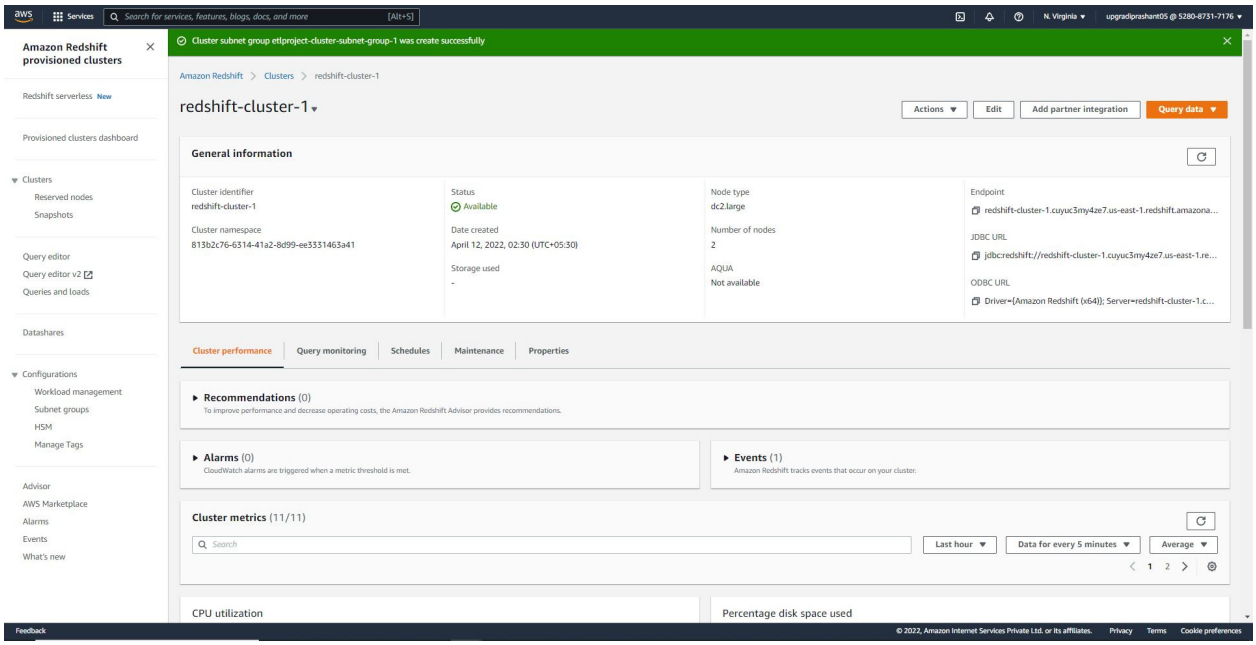
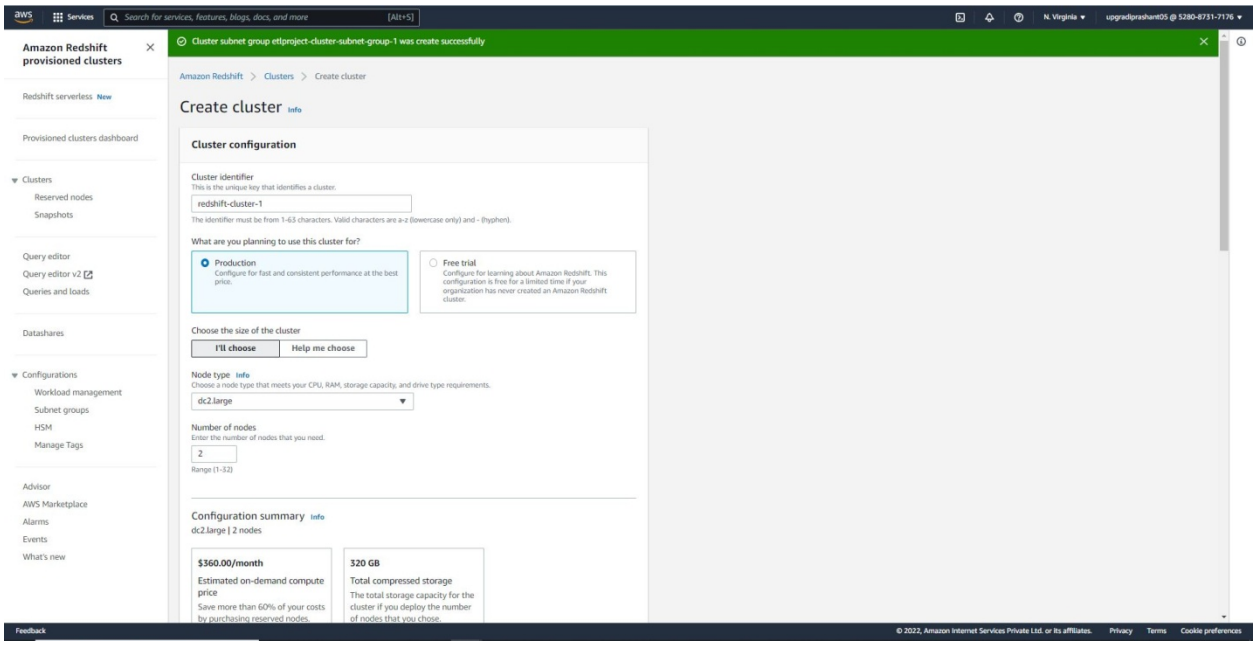
Creation of a Redshift Cluster

