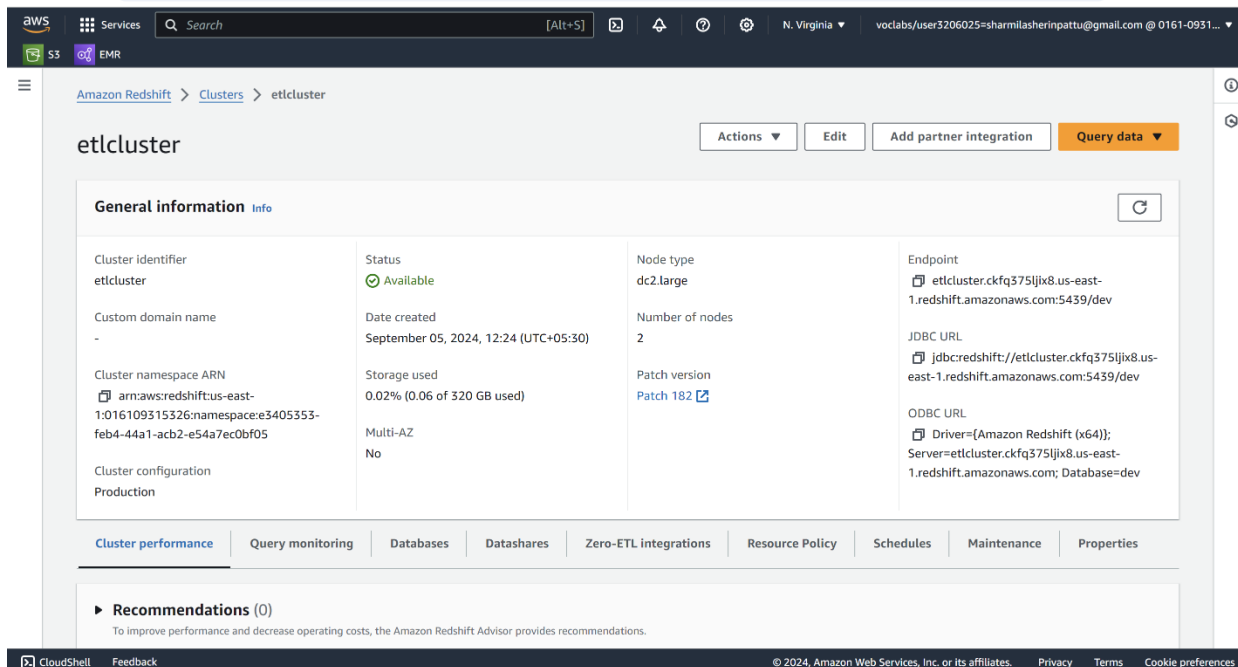


Creation of a Redshift Cluster

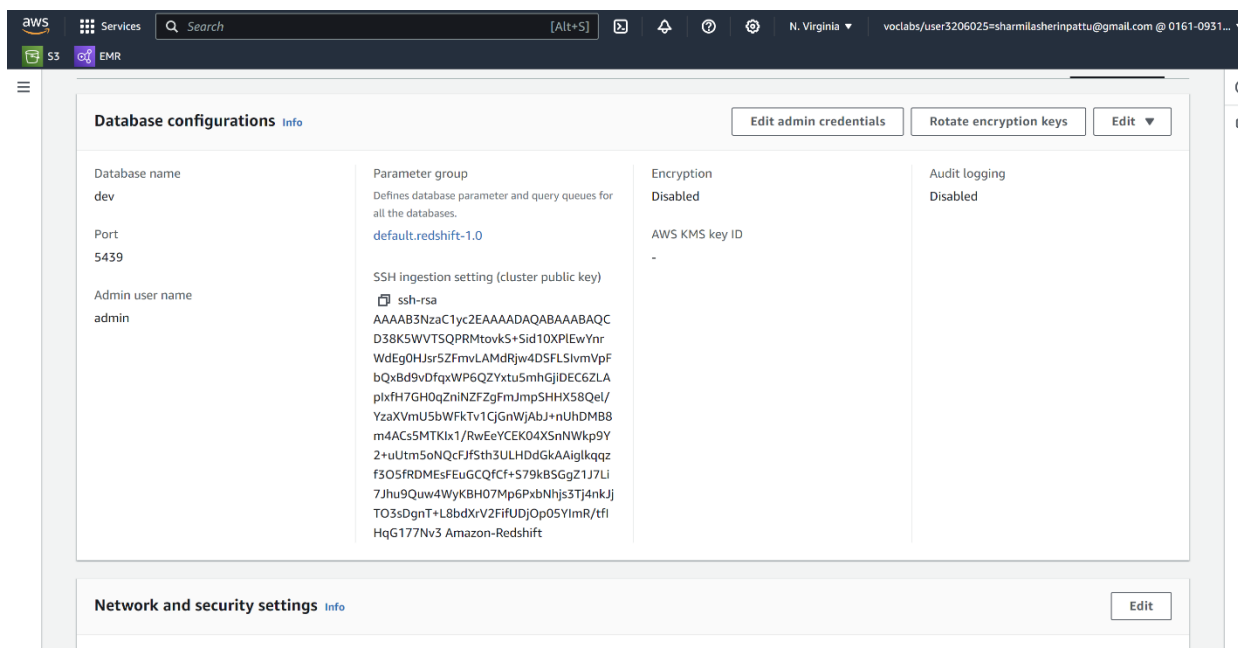
Screenshots of the configuration of the Redshift cluster that I have created:



The screenshot displays the AWS Management Console for the Redshift cluster 'etlcluster'. The cluster is in an 'Available' state. The configuration details are as follows:

Cluster identifier	Status	Node type	Endpoint
etlcluster	Available	dc2.large	etlcluster.ckfq375ljx8.us-east-1.redshift.amazonaws.com:5439/dev
Custom domain name	Date created	Number of nodes	JDBC URL
-	September 05, 2024, 12:24 (UTC+05:30)	2	jdbcredshift://etlcluster.ckfq375ljx8.us-east-1.redshift.amazonaws.com:5439/dev
Cluster namespace ARN	Storage used	Patch version	ODBC URL
arn:aws:redshift:us-east-1:016109315326:namespace:e3405353-feb4-44a1-acb2-e54a7ec0bf05	0.02% (0.06 of 320 GB used)	Patch 182	Driver={Amazon Redshift (x64)}; Server=etlcluster.ckfq375ljx8.us-east-1.redshift.amazonaws.com; Database=dev
Cluster configuration	Multi-AZ		
Production	No		

The console also shows tabs for Cluster performance, Query monitoring, Databases, Datashares, Zero-ETL integrations, Resource Policy, Schedules, Maintenance, and Properties. A Recommendations section is also visible, indicating no recommendations at the moment.

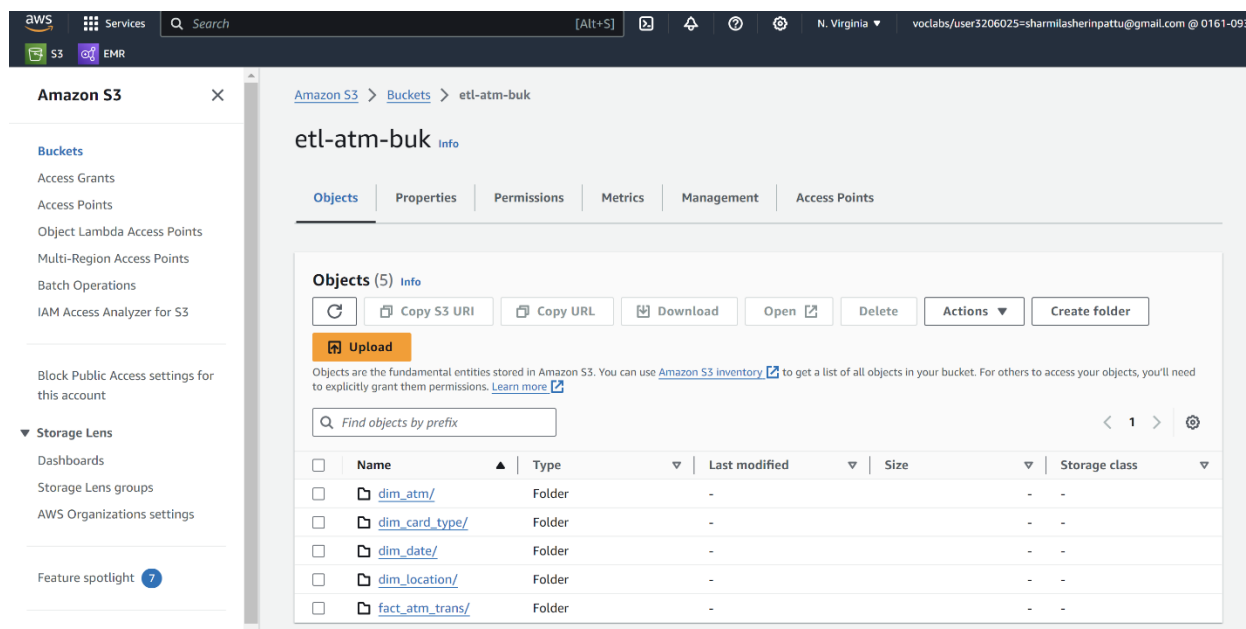


The screenshot displays the AWS Management Console for the database configurations of the Redshift cluster 'dev'. The configurations are as follows:

Database name	Parameter group	Encryption	Audit logging
dev	Defines database parameter and query queues for all the databases.	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5439	default.redshift-1.0	-	
Admin user name			
admin	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQACD38K5WVTSQPRMtvkS+Sid10XPIewYnrWdEg0HJsr5ZFmvLAMdRjw4D5FLslvmVpFbQxBd9vDfqxWP6QZYxtu5mhGjIDEC6ZLaplxfH7GH0qZniNZFZgFmJmpSHX58Qel/YzaXVmU5bWfKtV1CjGnWjAbJ+nUhDMB8m4ACs5MTKlx1/RwEeYCEK04XSnNWkp9Y2+uUtm5oNQcFJf5th3ULHDdGkAAiglkqzF3O5FRDMesFEuGCQCFc+S79kBSGgZ1J7Li7Jhu9Quw4WyKBH07Mp6PxbNhjs3Tj4nkJjTO3sDgnT+L8bdXrV2FifUDjOp05YImr/tfHqG177Nv3 Amazon-Redshift		

The console also shows tabs for Database configurations, Edit admin credentials, Rotate encryption keys, and Edit. A Network and security settings section is also visible, with an Edit button.

