

Creation of a Redshift Cluster

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

Amazon Redshift > Clusters > Create cluster

Create cluster [info](#)

Cluster configuration

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

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

Node type [info](#)

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

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

Create cluster subnet group

Cluster subnet group details

Name
You can't modify the name after your subnet group has been created.

The name must be 1-255 characters. Valid characters are A-Z, a-z, 0-9, space, hyphen (-), underscore (_), and period (.).

Description

Add subnets

VPC
Choose the VPC that contains the subnets that you want to include in your cluster subnet group.

Availability Zone

Subnet

Subnets in this cluster subnet group: (1)

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

Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

aditya_vpc

vpc-0f4a451185c3eb00

 You can't change the VPC associated with this cluster after the cluster has been created. [Learn more](#) 

VPC security groups

This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.

Choose one or more security groups

default

sg-046d549596871b5c7d

Cluster subnet group

Choose the Amazon Redshift subnet group to launch the cluster in.

etl-project-cluster-subnet-group-1

Availability Zone

Specify the Availability Zone that you want the cluster to be created in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

No preference

Enhanced VPC routing

Enabling this option forces network traffic between your cluster and data repositories through a VPC, instead of the

▼ Database configurations

Database name

Specify a database name to create an additional database.

etl

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a **reserved word**.

Database port

Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.

2701

The port must be numeric (1150-65535).

Parameter groups

Defines database parameter and query queues for all the databases.

Encryption

Encrypt all data on your cluster.

☒ Disabled

☐ Use AWS Key Management Service (AWS KMS)

☐ Use a hardware security module (HSM)

Database configurations

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

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

☐ **Show 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 "@".

Associate IAM roles

IAM roles (2)
Choose from existing IAM roles. You can associate up to 50 IAM roles with this cluster.

 < 1 >

- ☐ IAM roles
- ☐ AWSServiceRoleForRedshift
- ☐ ETL_project

Cancel Associate IAM roles

Screenshots of the configuration of the IAM Role:

Create role

1234

▼ Attach permissions policies

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

Create policy

Filter policies Showing 9 results

	Policy name ▼	Used as
<input type="checkbox"/>	AmazonDMSRedshiftS3Role	None
<input checked="" type="checkbox"/>	AmazonS3FullAccess	Permissions policy (1)
<input type="checkbox"/>	AmazonS3ObjectLambdaExecutionRolePolicy	None
<input type="checkbox"/>	AmazonS3OutpostsFullAccess	None
<input type="checkbox"/>	AmazonS3OutpostsReadOnlyAccess	None
<input type="checkbox"/>	AmazonS3ReadOnlyAccess	None
<input type="checkbox"/>	fsxRecordToS3	None
<input type="checkbox"/>	QuickSightAccessForS3StorageManagementAnalyticsReadOnly	None

► Set permissions boundary

Review

Provide the required information below and review this role before you create it.

Role name*

ETL_project

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

Role description

Allows Redshift clusters to access S3

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

Trusted entities

AWS service: redshift.amazonaws.com

Policies

AmazonS3FullAccess

Permissions boundary

Permissions boundary is not set

No tags were added.

Final Configuration

etl-project-redshift-cluster-1

Actions

Edit

Add partner integration

Query data

General information

Cluster identifier

etl-project-redshift-cluster-1

Cluster namespace

8ea053f7-f3e3-4cbd-9028-bae161987137

Status

Available

Date created

January 24, 2022, 20:57 (UTC+05:30)

Storage used

0.07% (0.24 of 320 GB used)

Node type

dc2.large

Number of nodes

2

AQUA

Not available

Endpoint

etl-project-redshift-cluster...

JDBC URL

jdbc:redshift://etl-project-r...

ODBC URL

Driver=(Amazon Redshift (...)

Database configurations

Change admin user password

Rotate encryption keys

Edit

Database name

etl

Parameter group

Defines database parameter and query queries for all the databases.

default:redshift-1.0

Encryption

Disabled

Audit logging

Disabled

Port

2701

Admin user name

awsuser

SSH ingestion setting (cluster public key)

ssh-rsa AAAAB3NzaC1yc3EAAA...

Network and security settings

Edit

Virtual private cloud (VPC)

vpc-0f8a461185c3ab000

Availability Zone

us-east-1f

VPC security group

Specify which instances and devices can connect to the cluster

sg-046d349d3697b5c7d

Publicly accessible

Allow instances and devices outside the VPC to connect to your database through the cluster endpoint.

