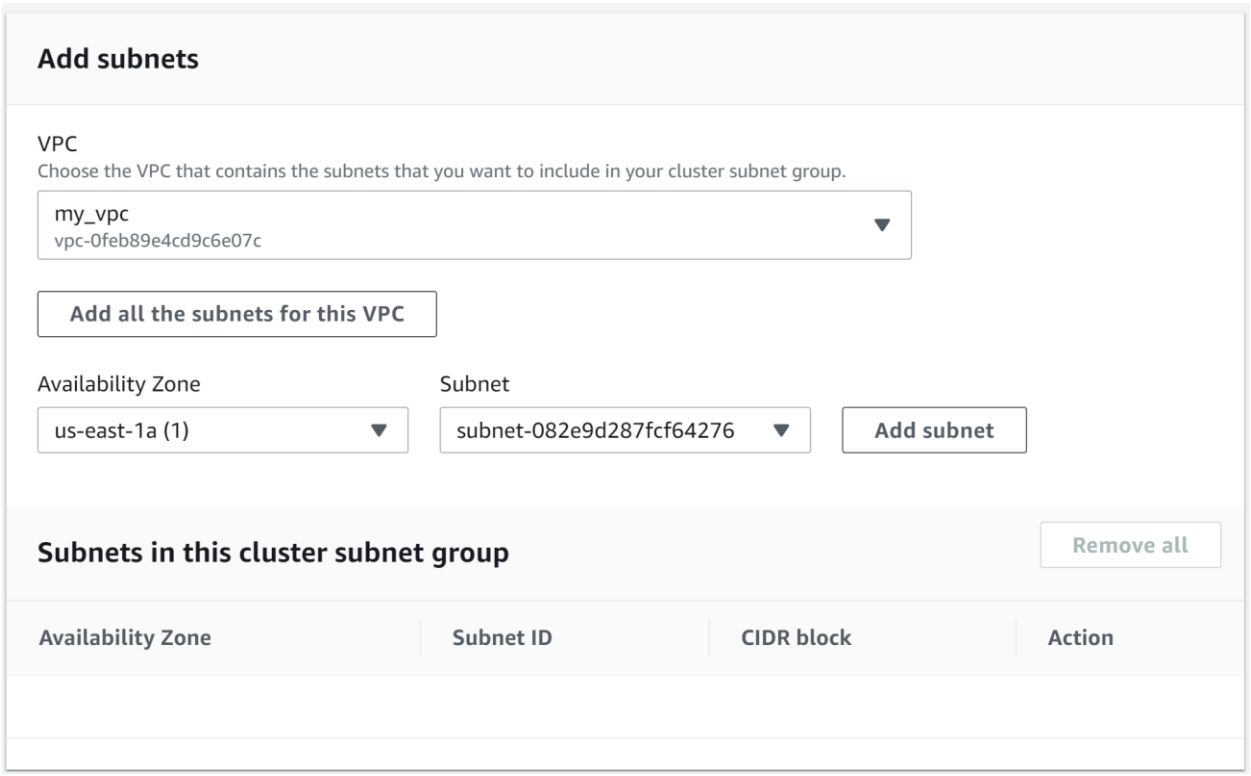
Creation of a RedShift Cluster

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 > Configurations > Subnet groups > Create subnet group

Create cluster subnet group

Cluster subnet group details

Name
You can't modify the name after your subnet group has been created.

cluster-subnet-group-1

The name must be 1-255 characters. Valid characters are A-Z, a-z, 0-9, space, hyphen(-), underscore (_), and period (.).

