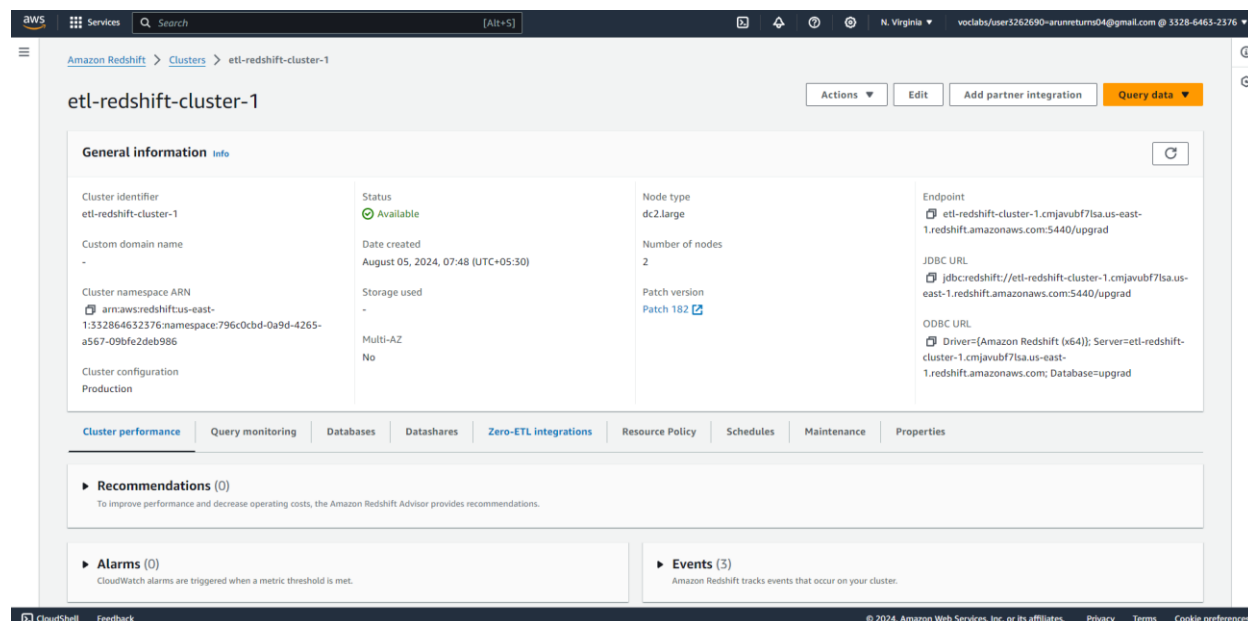


## Creation of a Redshift Cluster

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

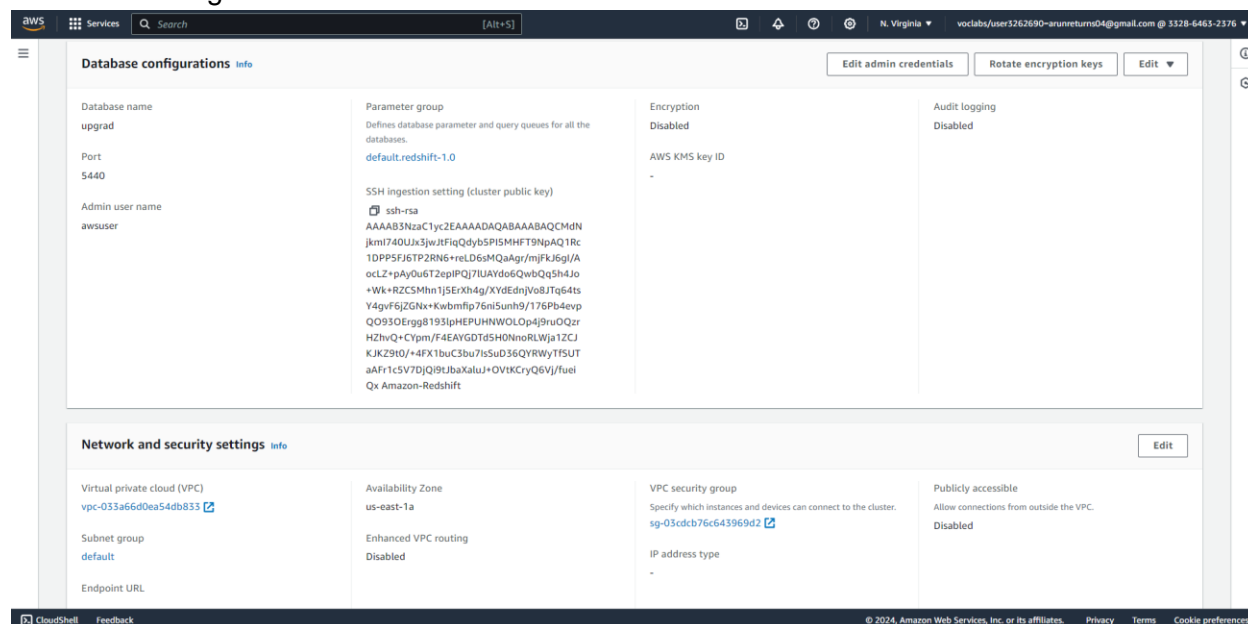


The screenshot displays the AWS Redshift console for the cluster 'etl-redshift-cluster-1'. The 'General information' tab is active, showing the following details:

- Cluster identifier:** etl-redshift-cluster-1
- Status:** Available
- Node type:** dc2.large
- Endpoint:** etl-redshift-cluster-1.cmjavubf7l5a.us-east-1.redshift.amazonaws.com:5440/upgrad
- Custom domain name:** -
- Date created:** August 05, 2024, 07:48 (UTC+05:30)
- Number of nodes:** 2
- Cluster namespace ARN:** arn:aws:redshift:us-east-1:332864632376:namespace:796c0cbd-0a9d-4265-a567-09bfe2deb986
- Patch version:** Patch 182
- Storage used:** -
- Multi-AZ:** No
- Cluster configuration:** Production