Disabled

Subnet

etl-project-cluster-subnet-group-1

Enhanced VPC routing

Disabled

Endpoint URL

-

Associated IAM roles (1)

Info

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

IAM roles

Status

Role type

ETL_project

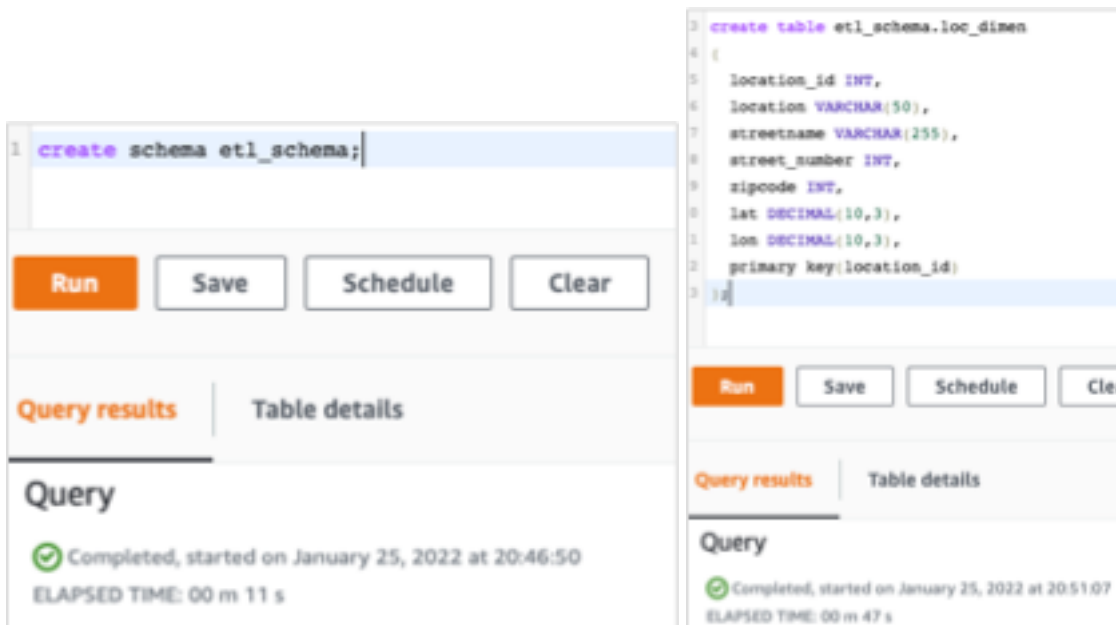
in-sync

2. 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:

Query to create a schema:

create schema etl_schema;



Query to create a location dimension table:

create table etl_schema.loc_dimen (

location_id INT,
location VARCHAR(50), streetname VARCHAR(255),
street_number INT,
zipcode INT,
lat DECIMAL(10,3),
lon DECIMAL(10,3),
primary key(location_id)

);

```
create table etl_schema.date_dimen
(
    date_id INT,
    full_date_time TIMESTAMP,
    year INT,
    month VARCHAR(20),
    day INT,
    hour INT,
    weekday VARCHAR(20),
    primary key(date_id)
);
```

Run Save Schedule Clear

Query results Table details

Query

Completed, started on January 25, 2022 at 20:52:59
ELAPSED TIME: 00 m 07 s

Query to create a date dimension table:

```
create table etl_schema.date_dimen (
date_id INT,
full_date_time TIMESTAMP, year INT,
month VARCHAR(20),
day INT,
hour INT,
weekday VARCHAR(20), primary key(date_id)
);
```

Query to create a card type dimension table:

```
create table etl_schema.card_type_dimen (
card_type_id INT, card_type VARCHAR(30), primary
key(card_type_id)
);
```

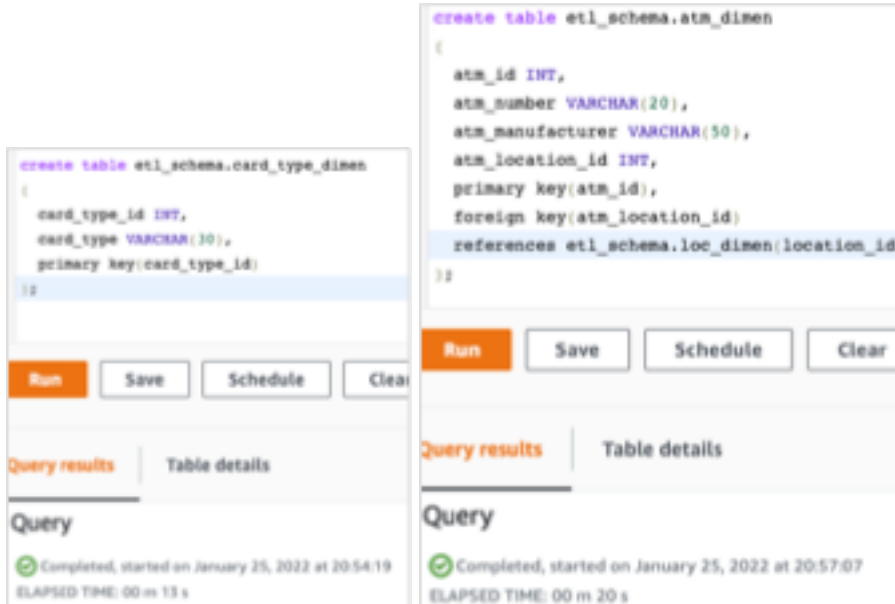
Query to create an atm dimension table:

```
create table etl_schema.atm_dimen (
atm_id INT,
atm_number VARCHAR(20), atm_manufacturer VARCHAR(50),
atm_location_id INT,
```

```

primary key(atm_id),
foreign key(atm_location_id)
references etl_schema.loc_dimen(location_id)
);

