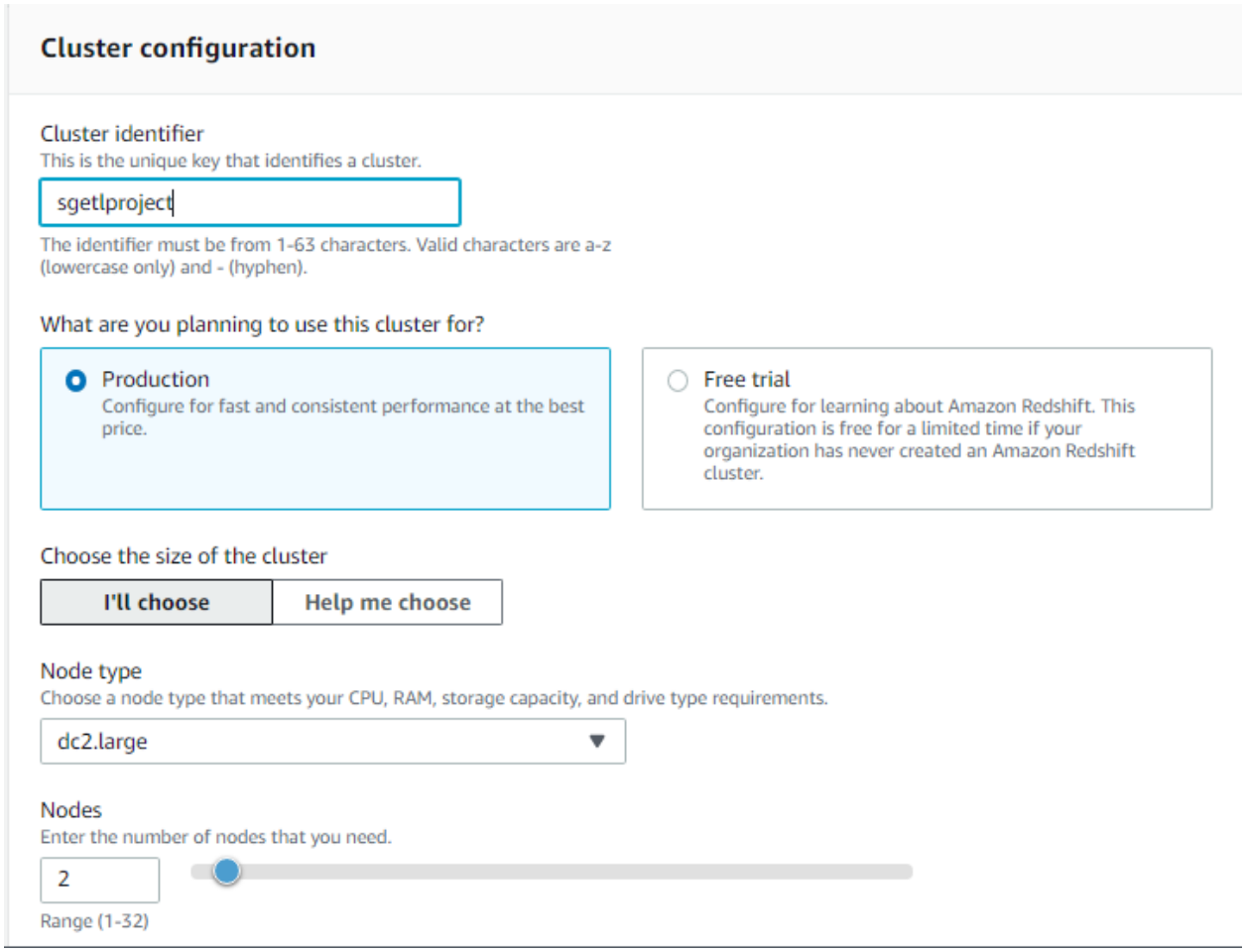
Creation of a RedShift Cluster

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

<Screenshot of the type of machine used along with number of nodes>



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

srivalli_vpc
vpc-0730991434fa98ee8

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-023f55e61179fb3d6

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

sgetlproject

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

us-east-1b

Database configurations

Database name
Specify a database name to create an additional database.

dev

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.