- JDBC URL:** jdbc:redshift://etl-redshift-cluster-1.cmjavubf7l5a.us-east-1.redshift.amazonaws.com:5440/upgrad
- ODBC URL:** Driver={Amazon Redshift (x64)}; Server=etl-redshift-cluster-1.cmjavubf7l5a.us-east-1.redshift.amazonaws.com; Database=upgrad

Below the general information, there are tabs for 'Cluster performance', 'Query monitoring', 'Databases', 'Datashares', 'Zero-ETL integrations', 'Resource Policy', 'Schedules', 'Maintenance', and 'Properties'. The 'Recommendations' section shows 0 recommendations. The 'Alarms' section shows 0 alarms. The 'Events' section shows 3 events.

## Database configurations



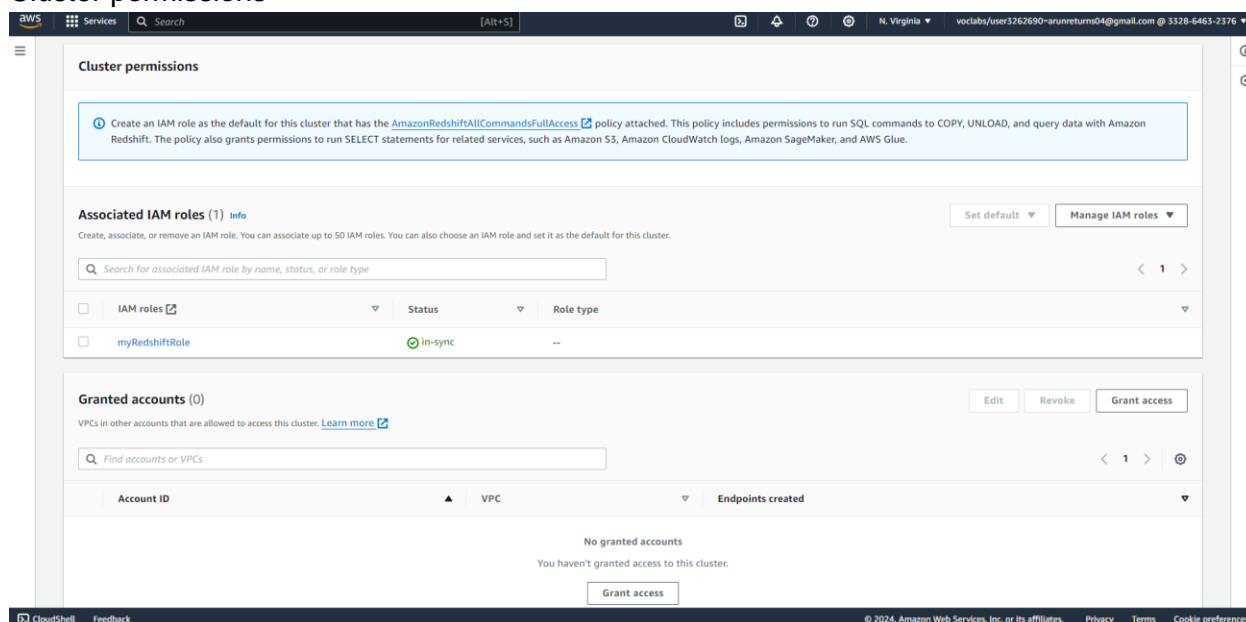
The screenshot displays the AWS Redshift console for the database 'upgrad'. The 'Database configurations' tab is active, showing the following details:

