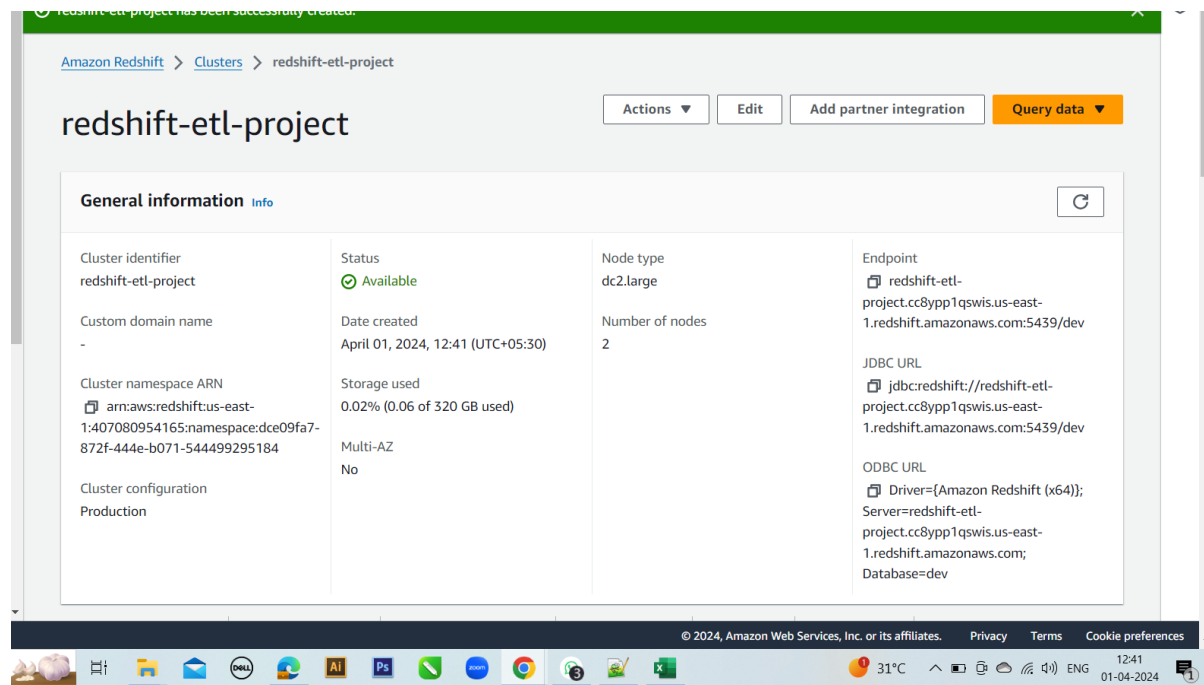


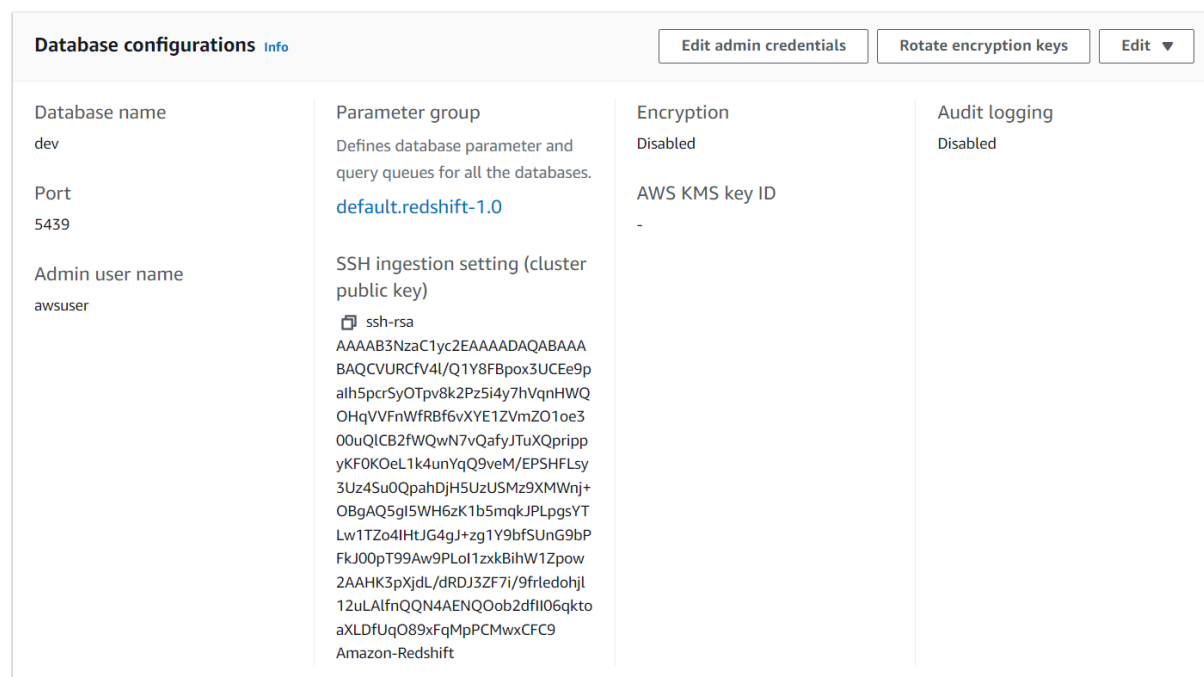
Creation of a Redshift Cluster

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



The screenshot shows the Amazon Redshift console interface for a cluster named 'redshift-etl-project'. The cluster is in an 'Available' state. The configuration details are as follows:

General information			
Cluster identifier redshift-etl-project	Status Available	Node type dc2.large	Endpoint redshift-etl-project.cc8ypp1qswis.us-east-1.redshift.amazonaws.com:5439/dev
Custom domain name -	Date created April 01, 2024, 12:41 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-etl-project.cc8ypp1qswis.us-east-1.redshift.amazonaws.com:5439/dev
Cluster namespace ARN arn:aws:redshift:us-east-1:407080954165:namespace:dce09fa7-872f-444e-b071-544499295184	Storage used 0.02% (0.06 of 320 GB used)		ODBC URL Driver={Amazon Redshift (x64)}; Server=redshift-etl-project.cc8ypp1qswis.us-east-1.redshift.amazonaws.com; Database=dev
Cluster configuration Production	Multi-AZ No		



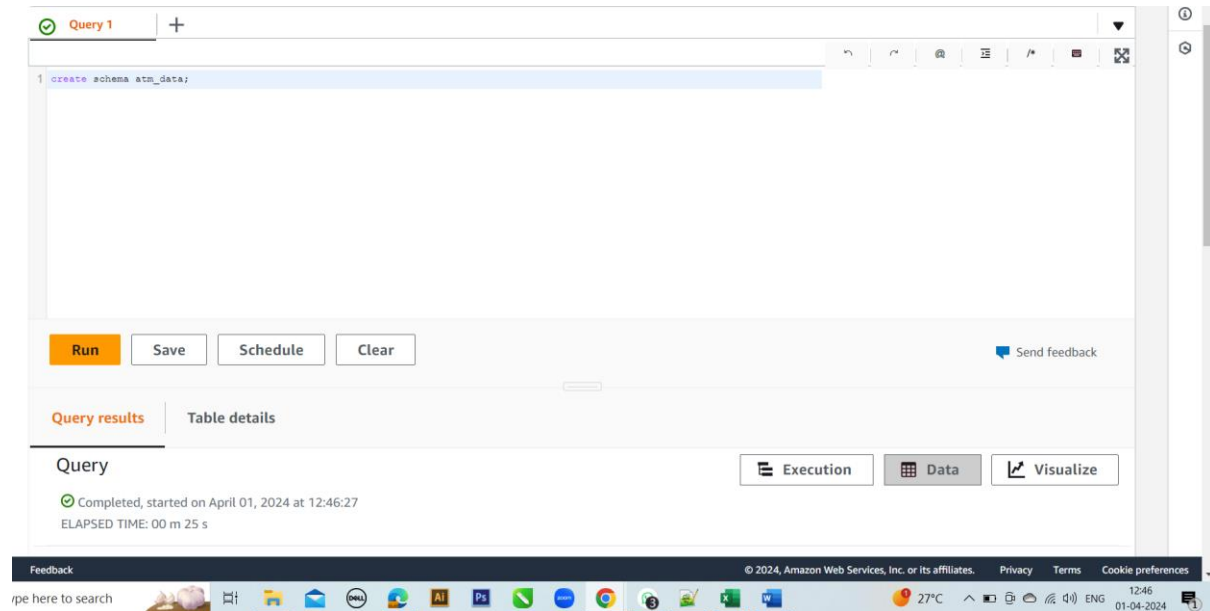
The screenshot shows the 'Database configurations' section of the Amazon Redshift console. The configurations are as follows:

Database configurations			
Database name dev	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5439	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQBAQCVURCFV4I/Q1Y8FBpox3UCEe9p alh5PcrSyOTpv8k2Pz5i4y7hVqnHWQ OHqVVFfWfRBF6vXEY1ZVvmZO1oe3 00uQICB2FWQwN7vQafyJTUxQprpp yKF0KOeL1k4unYqQ9veM/EPShFLsy 3Uz4Su0QpahDjH5UzUSMz9XMWnj+ OBgAQ5gl5WH6zK1b5mqkJPLpgsYT Lw1TZo4IHtJG4gJ+zg1Y9bfSUnG9bP FkJ00pT99Aw9PLol1zxkBiHw1Zpow 2AAHK3pXjL/dRDJ3ZF7i/9frledohjl 12uLAlfnQQN4AENQOob2dfI06qkto aXLDfUqO89xFqMpPCMwxCFc9 Amazon-Redshift	AWS KMS key ID -	
Admin user name awsuser			

Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Creation of schema atm_data on the Redshift cluster

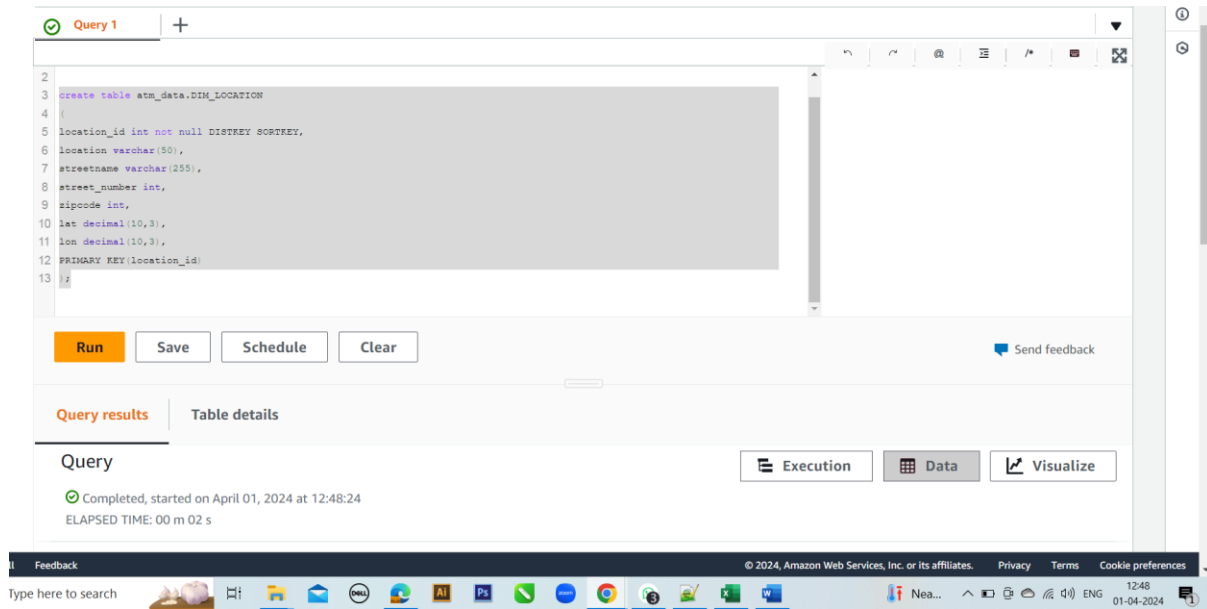
- create schema atm_data;



Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

Creation of location dimension table DIM_LOCATION

- 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)
);



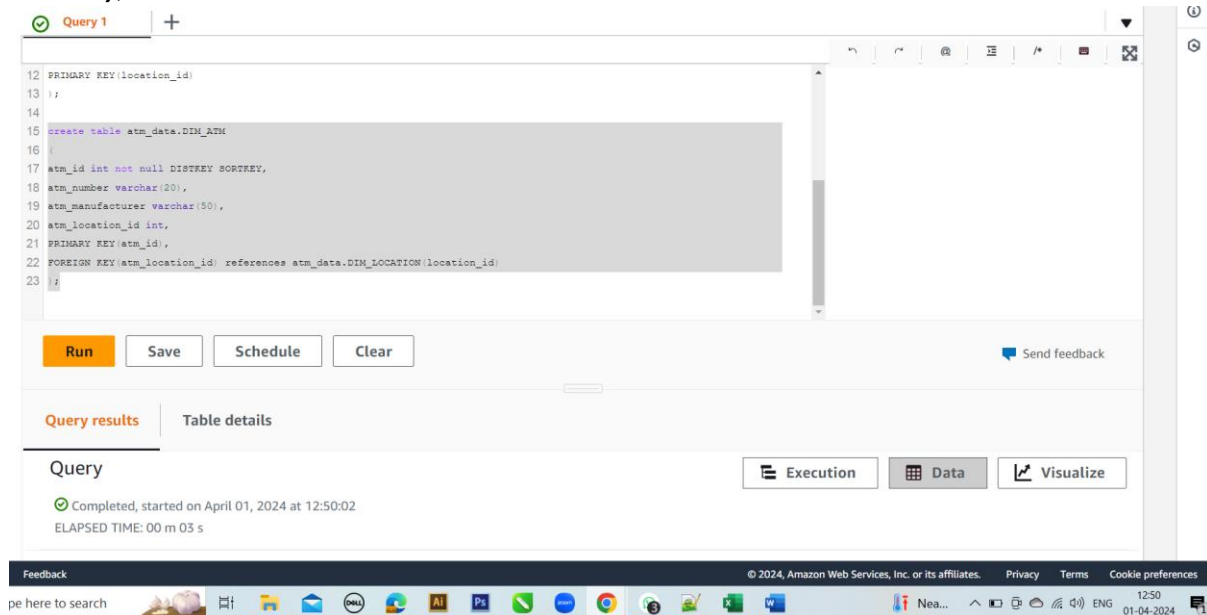
The screenshot shows the Amazon Redshift console interface. At the top, there's a tab labeled "Query 1". Below it, a SQL query is entered in a text area:

```
2
3 create table atm_data.DIM_LOCATION
4 (
5 location_id int not null DISTKEY SORTKEY,
6 location varchar(50),
7 streetname varchar(255),
8 street_number int,
9 zipcode int,
10 lat decimal(10,3),
11 lon decimal(10,3),
12 PRIMARY KEY(location_id)
13 );
```

Below the query area, there are buttons for "Run", "Save", "Schedule", and "Clear". To the right of these buttons is a "Send feedback" link. Below the buttons, there are two tabs: "Query results" (selected) and "Table details". Under the "Query results" tab, it shows "Query" and "Execution" buttons. Below that, it says "Completed, started on April 01, 2024 at 12:48:24" and "ELAPSED TIME: 00 m 02 s". At the bottom of the console, there's a taskbar with various application icons and a system clock showing 12:48 on 01-04-2024.

Creation of atm dimension table DIM_ATM

- **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)
);



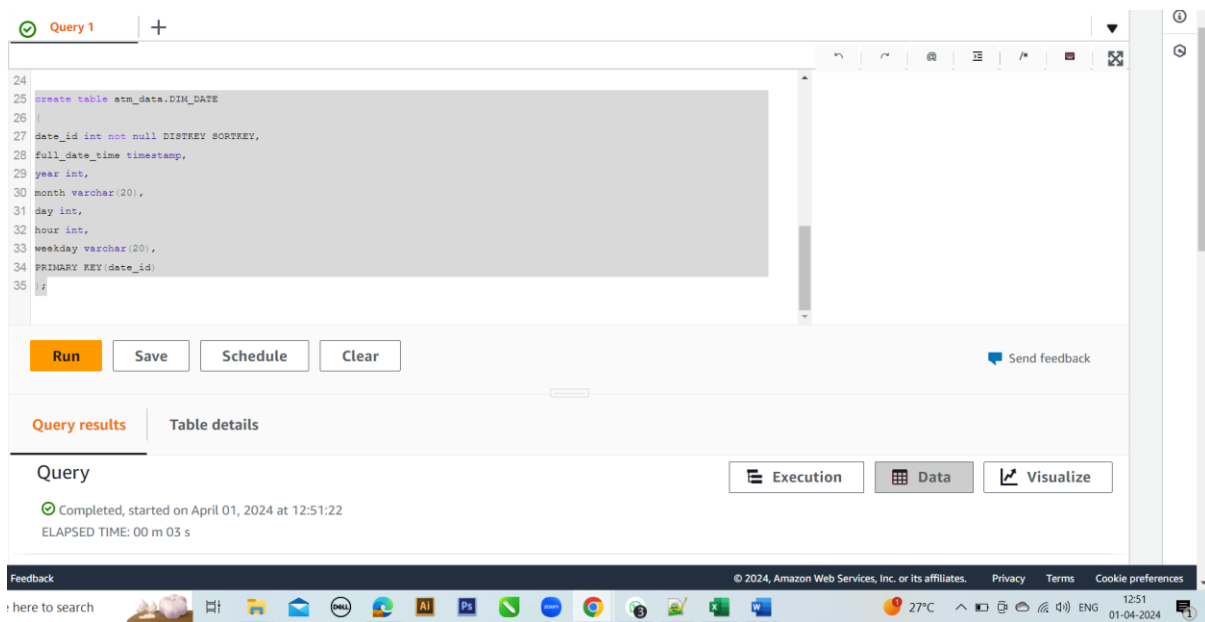
The screenshot shows the Amazon Redshift console interface. At the top, there's a tab labeled "Query 1". Below it, a SQL query is entered in a text area:

```
12 PRIMARY KEY(location_id)
13 );
14
15 create table atm_data.DIM_ATM
16 (
17 atm_id int not null DISTKEY SORTKEY,
18 atm_number varchar(20),
19 atm_manufacturer varchar(50),
20 atm_location_id int,
21 PRIMARY KEY(atm_id),
22 FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
23 );
```