```



Query to create a transaction fact table:

```

create table etl_schema.txn_fact
(
trans_id BIGINT,
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
etl_schema.loc_dimen(location_id), foreign key(atm_id)
references

```


etl_schema.atm_dimen(atm_id), foreign key(date_id) references
etl_schema.date_dimen(date_id),
foreign key(card_type_id)
references etl_schema.card_type_dimen(card_type_id)
);

```
create table etl_schema.txn_fact(  
    trans_id BIGINT,  
    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 etl_schema.loc_dimen(location_id),  
    foreign key(atm_id) references etl_schema.atm_dimen(atm_id),  
    foreign key(date_id) references etl_schema.date_dimen(date_id),  
    foreign key(card_type_id) references etl_schema.card_type_dimen(card_type_id));
```

Run Save Schedule Clear

Query results | Table details

Query

Completed, started on January 25, 2022 at 21:04:08
Elapsed Time: 05 m 02 s

3. 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 to load data into loc_dimen:

copy etl_schema.loc_dimen from 's3://etl-project-aditya/loc_dimen/part-00000-15ff3944-74a1-4f43-aa9c-060308042de8-c000.csv'

```
iam_role 'arn:aws:iam::083963476825:role/ETL_project'  
csv IGNOREHEADER 1 region 'us-east-1' ;
```

Query to load data into date_dimen:

```
copy etl_schema.date_dimen from 's3://etl-project-aditya/  
date_dimen/part-00000-5b1aa5ff-fa70-4202-87bc-35542c25a534-  
c000.csv'  
iam_role 'arn:aws:iam::083963476825:role/ETL_project'  
csv IGNOREHEADER 1 timeformat 'YYYY-MM-  
DDTHH:MI:SS' region 'us-east-1' ;
```



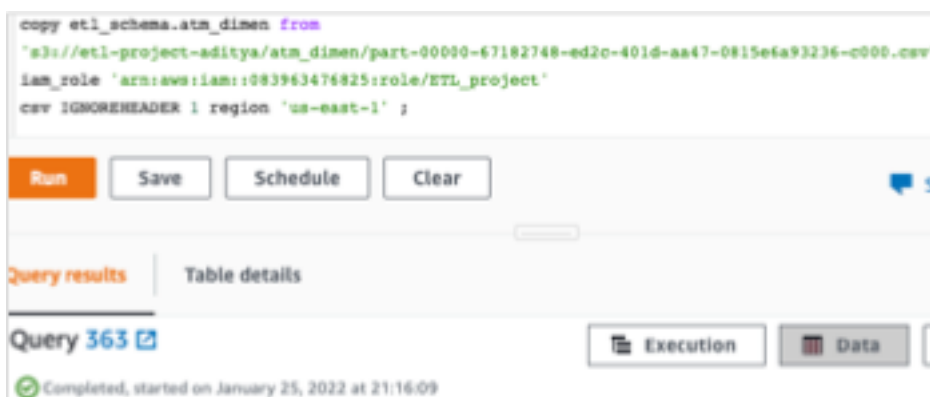
Query to load data into card_type_dimen:

```
copy etl_schema.card_type_dimen from 's3://etl-project-aditya/  
card_type_dimen/part-00000-e94963a7-5dbd-42f2-  
a1d3-6f847c0c41ff-c000.csv'  
iam_role 'arn:aws:iam::083963476825:role/ETL_project'  
csv IGNOREHEADER 1 region 'us-east-1' ;
```



Query to load data into atm_dimen:

```
copy etl_schema.atm_dimen from 's3://etl-project-aditya/
atm_dimen/part-00000-67182748-ed2c-401d- aa47-0815e6a93236-
c000.csv'
iam_role 'arn:aws:iam::083963476825:role/ETL_project'
csv IGNOREHEADER 1 region 'us-east-1' ;
```



Query to load data into txn_fact:

```
copy etl_schema.txn_fact from 's3://etl-project-aditya/txn_fact/
part-00000-2cb996f3-2589-4711-82b9-91299ee8ff19- c000.csv'
iam_role 'arn:aws:iam::083963476825:role/ETL_project'
csv IGNOREHEADER 1 region 'us-east-1' ;
```

4. Query to check number of record in each table:

```
select count(*) from etl_schema.txn_fact;
select count(*) from etl_schema.loc_dimen;
select count(*) from etl_schema.date_dimen; select count(*) from
etl_schema.atm_dimen; select count(*) from
etl_schema.card_type_dimen;
```

```
copy etl_schema.txn_fact from
's3://etl-project-aditya/txn_fact/part-00000-2cb996f3-2589-4711-82b9-91299ee8ff19-c000.csv'
iam_role 'arn:aws:iam::083963476825:role/ETL_project'
csv IGNOREHEADER 1 region 'us-east-1' ;
```

Run

Save

Schedule

Clear



Query results

Table details

Query

Execution

Data



Completed, started on January 25, 2022 at 21:07:42

ELAPSED TIME: 00 m 41 s