

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 > Clusters > etlprojectredshift

etlprojectredshift ▾

Actions ▾

Edit

Add partner integration

Query data ▾

General information

Cluster identifier

etlprojectredshift

Status

⏸ Paused

Node type

dc2.large

Endpoint

etlprojectredshift.crcivsfs7xh0.us-east-1...

Cluster namespace

ef69d3a7-fc9b-41d2-9428-0dd7d381b9e2

Date created

July 30, 2022, 16:21 (UTC+05:30)

Number of nodes

2

JDBC URL

jdbc:redshift://etlprojectredshift.crcivsfs...

Storage used

0.11% (0.35 of 320 GB used)

AQUA

Not available

ODBC URL

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

Cluster performance

Query monitoring

Schedules

Maintenance

Properties

Database configurations

Change admin user password

Rotate encryption keys

Edit ▾

Database name

etl_project_dev

Parameter group

Defines database parameter and query queues for all the databases.

default.redshift-1.0

Encryption

Disabled

AWS KMS key ID

-

Audit logging

Disabled

Port

5439

SSH ingestion setting (cluster public key)

ssh-rsa AAAAB3NzaC1yc2EAAAADAQ...

Admin user name

awsuser

Network and security settings

Edit

Virtual private cloud (VPC)

vpc-0d55b8b3b93aeb718

Availability Zone

us-east-1b

VPC security group

Specify which instances and devices can connect to the cluster.

sg-05bde3e674aa24da9

Publicly accessible

Allow instances and devices outside the VPC to connect to your database through the cluster endpoint.

Disabled

Subnet

malli-cluster-subnet-group-1

Enhanced VPC routing

Disabled

Endpoint URL

-

Cluster permissions

① 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) [Info](#)

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.

Set default ▼

Manage IAM roles ▼

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

< 1 >

<input type="checkbox"/>	IAM roles 🔗 ▼	Status ▼	Role type ▼
<input type="checkbox"/>	UpGrad_RedShift_S3	✔ in-sync	--

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 etl_project.CardType
```

```
(
```

```
card_type varchar(100)
```

```
,CardType_ID integer
```

```
);
```

```
create table etl_project.Location
```

```
(
```

```
atm_location varchar(100)
```

```
,atm_streetname varchar(300)
```

```
,atm_street_number integer
```

```
,atm_zipcode integer
```

```
,atm_lat decimal(10,3)
```

```
,atm_lon decimal(10,3)
```

```
,Location_ID integer not null distkey sortkey
```

```
);
```

```
create table etl_project.Date
```

```
(
```

```
Year integer
```

```
,Month varchar(100)
```

```
,Day integer
```

```
,Hour integer
```

,Weekday varchar(100)

,full_date_time varchar(100)

,Date_ID integer

);

create table etl_project.Atm

(

ATM_Dim_Id integer

,Atm_manufacturer varchar(100)

,atm_number varchar(100)

,atm_location_id integer

);

create table etl_project.Fact_Atm_Trans

(

Location_Id integer

,ATM_Dim_Id integer

,Date_Id integer

,CardType_Id integer

,atm_status varchar(50)

,currency varchar(50)

,service varchar(50)

,transaction_amount integer

```
,message_code varchar(500)

,message_text varchar(500)

,rain_3h varchar(50)

,clouds_all varchar(50)