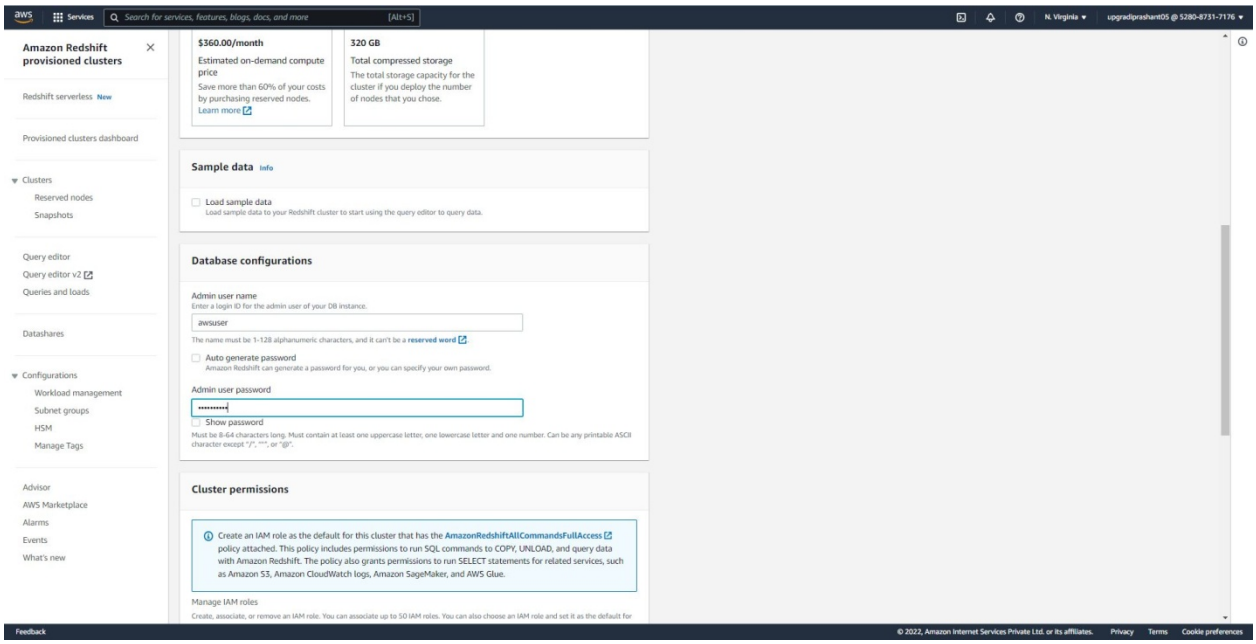
Screenshots of the configuration of the Redshift cluster that I have created:

