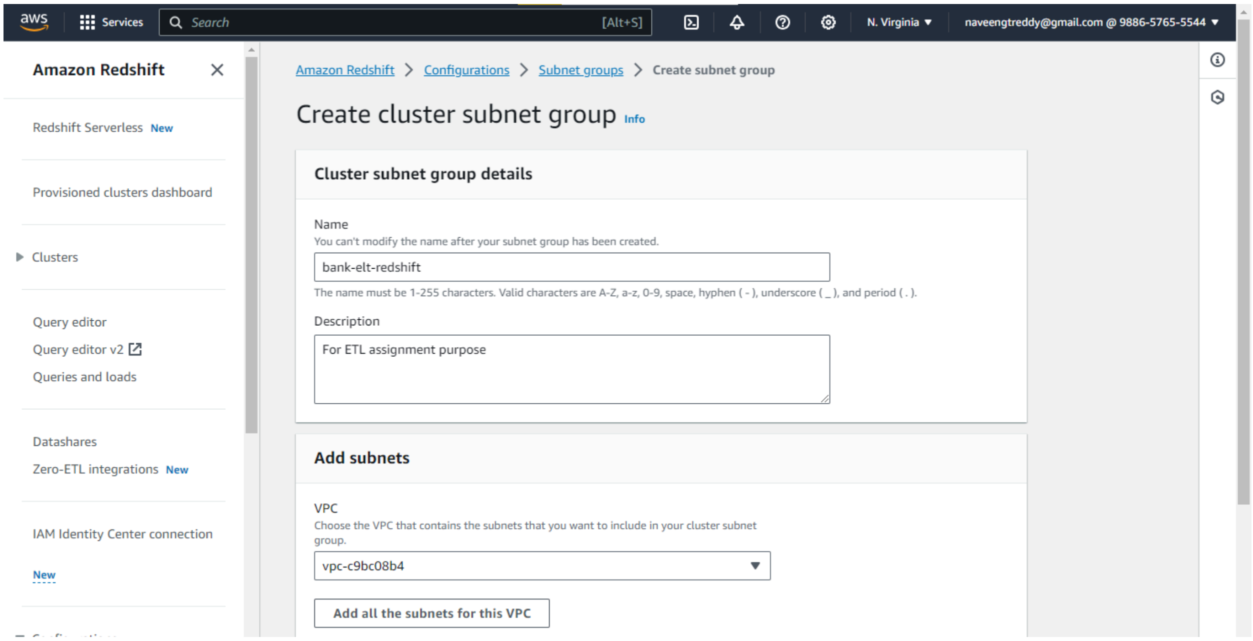
Creation of a Redshift Cluster

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

Create cluster subnet group -


aws
Services
Search
[Alt+S]
N. Virginia ▼
naveengtreddy@gmail.com @ 9886-5765-5544 ▼
Amazon Redshift
Redshift Serverless New
Provisioned clusters dashboard
Clusters
Query editor
Query editor v2
Queries and loads
Datashares
Zero-ETL integrations New
IAM Identity Center connection
New
Amazon Redshift > Configurations > Subnet groups > Create subnet group
Info

Create cluster subnet group

Cluster subnet group details

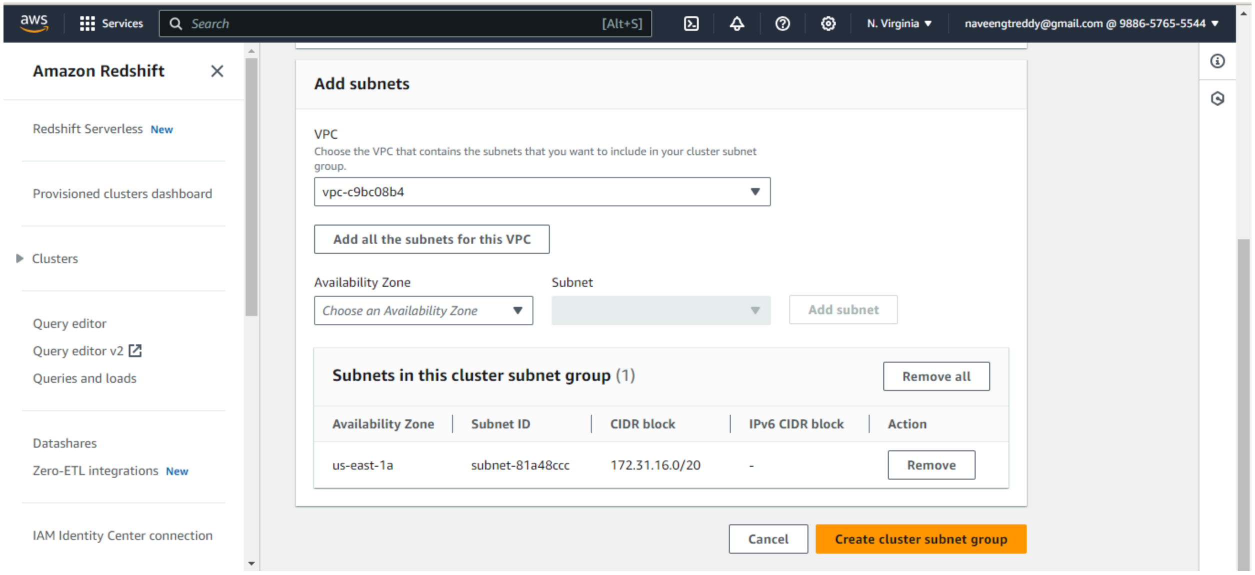
Name
You can't modify the name after your subnet group has been created.
bank-elt-redshift
The name must be 1-255 characters. Valid characters are A-Z, a-z, 0-9, space, hyphen (-), underscore (_), and period (.).

Description
For ETL assignment purpose

Add subnets

VPC
Choose the VPC that contains the subnets that you want to include in your cluster subnet group.
vpc-c9bc08b4

Add all the subnets for this VPC



Add subnets

VPC
Choose the VPC that contains the subnets that you want to include in your cluster subnet group.
vpc-c9bc08b4

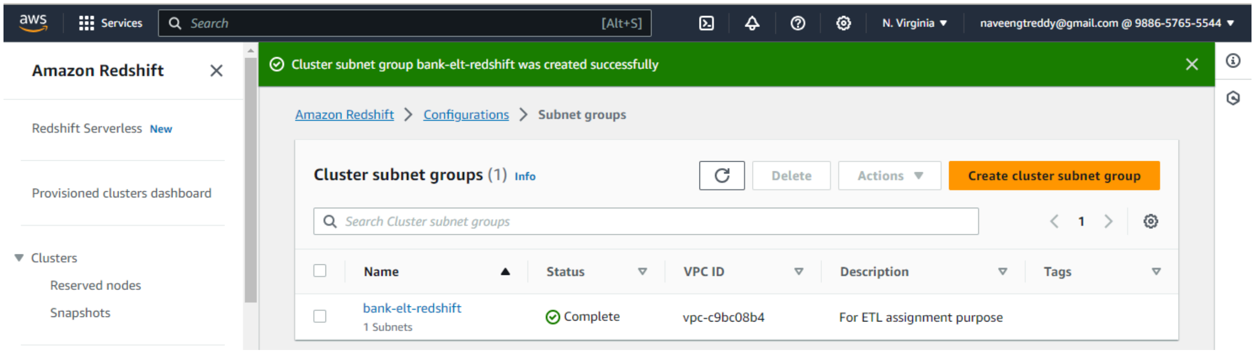
Add all the subnets for this VPC

Availability Zone: Choose an Availability Zone ▼
Subnet: ▼
Add subnet

Subnets in this cluster subnet group (1) — Remove all

Availability Zone	Subnet ID	CIDR block	IPv6 CIDR block	Action
us-east-1a	subnet-81a48ccc	172.31.16.0/20	-	Remove

Cancel Create cluster subnet group


aws
Services
Search [Alt+S]
N. Virginia
naveengtreddy@gmail.com @ 9886-5765-5544
Amazon Redshift
Redshift Serverless New
Provisioned clusters dashboard
Clusters
Reserved nodes
Snapshots
Amazon Redshift > Configurations > Subnet groups
Delete
Actions
Create cluster subnet group
Search Cluster subnet groups
1

