

Creation of a Redshift Cluster

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

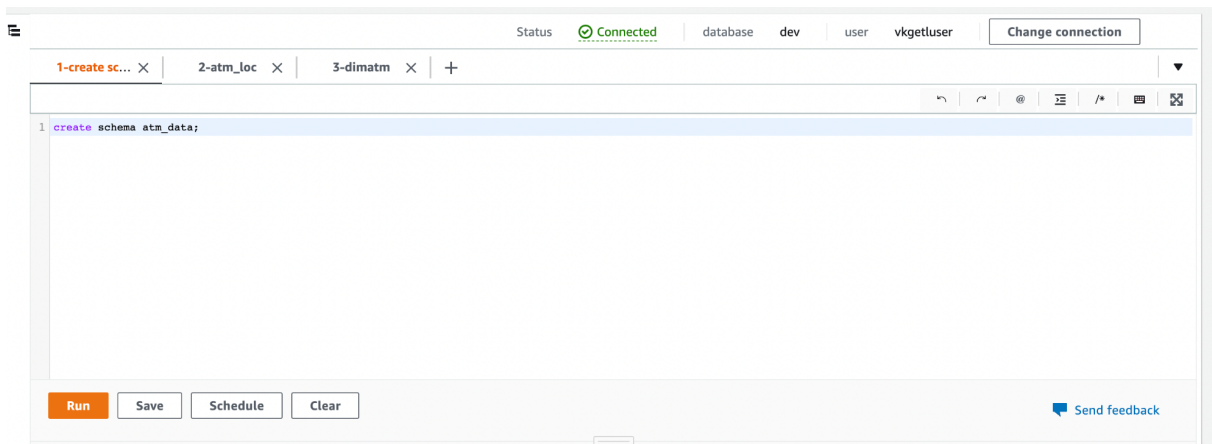
Clusters (1) Info									
<input type="text" value="Filter clusters by property or value"/> < 1 > ⓘ									
<input type="checkbox"/>	Cluster	Status	Cluster namespace	Availability Zone	Multi-AZ	Storage capacity us...	CPU utilization	Snapshots	
<input type="checkbox"/>	redshift-cluster-etlproject dc2.large 2 nodes 320 GB	⏸ Modifying Resuming	ea31e9e8-c6ff-4dd3-b32f-2e4...	us-east-1a	No			-	

redshift-cluster-etlproject			
Actions Edit Add partner integration Query data			
General information ⓘ			
Cluster identifier redshift-cluster-etlproject	Status 🟢 Available	Node type dc2.large	Endpoint 🔗 redshift-cluster-etlproject.c3o6rq28qakx.us-east-1.red...
Cluster namespace ea31e9e8-c6ff-4dd3-b32f-2e40be4559aa	Date created December 19, 2022, 22:12 (UTC+11:00)	Number of nodes 2	JDBC URL 🔗 jdbc:redshift://redshift-cluster-etlproject.c3o6rq28qak...
Cluster configuration Production	Storage used 0.07% (0.22 of 320 GB used)		ODBC URL 🔗 Driver={Amazon Redshift (x64)}; Server=redshift-cluste...
	Multi-AZ No		

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 schema atm_data;



```
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)  
)
```

Editor

Query history

Saved queries

Scheduled queries

Status

Connected

database

dev

user

vkgetluser

Change connection

1-create sc...

2-atm_loc

3-dimatm

+

```

7  --|-- lat: decimal(38,18) (nullable = true)
8  --|-- lon: decimal(38,18) (nullable = true)
9
10 create table atm_data.DIM_LOCATION(
11     location_id int not null distkey sortkey,
12     location varchar(50),
13     streetname varchar(255),
14     street_number int,
15     zipcode int,
16     lat decimal(10,3),
17     lon decimal(10,3),
18     primary key(location_id)
19 )

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

No resources

To view details, choose data from navigator.

```

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)
)

```

Editor

Query history

Saved queries

Scheduled queries

Status

Connected

database

dev

user

vkgetluser

Change connection

1-create sc... X

2-atm_loc X

3-dimatm X

+

```

1 -- root
2 --|-- atm_id: long (nullable = true)
3 --|-- atm_number: string (nullable = true)
4 --|-- atm_manufacturer: string (nullable = true)
5 --|-- atm_location_id: long (nullable = true)
6
7 create table atm_data.DIM_ATM(
8     atm_id int not null distkey sortkey,
9     atm_number varchar(20),
10    atm_manufacturer varchar(50),
11    atm_location_id int,
12    primary key (atm_id),
13    foreign key (atm_location_id) references atm_data.DIM_LOCATION(location_id)
14 )

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

No resources

To view details, choose data from navigator.

```

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)
);