Below the query area, there are buttons for "Run", "Save", "Schedule", and "Clear". To the right of these buttons is a "Send feedback" link. Below the buttons, there are two tabs: "Query results" (selected) and "Table details". Under the "Query results" tab, it shows "Query" and "Execution" buttons. Below that, it says "Completed, started on April 01, 2024 at 12:50:02" and "ELAPSED TIME: 00 m 03 s". At the bottom of the console, there's a taskbar with various application icons and a system clock showing 12:50 on 01-04-2024.

Creation of date dimension table DIM_DATE

- **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)
);



The screenshot displays a query execution interface. The query editor shows the following SQL code:

```

24
25 create table atm_data.DIM_DATE
26
27 date_id int not null DISTKEY SORTKEY,
28 full_date_time timestamp,
29 year int,
30 month varchar(20),
31 day int,
32 hour int,
33 weekday varchar(20),
34 PRIMARY KEY(date_id)
35 ;
  
```

Below the query editor, there are buttons for **Run**, **Save**, **Schedule**, and **Clear**. A **Send feedback** link is also present.

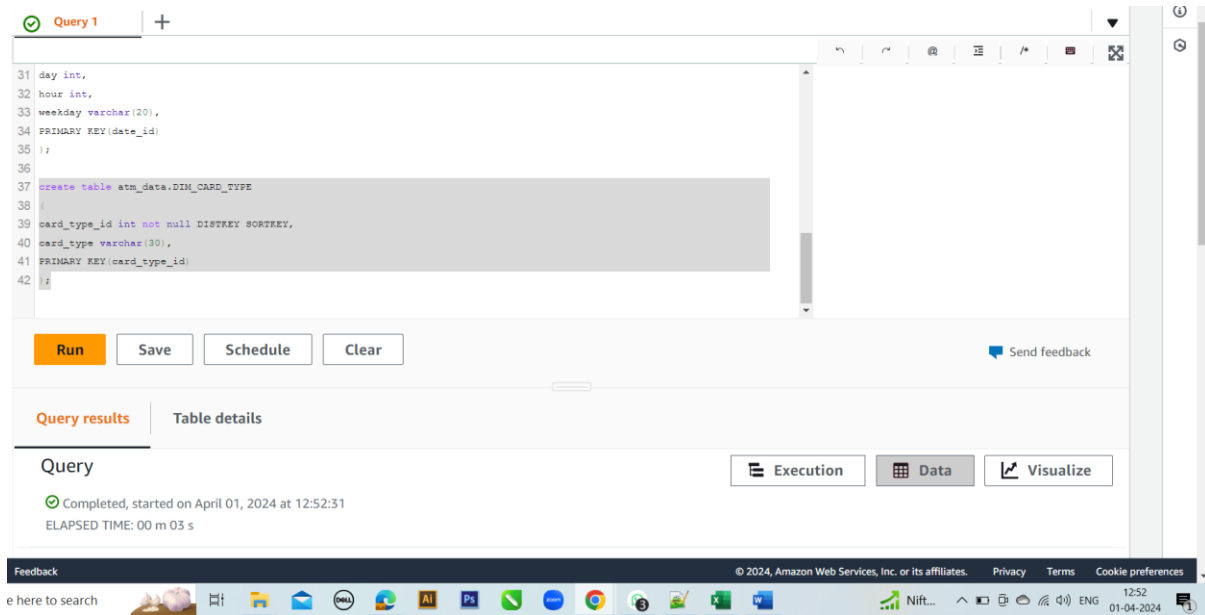
The **Query results** tab is active, showing the following information:

- Query**
- Completed, started on April 01, 2024 at 12:51:22
- ELAPSED TIME: 00 m 03 s

At the bottom of the interface, there are tabs for **Execution**, **Data**, and **Visualize**. The footer includes a feedback link, copyright information for Amazon Web Services, Inc. or its affiliates, and links for **Privacy**, **Terms**, and **Cookie preferences**.

