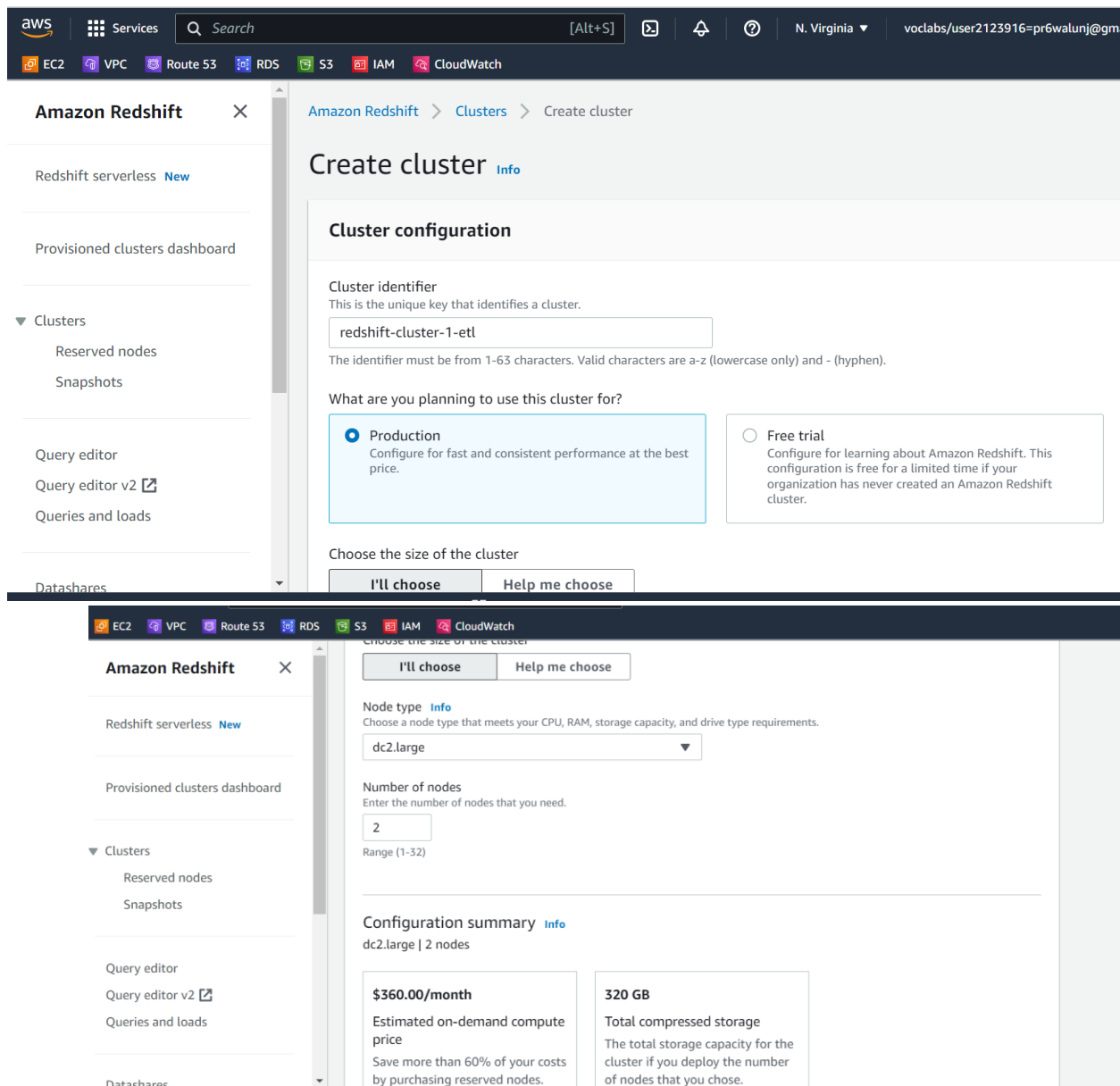


Creation of a Redshift Cluster

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



The first screenshot shows the 'Create cluster' page in the AWS Management Console. The 'Cluster identifier' is set to 'redshift-cluster-1-etl'. The 'What are you planning to use this cluster for?' section has 'Production' selected. The 'Choose the size of the cluster' section has 'I'll choose' selected.

