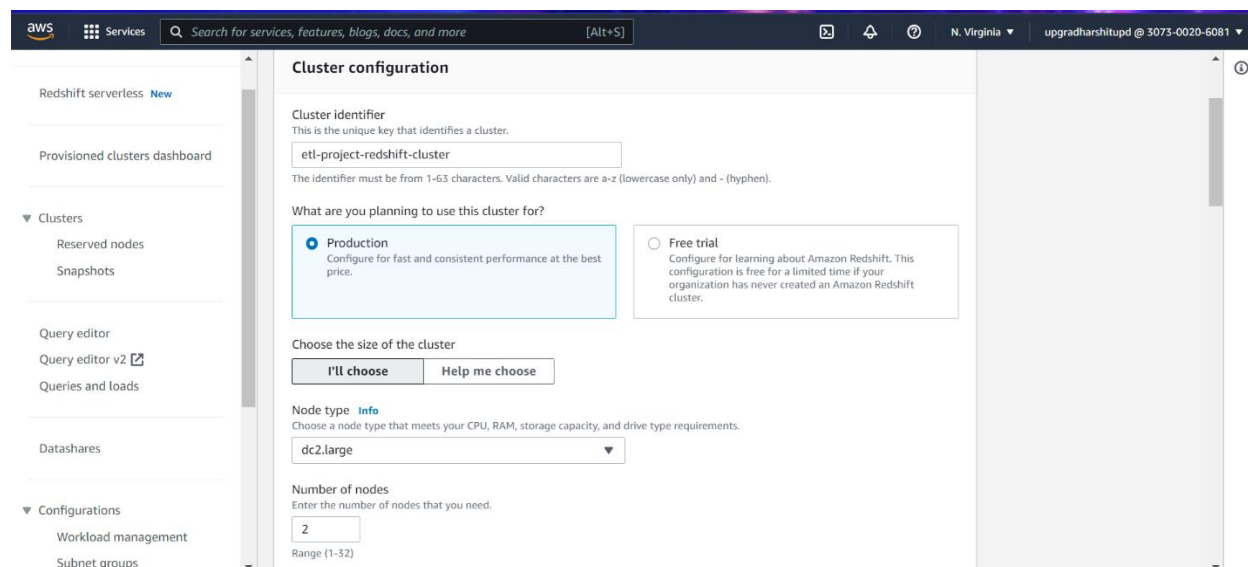


## Creation of a Redshift Cluster

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



**Cluster configuration**

**Cluster identifier**  
This is the unique key that identifies a cluster.  
etl-project-redshift-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?**

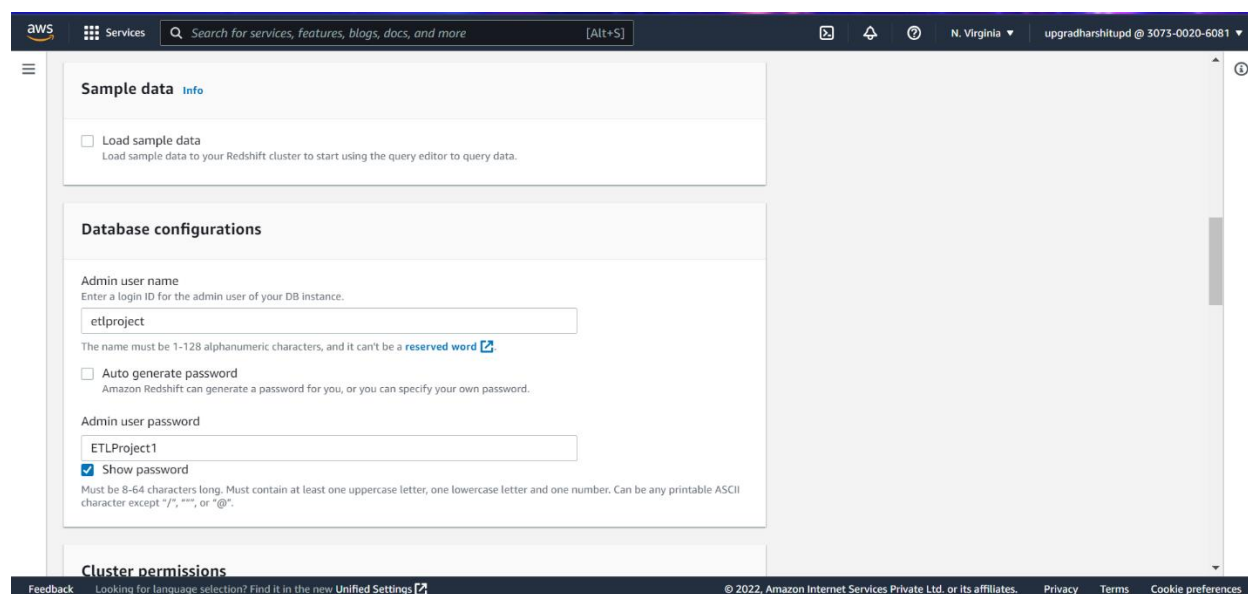
☒ **Production**  
Configure for fast and consistent performance at the best price.