```

Editor

Query history

Saved queries

Scheduled queries

Status

Connected

database

dev

user

vkgetluser

Change connection

1-create sc... X

2-atm_loc X

3-dimatm X

4-DIM_Date X

+

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

Send feedback

Query results

Table details

Query

Execution

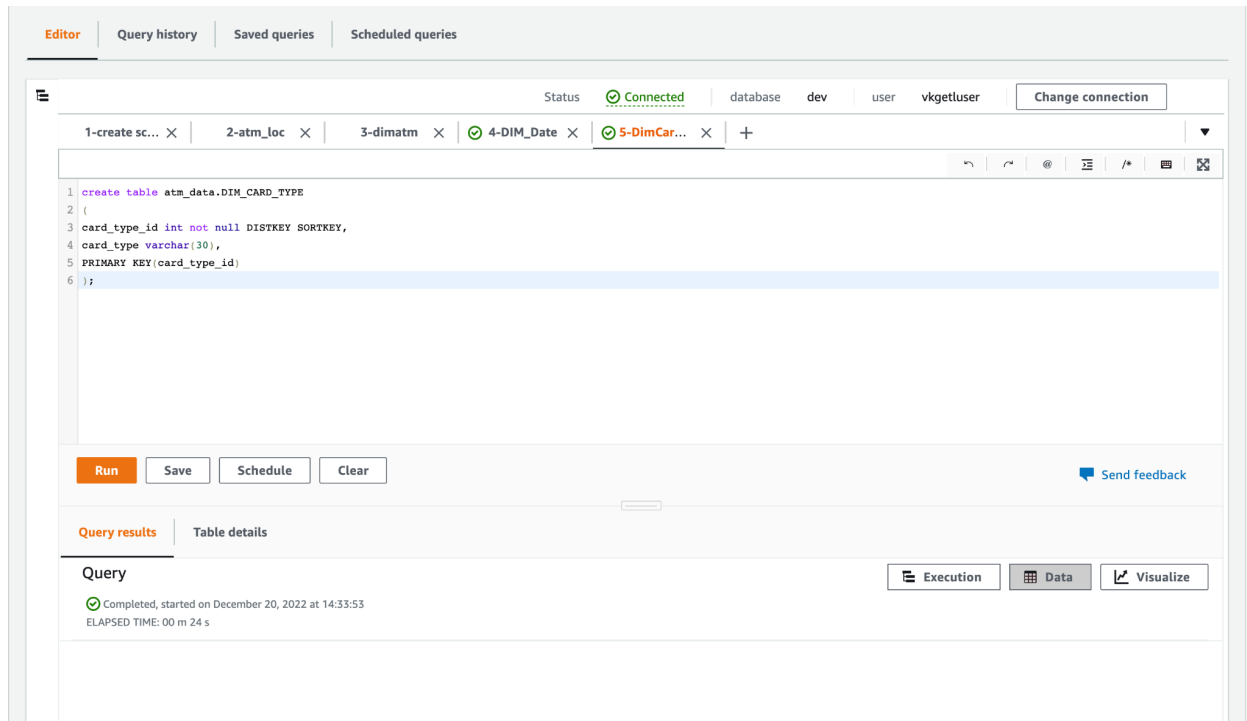
Data

Visualize

Completed, started on December 20, 2022 at 14:32:13

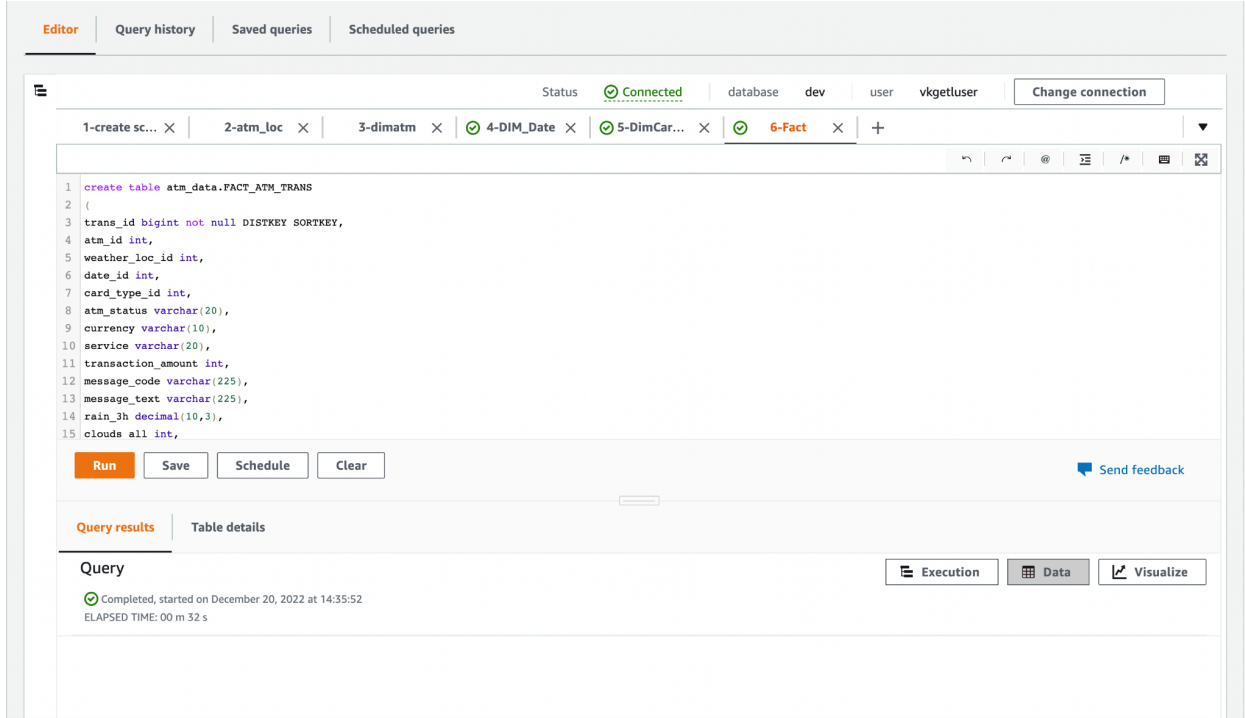
ELAPSED TIME: 00 m 30 s

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



```
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)
);



The screenshot displays the upGrad SQL Editor interface. At the top, there are tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. Below these, a status bar shows 'Status: Connected', 'database: dev', 'user: vkgetluser', and a 'Change connection' button. A tab bar at the top of the editor shows several open queries, with '6-Fact' currently selected. The main editor area contains a SQL query to create a table named 'atm_data.FACT_ATM_TRANS'. The query is as follows:

```
1 create table atm_data.FACT_ATM_TRANS
2 (
3   trans_id bigint not null DISTKEY SORTKEY,
4   atm_id int,
5   weather_loc_id int,
6   date_id int,
7   card_type_id int,
8   atm_status varchar(20),
9   currency varchar(10),
10  service varchar(20),
11  transaction_amount int,
12  message_code varchar(225),
13  message_text varchar(225),
14  rain_3h decimal(10,3),
15  clouds all int,
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The bottom section of the interface is titled 'Query results' and 'Table details'. Under 'Query results', there is a 'Query' section showing the execution status: 'Completed, started on December 20, 2022 at 14:35:52' and 'ELAPSED TIME: 00 m 32 s'. To the right of the query status, there are buttons for 'Execution', 'Data', and 'Visualize'.

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_location from 's3://vivek-etlprojectbucket01/location_dimension/'  
iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```

```
copy atm_data.dim_atm from 's3://vivek-etlprojectbucket01/atm_dimension/'  
iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```

```