The second screenshot shows the 'Node type' and 'Number of nodes' configuration. The 'Node type' is set to 'dc2.large' and the 'Number of nodes' is set to '2'. The 'Configuration summary' shows a total cost of '\$360.00/month' and '320 GB' of storage.

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2123916=pr6walunj@gr

EC2

VPC

Route 53

RDS

S3

IAM

CloudWatch

Amazon Redshift

Redshift serverless **New**

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datacharec

Database configurations

Admin user name

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

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

☐ Auto generate password

Amazon Redshift can generate a password for you, or you can specify your own password.

Admin user password

Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except "/", "", or "@".

☐ Show password

Cluster permissions

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2123916=pr6walunj@gmail.com @ 6498-5516-0313

EC2

VPC

Route 53

RDS

S3

IAM

CloudWatch

Amazon Redshift > Clusters > redshift-cluster-1-etl

redshift-cluster-1-etl

Actions

Edit

Add partner integration

Query data

General information

Cluster identifier

redshift-cluster-1-etl

Status

Available

Node type

dc2.large

Endpoint

redshift-cluster-1-etl.cabqcjcw5f8...

Cluster namespace

fce127ee-7a95-463a-bd9a-97554868a24b

Date created

November 08, 2022, 13:29 (UTC+05:30)

Number of nodes

2

JDBC URL

jdbc:redshift://redshift-cluster-1-etl...

Storage used

-

ODBC URL

Driver={Amazon Redshift (x64)}; Ser...

Cluster permissions

Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

Search for associated IAM role by name, status, or role type

☐ IAM roles

Status

Role type

☐ redshift_s3_fullaccess

in-sync

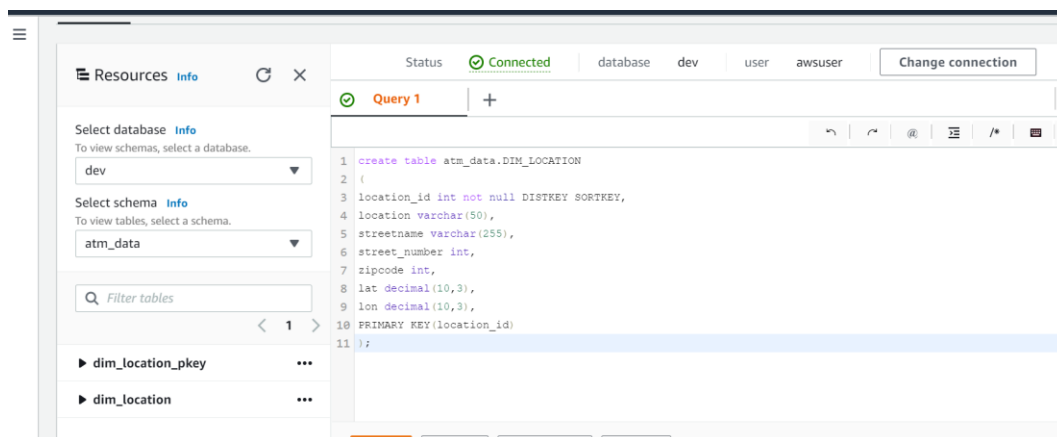
--

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

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:

```
create table atm_data.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)
);
```



2.]

```
create table atm_data.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_data.DIM_LOCATION(location_id)
```

Resources Info

Select database Info
To view schemas, select a database.
dev

Select schema Info
To view tables, select a schema.
atm_data

Filter tables

1

dim_atm_pkey ...

dim_location_pkey ...

dim_atm ...

dim_location ...

Query 1

```

1 create table atm_data.DIM_ATM
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );

```

Run Save Schedule Clear

Send feedback

3.]

```

create table atm_data.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)
);

```

To view schemas, select a database.
dev

Select schema Info
To view tables, select a schema.
atm_data

Filter tables

1

dim_atm_pkey ...

dim_date_pkey ...

dim_location_pkey ...

dim_atm ...

dim_date ...

dim_location ...

