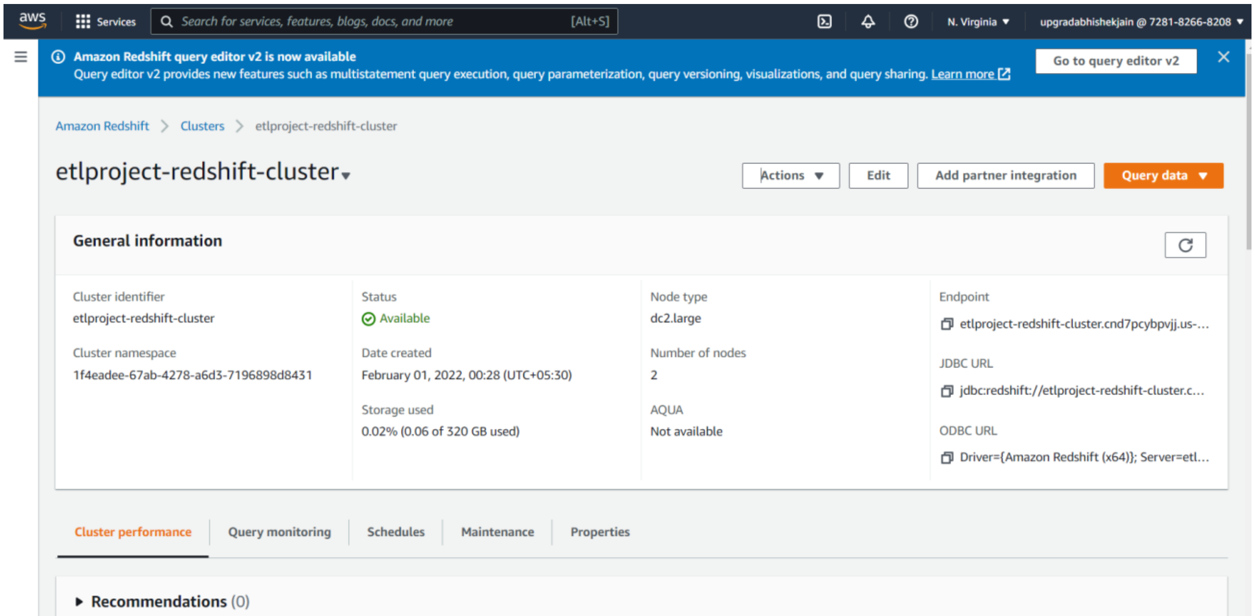
Creation of a Redshift Cluster

Screenshots of the configuration of the Redshift cluster that you have created:
Type of machine : dc2.large



Amazon Redshift query editor v2 is now available
Query editor v2 provides new features such as multistatement query execution, query parameterization, query versioning, visualizations, and query sharing. Learn more

etlproject-redshift-cluster

General information

Cluster identifier
etlproject-redshift-cluster

Cluster namespace
1f4eadee-67ab-4278-a6d3-7196898d8431

Status
Available

Date created
February 01, 2022, 00:28 (UTC+05:30)

Storage used
0.02% (0.06 of 320 GB used)

Node type
dc2.large

Number of nodes
2

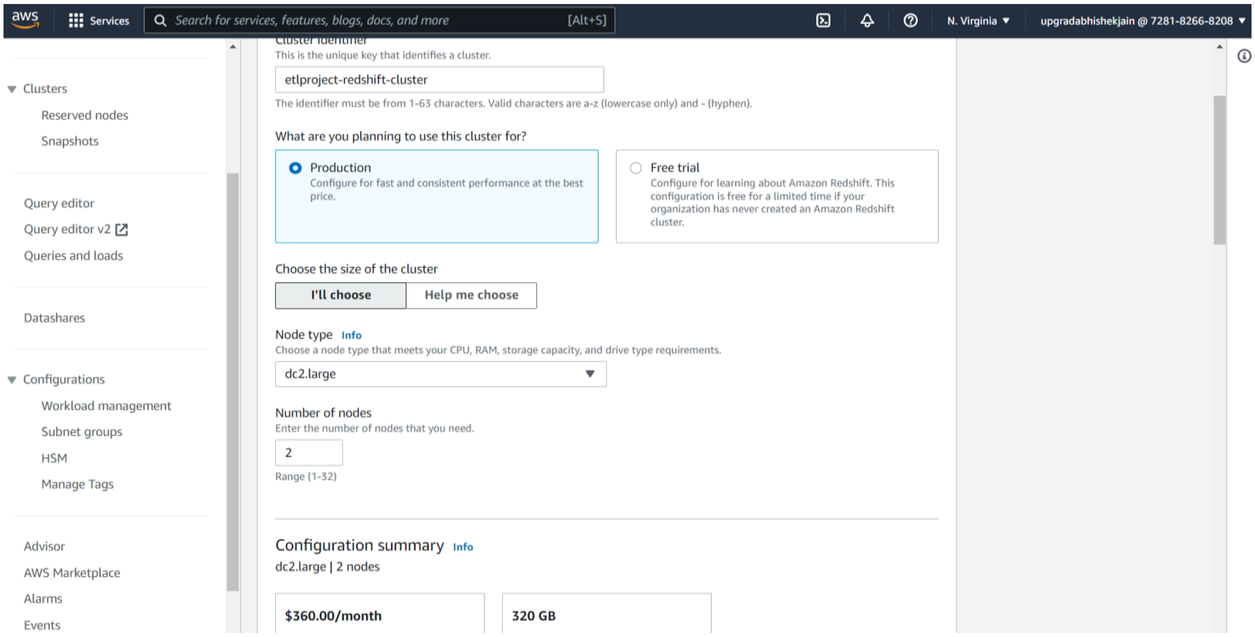
AQUA
Not available

Endpoint
etlproject-redshift-cluster.cnd7pcybpvjj.us-...

JDBC URL
jdbc:redshift://etlproject-redshift-cluster.c...

ODBC URL
Driver={Amazon Redshift (x64)}; Server=etl...

Screenshot of various configurations while setting up Red shift cluster
Cluster use and node type



Cluster identifier
This is the unique key that identifies a cluster.
etlproject-redshift-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. This 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

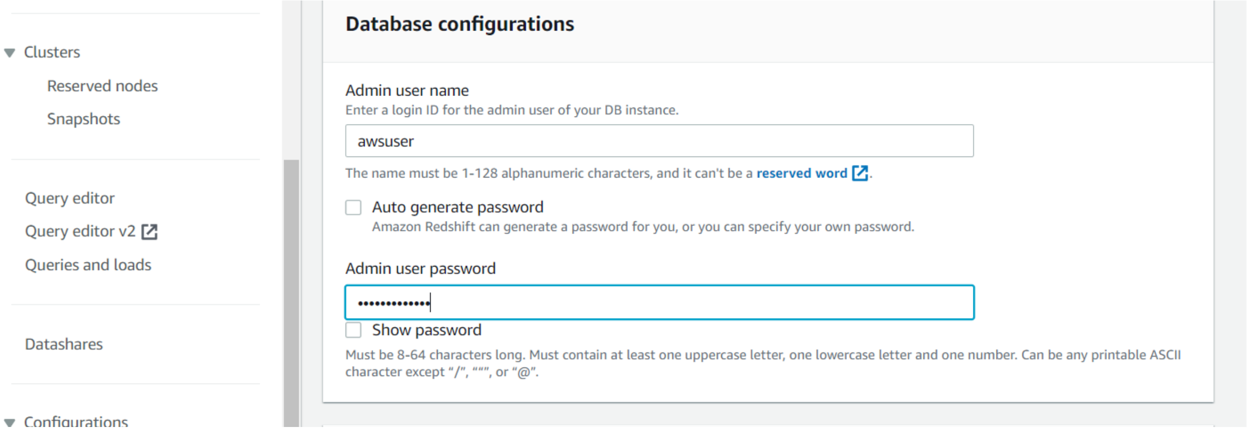
Node type
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.
dc2.large

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

Configuration summary
dc2.large | 2 nodes

$360.00/month
320 GB

Database Configurations



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.

