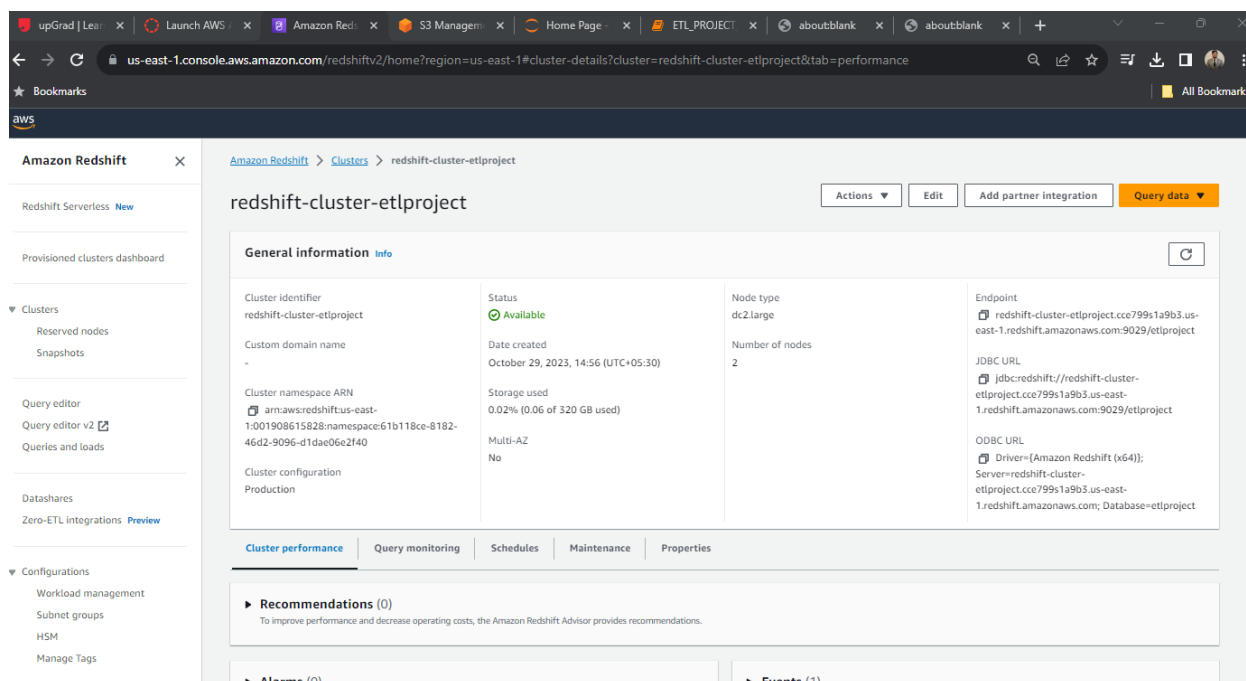


## Creation of a Redshift Cluster

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

Screenshot of the type of machine used along with number of nodes



The screenshot displays the AWS Redshift console interface. The left sidebar shows the navigation menu with options like 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Datashares', 'Configurations', and 'Workload management'. The main content area shows the details for the cluster 'redshift-cluster-etlproject'.

**General information**

Cluster identifier redshift-cluster-etlproject	Status Available	Node type dc2.large	Endpoint redshift-cluster-etlproject.cce799s1a9b3.us-east-1.redshift.amazonaws.com:9029/etlproject
Custom domain name -	Date created October 29, 2023, 14:56 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-etlproject.cce799s1a9b3.us-east-1.redshift.amazonaws.com:9029/etlproject
Cluster namespace ARN arn:aws:redshift:us-east-1:001908615828:namespace:61b118ce-8182-46d2-9096-d1dae06e2f40	Storage used 0.02% (0.06 of 320 GB used)		ODBC URL Driver=(Amazon Redshift (x64)); Server=redshift-cluster-etlproject.cce799s1a9b3.us-east-1.redshift.amazonaws.com; Database=etlproject
Cluster configuration Production	Multi-AZ No		