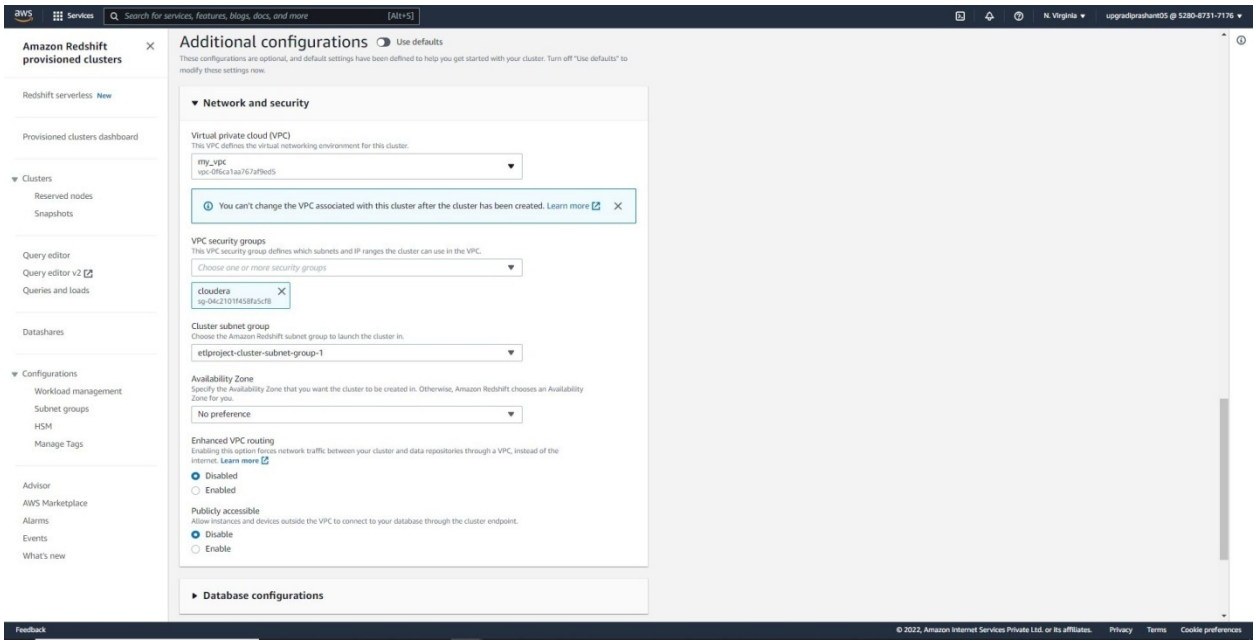
<Screenshot of the type of machine used along with number of nodes>