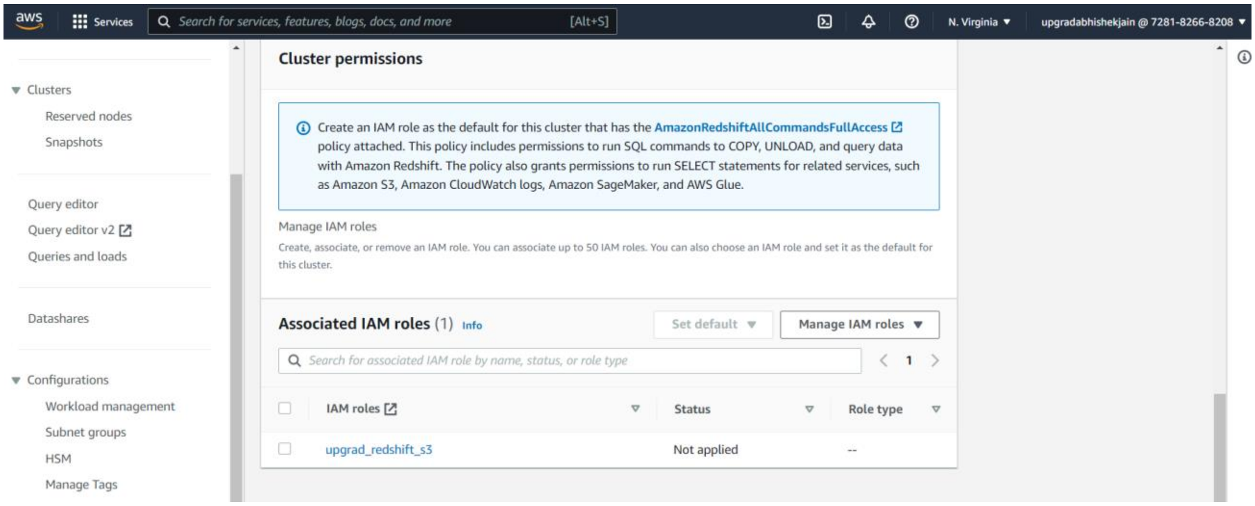
Admin user password

•••••••••••••

□ 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 "@".

IAM Roles



Cluster permissions

Create an IAM role as the default for this cluster that has the AmazonRedshiftAllCommandsFullAccess policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Manage IAM roles

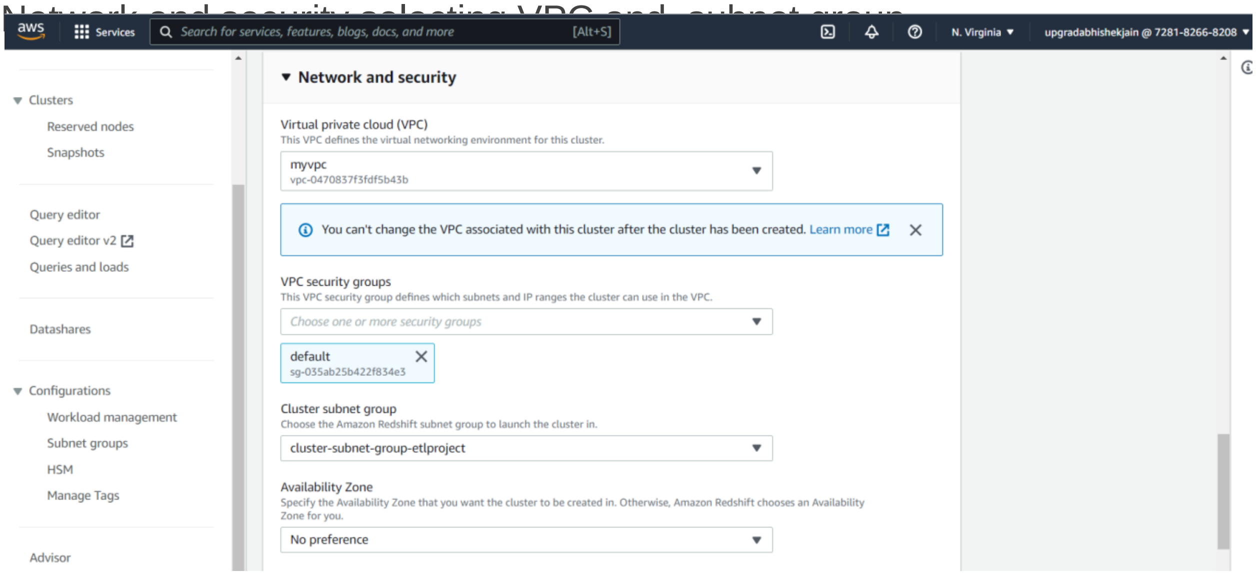
Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

Associated IAM roles (1) Info

Set default
Manage IAM roles
Search for associated IAM role by name, status, or role type
1

□	IAM roles	Status	Role type
□	upgrad_redshift_s3	Not applied	--

Network and security selecting VPC and subnet group



▼ Network and security

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.
myvpc
vpc-0470837f3fdf5b43b

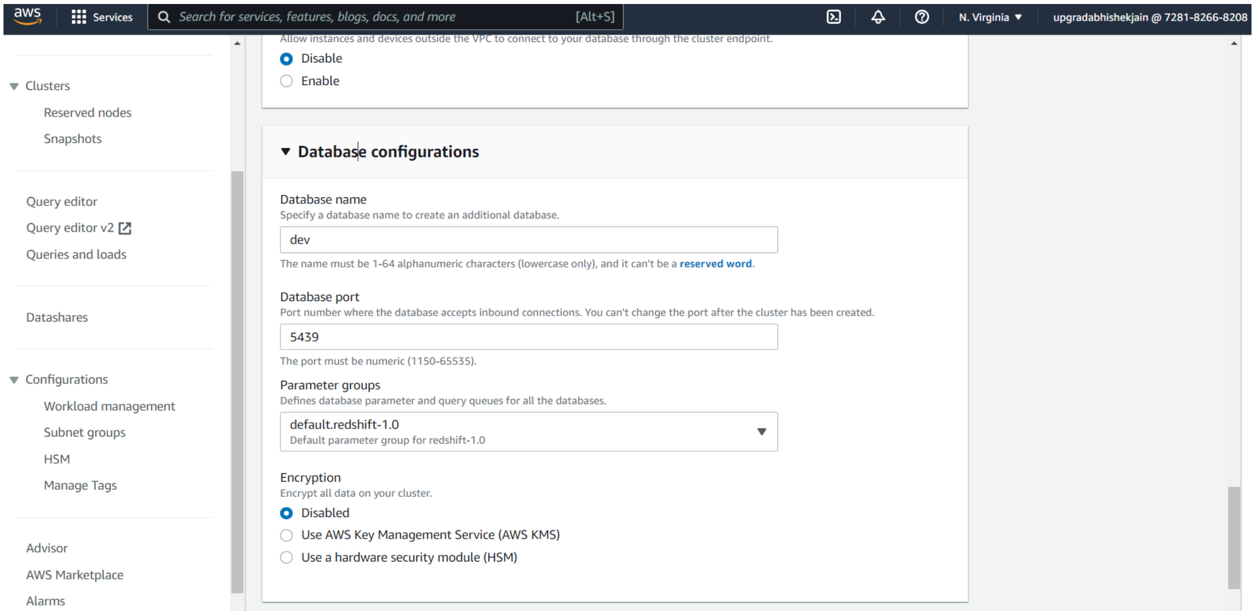
You can't change the VPC associated with this cluster after the cluster has been created. Learn more

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-035ab25b422f834e3

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

Availability Zone
Specify the Availability Zone that you want the cluster to be created in. Otherwise, Amazon Redshift chooses an Availability Zone for you.
No preference

Configuring DB port



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.