☐ **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.  
dc2.large

**Number of nodes**  
Enter the number of nodes that you need.  
2  
Range (1-32)



**Sample data** [Info](#)

☐ **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.  
etlproject  
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**  
ETLProject1  
☒ **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**



Services

Search for services, features, blogs, docs, and more

[Alt+S]

N. Virginia

upgradharshitupd @ 3073-0020-6081

▼ Network and security info

Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

my\_vpc-vpc

vpc-0d2627baea90424bb

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

MySecGrp

sg-0e16fcd04ef3df1b5

Cluster subnet group

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

eti-project-subnet-group

Availability Zone


Specify the Availability Zone that you want the cluster to be created in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

No preference

Enhanced VPC routing

Enabling this option forces network traffic between your cluster and data repositories through a VPC, instead of the internet. [Learn more](#)

☒ Turn off



Services

Search for services, features, blogs, docs, and more

[Alt+S]

N. Virginia

upgradharshitupd @ 3073-0020-6081

► Network and security info

▼ Database configurations

Database name

Specify a database name to create an additional database.

etlproject

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.

2222

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)

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

etl-project-redshift-cluster

Actions Edit Add partner integration Query data

**General information**

Cluster identifier etl-project-redshift-cluster	Status Available	Node type dc2.large	Endpoint etl-project-redshift-cluster.ct0sxrei7gz1.u...
Cluster namespace afb2ae87-5764-4208-b391-d06e7dca683f	Date created August 10, 2022, 13:39 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://etl-project-redshift-cluster....
	Storage used -	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=e...

Cluster performance Query monitoring Schedules Maintenance Properties

**Database configurations**

Change admin user password Rotate encryption keys Edit

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

Associate IAM roles

**IAM roles (1/3)**  
Choose from existing IAM roles. You can associate up to 50 IAM roles with this cluster.

Search for IAM role to associate

☐ AWSServiceRoleForRedshift

☒ etl\_project\_iam

☐ upGrad\_RedShift\_S3

Cancel Associate IAM roles

Grant access

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

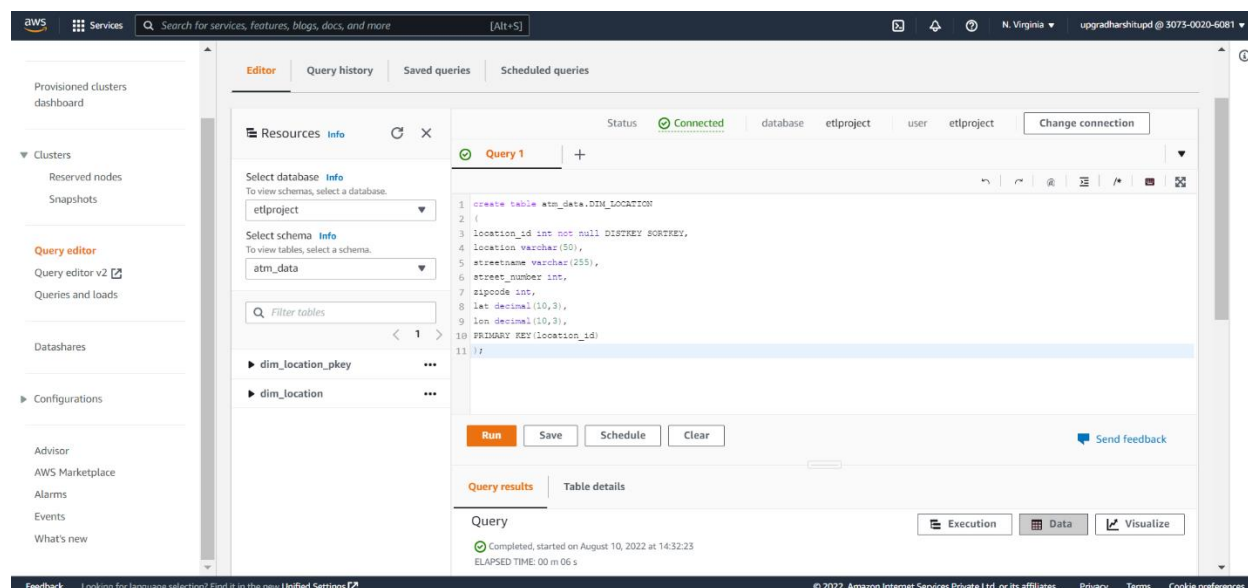
© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences

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:**

### 1. 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 console interface. On the left, there is a navigation pane with options like 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main area is titled 'Query editor' and shows a SQL query being executed. The query is:
 

```
1 create table atm_data.DIM_LOCATION
2 (
3 location_id int not null DISTKEY SORTKEY,
4 location varchar(50),
5 streetname varchar(255),
6 street_number int,
7 zipcode int,
8 lat decimal(10,3),
9 lon decimal(10,3),
10 PRIMARY KEY(location_id)
11 );
```

 The query is labeled 'Query 1' and has a status of 'Connected'. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the query, there is a section for 'Query results' and 'Table details'. The 'Query results' section shows the query status as 'Completed, started on August 10, 2022 at 14:32:23' and 'ELAPSED TIME: 00 m 06 s'. The 'Table details' section shows the table name 'dim\_location' and its schema 'atm\_data'.

## 2. CREATING ATM DIMENSION TABLE

create table atm\_data.DIM\_ATM

(

atm\_id int not null DISTKEY SORTKEY,

atm\_number varchar(20),

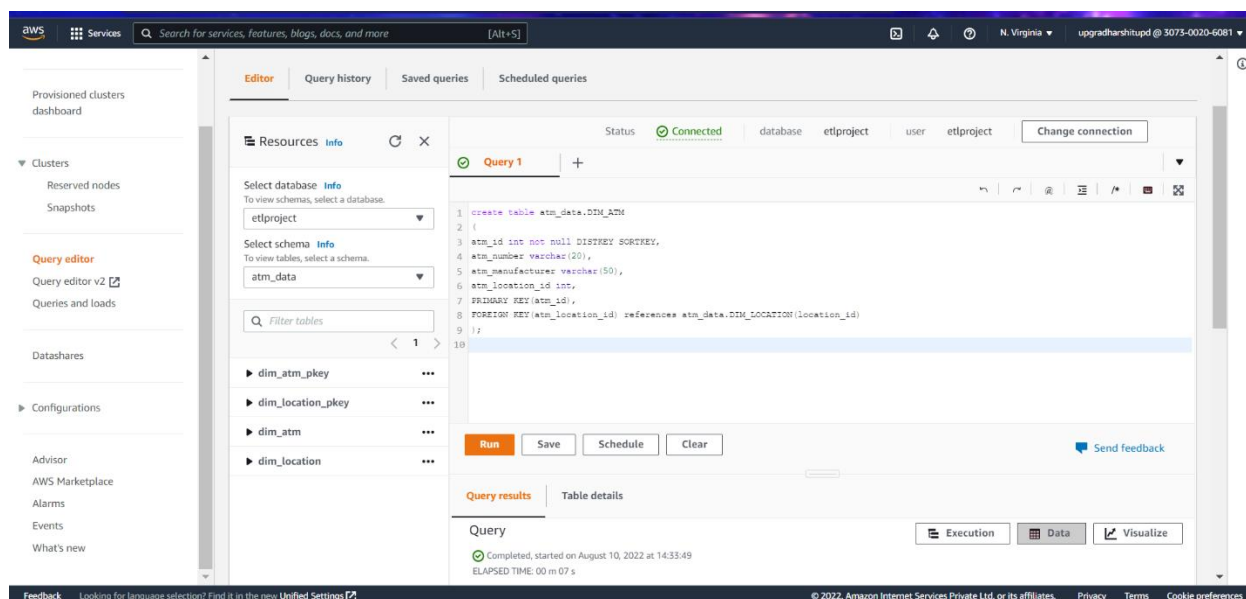
atm\_manufacturer varchar(50),

atm\_location\_id int,

PRIMARY KEY(atm\_id),

FOREIGN KEY(atm\_location\_id) references atm\_data.DIM\_LOCATION(location\_id)