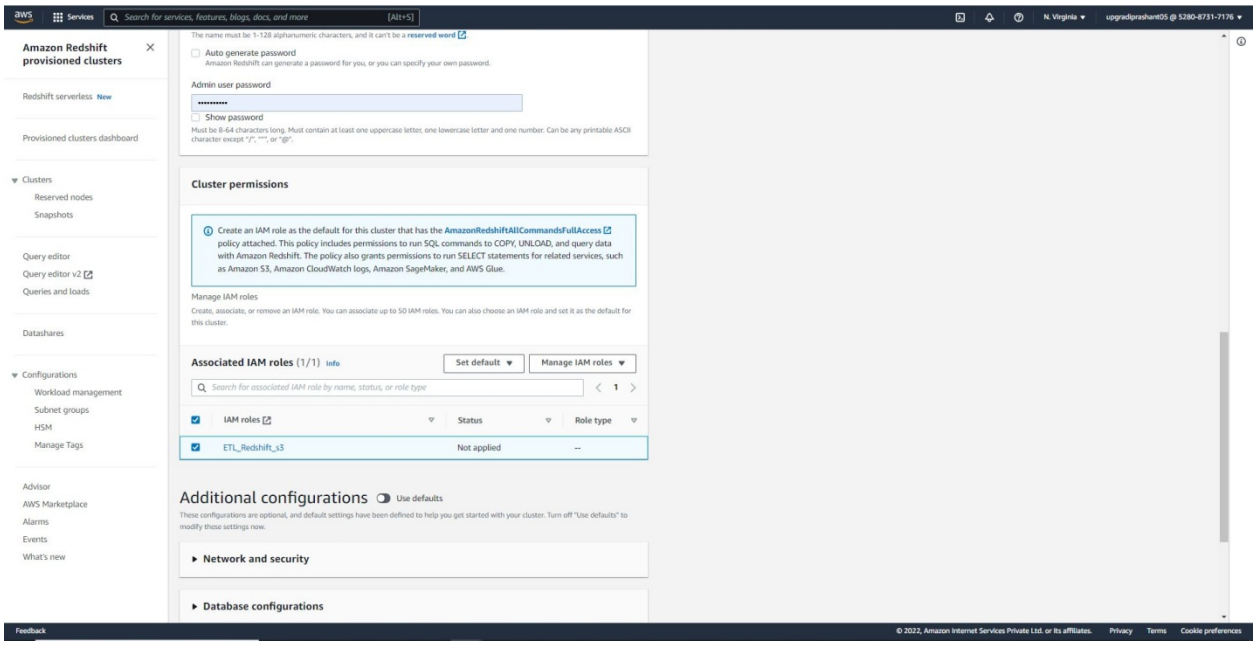


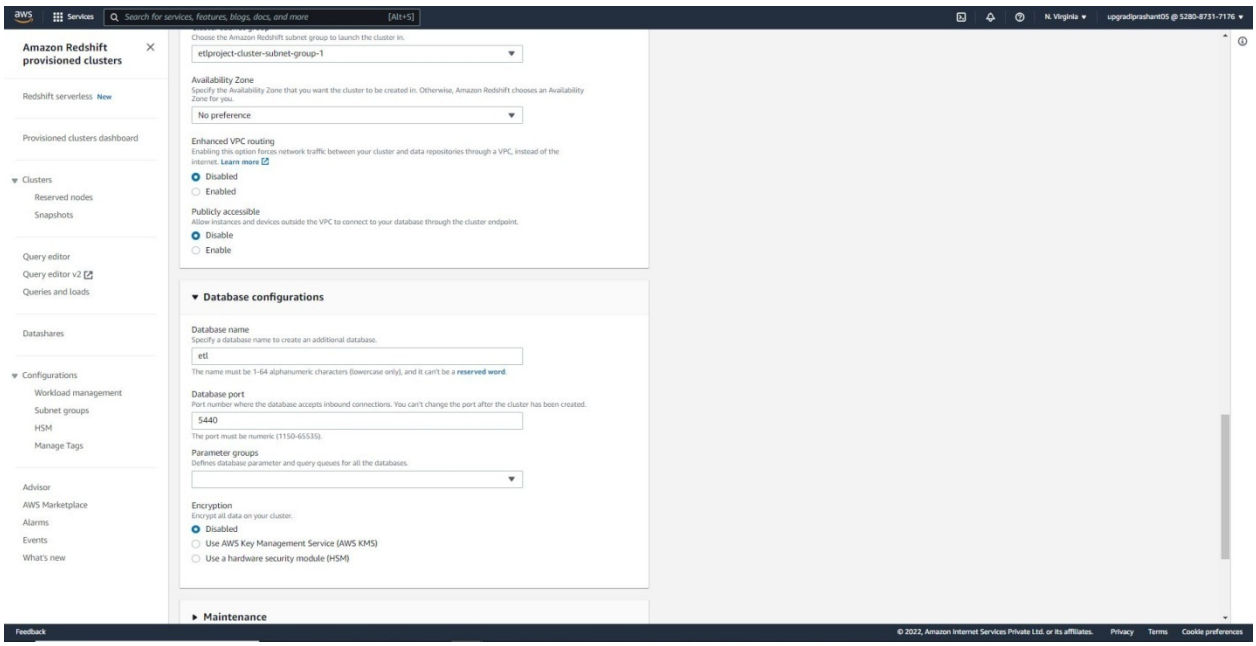
Screenshots of the various configurations associated with cluster creation:




Amazon Redshift provisioned clusters
Redshift serverless New
Provisioned clusters dashboard
Clusters
Reserved nodes
Snapshots
Query editor
Query editor v2
Queries and loads
Datashares
Configurations
Workload management
Subnet groups
HSM
Manage Tags
Advisor
AWS Marketplace
Alarms
Events
What's new
$360.00/month
Estimated on-demand compute price
Save more than 60% of your costs by purchasing reserved nodes.
Learn more
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.
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
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
Feedback
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