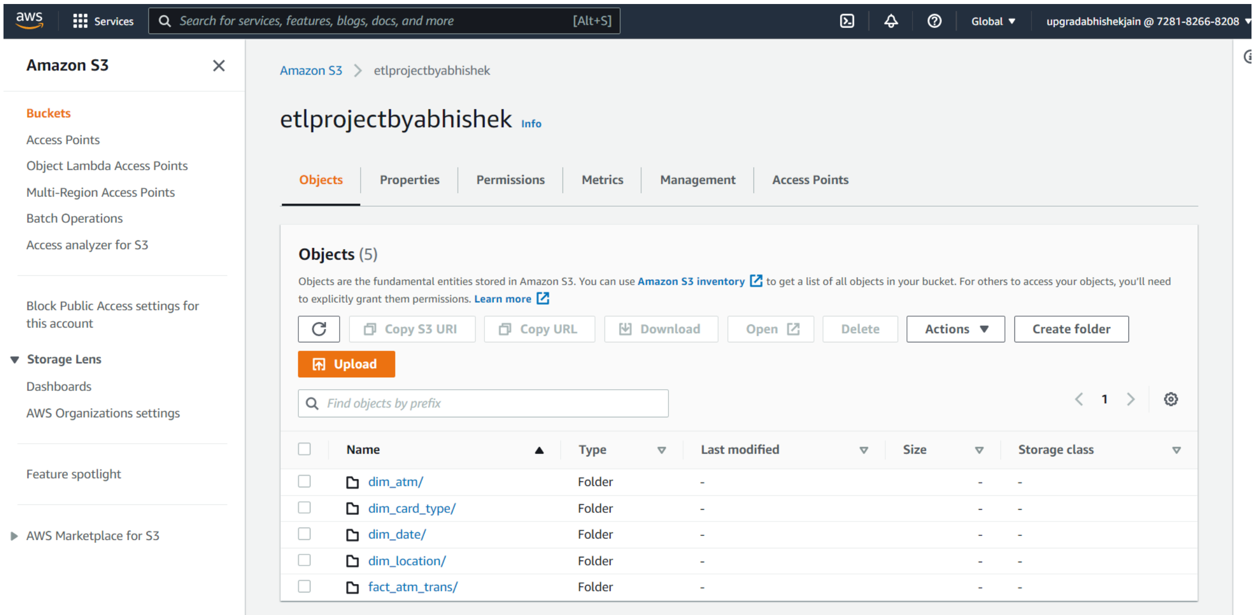
Database port
Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.
5439
The port must be numeric (1150-65535).

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)

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

View all the data in Amazon S3 Bucket:



Amazon S3 > etlprojectbyabhishek

etlprojectbyabhishek Info

Objects (5)

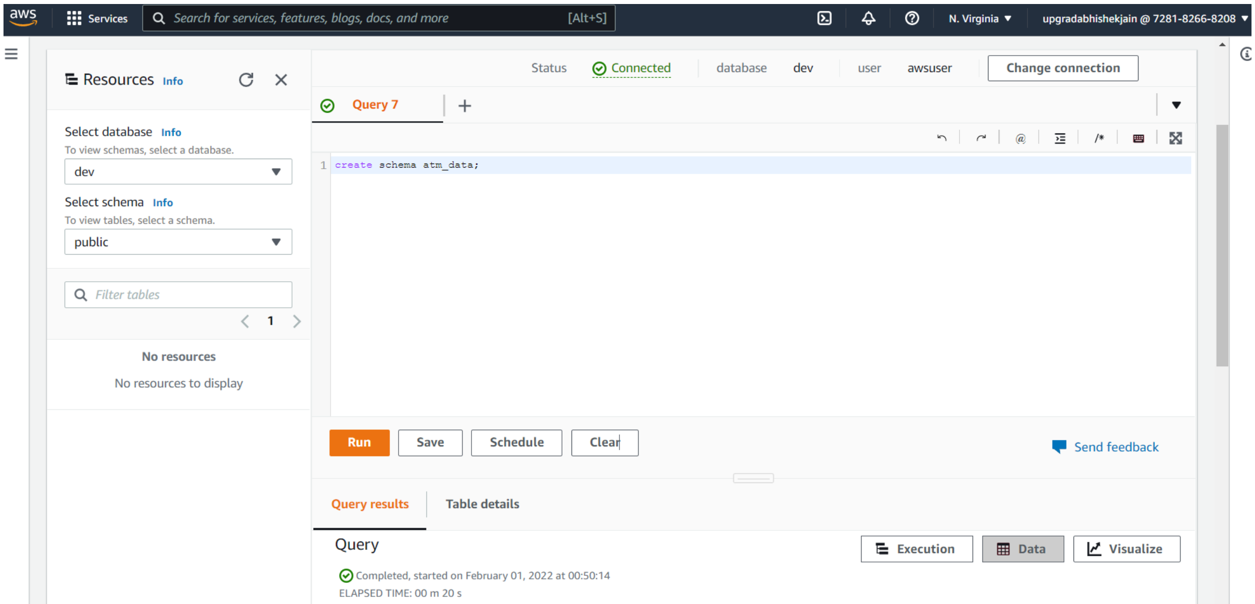
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

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

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

Create a schema for dimension and fact table

- **create schema atm_data;**



Status Connected database dev user awsuser

Select database: dev

Select schema: public

No resources

No resources to display

```
create schema atm_data;
```

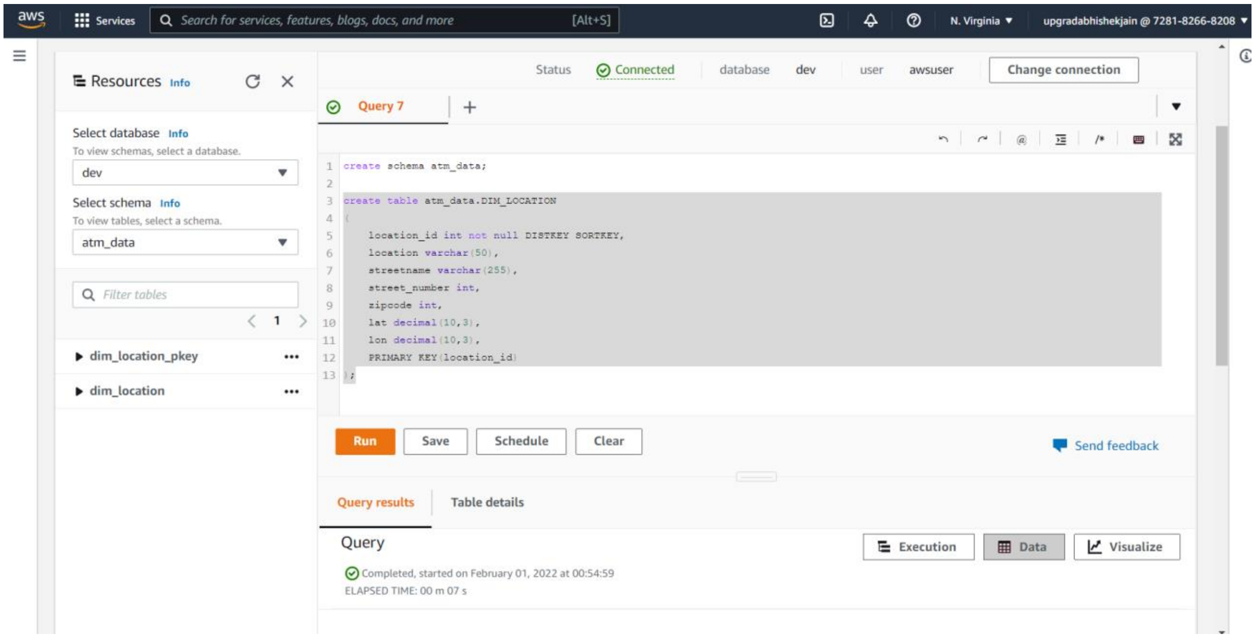
Query

Completed, started on February 01, 2022 at 00:50:14

ELAPSED TIME: 00 m 20 s

- **Create Dim_location 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)
);
```