Checking the s3 bucket for the five pushed csv files:



Amazon S3 > Buckets > etl-atm-buk

etl-atm-buk info

[Objects](#) [Properties](#) [Permissions](#) [Metrics](#) [Management](#) [Access Points](#)

Objects (5) Info

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#)

[Upload](#)

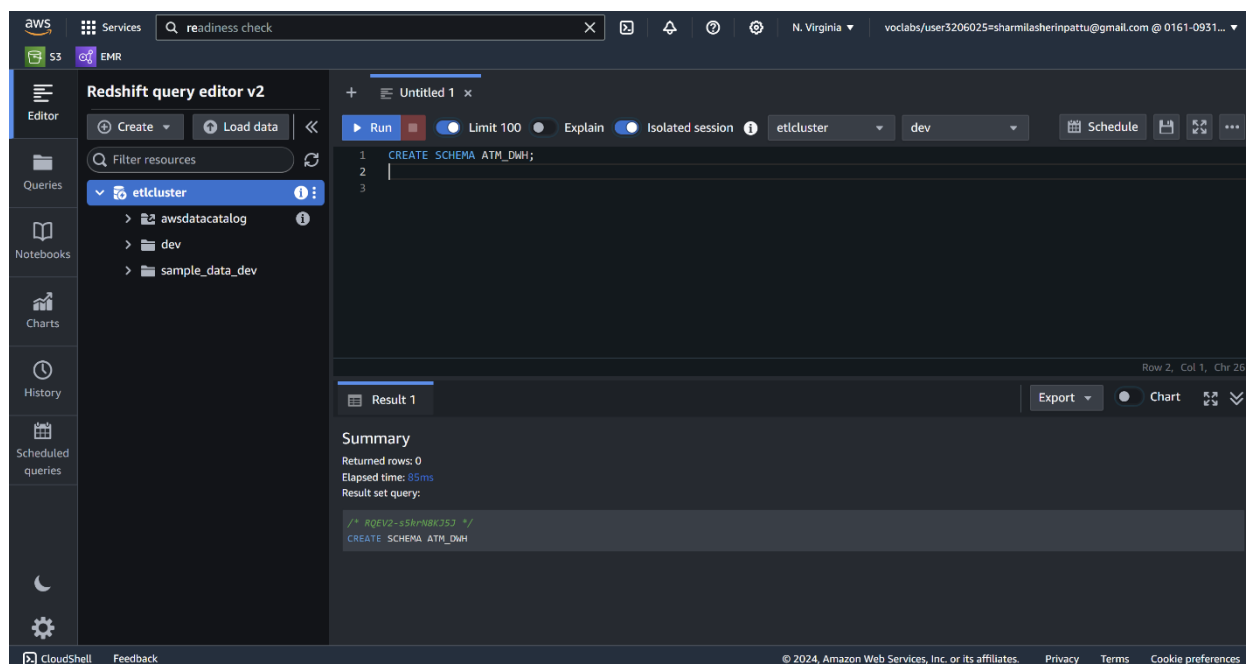
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

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

Create schema:

Query: **create schema ATM_DWH;**



Readiness check

Redshift query editor v2

[Create](#) [Load data](#) [Run](#) [Limit 100](#) [Explain](#) [Isolated session](#) [etlcluster](#) [dev](#) [Schedule](#) [Export](#) [Chart](#)

etlcluster

- awsdatacatalog
- dev
- sample_data_dev

```

1 CREATE SCHEMA ATM_DWH;
2
3

```

Row 2, Col 1, Chr 26

Result 1 [Export](#) [Chart](#)

Summary

Returned rows: 0
Elapsed time: 85ms
Result set query:

```

/* RDEV2--S3/NB/K353 */
CREATE SCHEMA ATM_DWH

```

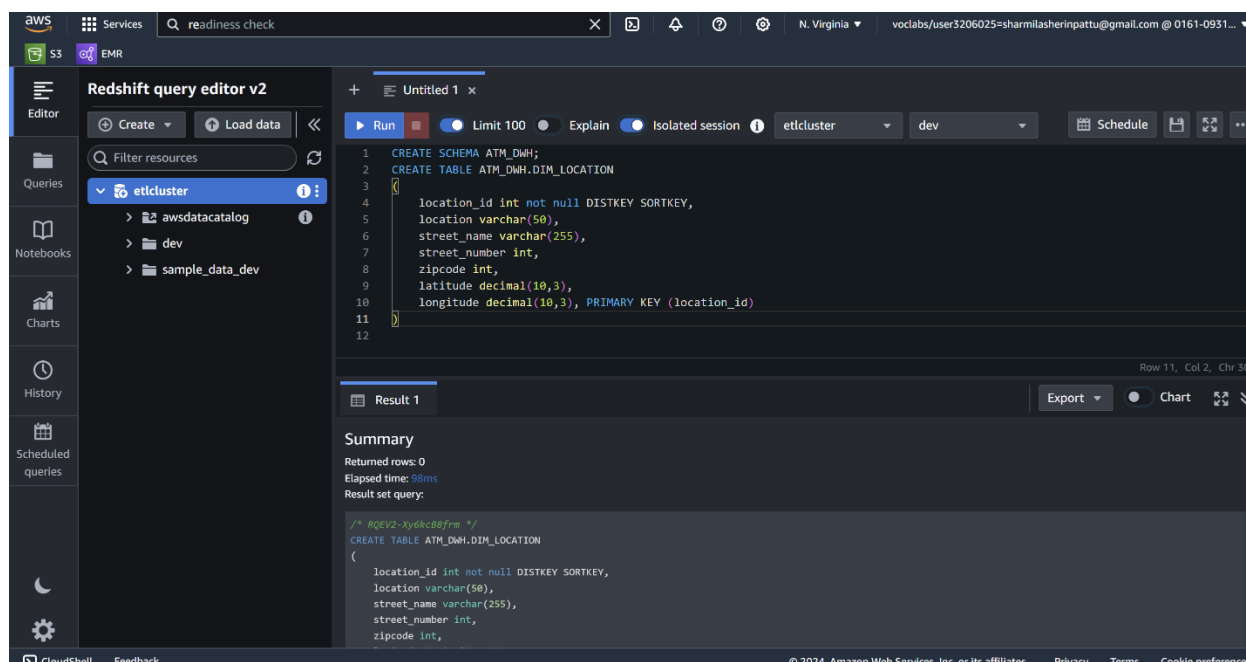
CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Create DIM LOCATION table on Redshift:

Query:

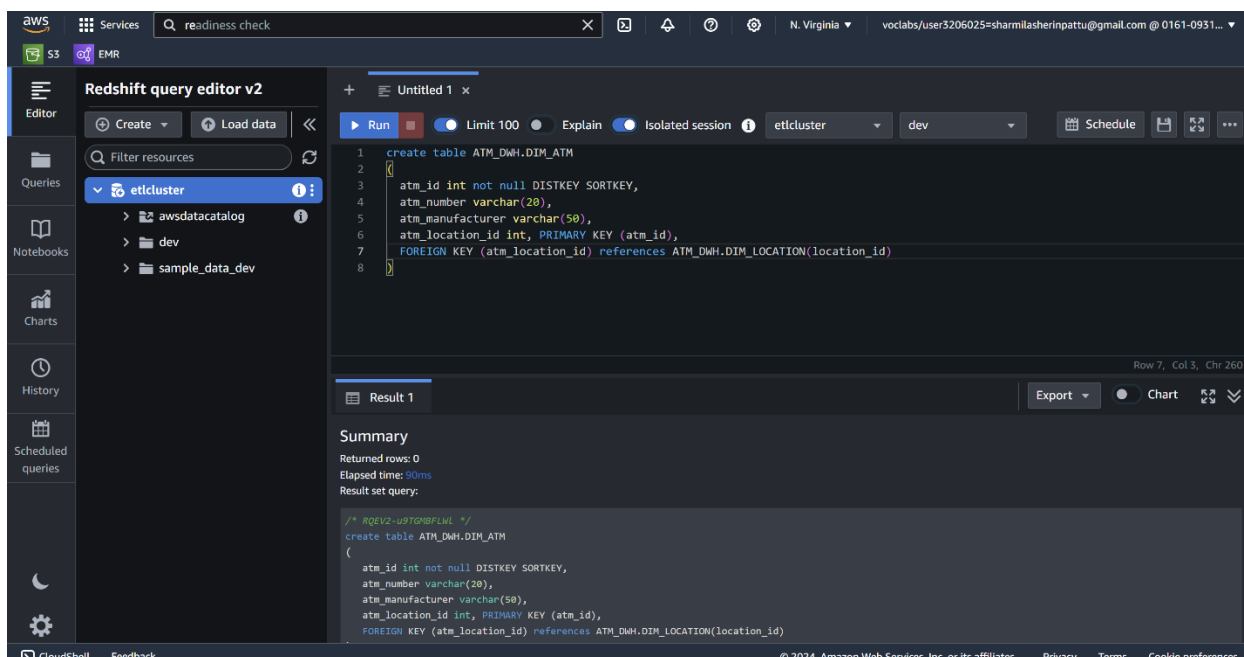
```
create table ATM_DWH.DIM_LOCATION
(
location_id int not null DISTKEY SORTKEY,
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int, lat
decimal(10,3), lon
decimal(10,3),
PRIMARY KEY (location_id));
```



Create DIM ATM table on Redshift:

Query:

```
create table ATM_DWH.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_DWH.DIM_LOCATION(location_id)
);
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area displays a SQL query to create a table named `ATM_DWH.DIM_ATM`. The query is as follows:

```
1 create table ATM_DWH.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, PRIMARY KEY (atm_id),
7   FOREIGN KEY (atm_location_id) references ATM_DWH.DIM_LOCATION(location_id)
8 }
```

The bottom section shows the query execution summary:

- Returned rows: 0
- Elapsed time: 80ms
- Result set query:

Create DIM DATE on Redshift:

Query:

```
create table ATM_DWH.DIM DATE
```

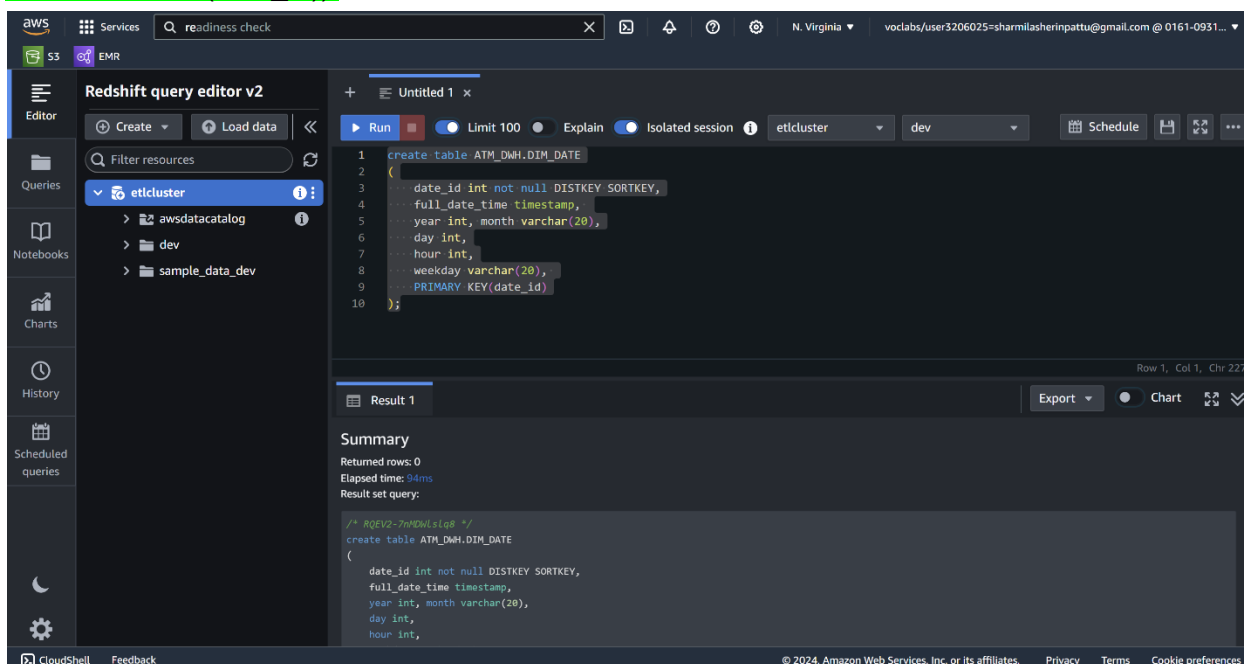
```
date_id int not null DISTKEY SORTKEY,
```