Creation of card type dimension table DIM_CARD_TYPE

- **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 SQL query editor with a query titled "Query 1". The query is as follows:

```

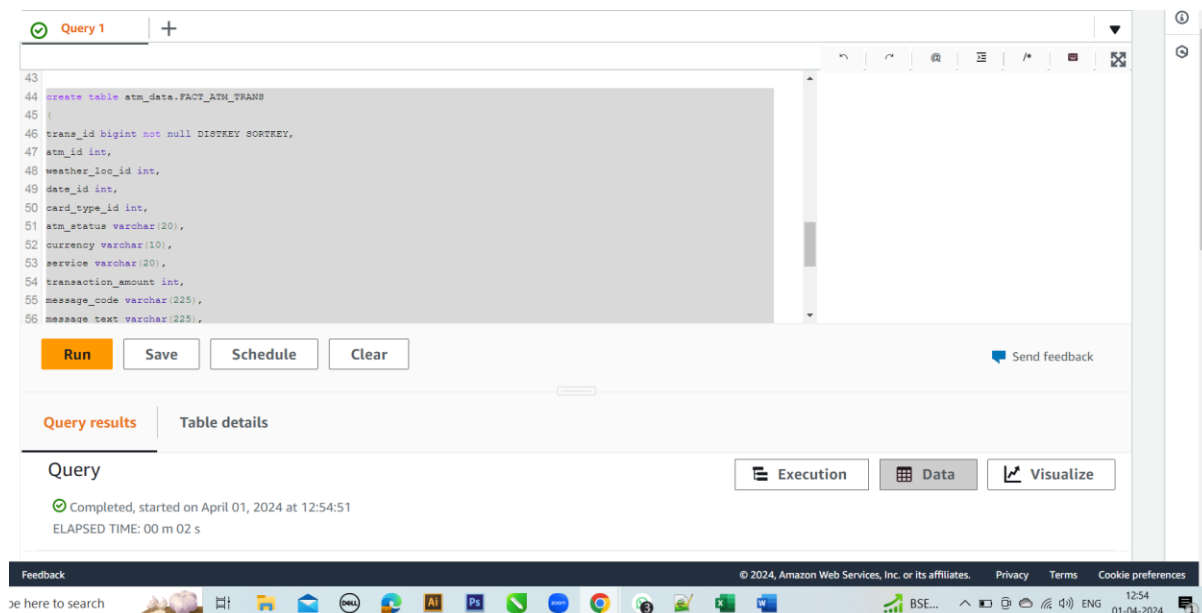
31 day int,
32 hour int,
33 weekday varchar(20),
34 PRIMARY KEY(date_id)
35 );
36
37 create table atm_data.DIM_CARD_TYPE
38 (
39 card_type_id int not null DISTKEY SORTKEY,
40 card_type varchar(30),
41 PRIMARY KEY(card_type_id)
42 );

```

Below the query editor, there are buttons for "Run", "Save", "Schedule", and "Clear". A "Send feedback" link is also present. The "Query results" tab is selected, showing a status of "Completed, started on April 01, 2024 at 12:52:31" and "ELAPSED TIME: 00 m 03 s". The "Table details" tab is also visible. The bottom of the interface shows a taskbar with various application icons and a system tray with the date "01-04-2024" and time "12:52".

Creation of atm transaction fact table FACT_ATM_TRANS

- **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)
);



Query 1

```

43
44 create table atm_data.FACT_ATM_TRANS
45 (
46 trans_id bigint not null DISTKEY SORTKEY,
47 atm_id int,
48 weather_loc_id int,
49 date_id int,
50 card_type_id int,
51 atm_status varchar(20),
52 currency varchar(10),
53 service varchar(20),
54 transaction_amount int,
55 message_code varchar(225),
56 message_text varchar(225),

```

Run Save Schedule Clear Send feedback

Query results Table details

Query

Completed, started on April 01, 2024 at 12:54:51
ELAPSED TIME: 00 m 02 s

Execution Data Visualize

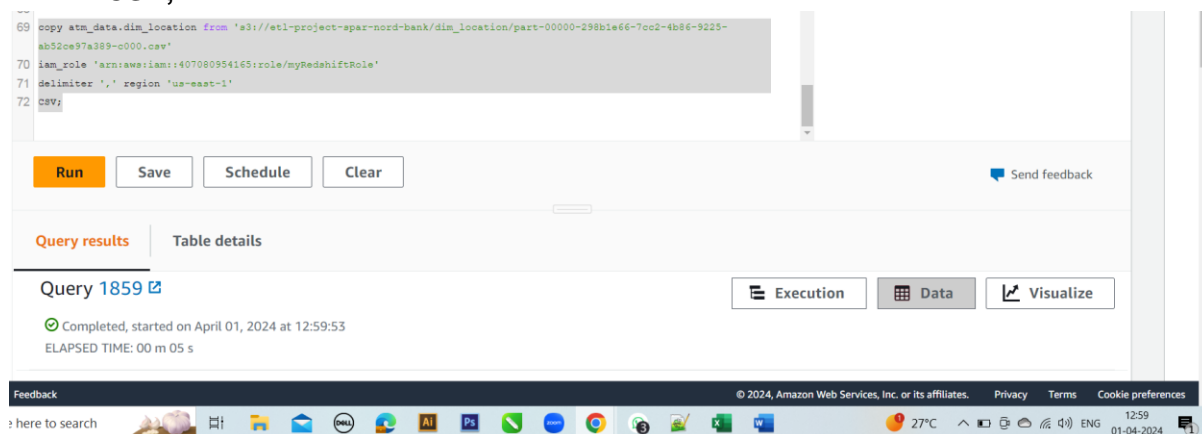
Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 query for DIM_LOCATION table

- **copy atm_data.dim_location from 's3://etl-project-spar-nord-bank/dim_location/part-00000-298b1e66-7cc2-4b86-9225-ab52ce97a389-c000.csv' iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;**



```

69 copy atm_data.dim_location from 's3://etl-project-spar-nord-bank/dim_location/part-00000-298b1e66-7cc2-4b86-9225-
70 ab52ce97a389-c000.csv'
71 iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
72 delimiter ',' region 'us-east-1'
73 CSV;