);



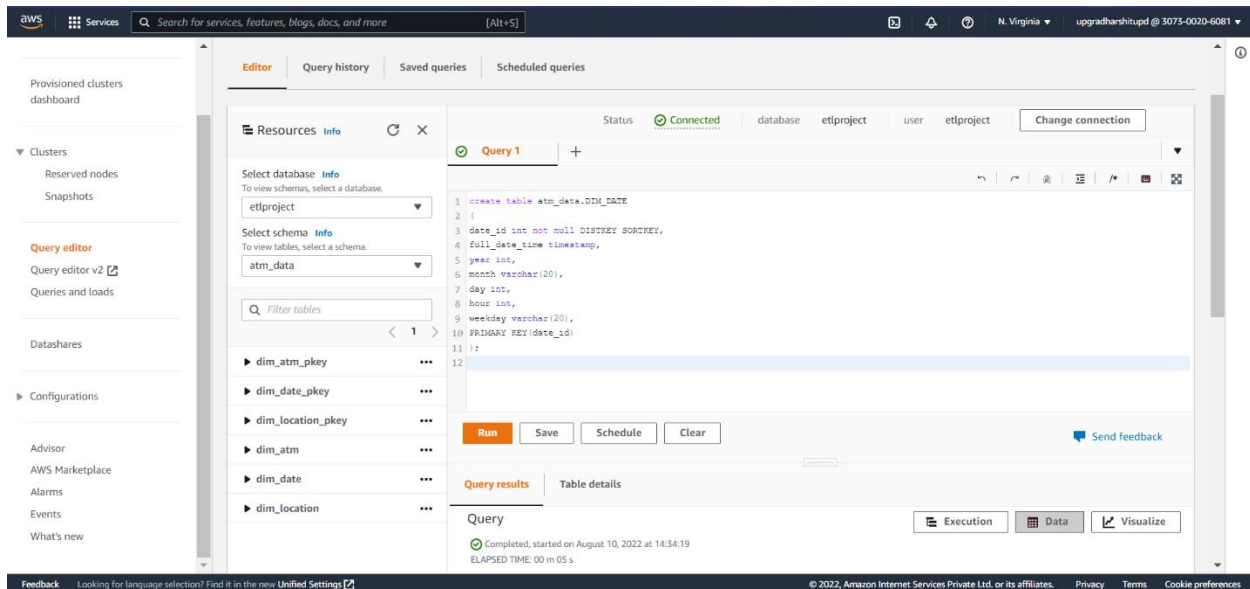
The screenshot shows the AWS Glue console interface. On the left, there is a navigation pane with options like 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main area is titled 'Editor' and shows a query editor for 'Query 1'. The query is as follows:

```
1 create table atm_data.DIM_ATM
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' section shows that the query is 'Completed, started on August 10, 2022 at 14:33:49' with an 'ELAPSED TIME: 00 m 07 s'. The 'Table details' section is also visible.

### 3. 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)
);
```



The screenshot shows the AWS Glue console interface. On the left, there is a navigation menu with options like 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main area is titled 'Editor' and shows a query editor for a database named 'etlproject' and schema 'atm\_data'. The query being edited is:

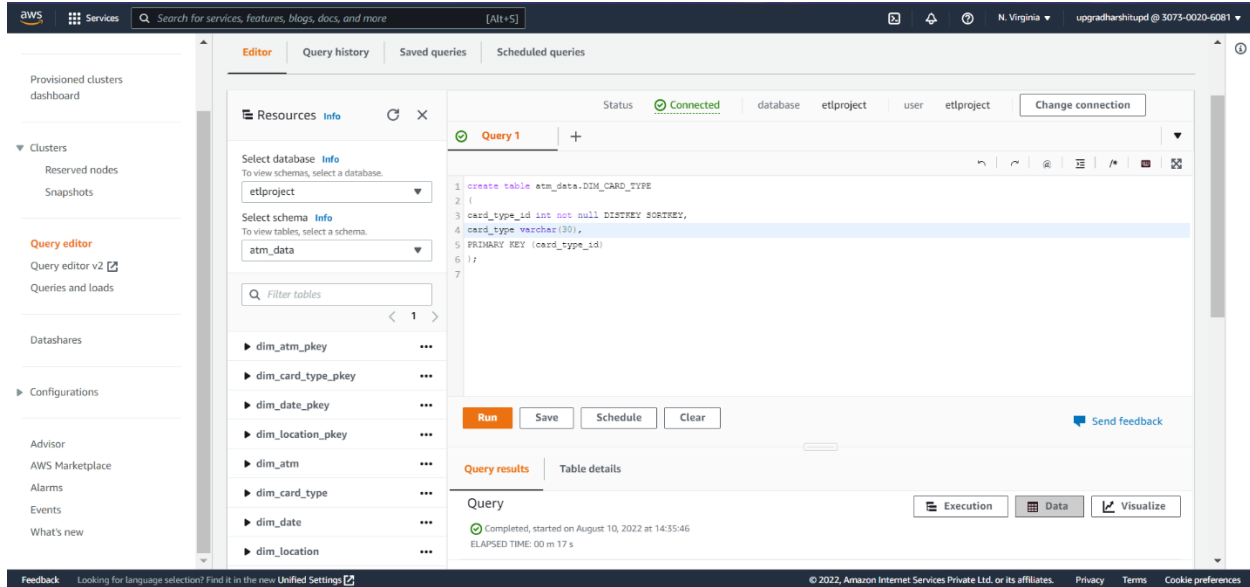
```
1 create table atm_data.DIM_DATE
2 (
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int,
6   month varchar(20),
7   day int,
8   hour int,
9   weekday varchar(20),
10  PRIMARY KEY(date_id)
11 )
12
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a message: 'Query Completed, started on August 10, 2022 at 14:34:19 ELAPSED TIME: 00 m 05 s'. The bottom of the console shows a footer with 'Feedback', 'Looking for language selection? Find it in the new Unified Settings', and copyright information: '© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences'.

#### 4. 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)
);
```



