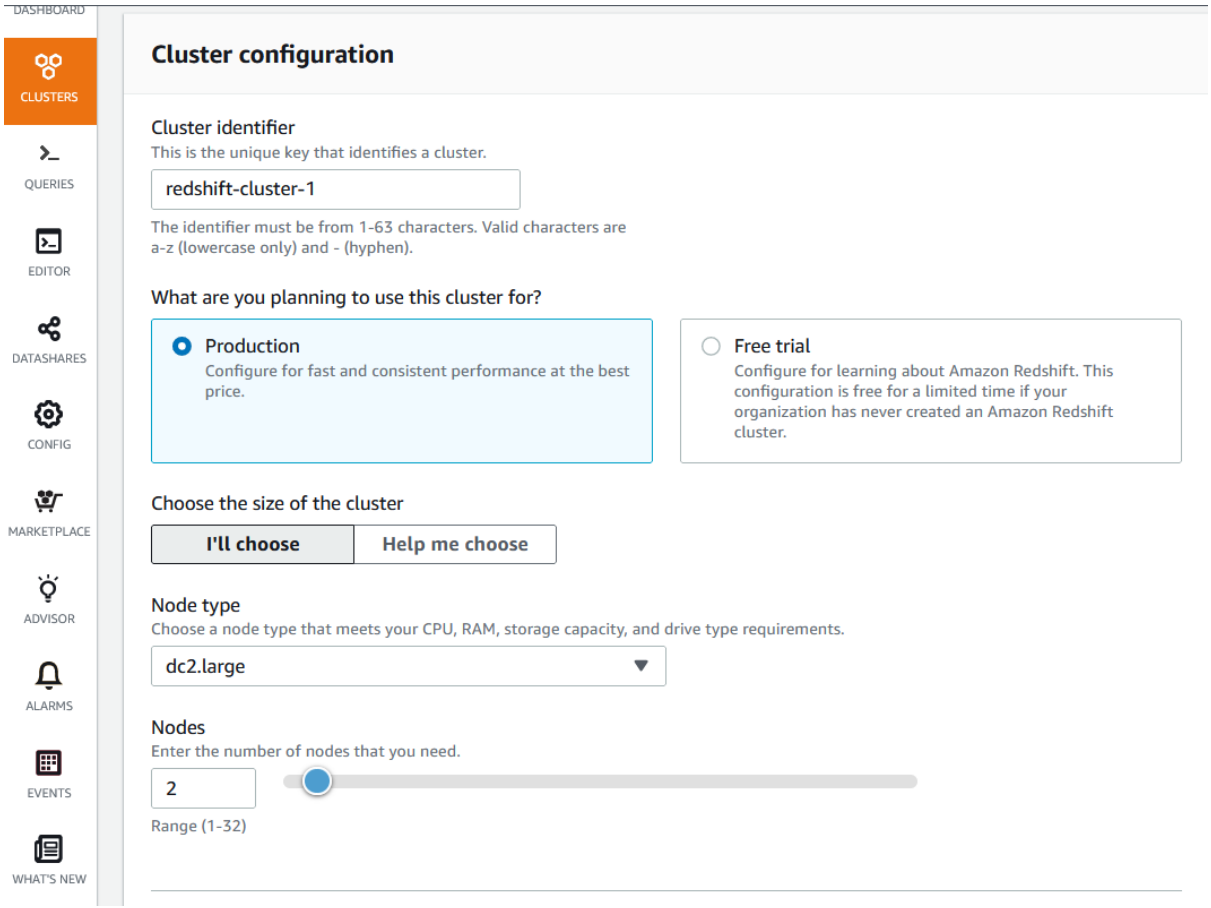
Creation of a RedShift Cluster

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



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 any printable ASCII character except "/", """, or "@".

▼ Cluster permissions

Your cluster needs permissions to access other AWS services on your behalf. For the required permissions, add IAM roles with the principal "redshift.amazonaws.com". You can associate up to 10 IAM roles with this cluster. Learn more

Available IAM roles

Choose an IAM role ▼ | ↻ | Associate IAM role

Associated IAM roles	Status	
Redshift_SGC arn:aws:iam::139471262780:role/Redshift_SGC	Not applied	Remove

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.

