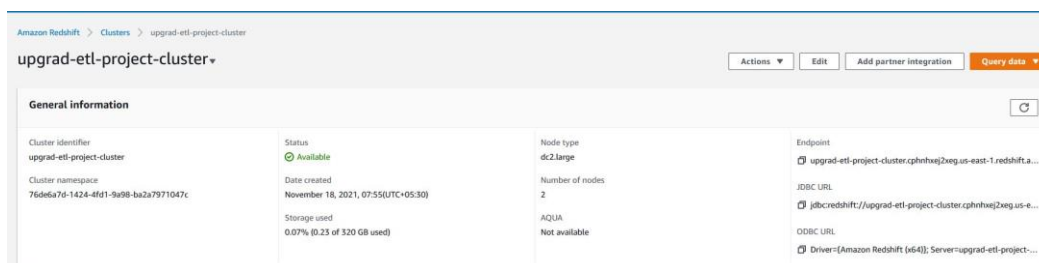


## Creation of a Redshift Cluster

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

1. Screenshot for type of machine and number of nodes used.



Amazon Redshift > Clusters > upgrad-etl-project-cluster

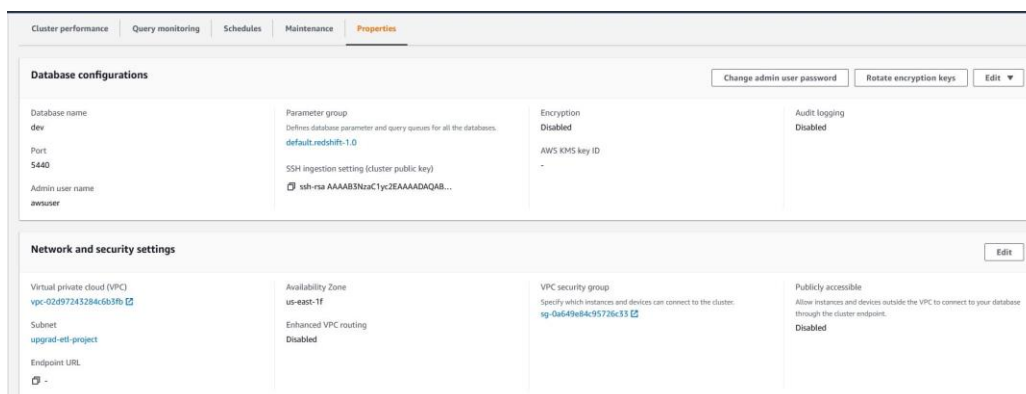
upgrad-etl-project-cluster

Actions Edit Add partner integration Query data

**General information**

Cluster identifier upgrad-etl-project-cluster	Status Available	Node type dc2.large	Endpoint upgrad-etl-project-cluster.cphnhwz2eq.us-east-1.redshift.a...
Cluster namespace 76de6a76-1424-4fd1-9a98-ba2a7971047c	Date created November 18, 2021, 07:55(UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://upgrad-etl-project-cluster.cphnhwz2eq.us-e...
Storage used 0.07% (0.23 of 320 GB used)		AQUA Not available	ODBC URL Driver=(Amazon Redshift (x64)); Server=upgrad-etl-project...

