Creation of a Redshift Cluster

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

- Setting up IAM role

IAM > Roles > Create role

Step 1
Select trusted entity

Step 2
Add permissions

Step 3
Name, review, and create

Select trusted entity Info

Trusted entity type

- ● **AWS service**
  Allow AWS services like EC2, Lambda, or others to perform actions in this account.

- ○ **AWS account**
  Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.

- ○ **Web identity**
  Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.

- ○ **SAML 2.0 federation**
  Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.

- ○ **Custom trust policy**
  Create a custom trust policy to enable others to perform actions in this account.

Use case
Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Common use cases

- ○ EC2
  Allows EC2 instances to call AWS services on your behalf.

- ○ Lambda
  Allows Lambda functions to call AWS services on your behalf.

Use cases for other AWS services:

Redshift ▼

- ● Redshift - Customizable
  Allows Redshift clusters to call AWS services on your behalf.

- ○ Redshift
  Allows Redshift clusters to call AWS services on your behalf.

- ○ Redshift - Scheduler
  Allow Redshift Scheduler to call Redshift on your behalf.

Cancel   **Next**

Add permissions Info

Permissions policies (Selected 1/808) Info

Choose one or more policies to attach to your new role.

Create policy ⧉

🔍 Filter policies by property or policy name and press enter.   1 match   < 1 > ⚙

"amazons3fullaccess" ✕   Clear filters

☑	Policy name ⧉	▽	Type ▽	Description
☑	⊞ 🎁 AmazonS3FullAccess		AWS m...	Provides full access to all buckets via the AWS Management Console.

▶ **Set permissions boundary - _optional_** Info

Set a permissions boundary to control the maximum permissions this role can have. This is not a common setting, but you can use it to delegate permission management to others.

Cancel   Previous   Next

IAM > Roles > Create role

Name, review, and create

Role details

Role name
Enter a meaningful name to identify this role.

    etl-redshift-s3-01

Maximum 64 characters. Use alphanumeric and '+=,.@-_' characters.

Description
Add a short explanation for this role.

    Allows Redshift clusters to call AWS services on your behalf.

Maximum 1000 characters. Use alphanumeric and '+=,.@-_' characters.

Step 1: Select trusted entities   [Edit]

```
1  {
2      "Version": "2012-10-17",
3      "Statement": [
4          {
5              "Effect": "Allow",
6              "Action": [
7                  "sts:AssumeRole"
8              ],
9              "Principal": {
10                 "Service": [
11                     "redshift.amazonaws.com"
12                 ]
13             }
14         }
15     ]
16 }
```

Step 2: Add permissions   [Edit]

Permissions policy summary

Policy name ⬈	Type	Attached as
AmazonS3FullAccess	AWS managed	Permissions policy

Tags

Add tags - *optional*   Info
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.

No tags associated with the resource.

[Add tag]

You can add up to 50 more tags.

Cancel   Previous   Create role

- Redshift Cluster Configuration

Create cluster Info

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

etl-redshift-cluster-1

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

⦿ **Production**
Configure for fast and consistent performance at the best price.

○ **Free trial**
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster
⦿ I'll choose
○ Help me choose

Node type Info
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

dc2.large ▼

Number of nodes
Enter the number of nodes that you need.

2

Range (1-32)

Configuration summary Info
dc2.large | 2 nodes

$360.00/month	320 GB
Estimated on-demand compute price	Total compressed storage
Save more than 60% of your costs by purchasing reserved nodes. Learn more ↗	The total storage capacity for the cluster if you deploy the number of nodes that you chose.

Sample data Info

☐ **Load sample data**
Load sample data to your Redshift cluster to start using the query editor to query data.

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a reserved word ↗.

☐ **Auto generate password**
Amazon Redshift can generate a password for you, or you can specify your own password.

Admin user password
Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except "/", """, or "@".

•••••••••

☐ Show password

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.

| Set default ▼ | Manage IAM roles ▼ |

🔍 Search for associated IAM role by name, status, or role type   ‹ **1** ›

☐	IAM roles 🗗	▽	Status	▽	Role type	▽
☐	etl-redshift-s3-01		Not applied		--	

Additional configurations ⬤ Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

▶ **Network and security** Info

▶ **Database configurations** Info

▶ **Maintenance**

▶ **Monitoring**

▶ **Backup**

Cancel   **Create cluster**

- Created Redshift cluster

etl-redshift-cluster-1

Actions ▼ | Edit | Add partner integration | Query data ▼

General information

Cluster identifier	Status	Node type	Endpoint
etl-redshift-cluster-1	⊘ Available	dc2.large	⎘ etl-redshift-cluster-1.cdakre...
Cluster namespace	Date created	Number of nodes	JDBC URL
6339124b-f4dc-4d70-80f7-9423099eb566	January 23, 2023, 14:29 (UTC+05:30)	2	⎘ jdbc:redshift://etl-redshift-cl...
Cluster configuration	Storage used		ODBC URL
Production	0.02% (0.07 of 320 GB used)		⎘ Driver={Amazon Redshift (x...
	Multi-AZ		
	No		