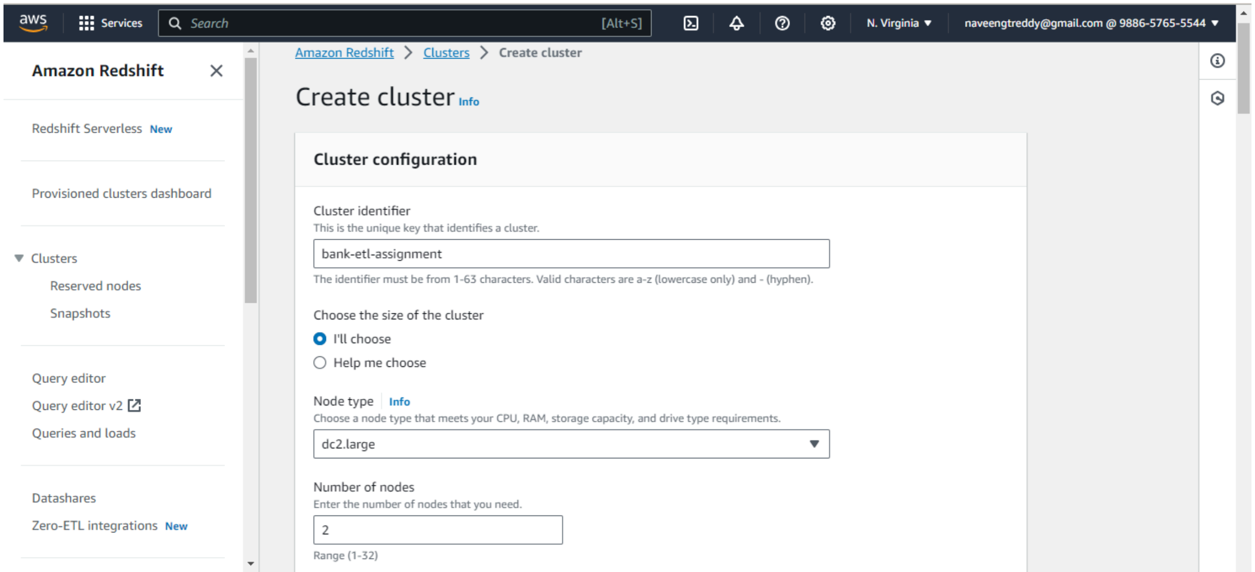
Cluster subnet group bank-elt-redshift was created successfully

Cluster subnet groups (1) Info

Name	Status	VPC ID	Description	Tags
bank-elt-redshift 1 Subnets	✓ Complete	vpc-c9bc08b4	For ETL assignment purpose	

Creating redshift cluster

Node type and number of nodes -



Create cluster Info

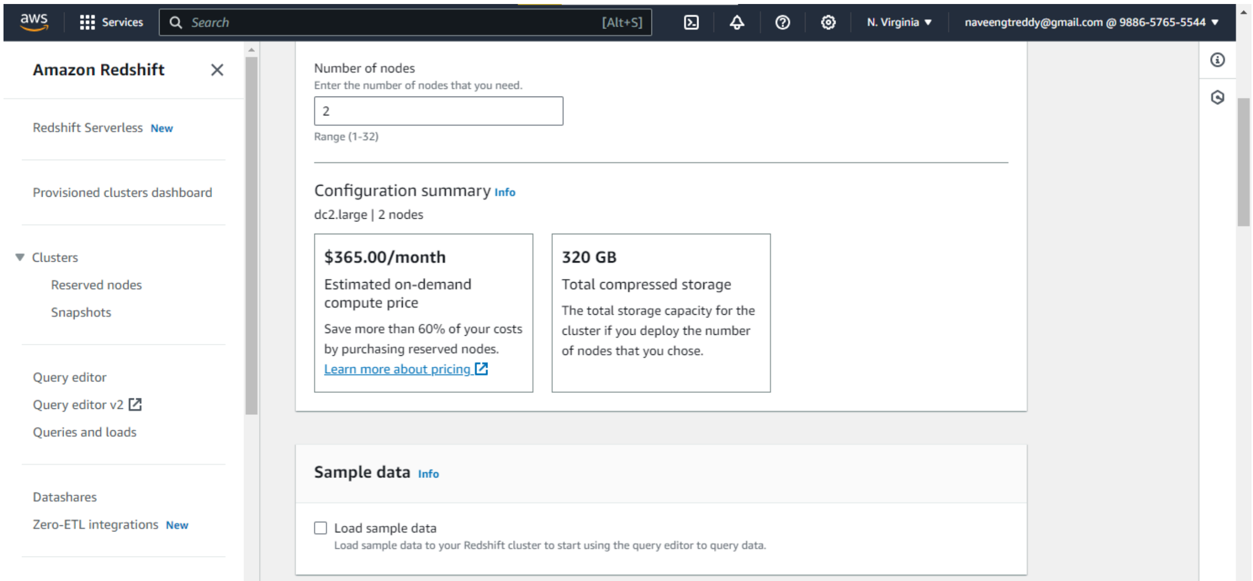
Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.
bank-etl-assignment
The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

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

Number of nodes
Enter the number of nodes that you need.
2
Range (1-32)



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

2

Range (1-32)

Configuration summary Info

dc2.large | 2 nodes

$365.00/month

Estimated on-demand compute price

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

Learn more about pricing

320 GB

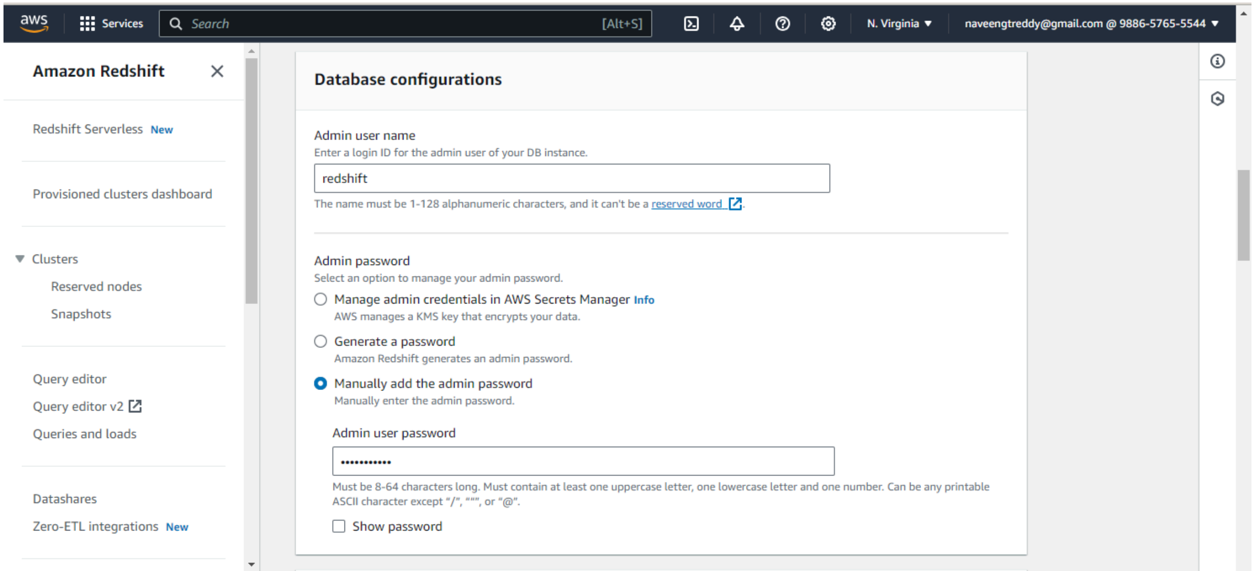
Total compressed storage

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.

Configuring database -



Database configurations

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

redshift

The name must be 1-128 alphanumeric characters, and it can't be a reserved word .

Admin password
Select an option to manage your admin password.

○ Manage admin credentials in AWS Secrets Manager Info
AWS manages a KMS key that encrypts your data.

○ Generate a password
Amazon Redshift generates an admin password.

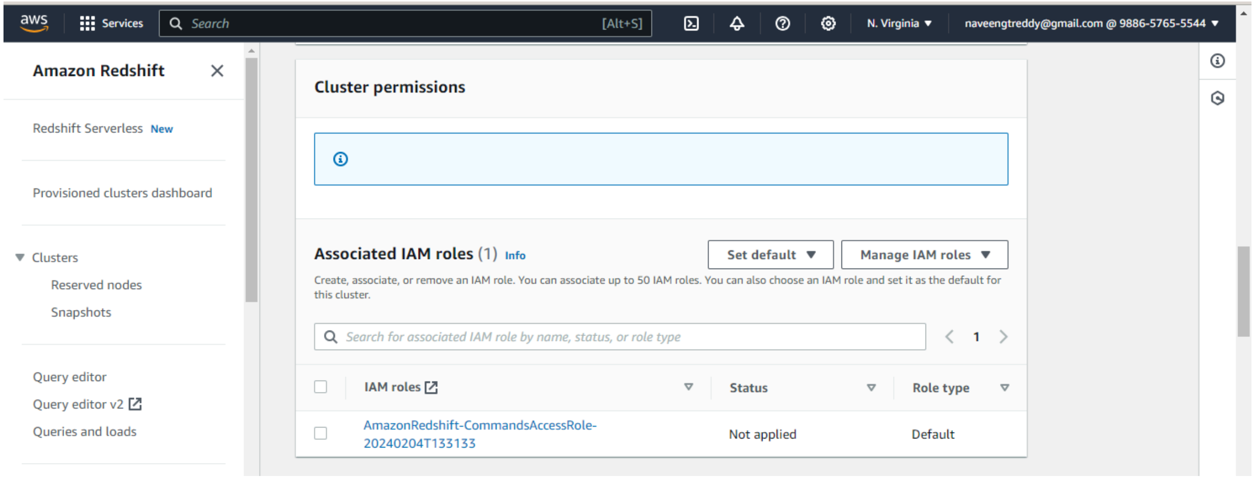
● Manually add the admin password
Manually enter the admin password.

