

## 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>

redshift-cluster-1

Actions

Edit

Add partner integration

Query cluster

General information

Cluster identifier

redshift-cluster-1

Cluster namespace

9304789e-e545-4b6b-a8c0-e671997b16f7

Status

Available

Date created

July 17, 2021, 08:52(UTC-04:00)

Storage used

0.07% (0.24 of 320 GB used)

Node type

dc2.large

Number of nodes

2

AQUA

Not available

Endpoint

redshift-cluster-1.cetb72igbpkc.us-east-1.re...

JDBC URL

jdbc:redshift://redshift-cluster-1.cetb72igbp...

ODBC URL

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

Cluster

Cluster namespace

Status

Storage capacity us...

CPU utilization

Snapshots

redshift-cluster-1

dc2.large | 2 nodes | 320 GB

9304789e-e545-4b6b...

Available

< 1%

3%

3 snapshots

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:**

<Queries>

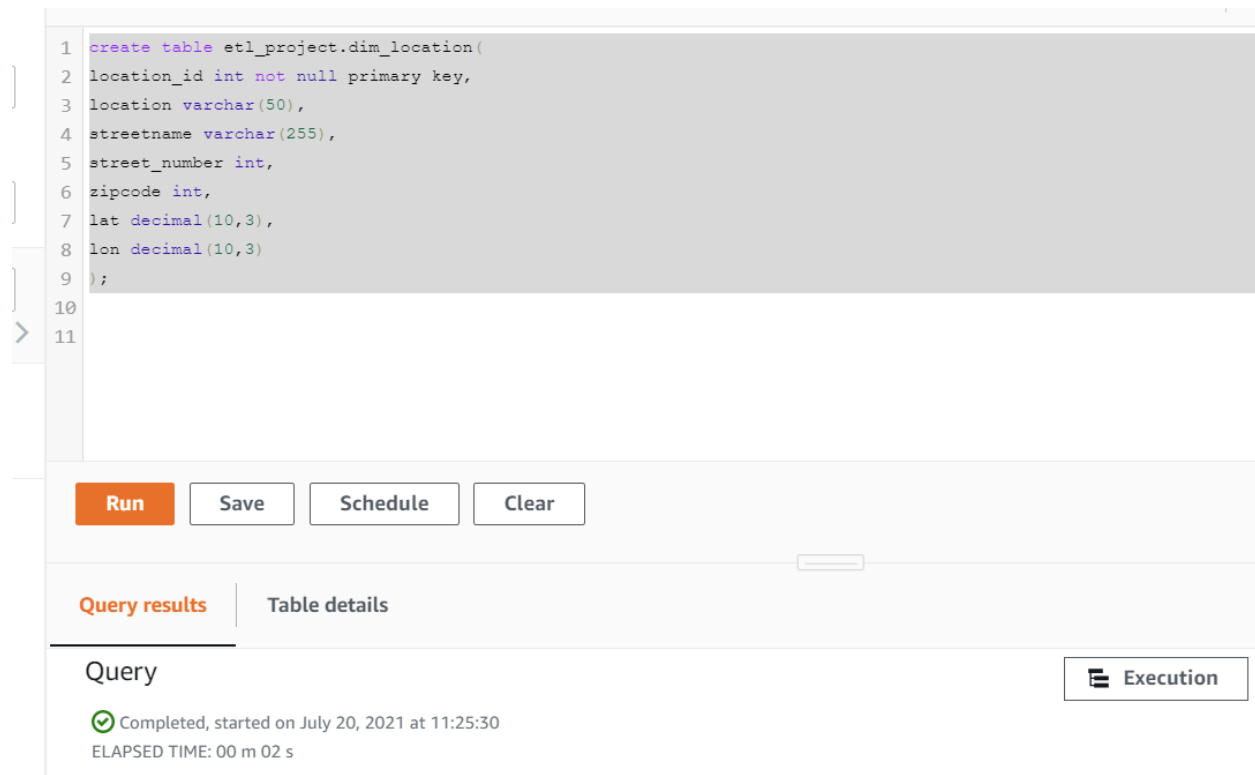
### Creating schema

-create schema etl\_project;

