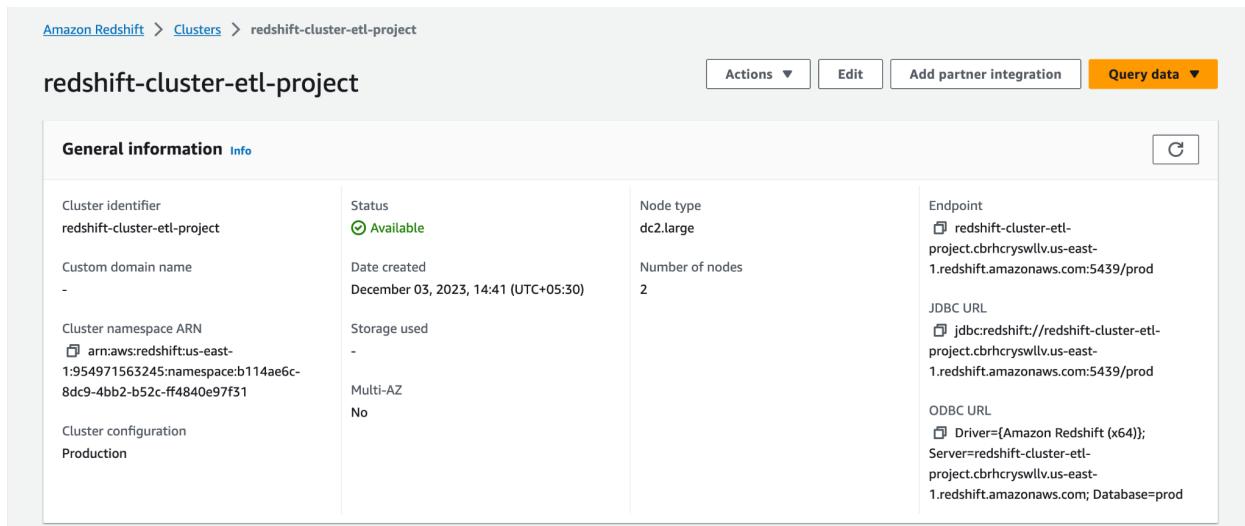


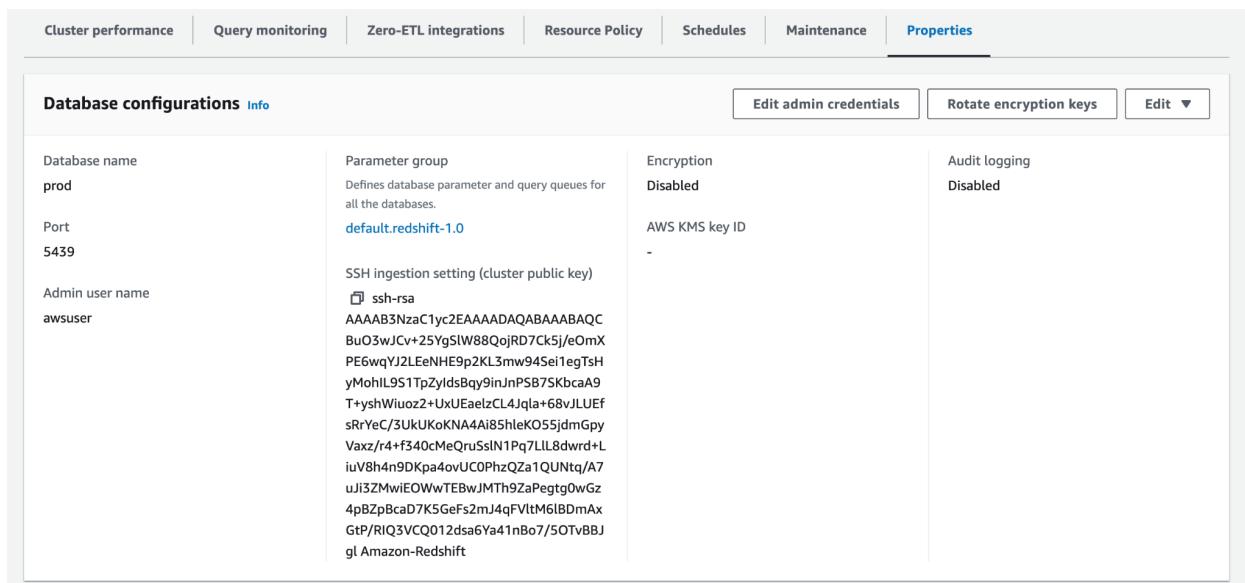
Creation of a Redshift Cluster

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



General information

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-etl-project	Available	dc2.large	redshift-cluster-etl-project.cbrhcrswlv.us-east-1.redshift.amazonaws.com:5439/prod
Custom domain name	Date created	Number of nodes	JDBC URL
-	December 03, 2023, 14:41 (UTC+05:30)	2	jdbc:redshift://redshift-cluster-etl-project.cbrhcrswlv.us-east-1.redshift.amazonaws.com:5439/prod
Cluster namespace ARN	Storage used		ODBC URL
arn:aws:redshift:us-east-1:954971563245:namespace:b114ae6c-8dc9-4bb2-b52c-ff4840e97f31	-		Driver={Amazon Redshift (x64)};Server=redshift-cluster-etl-project.cbrhcrswlv.us-east-1.redshift.amazonaws.com;Database=prod
Cluster configuration	Multi-AZ		
Production	No		