```
full_date time timestamp,
```

```
year int,
```

```
month varchar(20),day int,hour int,weekday varchar(20),
```

```
PRIMARY KEY(date_id));
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area displays a SQL query to create a table named `ATM_DWH.DIM_DATE`. The query is as follows:

```
1 create table ATM_DWH.DIM_DATE
2 {
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int, month varchar(20),
6   day int,
7   hour int,
8   weekday varchar(20),
9   PRIMARY KEY(date_id)
10 }
```

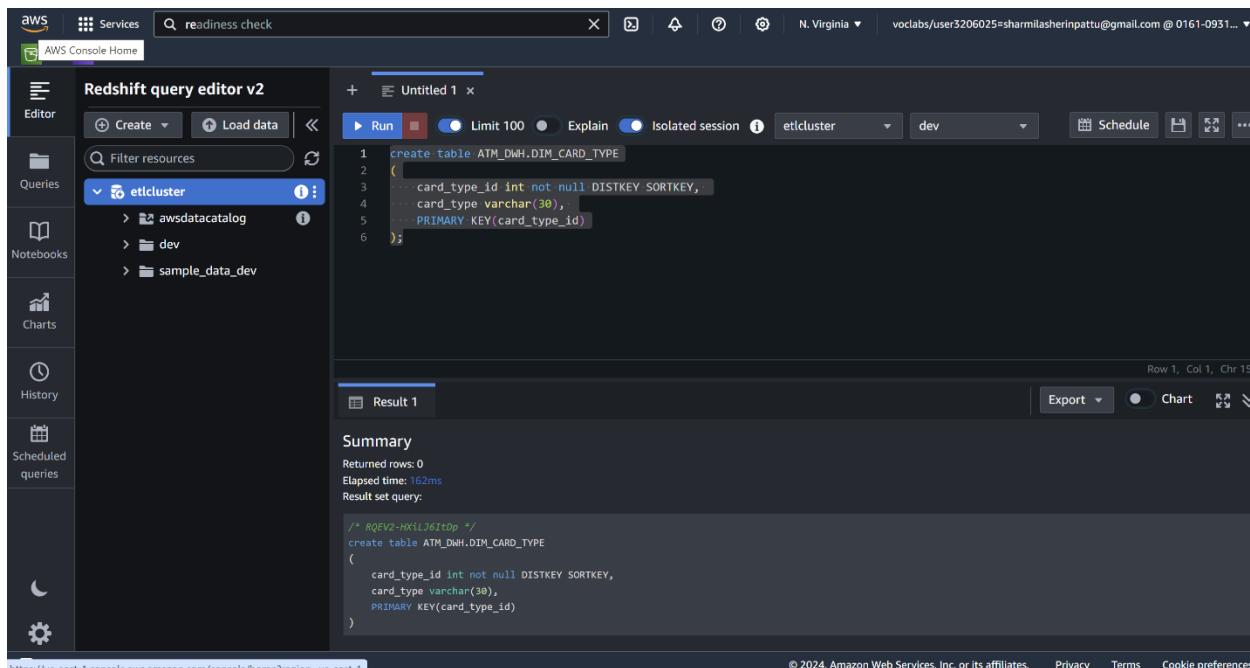
The bottom section shows the query execution summary:

- Returned rows: 0
- Elapsed time: 94ms
- Result set query:

Create DIM_CARD_TYPE on Redshift:

Query:

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

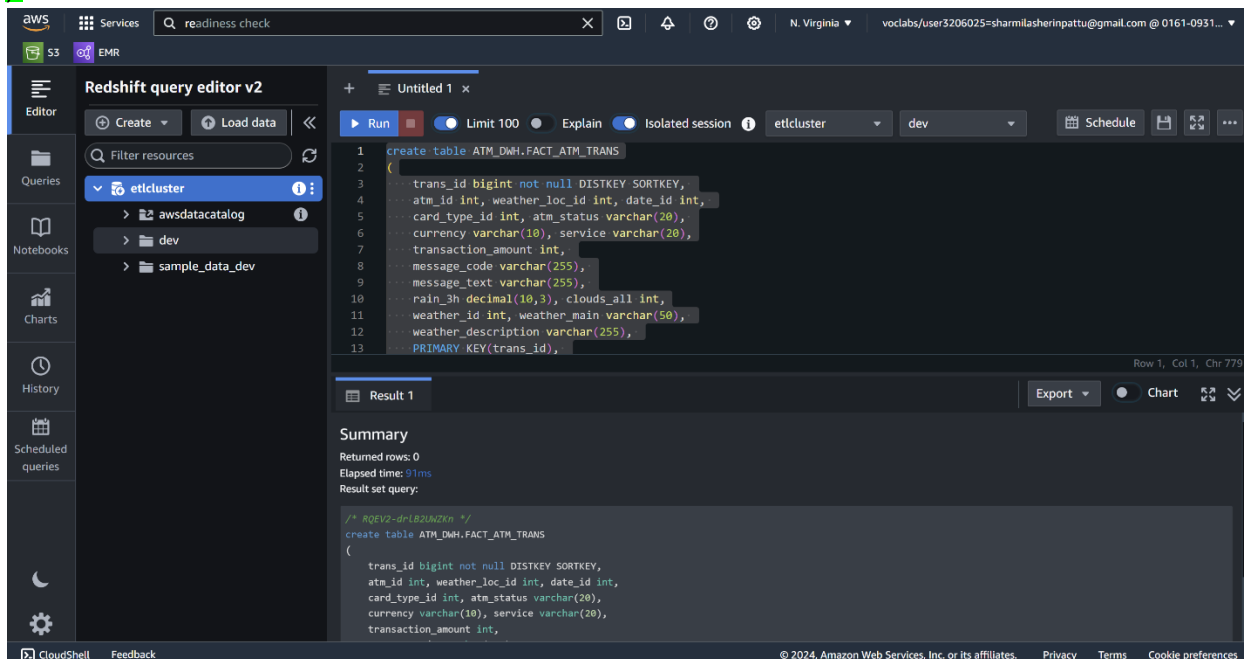


Create FACT_ATM_TRANS on Redshift:

Query:

```
create table ATM_DWH.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(255),
message_text varchar(255),
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_DWH.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references ATM_DWH.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references ATM_DWH.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references ATM_DWH.DIM_CARD_TYPE(card_type_id)
);
```

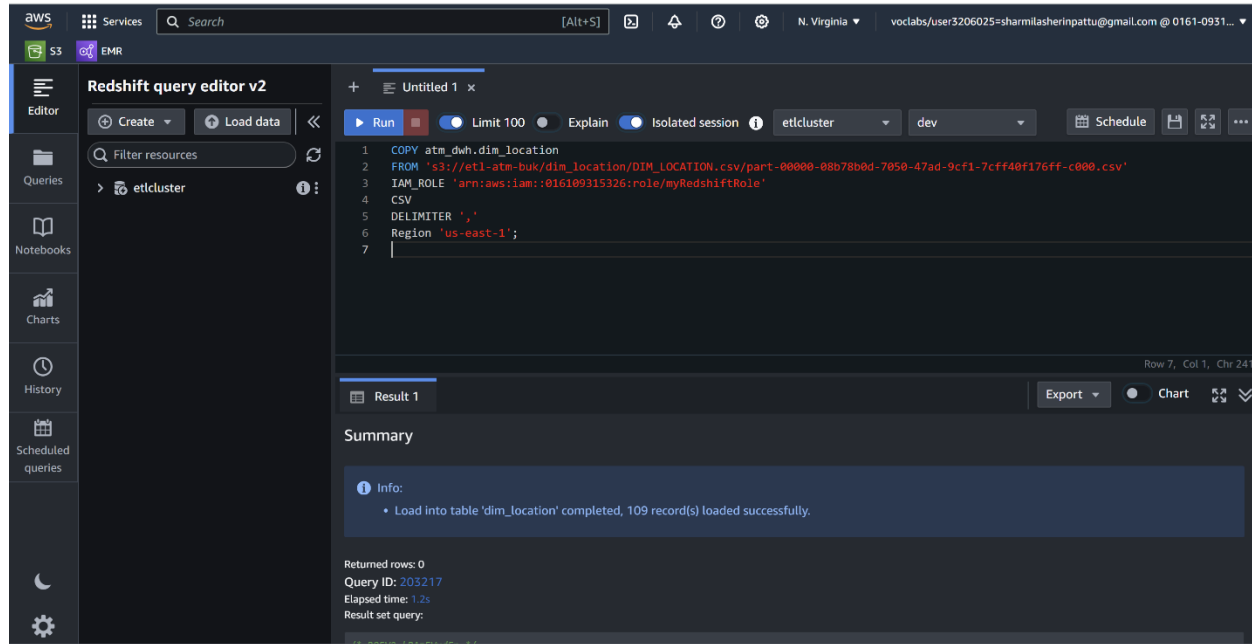


Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

• Load Data to DIM LOCATION:

Query:

```
COPY atm_dwh.dim_location
from 's3://etl-atm-buk/dim_location/DIM_LOCATION.csv/part-00000-08b78b0d-7050-47ad-9cf1-
7cff40f176ff-c000.csv'
iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
CSV
DELIMITER ','
Region 'us-east-1';
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area displays a SQL query in a dark theme. The query is as follows:

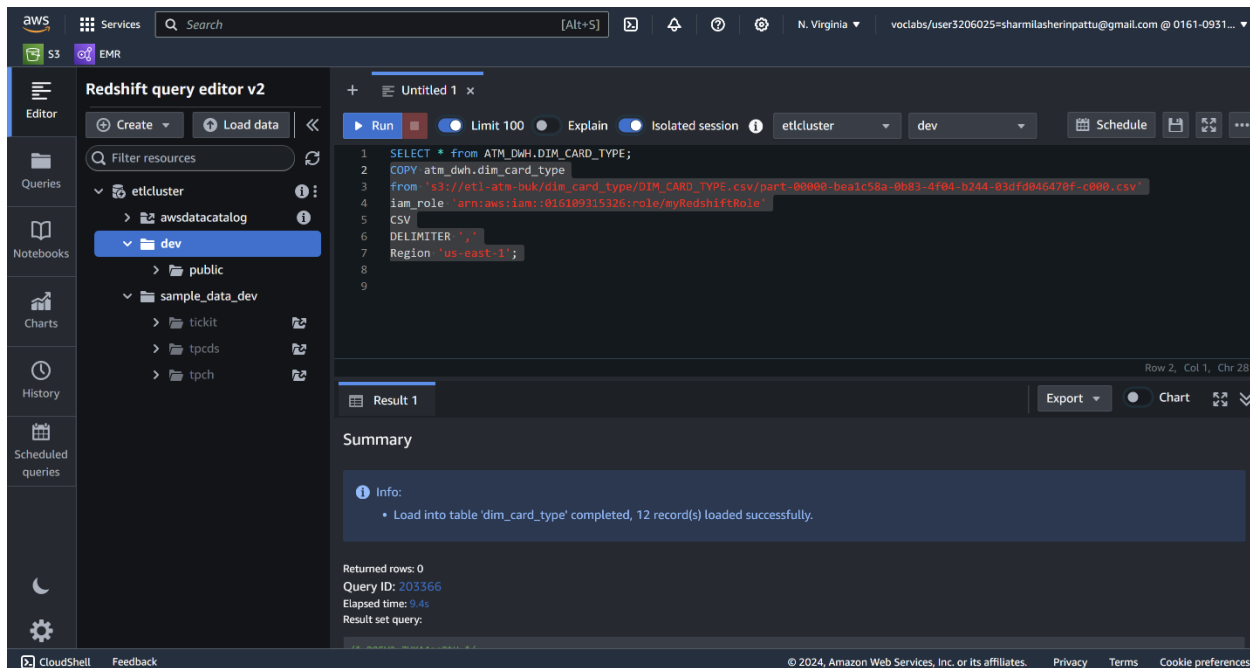
```
1 COPY atm_dwh.dim_location
2 FROM 's3://etl-atm-buk/dim_location/DIM_LOCATION.csv/part-00000-08b78b0d-7050-47ad-9cf1-7c9ff40f176ff-c000.csv'
3 IAM_ROLE 'arn:aws:iam::016109315326:role/myRedshiftRole'
4 CSV
5 DELIMITER ','
6 Region 'us-east-1';
7
```

The query has been executed, and the results are shown in the bottom panel. The summary indicates that the load into the table 'dim_location' was completed successfully, with 109 records loaded. The query ID is 203217, and the elapsed time is 1.2s.

Load Data to DIM_CARD_TYPE:

Query:

```
COPY atm_dwh.dim_card_type
from 's3://etl-atm-buk/dim_card_type/DIM_CARD_TYPE.csv/part-00000-bea1c58a-0b83-4f04-
b244-03dfd046470f-c000.csv'
iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
CSV
DELIMITER ','
Region 'us-east-1';
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area displays a SQL query in a dark theme. The query is as follows:

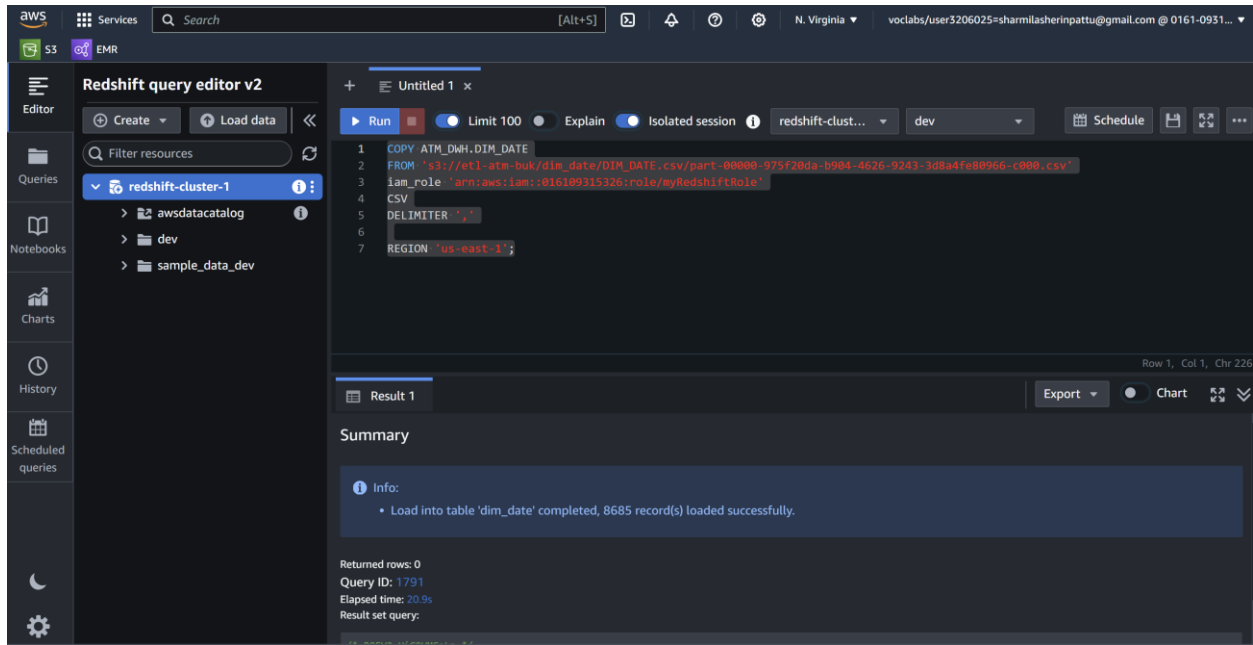
```
1 SELECT * from ATM_DWH.DIM_CARD_TYPE;
2 COPY atm_dwh.dim_card_type
3 from 's3://etl-atm-buk/dim_card_type/DIM_CARD_TYPE.csv/part-00000-bea1c58a-0b83-4f04-b244-83dfd046470f-c000.csv'
4 iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
5 CSV
6 DELIMITER ','
7 Region 'us-east-1';
8
9
```

The query has been executed, and the results are displayed in the bottom section. The summary indicates that the load into the table 'dim_card_type' was completed successfully, with 12 record(s) loaded. The query ID is 203366, and the elapsed time is 9.4s.

