# ETL CASESTUDY – Redshift setup

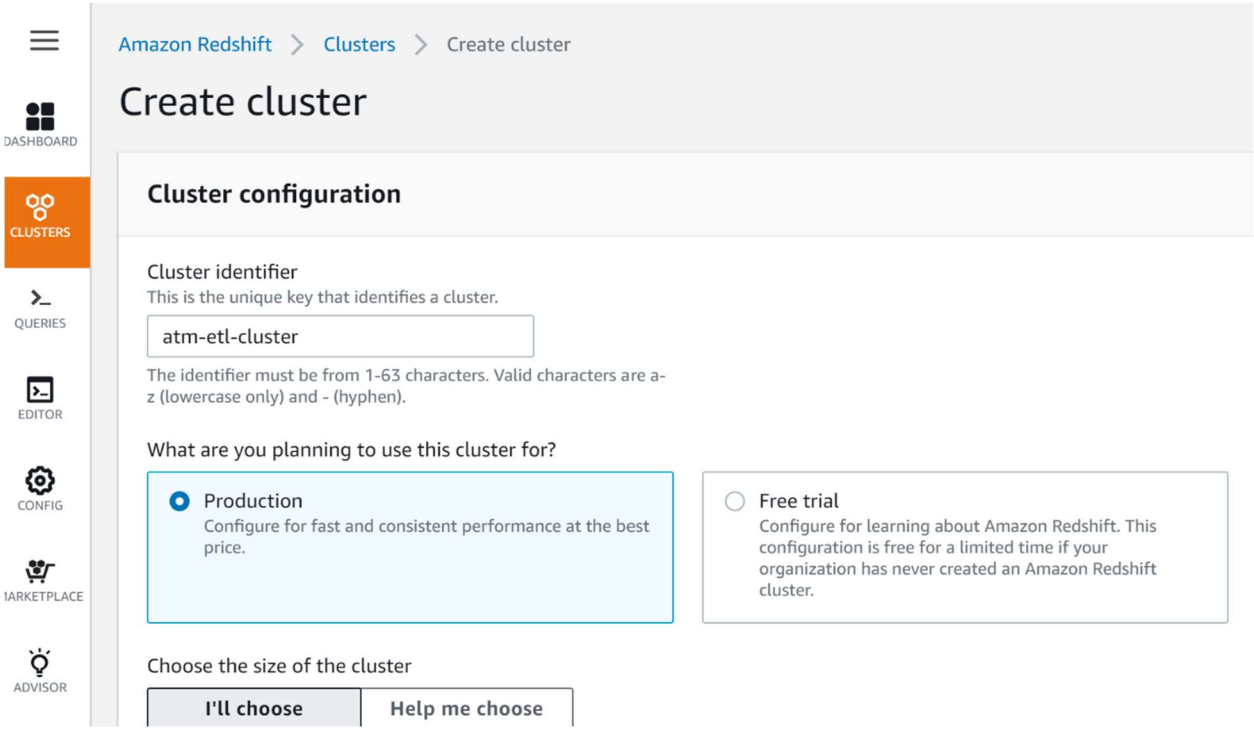
# Contents

## Amazon redshift cluster setup

As a part of the ETL casestudy data is loaded into redshift warehouse to run analytical queries and derive insights.

### Step 1 – Cluster identifier

Navigate to Redshift dashboard on AWS console and click on create cluster.
Enter cluster identifier.




### Step 2- Node type

Choose the data node type as DC2.large with 2 nodes.

Choose the size of the cluster

I'll choose | Help me choose

Node type

Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

dc2.large

Nodes

Enter the number of nodes that you need.

2

Range (1-32)

Configuration summary

dc2.large | 2 nodes

| $360.00/month | 320 GB |
| --- | --- |
| Estimated on-demand compute | Total compressed storage |

## Step 3 – Database configuration

Provide database details. Changed the default port number for security purposes.

**Database configurations**

Database name (optional)

Specify a database name to create an additional database.

atmtransdb

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a **reserved word**.

Database port (optional)

Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.

5300

The port must be numeric (1150-65535).

Master user name

Enter a login ID for the master user of your DB instance.

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a **reserved word**.