```
create 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)
);
```

Status: Connected | database: dev | user: awsuser

Select database: dev

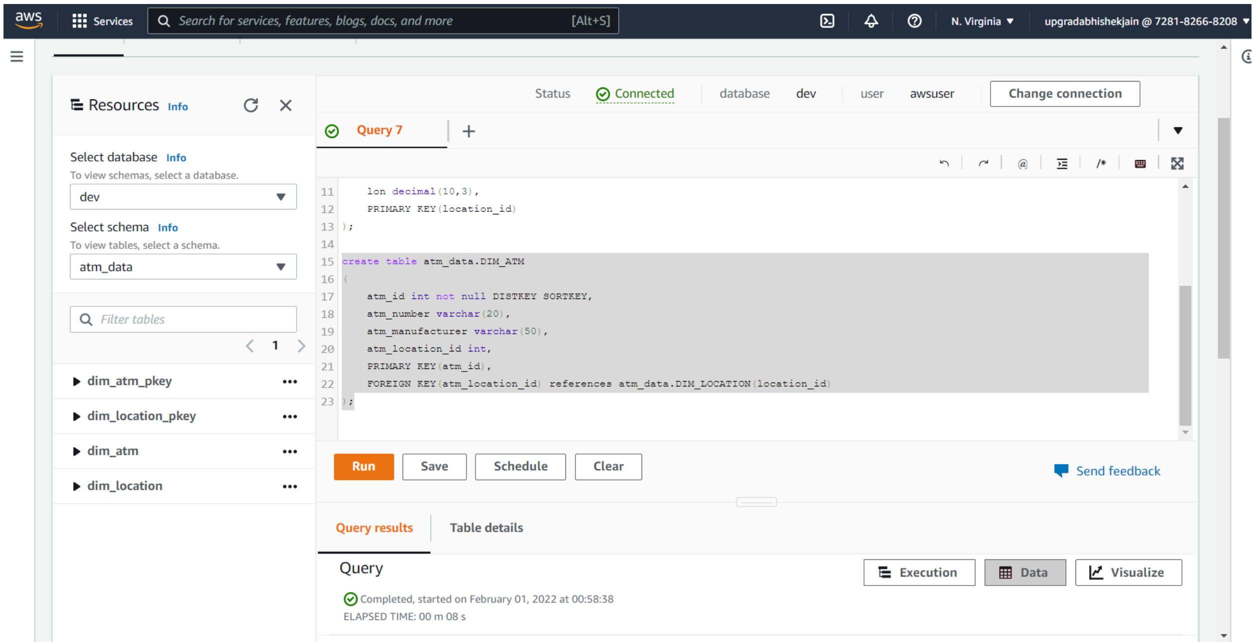
Select schema: atm_data

Query

Completed, started on February 01, 2022 at 00:54:59
ELAPSED TIME: 00 m 07 s

- **Create Dim_ATM 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 dev user awsuser Change connection

Select database: dev (To view schemas, select a database.)

Select schema: atm_data (To view tables, select a schema.)

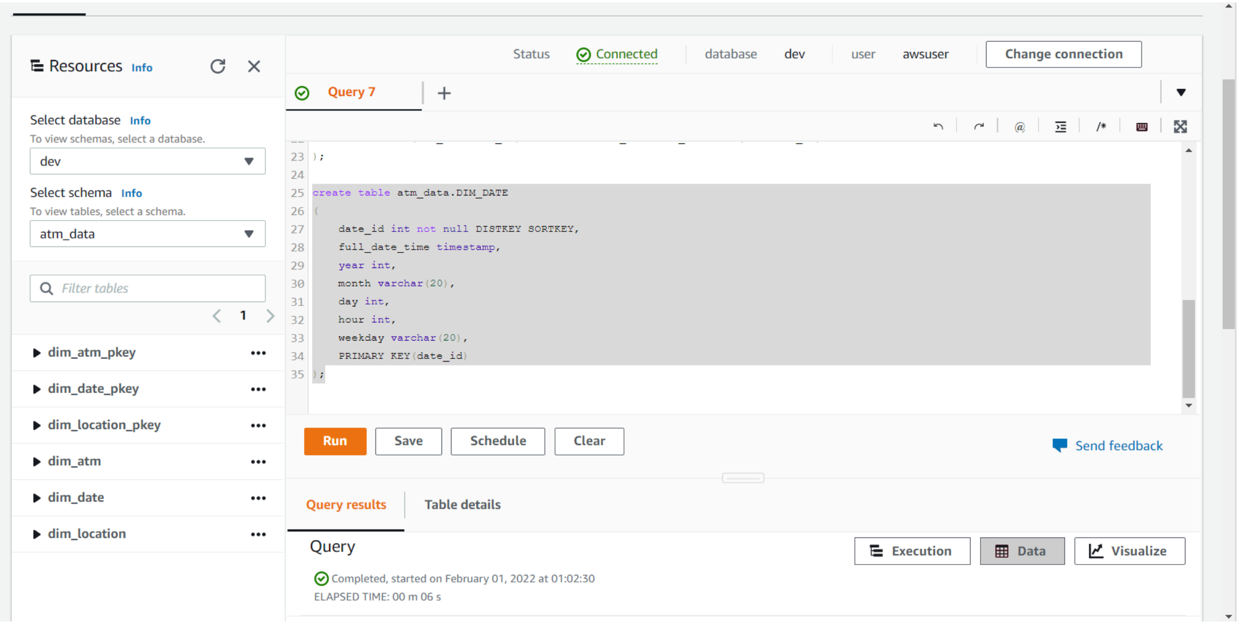
```
    lon decimal(10,3),
    PRIMARY KEY(location_id)
);

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 February 01, 2022 at 00:58:38
ELAPSED TIME: 00 m 08 s