copy atm_data.dim_card_type from 's3://vivek-etlprojectbucket01/card_type_dimension/'  
iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```

```
copy atm_data.dim_date from 's3://vivek-etlprojectbucket01/date_dimension/'  
iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```

```
copy atm_data.fact_atm_trans from 's3://vivek-etlprojectbucket01/atm_transaction_fact/'  
iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'  
delimiter ',' region 'us-east-1'  
CSV;
```


Editor
Query history
Saved queries
Scheduled queries

Status
Connected
database
dev
user
vkgetluser
Change connection

1-create sc... X
2-atm_loc X
3-dimatm X
4-DIM_Date X
5-DimCar... X
6-Fact X
7-copy-di... X
+

```

1 copy_atm_data.dim_location from 's3://vivek-etlprojectbucket01/location_dimension/'
2 iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'
3 delimiter ',' region 'us-east-1'
4 CSV;

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query 200296
Execution
Data
Visualize

Completed, started on December 20, 2022 at 14:39:05
ELAPSED TIME: 00 m 33 s

Editor
Query history
Saved queries
Scheduled queries

Status
Connected
database
dev
user
vkgetluser
Change connection

1-create sc... X
2-atm_loc X
3-dimatm X
4-DIM_Date X
5-DimCar... X
6-Fact X
7-copy-di... X
+

```

1 copy_atm_data.dim_location from 's3://vivek-etlprojectbucket01/location_dimension/'
2 iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'
3 delimiter ',' region 'us-east-1'
4 CSV;

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query 200296
Execution
Data
Visualize

Completed, started on December 20, 2022 at 14:39:05
ELAPSED TIME: 00 m 33 s

Editor
Query history
Saved queries
Scheduled queries

Status

Connected

database

dev

user

vkgetluser

Change connection

1-create sc... X

2-atm_loc X

3-dimatm X

4-DIM_Date X

5-DimCar... X

6-Fact X

7-copy-di... X

8-dimatm X

+

▼

```

1 copy atm_data.dim_atm from 's3://vivek-etlprojectbucket01/atm_dimension/'
2 iam_role 'arn:aws:iam::710861875122:role/redshift_s3_full_access'
3 delimiter ',' region 'us-east-1'
4 CSV;

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 200331

Execution

Data

Visualize

Completed, started on December 20, 2022 at 14:41:30
ELAPSED TIME: 00 m 21 s