Default VPC
vpc-8e850af3 ▼

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-82cf8680

default
sg-82cf8680

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

default

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-1a

Enhanced VPC routing
Enabling this option forces network traffic between your cluster and data repositories through a VPC, instead of the internet. **Learn more**

- (•) Disabled
- ( ) Enabled

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

- (•) Disable
- ( ) Enable

▼ 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.

5439

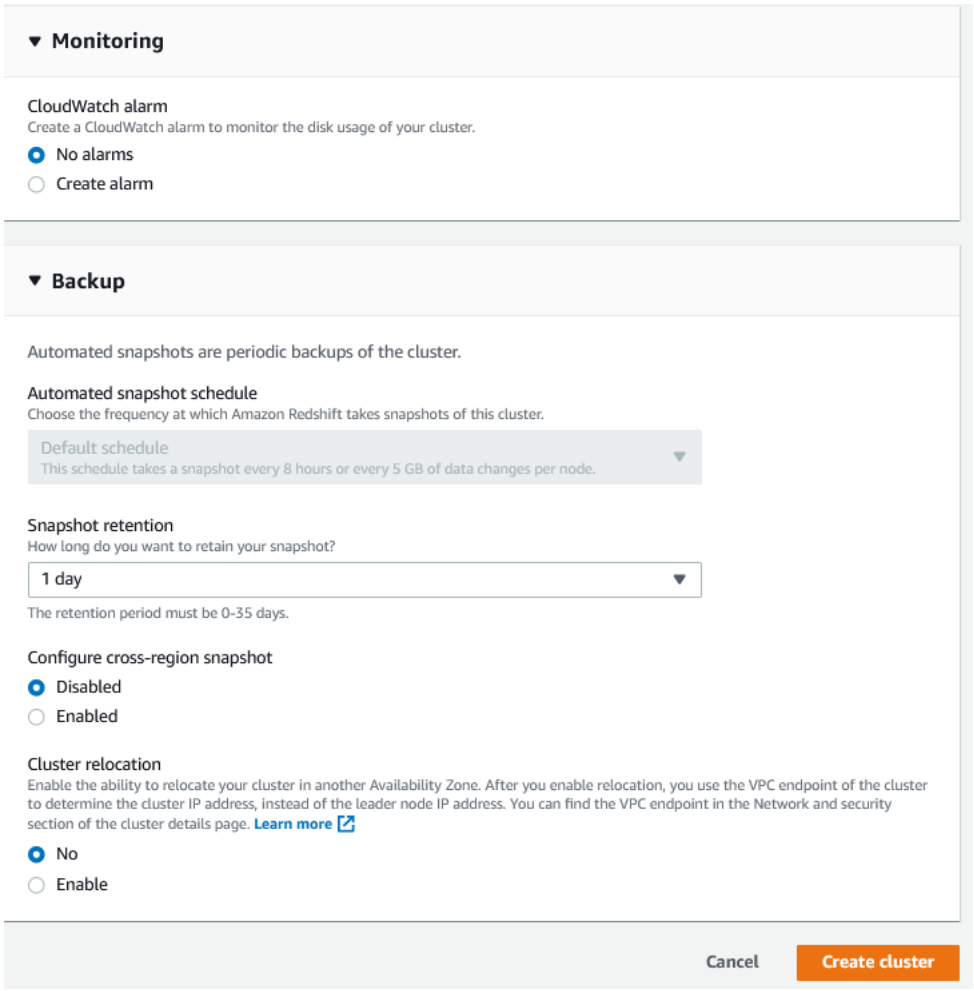
The port must be numeric (1150-65535).

