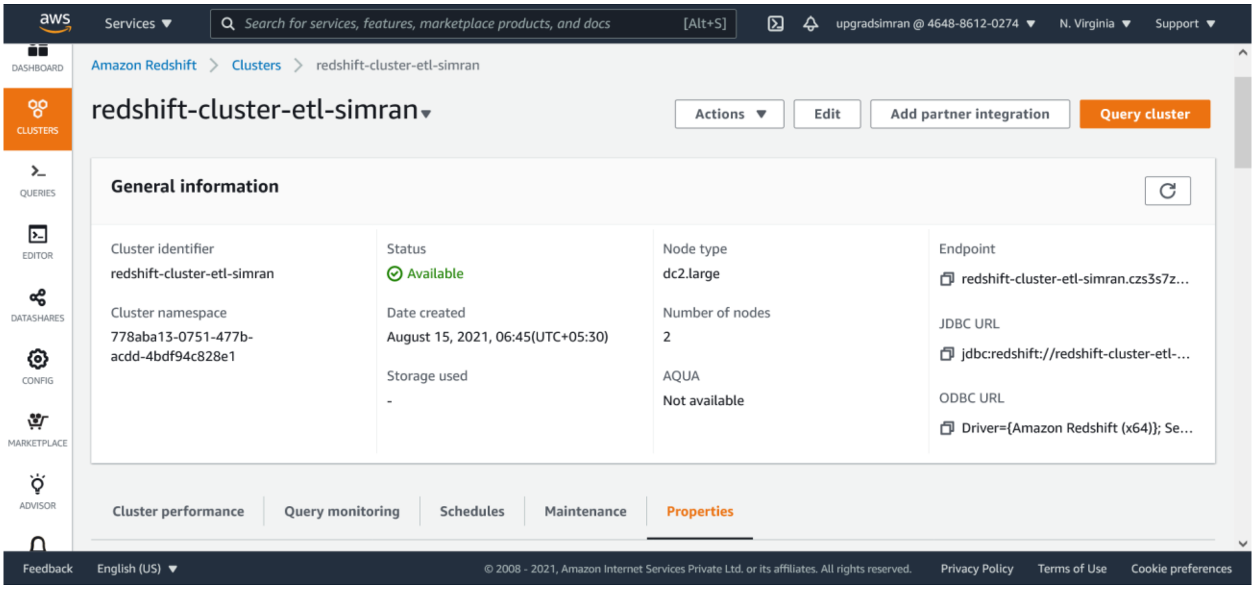
Creation of a RedShift Cluster

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

Screenshot of the type of machine used along with number of nodes:



redshift-cluster-etl-simran

General information

Cluster identifier: redshift-cluster-etl-simran

Cluster namespace: 778aba13-0751-477b-acdd-4bdf94c828e1

Status: Available

Date created: August 15, 2021, 06:45(UTC+05:30)

Storage used: -

Node type: dc2.large

Number of nodes: 2

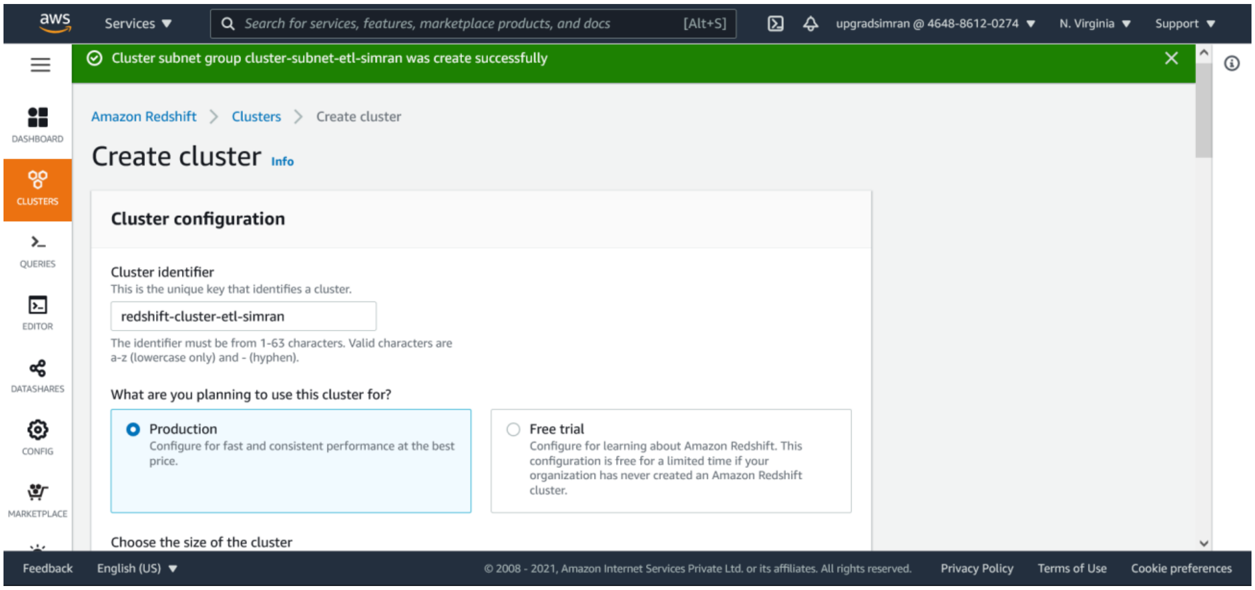
AQUA: Not available

Endpoint: redshift-cluster-etl-simran.czs3s7z...

JDBC URL: jdbc:redshift://redshift-cluster-etl-...

ODBC URL: Driver={Amazon Redshift (x64)}; Se...

Screenshots of the various configurations associated with cluster creation:


aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
DASHBOARD
CLUSTERS
QUERIES
EDITOR
DATASHARES
CONFIG
MARKETPLACE
Amazon Redshift > Clusters > Create cluster
Info
Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences

Cluster subnet group cluster-subnet-etl-simran was create successfully

Create cluster

Cluster configuration

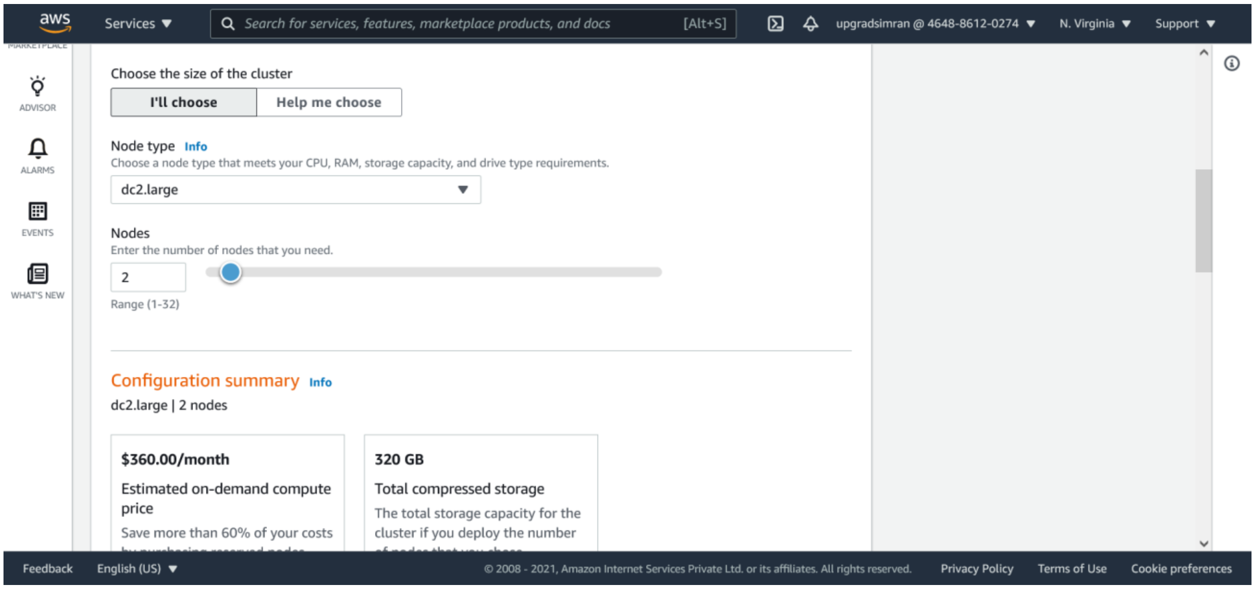
Cluster identifier
This is the unique key that identifies a cluster.
redshift-cluster-etl-simran
The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

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


aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
ADVISOR
ALARMS
EVENTS
WHAT'S NEW

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 ▼

Nodes

Enter the number of nodes that you need.

2

Range (1-32)

Configuration summary Info

dc2.large | 2 nodes

$360.00/month

Estimated on-demand compute price

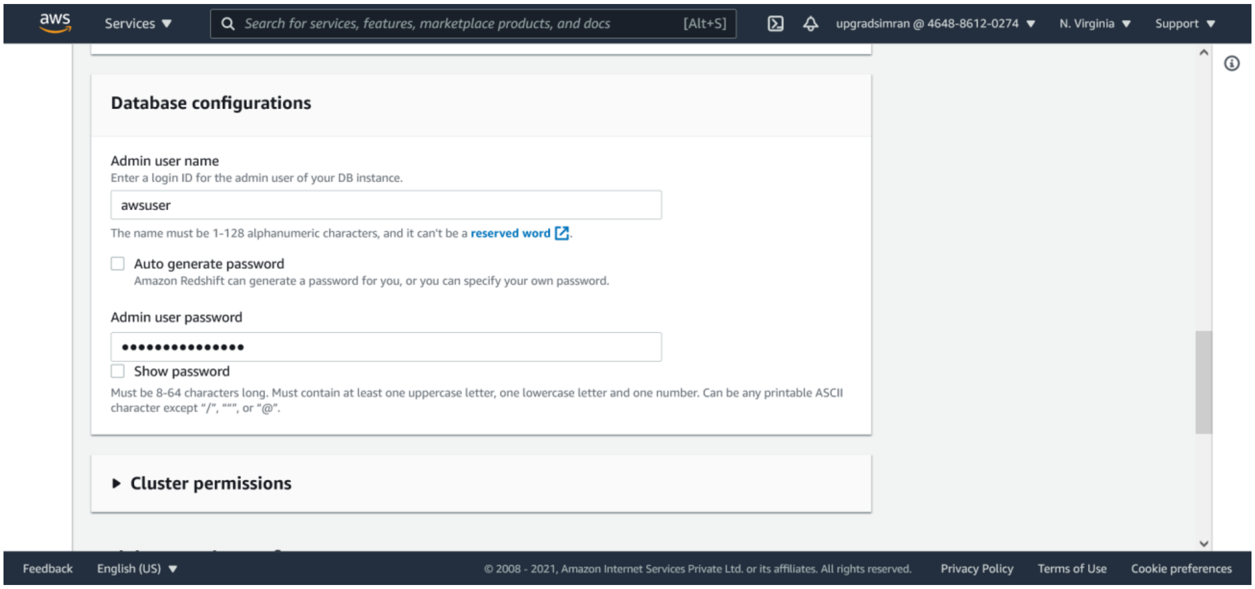
Save more than 60% of your costs

320 GB

Total compressed storage

The total storage capacity for the cluster if you deploy the number

Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences



aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences

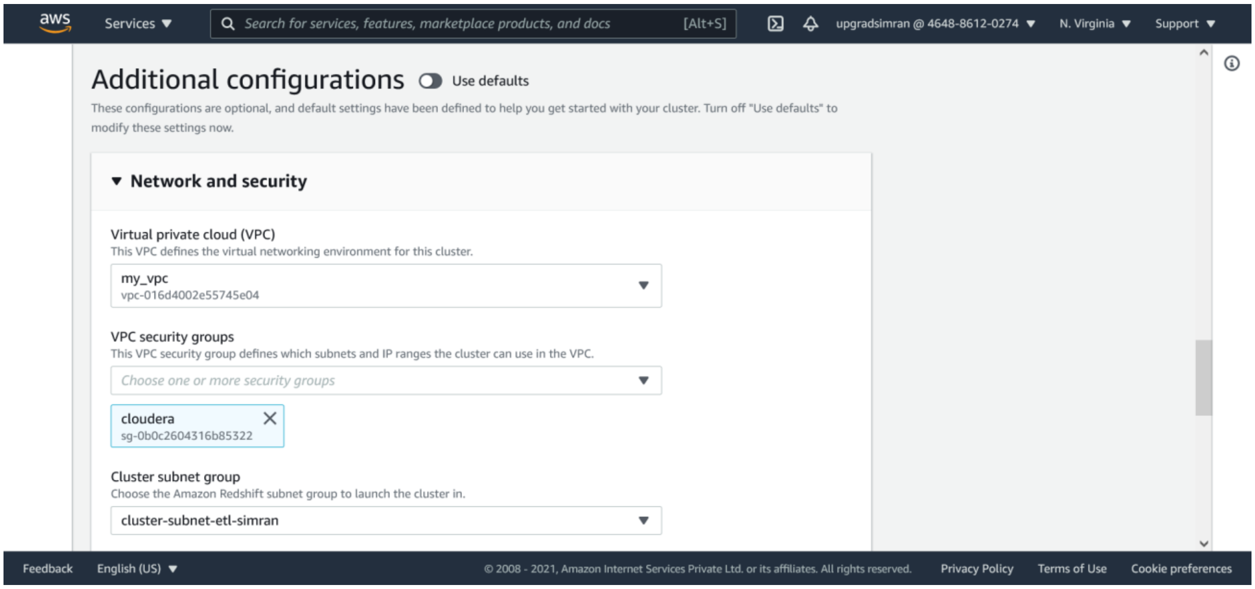
Database configurations

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

▶ Cluster permissions



aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences

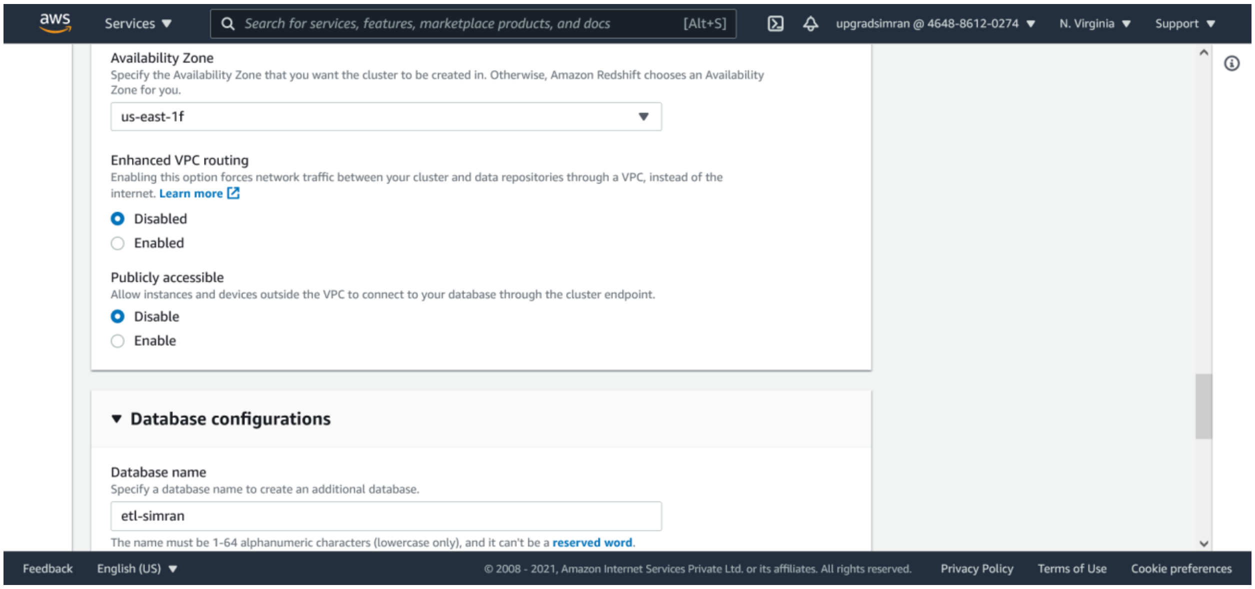
Additional configurations ○ Use defaults
These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

▼ Network and security

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.
my_vpc
vpc-016d4002e55745e04

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-0b0c2604316b85322

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


aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences

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-1f ▼

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.

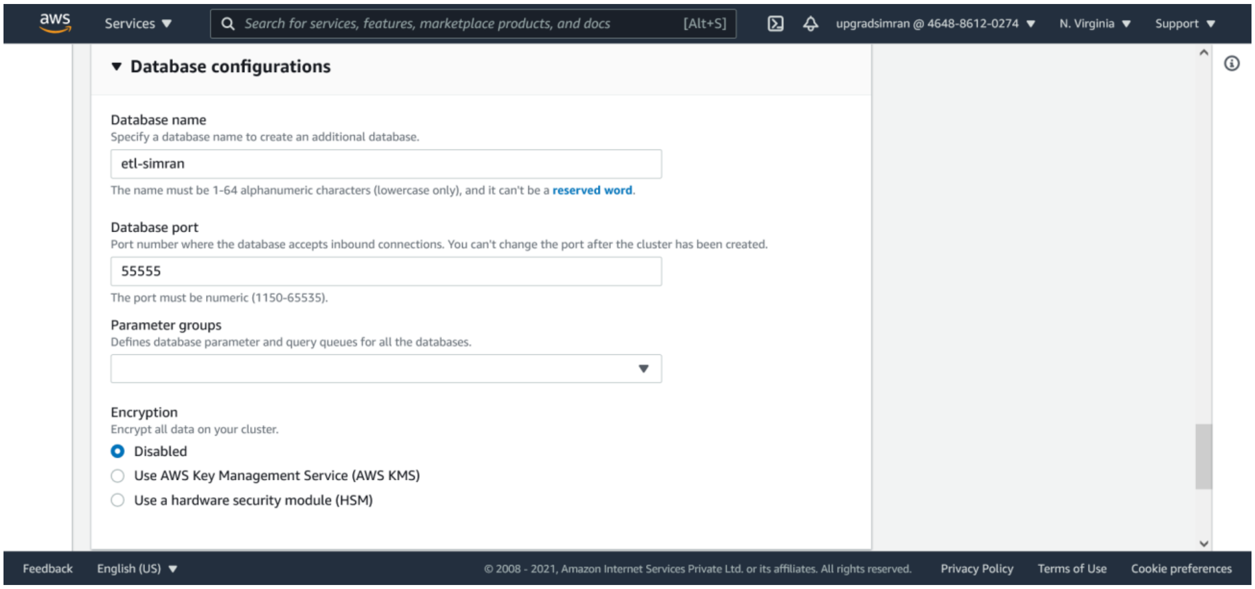
● Disable
○ Enable

▼ Database configurations

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

etl-simran

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a **reserved word**.


aws
Services ▼
Search for services, features, marketplace products, and docs [Alt+S]
upgradsimran @ 4648-8612-0274 ▼
N. Virginia ▼
Support ▼
Feedback
English (US) ▼
© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Cookie preferences

▼ Database configurations

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

etl-simran

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.

55555

The port must be numeric (1150-65535).

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

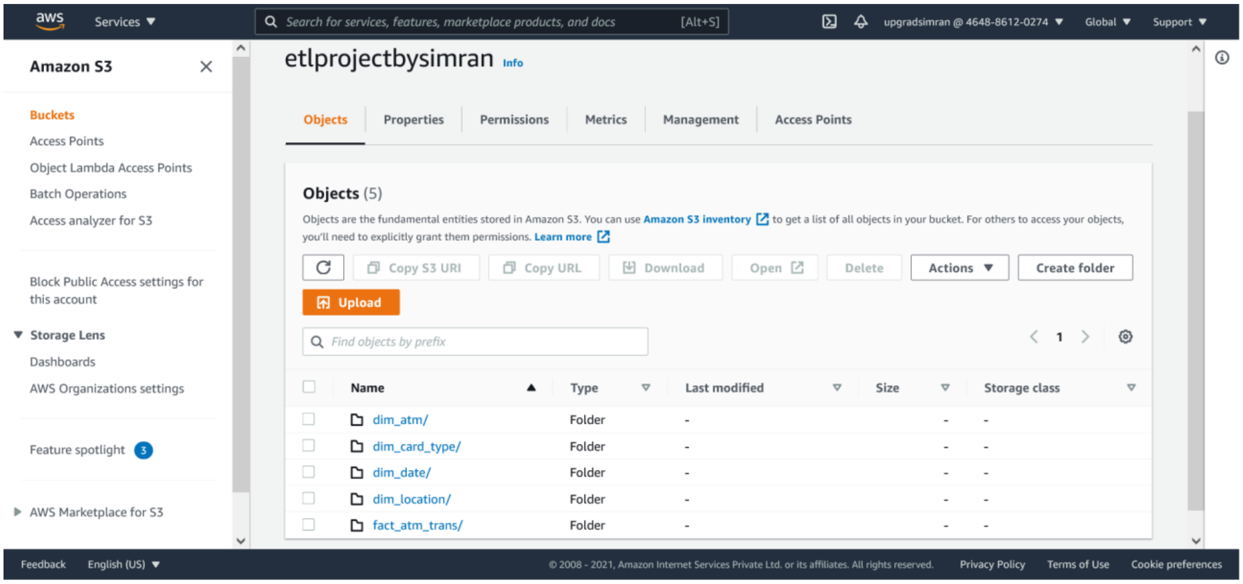
▼

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

Viewing all the data in Amazon S3 bucket:



etlprojectbysimran Info

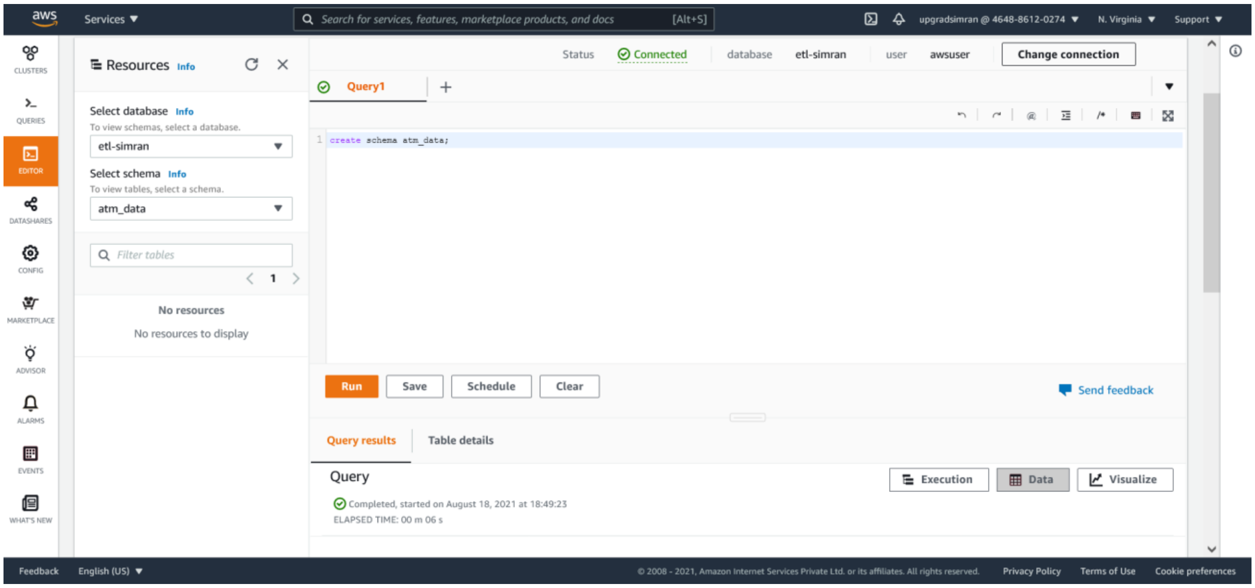
Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Name	Type	Last modified	Size	Storage class
dim_atm/	Folder	-	-	-
dim_card_type/	Folder	-	-	-
dim_date/	Folder	-	-	-
dim_location/	Folder	-	-	-
fact_atm_trans/	Folder	-	-	-

Query to create a schema for the dimension and fact tables:

create schema atm_data;



Status Connected database etl-simran user awsuser

Select database Info
To view schemas, select a database.
etl-simran

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

Filter tables

No resources
No resources to display

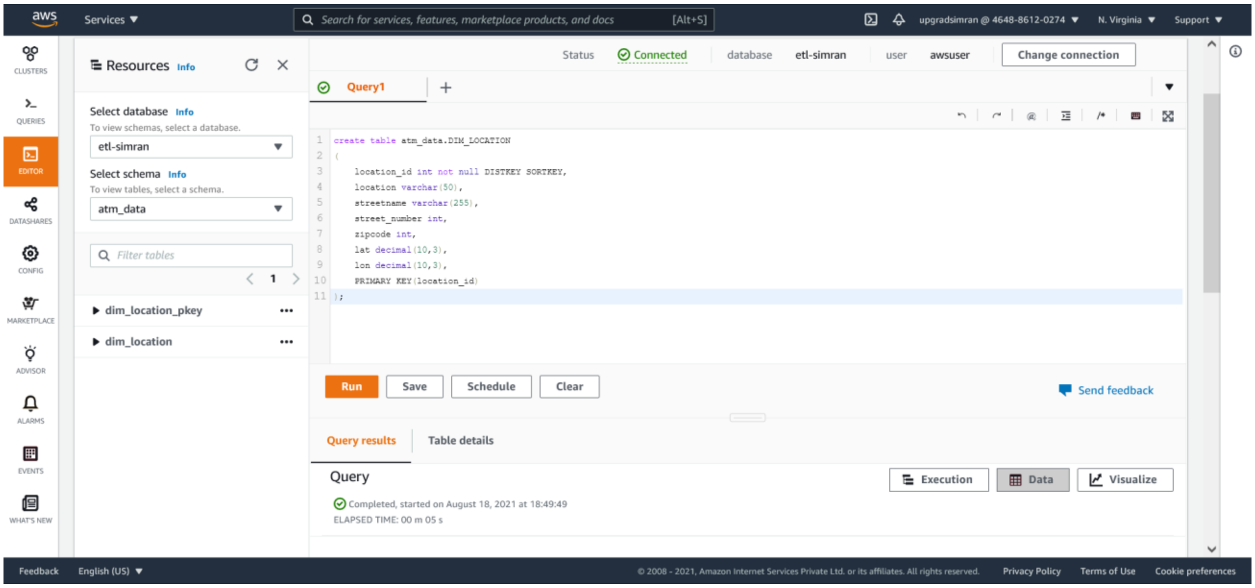
```
create schema atm_data;
```

Query
Completed, started on August 18, 2021 at 18:49:23
ELAPSED TIME: 00 m 06 s

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

- **Creating location dimension table**

```
create table atm_data.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 etl-simran user awsuser

Select database Info
To view schemas, select a database.
etl-simran

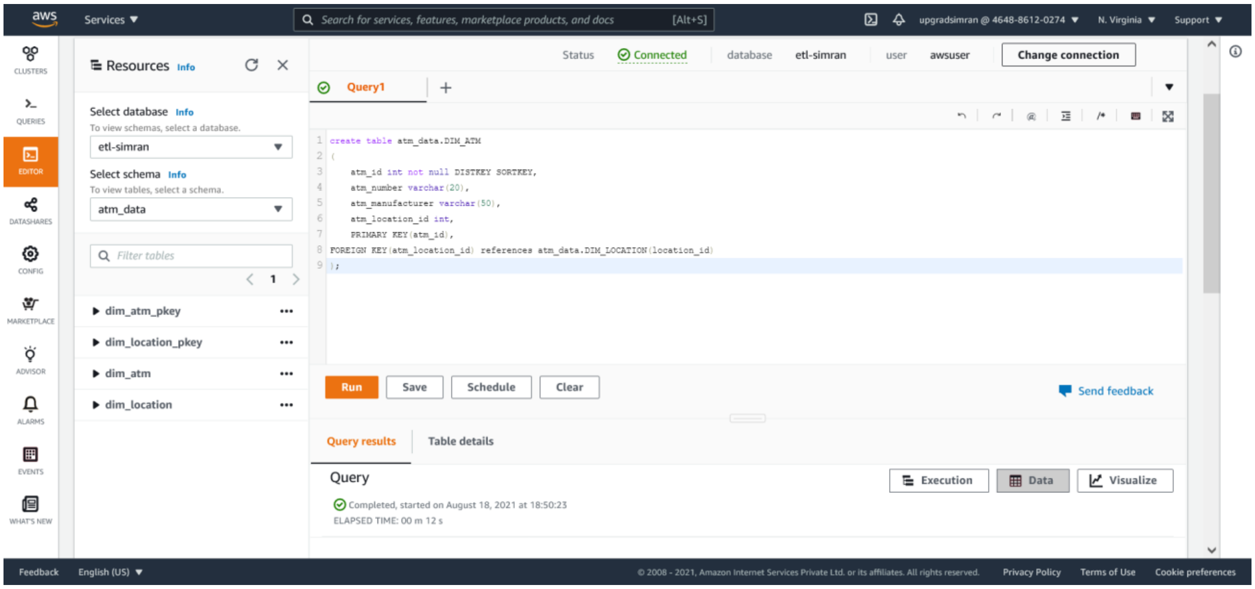
Select schema Info
To view tables, select a schema.
atm_data

```
create table atm_data.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)
);
```

Query
✓ Completed, started on August 18, 2021 at 18:49:49
ELAPSED TIME: 00 m 05 s

- **Creating atm dimension table**

```
create table atm_data.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 atm_data.DIM_LOCATION(location_id)
);
```



Status: Connected | database: etl-simran | user: awsuser

Select database: etl-simran

Select schema: atm_data

```
create table atm_data.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 atm_data.DIM_LOCATION(location_id)
);
```

Query

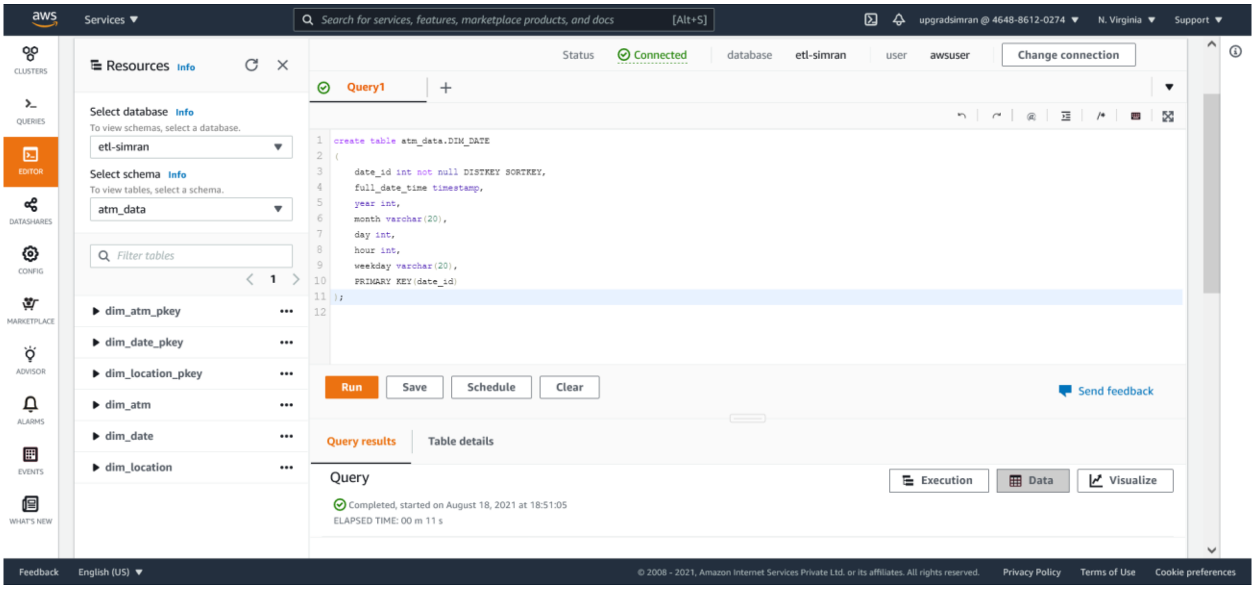
Completed, started on August 18, 2021 at 18:50:23

ELAPSED TIME: 00 m 12 s

- **Creating date dimension table**

```
create table atm_data.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)

);
```



Status: Connected | database: etl-simran | user: awsuser

Select database: etl-simran

Select schema: atm_data

```
create table atm_data.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)
);
```

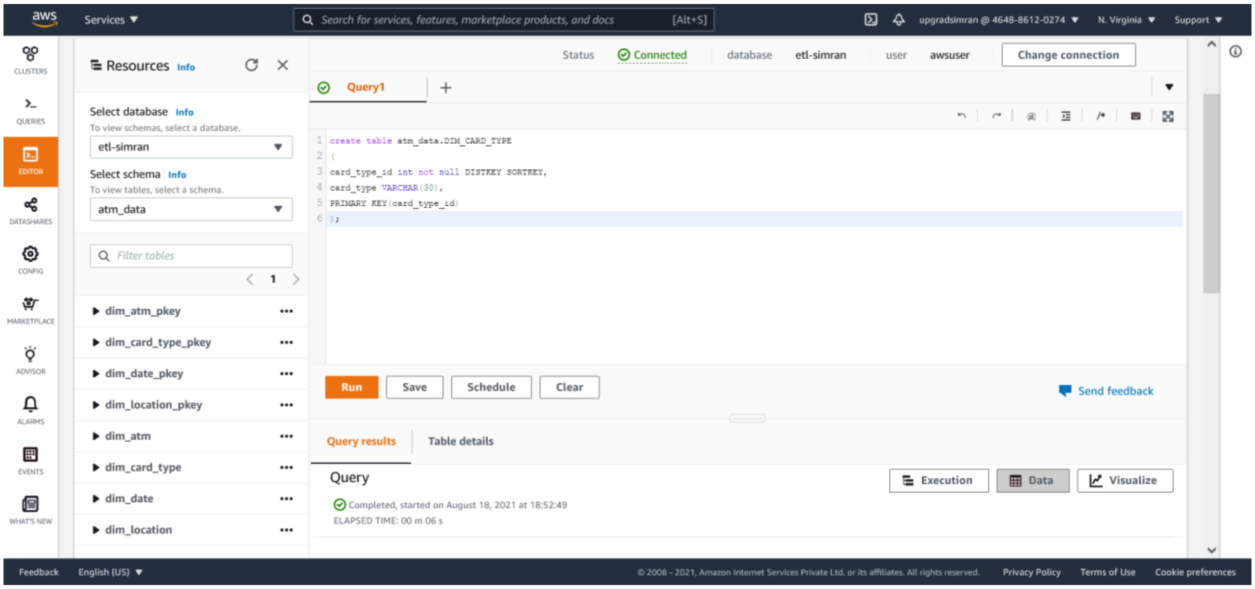
Query

Completed, started on August 18, 2021 at 18:51:05

ELAPSED TIME: 00 m 11 s

- **Creating card type dimension table**

create table atm_data.DIM_CARD_TYPE
(
    card_type_id int not null DISTKEY SORTKEY,
    card_type varchar(30)
    PRIMARY KEY(card_type_id)
);



Status: Connected | database: etl-simran | user: awsuser

Select database: etl-simran

Select schema: atm_data

```
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type VARCHAR(30),
PRIMARY KEY(card_type_id)
);
```

Query

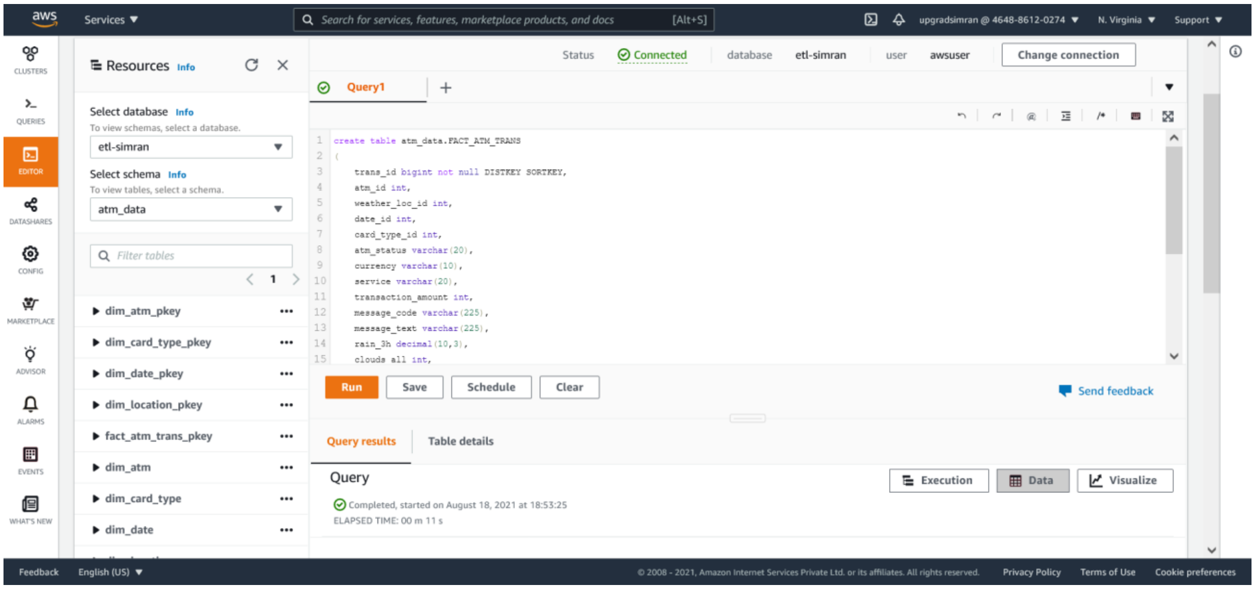
Completed, started on August 18, 2021 at 18:52:49

ELAPSED TIME: 00 m 06 s

- **Creating atm transactions fact table**

create table atm_data.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 atm_data.DIM_LOCATION(location_id),
        FOREIGN KEY(atm_id) references atm_data.DIM_DATA(atm_id),
        FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
        FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
);
```



Status: ✓ Connected | database: etl-simran | user: awsuser

Select database: etl-simran

Select schema: atm_data

▶ dim_atm_pkey •••

▶ dim_card_type_pkey •••

▶ dim_date_pkey •••

▶ dim_location_pkey •••

▶ fact_atm_trans_pkey •••

▶ dim_atm •••

▶ dim_card_type •••

▶ dim_date •••

```
create table atm_data.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,
```

Query

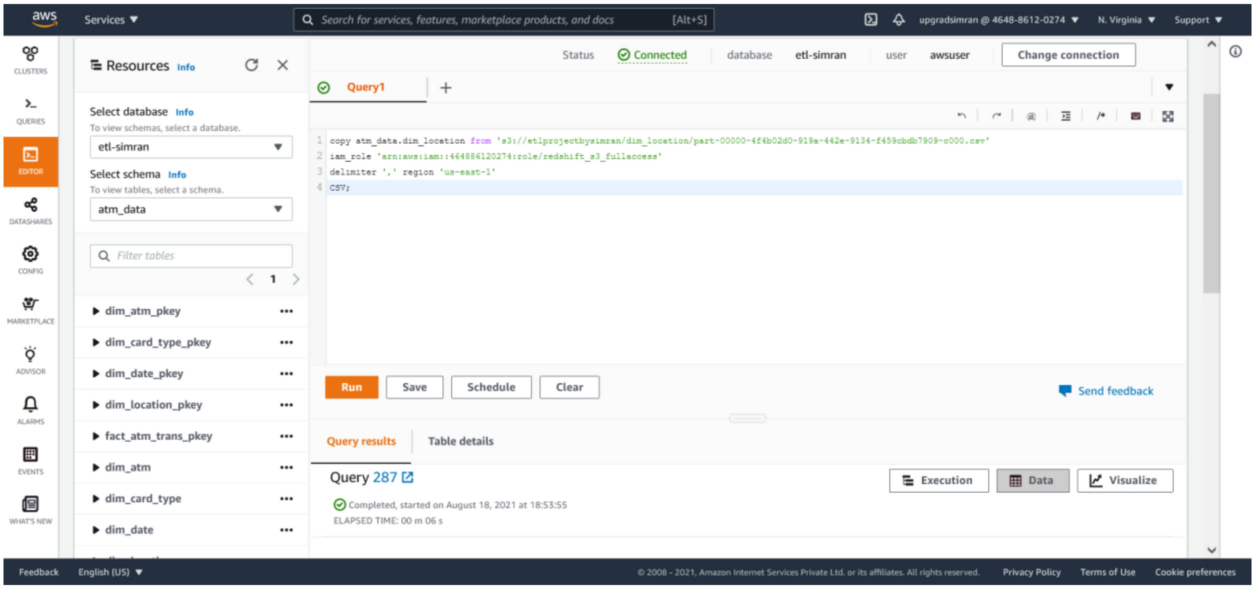
✓ Completed, started on August 18, 2021 at 18:53:25
ELAPSED TIME: 00 m 11 s

Loading data into a RedShift cluster from Amazon S3 bucket

Queries to copy the data from S3 bucket to the RedShift cluster in the appropriate tables:

- **Copying the data to dim_location table**

copy atm_data.dim_location from 's3://etlprojectbysimran/dim_location/part-00000-4f4b02d0-919a-442e-9134-f459cbdb7909-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;



```
copy atm_data.dim_location from 's3://etlprojectbysimran/dim_location/part-00000-4f4b02d0-919a-442e-9134-f459cbdb7909-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Status: Connected | database: etl-simran | user: awsuser

Select database: etl-simran

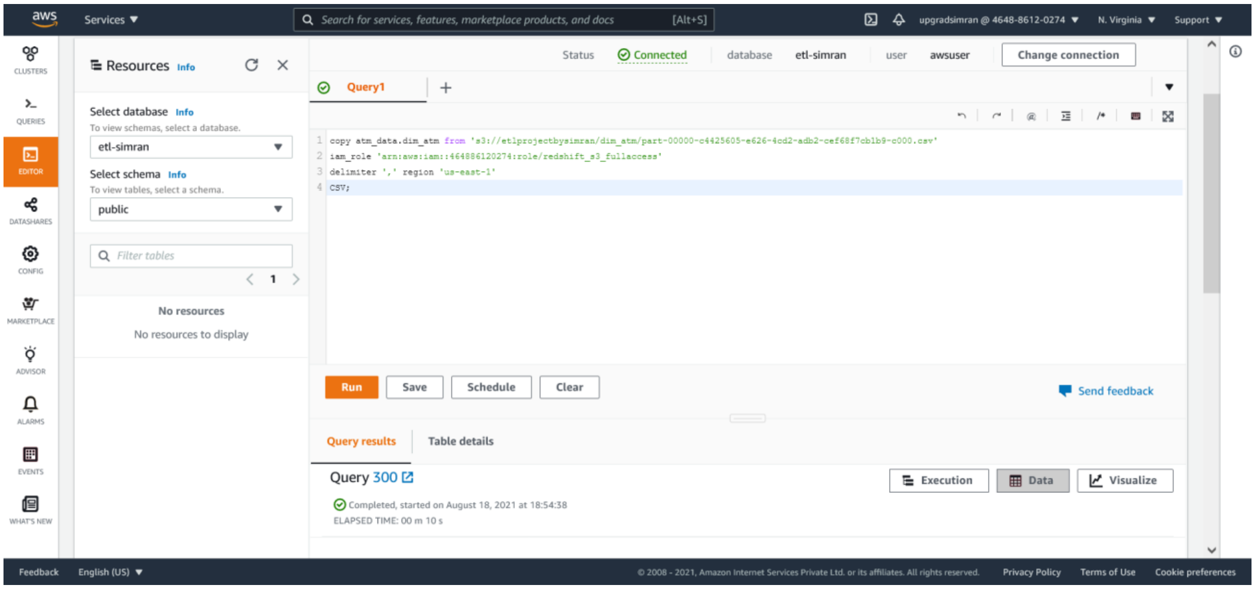
Select schema: atm_data

Query 287

Completed, started on August 18, 2021 at 18:53:55
ELAPSED TIME: 00 m 06 s

- **Copying the data to dim_atm table**

copy atm_data.dim_atm from ' s3://etlprojectbysimran/dim_atm/part-00000-c4425605-e626-4cd2-adb2-cef68f7cb1b9-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;

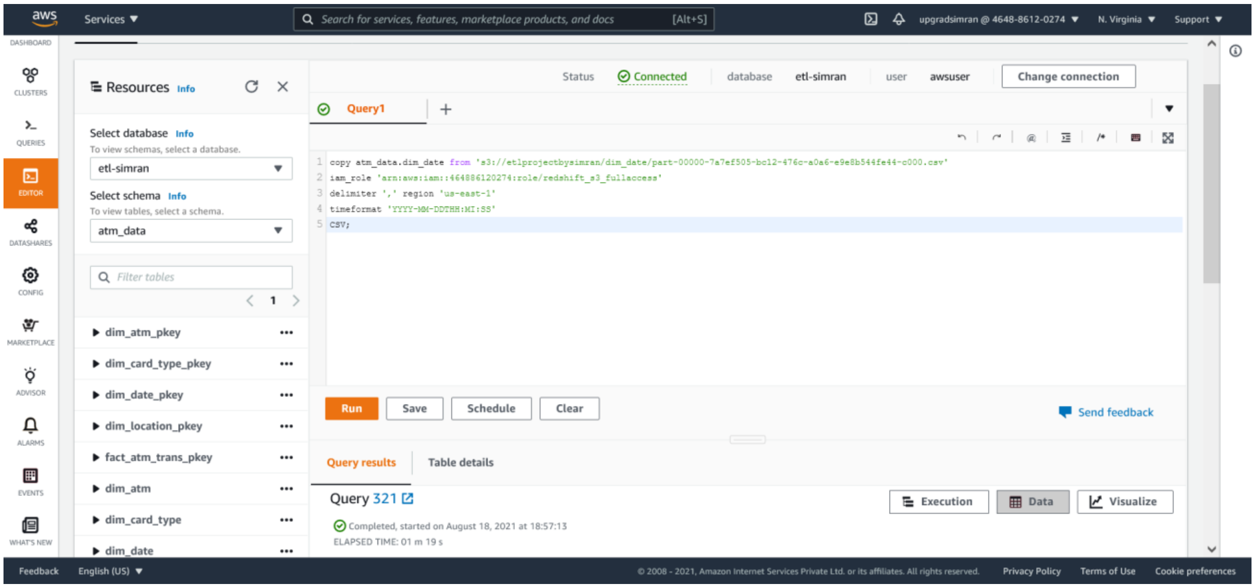


```
copy atm_data.dim_atm from 's3://etlprojectbysimran/dim_atm/part-00000-c4425605-e626-4cd2-adb2-cef68f7cb1b9-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 300
Completed, started on August 18, 2021 at 18:54:38
ELAPSED TIME: 00 m 10 s

- **Copying the data to dim_date table**

copy atm_data.dim_date from ' s3://etlprojectbysimran/dim_date/part-00000-7a7ef505-bc12-476c-a0a6-e9e8b544fe44-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;



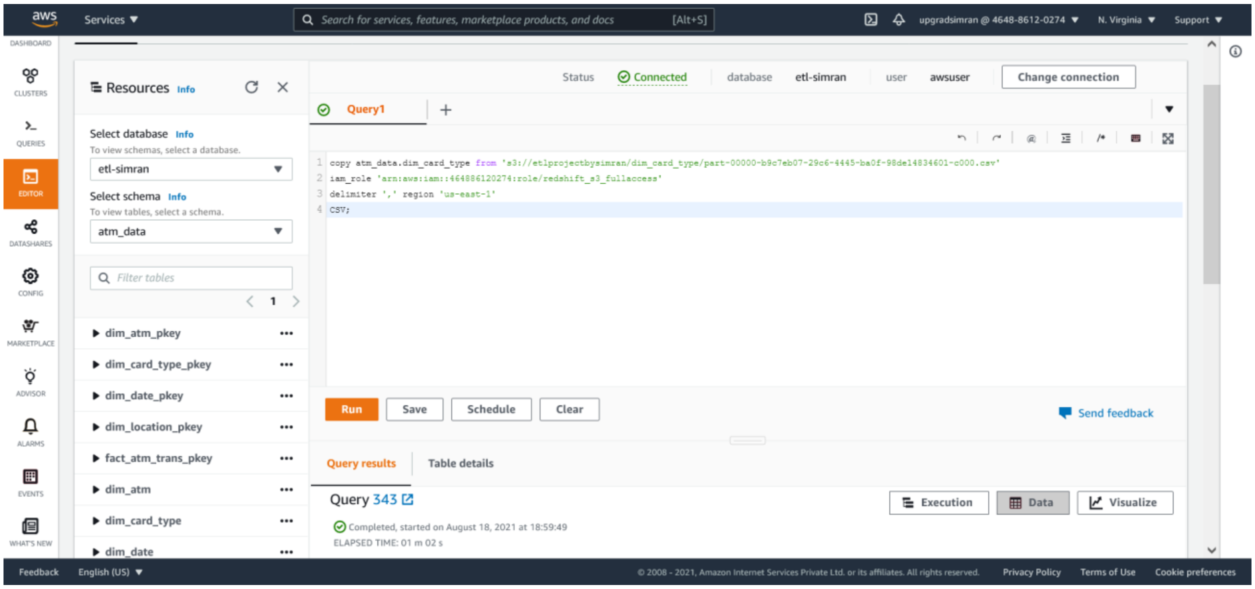
```
copy atm_data.dim_date from 's3://etlprojectbysimran/dim_date/part-00000-7a7ef505-bc12-476c-a0a6-e9e8b544fe44-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```

Query 321
Completed, started on August 18, 2021 at 18:57:13
ELAPSED TIME: 01 m 19 s

- **Copying the data to dim_card_type table**

copy atm_data.dim_card_type from ' s3://etlprojectbysimran/dim_card_type/part-00000-b9c7eb07-29c6-4445-ba0f-98de14834601-c000.csv'

iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;



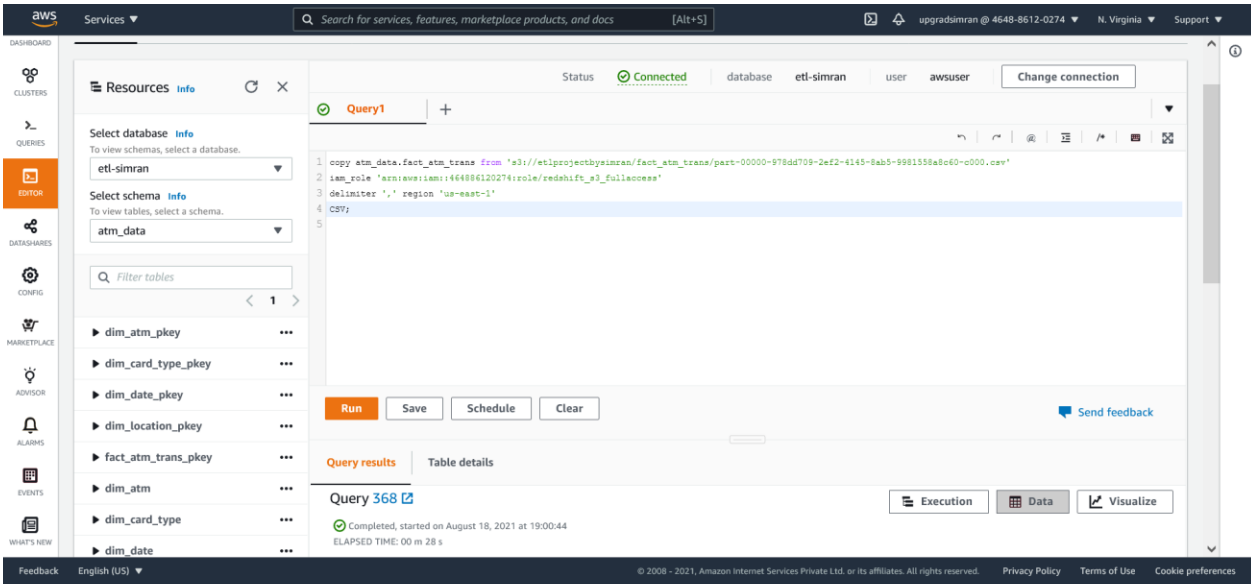
```
copy atm_data.dim_card_type from 's3://etlprojectbysimran/dim_card_type/part-00000-b9c7eb07-29c6-4445-ba0f-98de14834601-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 343

Completed, started on August 18, 2021 at 18:59:49
ELAPSED TIME: 01 m 02 s

- **Copying the data to fact_atm_trans table**

copy atm_data.fact_atm_trans from ' s3://etlprojectbysimran/fact_atm_trans/part-00000-978dd709-2ef2-4145-8ab5-9981558a8c60-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;



```
copy atm_data.fact_atm_trans from 's3://etlprojectbysimran/fact_atm_trans/part-00000-978dd709-2ef2-4145-8ab5-9981558a8c60-c000.csv'
iam_role 'arn:aws:iam::464886120274:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 368

Completed, started on August 18, 2021 at 19:00:44
ELAPSED TIME: 00 m 28 s