Properties

Database configurations

Database name	Parameter group	Encryption	Audit logging
prod	Defines database parameter and query queues for all the databases. default.redshift-1.0	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5439	ssh-rsa AAAAAB3NzaC1yc2EAAAQABAAQCBuO3wJCv+2SYgSlW88QojRD7Ck5j/eOmXPE6wqYJ2LEeNHE9p2KL3mw94Sei1egTshyMolL951TpZyldsbqy9inJnPSB7SKbca9T+yshWiuoz2+UxEaelzCL4Jqla+68vJLUEfsRyEc/3UkUKoKNA4A18ShleKO55jdjmGpyVaxz/r4+f340cMeQuSsln1Pq7LL8dwrd+Liuv8h4n9DKpa4ovUC0PhzQzaPegtg0wGz4pB2p8cad7K5GefS2mJ4qFVltM6IBdmaxGtp/RIQ3VCQ012ds6Ya41nBo7/SOTvBBJgl Amazon-Redshift	-	
Admin user name			
awsuser			

Network and security settings Info

				<small>Edit</small>
Virtual private cloud (VPC) vpc-0f76c891fe2e2c7905	Availability Zone us-east-1a	VPC security group Specify which instances and devices can connect to the cluster. sg-0fb725828f42db10	Publicly accessible Allow connections from outside the VPC. Disabled	
Subnet group cluster-subnet-group-etl-project	Enhanced VPC routing Disabled	IP address type -		
Endpoint URL -				

Cluster permissions

ⓘ Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1) Info

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

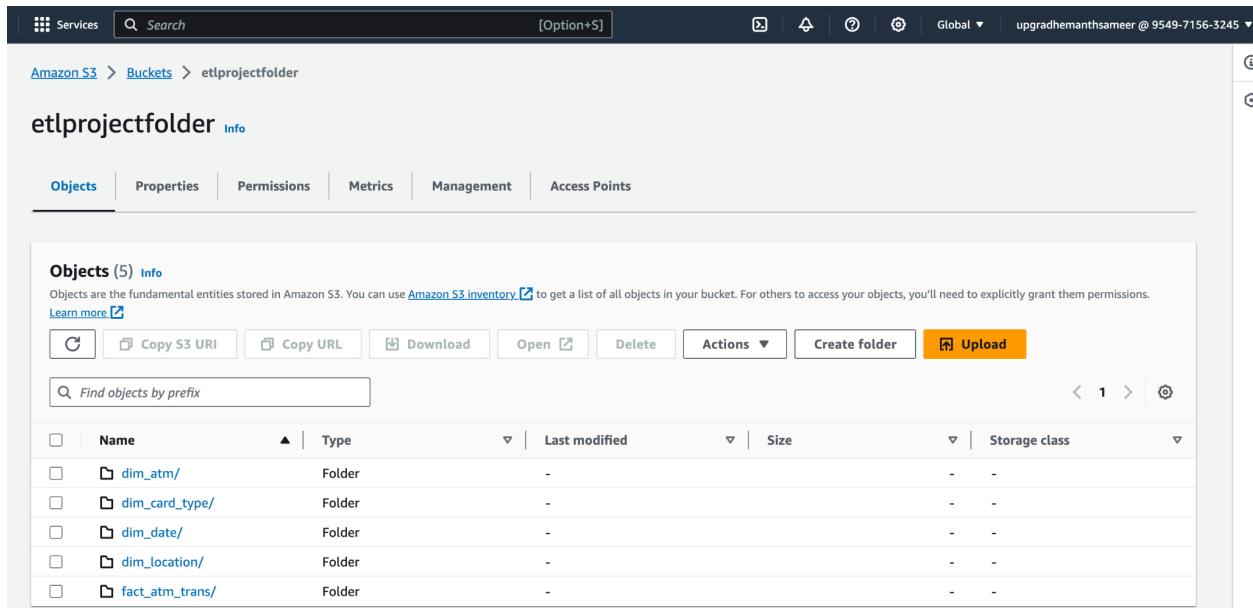
<input type="checkbox"/>	<small>IAM roles</small>	<small>Status</small>	<small>Role type</small>
<input type="checkbox"/>	redshift_s3_fullaccess	<small>in-sync</small>	--

Node IP addresses (3)

Node role	Public IP address	Private IP address
Leader	54.235.151.187	10.0.3.188
Compute-0	3.216.130.77	10.0.3.39
Compute-1	54.157.184.209	10.0.4.153

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

S3 bucket containing files:



Objects (5) [Info](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions.

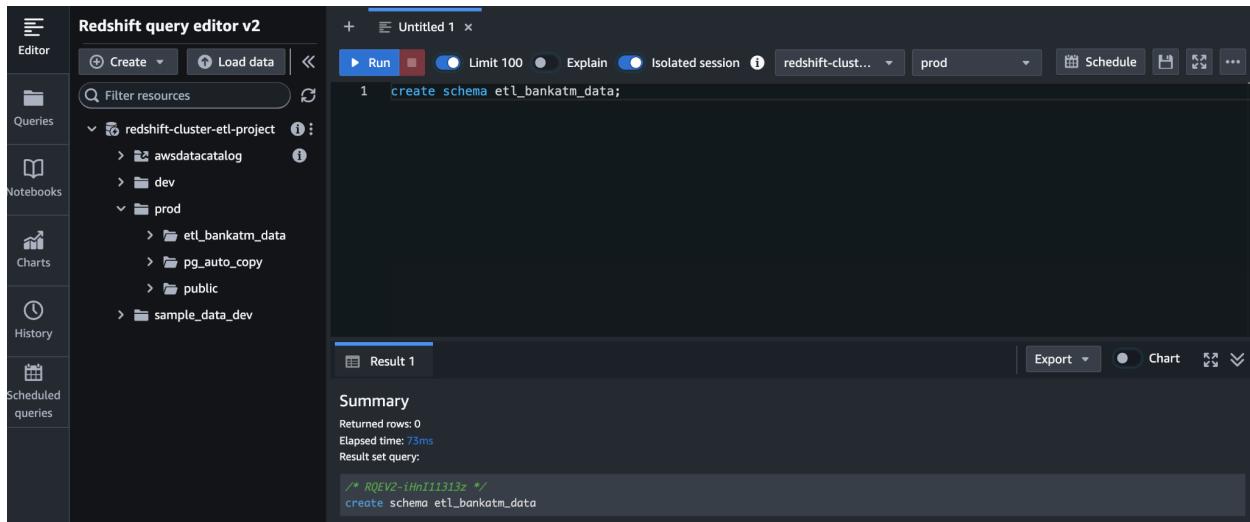
[Learn more](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dim_atm/	Folder	-	-	-
<input type="checkbox"/>	dim_card_type/	Folder	-	-	-
<input type="checkbox"/>	dim_date/	Folder	-	-	-
<input type="checkbox"/>	dim_location/	Folder	-	-	-
<input type="checkbox"/>	fact_atm_trans/	Folder	-	-	-