Query 1

```

1 create table atm_data.DIM_DATE
2 (
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int,
6   month varchar(20),
7   day int,
8   hour int,
9   weekday varchar(20),
10  PRIMARY KEY(date_id)
11 );

```

Run Save Schedule Clear

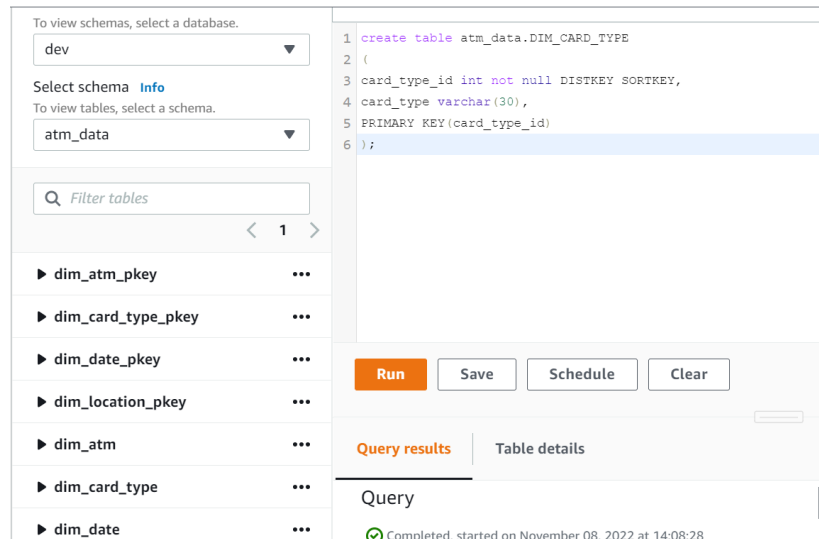
Query results Table details

Query

Execution Data

4.]

```
create table atm_data.DIM_CARD_TYPE
(
  card_type_id int not null DISTKEY SORTKEY,
  card_type varchar(30),
  PRIMARY KEY(card_type_id)
);
```

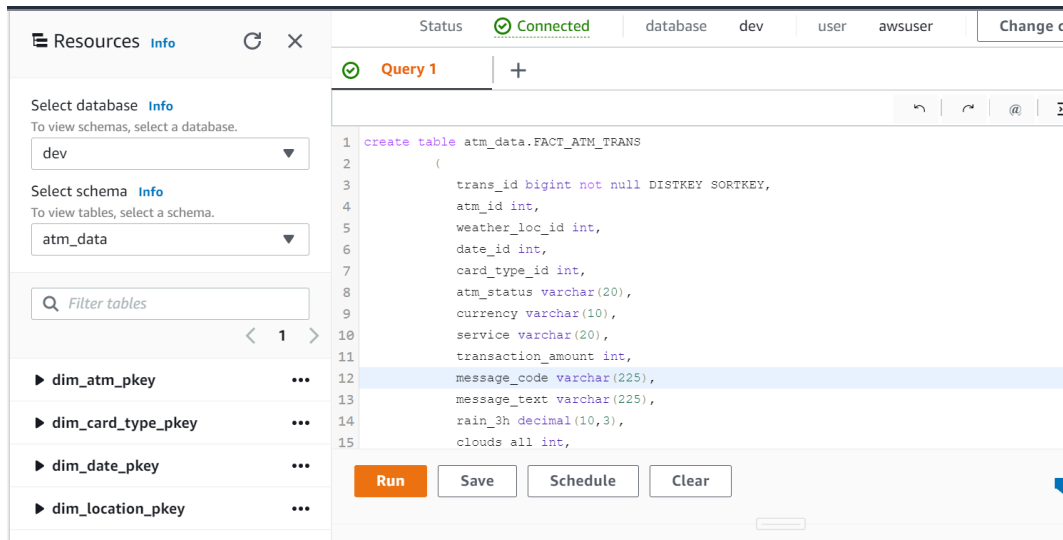


The screenshot shows a database management interface. On the left, there's a sidebar with a dropdown menu set to 'dev' and 'atm_data' selected. Below this is a search bar labeled 'Filter tables' and a list of tables including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'dim_atm', 'dim_card_type', and 'dim_date'. On the right, a SQL query is entered in a text area: `create table atm_data.DIM_CARD_TYPE (card_type_id int not null DISTKEY SORTKEY, card_type varchar(30), PRIMARY KEY(card_type_id));`. Below the query area are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there's a status bar indicating 'Query results' and 'Table details' tabs, with a message 'Query Completed, started on November 08, 2022 at 14:08:28'.