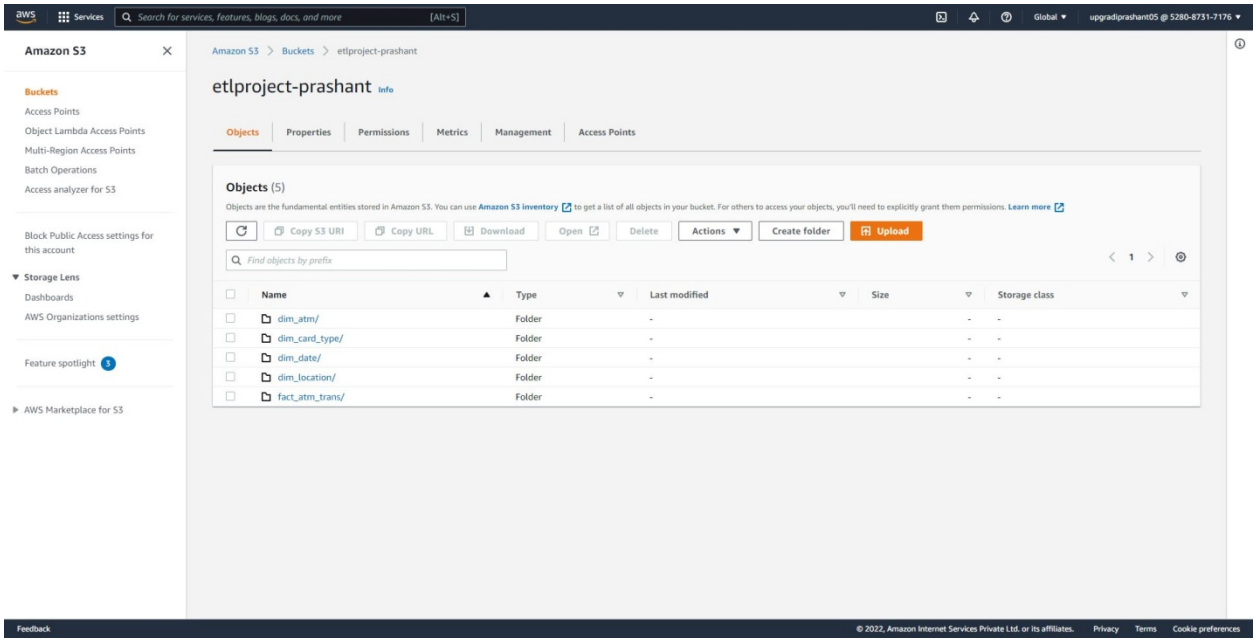

Amazon Redshift provisioned clusters
Redshift serverless New
Provisioned clusters dashboard
Clusters
Reserved nodes
Snapshots
Query editor
Query editor v2
Queries and loads
Datashares
Configurations
Workload management
Subnet groups
HSM
Manage Tags
Advisor
AWS Marketplace
Alarms
Events
What's new
Additional 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-0f6ca1aa767af9ed5
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-04c2101f458fa5cf8
Cluster subnet group
Choose the Amazon Redshift subnet group to launch the cluster in.
etlproject-cluster-subnet-group-1
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
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
Feedback
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences


Amazon Redshift provisioned clusters
Redshift serverless New
Provisioned clusters dashboard
Clusters
Reserved nodes
Snapshots
Query editor
Query editor v2
Queries and loads
Datashares
Configurations
Workload management
Subnet groups
HSM
Manage Tags
Advisor
AWS Marketplace
Alarms
Events
What's new
Auto generate password
Admin user password
Show password
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
Associated IAM roles (1/1) Info
Set default
Manage IAM roles
IAM roles
Status
Role type
ETL_Redshift_s3
Not applied
Additional configurations
Use defaults
Network and security
Database configurations
Feedback
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences


Amazon Redshift provisioned clusters
Redshift serverless New
Provisioned clusters dashboard
Clusters
Reserved nodes
Snapshots
Query editor
Query editor v2
Queries and loads
Datashares
Configurations
Workload management
Subnet groups
HSM
Manage Tags
Advisor
AWS Marketplace
Alarms
Events
What's new
etlproject-cluster-subnet-group-1
Availability Zone
No preference
Enhanced VPC routing
Disabled
Enabled
Publicly accessible
Disable
Enable
Database configurations
Database name
etl
Database port
5440
Parameter groups
Encryption
Disabled
Use AWS Key Management Service (AWS KMS)
Use a hardware security module (HSM)
Maintenance
Feedback
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

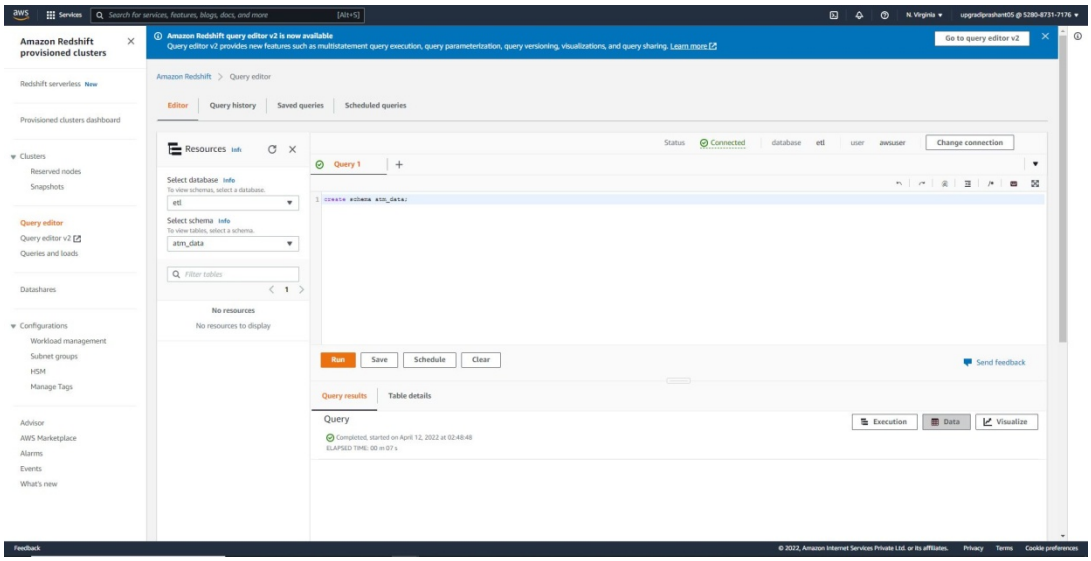
Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Viewing all the data in Amazon S3 bucket:



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

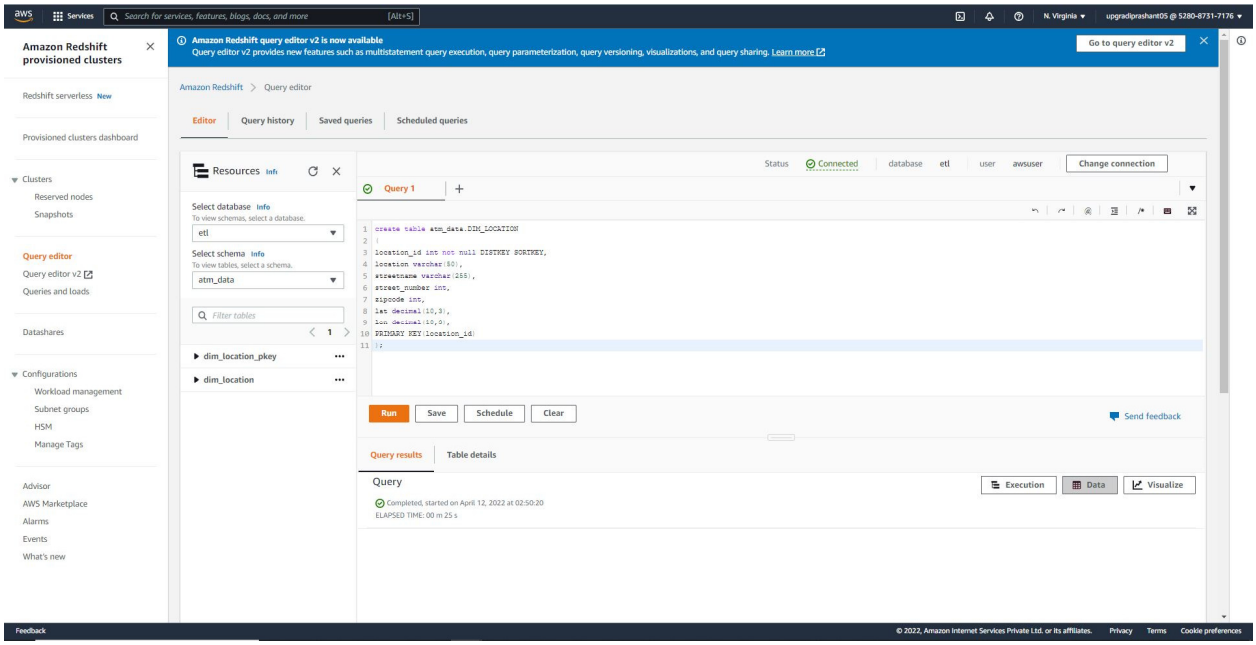
create schema atm_data;



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

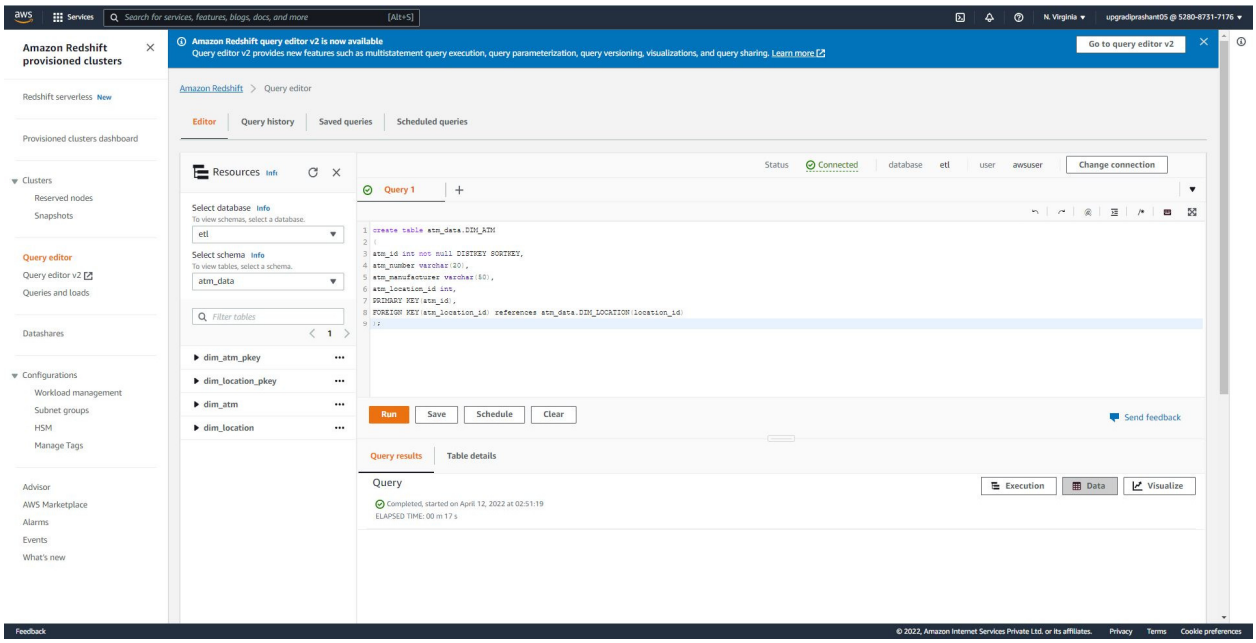
- **Creating location dimension table**

```
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)
);
```