Description

upgrad redshift cluster

Add subnets

VPC
Choose the VPC that contains the subnets that you want to include in your cluster subnet group.

my_vpc
vpc-0e837d07bf3da9ce9

Add all the subnets for this VPC

Availability Zone

us-east-1e (1)

Subnet

subnet-0fe761786b22381b7

Add subnet

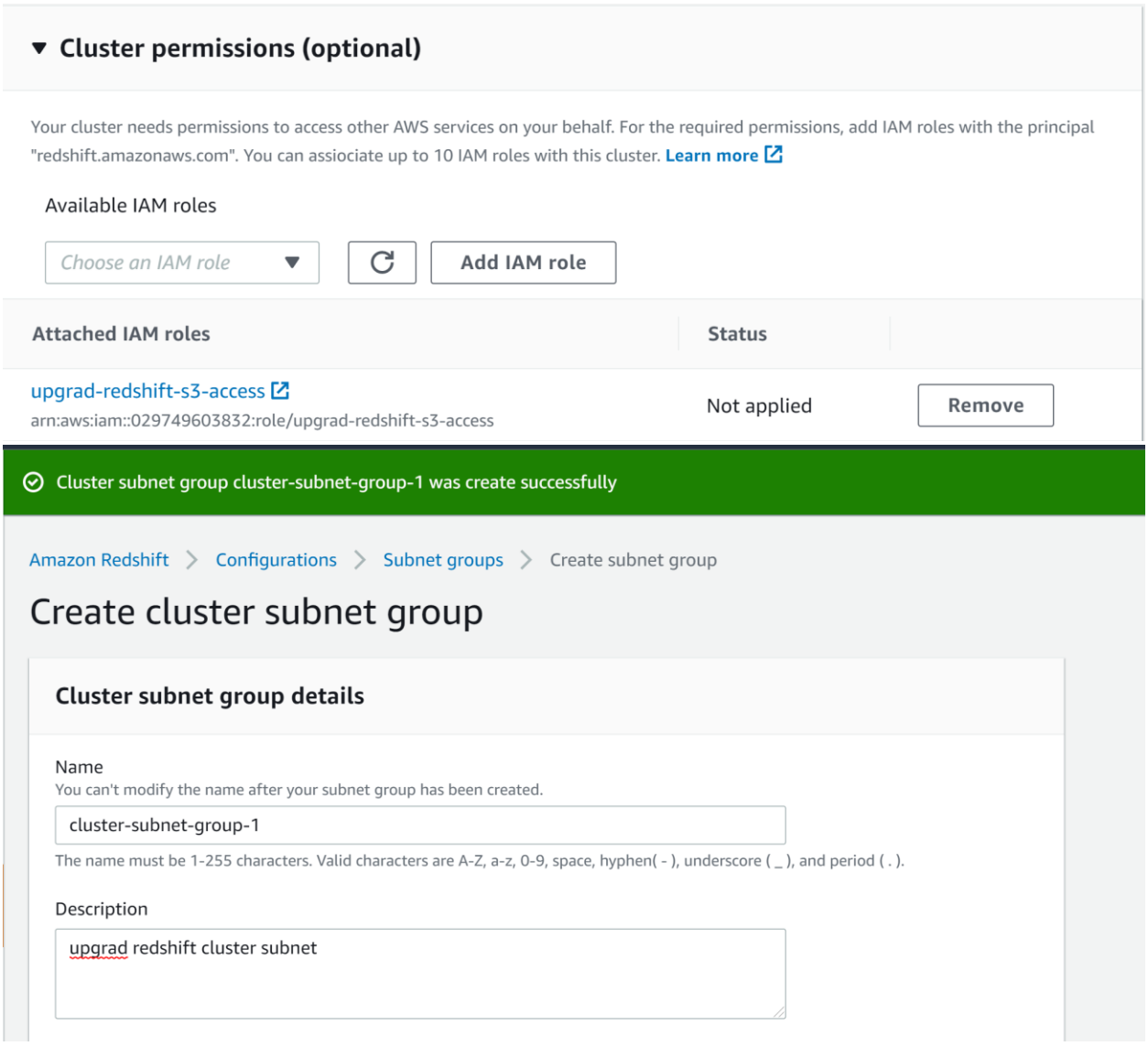
Subnets in this cluster subnet group

Remove all

Availability Zone	Subnet ID	CIDR block	Action
us-east-1e	subnet-0fe761786b22381b7	10.0.0.0/24	Remove

Cancel

Create cluster subnet group



Cluster configuration

Cluster identifier

This is the unique key that identifies a cluster.

redshift-cluster-etl-proj

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]

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)

Database configurations

Database name (optional)
Specify a database name to create an additional database.

etlprojdb

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.