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)


Monitoring
CloudWatch alarm
Create a CloudWatch alarm to monitor the disk usage of your cluster.
No alarms
Create alarm
Backup
Automated snapshots are periodic backups of the cluster.
Automated snapshot schedule
Choose the frequency at which Amazon Redshift takes snapshots of this cluster.
Default schedule
This schedule takes a snapshot every 8 hours or every 5 GB of data changes per node.
Snapshot retention
How long do you want to retain your snapshot?
1 day
The retention period must be 0-35 days.
Configure cross-region snapshot
Disabled
Enabled
Cluster relocation
Enable the ability to relocate your cluster in another Availability Zone. After you enable relocation, you use the VPC endpoint of the cluster to determine the cluster IP address, instead of the leader node IP address. You can find the VPC endpoint in the Network and security section of the cluster details page. Learn more
No
Enable
Cancel
Create cluster

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:

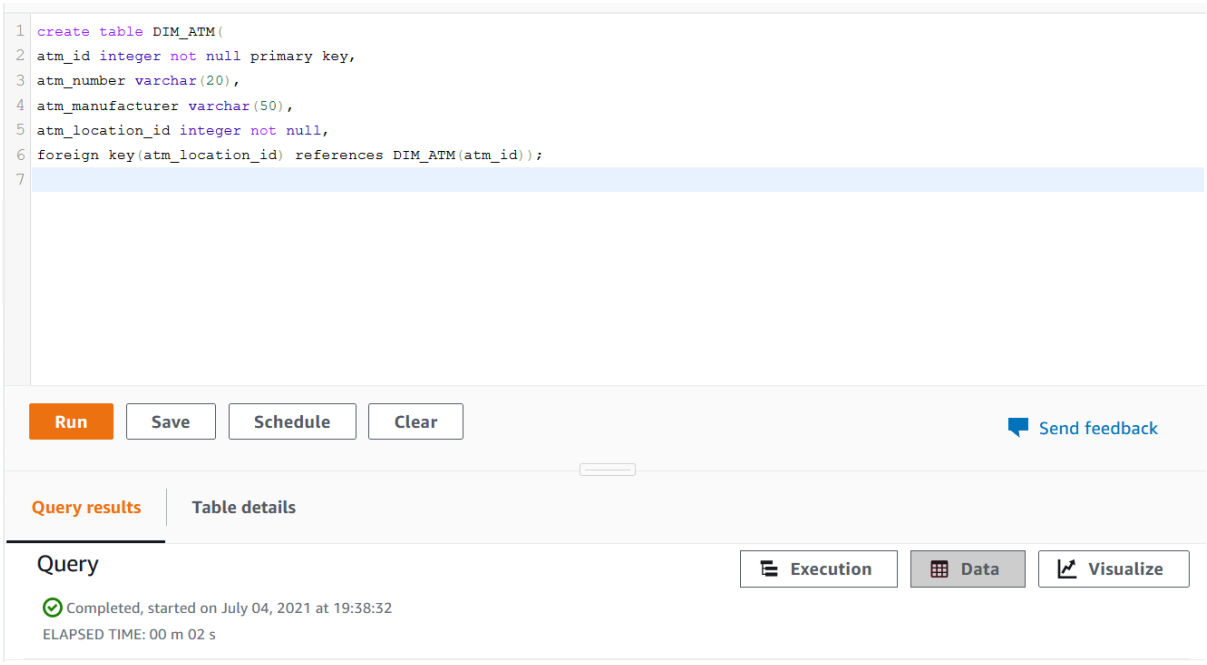
```
create table DIM_LOCATION(
location_id integer not null primary key,
location varchar(50),
streetname varchar(255),
street_number integer,
zipcode integer,
lat decimal(10,3),
lon decimal(10,3));

create table DIM_ATM(
atm_id integer not null primary key,
atm_number varchar(20),
atm_manufacturer varchar(50),
atm_location_id integer not null,
foreign key(atm_location_id) references DIM_ATM(atm_id));

create table DIM_DATE(
date_id integer not null primary key,
full_date_time timestamp,
year integer,
month varchar(20),
day integer,
hour integer,
weekday varchar(20));

create table DIM_CARD_TYPE(
card_type_id integer not null primary key,
card_type varchar(23));
```

```
create table FACT_ATM_TRANS(
trans_id BIGINT not null primary key,
atm_id integer not null,
weather_loc_id integer not null,
date_id integer not null,
card_type_id integer not null,
atm_status VARCHAR(20),
currency VARCHAR(10),
service VARCHAR(20),
transaction_amount integer,
message_code VARCHAR(255),
message_text VARCHAR(255),
rain_3h DECIMAL(10,3),
clouds_all integer,
weather_id integer,
weather_main VARCHAR(50),
weather_description VARCHAR(255),
foreign key(atm_id) references DIM_ATM(atm_id),
foreign key(date_id) references DIM_DATE(date_id),
foreign key(card_type_id) references DIM_CARD_TYPE(card_type_id),
foreign key(weather_loc_id) references DIM_LOCATION(location_id));
```



```
create table DIM_CARD_TYPE(
card_type_id integer not null primary key,
card_type varchar(23));
```

