

Creation of a Redshift Cluster

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

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

redshift-cluster-etl-project

Actions Edit Add partner integration Query data

General information info

Cluster identifier redshift-cluster-etl-project	Status Available	Node type dc2.large	Endpoint redshift-cluster-etl-project.ccrxw2ciwuo0.us-east-1.redshift.amazonaws.com:5439/dev
Custom domain name - new -	Date created July 18, 2023, 22:09 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-etl-project.ccrxw2ciwuo0.us-east-1.redshift.amazonaws.com:5439/dev
Cluster ARN arn:aws:redshift:us-east-1:749384479763:namespace:a5b2c402-0489-41b2-adf8-58b5b9de2589	Storage used 0.02% (0.06 of 320 GB used)		ODBC URL Driver=(Amazon Redshift (x64)); Server=redshift-cluster-etl-project.ccrxw2ciwuo0.us-east-1.redshift.amazonaws.com; Database=dev
Cluster configuration Production	Multi-AZ No		

Cluster performance Query monitoring Schedules Maintenance Properties

Cluster performance Query monitoring Schedules Maintenance **Properties**

Database configurations info

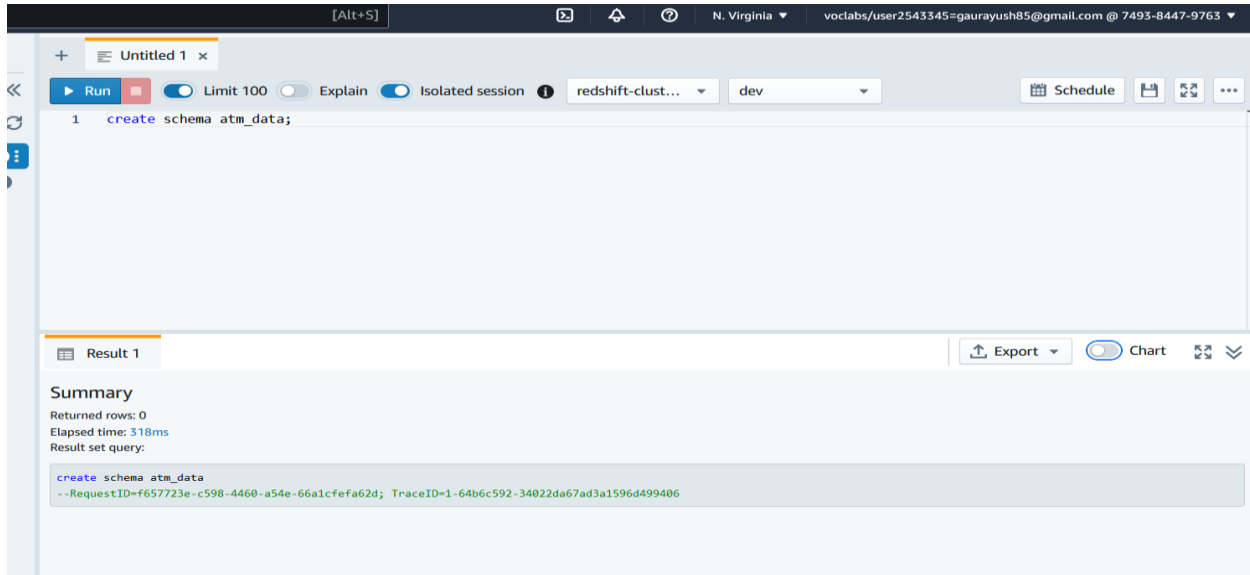
Edit admin credentials Rotate encryption keys Edit

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 AAAAB3NzaC1yc2EAAAADAQABAAQDBSW4vkVlgRHHZneNyh0vcpoxa7fmp6t87dPZMLz0akP2v1QaHm31omVdr6B790GwCnuBFJPDpq96hsujEiPnX4zHWgUOPI3cMMVmJfQzIuCEwPlsD+PmJ0khY5+cYaoKyFD4nSp5wywdbrtNF1jcBpiq49jppYdXilGUUsJc59G1afB9LbdYtrmcrib6UGsFW40JaGC1yms6XFUpHzJ59bKEZ8LBAwmrYt2l6bozfKbPWxpZd3ZbUdetqXS3TM3ePHRufdRagMdNDxZWtXszSD/zVcQmm02rtR07ZUjxZrKC+jGDx/czRhpdHwyebE1SYIMCES7fvm5KvPDAmazon-Redshift	AWS KMS key ID -	

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

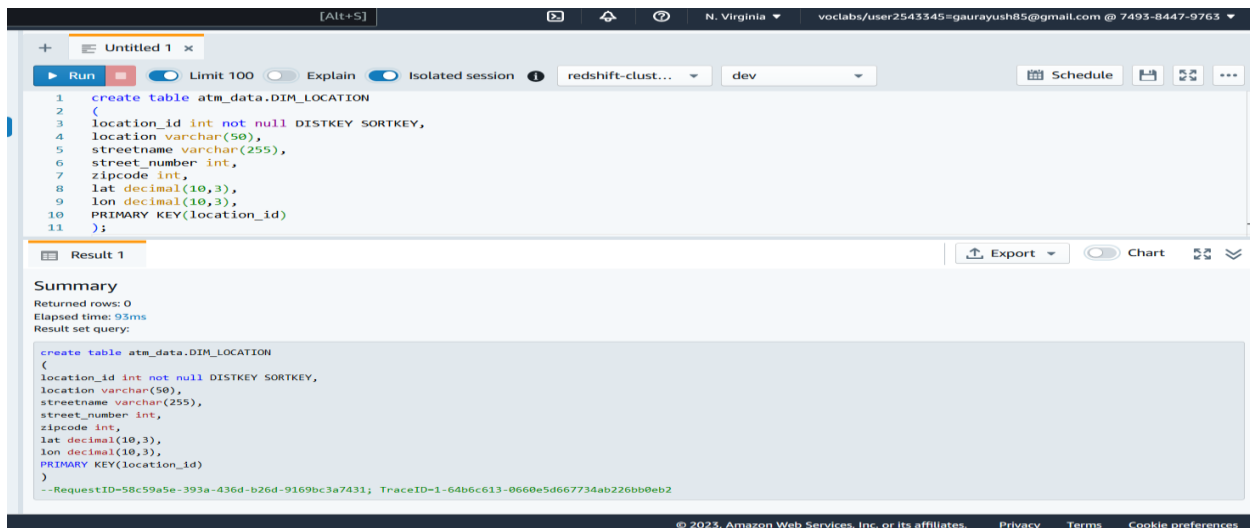
Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

Query for creating schema:
create schema atm_data;



The screenshot shows the AWS Redshift console interface. At the top, there's a navigation bar with the user's name 'voclabs/user2543345=gaurayush85@gmail.com' and a session ID. Below this, a toolbar contains buttons for 'Run', 'Limit 100', 'Explain', 'Isolated session', and a dropdown menu showing 'redshift-clust...' and 'dev'. The main area displays a single SQL query: `create schema atm_data;`. Below the query, the 'Result 1' tab is active, showing a 'Summary' section with the following details: 'Returned rows: 0', 'Elapsed time: 318ms', and 'Result set query:'. The query text is repeated in a light blue box: `create schema atm_data`. At the bottom, a status bar shows the request ID and trace ID.

- Creating location dimension table
- ```
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)
);
```

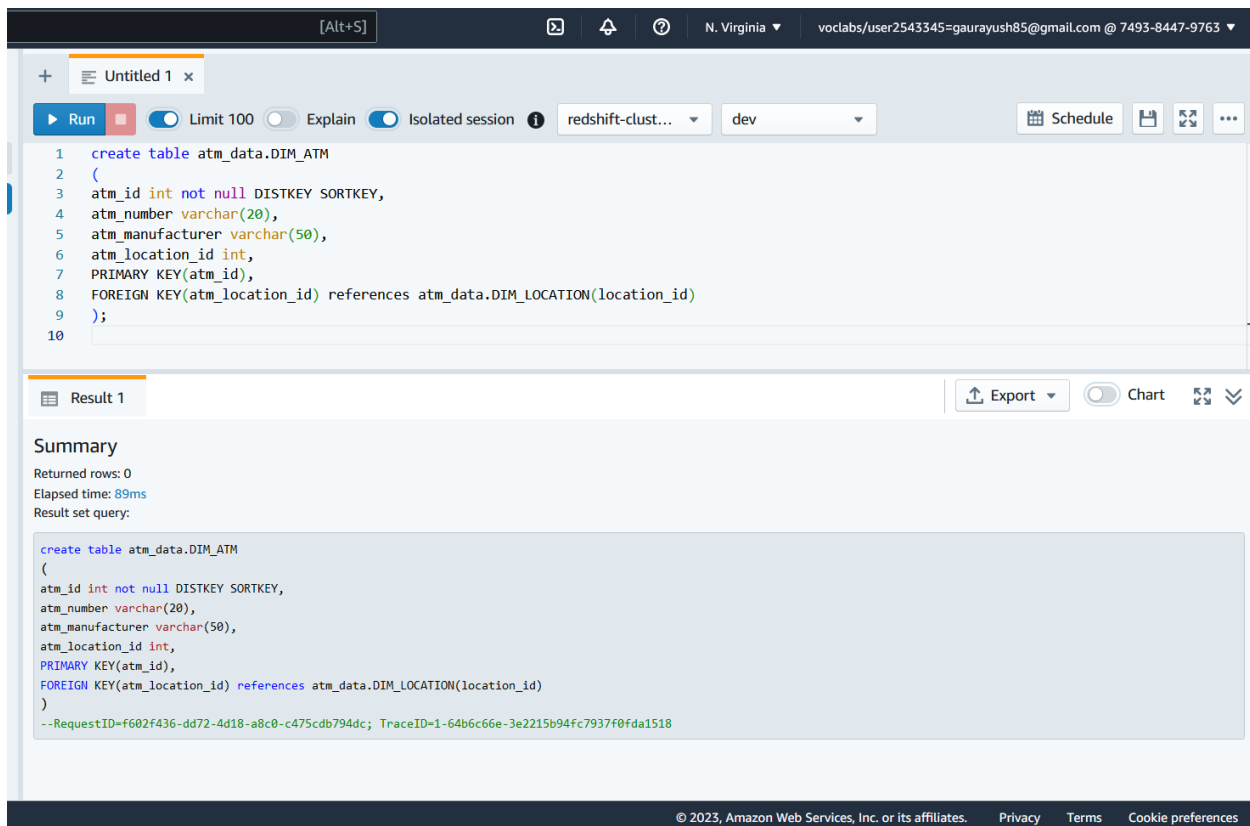


The screenshot shows the AWS Redshift console interface. The toolbar is similar to the previous screenshot. The main area displays a multi-line SQL query: `create table atm_data.DIM_LOCATION ( location_id int not null DISTKEY SORTKEY, location varchar(50), streetname varchar(255), street_number int, zipcode int, lat decimal(10,3), lon decimal(10,3), PRIMARY KEY(location_id) );`. Below the query, the 'Result 1' tab is active, showing a 'Summary' section with the following details: 'Returned rows: 0', 'Elapsed time: 93ms', and 'Result set query:'. The query text is repeated in a light blue box. At the bottom, a status bar shows the request ID and trace ID.

- Creating atm dimension table

```
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 a SQL query editor interface. At the top, there's a toolbar with buttons for 'Run', 'Limit 100', 'Explain', 'Isolated session', and a dropdown menu showing 'redshift-clust...'. Below the toolbar, the SQL query is entered in a text area:

```
1 create table atm_data.DIM_ATM
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 atm_data.DIM_LOCATION(location_id)
9);
10
```

Below the query editor, there's a 'Result 1' section. It shows a 'Summary' of the query execution:

- Returned rows: 0
- Elapsed time: 89ms
- Result set query:

The result set query is displayed in a code block, showing the same SQL query as above. At the bottom of the interface, there's a footer with copyright information: '© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

- Creating date dimension table

```
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)
)
```

```

[Alt+S]
N. Virginia | voclabs/user2543345=gaurayush85@gmail.com @ 7493-8447-9763
+ | Untitled 1 x
Run | Limit 100 | Explain | Isolated session | redshift-clust... | dev | Schedule | Export | Chart | ...
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);
Result 1 | Export | Chart | ...
Summary
Returned rows: 0
Elapsed time: 87ms
Result set query:
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)
)
--RequestID=39597d77-0a97-46ea-8817-868517fb02f6; TraceID=1-64b6c6a1-37e0305169f63acc1d7a530a
© 2023, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

```

- Creating card type dimension table
- ```

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

```

```

[Alt+S]
N. Virginia | voclabs/user2543345=gaurayush85@gmail.com @ 7493-8447-9763
+ | Untitled 1 x
Run | Limit 100 | Explain | Isolated session | redshift-clust... | dev | Schedule | Export | Chart | ...
1 create table atm_data.DIM_CARD_TYPE
2 (
3 card_type_id int not null DISTKEY SORTKEY,
4 card_type varchar(30),
5 PRIMARY KEY(card_type_id)
6 );
7
Result 1 | Export | Chart | ...
Summary
Returned rows: 0
Elapsed time: 100ms
Result set query:
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
)
--RequestID=a420d5a1-2c0f-4e6d-8aa2-fb374146c7b0; TraceID=1-64b6c6c7-3e2ee04b55eb8ef1404cd513
© 2023, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences

```

- Creating atm transactions fact table

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),

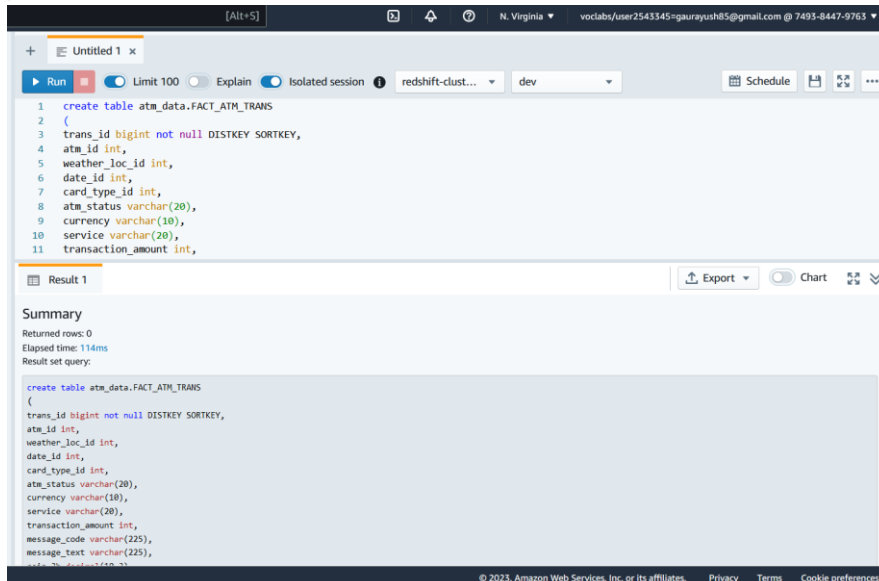
FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),

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 shows a web-based SQL editor interface. The top bar includes a search icon, a notification bell, and user information: "H. Virginia" and "voclabs/user2543145-gaurayush85@gmail.com @ 7493-4447-9763". Below the bar, there's a toolbar with buttons for "Run", "Limit 100", "Explain", "Isolated session", and a dropdown menu showing "redshift-clust..." and "dev". The main area contains a SQL query:

```
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,
```

. Below the query, there's a "Result 1" section with a "Summary" box. The summary indicates "Returned rows: 0", "Elapsed time: 114ms", and "Result set query:". The query text is repeated in the result area. At the bottom, there's a footer with "© 2023, Amazon Web Services, Inc. or its affiliates." and links for "Privacy", "Terms", and "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

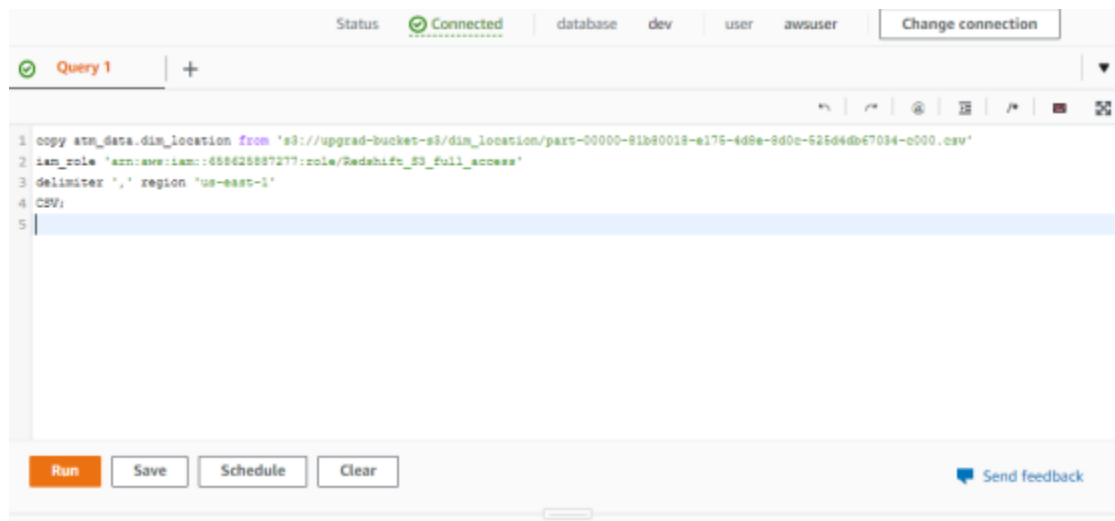
- Copying the data to dim_location table

copy atm_data.dim_location from 's3://upgrad-bucket-s3/dim_location/part-00000-81b80018-e175-4d8e-8d0c-525d4db67034-c000.csv'

iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'

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

CSV;



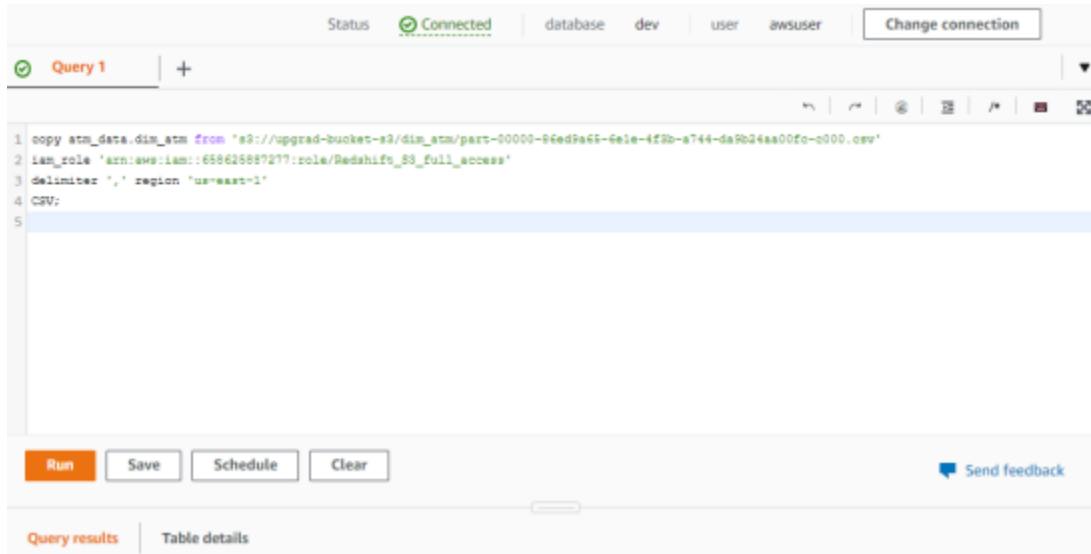
- Copying the data to dim_atm table

copy atm_data.dim_atm from 's3://upgrad-bucket-s3/dim_atm/part-00000-86ed9a65-6e1e 4f3b-a744-da9b24aa00fc-c000.csv'

iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'

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

CSV



```

1 copy atm_data.dim_atm from 's3://upgrad-bucket-s3/dim_atm/part-00000-86ed9a65-6e1e-4f2b-a744-da9b24aa00fc-c000.csv'
2 iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
3 delimiter ',' region 'us-east-1'
4 CSV;
5

```

Run Save Schedule Clear Send feedback

Query results Table details

- Copying the data to dim_date table

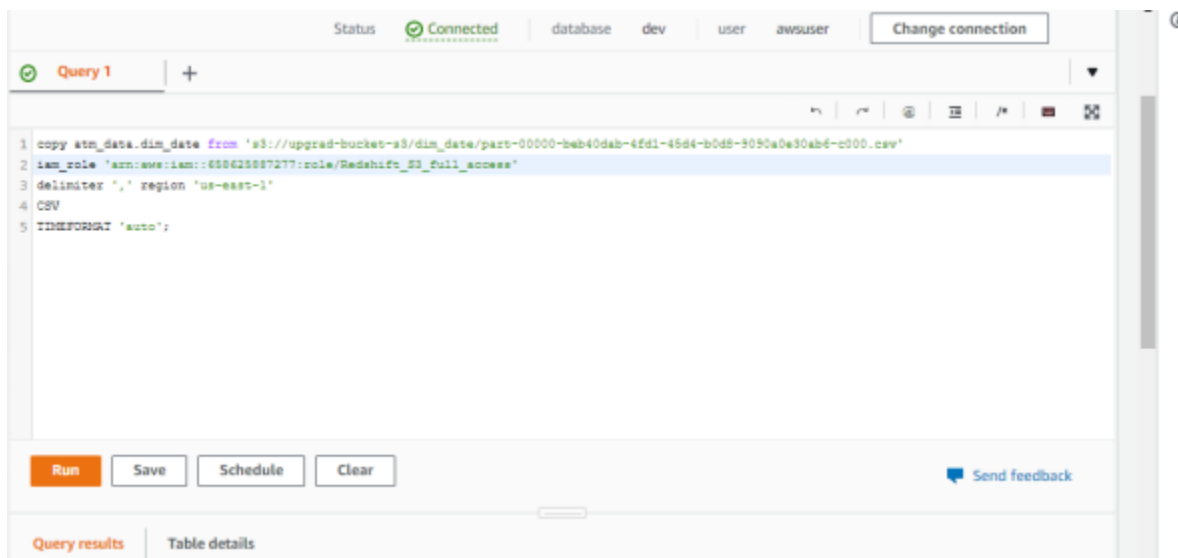
copy atm_data.dim_date from 's3://upgrad-bucket-s3/dim_date/part-00000-beb40dab 4fd1-45d4-b0d8-9090a0e30ab6-c000.csv'

iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'

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

CSV

TIMEFORMAT 'auto';



```

1 copy atm_data.dim_date from 's3://upgrad-bucket-s3/dim_date/part-00000-beb40dab-4fd1-45d4-b0d8-9090a0e30ab6-c000.csv'
2 iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';

```

Run Save Schedule Clear Send feedback

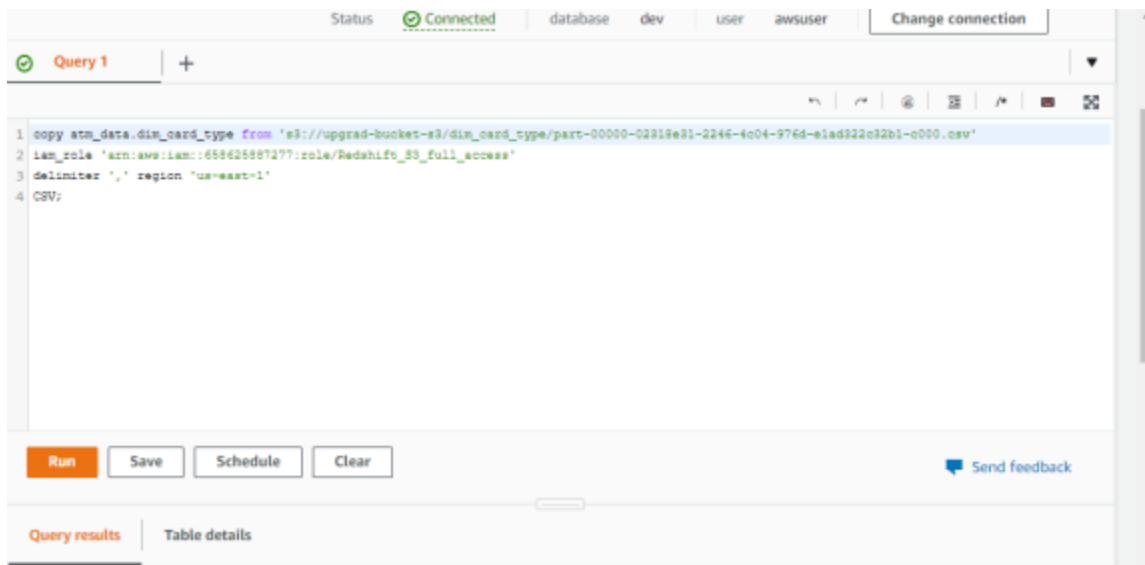
Query results Table details

- Copying the data to dim_card_type table

copy atm_data.dim_card_type from 's3://upgrad-bucket-s3/dim_card_type/part-00000-02318e31-2246-4c04-976d-e1ad322c32b1-c000.csv'

iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'

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



• Copying the data to fact_atm_trans table

copy atm_data.fact_atm_trans from 's3://upgrad-bucket-s3/fact_atm_trans/part-00000-fb72e768-f589-4caa-bd26-1bc6fcbbdb8a-c000.csv'
iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access'
delimiter ',' region 'us-east-1'
CSV;