5439

The port must be numeric (1150-65535).

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

etlproj

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

Master user password

Etlproj1

☑ 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.
- The master password can only contain ASCII characters (ASCII codes 33-126), except ' (single quotation mark), " (double quotation mark), /, \, or @.

Network and security

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster. Choose a VPC that has a subnet group. Only valid VPCs are enabled in the list.

my_vpc
vpc-0e837d07bf3da9ce9

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

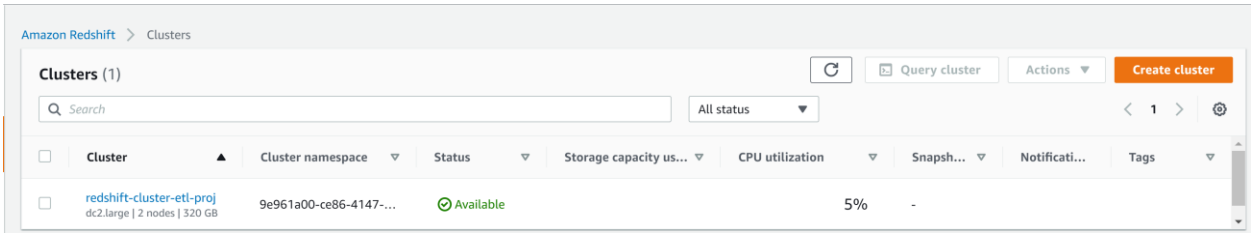
default
sg-0a08b9c056c03a39c

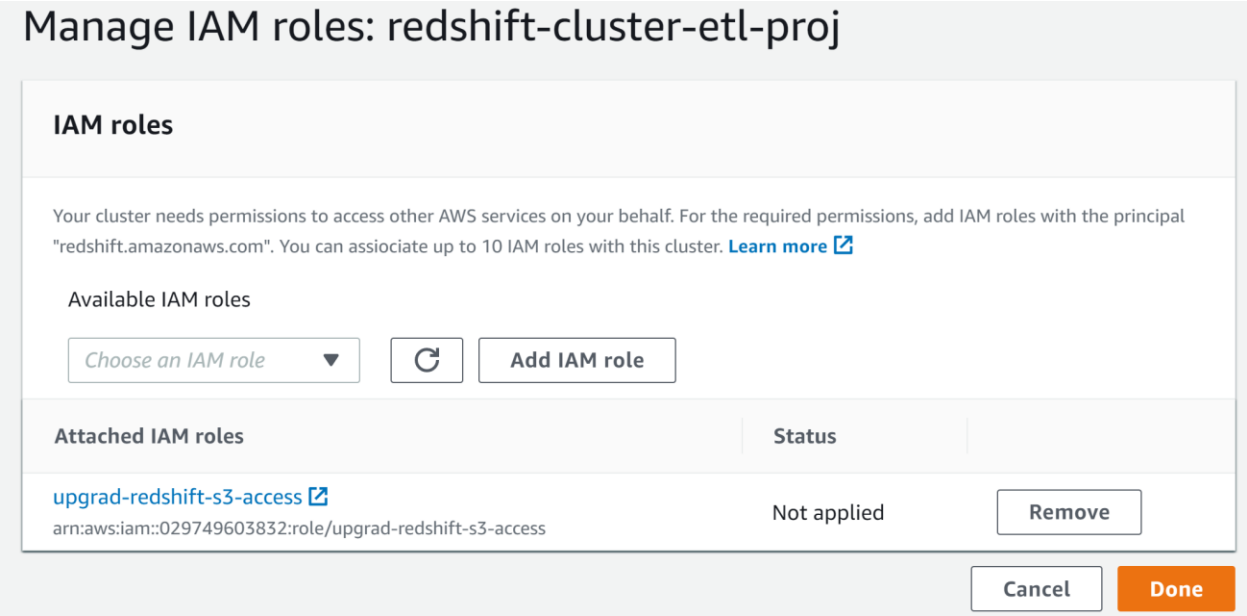
Cluster subnet group
Choose the Amazon Redshift subnet group to launch the cluster in.

