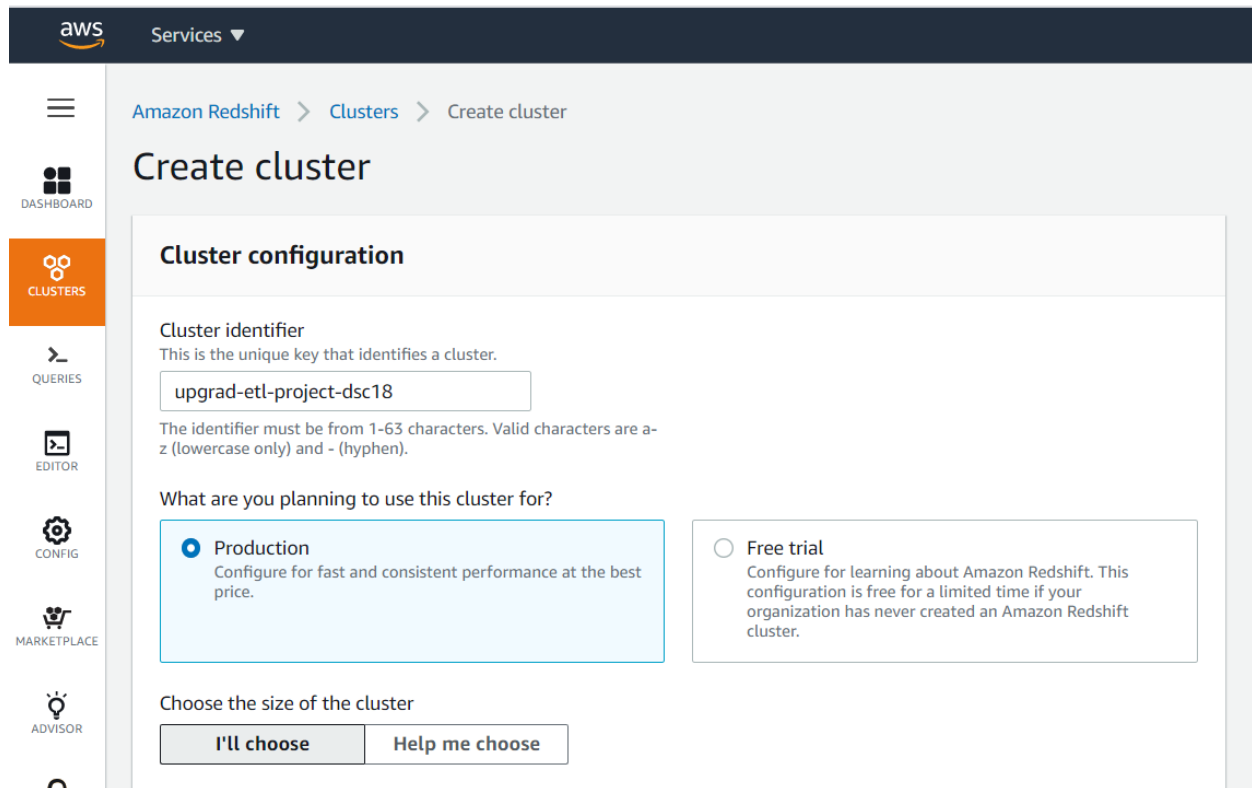


Creation of a RedShift Cluster

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



aws Services ▾

Amazon Redshift > Clusters > Create cluster

Create cluster

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

upgrad-etl-project-dsc18

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

ALARMS

EVENTS

WHAT'S NEW

Services ▼

EC2 Instance Types

Help me choose

Node type

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

Recommended

RA3


High performance with scalable managed storage

☐

ra3.4xlarge\$3.26/node/hour
Managed storage:
up to 64 TB/node\$0.024/GB/month

☐

ra3.16xlarge\$13.04/node/hour
Managed storage:
up to 64 TB/node\$0.024/GB/month



ra3.4xlarge
12 vCPU (gen 3)

DC2

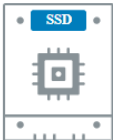
High performance with fixed local SSD storage

☒

dc2.large\$0.25/node/hour
Storage: 160 GB/node

☐

dc2.8xlarge\$4.80/node/hour
Storage: 2.6 TB/node



dc2.large
2 vCPU (gen 2)

Show legacy dense storage node types

Nodes

Enter the number of nodes that you need.