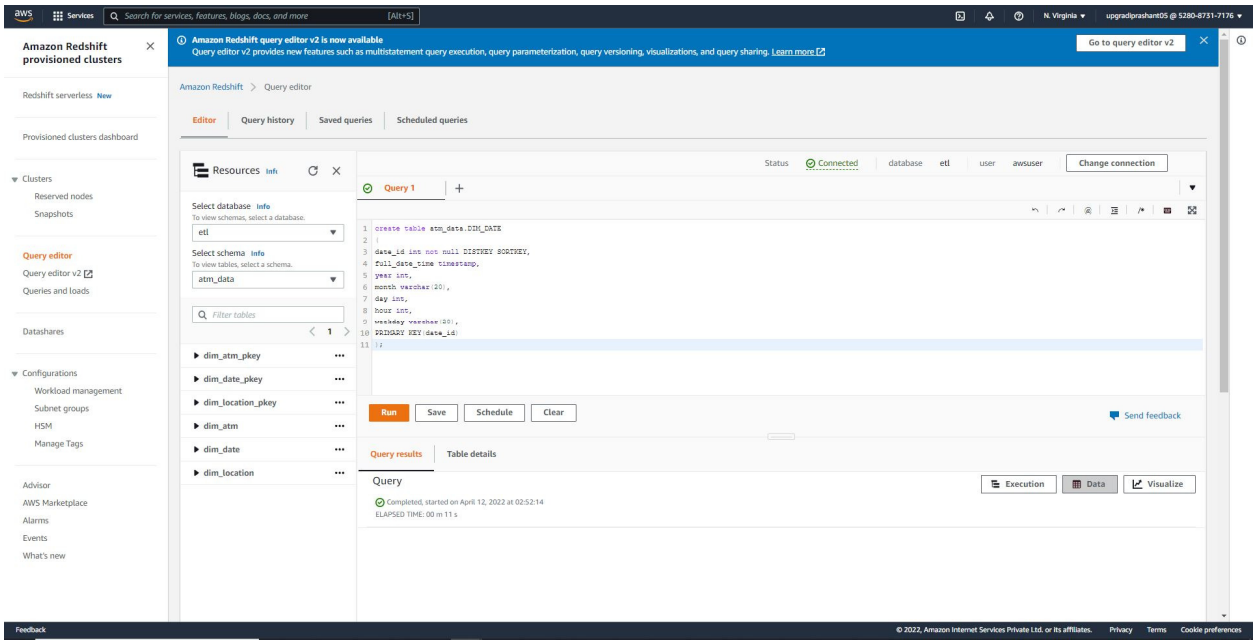
- **Creating atm dimension table**

```
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)
);
```