**Cluster performance** | Query monitoring | Schedules | Maintenance | Properties

**Recommendations (0)**  
To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

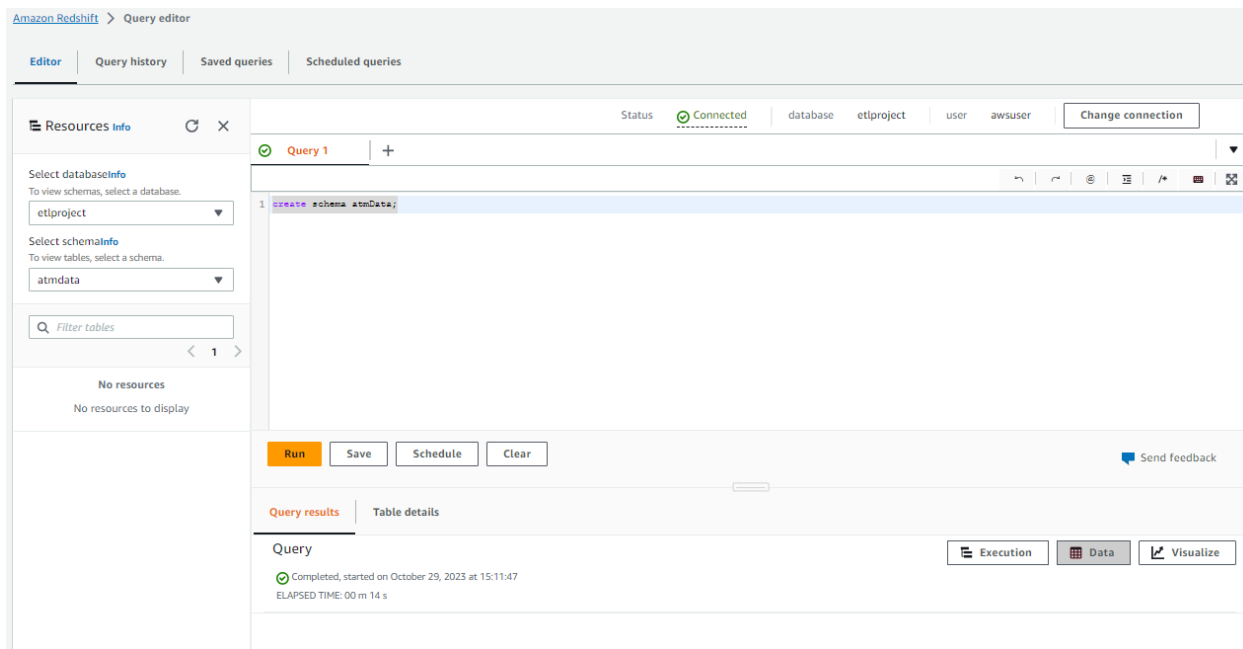
**Alarms (0)** | **Events (1)**

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

## Creating Schema:

### Query:

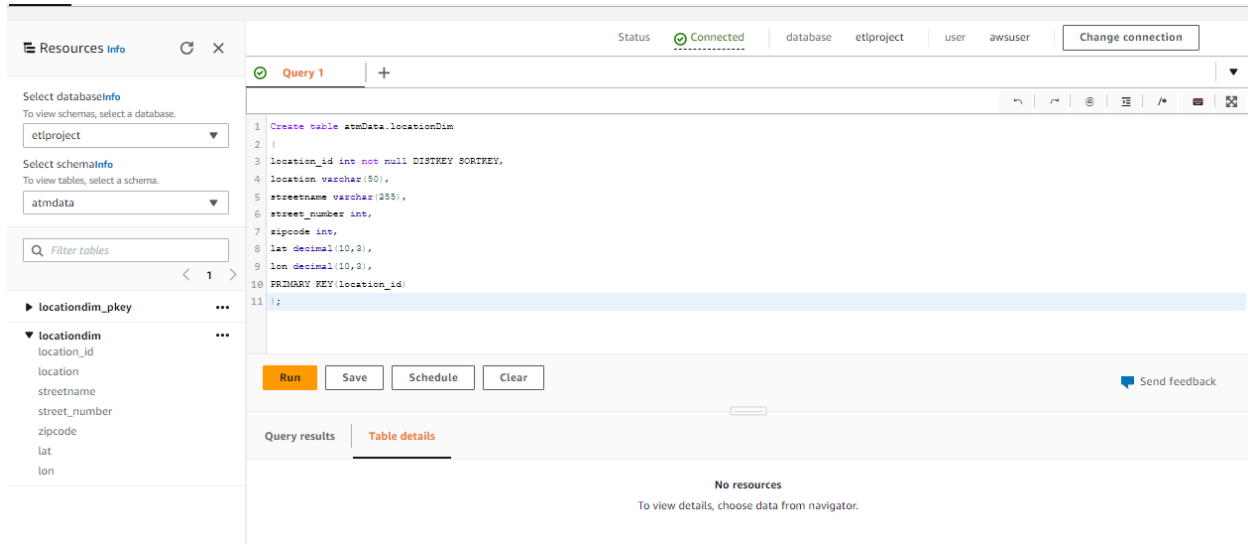
create schema atmData;



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

- Creating location dimension table:

```
Create table atmData.locationDimw
(
location varchar(255),
streetname varchar(255),
street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
location_id int not null DISTKEY SORTKEY,
PRIMARY KEY(location_id)
);
```



The screenshot shows a SQL IDE interface. On the left, there is a 'Resources Info' panel with a tree view showing the database structure. The 'locationdim' table is selected, and its columns (location\_id, location, streetname, street\_number, zipcode, lat, lon) are listed. The main area is a query editor with a SQL query to create a table named 'atmData.locationDim'. The query is as follows:

```
1 Create table atmData.locationDim
2 (
3 location_id int not null DISTKEY SORTKEY,
4 location varchar(50),
5 streetname varchar(255),
6 street_number int,
7 zipcode int,
8 lat decimal(10,5),
9 lon decimal(10,5),
10 PRIMARY KEY(location_id)
11 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there is a 'Query results' section with a 'Table details' tab, which currently shows 'No resources'.

- Creating atm dimension table

```
create table atmData.atmDim
