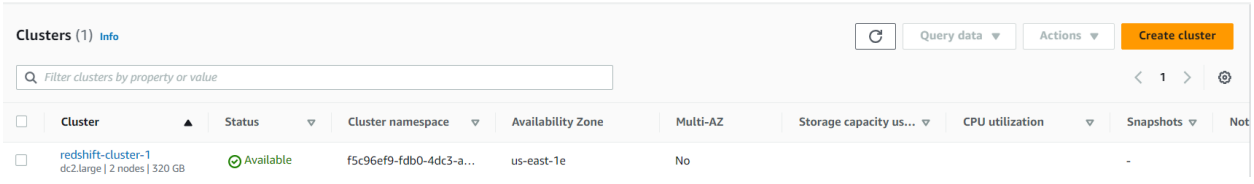
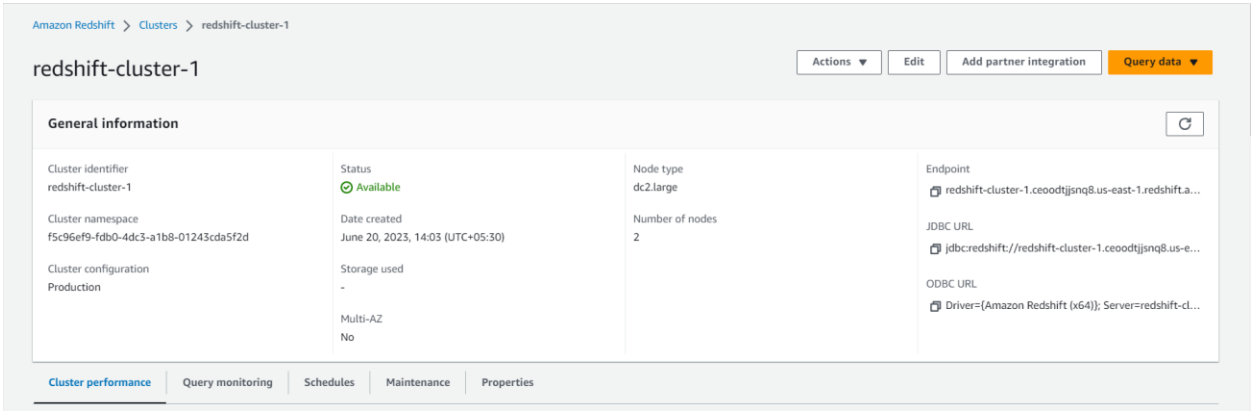
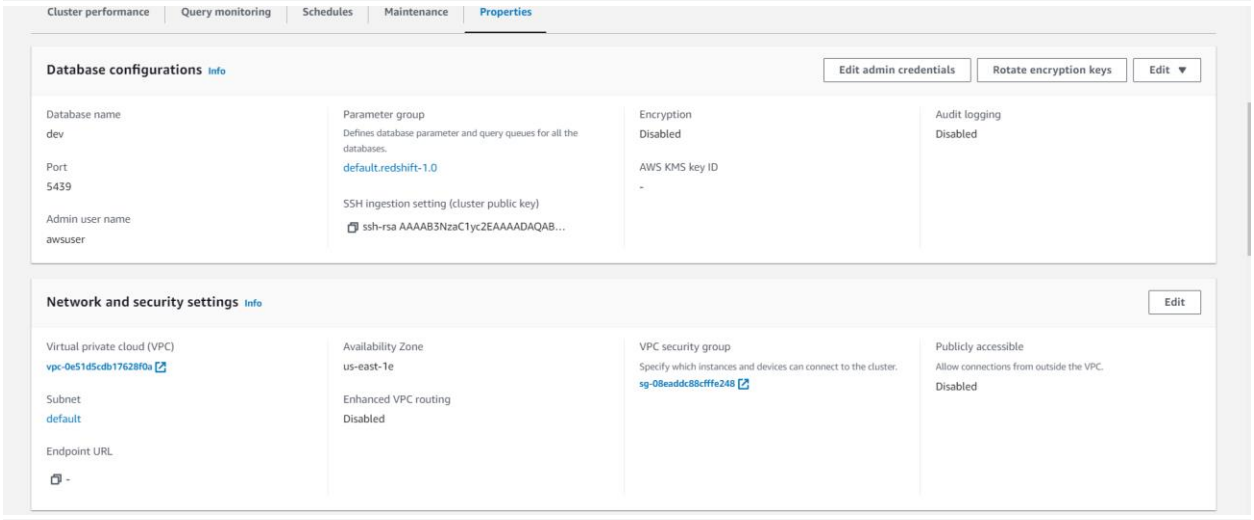
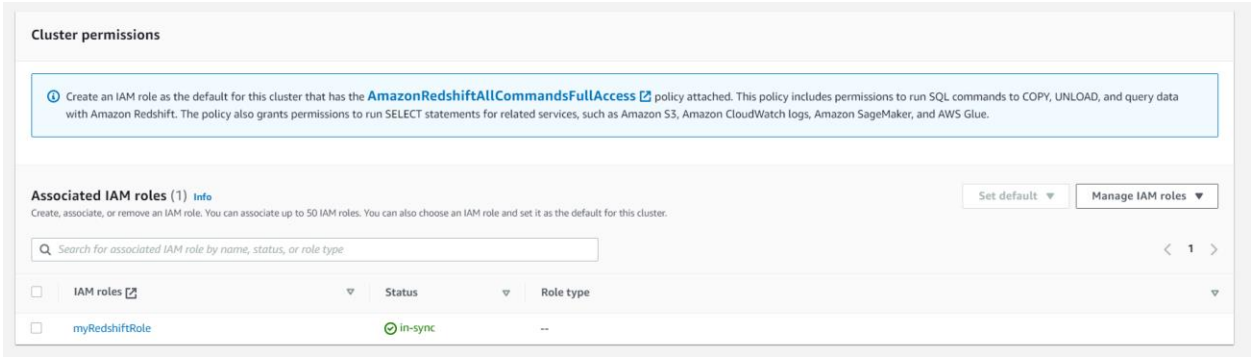
Creation of a Redshift Cluster