2. Screenshot for database configurations and network properties.



Cluster performance Query monitoring Schedules Maintenance Properties

**Database configurations**

Change admin user password Rotate encryption keys Edit

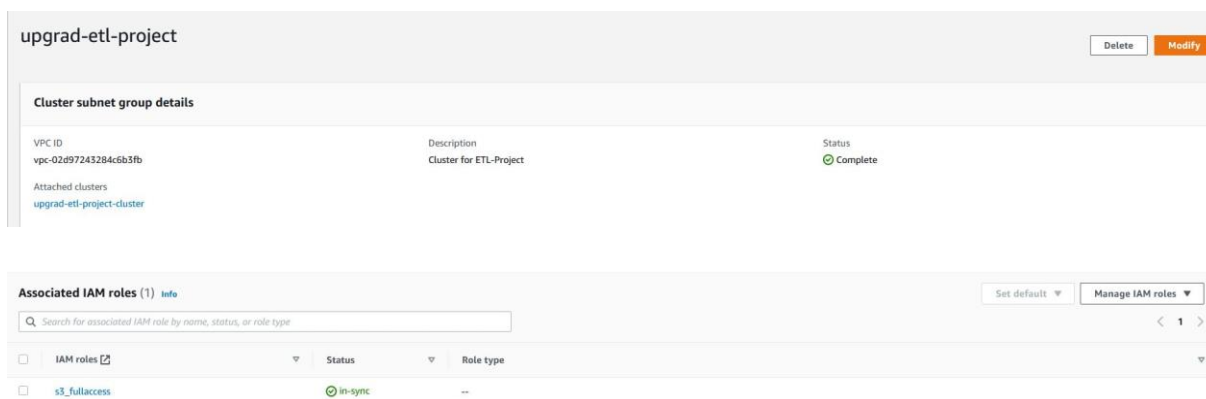
Database name dev	Parameter group default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5440	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...	AWS KMS key ID -	
Admin user name aweruser			

**Network and security settings**

Edit

Virtual private cloud (VPC) vpc-02d97243284c6b3fb	Availability Zone us-east-1f	VPC security group sg-0a649e84c9573ac33	Publicly accessible Disabled
Subnet upgrad-etl-project	Enhanced VPC routing Disabled		
Endpoint URL -			

3. Screenshots subnet group and associated IAM rules.



upgrad-etl-project

Delete Modify

**Cluster subnet group details**

VPC ID vpc-02d97243284c6b3fb	Description Cluster for ETL-Project	Status Complete
Attached clusters upgrad-etl-project-cluster		

**Associated IAM roles (1)**

Set default Manage IAM roles

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

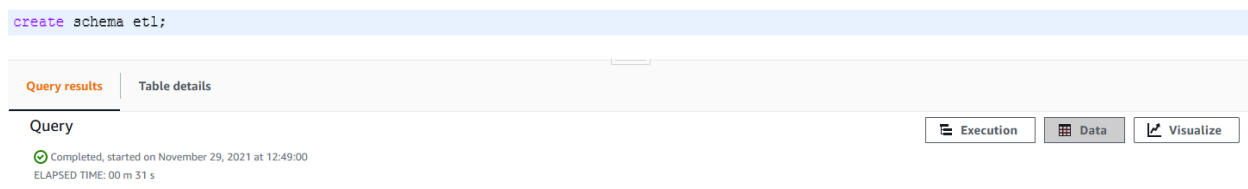
<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	s3_fullaccess	in-sync	--

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:**

1. Query to create schema.

```
create schema etl;
```

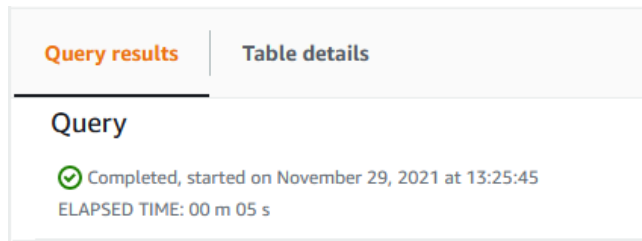


The screenshot shows the Redshift Query Editor interface. At the top, the SQL query `create schema etl;` is entered. Below the query, there are two tabs: "Query results" (active) and "Table details". Under the "Query results" tab, the status is "Query" and it shows a green checkmark indicating the query was completed. The execution details state: "Completed, started on November 29, 2021 at 12:49:00" and "ELAPSED TIME: 00 m 31 s". On the right side of the interface, there are three buttons: "Execution", "Data", and "Visualize".