2

Range (1-32)

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

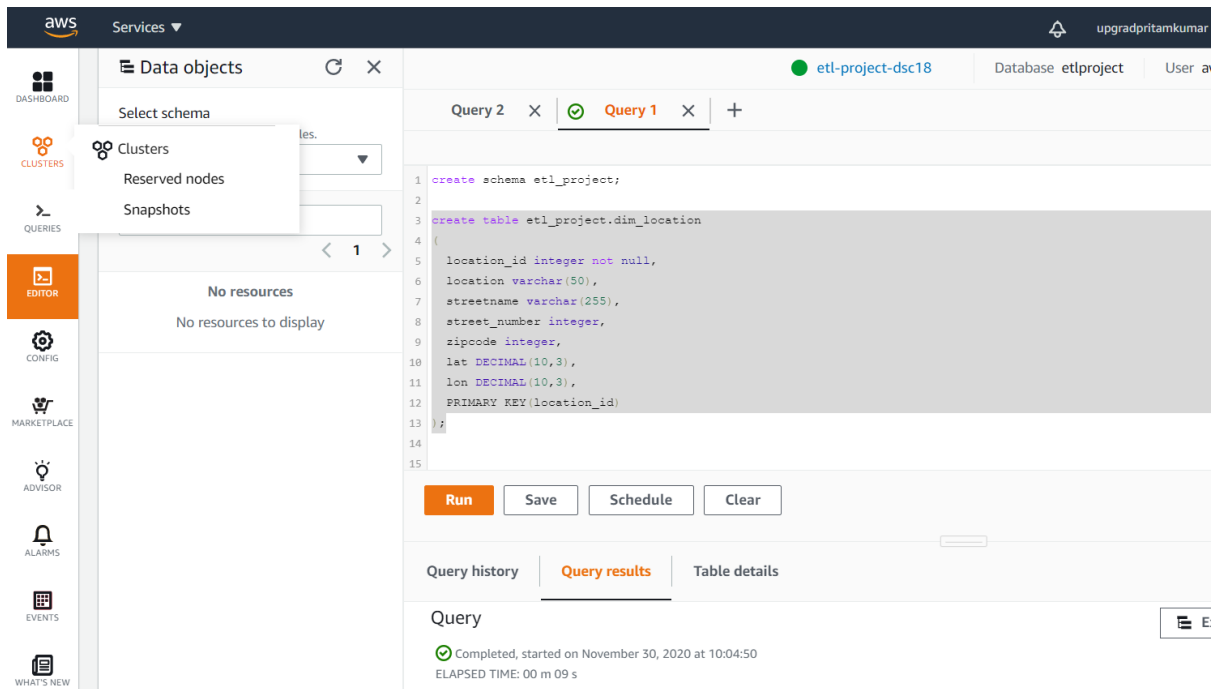
Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

1) **Schema creation** → create schema etl_project;