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

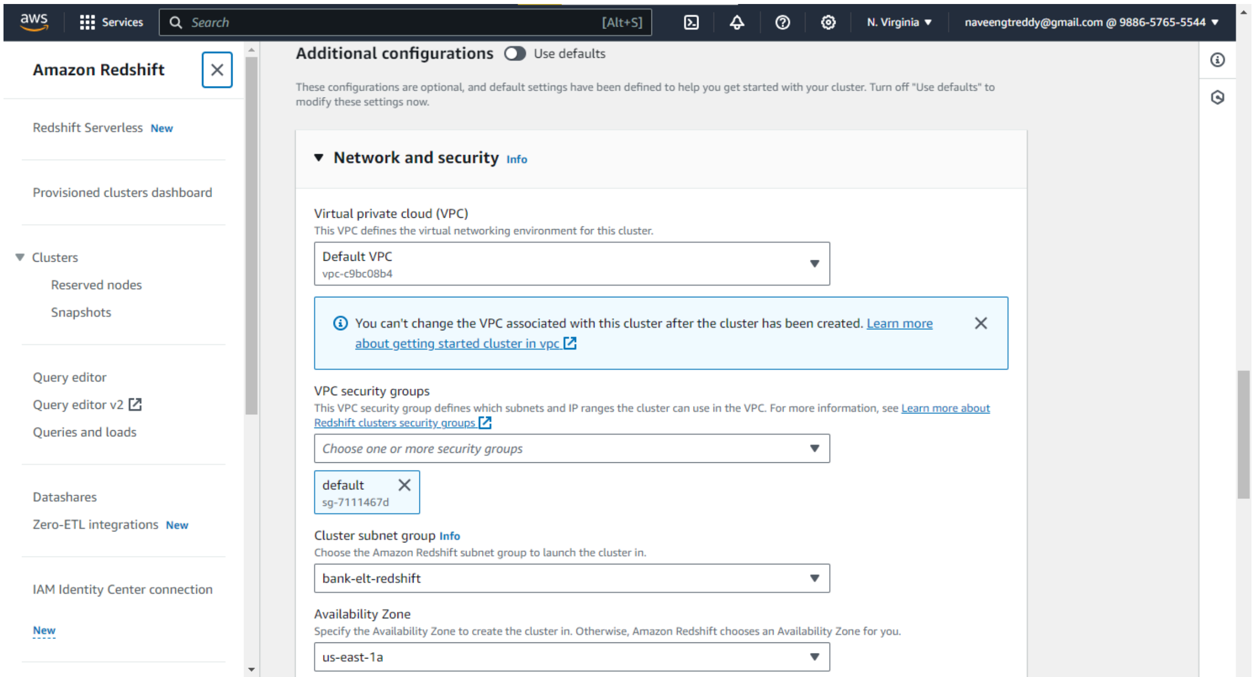


Cluster permissions

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.

□	IAM roles	Status	Role type
□	AmazonRedshift-CommandsAccessRole-20240204T133133	Not applied	Default



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 Info

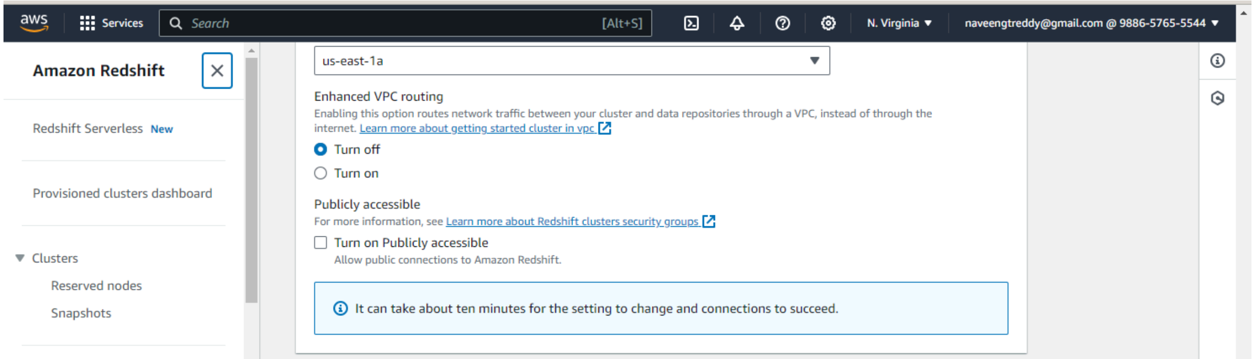
Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.
Default VPC
vpc-c9bc08b4

You can't change the VPC associated with this cluster after the cluster has been created. Learn more about getting started cluster in vpc

VPC security groups
This VPC security group defines which subnets and IP ranges the cluster can use in the VPC. For more information, see Learn more about Redshift clusters security groups
Choose one or more security groups
default
sg-7111467d

Cluster subnet group Info
Choose the Amazon Redshift subnet group to launch the cluster in.
bank-elt-redshift

Availability Zone
Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.
us-east-1a



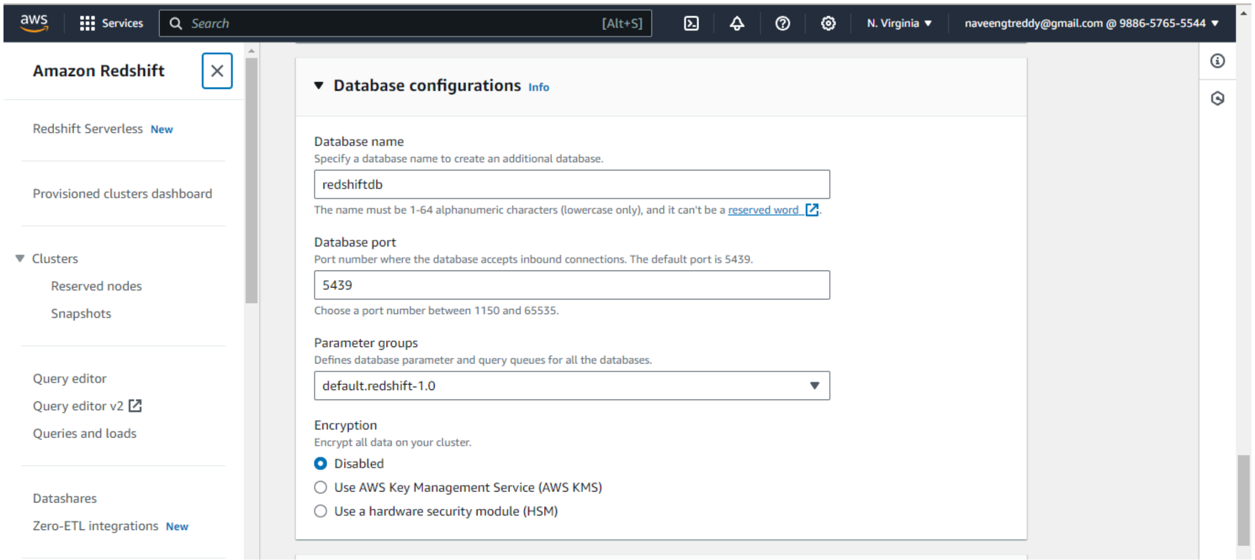
us-east-1a

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 about getting started cluster in vpc

● Turn off
○ Turn on

Publicly accessible
For more information, see Learn more about Redshift clusters security groups
□ Turn on Publicly accessible
Allow public connections to Amazon Redshift.

It can take about ten minutes for the setting to change and connections to succeed.