Load Data to DIM_DATE:

Query:

```
COPY ATM_DWH.DIM_DATE
FROM 's3://etl-atm-buk/dim_date/DIM_DATE.csv/part-00000-975f20da-b904-4626-9243-3d8a4fe80966-c000.csv'
iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
CSV
DELIMITER ','
REGION 'us-east-1';
```

Load Data to DIM_ATM

Query:

COPY atm_dwh.dim_atm

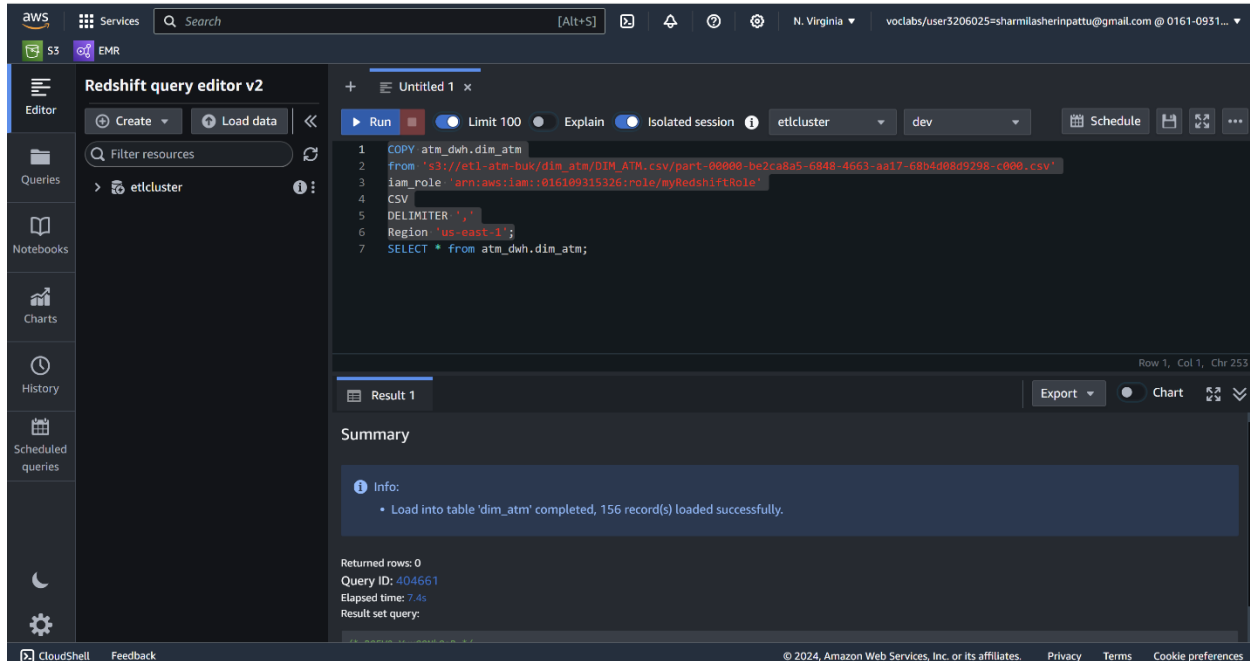
from 's3://etl-atm-buk/dim_atm/DIM_ATM.csv/part-00000-be2ca8a5-6848-4663-aa17-68b4d08d9298-c000.csv'

iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'

CSV

DELIMITER ','

Region 'us-east-1';



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar contains navigation options: Editor, Queries, Notebooks, Charts, History, and Scheduled queries. The main editor area displays a SQL query in a dark theme. The query is as follows:

```
1 COPY atm_dwh.dim_atm
2 from 's3://etl-atm-buk/dim_atm/DIM_ATM.csv/part-00000-be2ca8a5-6848-4663-aa17-68b4d08d9298-c000.csv'
3 iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
4 CSV
5 DELIMITER ','
6 Region 'us-east-1';
7 SELECT * from atm_dwh.dim_atm;
```

Below the query editor, the results section shows a summary of the query execution:

- Info:**
 - Load into table 'dim_atm' completed, 156 record(s) loaded successfully.
- Returned rows: 0
- Query ID: 404661
- Elapsed time: 7.4s
- Result set query:

The bottom of the interface shows the AWS CloudShell environment with the user 'voclabs/user3206025=sharmilasherinpattu@gmail.com' and the region 'N. Virginia'.

Load Data to FACT_ATM_TRANS:

Query:

```
COPY atm_dwh.fact_atm_trans
from 's3://etl-atm-buk/fact_atm_trans/FACT_ATM_TRANS/part-00000-e50ac4ea-5ad7-44b5-
ac5e-8b4b325854e2.csv'
iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
CSV
DELIMITER ','
Region 'us-east-1';
```

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3206025=sharmilasherinpattu@gmail.com @ 0161-0931...

S3

EMR

Editor

Filter resources

etlcluster

etlcluster

Notebooks

Charts

History

Scheduled queries

Untitled 1

Run

Limit 100

Explain

Isolated session

etlcluster

dev

Schedule

```
1 COPY atm_dwh.fact_atm_trans
2 from 's3://etl-atm-buk/fact_atm_trans/FACT_ATM_TRANS/part-00000-e58ac4ea-5ad7-44b5-ac5e-8b4b325854e2-c000.csv'
3 iam_role 'arn:aws:iam::016109315326:role/myRedshiftRole'
4 CSV
5 DELIMITER ','
6 Region 'us-east-1';
```

Result 1

Export

Chart

Summary

Info:

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

Returned rows: 0

Query ID: 404744

Elapsed time: 21.4s

Result set query: