

Creation of Redshift Cluster

Creation of IAM Role (for Redshift to get access to S3):

redshift_full_access_S3

Delete

Allows Redshift clusters to call AWS services on your behalf.

Summary

Edit

Creation date

December 17, 2022, 10:23 (UTC+05:30)

ARN

arn:aws:iam::360448222316:role/redshift_full_access_S3

Last activity

None

Maximum session duration

1 hour

Permissions

Trust relationships

Tags

Access Advisor

Revoke sessions

Permissions policies (1) Info

You can attach up to 10 managed policies.

Simulate Remove Add permissions

< 1 >

☐

Policy name

Type

Description

☐

AmazonS3FullAccess

AWS managed

Provides full access to all buckets via the AWS Management Console.

Permissions boundary - (not set) Info

Creation of Cluster Subnet Group

Cluster subnet group custom-cluster-subnet-group-1 was create successfully

Amazon Redshift

>

Configurations

>

Subnet groups

>

Subnet group

custom-cluster-subnet-group-1

Delete

Modify

Cluster subnet group details

VPC ID

Description

Status

vpc-5da27a20

CUSTOM-cluster-subnet-group-1

Complete

Attached clusters

This subnet group has no attached clusters

Subnets (1)

< 1 >

Availability Zone

Subnet ID

CIDR block

us-east-1a

subnet-b565ecea

172.31.32.0/20

Redshift Cluster Configuration:

Figure 1: Type of node (dc2.large) and number of nodes (2)

Amazon Redshift > Clusters > redshift-etl

redshift-etl

Actions Edit Add partner integration Query data

General information

| | | | |
|---|--|------------------------|---|
| Cluster identifier redshift-etl | Status Available | Node type dc2.large | Endpoint redshift-etl.cfamr9e9iyx.us-east-1.redshift.ama... |
| Cluster namespace 8eb23e14-df9c-4fa9-b03b-842d489622a3 | Date created December 17, 2022, 10:56 (UTC+05:30) | Number of nodes 2 | JDBC URL jdbc:redshift://redshift-etl.cfamr9e9iyx.us-east-1... |
| Cluster configuration Production | Storage used - | | ODBC URL Driver={Amazon Redshift (x64)}; Server=redshift-... |
| | Multi-AZ No | | |

Cluster performance Query monitoring Schedules Maintenance Properties

Figure 2: Database configuration, Network and Security Groups

Cluster performance Query monitoring Schedules Maintenance Properties

Database configurations

Edit admin credentials Rotate encryption keys Edit

| | | | |
|----------------------------|---|---|---------------------------|
| Database name etl_atm | Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0 | Encryption Disabled AWS KMS key ID - | Audit logging Disabled |
| Port 5990 | SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB... | | |
| Admin user name awsuser | | | |

Network and security settings

Edit

| | | | |
|---|----------------------------------|--|--|
| Virtual private cloud (VPC) vpc-5da27a20 | Availability Zone us-east-1a | VPC security group Specify which instances and devices can connect to the cluster. sg-abeeb79c | Publicly accessible Allow connections from outside the VPC. Disabled |
| Subnet custom-cluster-subnet-group-1 | Enhanced VPC routing Disabled | | |
| Endpoint URL - | | | |

Figure 3: IAM role associated with the cluster

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.

Set default Manage IAM roles

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

| <input type="checkbox"/> | IAM roles | Status | Role type |
|--------------------------|-------------------------|---------|-----------|
| <input type="checkbox"/> | redshift_full_access_S3 | In-sync | -- |

Figure 4: Two Compute nodes and one Leader node

Node IP addresses (3)

| Node role | Public IP address | Private IP address |
|-----------|-------------------|--------------------|
| Leader | 18.205.5.137 | 172.31.41.9 |
| Compute-0 | 44.213.51.175 | 172.31.35.243 |
| Compute-1 | 23.23.103.161 | 172.31.38.221 |

View the files created in S3 bucket:

5 folders are created, each for fact/dimension table

Amazon S3 > Buckets > atm-upgrad-proj

atm-upgrad-proj [Info](#)

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (5)

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](#)

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

| <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 | - | - | - |

ATM: Dimension table

Amazon S3 > Buckets > atm-upgrad-proj > dim_atm/

dim_atm/ [Copy S3 URI](#)

Objects | Properties

Objects (2)

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](#)

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

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|--|------|---|--------|---------------|
| <input type="checkbox"/> | _SUCCESS | - | December 17, 2022, 12:24:20 (UTC+05:30) | 0 B | Standard |
| <input type="checkbox"/> | part-00000-63c3a61e-ee5d-49e7-9e0e-5cd84c8b1f53-c000.csv | csv | December 17, 2022, 12:24:20 (UTC+05:30) | 2.4 KB | Standard |

Card Type: Dimension table

Amazon S3 > Buckets > atm-upgrad-proj > dim_card_type/

Copy S3 URI

dim_card_type/

Objects | Properties

Objects (2)

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](#)

Refresh

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

⚙

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|--|------|---|---------|---------------|
| <input type="checkbox"/> | _SUCCESS | - | December 17, 2022, 12:24:27 (UTC+05:30) | 0 B | Standard |
| <input type="checkbox"/> | part-00000-adf96d01-6ff5-478a-afe4-1465f81ba096-c000.csv | csv | December 17, 2022, 12:24:27 (UTC+05:30) | 183.0 B | Standard |

Date: Dimension table

Amazon S3 > Buckets > atm-upgrad-proj > dim_date/

Copy S3 URI

dim_date/

Objects | Properties

Objects (2)

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](#)

Refresh

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

⚙

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|--|------|---|----------|---------------|
| <input type="checkbox"/> | _SUCCESS | - | December 17, 2022, 12:24:24 (UTC+05:30) | 0 B | Standard |
| <input type="checkbox"/> | part-00000-f78dec33-e586-4af0-b409-f191a27758d7-c000.csv | csv | December 17, 2022, 12:24:24 (UTC+05:30) | 470.3 KB | Standard |

Location: Dimension table

Amazon S3 > Buckets > atm-upgrad-proj > dim_location/

Copy S3 URI

dim_location/

Objects | Properties

Objects (2)

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](#)

Refresh

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

⚙

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|--|------|---|--------|---------------|
| <input type="checkbox"/> | _SUCCESS | - | December 17, 2022, 12:24:17 (UTC+05:30) | 0 B | Standard |
| <input type="checkbox"/> | part-00000-b41f886c-e83d-4982-9d68-48a543b8cceb-c000.csv | csv | December 17, 2022, 12:24:17 (UTC+05:30) | 5.6 KB | Standard |

ATM Transactions: Fact table

Amazon S3 > Buckets > atm-upgrad-proj > fact_atm_trans/

fact_atm_trans/ Copy S3 URI

Objects | Properties

Objects (2)
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](#)

Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|--|------|---|----------|---------------|
| <input type="checkbox"/> | _SUCCESS | - | December 17, 2022, 12:24:58 (UTC+05:30) | 0 B | Standard |
| <input type="checkbox"/> | part-00000-f05e0e52-e198-4d2c-bca2-9b3d3fd47f74-c000.csv | csv | December 17, 2022, 12:24:53 (UTC+05:30) | 201.4 MB | Standard |

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

1. Creation of Schema

```
create schema atm_info;
```

Query 1 × | Query 2 × | Query 3 × | ✓ Query 4 × | +

↶ ↷ @ ≡ / * 📄 🔍

```
1 create schema atm_info;
2
```

Run Save Schedule Clear Send feedback

Query results | Table details

Query Execution Data Visualize

✓ Completed, started on November 20, 2022 at 22:32:23
ELAPSED TIME: 00 m 23 s

2. Queries to create the various dimension and fact tables:

- Creation of DIM_LOCATION table

```
create table atm_info.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 a SQL query editor with a tab labeled 'Query 4'. The query text is as follows:

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

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is selected, showing the following information:

Query

Completed, started on November 20, 2022 at 22:33:53
ELAPSED TIME: 00 m 10 s

At the bottom, there are three tabs: 'Execution', 'Data', and 'Visualize'.

- Creation of DIM_ATM table

```
create table atm_info.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_info.DIM_LOCATION(location_id)
);
```

Query 1 X | Query 2 X | Query 3 X | ✓ Query 4 X | +

↶ ↷ @ ≡ /* 📄 🗨

10 lat decimal(10,3),
11 lon decimal(10,3),
12 PRIMARY KEY(location_id)
13);
14
15 create table atm_info.DIM_ATM(
16 atm_id int not null DISTKEY SORTKEY,
17 atm_number varchar(20),
18 atm_manufacturer varchar(50),
19 atm_location_id int,
20 PRIMARY KEY(atm_id),
21 FOREIGN KEY(atm_location_id) references atm_info.DIM_LOCATION(location_id)
22);

Run Save Schedule Clear

Send feedback

Query results | Table details

Query

Execution Data Visualize

✓ Completed, started on November 20, 2022 at 22:34:41
ELAPSED TIME: 00 m 03 s

- Creation of DIM_DATE table

```
create table atm_info.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 X | Query 2 X | Query 3 X | **Query 4** X | +

```
22 );
23
24 create table atm_info.DIM_DATE
25 (
26     date_id int not null DISTKEY SORTKEY,
27     full_date_time timestamp,
28     year int,
29     month varchar(20),
30     day int,
31     hour int,
32     weekday varchar(20),
33     PRIMARY KEY(date_id)
34 );
```

Run Save Schedule Clear [Send feedback](#)

Query results | Table details

Query

Completed, started on November 20, 2022 at 22:35:23
ELAPSED TIME: 00 m 03 s

Execution Data Visualize

- Creation of DIM_CARD_TYPE table

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

Query 1 X | Query 2 X | Query 3 X | **Query 4** X | +

```
29     month varchar(20),
30     day int,
31     hour int,
32     weekday varchar(20),
33     PRIMARY KEY(date_id)
34 );
35
36 create table atm_info.DIM_CARD_TYPE
37 (
38     card_type_id int not null DISTKEY SORTKEY,
39     card_type varchar(30),
40     PRIMARY KEY(card_type_id)
41 );
```

Run Save Schedule Clear [Send feedback](#)

Query results | Table details

Query

Completed, started on November 20, 2022 at 22:36:04
ELAPSED TIME: 00 m 03 s

Execution Data Visualize

- Creation of FACT_ATM_TRANS table


```

create table atm_info.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_info.DIM_LOCATION(location_id),
    FOREIGN KEY(atm_id) REFERENCES atm_info.DIM_ATM (atm_id),
    FOREIGN KEY(date_id) REFERENCES atm_info.DIM_DATE (date_id)
);

```

```

43 create table atm_info.FACT_ATM_TRANS
44 (
45     trans_id bigint not null DISTKEY SORTKEY,
46     atm_id int,
47     weather_loc_id int,
48     date_id int,
49     card_type_id int,
50     atm_status varchar(20),
51     currency varchar(10),
52     service varchar(20),
53     transaction_amount int,
54     message_code varchar(255),
55     message_text varchar(255),
56     rain_3h decimal(10,3),
57     clouds_all int,
58     weather_id int,
59     weather_main varchar(50),
60     weather_description varchar(255),
61     PRIMARY KEY(trans_id),
62     FOREIGN KEY(weather_loc_id) REFERENCES atm_info.DIM_LOCATION(location_id),
63     FOREIGN KEY(atm_id) REFERENCES atm_info.DIM_ATM (atm_id),
64     FOREIGN KEY(date_id) REFERENCES atm_info.DIM_DATE (date_id)

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

Visualize

Completed, started on November 20, 2022 at 22:37:33

ELAPSED TIME: 00 m 28 s

3. Loading data from Amazon S3 buckets in the appropriate tables in a Redshift cluster

- Load data into DIM_LOCATION table

```
copy atm_info.DIM_LOCATION
from 's3://atm-upgrad-proj/dim_location/part-00000-b41f886c-e83d-4982-9d68-48a543b8cceb-
c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```

The screenshot displays the Amazon Redshift console interface. On the left, the 'Resources' panel shows the database 'etl_atm' and schema 'atm_info' selected. The main query editor area contains a SQL statement for loading data from an S3 bucket into the 'DIM_LOCATION' table. The query is as follows:

```
1 copy atm_info.DIM_LOCATION
2 from 's3://atm-upgrad-proj/dim_location/part-00000-aebbd4c8-4639-4321-a46b-2315d3e7dd2e-c000.csv'
3 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
4 delimiter ',' region 'us-east-1'
5 CSV;
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is active, showing that the query (Query 243) was completed successfully on December 17, 2022, at 11:03:46, with an elapsed time of 00 m 25 s. The 'Table details' tab is also visible.

- Load data into DIM_ATM table

```
copy atm_info.DIM_ATM
from 's3://atm-upgrad-proj/dim_atm/part-00000-63c3a61e-ee5d-49e7-9e0e-5cd84c8b1f53-
c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```

Query 4 X | Query 5 X | Query 1 X | Query 2 X | Query 3 X | + | ▼

```
1 copy atm_info.DIM_LOCATION
2 from 's3://atm-upgrad-proj/dim_location/part-00000-aebbd4c8-4639-4321-a46b-2315d3e7dd2e-c000.csv'
3 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
4 delimiter ',' region 'us-east-1'
5 CSV;
6
7 copy atm_info.DIM_ATM
8 from 's3://atm-upgrad-proj/dim_atm/part-00000-4ab342ab-048c-49e7-be9b-16786eb21393-c000.csv'
9 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
10 delimiter ',' region 'us-east-1'
11 CSV;
12
```

Run Save Schedule Clear

Query results | Table details

Query 274

Execution Data Visualize

Completed, started on December 17, 2022 at 11:04:29
ELAPSED TIME: 00 m 04 s

- Load data into DIM_DATE table

```
copy atm_info.DIM_DATE
from 's3://atm-upgrad-proj/dim_date/part-00000-f78dec33-e586-4af0-b409-f191a27758d7-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV timeformat 'auto';
```

Query 4 X | Query 5 X | Query 1 X | Query 2 X | Query 3 X | +

```
6
7 copy atm_info.DIM_ATM
8 from 's3://atm-upgrad-proj/dim_atm/part-00000-4ab342ab-048c-49e7-be9b-16786eb21393-c000.csv'
9 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
10 delimiter ',' region 'us-east-1'
11 CSV;
12
13 copy atm_info.DIM_DATE
14 from 's3://atm-upgrad-proj/dim_date/part-00000-fc9d30f4-f60f-45d8-9965-cd0038d8b543-c000.csv'
15 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
16 delimiter ',' region 'us-east-1'
17 CSV timeformat 'auto';
18
```

Run Save Schedule Clear

Query results | Table details

Query **297**

Execution Data Visualize

Completed, started on December 17, 2022 at 11:05:09
ELAPSED TIME: 00 m 05 s

- Load data into DIM_CARD_TYPE table

```
copy atm_info.DIM_CARD_TYPE
from 's3://atm-upgrad-proj/dim_card_type/part-00000-adf96d01-6ff5-478a-afe4-1465f81ba096-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```

- Query 4

Query 5

Query 1

Query 2

Query 3

+

↶

↷

@

☰

/*

📄

✕

```

10 delimiter ',' region 'us-east-1'
11 CSV;
12
13 copy atm_info.DIM_DATE
14 from 's3://atm-upgrad-proj/dim_date/part-00000-fc9d30f4-f60f-45d8-9965-cd0038d8b543-c000.csv'
15 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
16 delimiter ',' region 'us-east-1'
17 CSV timeformat 'auto';
18
19 copy atm_info.DIM_CARD_TYPE
20 from 's3://atm-upgrad-proj/dim_card_type/part-00000-3009215b-dbf4-490f-99b8-af0ddbe0e067-c000.csv'
21 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
22 delimiter ',' region 'us-east-1'
23 CSV;
24

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 332

Completed, started on December 17, 2022 at 11:05:42

ELAPSED TIME: 00 m 02 s

Execution

Data

Visualize

Query 4 X

Query 5 X

Query 1 X

Query 2 X

Query 3 X

+

▼

↶

↷

@

≡

/ *

```
10 delimiter ',' region 'us-east-1'
17 CSV timeformat 'auto';
18
19 copy atm_info.DIM_CARD_TYPE
20 from 's3://atm-upgrad-proj/dim_card_type/part-00000-3009215b-dbf4-490f-99b8-af0ddbe0e067-c000.csv'
21 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
22 delimiter ',' region 'us-east-1'
23 CSV;
24
25 copy atm_info.FACT_ATM_TRANS
26 from 's3://atm-upgrad-proj/fact_atm_trans/part-00000-f404c663-6e96-4cb6-a254-42e2732102dc-c000.csv'
27 iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
28 delimiter ',' region 'us-east-1'
29 CSV;
30
```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 358

Execution

Data

Visualize

Completed, started on December 17, 2022 at 11:06:29

ELAPSED TIME: 00 m 14 s