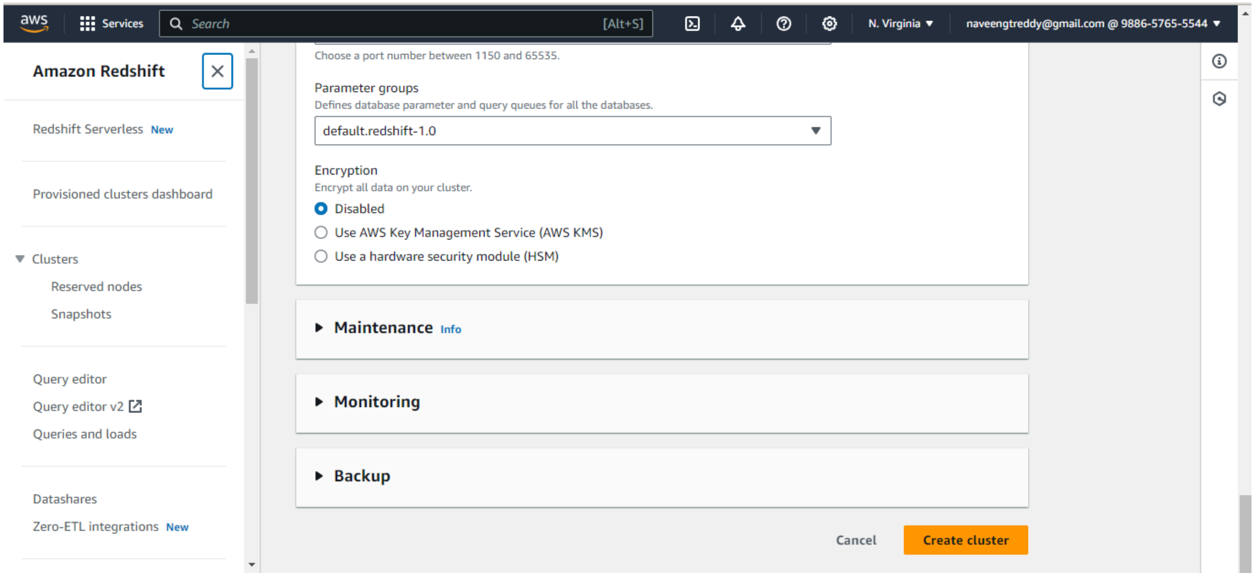
Database configurations Info

Database name
Specify a database name to create an additional database.
redshiftdb
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. The default port is 5439.
5439
Choose a port number between 1150 and 65535.

Parameter groups
Defines database parameter and query queues for all the databases.
default.redshift-1.0

Encryption
Encrypt all data on your cluster.
● Disabled
○ Use AWS Key Management Service (AWS KMS)
○ Use a hardware security module (HSM)

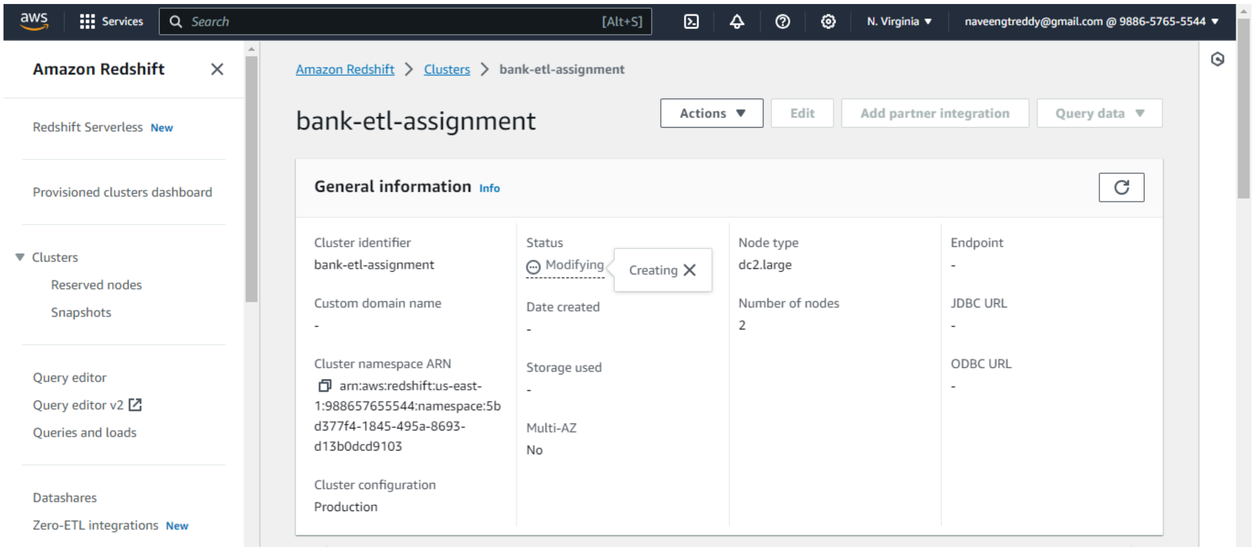


Choose a port number between 1150 and 65535.

Parameter groups
Defines database parameter and query queues for all the databases.
default.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 cluster under creation status



bank-etl-assignment

Cluster identifier
bank-etl-assignment

Custom domain name
-

Cluster namespace ARN
arn:aws:redshift:us-east-1:988657655544:namespace:5bd377f4-1845-495a-8693-d13b0dcd9103

Cluster configuration
Production

Status
Modifying
Creating

Date created
-

Storage used
-

Multi-AZ
No

Node type
dc2.large

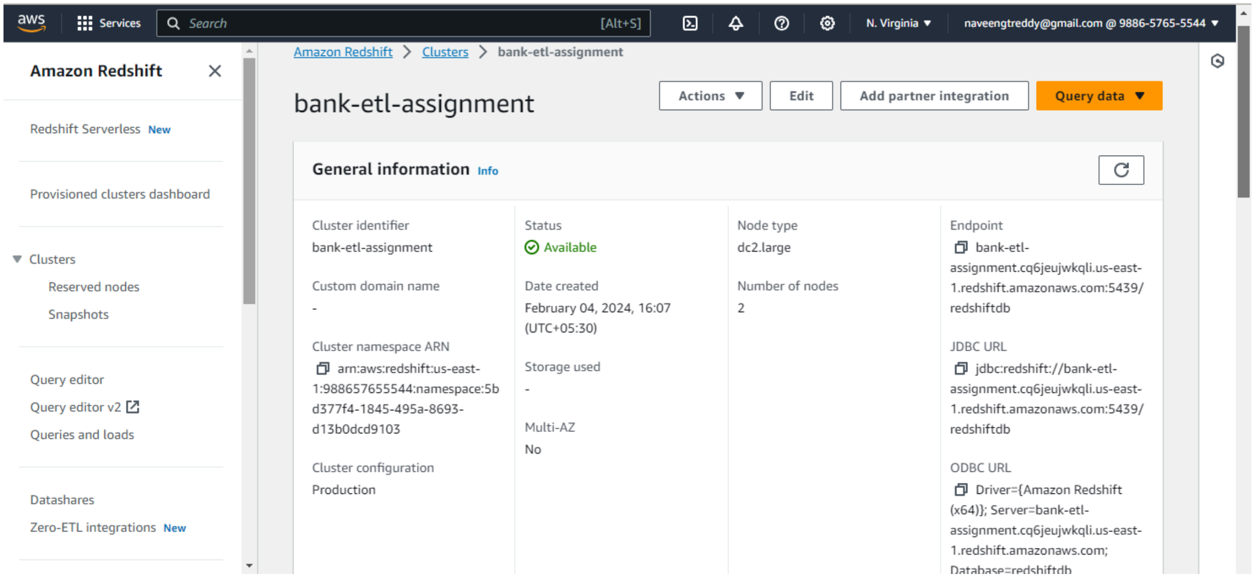
Number of nodes
2

Endpoint
-

JDBC URL
-

ODBC URL
-

Redshift cluster under available status -



bank-etl-assignment

General information

Cluster identifier
bank-etl-assignment

Custom domain name
-

Cluster namespace ARN
arn:aws:redshift:us-east-1:988657655544:namespace:5bd377f4-1845-495a-8693-d13b0dcd9103

Cluster configuration
Production

Status
Available

Date created
February 04, 2024, 16:07 (UTC+05:30)

Storage used
-

Multi-AZ
No

Node type
dc2.large

Number of nodes
2

Endpoint
bank-etl-assignment.cq6jeujwkqli.us-east-1.redshift.amazonaws.com:5439/redshiftdb

JDBC URL
jdbc:redshift://bank-etl-assignment.cq6jeujwkqli.us-east-1.redshift.amazonaws.com:5439/redshiftdb

ODBC URL
Driver={Amazon Redshift (x64)}; Server=bank-etl-assignment.cq6jeujwkqli.us-east-1.redshift.amazonaws.com; Database=redshiftdb

Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Loading data into a Redshift cluster from Amazon S3 bucket