### Creating DIM\_Location Table

```
create table etl_project.dim_location(  
location_id int not null primary key,  
location varchar(50),  
streetname varchar(255),  
street_number int,  
zipcode int,  
lat decimal(10,3),  
lon decimal(10,3)  
);
```

### Screenshot



The screenshot displays a SQL query editor interface. The query text is as follows:

```
1 create table etl_project.dim_location(  
2 location_id int not null primary key,  
3 location varchar(50),  
4 streetname varchar(255),  
5 street_number int,  
6 zipcode int,  
7 lat decimal(10,3),  
8 lon decimal(10,3)  
9 );  
10  
11
```

Below the query editor, there are four buttons: **Run**, **Save**, **Schedule**, and **Clear**. The **Run** button is highlighted in orange.

Underneath the buttons, there are two tabs: **Query results** (active) and **Table details**.

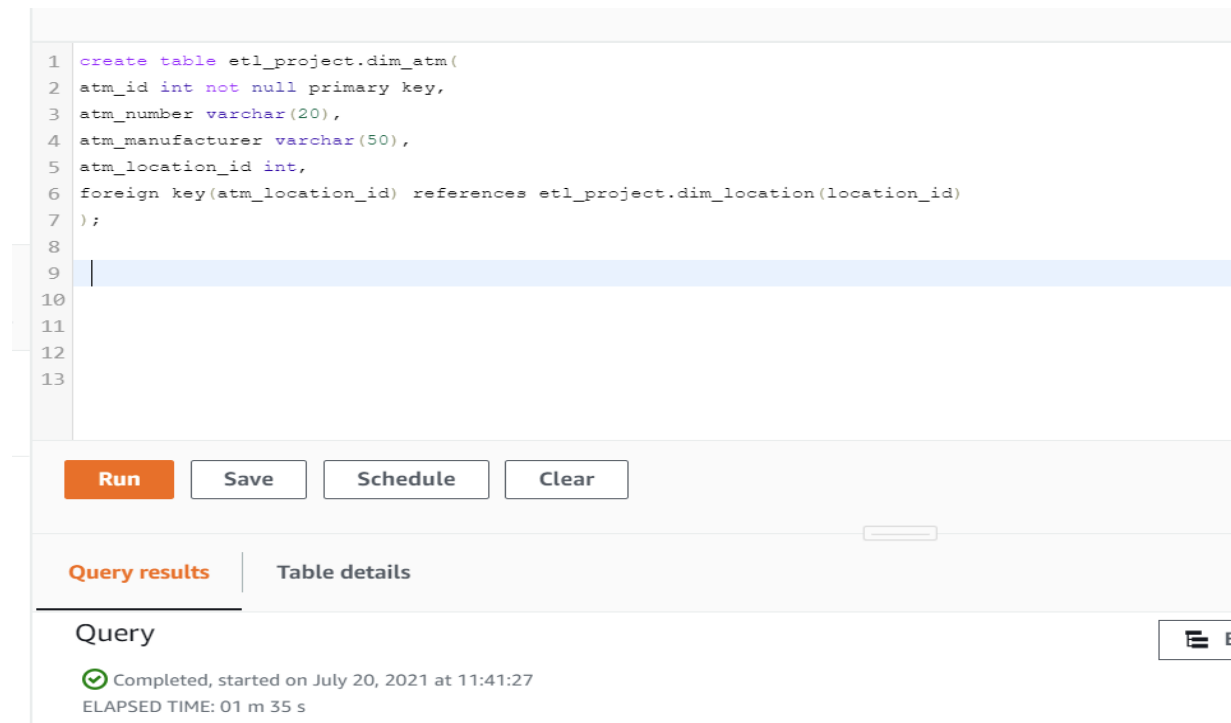
The **Query results** tab shows the following information:

- Query**
- Execution** (indicated by a green checkmark icon)
- Completed, started on July 20, 2021 at 11:25:30
- ELAPSED TIME: 00 m 02 s

