







# RedShift Setup

## Redshift Setup

### ■ Creating a SUBNET GROUP

Cluster subnet groups (1)							   	
<input type="text" value="Search"/>							< 1 > 	
<input type="checkbox"/>	Name		Status	VPC ID	Description	Tags		
<input type="checkbox"/>	cluster-subnet-group-1 1 Subnets		 Complete	vpc-04bd5aaa019c2bc5d	Upgrad Practice Redshift Cluster			

## Creating Cluster

### 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. 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

# RedShift Setup

## Choosing Node type – dc2.large

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)

## Database configuration – username and password

### 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

.....

#### ☐ Show password

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

# RedShift Setup

Attaching 'IAM' Role(redshift\_s3\_fullaccess)

For s3 access to cluster.

Associate IAM roles

IAM roles (1/2)

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

Q Search for IAM role to associate

< 1 >

IAM roles

☐

AWSServiceRoleForRedshift

☒

redshift\_s3\_fullaccess

Cancel

Associate IAM roles

# RedShift Setup




## Network and Security Configuration

### ▼ Network and security

#### Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

my\_vpc  
vpc-04bd5aaa019c2bc5d

 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

cloudera  
sg-0ba3dc75048194a19

#### Cluster subnet group

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

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.

# RedShift Setup

Database configurations – under additional configurations and here, set a database name. The default port is 5439, which is known globally. So, the best practice is to change the port number and set it to any one between 1150 and 65535. After this you can click on the Create Cluster button

## ▼ Database configurations

### Database name

Specify a database name to create an additional database.

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.

The port must be numeric (1150-65535).

### Parameter groups

Defines database parameter and query queues for all the databases.

Default parameter group for redshift-1.0



### Encryption

Encrypt all data on your cluster.

- ☒ Disabled
- ☐ Use AWS Key Management Service (AWS KMS)
- ☐ Use a hardware security module (HSM)

# RedShift Setup

## Schema and Tables Creation on RedShift

create schema ATM;

### – Creating table Location Dimension table

```
create table ATM.DIM_LOCATION(  
location_id VARCHAR(20) distkey sortkey primary key,  
location VARCHAR(50),  
streetname VARCHAR(255),  
street_number INT,  
zipcode INT,  
lat DECIMAL(10,3),  
lon DECIMAL(10,3)  
);
```

### – Loading data from S3 bucket

```
copy ATM.DIM_LOCATION from  
's3://etl-upgrad/DIM_LOCATION/'  
iam_role 'arn:aws:iam::730892717471:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1' IGNOREHEADER 1 removequotes;
```

<input type="text" value="Search rows"/>							< 1 > ⚙
location_id ▾	location ▾	streetname ▾	street_number ▾	zipcode ▾	lat ▾	lon ▾	
location10	Aars	Himmerlandsgade	70	9600	56.803	9.518	
location100	Taars	Bredgade	91	9830	57.385	10.116	
location102	Vadum	Ellehammersvej	43	9430	57.118	9.861	
location106	Vinderup	SÃfÃ, ndergade	5	7830	56.481	8.779	
location11	Abildgaard	HjÃfÃ, rringvej	144	9900	57.447	10.506	
location16	Brugsen i Breum	AakjÃfÃ,rsvej	1	7870	56.688	9.069	
location22	Durup	Torvet	4	7870	56.745	8.949	
location26	Frederiksberg	Gammel Kongevej	157	1850	55.677	12.537	
location3	Aalborg Storcenter Afd	Hobrovej	452	9200	57.005	9.876	
location30	GlyngÃfÃ, re	FÃfÃ,rsvej	1	7870	56.762	8.867	

# RedShift Setup

## – Creating Atm dimension table

```
create table ATM.DIM_ATM(  
  atm_id INT distkey sortkey primary key,  
  atm_number varchar(20),  
  atm_manufacturer VARCHAR(50),  
  atm_location_id VARCHAR(50),  
  foreign key(atm_location_id) references ATM.DIM_LOCATION(location_id)  
);
```

## – Loading data from S3 bucket

```
copy ATM.DIM_ATM from  
's3://etl-upgrad/DIM_ATM/'  
iam_role 'arn:aws:iam::730892717471:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1' IGNOREHEADER 1 removequotes;
```