(
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 atmData.locationDim(location_id)
);
```

Amazon Redshift > Query editor

Editor | Query history | Saved queries | Scheduled queries

Status: Connected | database: etlproject | user: awsuser | [Change connection](#)

Resources Info

Select database info  
To view schemas, select a database.  
etlproject

Select schema info  
To view tables, select a schema.  
atmdata

Filter tables

atmdim\_pkey  
locationdim\_pkey  
atmdim  
atm\_id  
atm\_number  
atm\_manufacturer  
atm\_location\_id  
locationdim

Query 1

```

1 create table atmData.atmDim
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atmData.locationDim(location_id)
9 );
10

```

Run Save Schedule Clear

Send feedback

Query results | Table details

No resources  
To view details, choose data from navigator.

- Creating date dimension table  
create table atmData.dateDim  
(  
year int,  
month varchar(20),  
day int,  
hour int,  
weekday varchar(20),  
full\_date\_time timestamp,  
date\_id int not null DISTKEY SORTKEY,  
PRIMARY KEY(date\_id)  
);

Amazon Redshift > Query editor

Editor | Query history | Saved queries | Scheduled queries

Resources Info

Select database Info  
To view schemas, select a database.  
etlproject

Select schema Info  
To view tables, select a schema.  
atmdata

Filter tables

atmdim\_pkey  
datedim\_pkey  
locationdim\_pkey  
atmdim  
datedim  
date\_id  
full\_date\_time  
year  
month  
day  
hour  
weekday  
locationdim

Status Connected database etlproject user awsuser Change connection

Query 1