### Creating DIM\_ATM

```
create table etl_project.dim_atm(  
atm_id int not null primary key,  
atm_number varchar(20),  
atm_manufacturer varchar(50),  
atm_location_id int,  
foreign key(atm_location_id) references etl_project.dim_location(location_id)  
);
```

### Screenshot



The screenshot displays a SQL query editor and execution interface. The query text is as follows:

```
1 create table etl_project.dim_atm(  
2 atm_id int not null primary key,  
3 atm_number varchar(20),  
4 atm_manufacturer varchar(50),  
5 atm_location_id int,  
6 foreign key(atm_location_id) references etl_project.dim_location(location_id)  
7 );  
8  
9  
10  
11  
12  
13
```

Below the query editor, there are four buttons: **Run** (orange), **Save**, **Schedule**, and **Clear**. A progress bar is visible below these buttons.

The interface has two tabs: **Query results** (active) and **Table details**. Under the **Query results** tab, the word **Query** is displayed. To the right of the query text, there is a button with a list icon and the letter **E**.

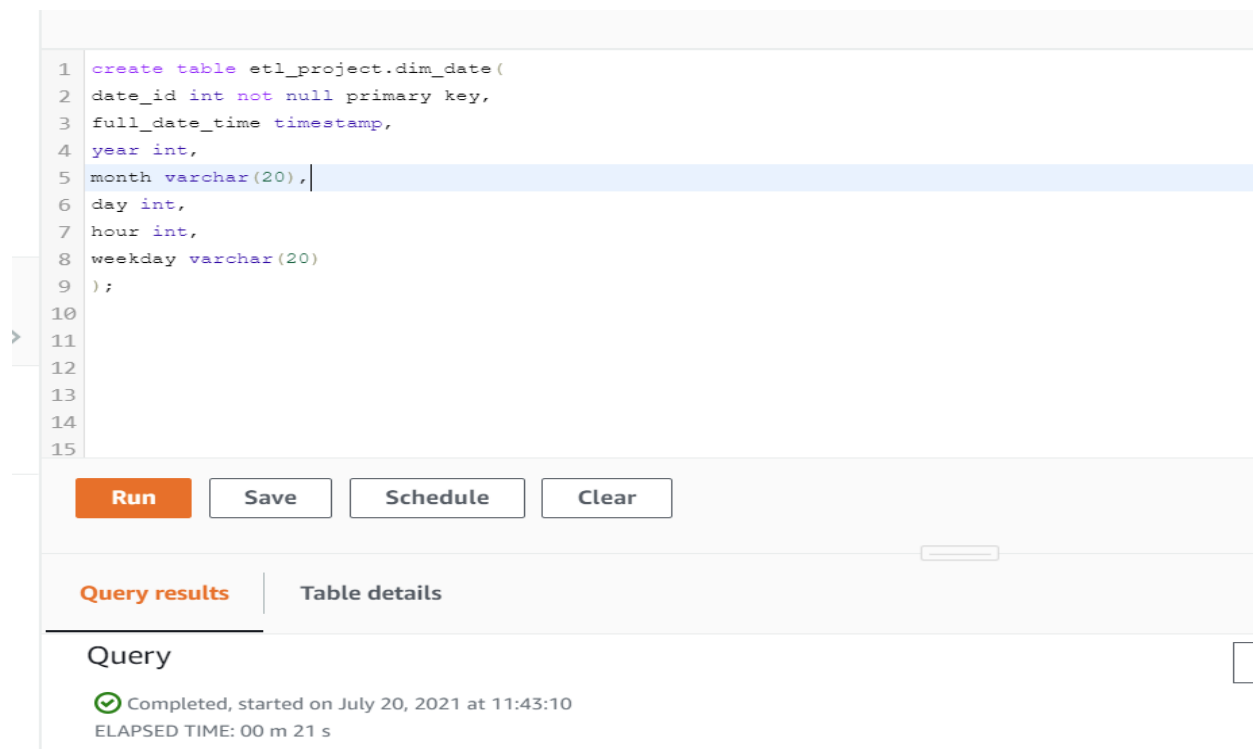
A status message indicates the query was successful: **Completed, started on July 20, 2021 at 11:41:27**. Below this, the **ELAPSED TIME: 01 m 35 s** is shown.