2) **Dim_Location Table creation** →

create table etl_project.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)
);
```

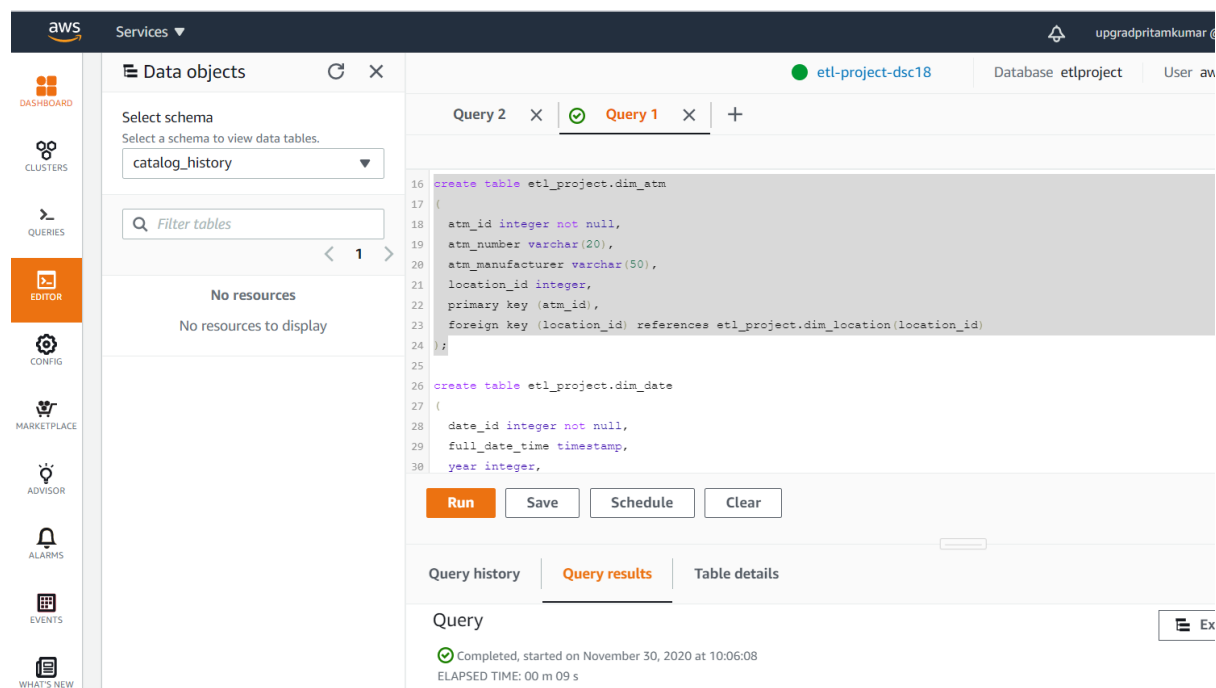


The screenshot shows the AWS Data console interface. On the left, there is a navigation menu with options like DASHBOARD, CLUSTERS, QUERIES, EDITOR, CONFIG, MARKETPLACE, ADVISOR, ALARMS, EVENTS, and WHAT'S NEW. The main area displays the 'Data objects' section for the 'etl-project-dsc18' database. A query editor is open, showing the following SQL code:

```
1 create schema etl_project;
2
3 create table etl_project.dim_location
4 (
5   location_id integer not null,
6   location varchar(50),
7   streetname varchar(255),
8   street_number integer,
9   zipcode integer,
10  lat DECIMAL(10,3),
11  lon DECIMAL(10,3),
12  PRIMARY KEY(location_id)
13 );
14
15
```

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 November 30, 2020 at 10:04:50. ELAPSED TIME: 00 m 09 s'.

3) DIM_ATM table creation →
 create table etl_project.dim_atm
 (
 atm_id integer not null,
 atm_number varchar(20),
 atm_manufacturer varchar(50),
 location_id integer,
 primary key (atm_id),
 foreign key (location_id) references etl_project.dim_location(location_id)
);



The screenshot shows the AWS Glue console interface. On the left, there is a sidebar with navigation options: DASHBOARD, CLUSTERS, QUERIES, EDITOR (highlighted), CONFIG, MARKETPLACE, ADVISOR, ALARMS, EVENTS, and WHAT'S NEW. The main area is titled 'Data objects' and shows a 'Select schema' dropdown set to 'catalog_history'. Below this, it says 'No resources' and 'No resources to display'. On the right, there is a 'Query 1' editor with the following SQL code:

```

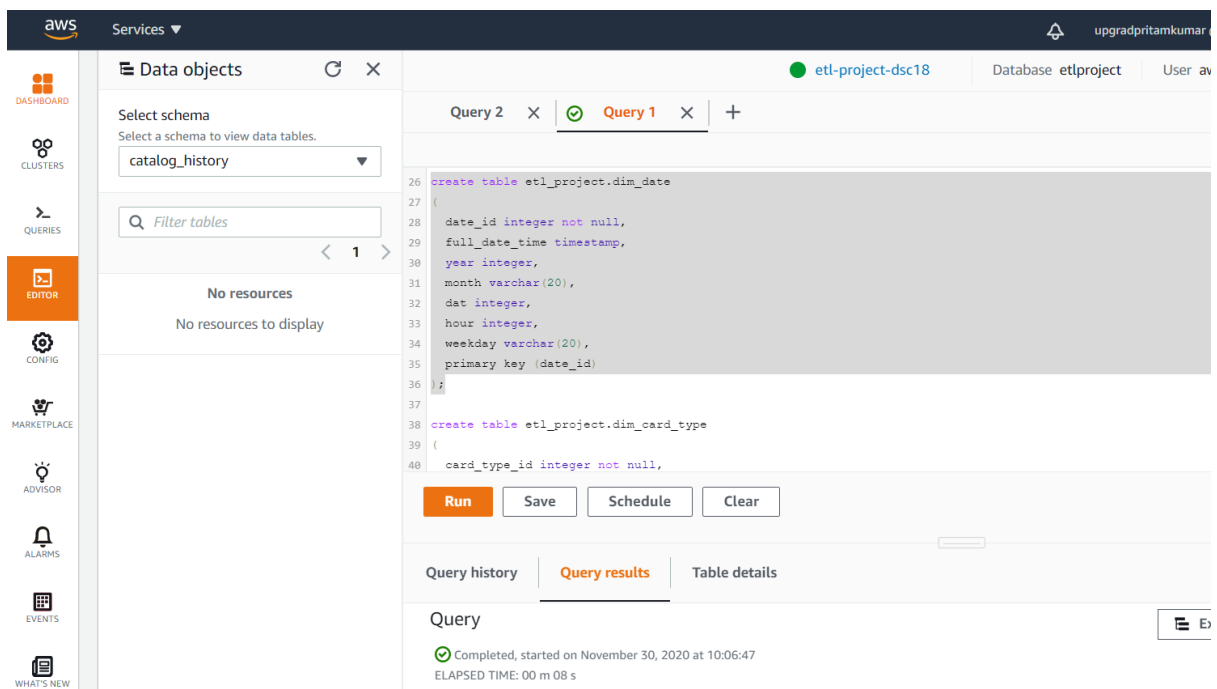
16 create table etl_project.dim_atm
17 (
18   atm_id integer not null,
19   atm_number varchar(20),
20   atm_manufacturer varchar(50),
21   location_id integer,
22   primary key (atm_id),
23   foreign key (location_id) references etl_project.dim_location(location_id)
24 );
25
26 create table etl_project.dim_date
27 (
28   date_id integer not null,
29   full_date_time timestamp,
30   year integer,
  
```

Below the code editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there are tabs for 'Query history', 'Query results' (selected), and 'Table details'. The 'Query results' tab shows a status of 'Completed, started on November 30, 2020 at 10:06:08' and 'ELAPSED TIME: 00 m 09 s'.

4) DIM_DATE table creation →

create table etl_project.dim_date

(
 date_id integer not null,
 full_date_time timestamp,
 year integer,
 month varchar(20),
 dat integer,
 hour integer,
 weekday varchar(20),
 primary key (date_id)
);