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:

create schema atm;

```
1  create schema atm;
2
3
4
5
6
7
8
9
10
11
12
13
14
15
```

Run   Save   Schedule   Clear

Query results | Table details

Query

✓ Completed, started on January 24, 2023 at 11:58:30
ELAPSED TIME: 00 m 56 s

```
create table atm.dim_loc
(
  location_id          int primary key,
  location             varchar(30),
  streetname           varchar(30),
  street_number        int,
  zipcode int,
  lat                  decimal(10,3),
  lon                  decimal(10,3)
);
```

Select database Info
To view schemas, select a database.

| dev ▼ |

Select schema Info
To view tables, select a schema.

| atm ▼ |

| Q Filter tables |

< 1 >

▶ dim_loc_pkey   •••

▶ dim_loc   •••

⊘ creation-i... ✕ | ⊘ dql-tables ✕ | +

```
1   create schema atm;
2
3   create table atm.dim_loc
4   (
5     location_id int primary key,
6     location varchar(30),
7     streetname varchar(30),
8     street_number int,
9     zipcode int,
10    lat decimal(10,3),
11    lon decimal(10,3)
12  );
13
14
15
```

[Run] [Save] [Schedule] [Clear]

Query results | Table details

Query

⊘ Completed, started on January 24, 2023 at 11:59:42
ELAPSED TIME: 00 m 16 s

```
create table atm.dim_atm
(
  atm_id               int primary key,
  atm_number           varchar(20),
  atm_manufacturer     varchar(20),
  atm_location_id      int references atm.dim_loc(location_id)
);
```

Select database Info
To view schemas, select a database.

dev ▼

Select schema Info
To view tables, select a schema.

atm ▼

🔍 Filter tables

< 1 >

▶ dim_atm_pkey   •••

▶ dim_loc_pkey   •••

▶ dim_atm   •••

▶ dim_loc   •••

⊘ creation-i... ✕ | ⊘ dql-tables ✕ | +

```
 6    location varchar(30),
 7    streetname varchar(30),
 8    street_number int,
 9    zipcode int,
10    lat decimal(10,3),
11    lon decimal(10,3)
12  );
13
14  create table atm.dim_atm
15  (
16    atm_id int primary key,
17    atm_number varchar(20),
18    atm_manufacturer varchar(20),
19    atm_location_id int references atm.dim_loc(location_id)
20  );
```

| Run | Save | Schedule | Clear |

Query results | **Table details**

Query

⊘ Completed, started on January 24, 2023 at 12:01:09
ELAPSED TIME: 00 m 07 s

```
create table atm.dim_date
(
  date_id              int primary key,
  full_date_time       timestamp,
  "year"               int,
  month                varchar(20),
  day                  int,
  hour                 int,
  weekday              varchar(20)
);
```

Select database Info
To view schemas, select a database.

dev ▼

Select schema Info
To view tables, select a schema.

atm ▼

🔍 Filter tables

‹ 1 ›

▶ dim_atm_pkey   •••

▶ dim_date_pkey   •••

▶ dim_loc_pkey   •••

▶ dim_atm   •••

▶ dim_date   •••

▶ dim_loc   •••

⊘ creation-i... ✕ | ⊘ dql-tables ✕ | +

```
19   atm_location_id int references atm.dim_loc(location_id)
20 );
21
22 create table atm.dim_date
23 (
24   date_id int primary key,
25   full_data_time timestamp,
26   "year" int,
27   month varchar(20),
28   day int,
29   hour int,
30   weekday varchar(20)
31 );
32
33
34
```

Run   Save   Schedule   Clear

Query results | Table details

Query

⊘ Completed, started on January 24, 2023 at 12:06:24
ELAPSED TIME: 00 m 15 s

```
create table atm.dim_card_type
(
  card_type_id        int primary key,
  card_type           varchar(30)
);
```

```
create table atm.fact_atm_tran
(
  tran_id                bigint primary key,
  atm_id                 int references atm.dim_atm(atm_id),
  weather_loc_id         int references atm.dim_loc(location_id),
  date_id                int references atm.dim_date(date_id),
  card_type_id           int references atm.dim_card_type(card_type_id),
  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)
);
```

Select database **Info**
To view schemas, select a database.

dev ▼

Select schema **Info**
To view tables, select a schema.