2. Query to create LOCATION dimension table.

```
create table etl.loc(
location_id integer not null,
atm_location varchar(50),
atm_streetname varchar(255),
atm_street_number integer,
atm_zipcode integer,
atm_lat NUMERIC(10,3),
atm_lon NUMERIC(10,3),
primary key(location_id));
```

```
create table etl.loc(
location_id integer not null,
atm_location varchar(50),
atm_streetname varchar(255),
atm_street_number integer,
atm_zipcode integer,
atm_lat NUMERIC(10,3),
atm_lon NUMERIC(10,3),
primary key(location_id));
```



The screenshot shows the Redshift Query Editor interface for the second query. The "Query results" tab is active, showing a green checkmark and the status "Query". The execution details state: "Completed, started on November 29, 2021 at 13:25:45" and "ELAPSED TIME: 00 m 05 s". The "Table details" tab is also visible but not selected.

3. Query to create ATM dimension table.

```
create table etl.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.loc(location_id));
```

```
15 | create table etl.atm(  
16 |   atm_id integer not null,  
17 |   atm_number varchar(20),  
18 |   atm_manufacturer varchar(50),  
19 |   location_id integer,  
20 |   primary key(atm_id),  
21 |   foreign key(location_id) references etl.loc(location_id));  
22 |
```

Run Save Schedule Clear

Query results | Table details

### Query

✔ Completed, started on November 29, 2021 at 12:51:46  
ELAPSED TIME: 00 m 13 s

#### 4. Query to create DATE dimension table.

```
create table etl.date(
year integer,
month varchar(20),
day integer,
hour integer,
weekday varchar(20),
full_date_time timestamp,
date_id integer,
primary key(date_id));
```

```
23 | create table etl.date(
24 | year integer,
25 | month varchar(20),
26 | day integer,
27 | hour integer,
28 | weekday varchar(20),
29 | full_date_time timestamp,
30 | date_id integer,
31 | primary key(date_id));
32 |
```

Run
Save
Schedule
Clear

Query results
Table details

### Query

✓ Completed, started on November 29, 2021 at 12:52:22  
 ELAPSED TIME: 00 m 05 s

#### 5. Query to create CARD dimension table.

```
create table etl.card(
card_type varchar(23),
card_type_id integer,
primary key(card_type_id));
```

```
33 | create table etl.card(
34 | card_type varchar(23),
35 | card_type_id integer,
36 | primary key(card_type_id));
37 |
```

Run
Save
Schedule
Clear

Query results
Table details

### Query

✓ Completed, started on November 29, 2021 at 12:53:04  
 ELAPSED TIME: 00 m 08 s

## 6. Query to create FACT\_ATM\_TRANS table.

```
create table etl.FACT_ATM_TRANS(
trans_id BIGINT not null,
atm_id integer,
location_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 NUMERIC(10,3),
clouds_all integer,
weather_id integer,
weather_main varchar(50),
weather_description varchar(255),
primary key(trans_id),
foreign key(atm_id) references etl.atm(atm_id),
foreign key(location_id) references etl.loc(location_id),
foreign key(date_id) references etl.date(date_id),
foreign key(card_type_id) references etl.card(card_type_id));
```

```
38 create table etl.FACT_ATM_TRANS(
39 trans_id BIGINT not null,
40 atm_id integer,
41 location_id integer,
42 date_id integer,
43 card_type_id integer,
44 atm_status varchar(20),
45 currency varchar(10),
46 service varchar(20),
47 transaction_amount integer,
48 message_code varchar(255),
49 message_text varchar(255),
50 rain_3h NUMERIC(10,3),
51 clouds_all integer,
52 weather_id integer,
53 weather_main varchar(50),
54 weather_description varchar(255),
55 primary key(trans_id),
56 foreign key(atm_id) references etl.atm(atm_id),
57 foreign key(location_id) references etl.loc(location_id),
58 foreign key(date_id) references etl.date(date_id),
59 foreign key(card_type_id) references etl.card(card_type_id)
60 );
61
```

Run

Save

Schedule

Clear

Query results

Table details

Query



Completed, started on November 29, 2021 at 12:53:29

ELAPSED TIME: 00 m 03 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

#### 1. Query to load data into ATM dimension table.

copy etl.atm from

's3://upgrad-etl-project/DIM\_ATM/part-00000-f0b28bf8-c567-4676-812b-2aee41dc4378-c000.csv'

iam\_role 'arn:aws:iam::289714114644:role/s3\_fullaccess'