The screenshot shows the AWS Glue console interface. On the left is a navigation sidebar with options like DASHBOARD, CLUSTERS, QUERIES, EDITOR, CONFIG, MARKETPLACE, ADVISOR, ALARMS, EVENTS, and WHAT'S NEW. The main area is titled 'Data objects' and shows a 'Select schema' dropdown set to 'catalog_history'. Below this, it says 'No resources' and 'No resources to display'. On the right, a SQL query editor is open for 'Query 1' in the 'etl-project-dsc18' database. The query contains two CREATE TABLE statements. The first creates 'etl_project.dim_date' with columns: date_id (integer not null), full_date_time (timestamp), year (integer), month (varchar(20)), dat (integer), hour (integer), weekday (varchar(20)), and a primary key on date_id. The second creates 'etl_project.dim_card_type' with column: card_type_id (integer not null). Below the editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, a 'Query history' section shows the query as 'Completed, started on November 30, 2020 at 10:06:47' with an 'ELAPSED TIME: 00 m 08 s'.

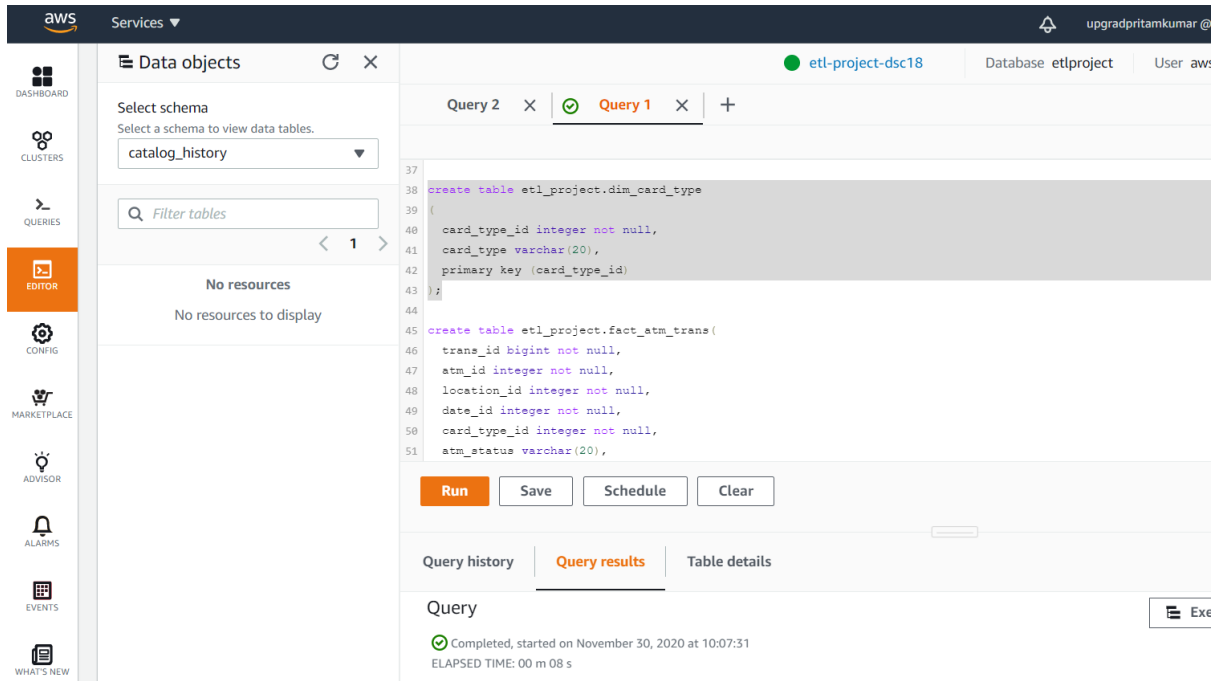
```