```

2 create table atmData.dateDim
3 |
4 date_id int not null DISTKEY SORTKEY,
5 full_date_time timestamp,
6 year int,
7 month varchar(20),
8 day int,
9 hour int,
10 weekday varchar(20),
11 PRIMARY KEY(date_id)
12 ;
13

```

Run Save Schedule Clear

Send feedback

Query results | Table details

No resources  
To view details, choose data from navigator.

- Creating card type dimension table  
create table atmData.cardTypeDim  
(  
card\_type\_id int not null DISTKEY SORTKEY,  
card\_type varchar(30),  
PRIMARY KEY(card\_type\_id)  
);

Amazon Redshift > Query editor

Editor | Query history | Saved queries | Scheduled queries

Status: Connected | database: etlproject | user: awsuser

Resources info

Select database: etlproject

Select schema: atmdata

Filter tables

atmdim\_pkey  
cardtypedim\_pkey  
datedim\_pkey  
locationdim\_pkey  
atmdim  
cardtypedim  
datedim  
locationdim

```

1 create table atmData.cardTypeDim
2 {
3 card_type_id int not null DISTKEY SORTKEY,
4 card_type varchar(30),
5 PRIMARY KEY (card_type_id)
6 ;
7

```

Run | Save | Schedule | Clear

Query results | Table details

No resources  
To view details, choose data from navigator.

- Creating atm transactions fact table  
create table atmData.atmTransFact  
(  
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 atmData.locationDim (location\_id),  
FOREIGN KEY(atm\_id) references atmData.atmDim (atm\_id),

FOREIGN KEY(date\_id) references atmData.dateDim (date\_id),  
FOREIGN KEY(card\_type\_id) references atmData.cardTypeDim (card\_type\_id)  
);

Amazon Redshift > Query editor

Editor | Query history | Saved queries | Scheduled queries

Status Connected | database etlproject | user awsuser | [Change con](#)

**Resources Info**

Select database [Info](#)  
To view schemas, select a database.  
etlproject

Select schema [Info](#)  
To view tables, select a schema.  
atmdata

1

- ▶ atmdim\_pkey
- ▶ atmtransfact\_pkey
- ▶ cardtypedim\_pkey
- ▶ datedim\_pkey
- ▶ locationdim\_pkey
- ▶ atmdim
- ▼ atmtransfact
  - trans\_id
  - atm\_id
  - weather\_loc\_id
  - date\_id
  - card\_type\_id
  - atm\_status
  - currency
  - service
  - transaction\_amount
  - message\_code

**Query 1**