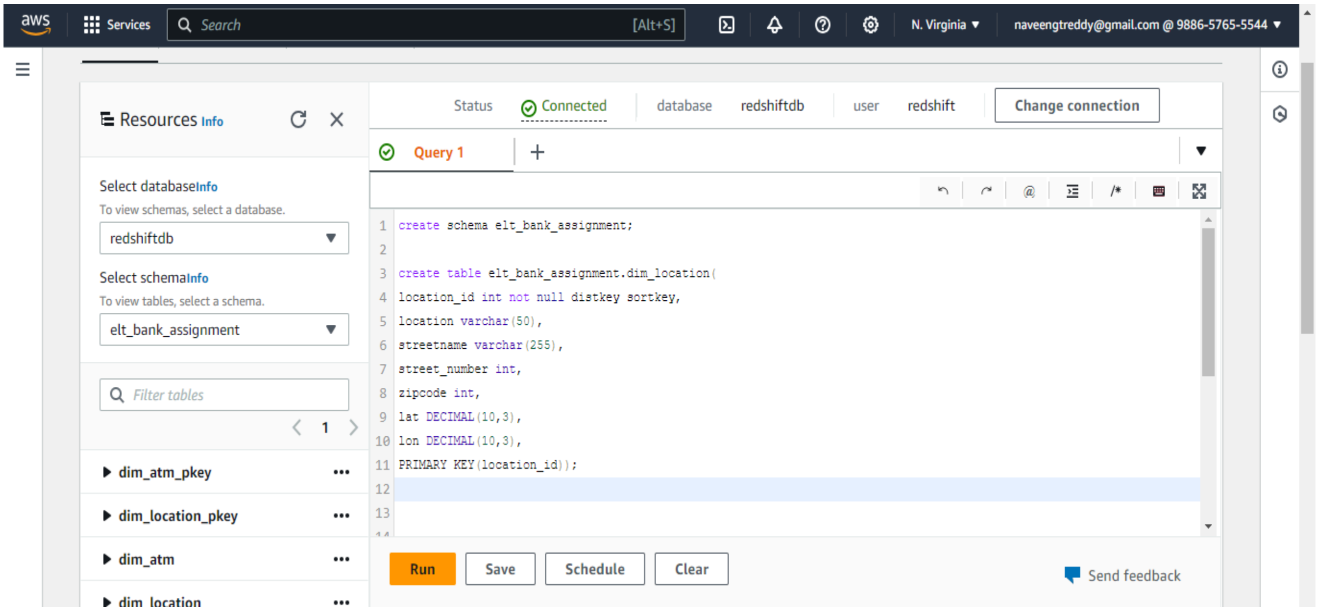
Queries to create the dimension and fact tables with appropriate primary and foreign keys

- **Creating Schema :**

```
create schema elt_bank_assignment;
```

- **Creating locatioin dimention Table.**

```
create table elt_bank_assignment.dim_location(
location_id int not null distkey sortkey,
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int,
lat DECIMAL(10, 3),
lon DECIMAL(10, 3),
PRIMARY KEY(location_id)
);
```



Status Connected database redshiftdb user redshift

Select database Info
To view schemas, select a database.
redshiftdb

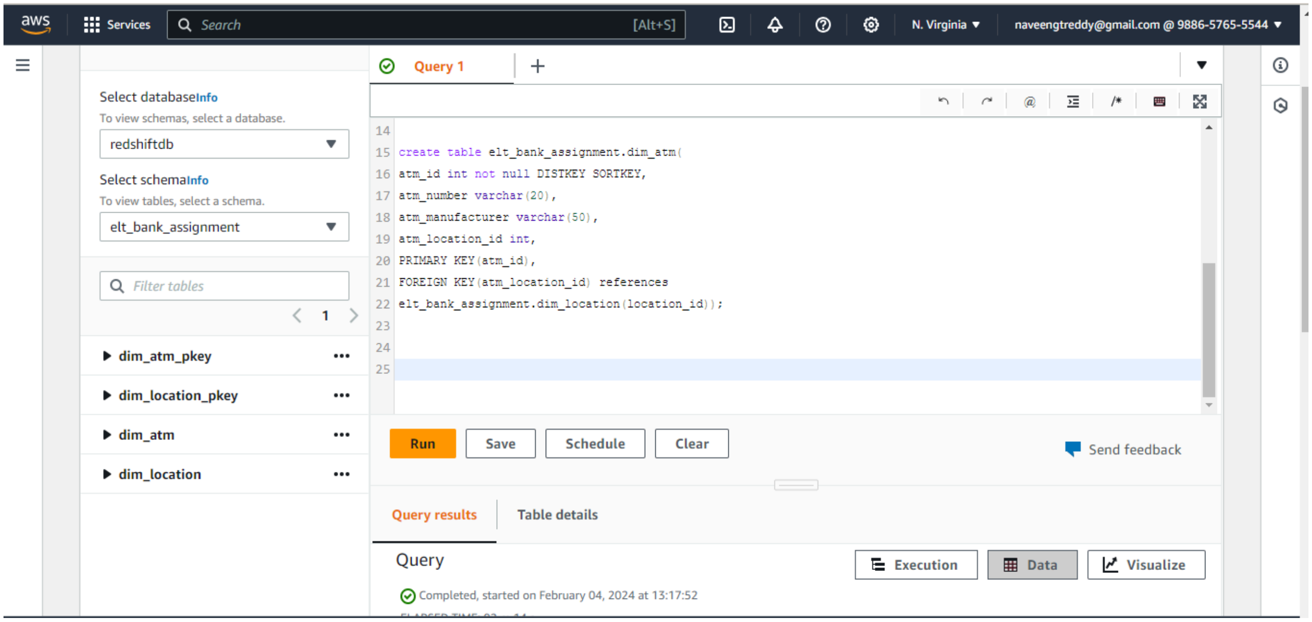
Select schema Info
To view tables, select a schema.
elt_bank_assignment

```
create schema elt_bank_assignment;

create table elt_bank_assignment.dim_location(
location_id int not null distkey sortkey,
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int,
lat DECIMAL(10,3),
lon DECIMAL(10,3),
PRIMARY KEY(location_id));
```

- **Creating ATM Dimention Table**

```
create table elt_bank_assignment.dim_atm(
atm_id int not null DISTKEY SORTKEY,
atm_number varchar(20),
atm_manufacturer varchar(50),
atm_location_id int,
PRIMARY KEY(atm_id),
FOREIGN KEY(atm_location_id) references
elt_bank_assignment.dim_location(location_id));
```



Select database Info
To view schemas, select a database.
redshiftdb

Select schema Info
To view tables, select a schema.
elt_bank_assignment

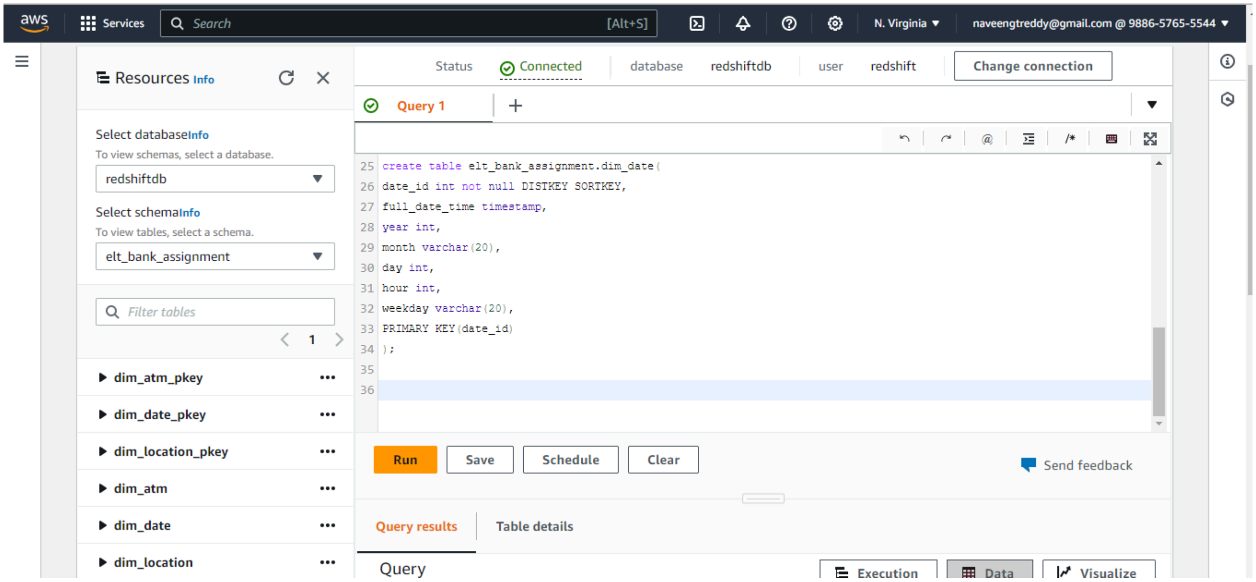
```
create table elt_bank_assignment.dim_atm(
atm_id int not null DISTKEY SORTKEY,
atm_number varchar(20),
atm_manufacturer varchar(50),
atm_location_id int,
PRIMARY KEY(atm_id),
FOREIGN KEY(atm_location_id) references
elt_bank_assignment.dim_location(location_id));
```