Run | Save | Schedule | Clear | Send feedback

Query results | Table details

Query

Execution | Data | Visualize

Completed, started on July 04, 2021 at 19:40:35

ELAPSED TIME: 00 m 02 s

```
create table DIM_DATE(
date_id integer not null primary key,
full_date_time timestamp,
year integer,
month varchar(20),
day integer,
hour integer,
weekday varchar(20));
```

Run | Save | Schedule | Clear | Send feedback

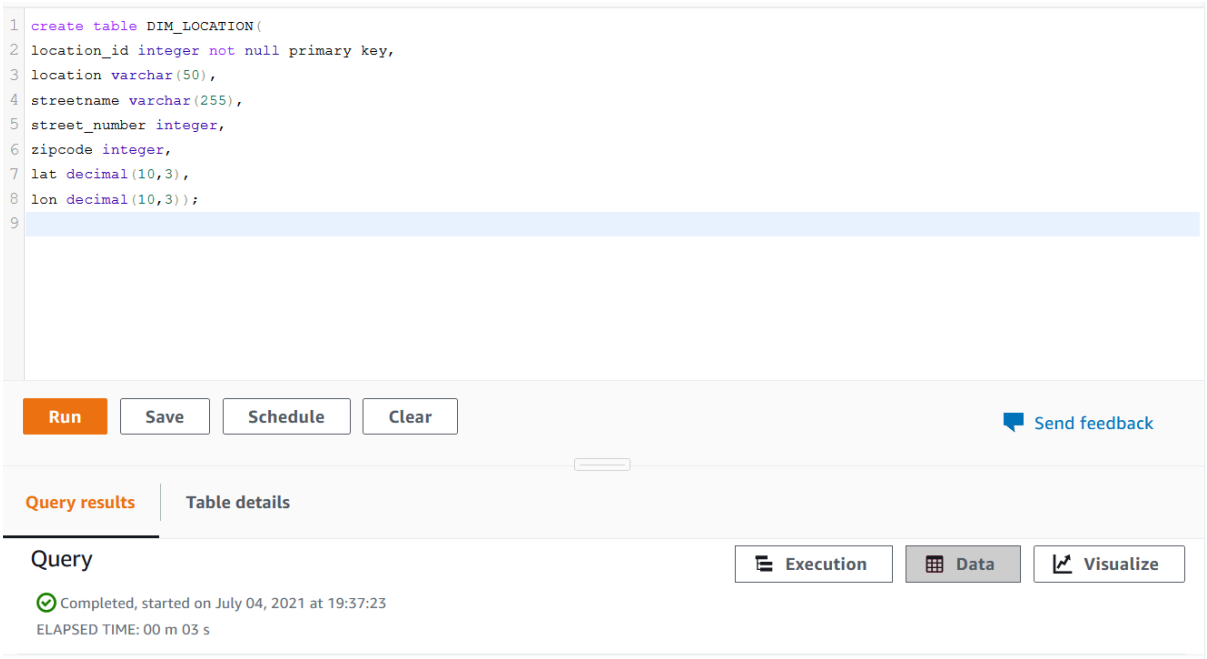
Query results | Table details

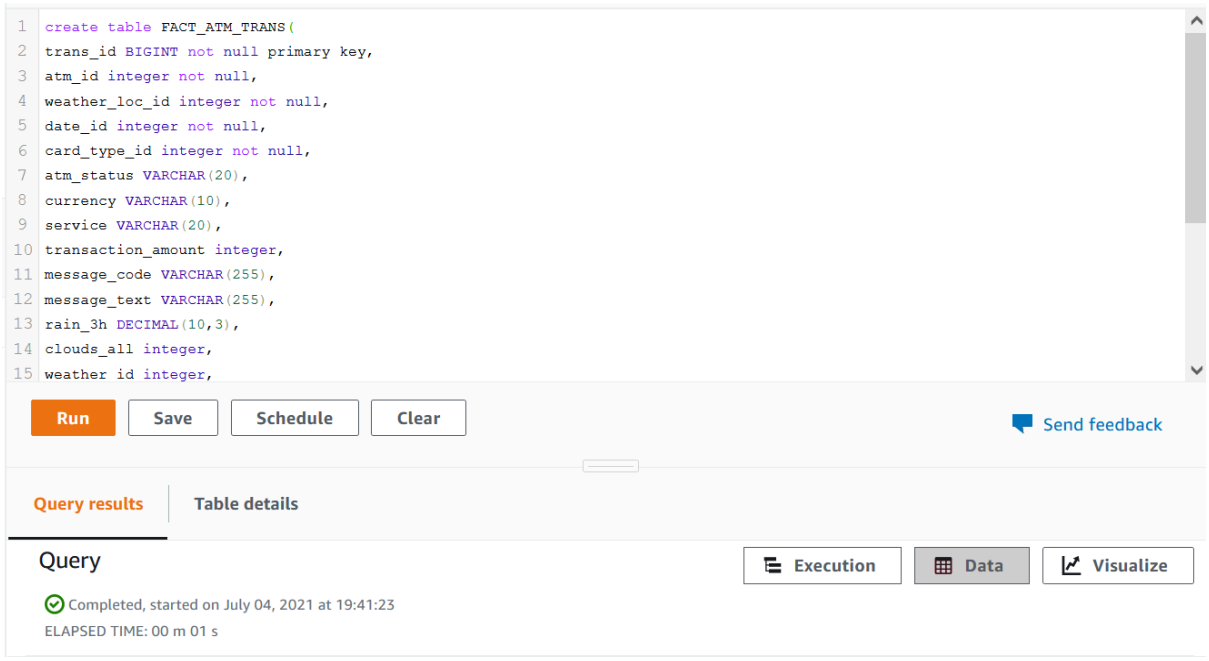
Query

Execution | Data | Visualize

Completed, started on July 04, 2021 at 19:39:43

ELAPSED TIME: 00 m 02 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

copy DIM_LOCATION from 's3://satheeshgopalan/dim_location/part-00000-ecc7cb4b-cb7b-4996-835c-a072b0e76d34-c000.csv'
credentials 'aws_iam_role=arn:aws:iam::139471262780:role/Redshift_SGC'
delimiter ',' region 'us-east-1';