```

14 main_2h decimal(10,2),
15 clouds_all int,
16 weather_id int,
17 weather_main varchar(50),
18 weather_description varchar(255),
19 PRIMARY KEY (trans_id),
20 FOREIGN KEY (weather_loc_id) references atmData.locationDim (location_id),
21 FOREIGN KEY (atm_id) references atmData.atmDim (atm_id),
22 FOREIGN KEY (date_id) references atmData.dateDim (date_id),
23 FOREIGN KEY (card_type_id) references atmData.cardTypeDim (card_type_id)
24 ;
25

```

[Run](#) [Save](#) [Schedule](#) [Clear](#)

Query results | **Table details**

No resources  
To view details, choose data from navigator.

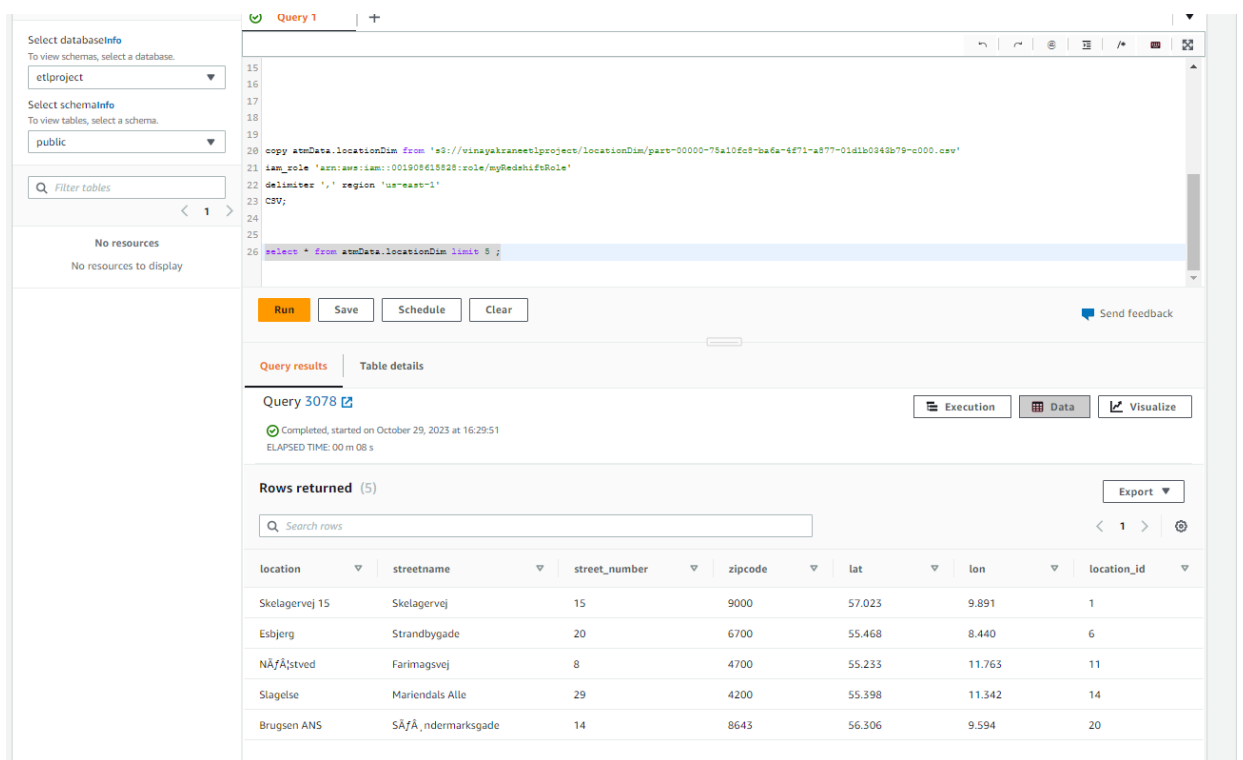
## Loading data into a Redshift cluster from Amazon S3 bucket

### Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

#### Queries:

- Copying the data to locationDim table

```
copy atmData.locationDim from 's3://vinayakraneetlproject/locationDim/part-00000-
b220c916-33e4-4550-bc76-777be5d02a7d-c000.csv'
iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