Query

Completed, started on February 04, 2024 at 13:17:52

- **Creating date dimension table**

```
create table elt_bank_assignment.dim_date(
date_id int not null DISTKEY SORTKEY,
full_date_time timestamp,
year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);
```



Select database: redshiftdb

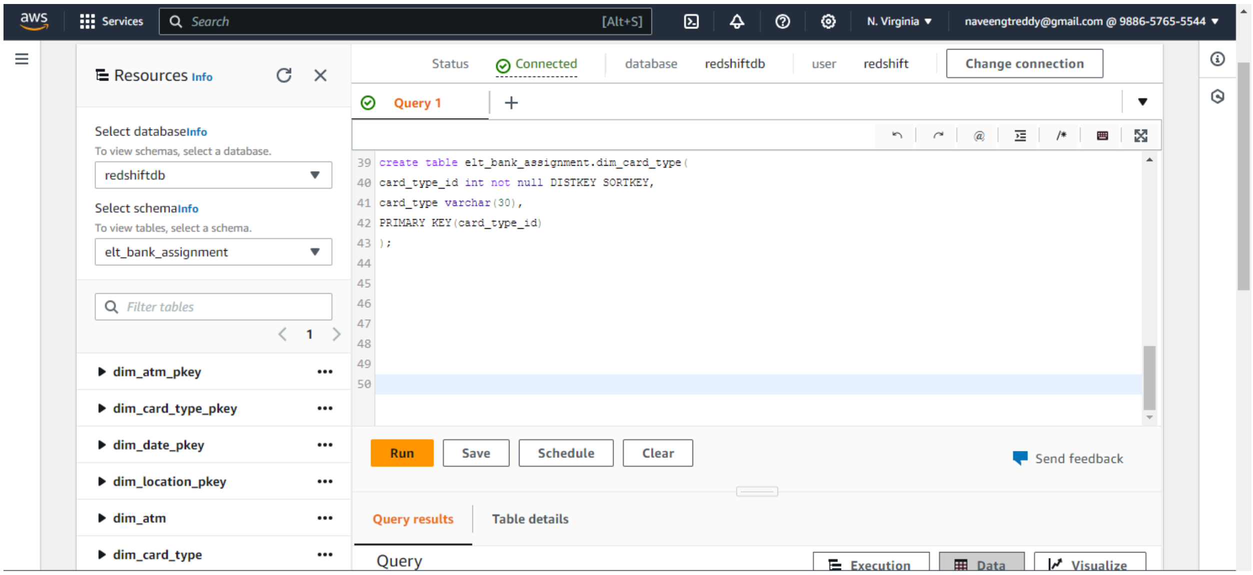
Select schema: elt_bank_assignment

Status: Connected | database: redshiftdb | user: redshift

```
create table elt_bank_assignment.dim_date(
date_id int not null DISTKEY SORTKEY,
full_date_time timestamp,
year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);
```

- **Creating card type dimension table**

```
create table elt_bank_assignment.dim_card_type(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);
```



Select database: redshiftdb

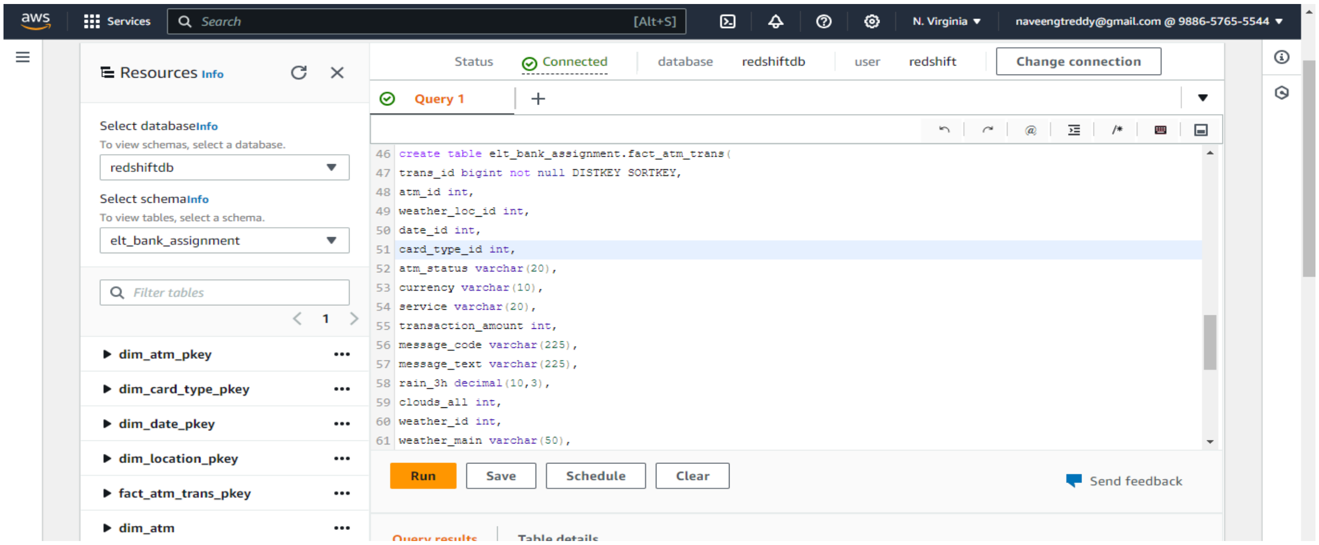
Select schema: elt_bank_assignment

Status: Connected | database: redshiftdb | user: redshift

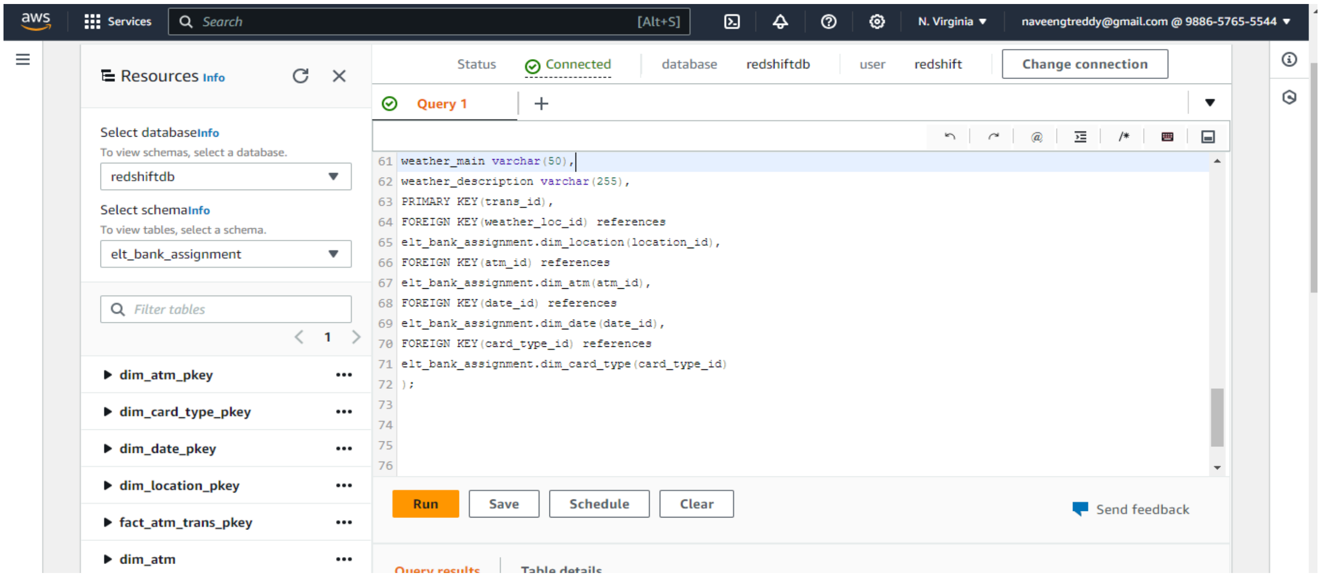
```
create table elt_bank_assignment.dim_card_type(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);
```

- **Creating atm transactions fact table**

```
create table elt_bank_assignment.fact_atm_trans(
trans_id bigint not null DISTKEY SORTKEY,
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(225),
message_text varchar(225),
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
elt_bank_assignment.dim_location(location_id),
FOREIGN KEY(atm_id) references
elt_bank_assignment.dim_atm(atm_id),
FOREIGN KEY(date_id) references
elt_bank_assignment.dim_date(date_id),
FOREIGN KEY(card_type_id) references
elt_bank_assignment.dim_card_type(card_type_id)
);
```



```
create table elt_bank_assignment.fact_atm_trans(
trans_id bigint not null DISTKEY SORTKEY,
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(225),
message_text varchar(225),
rain_3h decimal(10,3),
clouds_all int,
weather_id int,
weather_main varchar(50),
```



```
weather_main varchar(50),
weather_description varchar(255),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references
elt_bank_assignment.dim_location(location_id),
FOREIGN KEY(atm_id) references
elt_bank_assignment.dim_atm(atm_id),
FOREIGN KEY(date_id) references
elt_bank_assignment.dim_date(date_id),
FOREIGN KEY(card_type_id) references
elt_bank_assignment.dim_card_type(card_type_id)
);
```