26 create table etl_project.dim_date
27 (
28   date_id integer not null,
29   full_date_time timestamp,
30   year integer,
31   month varchar(20),
32   dat integer,
33   hour integer,
34   weekday varchar(20),
35   primary key (date_id)
36 );
37
38 create table etl_project.dim_card_type
39 (
40   card_type_id integer not null,

```

5) DIM_CARD_TYPE table creation →

```
create table etl_project.dim_card_type
(
  card_type_id integer not null,
  card_type varchar(20),
  primary key (card_type_id)
);
```



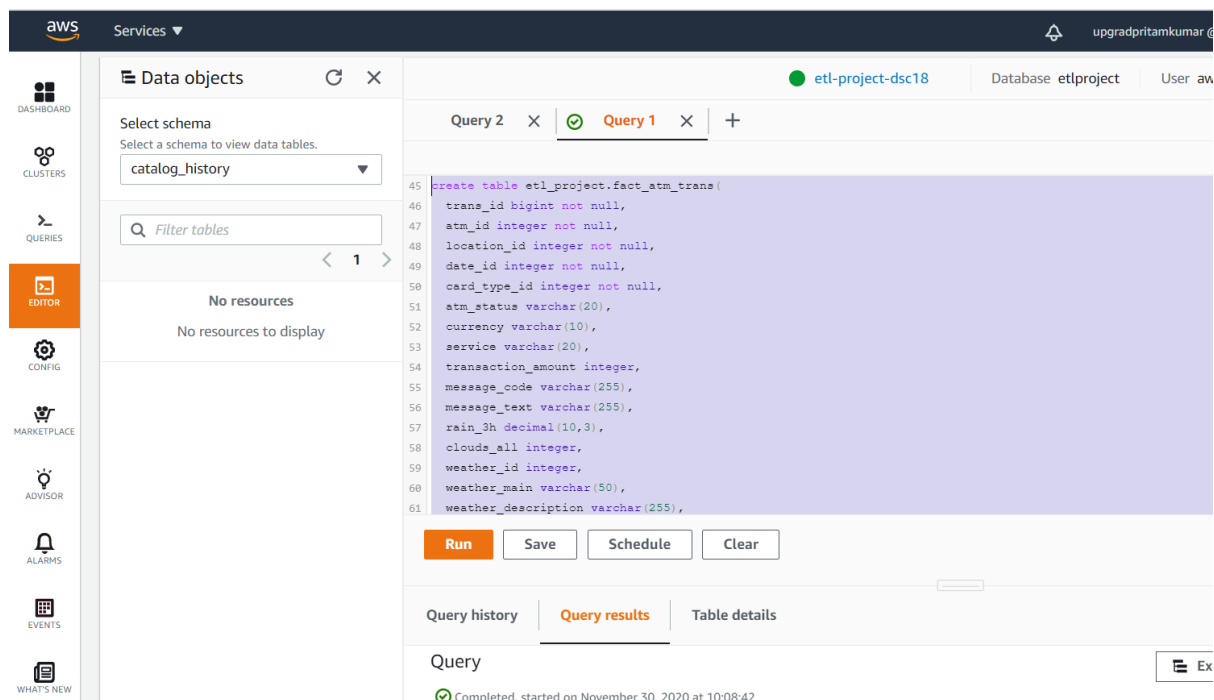
The screenshot shows the AWS Glue console interface. On the left, there is a sidebar with navigation options: DASHBOARD, CLUSTERS, QUERIES, EDITOR (highlighted), CONFIG, MARKETPLACE, ADVISOR, ALARMS, EVENTS, and WHAT'S NEW. The main area is titled 'Data objects' and shows a 'Select schema' dropdown set to 'catalog_history'. Below this, it says 'No resources' and 'No resources to display'. The right pane shows a SQL query editor with the following code:

```
37
38 create table etl_project.dim_card_type
39 (
40   card_type_id integer not null,
41   card_type varchar(20),
42   primary key (card_type_id)
43 )
44
45 create table etl_project.fact_atm_trans(
46   trans_id bigint not null,
47   atm_id integer not null,
48   location_id integer not null,
49   date_id integer not null,
50   card_type_id integer not null,
51   atm_status varchar(20),
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there are tabs for 'Query history', 'Query results' (selected), and 'Table details'. The 'Query results' tab shows a status message: 'Completed, started on November 30, 2020 at 10:07:31' and 'ELAPSED TIME: 00 m 08 s'.