### Creating DIM\_DATE

```
create table etl_project.dim_date(  
date_id int not null primary key,  
full_date_time timestamp,  
year int,  
month varchar(20),  
day int,  
hour int,  
weekday varchar(20)
```

);

## Screenshot



The screenshot shows a SQL query editor interface. The query text is as follows:

```
1 create table etl_project.dim_date(  
2 date_id int not null primary key,  
3 full_date_time timestamp,  
4 year int,  
5 month varchar(20),  
6 day int,  
7 hour int,  
8 weekday varchar(20)  
9 );  
10  
11  
12  
13  
14  
15
```

Below the query editor, there are four buttons: **Run** (orange), **Save**, **Schedule**, and **Clear**. Below these buttons, there are two tabs: **Query results** (active) and **Table details**. Under the **Query results** tab, the text "Query" is displayed. Below "Query", there is a green checkmark icon followed by the text "Completed, started on July 20, 2021 at 11:43:10" and "ELAPSED TIME: 00 m 21 s".

## Creating DIM\_CARD\_TYPE

```
create table etl_project.dim_card_type(  
card_type_id int not null primary key,  
card_type varchar(50)  
);
```

## Screenshot

```
10 create table etl_project.dim_card_type(  
11 card_type_id int not null primary key,  
12 card_type varchar(50)  
13 );  
14  
15  
16  
17  
18  
19
```