- Database name:** upgrad
- Port:** 5440
- Admin user name:** awsuser
- Parameter group:** default.redshift-1.0
- SSH ingestion setting (cluster public key):** ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQCMdNjkm1740UJx3jwJfFiQDyb5PI5MHFT9NpAQ1RcTDPpSFJ6TP2RN6+reLD6sMQaAgr/mjFk36gl/AocLZ+pAyOu6T2epIPQ7IUAyDo6QwbQq5h4Jo+Wk+RZCSMhn1j5ExXh4g/XYEdnjV08JTq64tsY4gvF6JZGNx+Kwbnfip76niSunh9/176Pb4evpQ93OErqg8193lpHEPUHNWOLop4j9ruOQzrH2hvQ+CYpm/F4EAYGDTd5H0NnoRLWja1ZCJJKZ9t0/+4FX1buC3bu7Isu036QYRWYtISUTaAFr1c5V7DQI9UJbaXalu+OVikCryQ6Vj/fuelQx Amazon-Redshift
- Encryption:** Disabled
- AWS KMS key ID:** -
- Audit logging:** Disabled

Below the database configurations, there are tabs for 'Network and security settings' and 'Edit'. The 'Network and security settings' tab is active, showing the following details:

- Virtual private cloud (VPC):** vpc-033a66d0ea54db833
- Subnet group:** default
- Endpoint URL:** -
- Availability Zone:** us-east-1a
- Enhanced VPC routing:** Disabled
- VPC security group:** sg-03cdcb76c643969d2
- IP address type:** -
- Publicly accessible:** Disabled

## Cluster permissions



**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.

Search for associated IAM role by name, status, or role type

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	myRedshiftRole	In-sync	--

**Granted accounts (0)** [Learn more](#)

VPCs in other accounts that are allowed to access this cluster.

Find accounts or VPCs

Account ID	VPC	Endpoints created
No granted accounts		

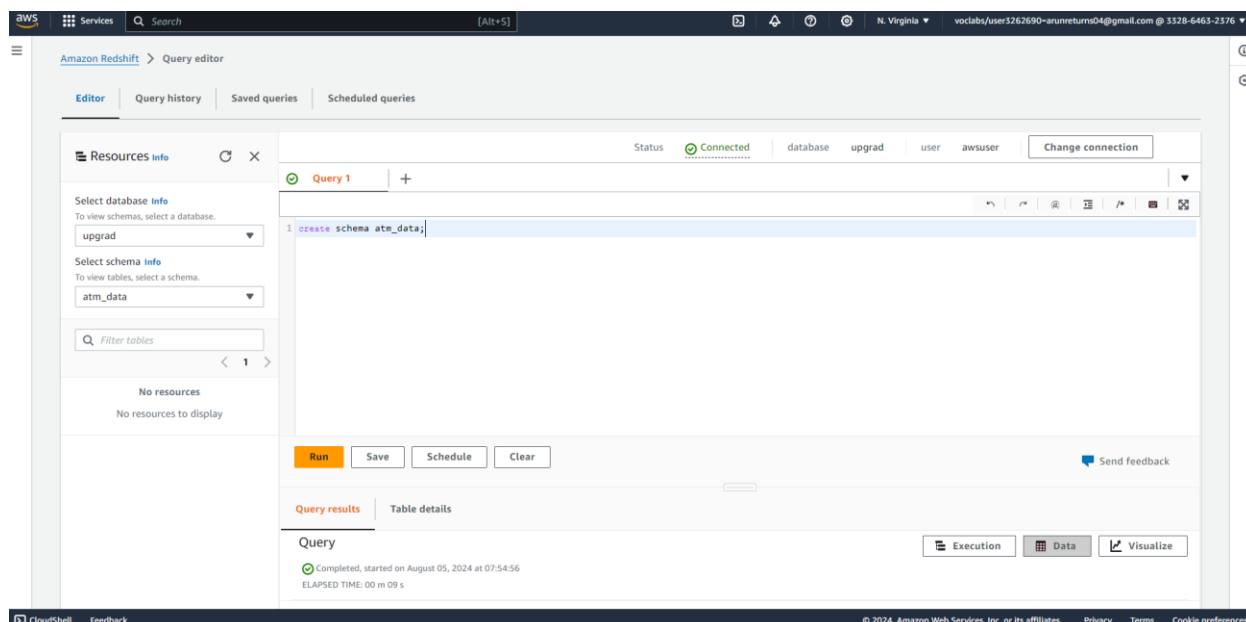
You haven't granted access to this cluster.