Master user password

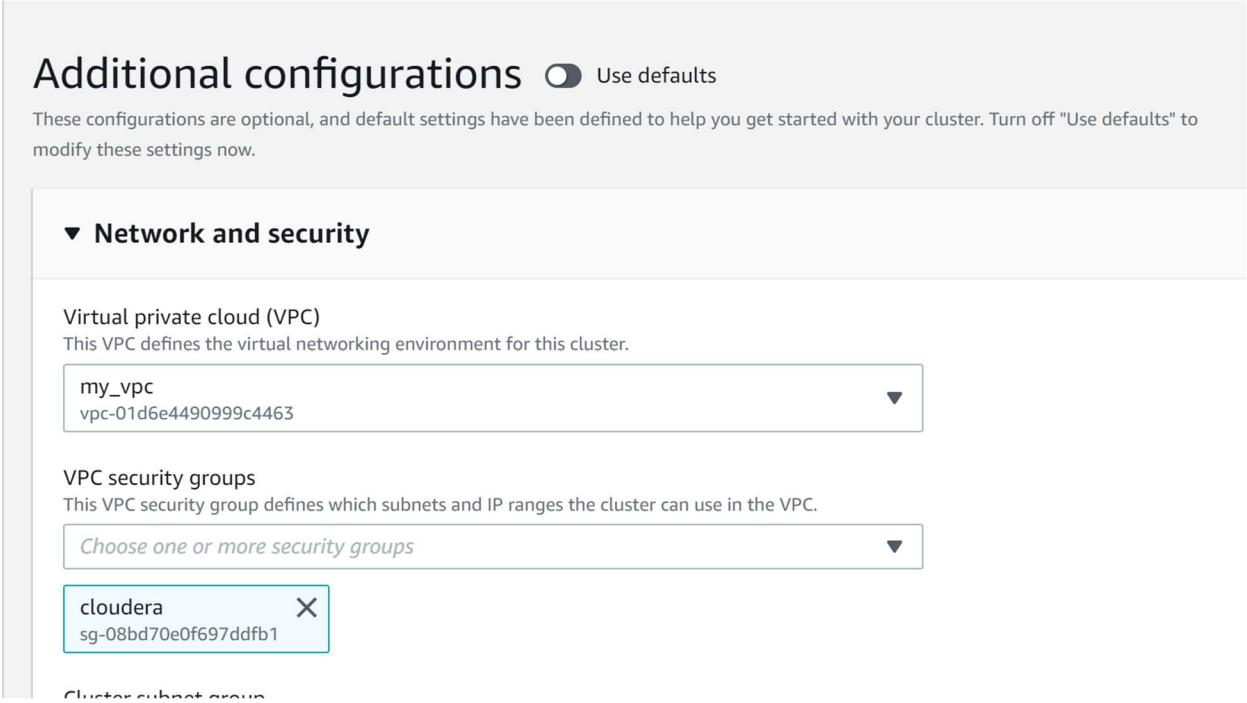
••••••••••

Show password

- The master password must be 8 - 64 characters.
- The value must contain at least one uppercase letter.
- The value must contain at least one lowercase letter.
- The value must contain at least one number.

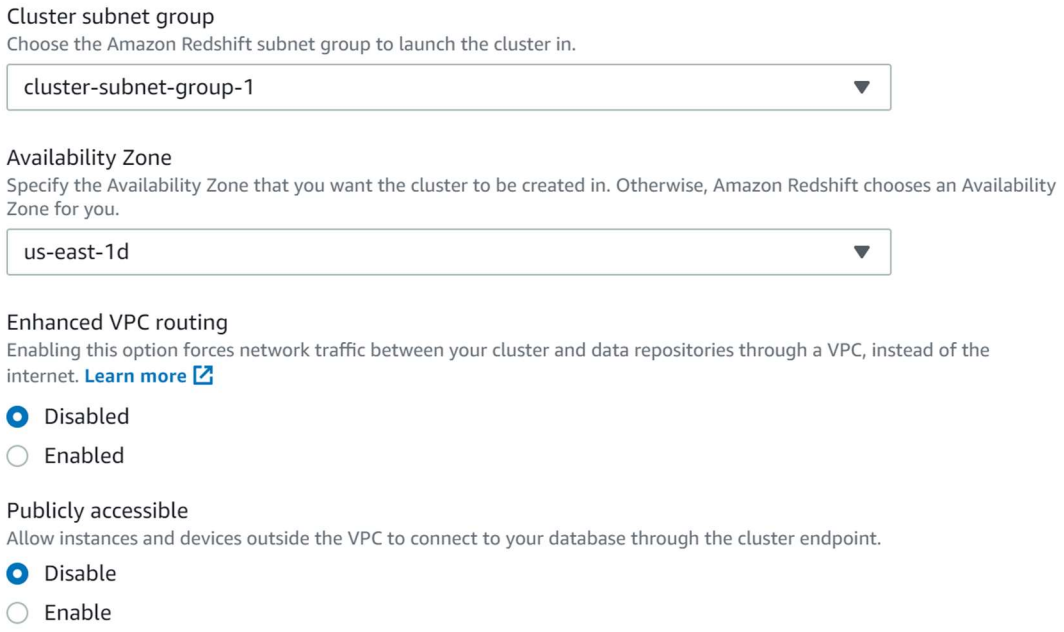
## Step 4 - VPC

In the additional configurations section provide VPC and security group settings.




## Step 5 – Availability zone

Further provide the availability zone and subnet group.