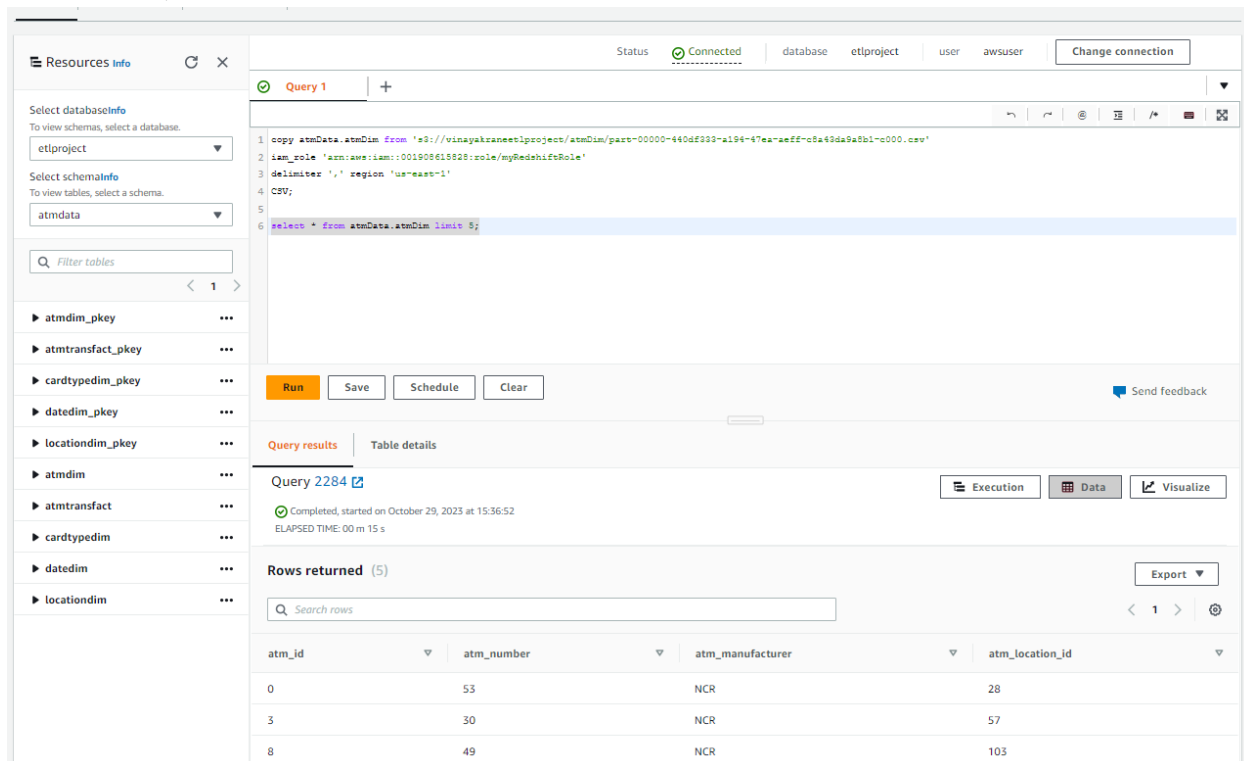
The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Select database' (etlproject) and 'Select schema' (public). The main area displays a SQL query for copying data from an S3 bucket into the 'atmData.locationDim' table. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 3078' which is 'Completed, started on October 29, 2023 at 16:29:51' with an 'ELAPSED TIME: 00 m 08 s'. It indicates 'Rows returned (5)'. A table with 7 columns (location, streetname, street\_number, zipcode, lat, lon, location\_id) displays 5 rows of data.

location	streetname	street_number	zipcode	lat	lon	location_id
Skelagervej 15	Skelagervej	15	9000	57.023	9.891	1
Esbjerg	Strandbygade	20	6700	55.468	8.440	6
NÅfÅstved	Farimagsvej	8	4700	55.233	11.763	11
Slagelse	Mariendals Alle	29	4200	55.398	11.342	14
Brugsen ANS	SÅfÅ, ndermarksgade	14	8643	56.306	9.594	20



- **Copying the data to atmDim table**

```
copy atmData.atmDim from 's3://vinayakraneetlproject/atmDim/part-00000-440df333-
a194-47ea-aeff-c8a43da9a8b1-c000.csv'
iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel shows the database 'etlproject' and schema 'atmdata'. The main query editor displays the following SQL query:

```
1 copy atmData.atmDim from 's3://vinayakraneetlproject/atmDim/part-00000-440df333-a194-47ea-aeff-c8a43da9a8b1-c000.csv'
2 iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1'
4 CSV;
5
6 select * from atmData.atmDim limit 5;
```

The query was executed successfully, as indicated by the 'Query 2284' status and the 'Completed' message. The results show 5 rows returned, with the following data:

atm_id	atm_number	atm_manufacturer	atm_location_id
0	53	NCR	28
3	30	NCR	57
8	49	NCR	103

- **Copying the data to dateDim table**