[Grant access](#)

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

Query for creating a schema:

```
create schema atm_data;
```



**Amazon Redshift > Query editor**

Editor | Query history | Saved queries | Scheduled queries

**Resources** [Info](#)

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

Select schema [Info](#)  
To view tables, select a schema.  
atm\_data

Filter tables

No resources  
No resources to display

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

Query 1

```
1 create schema atm_data;
```

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

[Send feedback](#)

**Query results** | Table details

**Query** [Execution](#) [Data](#) [Visualize](#)

Completed, started on August 05, 2024 at 07:54:56  
ELAPSED TIME: 00 m 09 s

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

- **Creating Local 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)
);
```

✓ Query 1 +

```
3 -- Creating Local dimension table
4 CREATE TABLE atm_data.DIM_LOCATION
5 (
6     location_id INT NOT NULL DISTKEY SORTKEY,
7     location VARCHAR(50),
8     streetname VARCHAR(255),
9     street_number INT,
10    zipcode INT,
11    lat DECIMAL(10,3),
12    lon DECIMAL(10,3),
13    PRIMARY KEY(location_id)
14 );
15
16 --Creating atm dimension table
```

Run

Save

Schedule

Clear

- **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)
);
```

✓ Query 1
+

```

15
16 --Creating atm dimension table
17
18 CREATE TABLE atm_data.DIM_ATM
19 (
20     atm_id INT NOT NULL DISTKEY SORTKEY,
21     atm_number VARCHAR(20),
22     atm_manufacturer VARCHAR(50),
23     atm_location_id INT,
24     PRIMARY KEY(atm_id),
25     FOREIGN KEY (atm_location_id) REFERENCES atm_data.DIM_LOCATION(location_id)
26 );
27

```

Run
Save
Schedule
Clear

- 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)
);

```

✓ Query 1
+

```

27
28 --Creating date dimension table
29
30 CREATE TABLE atm_data.DIM_DATE
31 (
32     date_id INT NOT NULL DISTKEY SORTKEY,
33     full_date_time TIMESTAMP,
34     year INT,
35     month VARCHAR(20),
36     day INT,
37     hour INT,
38     weekday VARCHAR(20),
39     PRIMARY KEY(date_id)
40 );

```

Run
Save
Schedule
Clear

- **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)
);
```

✓ Query 1

+

```
38     weekday VARCHAR(20),
39     PRIMARY KEY(date_id)
40 );
41
42 --Creating card type dimension table
43 CREATE TABLE atm_data.DIM_CARD_TYPE
44 (
45     card_type_id INT NOT NULL DISTKEY SORTKEY,
46     card_type VARCHAR(30),
47     PRIMARY KEY(card_type_id)
48 );
```

- **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)
);
```

```

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),
57     rain_3h DECIMAL(10,3),
58     clouds_all INT,
59     weather_id INT,
60     weather_main VARCHAR(50),
61     weather_description VARCHAR(255),
62     PRIMARY KEY (trans_id),
63     FOREIGN KEY (weather_loc_id) REFERENCES atm_data.DIM_LOCATION(location_id),
64     FOREIGN KEY (atm_id) REFERENCES atm_data.DIM_ATM(atm_id),
65     FOREIGN KEY (date_id) REFERENCES atm_data.DIM_DATE(date_id),
66     FOREIGN KEY (card_type_id) REFERENCES atm_data.DIM_CARD_TYPE(card_type_id)
67 );
68

```

Run

Save

Schedule

Clear

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-s3-1/dim_location/part-00000-6edaea43-bb5d-4253-a499-
c618a3e566a8-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```

- Copying the data to dim\_atm table

✓ Query 1 +