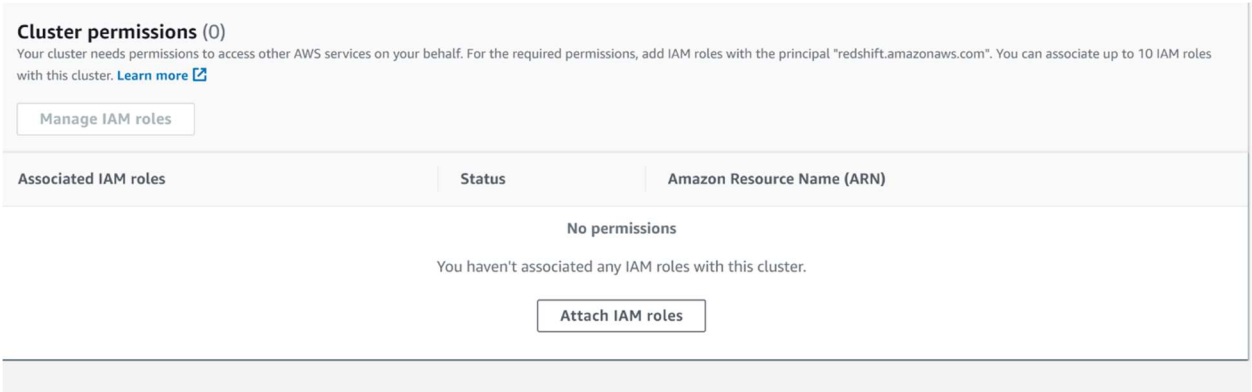


## Step 6 – Create cluster

Click on create cluster.

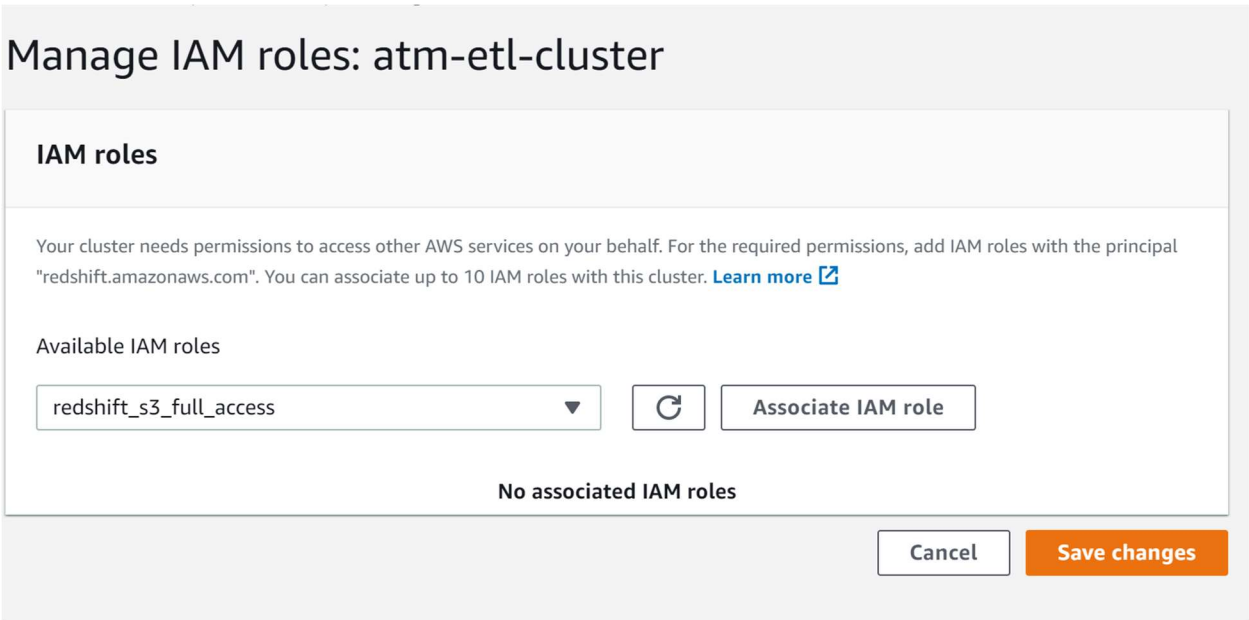

## Step 7- IAM role

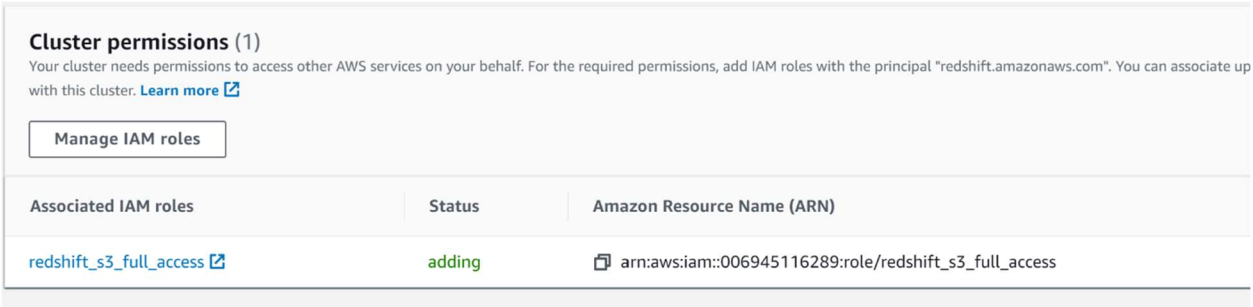
After the cluster is successfully created, associate the IAM role giving other AWS services permission to access redshift. Navigate to cluster properties tab to find Cluster permissions section, thenclick on Attach IAM roles.




## Step 8 – Associate IAM role

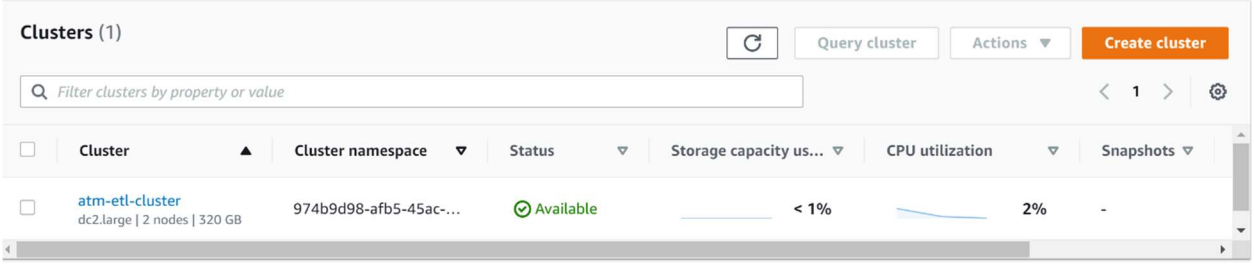
Select the IAM role and click on associate IAM role, then click on save changes.




## Step 9 – Redshift cluster

The redshift cluster is successfully created.

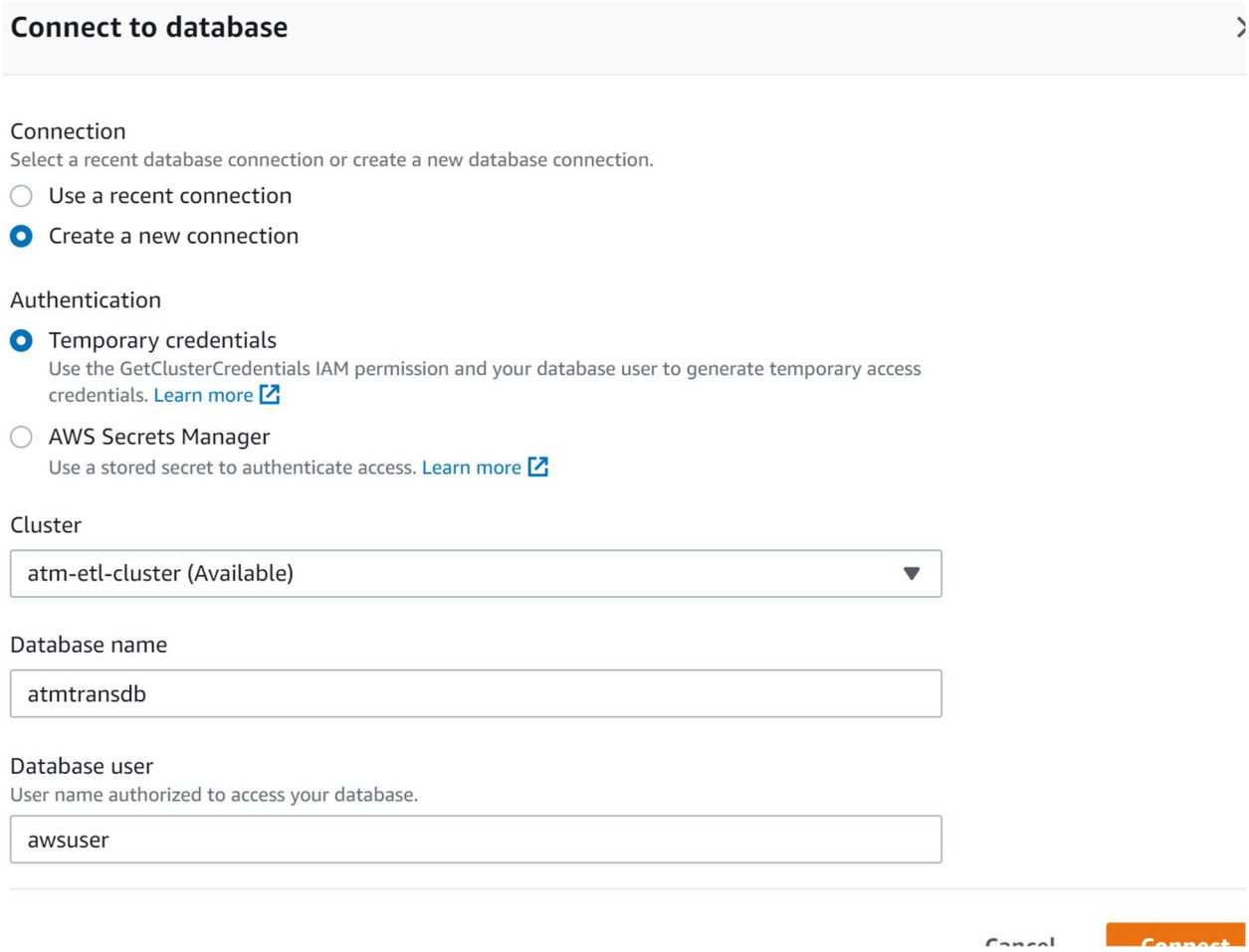


## Create Tables

The data now resides in S3 bucket. We create a schema in redshift cluster to load all the dimension and fact tables.

### Connect to database.

We need to connect to the database before we load any data.




Once connected to database we start by creating the schema.

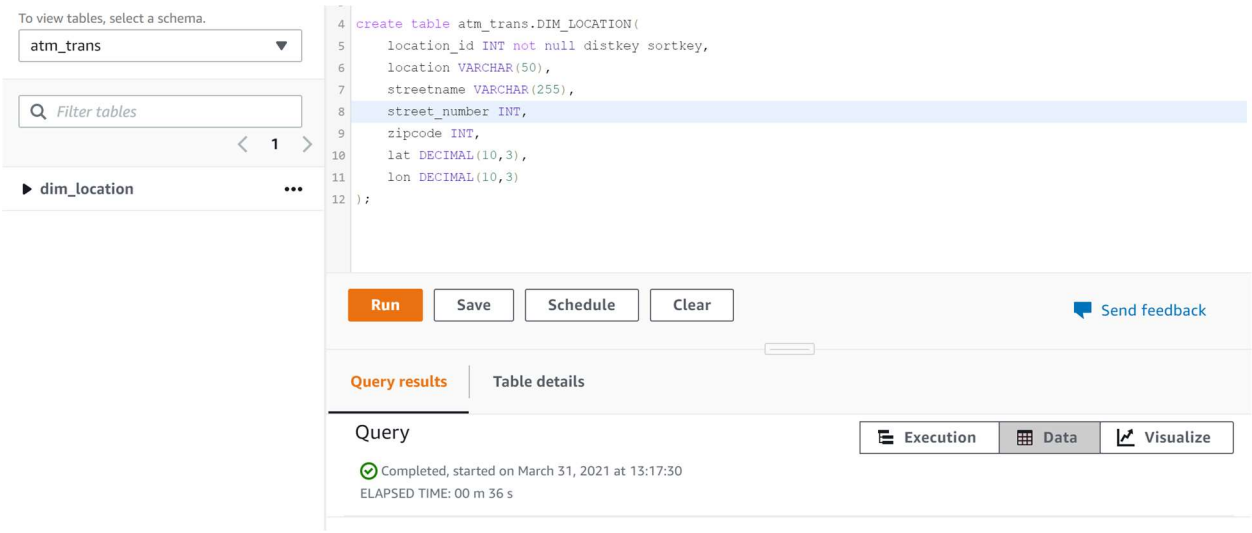
### Create schema.

*create schema atm_trans.*

### Creating location dimension

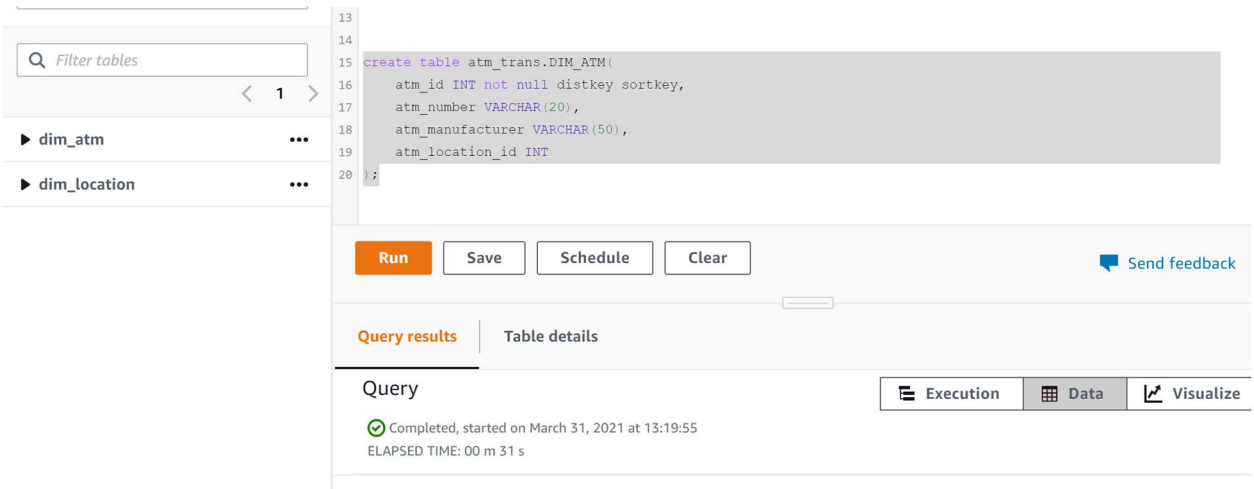
```
create table atm_trans.DIM_LOCATION(
        location_id INT not null distkey sortkey,
        location VARCHAR(50),
```

```
        streetname VARCHAR(255),
        street_number INT,
        zipcode INT,
        lat DECIMAL(10,3),
        lon DECIMAL(10,3),
        PRIMARY KEY(location_id)
);
```



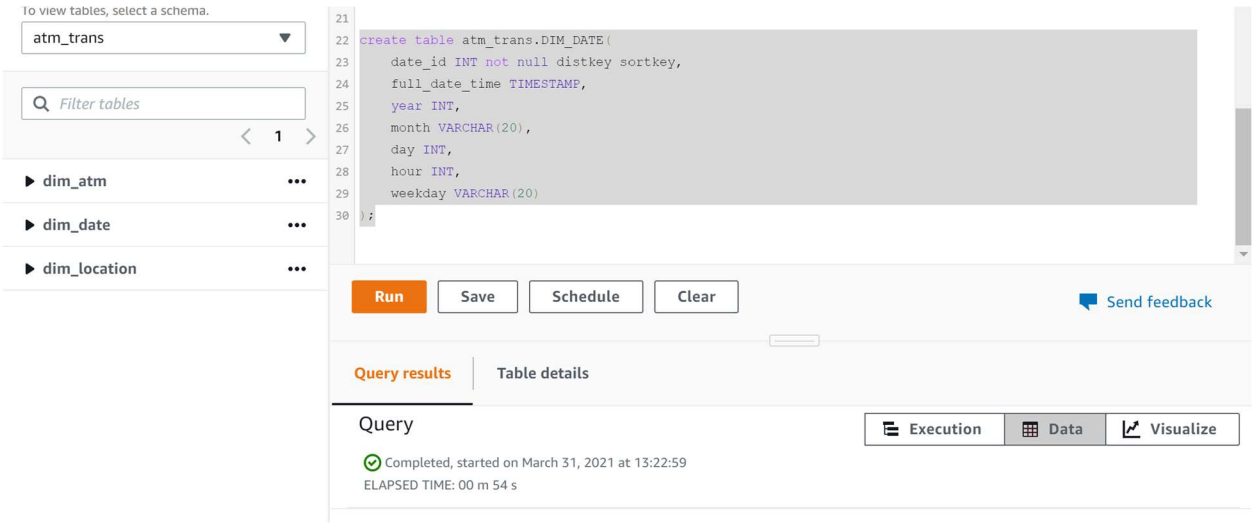

## Creating ATM dimension table

```
create table atm_trans.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_trans.DIM_LOCATION(location_id)
);
```



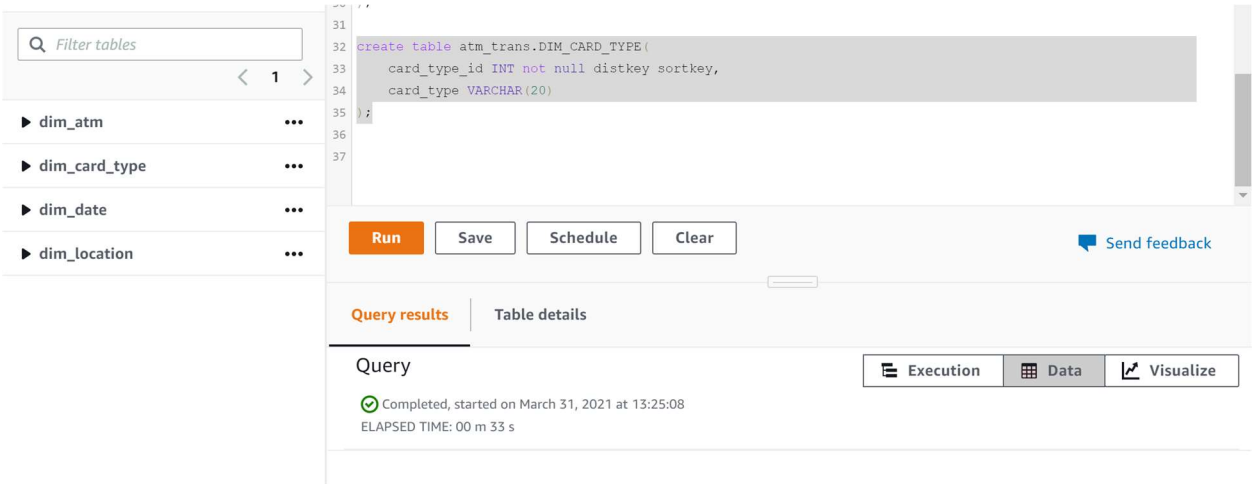

## Creating date dimension

```
create table atm_trans.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)
);
```




## Creating card type dimension

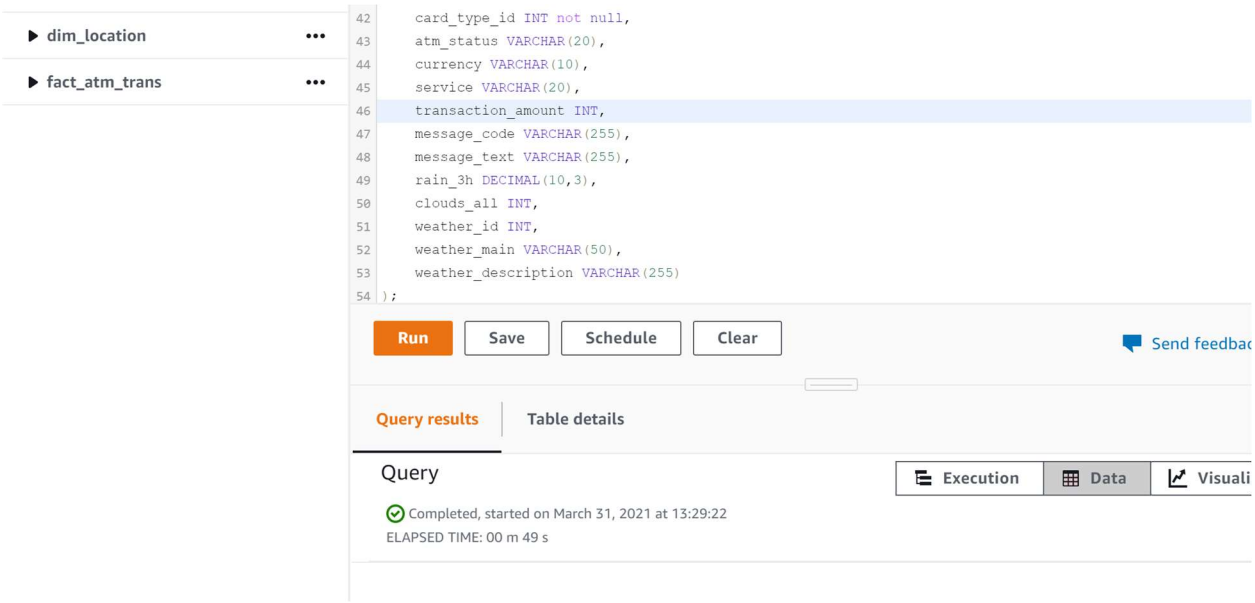
```
create table atm_trans.DIM_CARD_TYPE(
        card_type_id INT not null distkey sortkey,
        card_type VARCHAR(20),
        PRIMARY KEY(card_type_id)
);
```



Creating Fact table

```
create table atm_trans.FACT_ATM_TRANS(
        trans_id BIGINT not null distkey sortkey,
        atm_id INT not null,
         weather_loc_id INT not null,
         date_id INT not null,
        card_type_id INT not null,
        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
        atm_trans.DIM_LOCATION(location_id),
        FOREIGN KEY(atm_id) REFERENCES atm_trans.DIM_ATM(atm_id),
        FOREIGN KEY(date_id) REFERENCES atm_trans.DIM_DATE(date_id),
        FOREIGN KEY(card_type_id) REFERENCES atm_trans.DIM_CARD_TYPE
        (card_type_id)
);
```



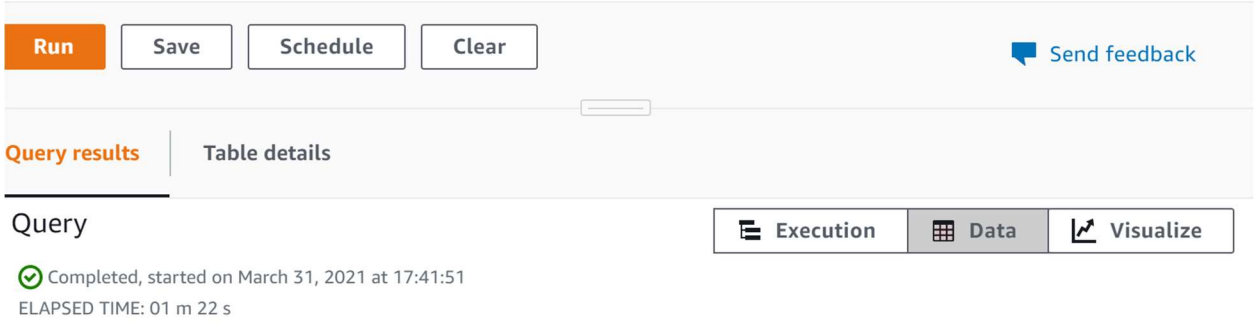

## Load Data

Once all the tables are created successfully, we can copy the data from S3 bucket into redshift.