- **Create Dimension date 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)
);
```



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

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

Filter tables

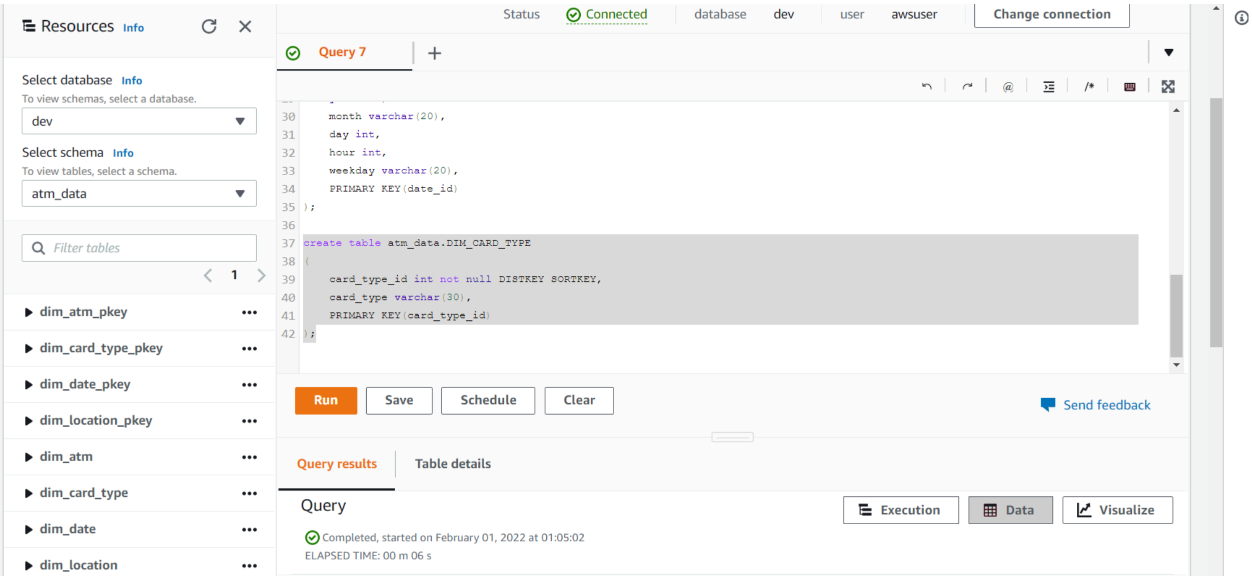
- dim_atm_pkey
- dim_date_pkey
- dim_location_pkey
- dim_atm
- dim_date
- dim_location

```
);

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 February 01, 2022 at 01:02:30
ELAPSED TIME: 00 m 06 s

- **Create 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)
  );



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

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

Filter tables

- dim_atm_pkey
- dim_card_type_pkey
- dim_date_pkey
- dim_location_pkey
- dim_atm
- dim_card_type
- dim_date
- dim_location

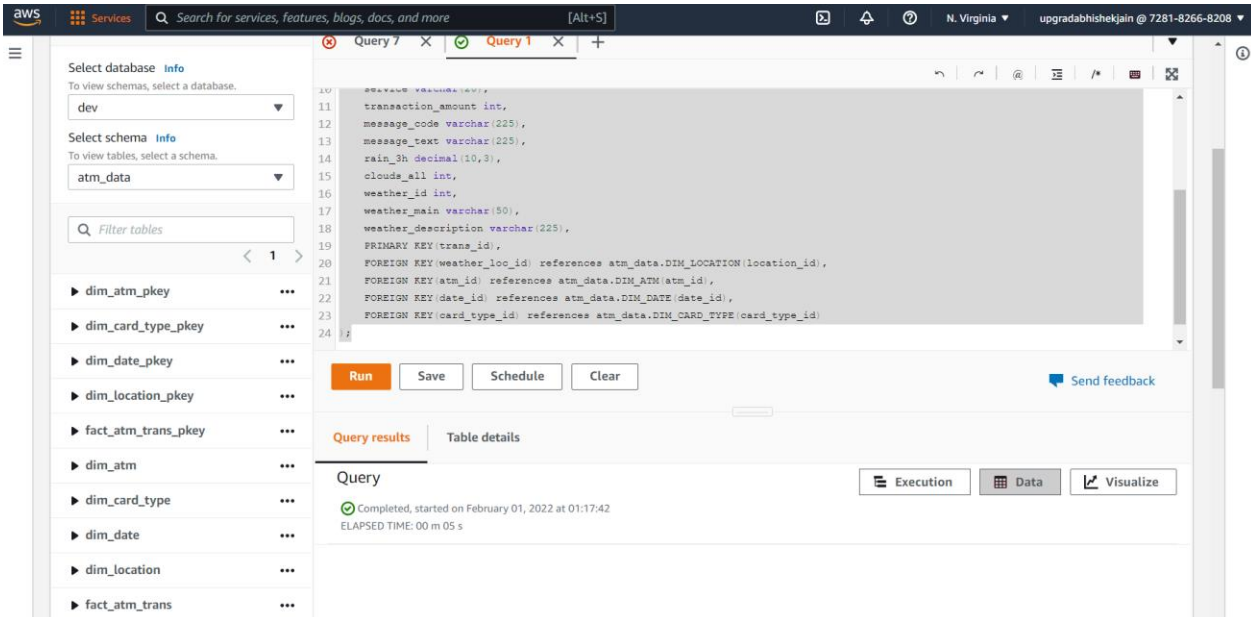
```
    month varchar(20),
    day int,
    hour int,
    weekday varchar(20),
    PRIMARY KEY(date_id)
);

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 February 01, 2022 at 01:05:02
ELAPSED TIME: 00 m 06 s

- Create atm transaction 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(225),
        PRIMARY KEY(trans_id),
        FOREIGN KEY(weather_loc_id) references
atm_data.DIM_LOCATION(location_id),
        FOREIGN KEY(atm_id) references atm_data.DIM_ATM(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)
);
```



```
  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(225),
  PRIMARY KEY(trans_id),
  FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
  FOREIGN KEY(atm_id) references atm_data.DIM_ATM(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)
);
```

Query

Completed, started on February 01, 2022 at 01:17:42
ELAPSED TIME: 00 m 05 s

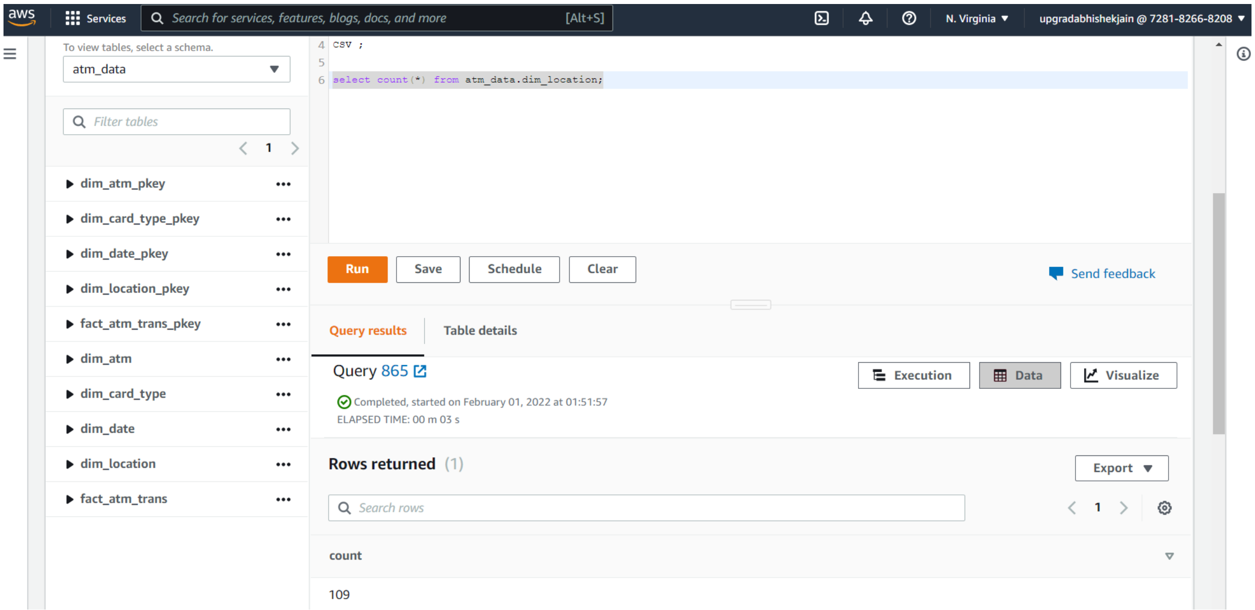
Loading data into a Redshift cluster from Amazon S3 bucket