cluster-subnet-group-1

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



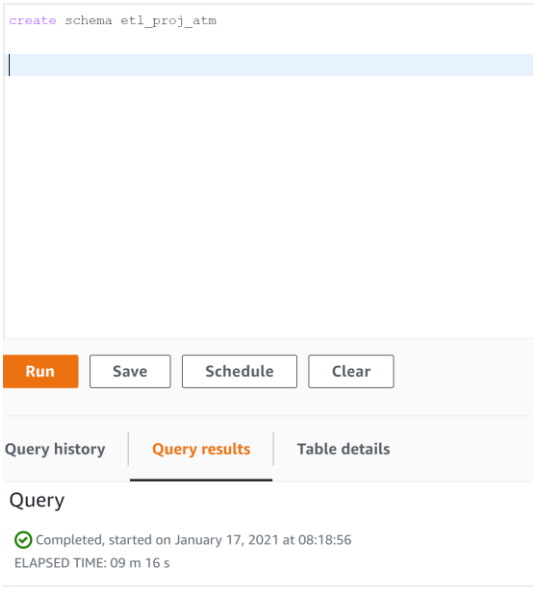


Setting up a database in the RedShift cluster and running queries to create the dimension and fact tables

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

Create schema:

Query: create schema etl_proj_atm



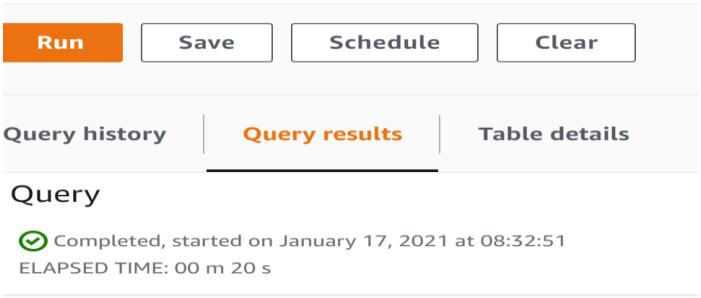
Create Dimension tables:

1) Dim_location:

Query :

```
create table etl_proj_atm.DIM_LOCATION
(location_id integer not null,
location VARCHAR(50),
streetname VARCHAR(255),
street_number integer,
zipcode integer,
lat DECIMAL(10,3),
lon DECIMAL(10,3),
PRIMARY KEY (location_id))
```

```
create table etl_proj_atm.DIM_LOCATION
(location_id integer not null,
location VARCHAR(50),
streetname VARCHAR(255),
street_number integer,
zipcode integer,
lat DECIMAL(10,3),
lon DECIMAL(10,3),
PRIMARY KEY (location_id))
```

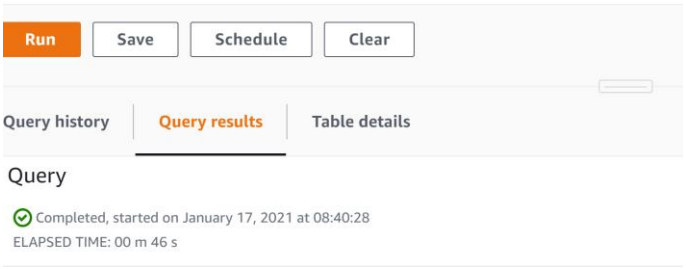


2) DIM_ATM

Query:

```
create table etl_proj_atm.DIM_ATM
(atm_id integer not null,
atm_number VARCHAR(20),
atm_manufacturer VARCHAR(50),
atm_location_id integer,
PRIMARY KEY (atm_id),
FOREIGN KEY (atm_location_id) REFERENCES etl_proj_atm.dim_location(location_id)
)
```

```
create table etl_proj_atm.DIM_ATM
(atm_id integer not null,
atm_number VARCHAR(20),
atm_manufacturer VARCHAR(50),
atm_location_id integer,
PRIMARY KEY (atm_id),
FOREIGN KEY (atm_location_id) REFERENCES etl_proj_atm.dim_location(location_id)
)
```