### Load data to Location dimension

```
copy atm_trans.DIM_LOCATION from
's3://atmtransetl/atm_trans/location/part-00000-ce21b2a1-bac6-4264-
96f5-acb2e0b66713-c000.csv'
iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access'
delimiter ',' region  'us-east-1' IGNOREHEADER 1;
```

```
copy atm_trans.DIM_LOCATION from
's3://atmtransetl/atm_trans/location/part-00000-ce21b2a1-bac6-4264-96f5-acb2e0b66713-c000.csv'
iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```




## Load data to ATM dimension

*copy atm_trans.DIM_ATM from 's3://atmtransetl/atm_trans/atm/part-00000-ed42b98c-7830-4575-9af1-313a5265fc8a-c000.csv' iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access' delimiter ',' region 'us-east-1' IGNOREHEADER 1;*

```
copy atm_trans.DIM_ATM from
's3://atmtransetl/atm_trans/atm/part-00000-ed42b98c-7830-4575-9af1-313a5265fc8a-c000.csv'
iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' IGNOREHEADER 1;
```

Run | Save | Schedule | Clear | Send feedback

Query results | Table details

Query

Execution | Data | Visualize

Completed, started on March 31, 2021 at 17:44:49

ELAPSED TIME: 00 m 19 s

## Load data into date dimension

*copy atm_trans.DIM_DATE from 's3://atmtransetl/atm_trans/date/part-00000-173910bb-4eee-4d8f-8783-8898367fe347-c000.csv' iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access' delimiter ',' region 'us-east-1' timeformat 'auto' IGNOREHEADER 1;*

```
copy atm_trans.DIM_DATE from
's3://atmtransetl/atm_trans/date/part-00000-173910bb-4eee-4d8f-8783-8898367fe347-c000.csv'
iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1' timeformat 'auto' IGNOREHEADER 1;
```

Run | Save | Schedule | Clear | Send feedback

Query results | Table details

Query

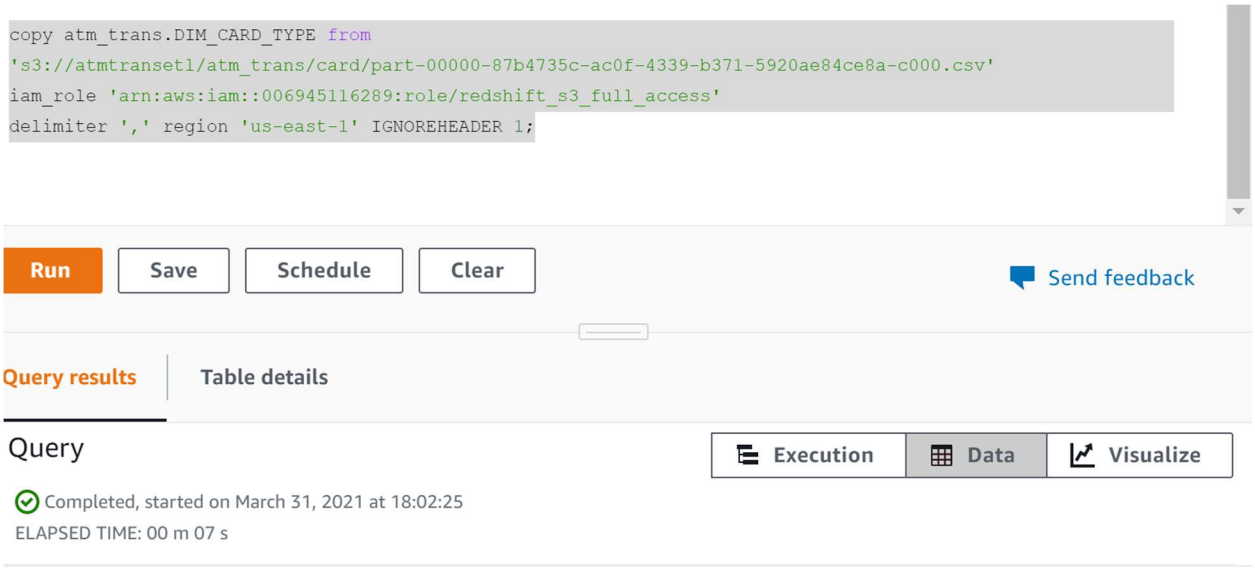
Execution | Data | Visualize

Completed, started on March 31, 2021 at 17:55:22

ELAPSED TIME: 00 m 13 s

## Load data into card type dimension

*copy atm_trans.DIM_CARD_TYPE from 's3://atmtransetl/atm_trans/card/part-00000-87b4735c-ac0f-4339-b371-5920ae84ce8a-c000.csv' iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access' delimiter ',' region 'us-east-1' IGNOREHEADER 1;*




## Load data into atm transaction fact table

*copy atm_trans.FACT_ATM_TRANS from 's3://atmtransetl/atm_trans/trans/part-00000-3ce97df6-9712-44ac-8fe6-908bfd6f8753-c000.csv' iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access' delimiter ',' region 'us-east-1'  IGNOREHEADER 1;*

```
copy atm_trans.FACT_ATM_TRANS from
's3://atmtransetl/atm_trans/trans/part-00000-3ce97df6-9712-44ac-8fe6-908bfd6f8753-c000.csv'
iam_role 'arn:aws:iam::006945116289:role/redshift_s3_full_access'
delimiter ',' region 'us-east-1'  IGNOREHEADER 1;
```

Run | Save | Schedule | Clear | Send feedback

Query results | Table details

Query

Execution | Data | Visualize

Completed, started on March 31, 2021 at 18:05:06

ELAPSED TIME: 00 m 24 s