atm ▼

🔍 Filter tables

⟨ 1 ⟩

▶ dim_atm_pkey ⋯

▶ dim_card_type_pkey ⋯

▶ dim_date_pkey ⋯

▶ dim_loc_pkey ⋯

▶ fact_atm_tran_pkey ⋯

▶ dim_atm ⋯

▶ dim_card_type ⋯

▶ dim_date ⋯

▶ dim_loc ⋯

▶ fact_atm_tran ⋯

⊘ creation-i... ✕ | ⊘ dql-tables ✕ | +

```
36   card_type varchar(30)
37 );
38
39
40 create table atm.fact_atm_tran
41 (
42   tran_id          bigint primary key,
43   atm_id           int references atm.dim_atm(atm_id),
44   weather_loc_id   int references atm.dim_loc(location_id),
45   date_id          int references atm.dim_date(date_id),
46   card_type_id     int references atm.dim_card_type(card_type_id),
47   atm_status       varchar(20),
48   currency         varchar(10),
49   service          varchar(20),
50   transaction_amount int,
51   message_code     varchar(255),
52   message_text     varchar(255),
53   rain_3h          decimal(10,3),
54   clouds_all       int,
55   weather_id       int,
56   weather_main     varchar(50),
57   weather_description varchar(255)
58 );
59 --Card Type
```

[Run] [Save] [Schedule] [Clear]

Query results | Table details

Query

⊘ Completed, started on January 24, 2023 at 12:09:44
ELAPSED TIME: 00 m 08 s

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

```
--Card Type
copy atm.dim_card_type
from 's3://sharads-etl-project/Data/redshift_csv/dim_card_type/part-00000-da182289-b9ce-
47d1-8e96-920498350788-c000.csv'
iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
region 'us-east-1'
CSV;
```

--Location
copy atm.dim_loc
from 's3://sharads-etl-project/Data/redshift_csv/dim_loc/part-00000-ea0bbee5-9124-4aff-95ed-ffc8c7a9d5fc-c000.csv'
iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
region 'us-east-1'
CSV;

```
     │ ⊘ creation-i...  ✕  │  ⊘ dql-tables  ✕  │  +
     │
61   from 's3://sharads-etl-project/Data/redshift_csv/dim_card_type/part-00000-da182285-b9cc-47d1-8c90-929490090700-c000.
62   iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
63   region 'us-east-1'
64   CSV;
65
66
67
68   --Location
69   copy atm.dim_loc
70   from 's3://sharads-etl-project/Data/redshift_csv/dim_loc/part-00000-ea0bbee5-9124-4aff-95ed-ffc8c7a9d5fc-c000.csv'
71   iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
72   region 'us-east-1'
73   CSV;
74
75
76
```

Run **Save** **Schedule** **Clear**

Query results | **Table details**

Query 964 ☑

⊘ Completed, started on January 24, 2023 at 12:18:17
ELAPSED TIME: 00 m 03 s

--ATM
copy atm.dim_atm
from 's3://sharads-etl-project/Data/redshift_csv/dim_atm/part-00000-fae2a521-7e94-4db8-bf64-6f1478f17e06-c000.csv'
iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
region 'us-east-1'
CSV;

--Date
copy atm.dim_date
from 's3://sharads-etl-project/Data/redshift_csv/dim_date/part-00000-8e1c5caf-00d4-4f10-9c27-04d48769fb1f-c000.csv'
iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
region 'us-east-1'
CSV;

Run Save Schedule Clear

Query results | Table details

Query 1066 ↗

✓ Completed, started on January 24, 2023 at 12:20:53
ELAPSED TIME: 00 m 09 s

--ATM Transaction
copy atm.fact_atm_tran from 's3://sharads-etl-project/Data/redshift_csv/fact_atm_tran1/part-00000-77c6b699-61fb-4ea8-8bc7-d3ae29187fd2-c000.csv'
iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
region 'us-east-1'
CSV;

```
     creation-i... ✕    dql-tables ✕    +

87  copy atm.dim_date
88  from 's3://sharads-etl-project/Data/redshift_csv/dim_date/part-00000-8e1c5caf-00d4-4f10-9c27-04d48769fb1f-c000.csv'
89  iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
90  region 'us-east-1'
91  CSV;
92
93
94
95  --ATM Transaction
96  copy atm.fact_atm_tran from 's3://sharads-etl-project/Data/redshift_csv/fact_atm_tran1/part-00000-77c6b699-61fb-4ea8-8bc7-d3ae29187fd2-c000.csv'
97  iam_role 'arn:aws:iam::732185110416:role/etl-redshift-s3-01'
98  region 'us-east-1'
99  CSV;
```

Run   Save   Schedule   Clear

Query results   Table details

Query 1103 ☑

✓ Completed, started on January 24, 2023 at 12:22:13
ELAPSED TIME: 00 m 20 s