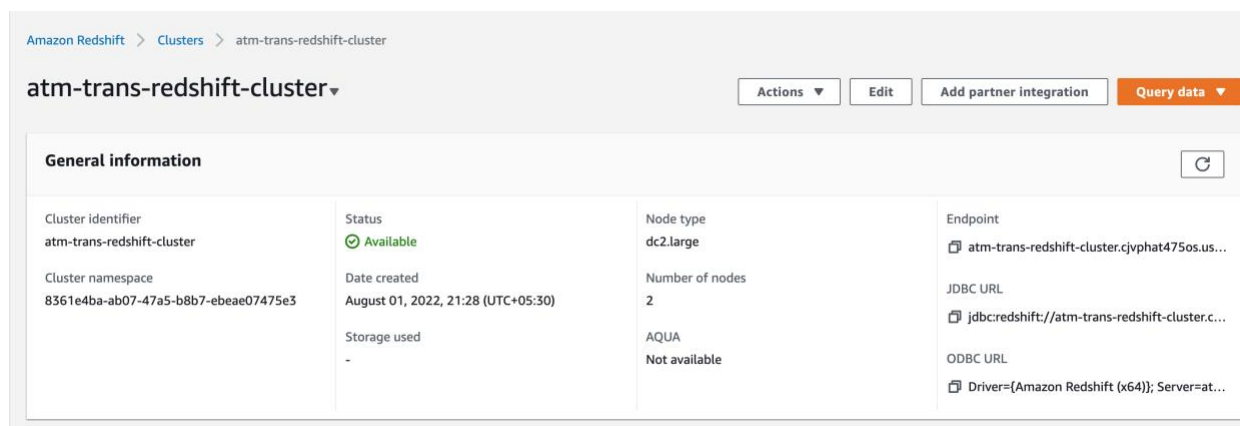


Creation of a Redshift Cluster

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

<Screenshot of the type of machine used along with number of nodes>



Amazon Redshift > Clusters > atm-trans-redshift-cluster			
atm-trans-redshift-cluster▼			
<div>Actions ▼ Edit Add partner integration Query data ▼</div>			
General information ↻			
Cluster identifier atm-trans-redshift-cluster	Status Available	Node type dc2.large	Endpoint atm-trans-redshift-cluster.cjvphat475os.us...
Cluster namespace 8361e4ba-ab07-47a5-b8b7-ebeae07475e3	Date created August 01, 2022, 21:28 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://atm-trans-redshift-cluster.c...
	Storage used -	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=at...

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

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

Schema creation:

```
create schema atmTransData;
```

Location Dimension table creation:

```
create table atmTransData.dim_location(
location_id INT,
location VARCHAR(50),
streetname VARCHAR(255),
street_number INT,
zipcode INT,
lat DECIMAL(10,3),
lon DECIMAL(10,3));
```

Atm Dimension table creation:

```
create table atmTransData.dim_atm(  
atm_id INT,  
atm_number VARCHAR(20),  
atm_manufacturer VARCHAR(50),  
atm_location_id INT);
```

Date Dimension table creation:

```
create table atmTransData.dim_date(  
date_id INT,  
full_date_time TIMESTAMP,  
year INT,  
month VARCHAR(20),  
day INT,  
hour INT,  
weekday VARCHAR(20));
```