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

- **Copy data to DIM_LOCATION and DIM_ATM table**

copy atm_data.dim_location from 's3://etlprojectbyabhishek/dim_location/part-00000-2c3f4932-5d7f-4a0f-b030-9b3283149f77-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
CSV;

Count of record : 109



```
csv ;

select count(*) from atm_data.dim_location;
```

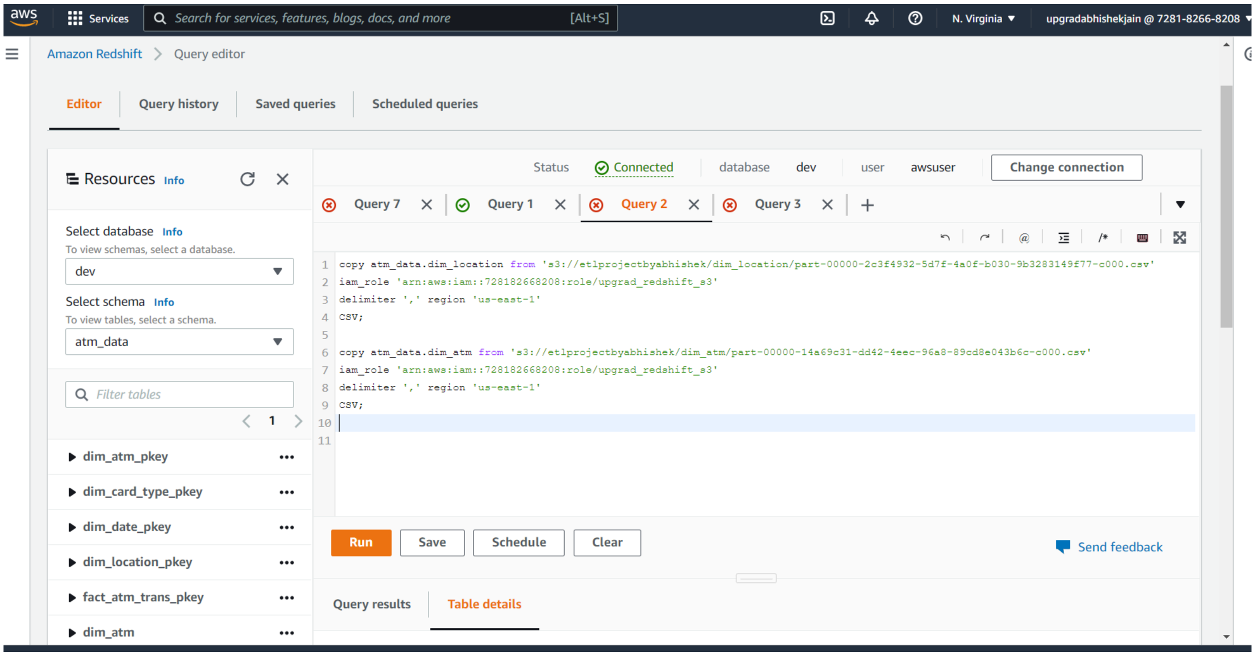
Query 865

Completed, started on February 01, 2022 at 01:51:57
ELAPSED TIME: 00 m 03 s

Rows returned (1)

count
109

copy atm_data.dim_atm from 's3://etlprojectbyabhishek/dim_atm/part-00000-14a69c31-dd42-4eec-96a8-89cd8e043b6c-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
CSV;



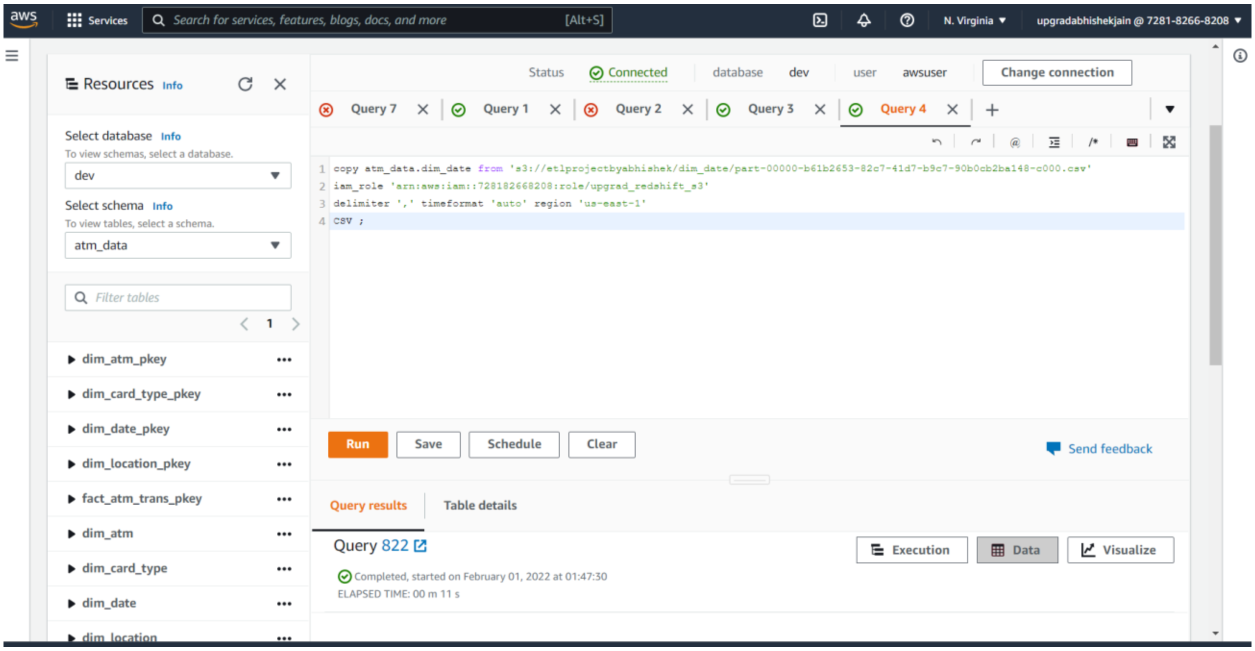
Status: Connected | database: dev | user: awsuser

Select database: dev

Select schema: atm_data

```
copy atm_data.dim_location from 's3://etlprojectbyabhishek/dim_location/part-00000-2c3f4932-5d7f-4a0f-b030-9b3283149f77-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
csv;

copy atm_data.dim_atm from 's3://etlprojectbyabhishek/dim_atm/part-00000-14a69c31-dd42-4eec-96a8-89cd8e043b6c-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
csv;
```

- **Copy data to DIM_DATE table**
  copy atm_data.dim_date from 's3://etlprojectbyabhishek/dim_date/part-00000-b61b2653-82c7-41d7-b9c7-90b0cb2ba148-c000.csv'
  iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
  delimiter ',' timeformat 'auto' region 'us-east-1'
  CSV ;