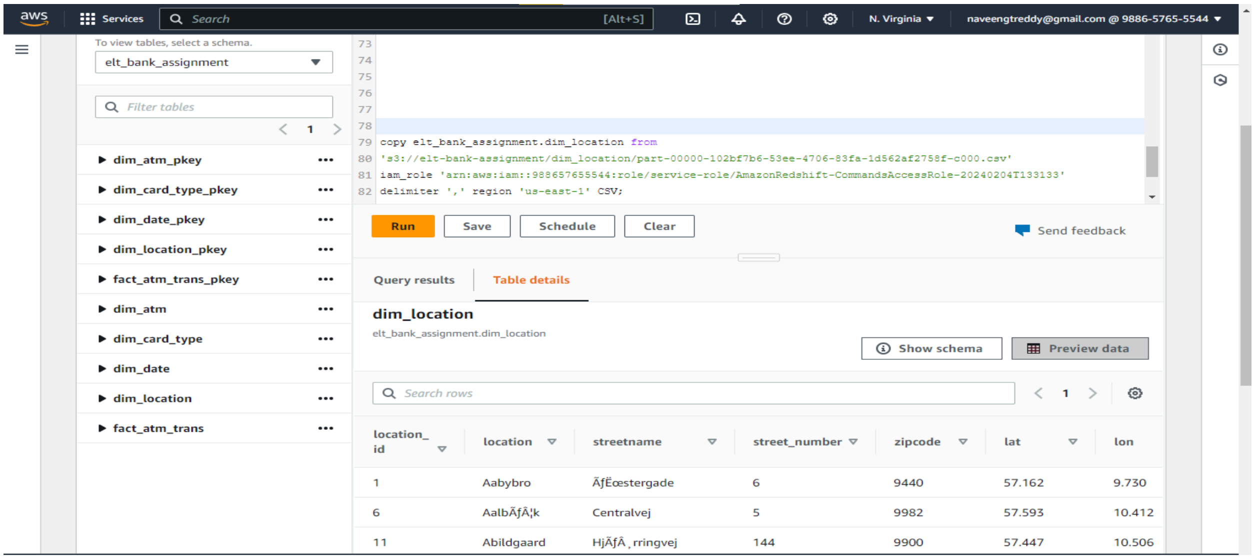
Loading data into a Redshift cluster from Amazon S3 bucket

Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

- **Copying the data to dim_location table**

```
copy elt_bank_assignment.dim_location from
's3://elt-bank-assignment/dim_location/part-00000-102bf7b6-53ee
-4706-83fa-1d562af2758f-c000.csv'
```

```
iam_role
'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1' CSV;
```



```
copy elt_bank_assignment.dim_location from
's3://elt-bank-assignment/dim_location/part-00000-102bf7b6-53ee-4706-83fa-1d562af2758f-c000.csv'
iam_role 'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1' CSV;
```

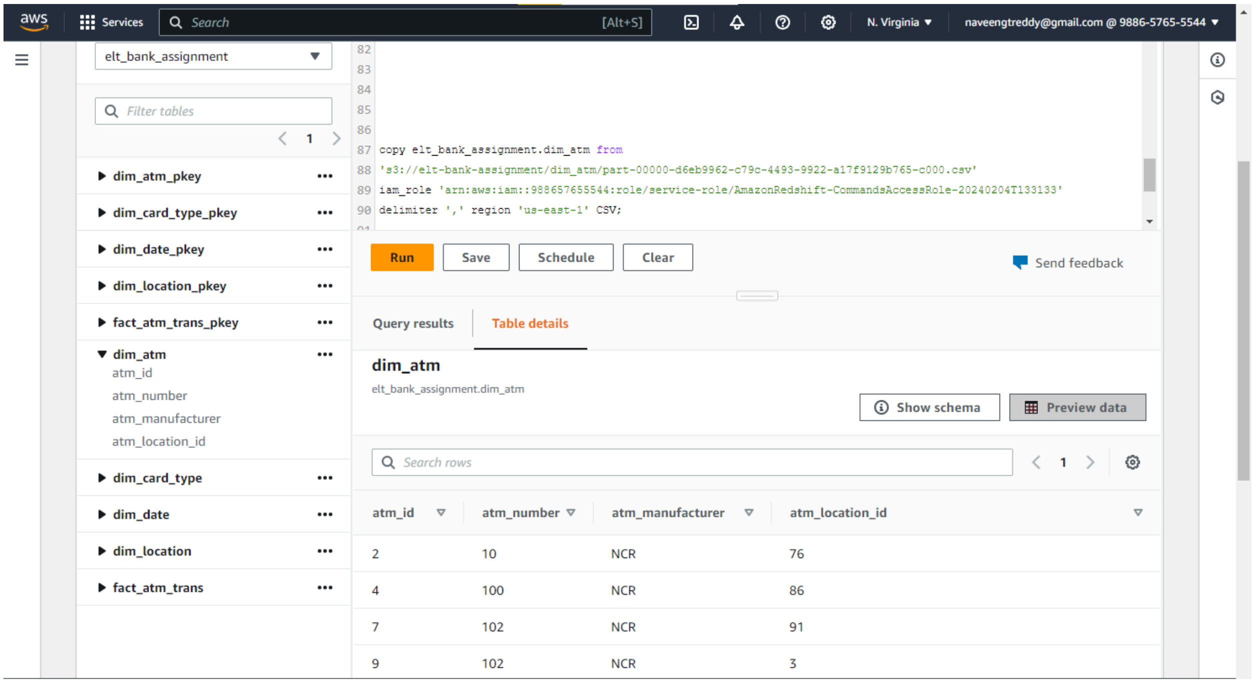
dim_location

elt_bank_assignment.dim_location

location_id	location	streetname	street_number	zipcode	lat	lon
1	Aabybro	ÃƒËœstergade	6	9440	57.162	9.730
6	AalbÃƒÂ¦k	Centralvej	5	9982	57.593	10.412
11	Abildgaard	HjÃƒÂ¸rringvej	144	9900	57.447	10.506

- **Copying the data to dim_atm table**

```
copy elt_bank_assignment.dim_atm from
's3://elt-bank-assignment/dim_atm/part-00000-d6eb9962-c79c-4493-9922-a17f9129b765-c000.csv'
iam_role
'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1' CSV;
```



```
copy elt_bank_assignment.dim_atm from
's3://elt-bank-assignment/dim_atm/part-00000-d6eb9962-c79c-4493-9922-a17f9129b765-c000.csv'
iam_role 'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1' CSV;
```

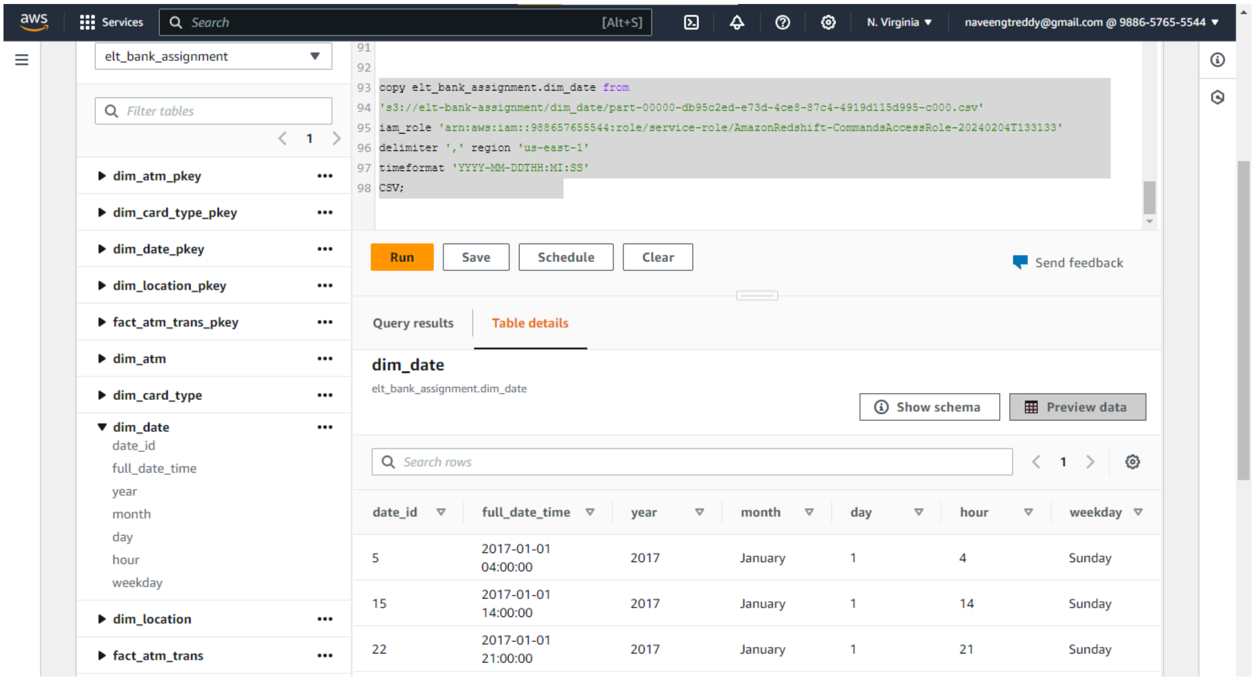
dim_atm

elt_bank_assignment.dim_atm

atm_id	atm_number	atm_manufacturer	atm_location_id
2	10	NCR	76
4	100	NCR	86
7	102	NCR	91
9	102	NCR	3