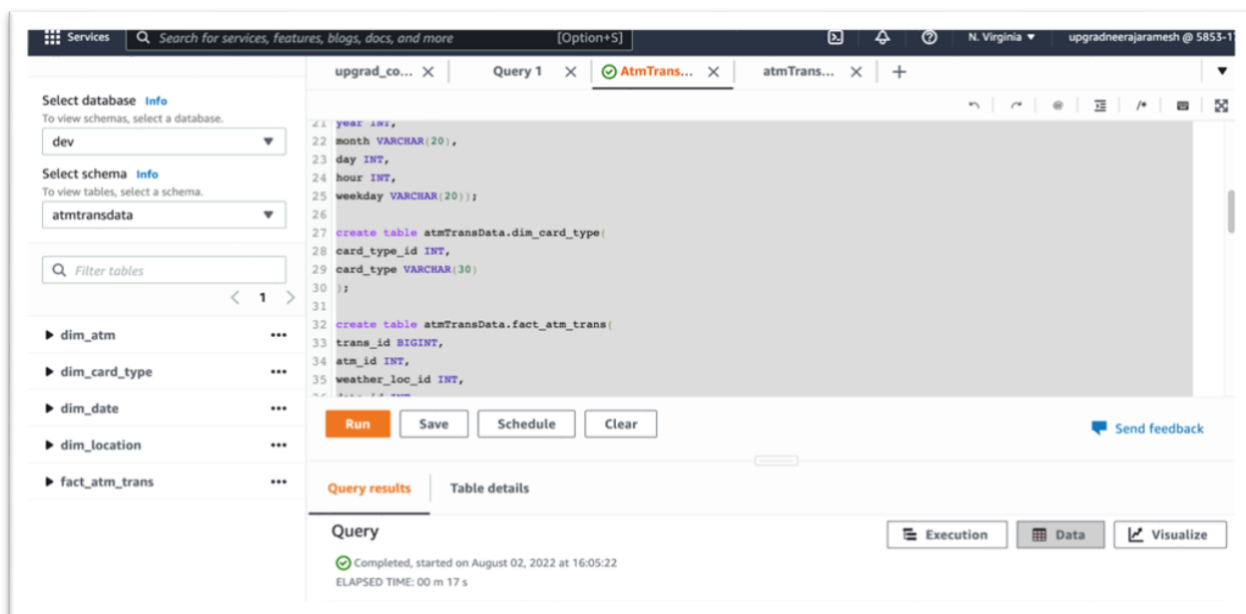
Card Dimension table creation:

```
create table atmTransData.dim_card_type(  
card_type_id INT,  
card_type VARCHAR(30)  
);
```

Fact_Atm_Trans fact table creation:

```
create table atmTransData.fact_atm_trans(  
trans_id BIGINT,  
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(1000),  
message_text VARCHAR(1000),  
rain_3h DECIMAL(10,3),  
clouds_all INT,  
weather_id INT,  
weather_main VARCHAR(50),  
weather_description VARCHAR(255));
```

Screenshot of success state of query



The screenshot displays the upGrad SQL IDE interface. On the left, the 'Select database' dropdown is set to 'dev', and the 'Select schema' dropdown is set to 'atmtransdata'. Below these, a list of tables is shown, including 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main editor area contains a SQL query for creating two tables: 'atmTransData.dim_card_type' and 'atmTransData.fact_atm_trans'. The query is as follows:

```

41 year INT,
22 month VARCHAR(20),
23 day INT,
24 hour INT,
25 weekday VARCHAR(20));
26
27 create table atmTransData.dim_card_type(
28 card_type_id INT,
29 card_type VARCHAR(30)
30 );
31
32 create table atmTransData.fact_atm_trans(
33 trans_id BIGINT,
34 atm_id INT,
35 weather_loc_id INT,

```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link. Below the buttons, the 'Query results' tab is active, showing a status message: 'Completed, started on August 02, 2022 at 16:05:22' and 'ELAPSED TIME: 00 m 17 s'. At the bottom right, there are three tabs: 'Execution', 'Data', and 'Visualize'.

3. 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 atmTransData.dim_location from
's3://atmtransdata/dimLoc/part-00000-16949c74-9c28-493b-af2e-f818c66388e4-c000.csv'
iam_role 'arn:aws:iam::585317706400:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

```
copy atmTransData.dim_atm from
's3://atmtransdata/dimAtm/part-00000-2d0bee25-6e48-4564-a96d-77c320ea95a3-c000.csv'
iam_role 'arn:aws:iam::585317706400:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

```
copy atmTransData.dim_card_type from
's3://atmtransdata/dimCard/part-00000-f764053d-601d-436e-b42b-19942e3d1e3d-c000.csv'
iam_role 'arn:aws:iam::585317706400:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