Status: Connected | database: dev | user: awsuser

Select database: dev

Select schema: atm_data

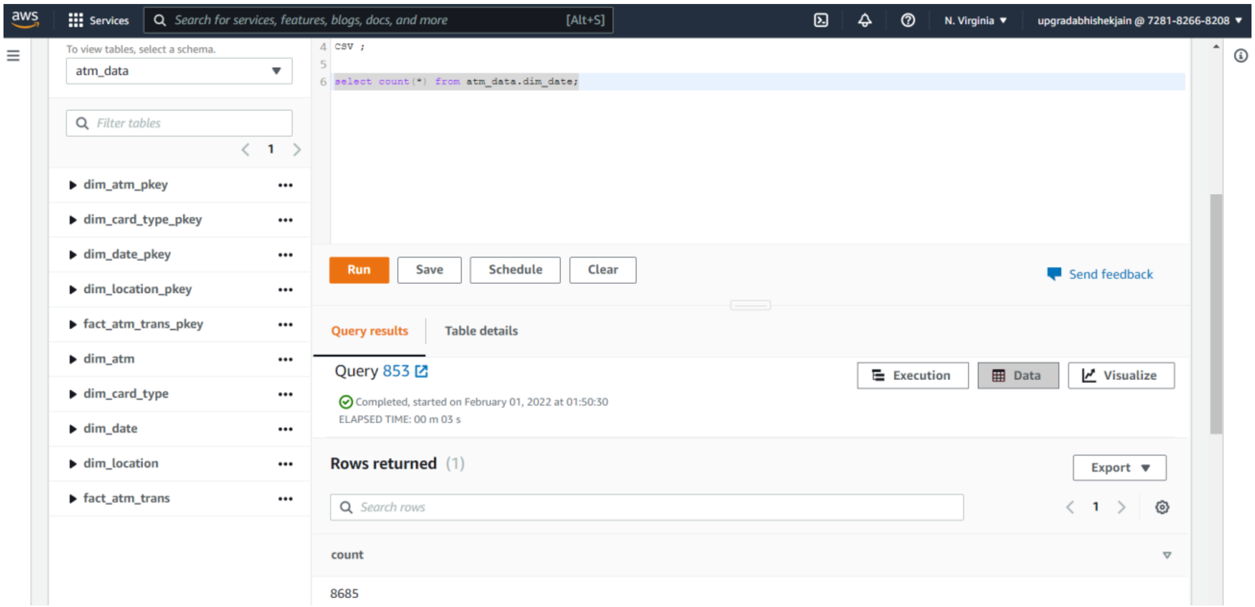
```
copy atm_data.dim_date from 's3://etlprojectbyabhishek/dim_date/part-00000-b61b2653-82c7-41d7-b9c7-90b0cb2ba148-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' timeformat 'auto' region 'us-east-1'
csv ;
```

Query 822

Completed, started on February 01, 2022 at 01:47:30

ELAPSED TIME: 00 m 11 s

Count of record :8685



```
CSV ;

select count(*) from atm_data.dim_date;
```

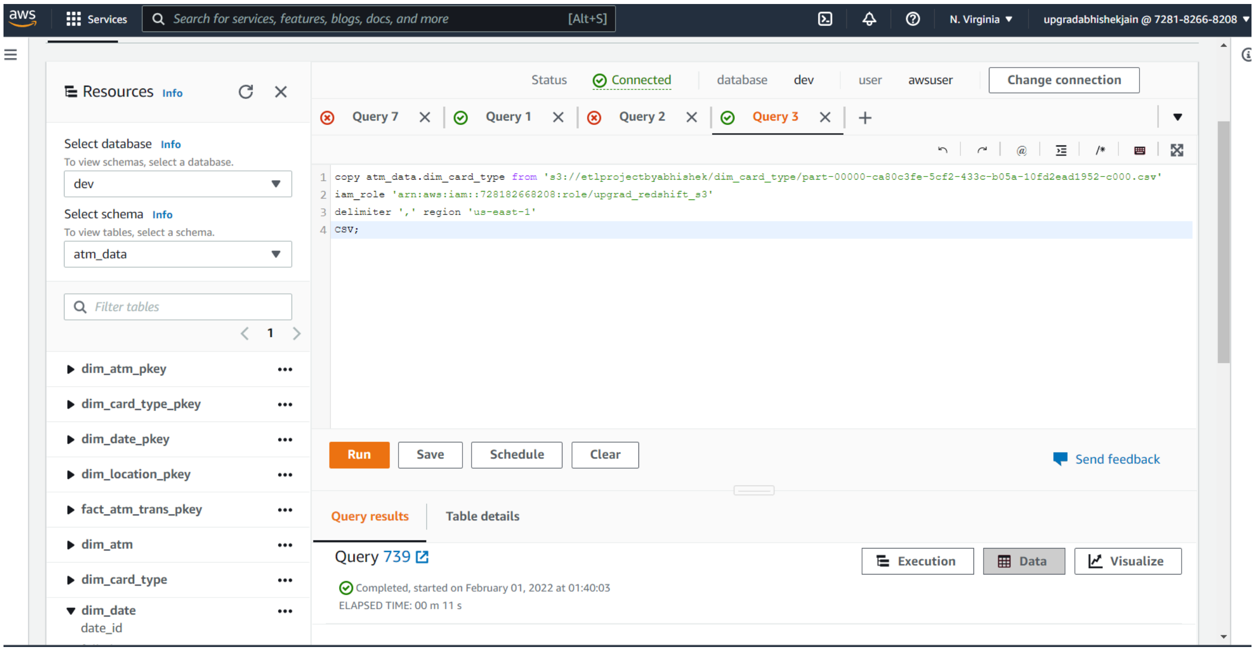
Query 853

Completed, started on February 01, 2022 at 01:50:30
ELAPSED TIME: 00 m 03 s

Rows returned (1)

count
8685

- **Copy data to DIM_CARD_TYPE table**
  copy atm_data.dim_card_type from 's3://etlprojectbyabhishek/dim_card_type/part-00000-ca80c3fe-5cf2-433c-b05a-10fd2ead1952-c000.csv'
  iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
  delimiter ',' region 'us-east-1'
  CSV;



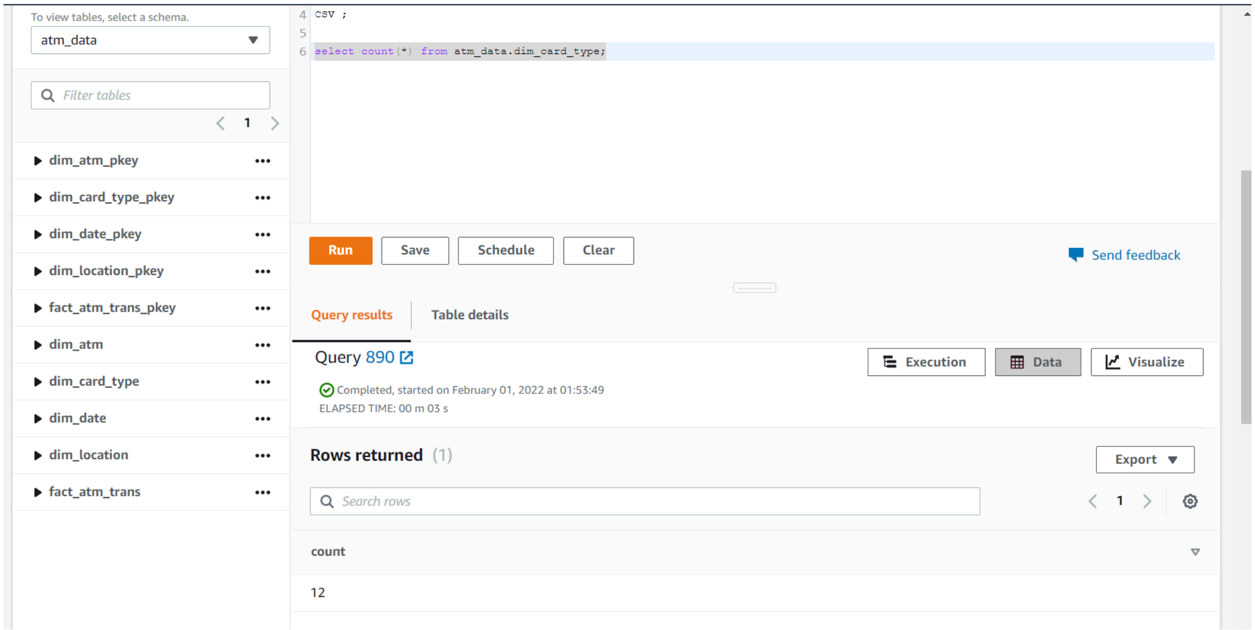
Status Connected database dev user awsuser

```
copy atm_data.dim_card_type from 's3://etlprojectbyabhishek/dim_card_type/part-00000-ca80c3fe-5cf2-433c-b05a-10fd2ead1952-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
CSV;
```