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

Query for creating schema:

```
create schema etl_bankatm_data;
```



The screenshot shows the Redshift query editor interface. On the left, there's a sidebar with options like Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main area shows a query in the editor pane:

```
1  create schema etl_bankatm_data;
```

The results pane below shows a summary:

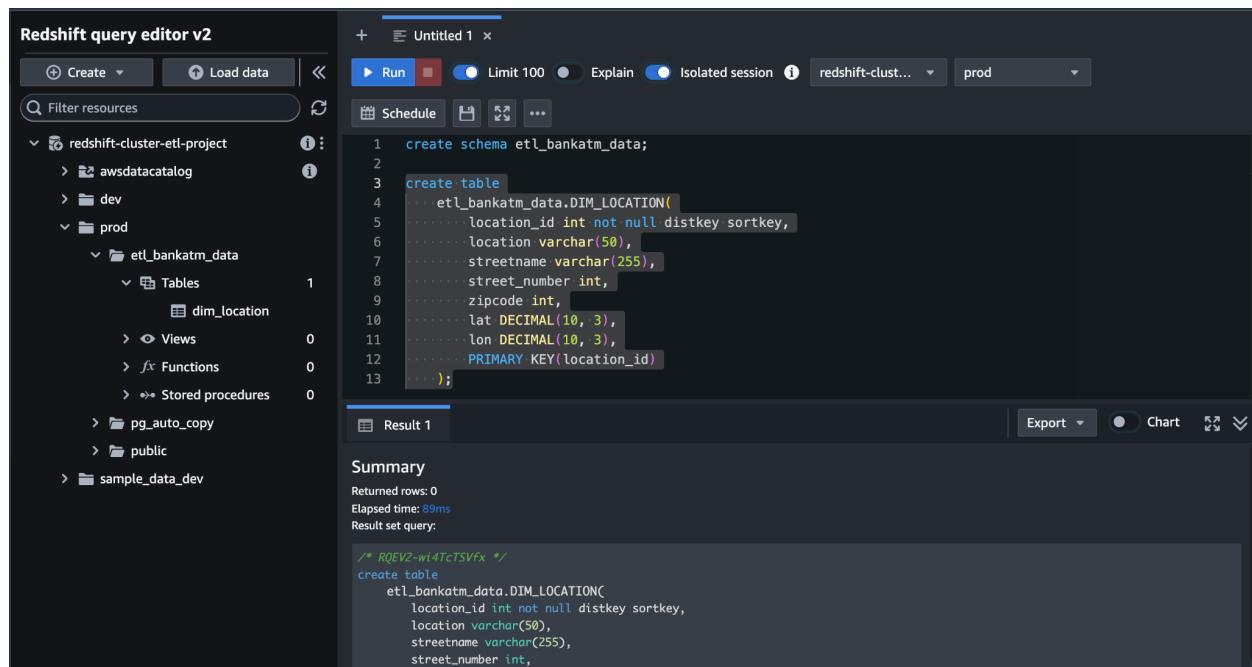
Summary
 Returned rows: 0
 Elapsed time: 73ms
 Result set query:
`/* RQEY2-iHnII113z */
create schema etl_bankatm_data`

Query for creating location dimension table:

```
create table
    etl_bankatm_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 Redshift query editor interface. On the left, the AWS Data Catalog sidebar displays the schema structure under the 'redshift-cluster-etl-project'. A table named 'dim_location' is selected. The main pane shows the SQL code for creating the table:

```

1  create schema etl_bankatm_data;
2
3  create table
4      etl_bankatm_data.DIM_LOCATION(
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     );
  
```

The code is highlighted in purple and blue. Below the code, the 'Result 1' tab shows the summary of the query execution:

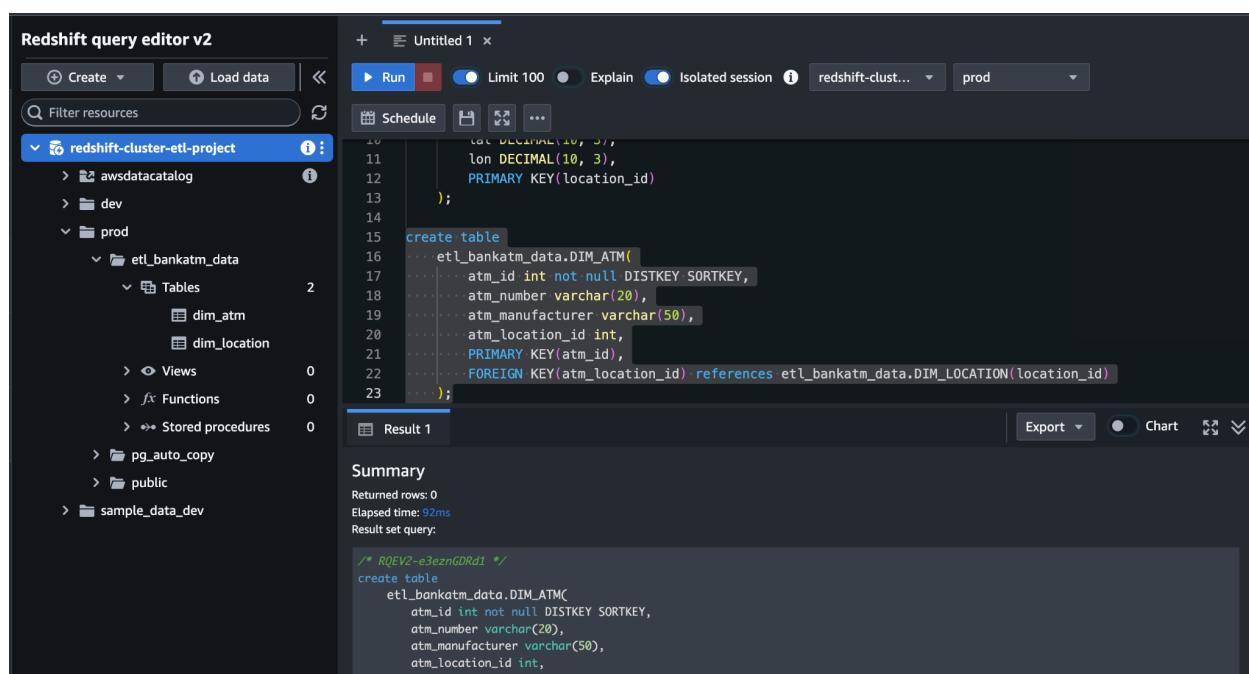
Summary
 Returned rows: 0
 Elapsed time: 89ms
 Result set query:
`/* RQE2-wi4TcTSVfx */
create table
etl_bankatm_data.DIM_LOCATION
location_id int not null distkey sortkey,
location varchar(50),
streetname varchar(255),
street_number int,`

Query for creating atm dimension table :

```
create table
```

```

etl_bankatm_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
etl_bankatm_data.DIM_LOCATION(location_id)
);
  
```



The screenshot shows the Redshift query editor interface. On the left, the file tree displays a project structure under 'redshift-cluster-etl-project'. The 'prod' folder contains an 'etl_bankatm_data' database with two tables: 'dim_atm' and 'dim_location'. The current tab is 'Untitled 1', which contains the SQL code for creating the 'DIM_ATM' table. The code is identical to the one shown above. The 'Run' button is highlighted, indicating the query has been executed. Below the code, the 'Result 1' tab shows a summary: 'Returned rows: 0' and 'Elapsed time: 92ms'. The result set query output is also visible.

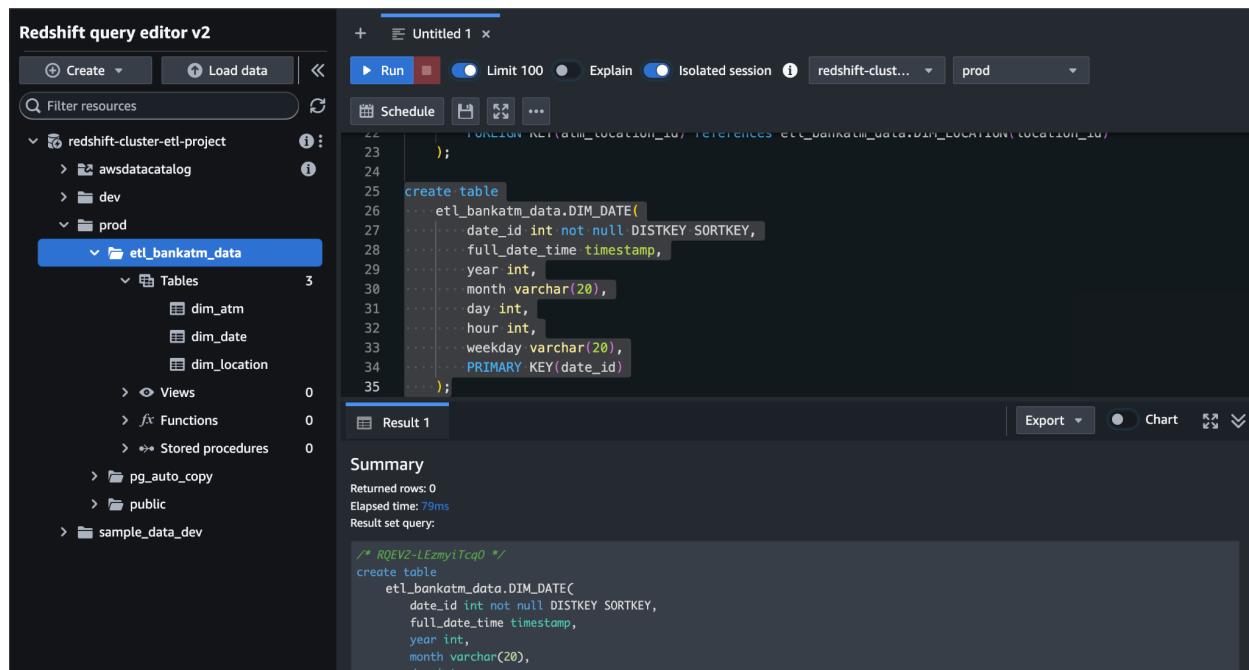
Query for creating date dimension table:

```

create table
etl_bankatm_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 shows the Redshift query editor interface. On the left, the navigation pane displays a database structure under 'redshift-cluster-etl-project'. A table named 'etl_bankatm_data' is selected, and its 'Tables' section is expanded, showing three tables: 'dim_atm', 'dim_date', and 'dim_location'. The main panel contains a code editor with the following SQL query:

```

create table etl_bankatm_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)
);
  
```

Below the code editor, the 'Summary' section indicates 0 returned rows and an elapsed time of 79ms. The result set query is also displayed.

Query for creating card type dimension table:

```

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

Redshift query editor v2

Untitled 1

Run | Limit 100 | Explain | Isolated session | redshift-clust... | prod

Schedule | ...

```

year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);

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

```

Result 1

Summary

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

```

/* RQEV2-RbuoftasRp */
create table
etl_bankatm_data.DIM_CARD_TYPE(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);

```

Query for creating atm transaction fact table:

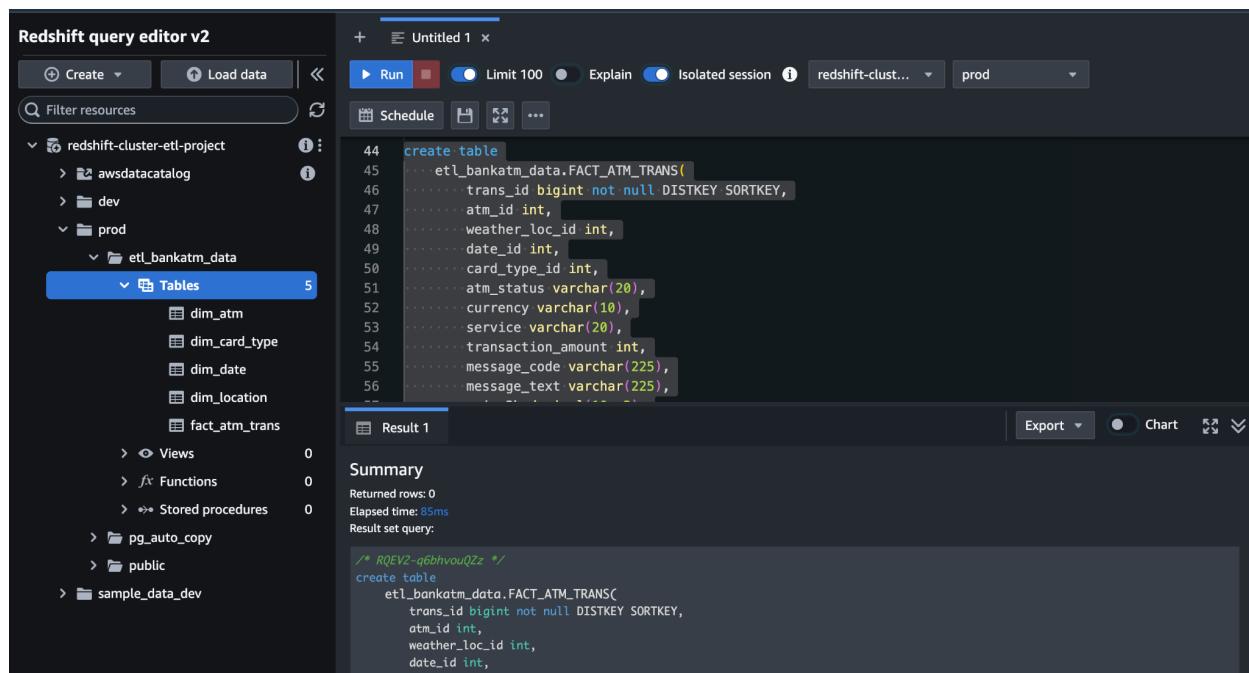
```

create table
etl_bankatm_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
etl_bankatm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references etl_bankatm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references etl_bankatm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references
etl_bankatm_data.DIM_CARD_TYPE(card_type_id)
);
  
```



The screenshot shows the Redshift query editor interface. On the left, the AWS Data Catalog sidebar displays the schema structure of the 'prod' database, specifically the 'etl_bankatm_data' schema, which contains tables like dim_atm, dim_card_type, dim_date, dim_location, and fact_atm_trans. The main query editor window shows the SQL code for creating the 'FACT_ATM_TRANS' table:

```

44  create table
45    etl_bankatm_data.FACT_ATM_TRANS(
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),
  
```

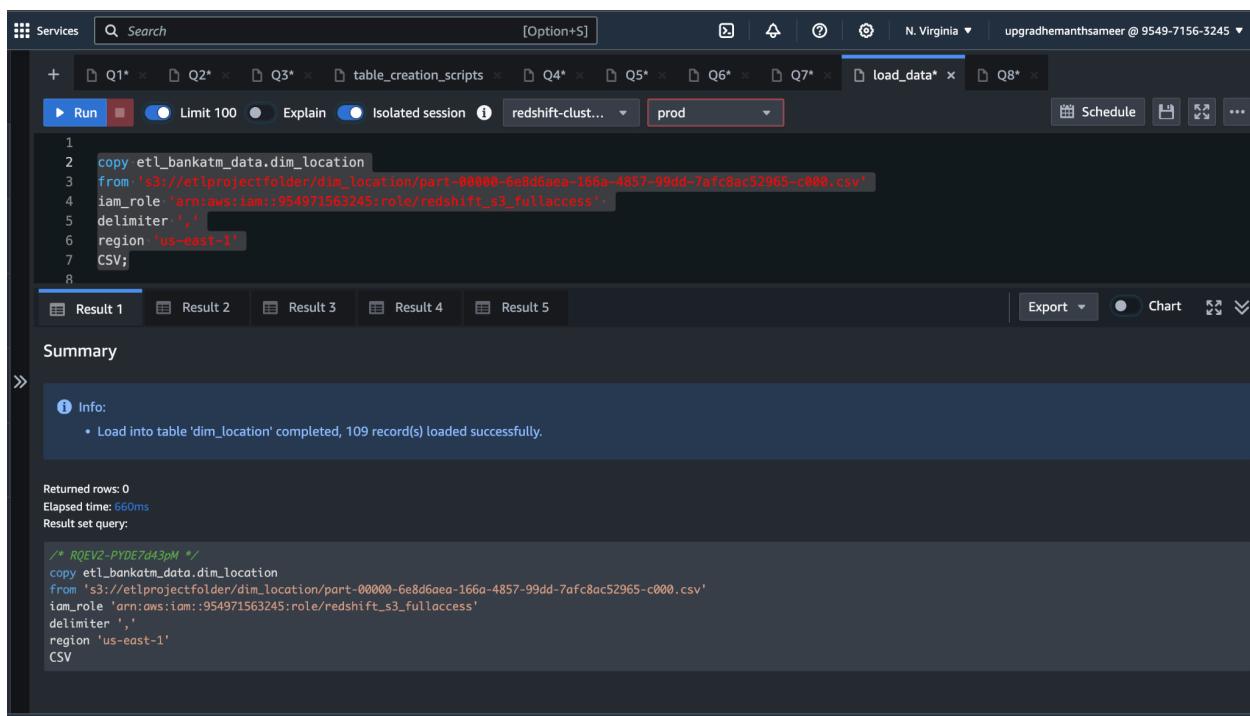
The 'Result 1' tab at the bottom indicates that the query was run successfully with 0 returned rows and an elapsed time of 85ms.

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

Query for copying data to dim_location table:

```
copy etl_bankatm_data.dim_location
from
's3://etlprojectfolder/dim_location/part-00000-6e8d6aea-166a-4857-99dd-7afc8ac52965-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift console interface. The query editor window contains the following SQL code:

```
copy etl_bankatm_data.dim_location
from
's3://etlprojectfolder/dim_location/part-00000-6e8d6aea-166a-4857-99dd-7afc8ac52965-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
CSV;
```

The session dropdown at the top right is set to "prod". Below the code, the "Summary" section displays the execution results:

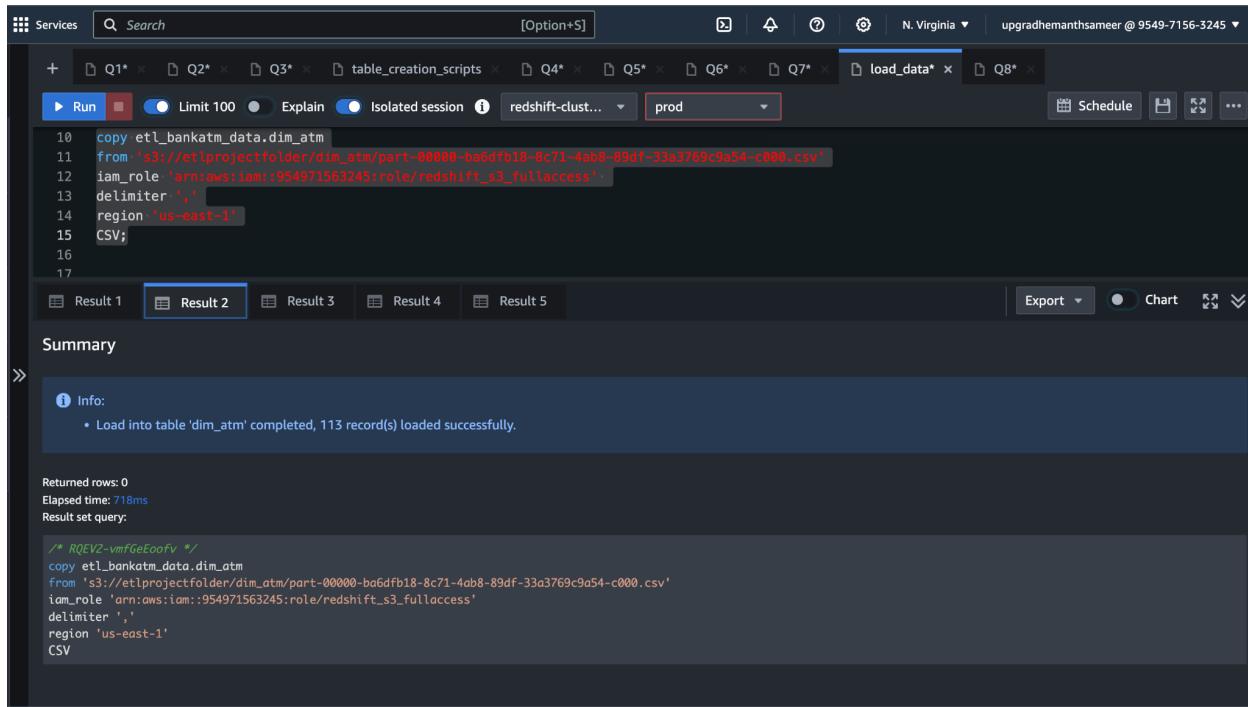
- Info:**
 - Load into table 'dim_location' completed, 109 record(s) loaded successfully.

Details of the execution are shown below, including the number of rows returned (0), elapsed time (660ms), and the full query text again.

Query for copying data to dim_atm table:

```
copy etl_bankatm_data.dim_atm
from
's3://etlprojectfolder/dim_atm/part-00000-ba6dfb18-8c71-4ab8-89df-33a3769c9a54-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
```

CSV;



The screenshot shows the AWS Redshift console interface. A query window titled 'load_data*' is open, displaying a COPY command to load data from an S3 bucket into a 'dim_atm' table. The command includes parameters for IAM role, delimiter, region, and file format. Below the command, the 'Result 2' tab is selected, showing a summary of the operation: 'Load into table 'dim_atm' completed, 113 record(s) loaded successfully.' The summary also indicates 0 returned rows and a 718ms elapsed time.

```

copy etl_bankatm_data.dim_atm
from 's3://etlprojectfolder/dim_atm/part-00000-ba6dfb18-8c71-4ab8-89df-33a3769c9a54-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
CSV;
    
```