```
copy atmData.dateDim from 's3://vinayakraneetlproject/dateDim/part-00000-b4c8aec4-
3965-41f6-87b9-f47a3b9eedbe-c000.csv'
iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
timeformat 'auto'
CSV;
```

Select schemaInfo

To view tables, select a schema.

public

Filter tables

No resources

No resources to display

```

16
17
18 copy atmData.dateDim from 's3://vinayakraneetlproject/dateDim/part-00000-b4e8ae9d-9965-41f6-87b9-f47a2b5eedb8-c000.csv'
19 iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
20 delimiter ',' region 'us-east-1'
21 timeformat 'auto'
22 CSV;
23
24
25 select * from atmData.dateDim limit 5;

```

Run Save Schedule Clear

Send feedback

Query results Table details

Query 3175

Completed, started on October 29, 2023 at 16:36:02  
ELAPSED TIME: 00 m 15 s

Execution Data Visualize

Rows returned (5)

Export

Search rows

year	month	day	hour	weekday	full_date_time	date_id
2017	July	27	21	Thursday	2017-07-27 21:00:00	1
2017	July	11	12	Tuesday	2017-07-11 12:00:00	6
2017	September	18	4	Monday	2017-09-18 04:00:00	11
2017	August	8	0	Tuesday	2017-08-08 00:00:00	14
2017	September	16	17	Saturday	2017-09-16 17:00:00	20

- **Copying the data to cardTypeDim table**

```

copy atmData.cardTypeDim from 's3://vinayakraneetlproject/cardTypeDim/part-00000-
ba678b24-718d-44a8-8ce9-a8ae9e3cfbd5-c000.csv'
iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;

```

Editor

Query history

Saved queries

Scheduled queries

Resources info

Select database info

To view schemas, select a database.

etlproject

Select schema info

To view tables, select a schema.

public

Filter tables

< 1 >

No resources

No resources to display

Status

Connected

database

etlproject

user

awsuser

Change connection

Query 1

+

1

copy atmData.cardTypeDim from 's3://vinayakraneetlproject/cardTypeDim/part-00000-ba678b24-718d-44a8-8ce9-a8ae9e9cfed8-c000.csv'

2

iam\_role 'arn:aws:iam::001908615828:role/myRedshiftRole'

3

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

4

CSV;

5

6

select \* from atmData.cardTypeDim limit 5;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 2466

Execution

Data

Visualize

Completed, started on October 29, 2023 at 15:46:53

ELAPSED TIME: 00 m 08 s

Rows returned (5)

Export

Search rows

< 1 >

card_type_id	card_type
5	Visa Dankort
1	Mastercard - on-us

- Copying the data to atmTransFact table

```
copy atmData.atmTransFact from 's3://vinayakraneetlproject/atmTransFact/part-00000-9b72c7a0-0fa7-4260-9399-31c61edd9b6b-c000.csv'
iam_role 'arn:aws:iam::001908615828:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

Select databaseInfo

To view schemas, select a database.

etlproject

Select schemaInfo

To view tables, select a schema.

public

Filter tables

< 1 >

No resources

No resources to display

Query 1

+

1

copy atmData.atmTransFact from 's3://vinayakranestlproject/atmTransFact/part-00000-9b72c7a0-0fa7-4260-9399-31c61edd9b6b-c000.csv'

2

iam\_role 'arn:aws:iam::001908615528:role/myRedshiftRole'

3

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

4

CSV;

5

6

select \* from atmData.atmTransFact limit 5;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 2501

Execution

Data

Visualize

Completed, started on October 29, 2023 at 15:49:17

ELAPSED TIME: 00 m 15 s

Rows returned (5)

Export

Search rows

< 1 >

trans_id	atm_id	weather_loc_id	date_id	card_type_id	atm_status	currency	service	transaction_amount	me
2	19	74	0	5	Active	DKK	Withdrawal	5428	
4	19	74	0	5	Active	DKK	Withdrawal	3829	
7	112	37	1	1	Inactive	DKK	Withdrawal	1594	
9	112	37	1	1	Inactive	DKK	Withdrawal	9664	