```

Run Save Schedule Clear Send feedback

Query results Table details

Query 1859

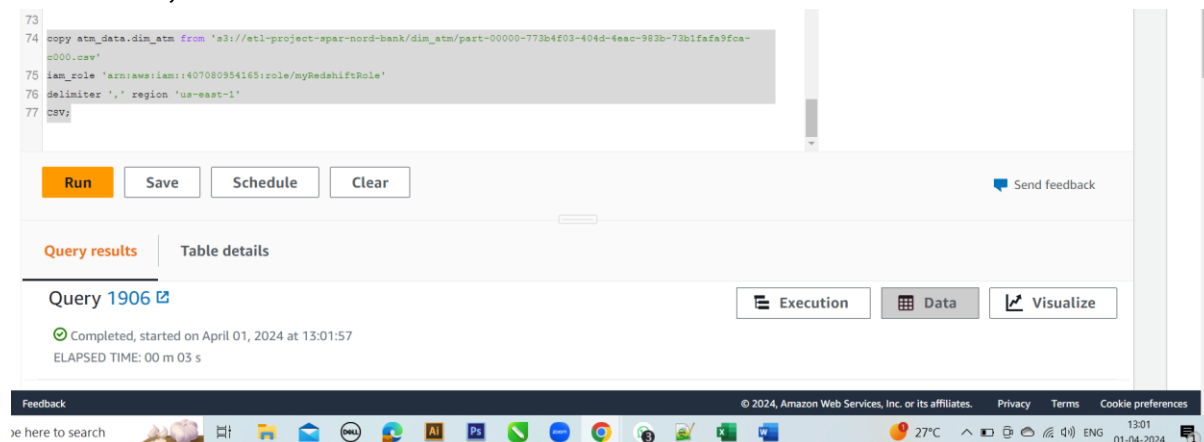
Completed, started on April 01, 2024 at 12:59:53
ELAPSED TIME: 00 m 05 s

Execution Data Visualize

Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Copy query for DIM_ATM table

- **copy atm_data.dim_atm from 's3://etl-project-spar-nord-bank/dim_atm/part-00000-773b4f03-404d-4eac-983b-73b1fafa9fca-c000.csv'**
iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



73
74 copy atm_data.dim_atm from 's3://etl-project-spar-nord-bank/dim_atm/part-00000-773b4f03-404d-4eac-983b-73b1fafa9fca-c000.csv'
75 iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
76 delimiter ',' region 'us-east-1'
77 CSV;

Run Save Schedule Clear Send feedback

Query results Table details

Query 1906

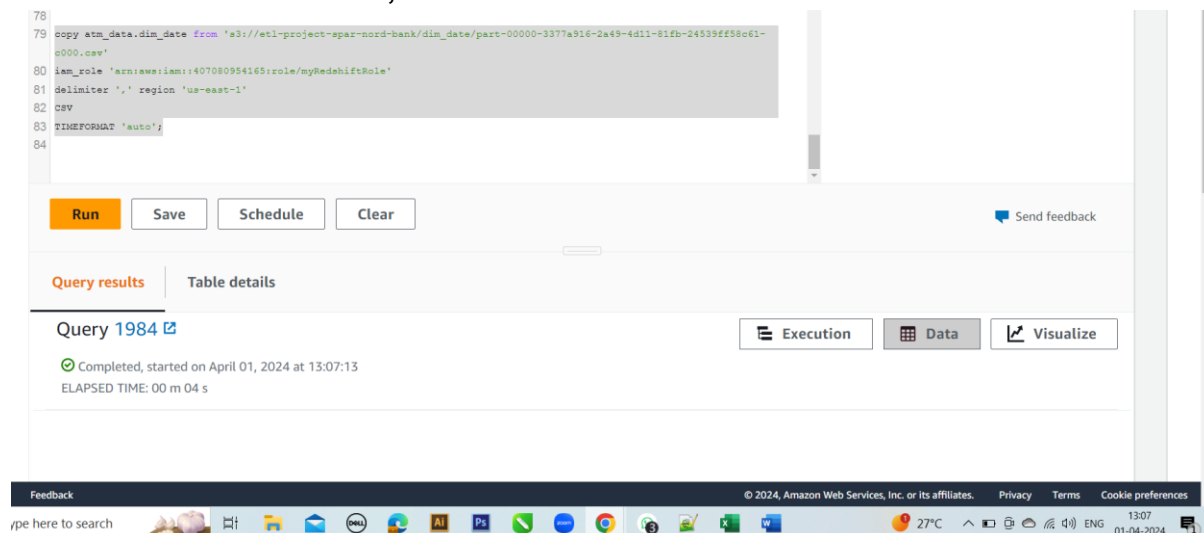
Completed, started on April 01, 2024 at 13:01:57
ELAPSED TIME: 00 m 03 s

Execution Data Visualize

Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Copy query for DIM_DATE table

- **copy atm_data.dim_date from 's3://etl-project-spar-nord-bank/dim_date/part-00000-3377a916-2a49-4d11-81fb-24539ff58c61-c000.csv'**
iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';



78
79 copy atm_data.dim_date from 's3://etl-project-spar-nord-bank/dim_date/part-00000-3377a916-2a49-4d11-81fb-24539ff58c61-c000.csv'
80 iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
81 delimiter ',' region 'us-east-1'
82 CSV
83 TIMEFORMAT 'auto';
84

Run Save Schedule Clear Send feedback

Query results Table details

Query 1984

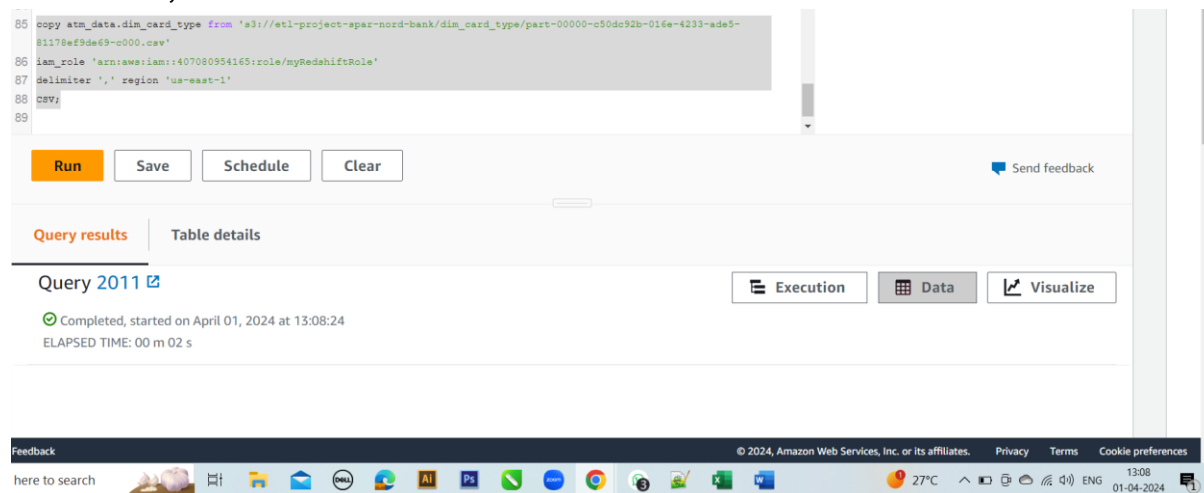
Completed, started on April 01, 2024 at 13:07:13
ELAPSED TIME: 00 m 04 s

Execution Data Visualize

Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Copy query for DIM_CARD_TYPE table

- **copy atm_data.dim_card_type from 's3://etl-project-spar-nord-bank/dim_card_type/part-00000-c50dc92b-016e-4233-ade5-81178ef9de69-c000.csv'**
iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



Query 2011

Completed, started on April 01, 2024 at 13:08:24
ELAPSED TIME: 00 m 02 s

Execution Data Visualize

Feedback

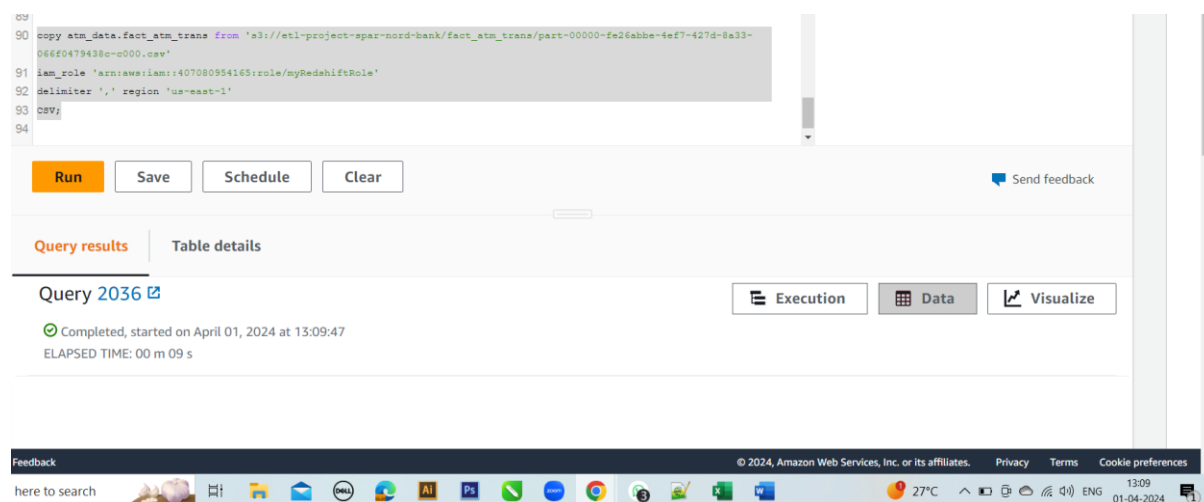
here to search

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

27°C 13:08 01-04-2024

Copy query for FACT_ATM_TRANS table

- **copy atm_data.fact_atm_trans from 's3://etl-project-spar-nord-bank/fact_atm_trans/part-00000-fe26abbe-4ef7-427d-8a33-066f0479438c-c000.csv'**
iam_role 'arn:aws:iam::407080954165:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