```
//
78 --* Copying the data to dim_location table
79
80 FROM 's3://upgrad-bucket-s3-1/dim_location/part-00000-340cf3f0-da30-4d93-b125-bc3c543a4830-c000.csv'
81 IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
82 DELIMITER ','
83 REGION 'us-east-1'
84 CSV;
```

Query results Table details

dim\_location  
atm\_data.dim\_location

Show schema Preview data

location_id	location	streetname	street_number	zipcode	lat	lon
0	Kolding	Vejlevej	135	6000	55.505	9.457
3	Odense	FÅrÅlledvej	3	5000	55.394	10.370
8	Skive	Adelgade	8	7800	56.567	9.027
12	Hobro	Adelgade	31	9500	56.638	9.794
13	Nordkraft	Kjellerups Torv	1	9000	57.047	9.932
17	Hasseris	Hasserisvej	113	9000	57.044	9.898
21	Silkeborg	Borgergade	36	8600	56.179	9.552
26	NykÅrÅ_bing Mors	Kirketorvet	1	7900	56.795	8.860
30	Menu KÅrÅ_bmand Klarup	Klarupvej	52	9270	57.013	10.046
31	Aarhus	SÅrÅ_nder Alle	11	8000	56.153	10.206

```
COPY atm_data.dim_atm
FROM 's3://upgrad-bucket-s3-1/dim_atm/part-00000-5bdb3bbc-35fc-44c2-ac36-05aea46e0bf8-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;
```

```
COPY atm_data.dim_atm
FROM 's3://upgrad-bucket-s3-1/dim_atm/part-00000-e95db0ed-a34e-4d59-a11d-2fdd5798ee9c-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;
```

- Copying the data to dim\_date table

```
COPY atm_data.dim_date
FROM 's3://upgrad-bucket-s3-1/dim_date/part-00000-10328b66-88c7-402f-ac6f-cd69109ef5bd-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV
TIMEFORMAT 'auto';
```

Query 1

```

90 DELIMITER ','
91 REGION 'us-east-1'
92 CSV;
93
94 --Copying the data to dim_date table
95 COPY atm_data.dim_date
96 FROM 's3://upgrad-bucket-s3-1/dim_date/part-00000-b2e79b52-90df-4034-a07c-f7918d81ddbf-c000.csv'
97 IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
98 DELIMITER ','
99 REGION 'us-east-1'
100 CSV
101 TIMEFORMAT 'auto';

```

- Copying the data to dim\_card\_type table

```

COPY atm_data.dim_card_type
FROM 's3://upgrad-bucket-s3-1/dim_card_type/part-00000-3529d67c-af4c-4310-98a5-624c707e575c-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```

```

103 --Copying the data to dim_card_type table
104 COPY atm_data.dim_card_type
105 FROM 's3://upgrad-bucket-s3-1/dim_card_type/part-00000-a5e1a0a1-72a8-42cc-9d45-9953f8db7bfe-c000.csv'
106 IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
107 DELIMITER ','
108 REGION 'us-east-1'
109 CSV;
110
111 --

```

Run

Save

Schedule

Clear

- Copying the data to fact\_atm\_trans table

```

COPY atm_data.fact_atm_trans
FROM 's3://upgrad-bucket-s3-1/fact_atm_trans/part-00000-b00b2638-1f9b-4e95-8ec2-52c4936926ef-c000.csv'
IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```



```

111 --Copying the data to fact_atm_trans table
112 COPY atm_data.fact_atm_trans
113 FROM 's3://upgrad-bucket-s3-1/fact_atm_trans/part-00000-38e20993-698e-4785-b511-38fef6b4559f-c000.csv'
114 IAM_ROLE 'arn:aws:iam::332864632376:role/myRedshiftRole'
115 DELIMITER ','
116 REGION 'us-east-1'
117 CSV;

```

Run
Save
Schedule
Clear

## Checking whether the data is loaded or not

```

select * from atm_data.fact_atm_trans
limit 1;

```

```

119 --Checking whether the data is loaded or not
120 select * from atm_data.fact_atm_trans
121 limit 3;

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query 2072
Execution
Data
Visualize

Completed, started on August 05, 2024 at 08:11:35  
ELAPSED TIME: 00 m 02 s

Rows returned (3)
Export

Search rows

trans_id	atm_id	weather_loc_id	date_id	card_type_id	atm_status	currency	service	transaction_amount	message_c
5	133	56	0	1	Active	DKK	Withdrawal	8981	
15	142	90	0	5	Active	DKK	Withdrawal	5508	