Query for copying data to dim_date table:

```

copy etl_bankatm_data.dim_date

from
's3://etlprojectfolder/dim_date/part-00000-fec0aa66-b69c-46d0-aa8d-1656a84a3a6a-c000.c
sv'

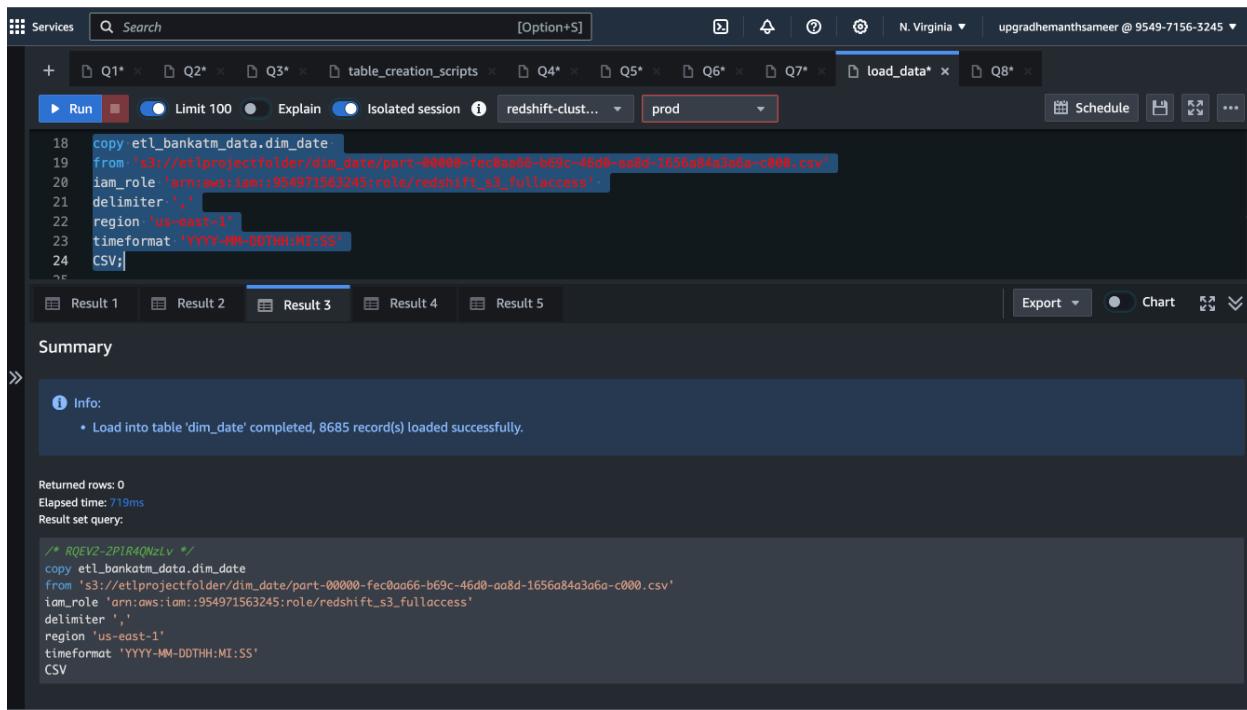
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'

delimiter ','

region 'us-east-1'

timeformat 'YYYY-MM-DDTHH:MI:SS'

CSV;
    
```



The screenshot shows the AWS Redshift console interface. A query named 'load_data*' is being run against a cluster named 'prod'. The query itself is:

```

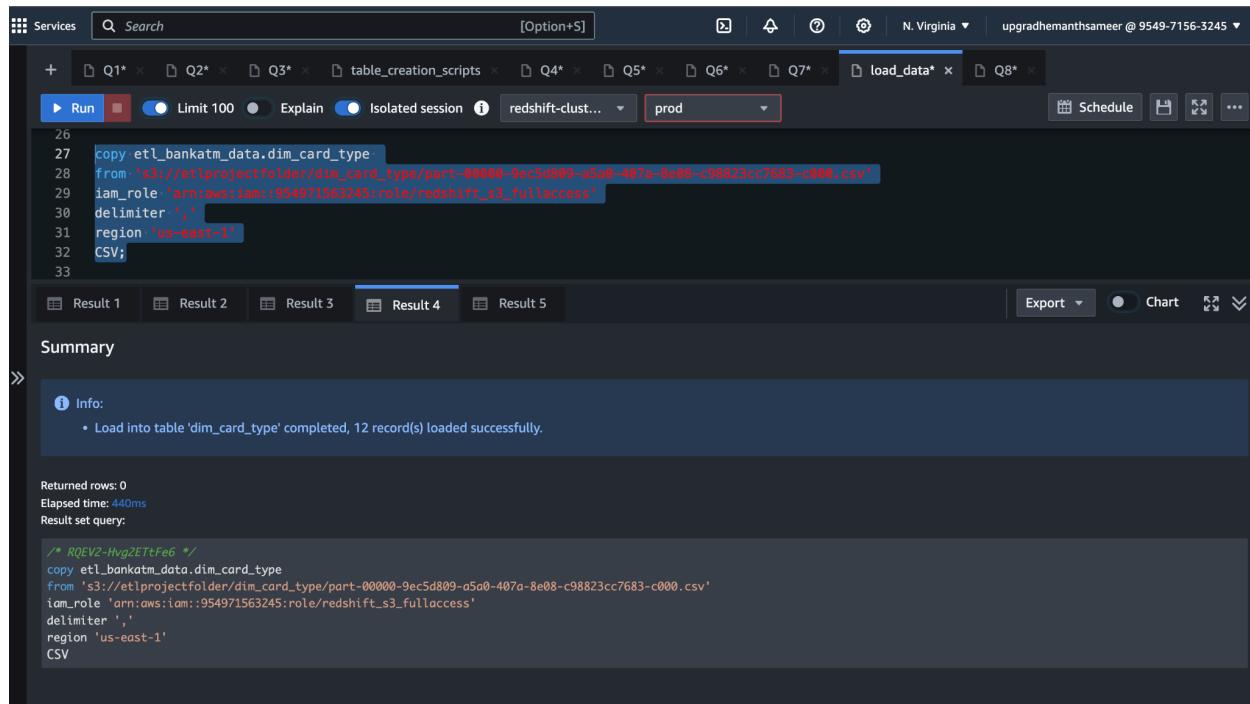
18 copy etl_bankatm_data.dim_date
19 from 's3://etlprojectfolder/dim_date/part-00000-fec0aa66-b69c-46d0-aa8d-1656a84a3a6a-c000.csv'
20 iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
21 delimiter ','
22 region 'us-east-1'
23 timeformat 'YYYY-MM-DDTHH:MI:SS'
24 CSV;
    
```

The results section shows a summary message: "Load into table 'dim_date' completed, 8685 record(s) loaded successfully." Below this, it provides timing information: "Elapsed time: 719ms".