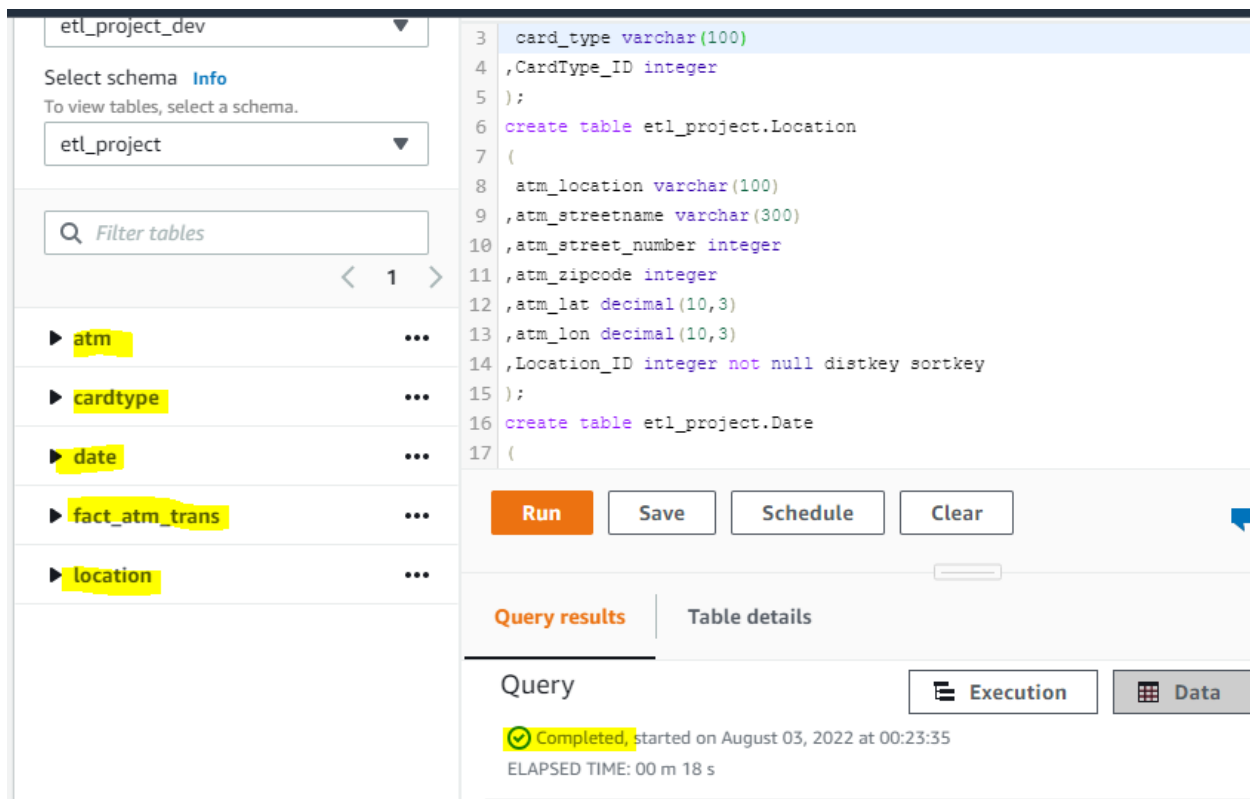
,weather_id varchar(50)

,weather_main varchar(50)

,weather_description varchar(500)

,trans_id integer

);
```



The screenshot shows a Redshift console interface. On the left, a sidebar displays the database schema 'etl_project_dev' and a list of tables: 'atm', 'cardtype', 'date', 'fact_atm_trans', and 'location'. The main area shows a SQL query editor with the following code:

```
3 card_type varchar(100)
4 ,CardType_ID integer
5 );
6 create table etl_project.Location
7 (
8   atm_location varchar(100)
9   ,atm_streetname varchar(300)
10  ,atm_street_number integer
11  ,atm_zipcode integer
12  ,atm_lat decimal(10,3)
13  ,atm_lon decimal(10,3)
14  ,Location_ID integer not null distkey sortkey
15 );
16 create table etl_project.Date
17 (
```

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 results' and 'Table details'. The 'Query results' tab is active, showing a status of 'Completed' with a green checkmark, indicating the query was successful. The execution time is shown as 'ELAPSED TIME: 00 m 18 s'.

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 etl_project.CardType from

's3://mallietlprojects3bucket/Dimensions/CardType/part-00000-00e534dd-dc6a-4f5e-856b-1a8b37e92ee8-c000.csv'

iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'

delimiter ',' region 'us-east-1';

copy etl_project.Location from

's3://mallietlprojects3bucket/Dimensions/Location/part-00000-152a8dae-7a0f-401e-970b-9fb0d602597f-c000.csv'

iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'

delimiter ',' region 'us-east-1';

copy etl_project.Date from

's3://mallietlprojects3bucket/Dimensions/Date/part-00000-6a59cf56-156a-4700-9e9b-db3999a1ee25-c000.csv'

iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'

delimiter ',' region 'us-east-1';

copy etl_project.Atm from

's3://mallietlprojects3bucket/Dimensions/Atm/part-00000-9118bd77-dab3-4c74-9ba6-36e5fcbaeac1-c000.csv'

iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'

delimiter ',' region 'us-east-1';

copy etl_project.Fact_Atm_Trans from

's3://mallietlprojects3bucket/Facts/part-00000-8954d551-71c3-4222-9267-7b54cb62b887-c000.csv'

iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'

delimiter ',' region 'us-east-1';

```

13 iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'
14 delimiter ',' region 'us-east-1';
15
16 copy etl_project.Atm from
17 's3://mallietlprojects3bucket/Dimensions/Atm/part-00000-9118bd77-dab3-4c74-9ba6-36e5fcbaeac1-c000.c
18 iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'
19 delimiter ',' region 'us-east-1';
20
21 copy etl_project.Fact_Atm_Trans from
22 's3://mallietlprojects3bucket/Facts/part-00000-8954d551-71c3-4222-9267-7b54cb62b887-c000.csv'
23 iam_role 'arn:aws:iam::719946286510:role/UpGrad_RedShift_S3'
24 delimiter ',' region 'us-east-1';

```

Run

Save

Schedule

Clear

Query results

Table details

Query 722 [↗](#)

Execution

✓ Completed, started on August 03, 2022 at 00:30:24

ELAPSED TIME: 02 m 48 s