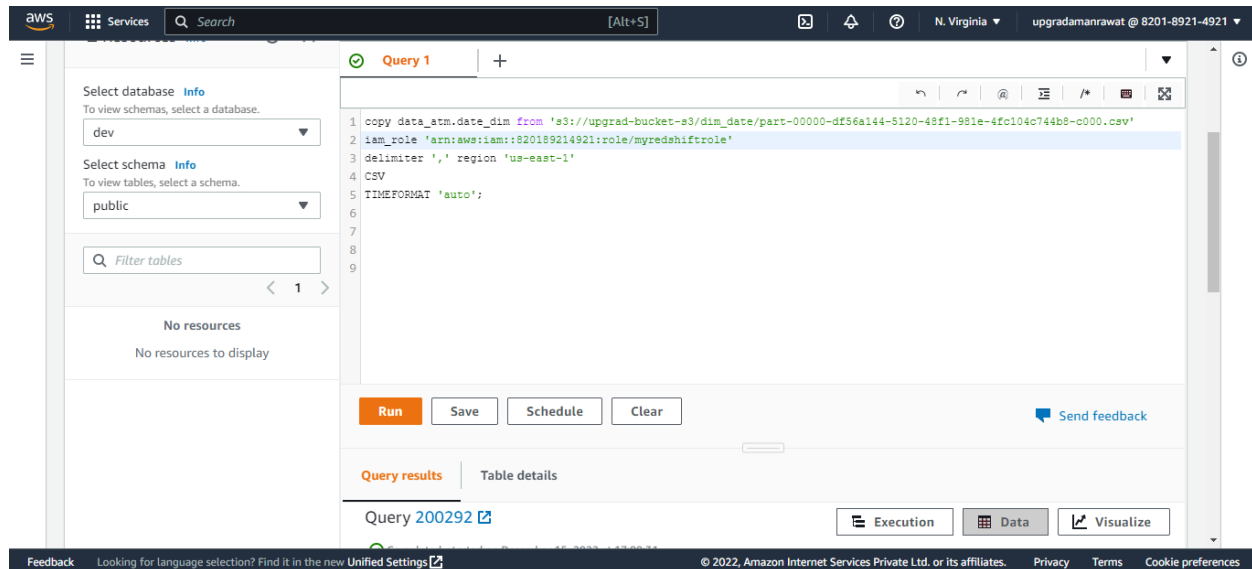
**copy data\_atm.atm\_dim from 's3://upgrad-bucket-s3/dim\_date/part-00000-df56a144-5120-48f1-981e-4fc104c744b8-c000.csv'**

**iam\_role 'arn:aws:iam::820189214921:role/myredshiftrole'**

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

**CSV TIMEFORMAT 'auto';**

**;**



The screenshot shows the AWS Glue console interface. On the left, there's a sidebar with 'Select database' (dev) and 'Select schema' (public). The main area displays a SQL query for 'Query 1':

```
1 copy_data_atm.date_dim from 's3://upgrad-bucket-s3/dim_date/part-00000-df56a144-5120-48f1-981e-4fc104c744b8-c000.csv'
2 iam_role 'arn:aws:iam::820189214921:role/myredshiftrole'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing 'Query 200292'. The bottom of the console shows the footer with '© 2022, Amazon Internet Services Private Ltd. or its affiliates.' and links for 'Privacy', 'Terms', and 'Cookie preferences'.

## • Copying the data to card\_type\_dim table

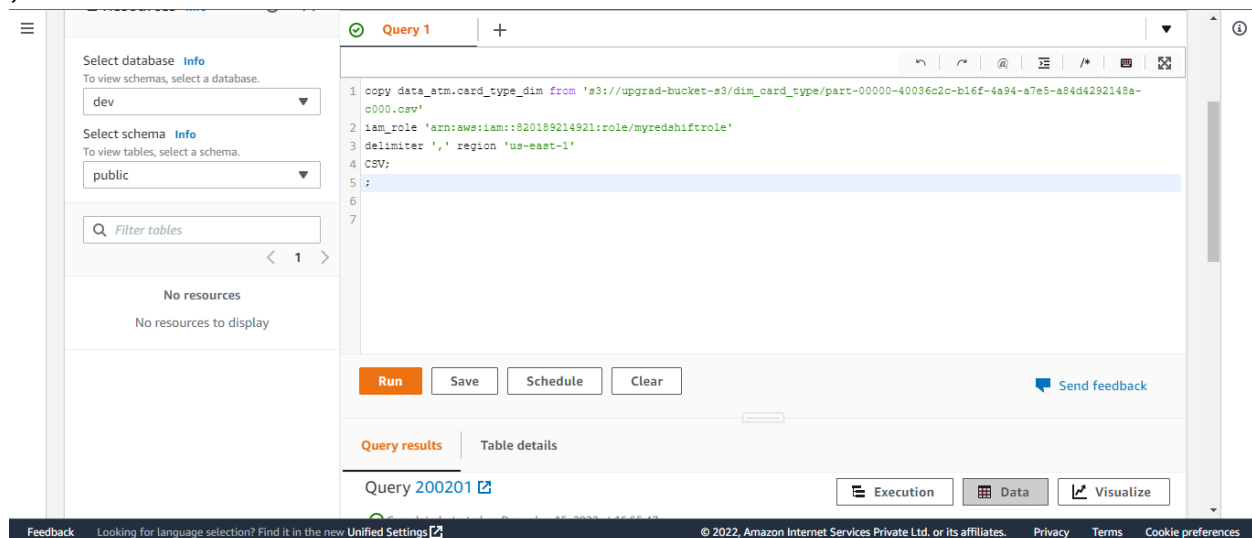
**copy\_data\_atm.card\_type\_dim from 's3://upgrad-bucket-s3/dim\_card\_type/part-00000-40036c2c-b16f-4a94-a7e5-a84d4292148a-c000.csv'**

**iam\_role 'arn:aws:iam::820189214921:role/myredshiftrole'**

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

**CSV;**

**;**



The screenshot shows the AWS Glue console interface. On the left, there's a sidebar with 'Select database' (dev) and 'Select schema' (public). The main area displays a SQL query for 'Query 1':