5.]

```
create table atm_data.FACT_ATM_TRANS
(
  trans_id bigint not null DISTKEY SORTKEY,
  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(225),
  message_text varchar(225),
  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_data.DIM_LOCATION(location_id),
  FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
```

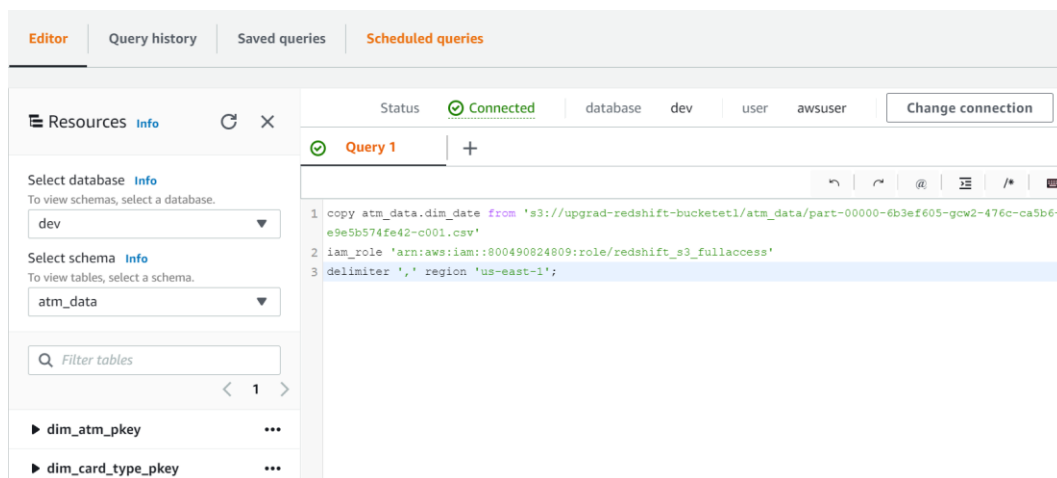
FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
);



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

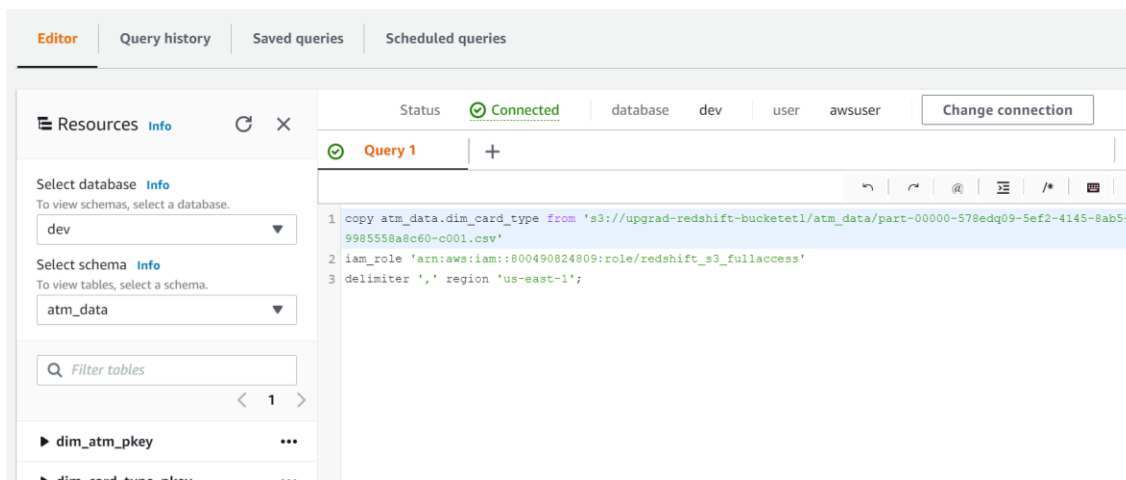
```
copy atm_data.dim_date from 's3://upgrad-redshift-bucketetl/atm_data/part-00000-6b3ef605-gcw2-476c-
ca5b6-e9e5b574fe42-c001.csv'
iam_role 'arn:aws:iam::800490824809:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1';
```