1150

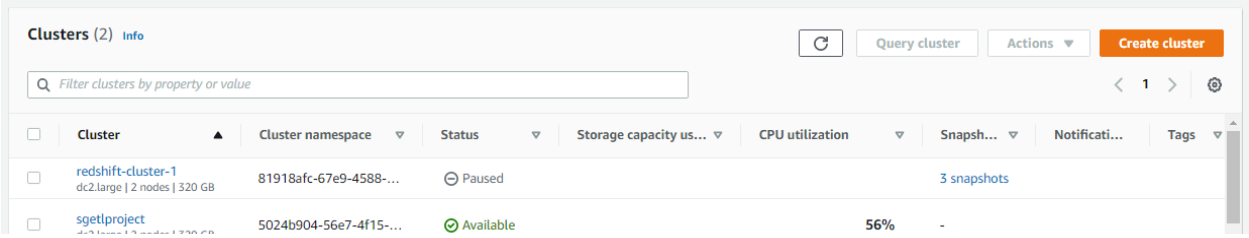
The port must be numeric (1150-65535).

Parameter groups
Defines database parameter and query queues for all the databases.

default.redshift-1.0
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)



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 foriegn keys:

```
create schema etl_project_schema;

create table etl_project_schema.dim_loc (
        atm_location varchar(113),
        atm_streetname varchar(83),
        atm_street_number,
        atm_zipcode numeric,
        atm_lat numeric,
        atm_lon numeric,
        atm_id varchar(113),
        location_id varchar(20),
        PRIMARY KEY (location_id));

create table etl_project_schema.dim_atm (
        atm_lat numeric,
        atm_lon numeric,
        atm_id varchar(113),
        atm_manufacturer varchar(113),
        PRIMARY KEY (atm_id));


create table etl_project_schema.dim_date (
        year int,
        month int,
        day int,
        hour int,
        weekday varchar(10),
        full_date_time varchar(100),
        date_id varchar(20),
```

```
        PRIMARY KEY (date_id));

create table etl_project_schema.dim_card (
        card_type varchar(100),
        card_type_id varchar(20),
    PRIMARY KEY (card_type_id));

create table etl_project_schema.fact_atm_trans (
        trans_id varchar(20),
        atm_prim_id varchar(20),
        location_id varchar(20),
        date_id varchar(20),
        card_type_id varchar(20),
        atm_status varchar(8),
        currency varchar(3),
        service varchar(10),
        transaction_amount int,
        message_code varchar(100),
        message_text varchar(100),
        rain_3h bigint,
        clouds_all int,
        weather_id varchar(20),
        weather_main varchar(15),
        weather_description varchar(32),
    foreign key(atm_prim_id) references etl_project_schema.dim_atm(atm_id),
    foreign key(location_id) references etl_project_schema.dim_loc(location_id),
    foreign key(date_id) references etl_project_schema.dim_date(date_id),
    foreign key(card_type_id) references etl_project_schema.dim_card(card_type_id));
```

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 etl_project_schema.dim_atm from 's3://sgetlproject/ETL/DIM_ATM/DIM_ATM.csv'
iam_role 'arn:aws:iam::893272193681:role/sgredshiftrole'
delimiter ',' region 'us-east-1' ignoreheader 1;

copy etl_project_schema.dim_loc from 's3://sgetlproject/ETL/DIM_LOC/DIM_LOC.csv'
iam_role 'arn:aws:iam::893272193681:role/sgredshiftrole'
delimiter ',' region 'us-east-1' ignoreheader 1;

copy etl_project_schema.dim_date from 's3://sgetlproject/ETL/DIM_DATE/DIM_DATE.csv'
iam_role 'arn:aws:iam::893272193681:role/sgredshiftrole'
```

```
delimiter ',' region 'us-east-1' ignoreheader 1;

copy etl_project_schema.dim_card from 's3://sgetlproject/ETL/DIM_CARD/DIM_CARD.csv'
iam_role 'arn:aws:iam::893272193681:role/sgredshiftrole'
delimiter ',' region 'us-east-1' ignoreheader 1;

copy etl_project_schema.fact_atm_trans from
's3://sgetlproject/ETL/FACT_ATM_TRANS/FACT_ATM_TRANS.csv'
iam_role 'arn:aws:iam::893272193681:role/sgredshiftrole'
delimiter ',' region 'us-east-1' ignoreheader 1
emptyasnull
blanksasnull
removequotes
escape;
```