```
1 copy_data_atm.card_type_dim from 's3://upgrad-bucket-s3/dim_card_type/part-00000-40036c2c-b16f-4a94-a7e5-a84d4292148a-c000.csv'
2 iam_role 'arn:aws:iam::820189214921:role/myredshiftrole'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing 'Query 200201'. The bottom of the console shows the footer with '© 2022, Amazon Internet Services Private Ltd. or its affiliates.' and links for 'Privacy', 'Terms', and 'Cookie preferences'.

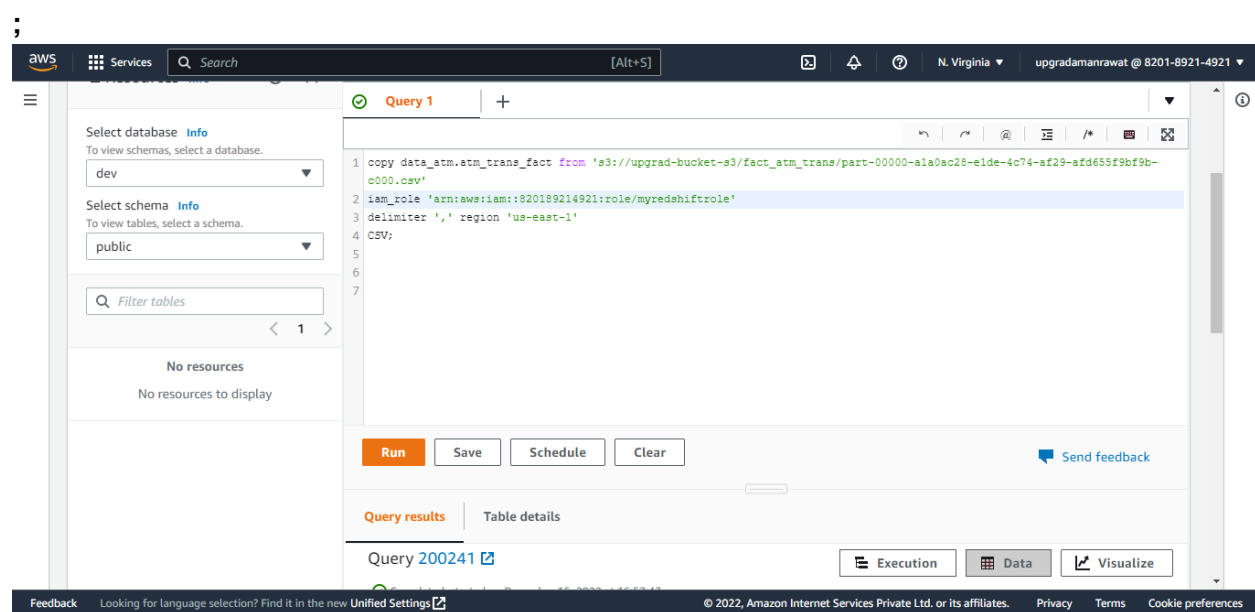
- Copying the data to atm\_trans\_fact table

copy data\_atm.atm\_trans\_fact from 's3://upgrad-bucket-s3/dim\_card\_type/part-00000-a1a0ac28-e1de-4c74-af29-afd655f9bf9b-c000.csv'

iam\_role 'arn:aws:iam::820189214921:role/myredshiftrole'

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

CSV;



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Select database' (set to 'dev') and 'Select schema' (set to 'public'). The main area displays a SQL query in a text editor:

```
1 copy data_atm.atm_trans_fact from 's3://upgrad-bucket-s3/dim_card_type/part-00000-a1a0ac28-e1de-4c74-af29-afd655f9bf9b-c000.csv'
2 iam_role 'arn:aws:iam::820189214921:role/myredshiftrole'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. At the bottom, there's a 'Query results' tab and a 'Table details' tab. The 'Query results' tab is active, showing 'Query 200241'.