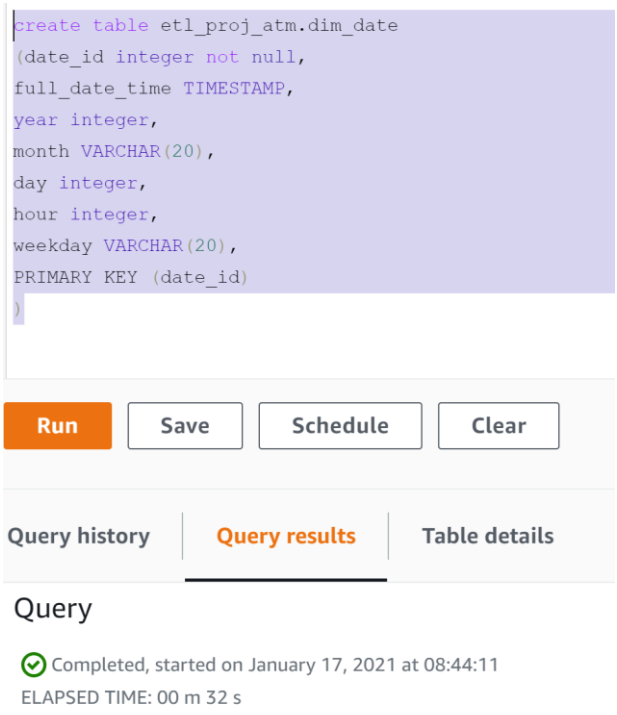


3) DIM_DATE:

Query:

```
create table etl_proj_atm.dim_date
(date_id integer not null,
full_date_time TIMESTAMP,
year integer,
month VARCHAR(20),
day integer,
hour integer,
```

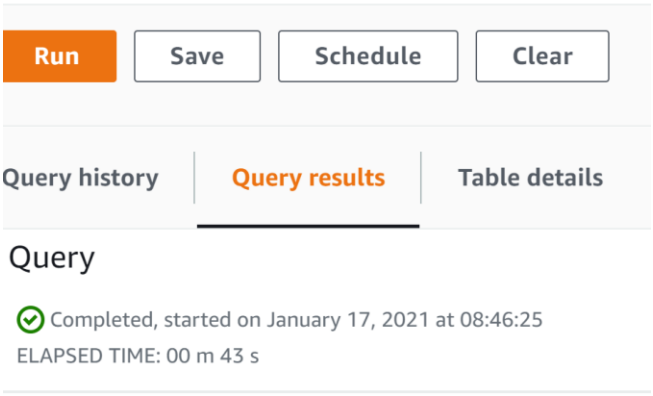
weekday VARCHAR(20),
PRIMARY KEY (date_id)
)



4) DIM_CARD_TYPE:

Qyery:
create table etl_proj_atm.dim_card_type
(card_type_id integer not null,
card_type VARCHAR(20),
PRIMARY KEY (card_type_id)
)

```
create table etl_proj_atm.dim_card_type
(card_type_id integer not null,
card_type VARCHAR(20),
PRIMARY KEY (card_type_id)
)
```



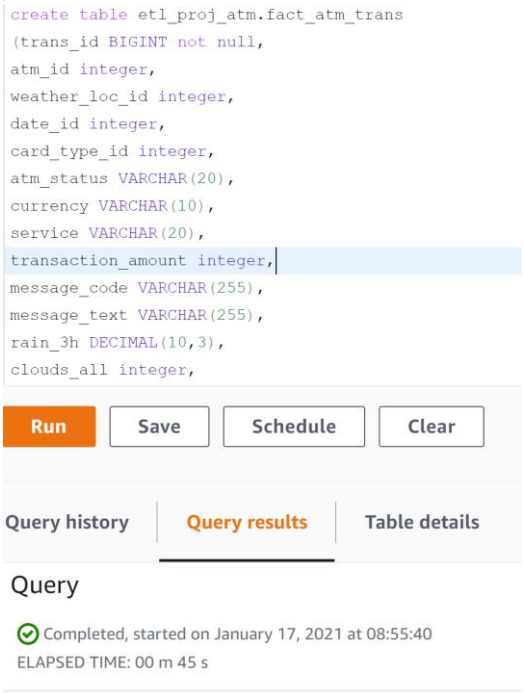
Note: While loading data into this table, was getting an error where data was larger than lenghth varchar(20) from column card_type. To accommodate this, table def was altered and lenth of card type was changed to varchar(30) as below:

ALTER TABLE etl_proj_atm.dim_card_type ALTER COLUMN card_type type varchar(30)

5) **FACT_ATM_TRANS**

Query: create table etl_proj_atm.fact_atm_trans
(trans_id BIGINT not null,
atm_id integer,
weather_loc_id integer,
date_id integer,
card_type_id integer,
atm_status VARCHAR(20),
currency VARCHAR(10),

service VARCHAR(20),
transaction_amount integer,
message_code VARCHAR(255),
message_text VARCHAR(255),
rain_3h DECIMAL(10,3),
clouds_all integer,
weather_id integer,
weather_main VARCHAR(50),
weather_description VARCHAR(255),
PRIMARY KEY (trans_id),
FOREIGN KEY (weather_loc_id) REFERENCES etl_proj_atm.dim_location(location_id),
FOREIGN KEY (atm_id) REFERENCES etl_proj_atm.dim_atm(atm_id),
FOREIGN KEY (date_id) REFERENCES etl_proj_atm.dim_date(date_id),
FOREIGN KEY (card_type_id) REFERENCES etl_proj_atm.dim_card_type(card_type_id)
)



Loading data into a RedShift cluster from Amazon S3 bucket

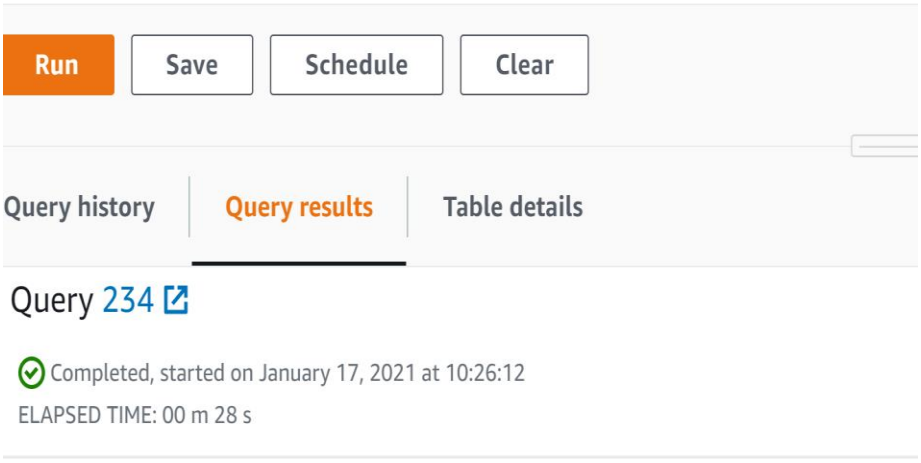
Queries to copy the data from S3 buckets to the RedShift cluster in the appropriate tables

1) **copy etl_proj_atm.dim_location**

FROM 's3://pshrutis3bucket/etlproj/dim_loc.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv

ignoreheader 1
region 'us-east-1' ;

