

## Creation of a Redshift Cluster

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

Amazon Redshift

Redshift serverless New

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2 [↗](#)

Queries and loads

Datashares

Configurations

Workload management

Create cluster [Info](#)

Cluster configuration

Cluster identifier

This is the unique key that identifies a cluster.

redshift-cluster-etl

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

Amazon Redshift

Redshift serverless New

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2 [↗](#)

Queries and loads

Datashares

Configurations

Workload management

Subnet groups

HSM

Number of nodes

Enter the number of nodes that you need.

1

Range (1-32)

Configuration summary [Info](#)

dc2.large | 1 node

\$180.00/month

Estimated on-demand compute price

Save more than 60% of your costs by purchasing reserved nodes.

[Learn more](#) [↗](#)

160 GB

Total compressed storage

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

Amazon Redshift

✕

Redshift serverless New

Provisioned clusters dashboard

▼ Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2 🔗

Queries and loads

Datashares

▼ Configurations

Workload management

Subnet groups

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

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](#)

Amazon Redshift

×

Redshift serverless New

Provisioned clusters dashboard

▼ Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2 🔗

Queries and loads

❗ 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 >

<input type="checkbox"/>	<a href="#">IAM roles</a> 🔗 ▼	Status ▼	Role type ▼
<input type="checkbox"/>	<a href="#">myRedshiftRole</a>	Not applied	--

Amazon Redshift

×

Redshift serverless New

Provisioned clusters dashboard

▼ Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

▼ Configurations

Workload management

Subnet groups

HSM

▼ Network and security Info

Virtual private cloud (VPC)  
This VPC defines the virtual networking environment for this cluster.

Default VPC  
vpc-02be644d479307f56

ⓘ 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-0cc02f38b87db5dba

×

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

cluster-subnet-group-1

Availability Zone  
Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

us-east-1a

Enhanced VPC routing  
Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the internet. [Learn more](#)

☒ Turn off

☐ Turn on

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 etlproject;
```

```
create table etlproject.dim_location  
(  
    location          varchar(50),  
    streetname        varchar(255),  
    street_number     int,  
    zipcode           int,  
    lat               decimal(10,3),  
    lon               decimal(10,3),  
    location_id       int  
);
```

```
create table etlproject.dim_atm  
(  
    atm_number        int,  
    atm_manufacturer  varchar(20),  
    atm_location_id   varchar(50),  
    atm_id            int  
);
```

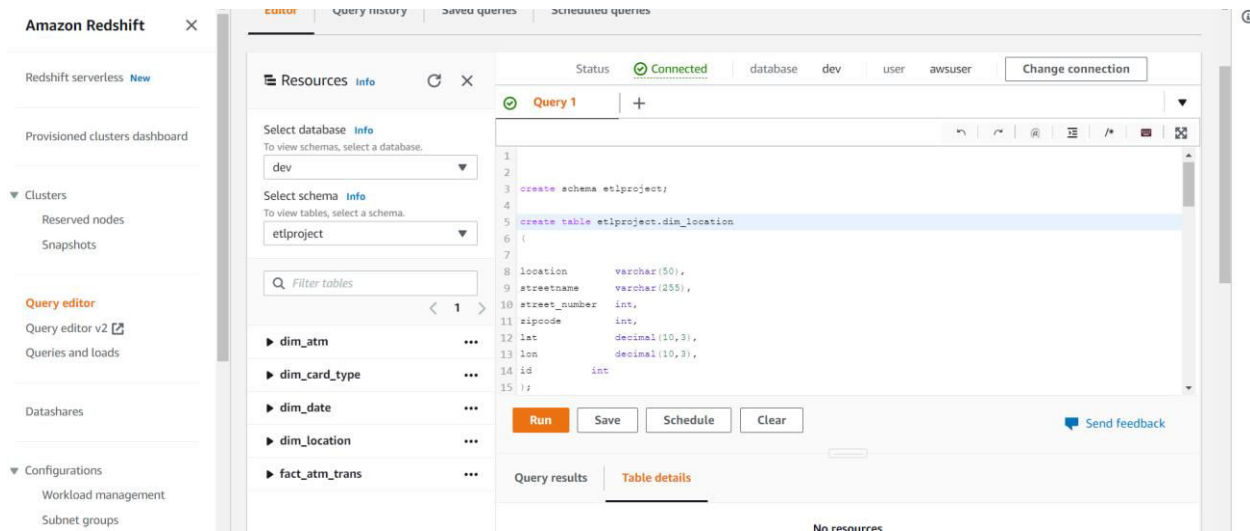
```
create table etlproject.dim_card_type  
(  
    card_type         varchar(30),  
    card_type_id      int  
);
```

```
create table etlproject.dim_date
```

```
(  
    year            int,  
    month           varchar(20) ,  
    day             int,  
    weekday         varchar(20) ,  
    hour            int,  
    full_date_time  varchar(30),  
    date_id         int  
);
```