6) FACT_ATM_TRANS table creation →

```
create table etl_project.fact_atm_trans(
  trans_id bigint not null,
  atm_id integer not null,
  location_id integer not null,
  date_id integer not null,
  card_type_id integer not null,
  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 (location_id) references etl_project.dim_location(location_id),
  foreign key (atm_id) references etl_project.dim_atm(atm_id),
  foreign key (date_id) references etl_project.dim_date(date_id),
  foreign key (card_type_id) references etl_project.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_project.dim_location from 's3://etl-project-dsc18/location-dim/part-00000-615b2af3-571b-4f08-99a2-2593db79654f-c000.csv'
iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv
IGNOREHEADER 1 delimiter ',' region 'us-east-1';
- 2) copy etl_project.dim_atm from 's3://etl-project-dsc18/atm-dim/part-00000-e9e6d7de-1169-466e-b67b-38a780751d41-c000.csv'
iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv
IGNOREHEADER 1 delimiter ',' region 'us-east-1';
- 3) copy etl_project.dim_card_type from 's3://etl-project-dsc18/card-type-dim/part-00000-d688fcb7-c70f-42a7-a262-987479a3adcb-c000.csv'
iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1
delimiter ',' region 'us-east-1';
- 4) copy etl_project.dim_date from 's3://etl-project-dsc18/date-dim/part-00000-f409154c-f0b1-447e-807b-9ff31ae4c6a8-c000.csv'
iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1
delimiter ',' region 'us-east-1';
- 5) copy etl_project.fact_atm_trans from 's3://etl-project-dsc18/fact-atm-trans/part-00000-44acf61e-3e37-4a6e-8a4f-53376f75524d-c000.csv'
iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1
delimiter ',' region 'us-east-1';

aws Services

upgradpritamkumar @ 1136-9228-6195 N. Virgin

DASHBOARD

CLUSTERS

QUERIES

EDITOR

CONFIG

MARKETPLACE

ADVISOR

ALARMS

EVENTS

WHAT'S NEW

Select schema

etl_project

Filter tables

1

dim_atm

dim_atm_pkey

dim_card_type

dim_card_type_pkey

dim_date

dim_date_pkey

dim_location

dim_location_pkey

fact_atm_trans

fact_atm_trans_pkey

Query 2

Query 1

```

66 foreign key (card_type_id) references etl_project.dim_card_type(card_type_id)
67 );
68
69 copy etl_project.dim_location from 's3://etl-project-dsc18/location-dim/part-00000-615b2af3-571b-4f08-99a2-2593db79654f-c000.csv'
70 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
71
72 copy etl_project.dim_atm from 's3://etl-project-dsc18/atm-dim/part-00000-e9e6d7de-1169-466e-b67b-38a780751d41-c000.csv'
73 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
74
75 copy etl_project.dim_card_type from 's3://etl-project-dsc18/card-type-dim/part-00000-d688fcb7-c70f-42a7-a262-987479a3adcb-c000.csv'
76 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
77
78 select count(*) from etl_project.dim_atm;
79
80 select * from etl_load_errors;

```

Run Save Schedule Clear

Send feedback

Query history Query results Table details

Query 665

Execution Data

Completed, started on November 30, 2020 at 10:46:03
ELAPSED TIME: 00 m 52 s

aws Services

upgradpritamkumar @ 1136-9228-6195 N. Virgin

DASHBOARD

CLUSTERS

QUERIES

EDITOR

CONFIG

MARKETPLACE

ADVISOR

ALARMS

EVENTS

WHAT'S NEW

Data objects

etl-project-dsc18

Database etlproject

User awsuser

Change connection

Select schema

etl_project

Filter tables

1

dim_atm

dim_atm_pkey

dim_card_type

dim_card_type_pkey

dim_date

dim_date_pkey

dim_location

dim_location_pkey

fact_atm_trans

fact_atm_trans_pkey

Query 2

Query 1

```

66 foreign key (card_type_id) references etl_project.dim_card_type(card_type_id)
67 );
68
69 copy etl_project.dim_location from 's3://etl-project-dsc18/location-dim/part-00000-615b2af3-571b-4f08-99a2-2593db79654f-c000.csv'
70 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
71
72 copy etl_project.dim_atm from 's3://etl-project-dsc18/atm-dim/part-00000-e9e6d7de-1169-466e-b67b-38a780751d41-c000.csv'
73 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
74
75 copy etl_project.dim_card_type from 's3://etl-project-dsc18/card-type-dim/part-00000-d688fcb7-c70f-42a7-a262-987479a3adcb-c000.csv'
76 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
77
78 copy etl_project.dim_date from 's3://etl-project-dsc18/date-dim/part-00000-f409154c-f0b1-447e-807b-9ff31ae4c6a8-c000.csv'
79 iam_role 'arn:aws:iam::113692286195:role/upgrad-etl-project-dsc18' csv IGNOREHEADER 1 delimiter ',' region 'us-east-1';
80
81 select count(*) from etl_project.dim_atm;

```

Run Save Schedule Clear

Send feedback

Query history Query results Table details

Query 716

Execution Data

Completed, started on November 30, 2020 at 10:48:13
ELAPSED TIME: 00 m 21 s