**Run** **Save** **Schedule** **Clear**

**Query results** | **Table details**

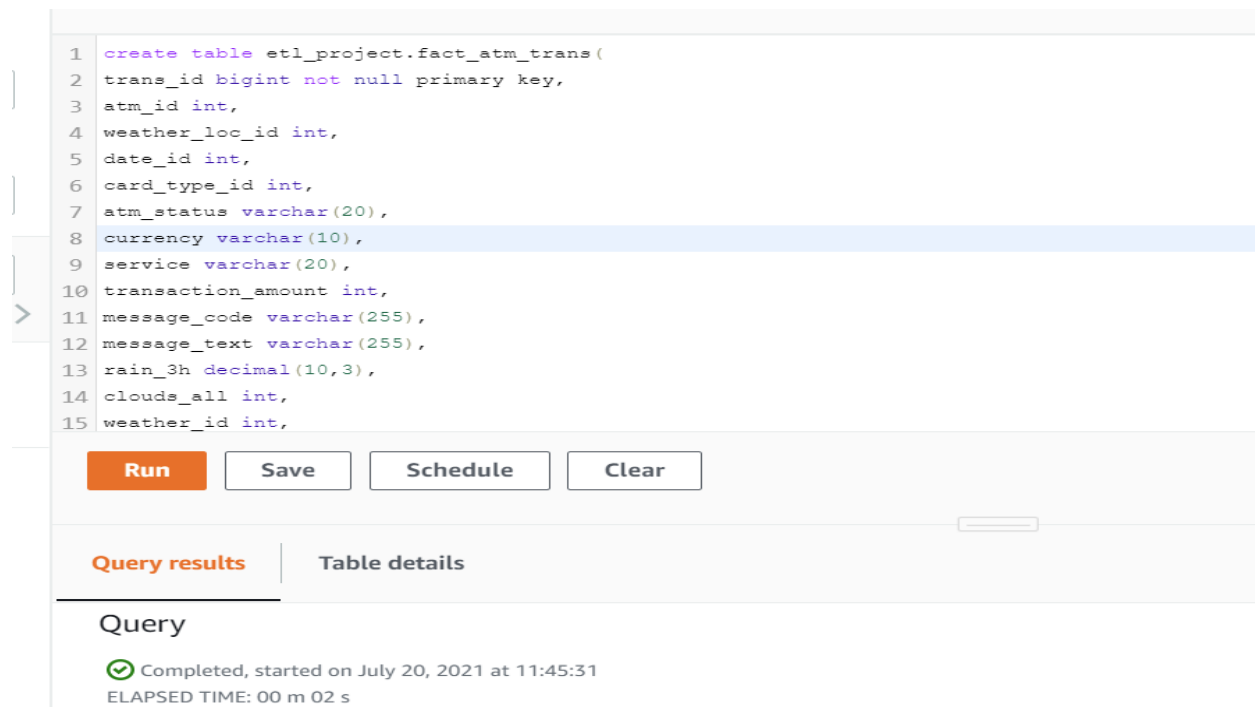
**Query**

✓ Completed, started on July 20, 2021 at 11:58:14  
ELAPSED TIME: 00 m 09 s

## Creating FACT\_ATM\_TRANS

```
create table etl_project.fact_atm_trans(  
trans_id bigint not null primary key,  
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(255),  
message_text varchar(255),  
rain_3h decimal(10,3),  
clouds_all int,  
weather_id int,  
weather_main varchar(50),  
weather_description varchar(255)  
);
```

## Screenshot



```

1 create table etl_project.fact_atm_trans (
2 trans_id bigint not null primary key,
3 atm_id int,
4 weather_loc_id int,
5 date_id int,
6 card_type_id int,
7 atm_status varchar(20),
8 currency varchar(10),
9 service varchar(20),
10 transaction_amount int,
11 message_code varchar(255),
12 message_text varchar(255),
13 rain_3h decimal(10,3),
14 clouds_all int,
15 weather_id int,

```

**Run** **Save** **Schedule** **Clear**

**Query results** **Table details**

**Query**

✓ Completed, started on July 20, 2021 at 11:45:31  
ELAPSED TIME: 00 m 02 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**

<Queries>

### Copying into DIM\_LOCATION Table

copy etl\_project.dim\_location from

's3://swathimss3bucket/ETL\_Project/LOCATION\_DIMENSION/LOCATION\_DIMENSION.csv'

iam\_role 'arn:aws:iam::063189671810:role/redshift\_s3\_fullaccess' delimiter ',' ignoreheader 1  
region 'us-east-1';

## Screenshot

```

1 copy etl_project.dim_location from
2 's3://swathimss3bucket/ETL_Project/LOCATION_DIMENSION/LOCATION_DIMENSION.csv/'
3 iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1 region 'us-east-1';
4
5
6
7
8
9
10
11

```

Run

Save

Schedule

Clear

Query results

Table details

Query 361

Execution

Completed, started on July 20, 2021 at 11:48:19

ELAPSED TIME: 00 m 48 s

## Copying into DIM\_ATM Table

```

copy etl_project.dim_atm from
's3://swathimss3bucket/ETL_Project/ATM_DIMENSION/ATM_DIMENSION.csv/'
iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1
region 'us-east-1';

```

## Screenshot

```
copy etl_project.dim_atm from  
's3://swathimss3bucket/ETL_Project/ATM_DIMENSION/ATM_DIMENSION.csv/'  
iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1 region 'us-east-1';
```

Run

Save

Schedule

Clear

Query results

Table details

Query [407](#)

Execution



Completed, started on July 20, 2021 at 11:51:25

ELAPSED TIME: 00 m 14 s

## Copying into DIM\_DATE Table

copy etl\_project.dim\_date from

's3://swathimss3bucket/ETL\_Project/DATETIME\_DIMENSION/DATETIME\_DIMENSION/'

iam\_role 'arn:aws:iam::063189671810:role/redshift\_s3\_fullaccess' delimiter ',' ignoreheader 1  
region 'us-east-1';



## Screenshot

5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15

```

copy etl_project.dim_date from
's3://swathimss3bucket/ETL_Project/DATETIME_DIMENSION/DATETIME_DIMENSION/'
iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1 region 'us-east-1';

```

Run

Save

Schedule

Clear

Query results

Table details

Query [431](#)

Execution

Completed, started on July 20, 2021 at 11:53:49

ELAPSED TIME: 00 m 21 s

## Copying into DIM\_CARD\_TYPE Table

```

copy etl_project.dim_card_type from
's3://swathimss3bucket/ETL_Project/CARD_TYPE_DIMENSION/CARD_TYPE_DIMENSION.csv'
iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1
region 'us-east-1';

```

## Screenshot

```

5
6 copy etl_project.dim_card_type from
7 's3://swathimss3bucket/ETL_Project/CARD_TYPE_DIMENSION/CARD_TYPE_DIMENSION.csv/'
8 iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1 region 'us-east-1';
9
10
11
12
13
14
15
16
17
18
19


```

Run Save Schedule Clear

Query results Table details

Query 493 [↗](#)

Execution

 Completed, started on July 20, 2021 at 11:59:09  
ELAPSED TIME: 00 m 14 s

## Copying into FACT\_ATM\_TRANS Table

```

copy etl_project.fact_atm_trans from
's3://swathimss3bucket/ETL_Project/FACT_ATM_TRANS/FACT_ATM_TRANS/'
iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1
quote '"' region 'us-east-1' csv;

```

## Screenshot

```

14
15 copy etl_project.fact_atm_trans from
16 's3://swathimss3bucket/ETL_Project/FACT_ATM_TRANS/FACT_ATM_TRANS/'
17 iam_role 'arn:aws:iam::063189671810:role/redshift_s3_fullaccess' delimiter ',' ignoreheader 1 quote '"' region 'us-east-1' csv;
18
19
20
21

```

Run Save Schedule Clear Send

Query results Table details

Query 538 [🔗](#) Execution Data

✓ Completed, started on July 20, 2021 at 12:03:21  
ELAPSED TIME: 01 m 04 s

## Screenshot of Dim\_Location count

```

5 select count(*) from etl_project.dim_location;
6
7
8
9
10
11

```

Run Save Schedule Clear

Query results Table details

Query 384 [🔗](#) Exec

✓ Completed, started on July 20, 2021 at 11:49:16  
ELAPSED TIME: 00 m 01 s

Rows returned (1)

Search rows

count
109

## Screenshot of Dim\_Atм Count

```

12 select count(*) from etl_project.dim_atm;
13
14
15

```

**Run** **Save** **Schedule** **Clear**

**Query results** | **Table details**

Query 418 [🔗](#)

✅ Completed, started on July 20, 2021 at 11:52:12  
ELAPSED TIME: 00 m 02 s

**Rows returned (1)**

count
156

## Screenshot of Dim\_Date count

```

13
14 select count(*) from etl_project.dim_date;
15
16

```

**Run** **Save** **Schedule** **Clear**

**Query results** | **Table details**

Query 442 [🔗](#)

✅ Completed, started on July 20, 2021 at 11:54:32  
ELAPSED TIME: 00 m 01 s

**Rows returned (1)**

count
8685

## Screenshot of Dim\_Card\_Type count

```

11
12 select count(*) from etl_project.dim_card_type;
13
14
15
16
17
18
19

```

**Run** **Save** **Schedule** **Clear**

**Query results** | **Table details**

Query [507](#)

✓ Completed, started on July 20, 2021 at 12:00:28  
ELAPSED TIME: 00 m 01 s

**Rows returned (1)**

🔍 Search rows

count
12

## Screenshot of Fact\_ATM\_Trans count

```

19
20 select count(*) from etl_project.fact_atm_trans;
21
22
23
24
25

```

**Run** **Save** **Schedule** **Clear**

**Query results** | **Table details**

Query [557](#)

✓ Completed, started on July 20, 2021 at 12:04:10  
ELAPSED TIME: 00 m 01 s

**Rows returned (1)**

🔍 Search rows

count
2468571

## Screenshot of the Schema Created

TO view tables, select a schema.

etl\_project ▼

Q Filter tables

< 1 >

▶ dim_atm_pkey	...
▶ dim_card_type_pkey	...
▶ dim_date_pkey	...
▶ dim_location_pkey	...
▶ fact_atm_trans_pkey	...
▶ dim_atm	...
▶ dim_card_type	...
▶ dim_date	...
▶ dim_location	...
▶ fact_atm_trans	...