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







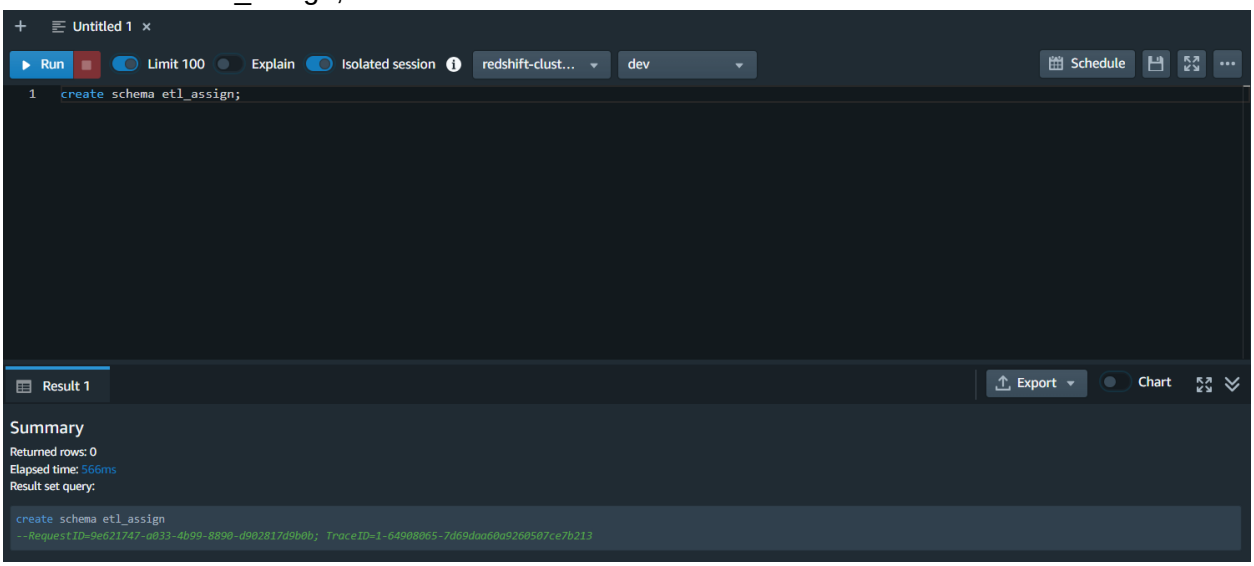


Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

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

Creating the schema –

create schema etl_assign;



Creating Dimension Tables –

1) **Creating location dimension table**

```
CREATE TABLE etl_assign.DIM_LOCATION (
     location_id int not null,
     location varchar(50),
     streetname varchar(255),
     street_number int,
     zipcode int,
     lat decimal(10,3),
     lon decimal(10,3),
     primary key(location_id)
     );
```