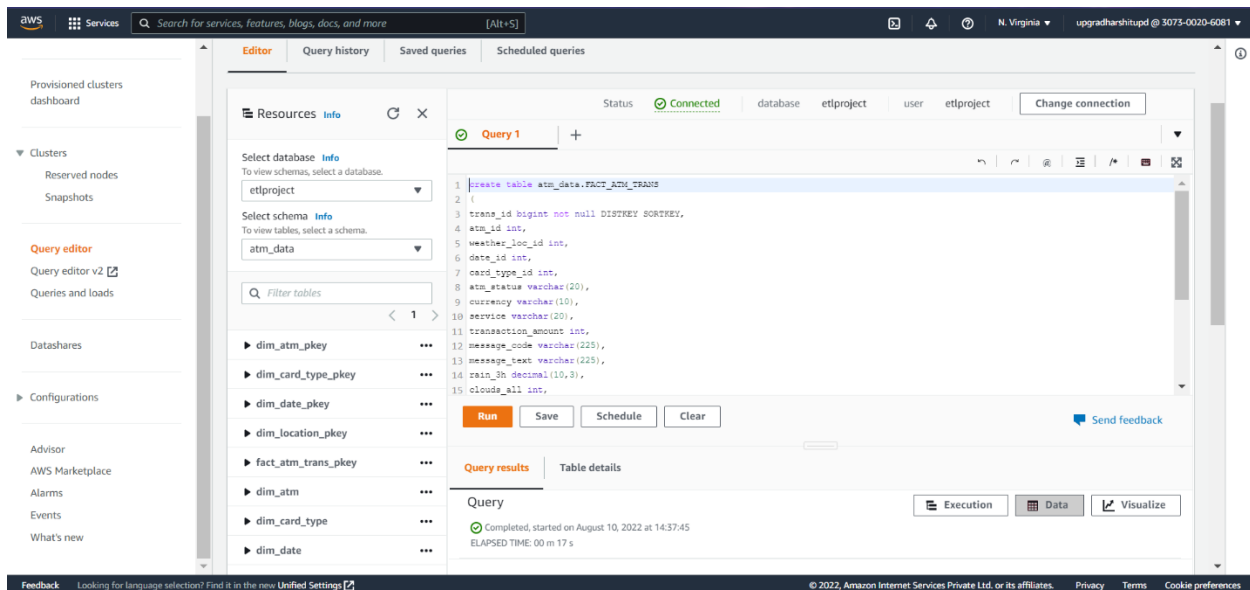
The screenshot shows the AWS Glue console interface. On the left, there is a navigation pane with options like 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main area is titled 'Editor' and shows a query editor for 'Query 1'. The query is as follows:

```
1 create table atm_data.DIM_CARD_TYPE
2 (
3   card_type_id int not null DISTKEY SORTKEY,
4   card_type varchar(30),
5   PRIMARY KEY (card_type_id)
6 );
7
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there is a section for 'Query results' and 'Table details'. The 'Query results' section shows the status 'Completed, started on August 10, 2022 at 14:35:46' and 'ELAPSED TIME: 00 m 17 s'. The 'Table details' section is currently empty. At the bottom of the console, there is a footer with 'Feedback', 'Looking for language selection? Find it in the new Unified Settings', and '© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences'.

## 5. 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 Glue console interface. On the left, there is a navigation pane with options like 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Databshares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main area is titled 'Editor' and shows a query editor for a database named 'etlproject' and schema 'atm\_data'. The query being edited is the SQL statement to create the 'FACT\_ATM\_TRANS' table, which is the same as the one provided in the previous block. The query is numbered 1 through 15. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of the query editor, there is a 'Query results' section showing the query execution status as 'Completed, started on August 10, 2022 at 14:37:45' and 'ELAPSED TIME: 00 m 17 s'. The bottom of the console shows a footer with 'Feedback', a language selection prompt, and copyright information for Amazon Internet Services Private Ltd. and upGrad Education Pvt. Ltd.



## 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. COPYING THE DATA TO DIM\_LOCATION TABLE

```
copy atm_data.dim_location from 's3://etl-project-  
s3bucket/etlprojecttables/dim_location/part-00000-67787b48-208d-461f-94d0-  
c13dae0aa728-c000.csv'  
iam_role 'arn:aws:iam::307300206081:role/etl_project_iam'  
delimiter ',' region 'us-east-1'  
CSV;
```

#### 2. COPYING THE DATA TO DIM\_ATM TABLE

```
copy atm_data.dim_atm from 's3://etl-project-s3bucket/etlprojecttables/dim_atm/part-  
00000-4e6df224-283c-4cb3-8450-07c43d796927-c000.csv'  
iam_role 'arn:aws:iam::307300206081:role/etl_project_iam'  
delimiter ',' region 'us-east-1'  
CSV;
```

#### 3. COPYING THE DATA TO DIM\_DATE TABLE

```
copy atm_data.dim_date from ' s3://etl-project-s3bucket/etlprojecttables/dim_date/part-  
00000-04a077c7-5703-4dfd-b874-ac598376e534-c000.csv'  
iam_role 'arn:aws:iam::307300206081:role/etl_project_iam'  
delimiter ',' region 'us-east-1'  
timeformat 'YYYY-MM-DDTHH:MI:SS'  
CSV;
```

#### 4. COPYING THE DATA TO DIM\_CARD\_TYPE TABLE

```
copy atm_data.dim_card_type from ' s3://etl-project-  
s3bucket/etlprojecttables/dim_card_type/part-00000-23fc3fe1-9f42-420c-a26a-  
fd6dc9450f13-c000.csv'  
iam_role 'arn:aws:iam::307300206081:role/etl_project_iam'  
delimiter ',' region 'us-east-1'  
CSV;
```

#### 5. COPYING THE DATA TO FACT\_ATM\_TRANS TABLE

```
copy atm_data.fact_atm_trans from ' s3://etl-project-  
s3bucket/etlprojecttables/fact_atm_trans/part-00000-3c559558-d3d1-4dda-92a3-  
84df7abf4849-c000.csv'  
iam_role 'arn:aws:iam::307300206081:role/etl_project_iam'  
delimiter ',' region 'us-east-1'  
CSV;
```