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

CSV;

```
62 copy etl.atm from
63 's3://upgrad-etl-project/DIM_ATM/part-00000-f0b28bf8-c567-4676-812b-2aee41dc4378-c000.csv'
64 iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
65 delimiter ',' region 'us-east-1' IGNOREHEADER 1
66 CSV;
67
```

Run Save Schedule Clear

Query results Table details

Query 1279 [🔗](#)

🟢 Completed, started on November 29, 2021 at 12:54:36  
ELAPSED TIME: 00 m 08 s

#### 2. Query to load data into DATE dimension table.

copy etl.date from

's3://upgrad-etl-project/DIM\_DATE/part-00000-20097453-36de-43e1-b3ea-8c8acfd92d67-c000.csv'

iam\_role 'arn:aws:iam::289714114644:role/s3\_fullaccess'

delimiter ',' IGNOREHEADER 1

timeformat 'auto'

region 'us-east-1';

```
68 copy etl.date from
69 's3://upgrad-etl-project/DIM_DATE/part-00000-20097453-36de-43e1-b3ea-8c8acfd92d67-c000.csv'
70 iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
71 delimiter ',' IGNOREHEADER 1
72 timeformat 'auto'
73 region 'us-east-1';
74
```

Run Save Schedule Clear

Query results Table details

Query 1291 [🔗](#)

🟢 Completed, started on November 29, 2021 at 12:55:03  
ELAPSED TIME: 00 m 03 s

### 3. Query to load data into LOCATION dimension table.

```
copy etl.loc from
's3://upgrad-etl-project/DIM_LOC/part-00000-22ba2432-9089-4824-8561-37246c056fd3-
c000.csv'
iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
delimiter ',' IGNOREHEADER 1
region 'us-east-1';
```

```

74
75 copy etl.loc from
76 's3://upgrad-etl-project/DIM_LOC/part-00000-22ba2432-9089-4824-8561-37246c056fd3-c000.csv'
77 iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
78 delimiter ',' IGNOREHEADER 1
79 region 'us-east-1';
80

```

Run

Save

Schedule

Clear

Query results

Table details

Query [1320](#)

Completed, started on November 29, 2021 at 12:56:51  
ELAPSED TIME: 00 m 04 s

### 4. Query to load data into CARD dimension table.

```
copy etl.card from
's3://upgrad-etl-project/DIM_CARD/part-00000-e944f841-ce2e-4175-9c53-5e5893a4ef8e-
c000.csv'
iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
delimiter ',' IGNOREHEADER 1
region 'us-east-1';
```

```

81 copy etl.card from
82 's3://upgrad-etl-project/DIM_CARD/part-00000-e944f841-ce2e-4175-9c53-5e5893a4ef8e-c000.csv'
83 iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
84 delimiter ',' IGNOREHEADER 1
85 region 'us-east-1';
86

```

Run

Save

Schedule

Clear

Query results

Table details

Query [1325](#)

Completed, started on November 29, 2021 at 12:57:36  
ELAPSED TIME: 00 m 23 s

5. Query to load data into FACT\_ATM\_TRANS table.

```
copy etl.FACT_ATM_TRANS from
's3://upgrad-etl-project/FACT_ATM_TRANS/part-00000-2b0eccf5-78f5-4bed-9d6e-
b2206b942fb2-c000.csv'
iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
delimiter ',' IGNOREHEADER 1
region 'us-east-1'
TRUNCATECOLUMNS
CSV;
```

```
88 copy etl.FACT_ATM_TRANS from
89 's3://upgrad-etl-project/FACT_ATM_TRANS/part-00000-2b0eccf5-78f5-4bed-9d6e-b2206b942fb2-c000.csv'
90 iam_role 'arn:aws:iam::289714114644:role/s3_fullaccess'
91 delimiter ',' IGNOREHEADER 1
92 region 'us-east-1'
93 TRUNCATECOLUMNS
94 CSV;
95
```

Run Save Schedule Clear

Query results Table details

**Query 1341** [🔗](#)  
✔ Completed, started on November 29, 2021 at 12:58:32  
ELAPSED TIME: 00 m 21 s