```
create table etlproject.fact_atm_trans
```

```
(  
  
    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),  
    card_type_id    int,  
    date_id         int,  
    atm_id          int,  
    weather_loc_id  int,  
    trans_id        bigint  
);
```



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

**copy etlproject.dim\_atm from 's3://etlprojectnew/data/dim\_atm/part-00000-e0053268-59dc-47ab-bf23-3cdd69e01f23-c000.csv' iam\_role  
'arn:aws:iam::274560308997:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV  
IGNOREHEADER as 1  
;**

ELAPSED TIME: 00 m 08 s

Rows returned (113) Export ▼

< 1 2 3 4 5 6 7 ... 12 >
 ⚙

atm_number ▼	atm_manufacturer ▼	atm_location ▼	atm_id ▼
1	NCR	74	1
10	NCR	76	2
100	NCR	56	3
101	NCR	17	4
102	NCR	3	5
103	Diebold Nixdorf	103	6
104	NCR	58	7
105	Diebold Nixdorf	76	8
106	NCR	55	9
107	Diebold Nixdorf	62	10

```
copy etlproject.dim_card_type from 's3://etlprojectnew/data/dim_card_type/part-00000-709454aa-f42c-41ec-b8f6-705ff18a208e-c000.csv' iam_role
'arn:aws:iam::274560308997:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV
IGNOREHEADER as 1
;
```

Query 1210

ExecutionData

Completed, started on December 14, 2022 at 18:30:50

ELAPSED TIME: 00 m 02 s

Rows returned (5)

card_type	card_type_id
CIRRUS	1
Dankort	2
Dankort - on-us	3
HÃfÃ¡vekort	4
HÃfÃ¡vekort - on-us	5

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Private



```
copy etlproject.dim_date from 's3://etlprojectnew/data/dim_date/part-00000-20bba22e-
cd33-41d2-9c42-cb90c656e149-c000.csv' iam_role
'arn:aws:iam::274560308997:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV
IGNOREHEADER as 1
;
```

Query results

Table details

Query [1162](#)

Execution

Data

Visualize

Completed, started on December 14, 2022 at 18:27:11

ELAPSED TIME: 00 m 08 s

Rows returned (5)


Export




Search rows


< 1 > ⚙

year	month	day	weekday	hour	full_date_time	date_id
2017	January	1	Sunday	0	2017-01-01T00:00:00.000Z	1
2017	January	1	Sunday	1	2017-01-01T01:00:00.000Z	2
2017	January	1	Sunday	2	2017-01-01T02:00:00.000Z	3
2017	January	1	Sunday	3	2017-01-01T03:00:00.000Z	4
2017	January	1	Sunday	4	2017-01-01T04:00:00.000Z	5


```
copy etlproject.dim_location from 's3://etlprojectnew/data/dim_location/part-00000-26c1b703-2b27-424b-ad58-0dc814f08346-c000.csv' iam_role
'arn:aws:iam::274560308997:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV
IGNOREHEADER as 1
;
```

Query [1278](#) 

 Execution
  Data
  Visualize



 Completed, started on December 14, 2022 at 18:34:43  
 ELAPSED TIME: 00 m 08 s

Rows returned (5) Export ▼

< 1 > 

location ▼	streetname ▼	street_number ▼	zipcode ▼	lat ▼	lon ▼	location_id
Aabybro	ÅfEøestergade	6	9440	57.162	9.730	1
Aalborg Hallen	Europa Plads	4	9000	57.044	9.913	2
Aalborg Storcenter Afd	Hobrovej	452	9200	57.005	9.876	3
Aalborg Storcenter indg. D	Hobrovej	452	9200	57.005	9.876	4
Aalborg Syd	Hobrovej	440	9200	57.005	9.881	5

```
copy etlproject.fact_atm_trans from 's3://etlprojectnew/data/fact_atm_trans/part-00000-c4a91d36-6042-45c5-b144-81cef9fa9ff1-c000.csv' iam_role
'arn:aws:iam::274560308997:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV
IGNOREHEADER as 1
;
```

Query results		Table details			
Query <a href="#">1352</a>		Execution		Data	Visualize
<div>  Completed, started on December 14, 2022 at 18:39:33            ELAPSED TIME: 00 m 08 s         </div>					
Rows returned (5)		Export ▼			
<input type="text" value="Search rows"/>		<div>             &lt; 1 &gt;              </div>			
currency ▼	service ▼	transaction_amount ▼	message_code ▼	message_text ▼	rain_3h ▼
DKK	Withdrawal	8819			0.000
DKK	Withdrawal	3376			0.000
DKK	Withdrawal	7812			0.000
DKK	Withdrawal	7751			0.000
DKK	Withdrawal	8488			0.000