Query 2036

Completed, started on April 01, 2024 at 13:09:47
ELAPSED TIME: 00 m 09 s

Execution Data Visualize

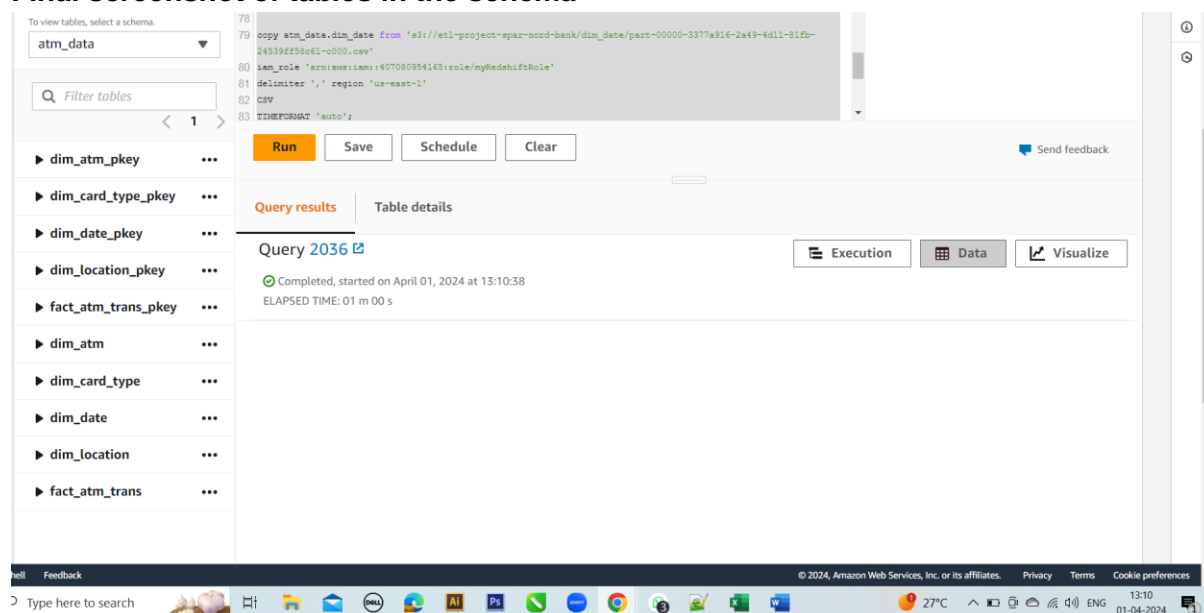
Feedback

here to search

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

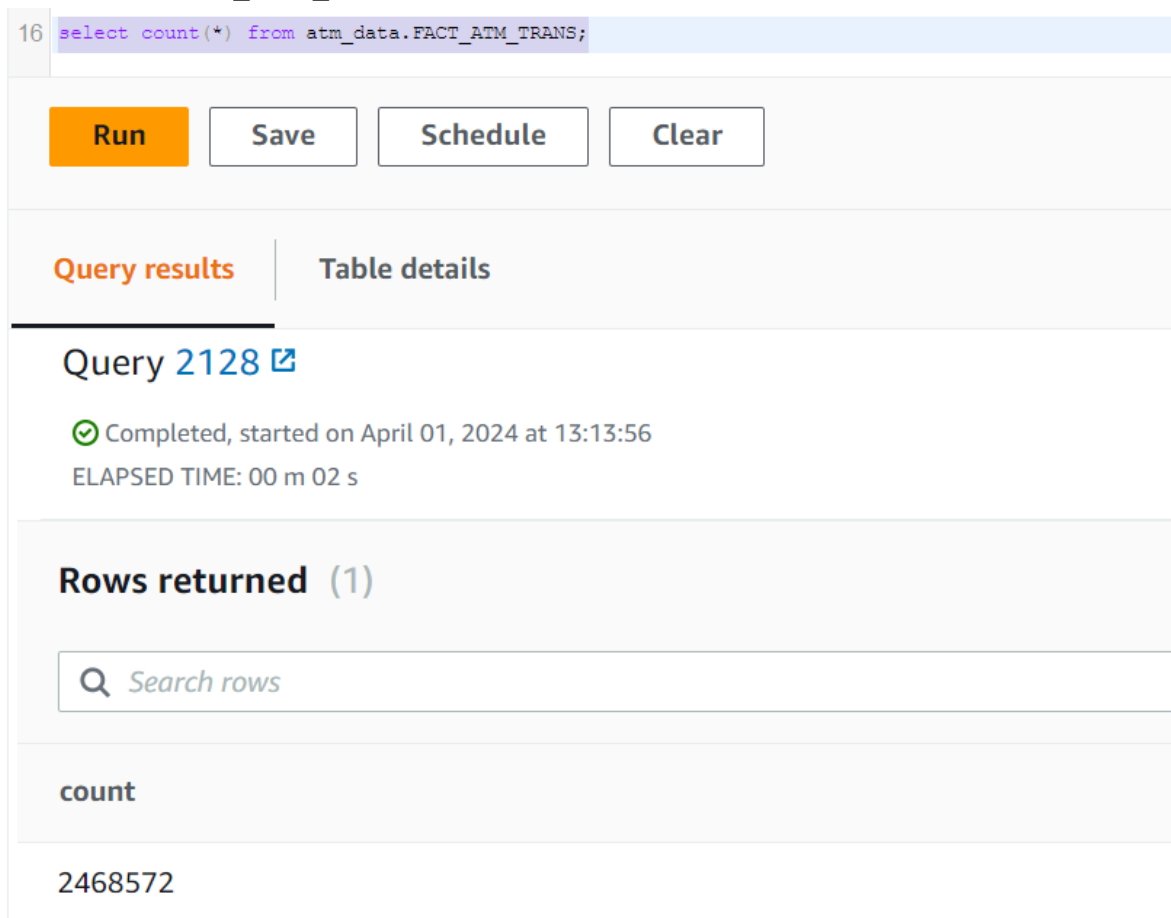
27°C 13:09 01-04-2024

Final screenshot of tables in the schema



Count verification in the tables

- Count of FACT_ATM_TRANS



16 `select count(*) from atm_data.FACT_ATM_TRANS;`

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

Query results | **Table details**

Query 2128

Completed, started on April 01, 2024 at 13:13:56
ELAPSED TIME: 00 m 02 s

Rows returned (1)

count
2468572

- Count of DIM_DATE

15

16

```
select count(*) from atm_data.DIM_DATE;
```

Run

Save

Schedule

Clear

Query results

Table details

Query 2144 [🔗](#)

✔ Completed, started on April 01, 2024 at 13:14:54

ELAPSED TIME: 00 m 02 s

Rows returned (1)

🔍 Search rows

count
8685

- Count of DIM_CARD_TYPE

15

16 `select count(*) from atm_data.DIM_CARD_TYPE;`

Run

Save

Schedule

Clear

Query results

Table details

Query 2158

✔

Completed, started on April 01, 2024 at 13:15:32

ELAPSED TIME: 00 m 02 s

Rows returned (1)

Search rows

count
12

- Count of DIM_ATM

15

16 `select count(*) from atm_data.DIM_ATM;`

Run

Save

Schedule

Clear

Query results

Table details

Query 2172 [↗](#)

✔ Completed, started on April 01, 2024 at 13:16:20

ELAPSED TIME: 00 m 03 s

Rows returned (1)

🔍

Search rows

count
113

- Count of DIM_LOCATION

```
16 select count(*) from atm_data.DIM_LOCATION;
```

Run

Save

Schedule

Clear

Query results

Table details

Query 2182 [🔗](#)

✅ Completed, started on April 01, 2024 at 13:16:48

ELAPSED TIME: 00 m 03 s

Rows returned (1)

🔍 *Search rows*

count

109

***** Thank you *****