- **Copying the data to dim_date table**

```
copy elt_bank_assignment.dim_date from
's3://elt-bank-assignment/dim_date/part-00000-db95c2ed-e73d-4ce8-87c4-4919d115d995-c000.csv'
iam_role
'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



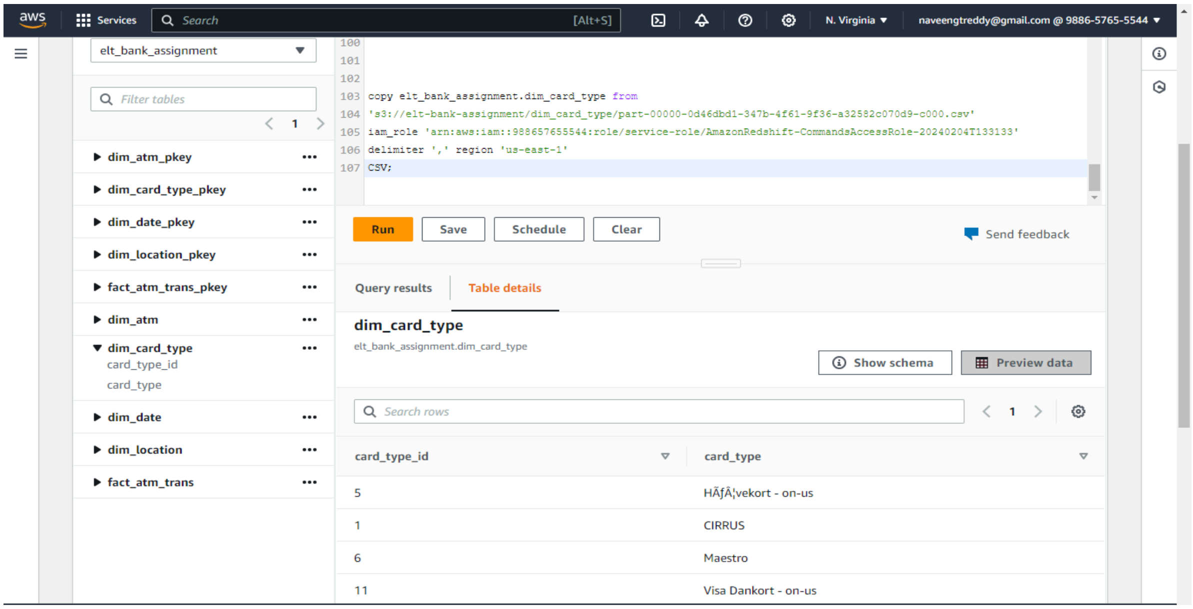
```
copy elt_bank_assignment.dim_date from
's3://elt-bank-assignment/dim_date/part-00000-db95c2ed-e73d-4ce8-87c4-4919d115d995-c000.csv'
iam_role 'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```

dim_date
elt_bank_assignment.dim_date

date_id	full_date_time	year	month	day	hour	weekday
5	2017-01-01 04:00:00	2017	January	1	4	Sunday
15	2017-01-01 14:00:00	2017	January	1	14	Sunday
22	2017-01-01 21:00:00	2017	January	1	21	Sunday

- **Copying the data to dim_card_type table**

```
copy elt_bank_assignment.dim_card_type from
's3://elt-bank-assignment/dim_card_type/part-00000-0d46dbd1-347
b-4f61-9f36-a32582c070d9-c000.csv'
iam_role
'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-Com
mandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
CSV;
```



```
copy elt_bank_assignment.dim_card_type from
's3://elt-bank-assignment/dim_card_type/part-00000-0d46dbd1-347b-4f61-9f36-a32582c070d9-c000.csv'
iam_role 'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
CSV;
```

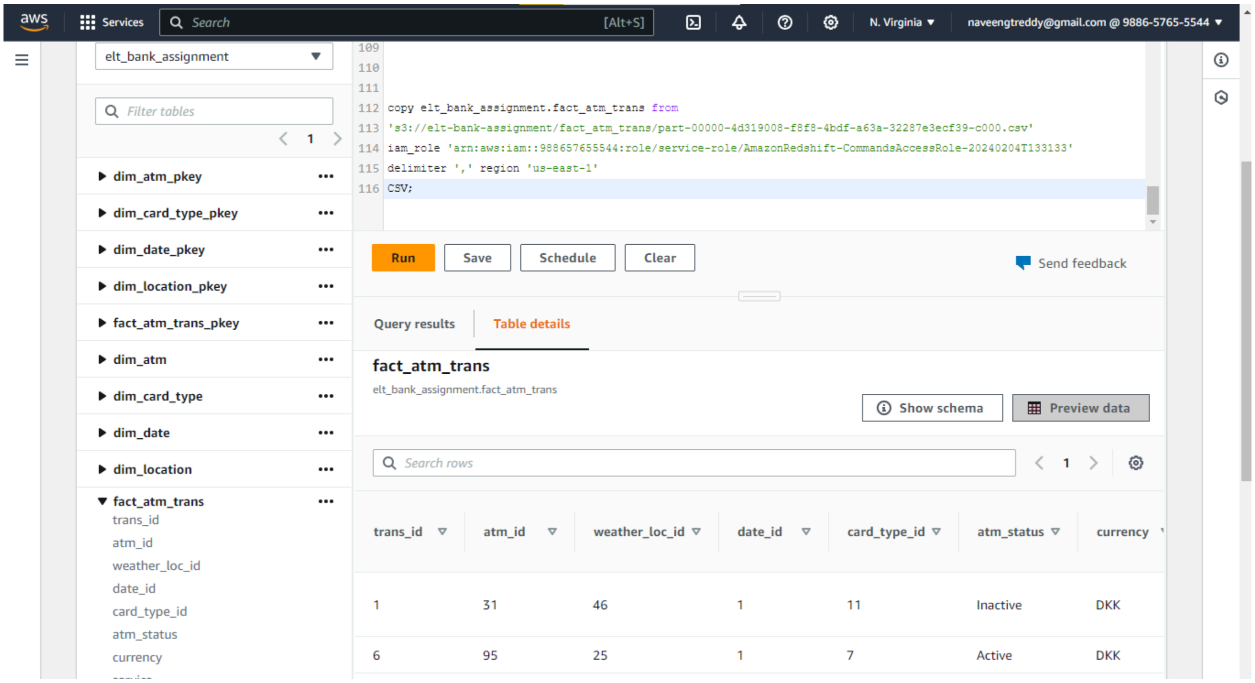
dim_card_type

elt_bank_assignment.dim_card_type

card_type_id	card_type
5	HÃƒÂ¦vekort - on-us
1	CIRRUS
6	Maestro
11	Visa Dankort - on-us

- **Copying the data to fact_atm_trans table**

```
copy elt_bank_assignment.fact_atm_trans from
's3://elt-bank-assignment/fact_atm_trans/part-00000-4d319008-f8
f8-4bdf-a63a-32287e3ecf39-c000.csv'
iam_role
'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-Com
mandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
CSV;
```



```
copy elt_bank_assignment.fact_atm_trans from
's3://elt-bank-assignment/fact_atm_trans/part-00000-4d319008-f8f8-4bdf-a63a-32287e3ecf39-c000.csv'
iam_role 'arn:aws:iam::988657655544:role/service-role/AmazonRedshift-CommandsAccessRole-20240204T133133'
delimiter ',' region 'us-east-1'
CSV;
```

fact_atm_trans

elt_bank_assignment.fact_atm_trans

trans_id	atm_id	weather_loc_id	date_id	card_type_id	atm_status	currency
1	31	46	1	11	Inactive	DKK
6	95	25	1	7	Active	DKK