2) **Creating atm dimension table**

```
CREATE TABLE etl_assign.DIM_ATM (
    atm_id int not null,
    atm_number VARCHAR(20),
    atm_manufacturer VARCHAR(50),
    atm_location_id INT,
    PRIMARY KEY(atm_id),
    foreign key(atm_location_id) references etl_assign.dim_location(location_id)
) ;
```

3) **Creating date dimension table**

```
CREATE TABLE etl_assign.DIM_DATE (
    date_id int not null,
    full_date_time TIMESTAMP,
    year int,
    month VARCHAR(20),
    day int,
    hour int,
    weekday varchar(20),
    primary key(date_id) );
```

4) **Creating card type dimension table**

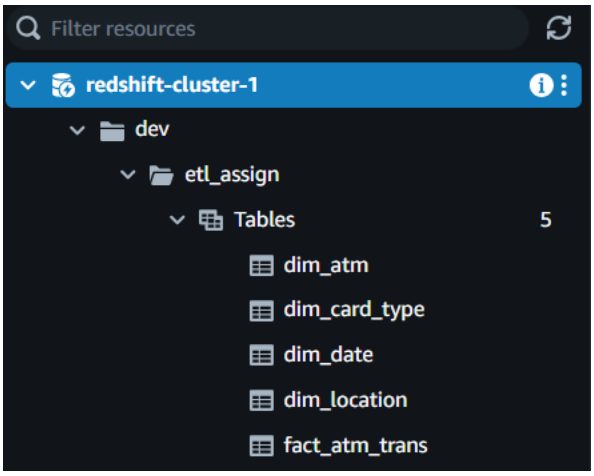
```
CREATE TABLE etl_assign.DIM_CARD_TYPE (
    card_type_id int,
    card_type varchar(30),
    primary key(card_type_id) );
```

5) **Creating atm transactions fact table**

```
CREATE TABLE etl_assign.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(255),
    message_text VARCHAR(255),
```

```
    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 etl_assign.DIM_LOCATION(location_id),
    FOREIGN KEY (atm_id) REFERENCES etl_assign.DIM_ATM(atm_id),
    FOREIGN KEY (date_id) REFERENCES etl_assign.DIM_DATE(date_id),
    FOREIGN KEY (card_type_id) REFERENCES etl_assign.DIM_CARD_TYPE(card_type_id)
);
```

Screen Shot of tables created -



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 etl_assign.dim_location from 's3://upgrad-de/dim_location/part-00000-
54a462c3-ca73-4a19-8c37-f76f1cc73397-c000.csv'
iam_role 'arn:aws:iam::355232156331:role/myRedshiftRole' IGNOREHEADER 1
delimiter ',' region 'us-east-1'
CSV;
```

2) Copying the data to dim_atm table

```
copy etl_assign.dim_atm from 's3://upgrad-de/dim_atm/part-00000-d2821730-86f6-
4e9a-8063-d505fa4cc15f-c000.csv'
iam_role 'arn:aws:iam::355232156331:role/myRedshiftRole' IGNOREHEADER 1
delimiter ',' region 'us-east-1';
```

3) Copying the data to dim_date table

```
copy etl_assign.DIM_DATE from 's3://upgrad-de/dim_date/part-00000-25ad5e65-7f39-
410e-bf94-07c56effd01d-c000.csv'
iam_role 'arn:aws:iam::355232156331:role/myRedshiftRole' IGNOREHEADER 1
delimiter ',' region 'us-east-1';
```

4) Copying the data to dim_card_type table

```
copy etl_assign.dim_card_type from 's3://upgrad-de/dim_card_type/part-00000-
2037cff7-0735-49ea-8309-6aef87a3c827-c000.csv'
iam_role 'arn:aws:iam::355232156331:role/myRedshiftRole' IGNOREHEADER 1
delimiter ',' region 'us-east-1';
```

5) Copying the data to fact_atm_trans table

```
copy etl_assign.fact_atm_trans from 's3://upgrad-de/fact_atm_trans/part-
00000-bc140ed4-e0ea-4f94-b9ed-1aebac930e5a-c000.csv'
iam_role 'arn:aws:iam::355232156331:role/myRedshiftRole' IGNOREHEADER 1
delimiter ',' region 'us-east-1'
CSV;
```