Copy the data to dim_card_type table

copy atm_data.dim_card_type from 's3://upgrad-redshift-bucketetl/atm_data/part-00000-578edq09-5ef2-4145-8ab5-9985558a8c60-c001.csv'

iam_role 'arn:aws:iam::800490824809:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1';



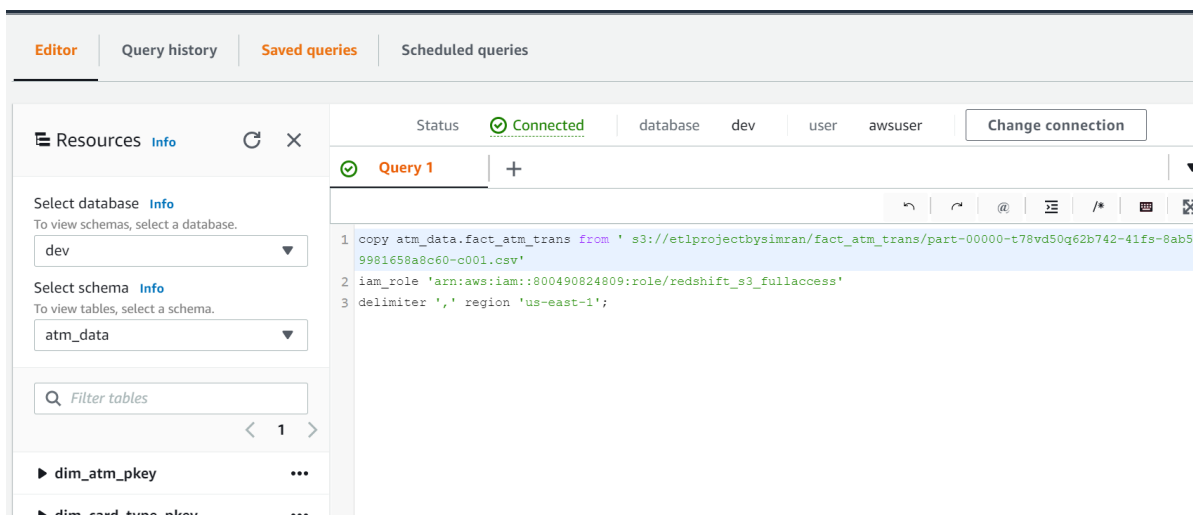
The screenshot shows the Redshift console interface. The 'Editor' tab is active, displaying a SQL query. The 'Resources' panel on the left shows the database 'dev' and schema 'atm_data'. The query editor shows the following SQL:

```
1 copy atm_data.dim_card_type from 's3://upgrad-redshift-bucketetl/atm_data/part-00000-578edq09-5ef2-4145-8ab5-9985558a8c60-c001.csv'
2 iam_role 'arn:aws:iam::800490824809:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1';
```

Copy the data to fact_atm_trans table

copy atm_data.fact_atm_trans from 's3://etlprojectbysimran/fact_atm_trans/part-00000-t78vd50q62b742-41fs-8ab5-9981658a8c60-c001.csv'

iam_role 'arn:aws:iam::800490824809:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1';



The screenshot shows the Redshift console interface. The 'Editor' tab is active, displaying a SQL query. The 'Resources' panel on the left shows the database 'dev' and schema 'atm_data'. The query editor shows the following SQL:

```
1 copy atm_data.fact_atm_trans from 's3://etlprojectbysimran/fact_atm_trans/part-00000-t78vd50q62b742-41fs-8ab5-9981658a8c60-c001.csv'
2 iam_role 'arn:aws:iam::800490824809:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1';
```