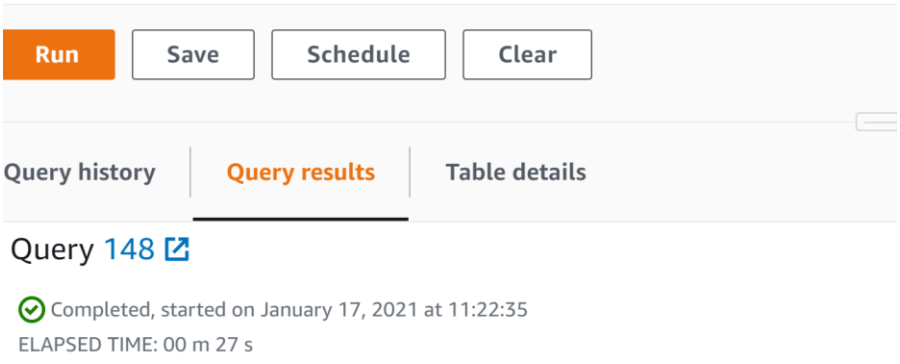
```
copy etl_proj_atm.dim_location
FROM 's3://pshrutis3bucket/etlproj/dim_loc.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```



2)
copy etl_proj_atm.dim_atm
FROM 's3://pshrutis3bucket/etlproj/dim_atm.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv

ignoreheader 1
region 'us-east-1' ;

```
copy etl_proj_atm.dim_atm
FROM 's3://pshrutis3bucket/etlproj/dim_atm.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```

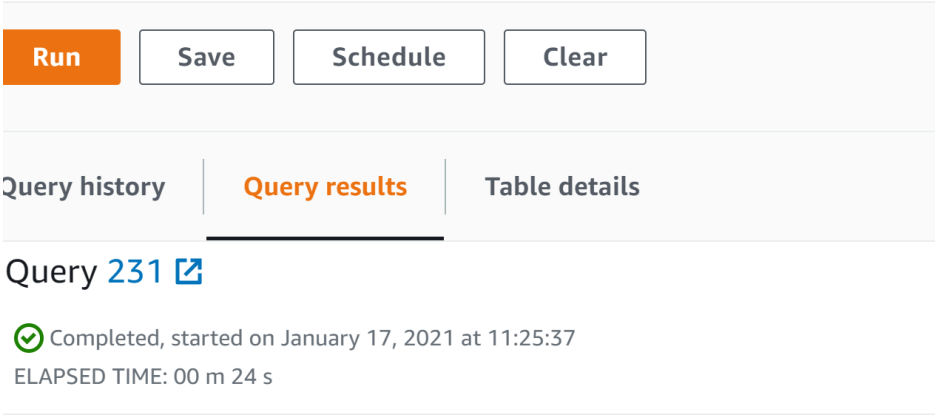


3)

copy etl_proj_atm.dim_date
FROM 's3://pshrutis3bucket/etlproj/dim_date.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'

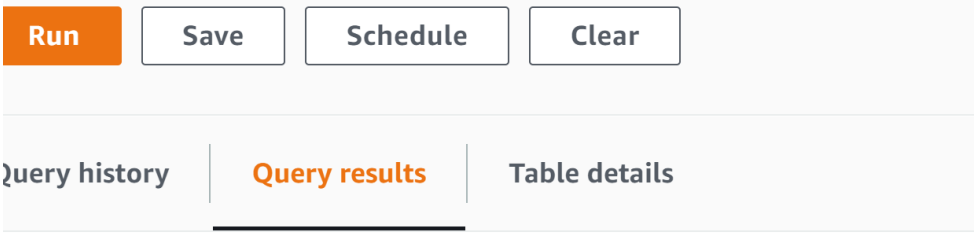
csv
ignoreheader 1
region 'us-east-1' ;

```
copy etl_proj_atm.dim_date
FROM 's3://pshrutis3bucket/etlproj/dim_date.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```



4)
copy etl_proj_atm.dim_card_type
FROM 's3://pshrutis3bucket/etlproj/dim_card_type.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;

```
copy etl_proj_atm.dim_card_type
FROM 's3://pshrutis3bucket/etlproj/dim_card_type.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```



Query 36

Completed, started on January 18, 2021 at 12:01:23
ELAPSED TIME: 00 m 13 s

5)

```
copy etl_proj_atm.fact_atm_trans
FROM 's3://pshrutis3bucket/etlproj/fact_dim_load.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```

```
copy etl_proj_atm.fact_atm_trans
FROM 's3://pshrutis3bucket/etlproj/fact_dim_load.csv'
iam_role 'arn:aws:iam::029749603832:role/upgrad-redshift-s3-access'
csv
ignoreheader 1
region 'us-east-1' ;
```