### dim\_atm

atm.dim\_atm

 Show schema

 Search rows

atm_id	atm_number	atm_manufacturer	atm_location_id
5	5	NCR	location69
15	15	NCR	location104
22	22	NCR	location79
23	23	Diebold Nixdorf	location107
24	24	NCR	location40
27	27	NCR	location34
28	28	NCR	location66
33	33	NCR	location102
35	35	NCR	location1
39	39	NCR	location95

Ac  
Go

# RedShift Setup

## – Creating Card type Dimension table


```
create table ATM.DIM_CARD_TYPE(  
card_type_id VARCHAR(20) distkey sortkey primary key,  
card_type VARCHAR(30)  
);
```

## – Loading data from S3 bucket

```
copy ATM.DIM_CARD_TYPE from  
's3://etl-upgrad/DIM_CARD_TYPE/'  
iam_role 'arn:aws:iam::730892717471:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1' IGNOREHEADER 1 removequotes;
```

### dim\_card\_type

atm.dim\_card\_type

 Search rows	
card_type_id	card_type
card1	CIRRUS
card10	Visa Dankort
card6	Maestro
card7	MasterCard
card11	Visa Dankort - on-us
card4	HÃfÃ'vekort
card5	HÃfÃ'vekort - on-us
card8	Mastercard - on-us
card9	VISA
card12	VisaPlus



# RedShift Setup

## – Creating Date dimension table

```
create table ATM.DIM_DATE(  
date_id VARCHAR(20) distkey sortkey primary key,  
full_date_time TIMESTAMP,  
year INT,  
month VARCHAR(20),  
day INT,  
hour INT,  
weekday VARCHAR(20)  
);
```

## – Loading data from S3 Bucket

```
copy ATM.DIM_DATE from  
's3://etl-upgrad/DIM_DATE/'  
iam_role 'arn:aws:iam::730892717471:role/redshift_s3_fullaccess'  
timeformat 'YYYY-MM-DDTHH:MI:SS'  
delimiter ',' region 'us-east-1' IGNOREHEADER 1 removequotes;
```

date_id	full_date_time	year	month	day	hour	weekday
date1009	2017-08-13 05:00:00	2017	August	13	5	Sunday
date1010	2017-08-13 06:00:00	2017	August	13	6	Sunday
date1013	2017-08-13 09:00:00	2017	August	13	9	Sunday
date1018	2017-08-13 14:00:00	2017	August	13	14	Sunday
date1021	2017-08-13 17:00:00	2017	August	13	17	Sunday
date1028	2017-08-14 00:00:00	2017	August	14	0	Monday
date1029	2017-08-14 01:00:00	2017	August	14	1	Monday
date1031	2017-08-14 03:00:00	2017	August	14	3	Monday
date104	2017-04-05 08:00:00	2017	April	5	8	Wednesday
date1045	2017-08-14 17:00:00	2017	August	14	17	Monday

# RedShift Setup

## – Creating Transaction Fact table

```
create table ATM.FACT_ATM_TRANS(  
trans_id VARCHAR(20) distkey sortkey primary key,  
atm_id INT ,  
weather_loc_id VARCHAR(20) ,  
date_id VARCHAR(20) ,  
card_type_id VARCHAR(20) ,  
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) ,  
foreign key(weather_loc_id) references ATM.DIM_LOCATION(location_id),  
foreign key(atm_id) references ATM.DIM_ATM(atm_id),  
foreign key(date_id) references ATM.DIM_DATE(date_id),  
foreign key(card_type_id) references ATM.DIM_CARD_TYPE(card_type_id)  
);
```

## – Loading data from S3 Bucket

```
copy ATM.FACT_ATM_TRANS from  
's3://etl-upgrad/FACT_ATM_TRANS/'  
iam_role 'arn:aws:iam::730892717471:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1' IGNOREHEADER 1 removequotes;
```

# RedShift Setup

trans_id ▾	atm_id ▾	weather_loc_id ▾	date_id ▾	card_type_id ▾	atm_status ▾	currency ▾	service
transaction100	35	location1		card10	Active	DKK	Withdraw
transaction10000	35	location1	date2447	card11	Active	DKK	Withdraw
transaction1000004	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000005	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000006	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000010	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000012	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000015	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000016	60	location37	date4361	card11	Active	DKK	Withdraw
transaction1000018	60	location37	date4386	card11	Active	DKK	Withdraw

Activate Windows  
Go to Settings to activate Windows.