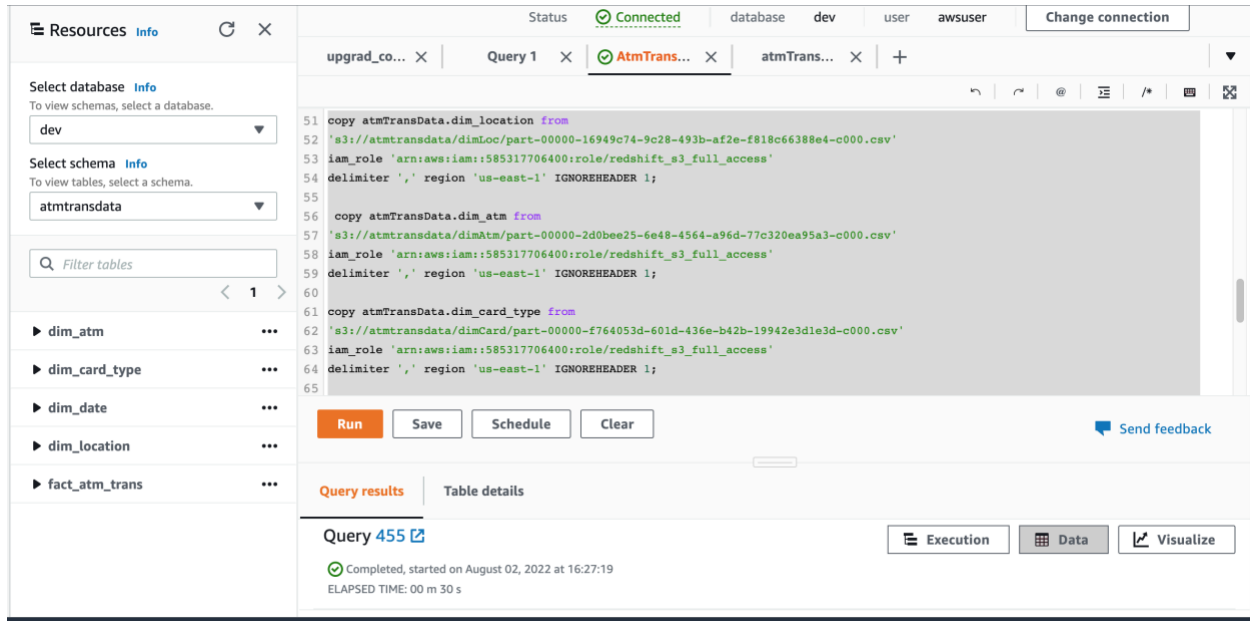
```
copy atmTransData.dim_date from
's3://atmtransdata/dimDate/part-00000-dadd0ed7-9b39-4db0-9ce1-dd50fae07258-c000.csv'
iam_role 'arn:aws:iam::585317706400:role/redshift_s3_full_access'
timeformat 'auto'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

```
copy atmTransData.fact_atm_trans from
's3://atmtransdata/factAtmTrans/part-00000-acb5589d-8ef4-475d-be0d-177eb086bdfd-c000.csv'
iam_role 'arn:aws:iam::585317706400:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

Queries to verify count in dimension and fact tables:

```
select count(1) from atmTransData.dim_location;
select count(1) from atmTransData.dim_card_type;
select count(1) from atmTransData.dim_atm;
select count(1) from atmTransData.dim_date;
select count(1) from atmTransData.fact_atm_trans;
```

Success screenshot:



The screenshot displays the upGrad SQL IDE interface. On the left, the 'Resources' panel shows the database 'dev' and schema 'atmtransdata'. The main editor area contains a SQL query with line numbers 51 to 65. The query is a multi-part INSERT statement using the 'copy' command to load data from S3 into three tables: 'dim_location', 'dim_atm', and 'dim_card_type'. The status bar at the top indicates 'Connected' and 'dev' database. The bottom panel shows 'Query results' and 'Table details' tabs, with the 'Query results' tab active, displaying 'Query 455' and its execution status: 'Completed, started on August 02, 2022 at 16:27:19, ELAPSED TIME: 00 m 30 s'.