Query for copying data to dim_card_type table:

```

copy etl_bankatm_data.dim_card_type
from
's3://etlprojectfolder/dim_card_type/part-00000-9ec5d809-a5a0-407a-8e08-c98823cc7683-c
000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
CSV;
    
```



The screenshot shows the AWS Redshift console interface. A query window is open with the following SQL code:

```

26
27 copy etl_bankatm_data.dim_card_type
28 from 's3://etlprojectfolder/dim_card_type/part-00000-9ec5d809-a5a0-407a-8e08-c98823cc7683-c000.csv'
29 iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
30 delimiter ','
31 region 'us-east-1'
32 CSV;
33

```

The status bar at the top indicates the session is connected to 'N. Virginia' and the user is 'upgradhemanthsameer @ 9549-7156-5245'. Below the code, the 'Result 4' tab is selected, showing the following summary:

Summary

- Info:**
 - Load into table 'dim_card_type' completed, 12 record(s) loaded successfully.

Returned rows: 0
Elapsed time: 440ms
Result set query:

```

/* ROEV2-HvgZETtFe6 */
copy etl_bankatm_data.dim_card_type
from 's3://etlprojectfolder/dim_card_type/part-00000-9ec5d809-a5a0-407a-8e08-c98823cc7683-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ','
region 'us-east-1'
CSV;

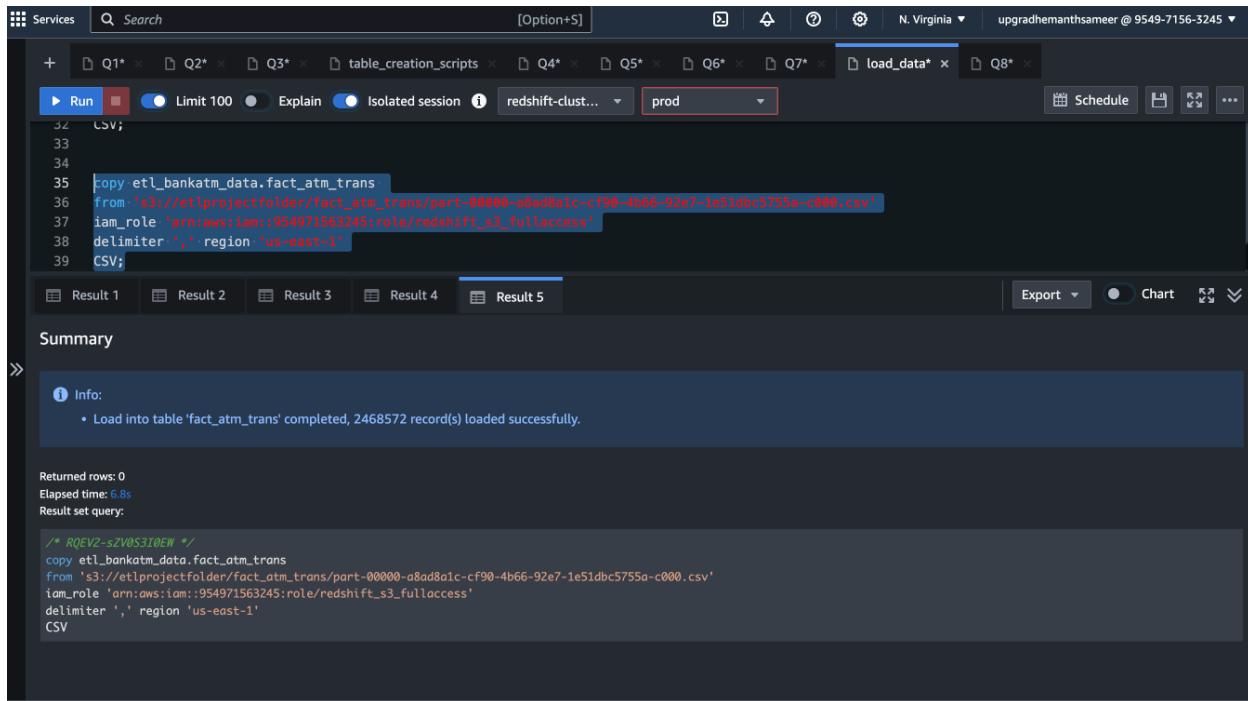
```

Query for copying data to fact_atm_trans table:

```

copy etl_bankatm_data.fact_atm_trans
from
's3://etlprojectfolder/fact_atm_trans/part-00000-a8ad8a1c-cf90-4b66-92e7-1e51dbc5755a-
c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;

```



The screenshot shows a Redshift Data Studio interface with a query editor and a summary panel.

Query Editor:

```
32 CSV;
33
34
35 copy etl_bankatm_data.fact_atm_trans
36 from 's3://etlprojectfolder/fact_atm_trans/part-00000-a8ad8a1c-cf90-4b66-92e7-1e51dbc5755a-c000.csv'
37 iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
38 delimiter ',' region 'us-east-1'
39 CSV;
```

Summary Panel:

Info:

- Load into table 'fact_atm_trans' completed, 2468572 record(s) loaded successfully.

Returned rows: 0
Elapsed time: 6.8s
Result set query:

```
/* RQEY2-sZV0S3IOEW */
copy etl_bankatm_data.fact_atm_trans
from 's3://etlprojectfolder/fact_atm_trans/part-00000-a8ad8a1c-cf90-4b66-92e7-1e51dbc5755a-c000.csv'
iam_role 'arn:aws:iam::954971563245:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV
```