Query 739

Completed, started on February 01, 2022 at 01:40:03
ELAPSED TIME: 00 m 11 s

Count of record :12



To view tables, select a schema.

atm_data

```
csv ;

select count(*) from atm_data.dim_card_type;
```

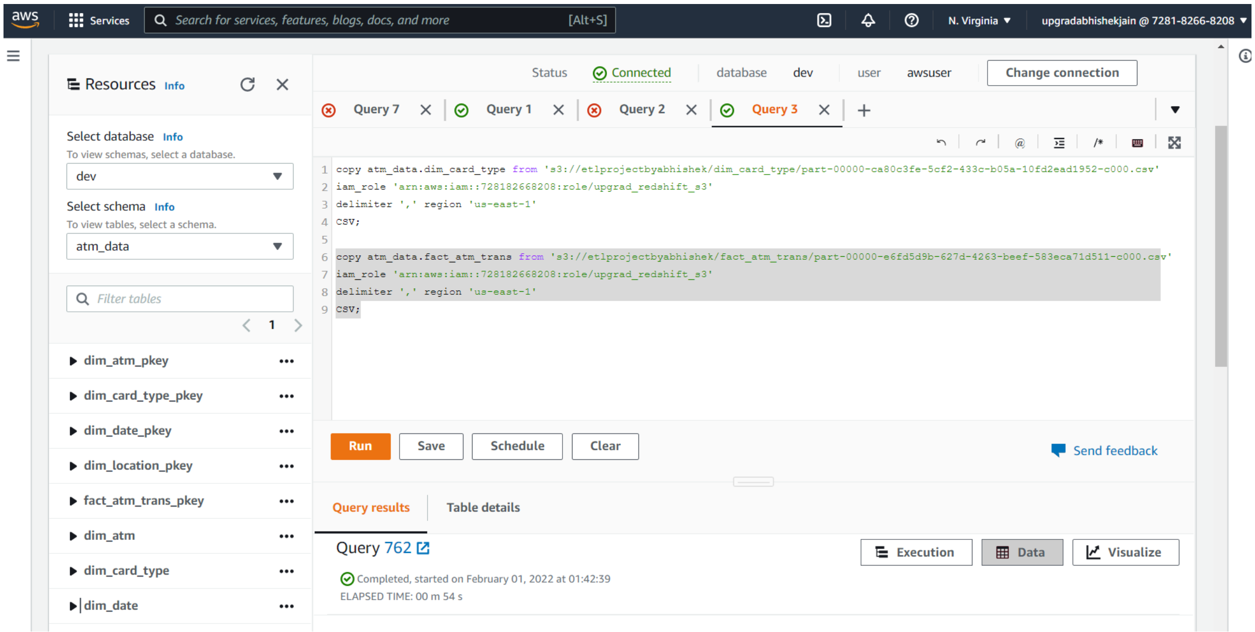
Query 890

Completed, started on February 01, 2022 at 01:53:49
ELAPSED TIME: 00 m 03 s

Rows returned (1)

count
12

- **Copy data to FACT_ATM_TRANSACTION table**
  copy atm_data.fact_atm_trans from 's3://etlprojectbyabhishek/fact_atm_trans/part-00000-e6fd5d9b-627d-4263-beef-583eca71d511-c000.csv'
  iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
  delimiter ',' region 'us-east-1'
  CSV;



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

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

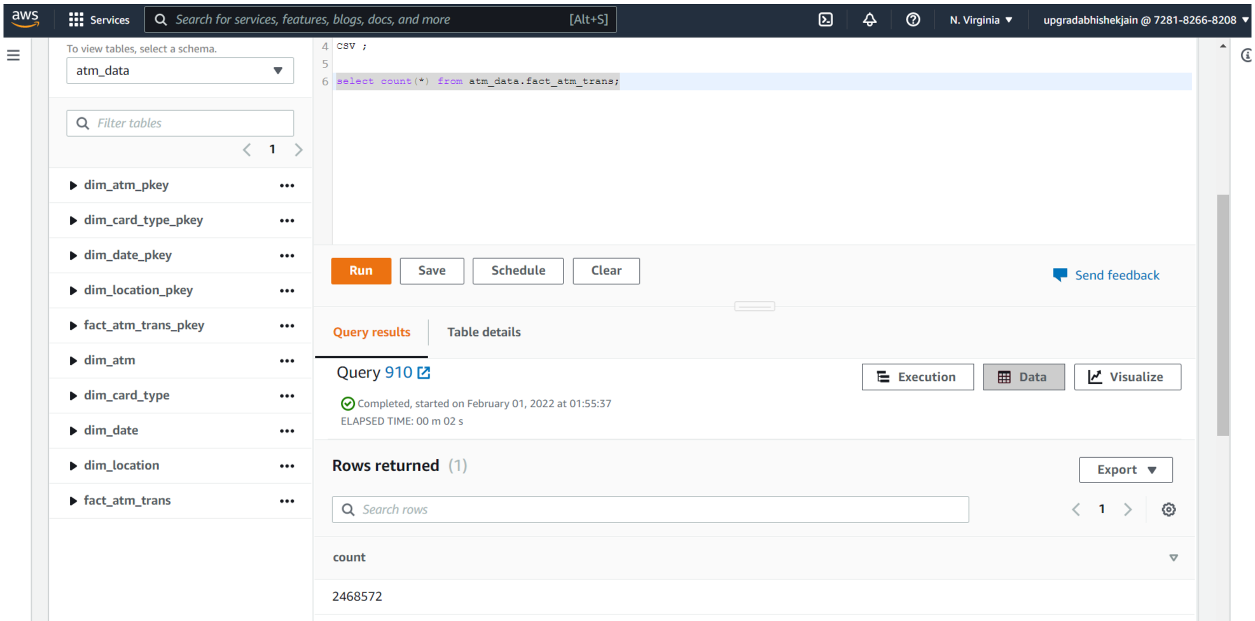
Status Connected database dev user awsuser

```
copy atm_data.dim_card_type from 's3://etlprojectbyabhishek/dim_card_type/part-00000-ca80c3fe-5cf2-433c-b05a-10fd2ead1952-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
csv;

copy atm_data.fact_atm_trans from 's3://etlprojectbyabhishek/fact_atm_trans/part-00000-e6fd5d9b-627d-4263-beef-583eca71d511-c000.csv'
iam_role 'arn:aws:iam::728182668208:role/upgrad_redshift_s3'
delimiter ',' region 'us-east-1'
csv;
```

Query 762

Completed, started on February 01, 2022 at 01:42:39
ELAPSED TIME: 00 m 54 s

Count of record : 2468572



```
csv ;

select count(*) from atm_data.fact_atm_trans;
```

Query 910

Completed, started on February 01, 2022 at 01:55:37
ELAPSED TIME: 00 m 02 s

Rows returned (1)

count
2468572