copy DIM_ATM from 's3://satheeshgopalan/dim_atm/part-00000-2e0c55f2-957f-4fb7-bba0-4e10e1117566-c000.csv'
credentials 'aws_iam_role=arn:aws:iam::139471262780:role/Redshift_SGC'
delimiter ',' region 'us-east-1';

copy DIM_DATE from 's3://satheeshgopalan/dim_date/part-00000-0b7e52ee-e32d-4c1a-b4ed-410ff5f4faf0-c000.csv'
credentials 'aws_iam_role=arn:aws:iam::139471262780:role/Redshift_SGC'
delimiter ',' timeformat 'auto' region 'us-east-1';

copy DIM_CARD_TYPE from 's3://satheeshgopalan/dim_card_type/part-00000-4474c06d-e07f-4de5-896a-168a83743af1-c000.csv'
credentials 'aws_iam_role=arn:aws:iam::139471262780:role/Redshift_SGC'
delimiter ',' region 'us-east-1';

copy FACT_ATM_TRANS from 's3://satheeshgopalan/fact_atm_trans/part-00000-0c0ba84b-188b-406b-8e94-0bbfdd84724e-c000.csv'
credentials 'aws_iam_role=arn:aws:iam::139471262780:role/Redshift_SGC'
delimiter ',' region 'us-east-1';