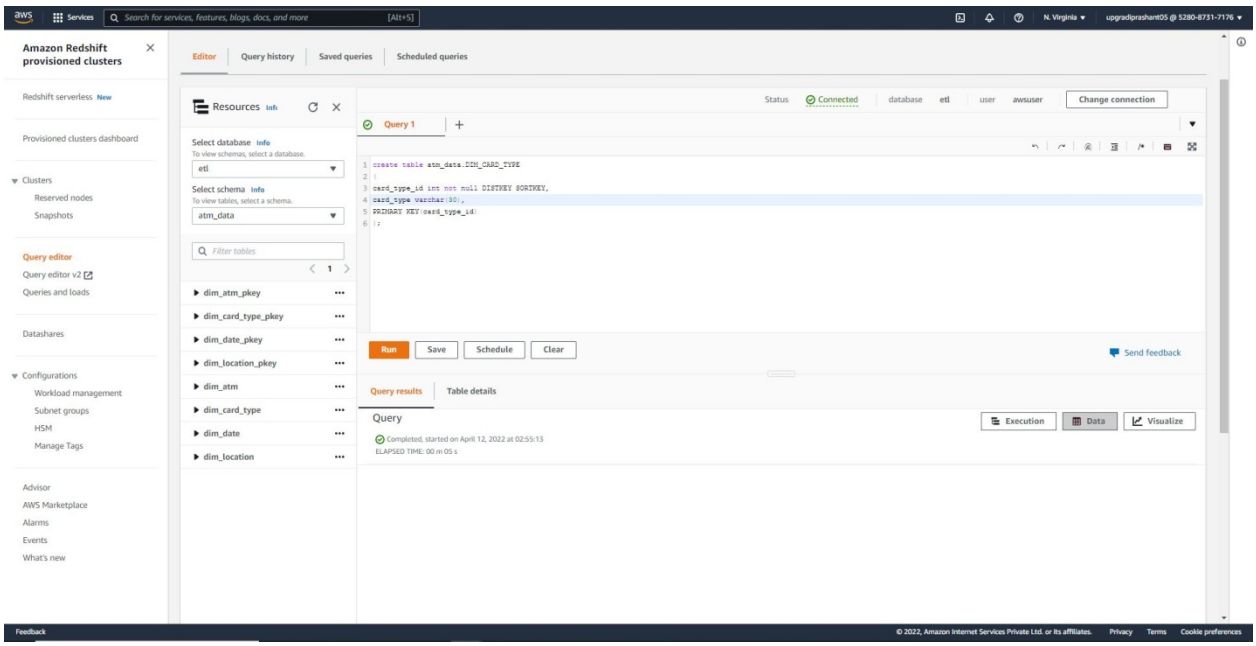
- **Creating date dimension table**

```
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)
);
```



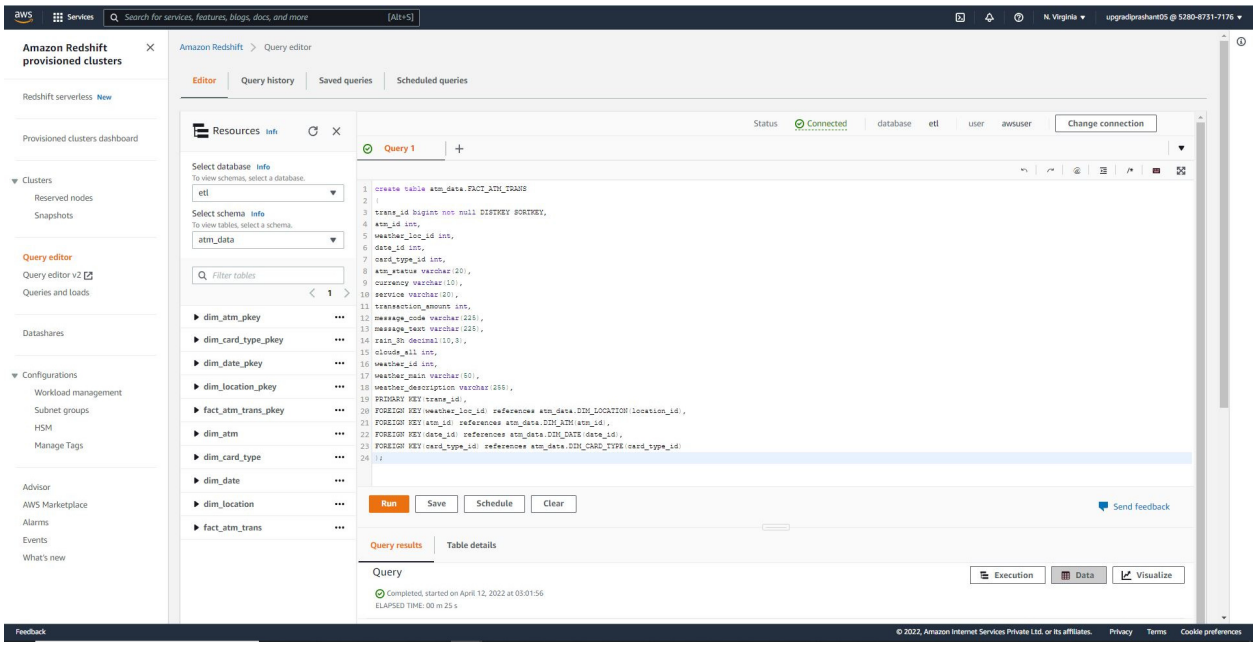
- **Creating card type dimension table**

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



- **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)
);
```

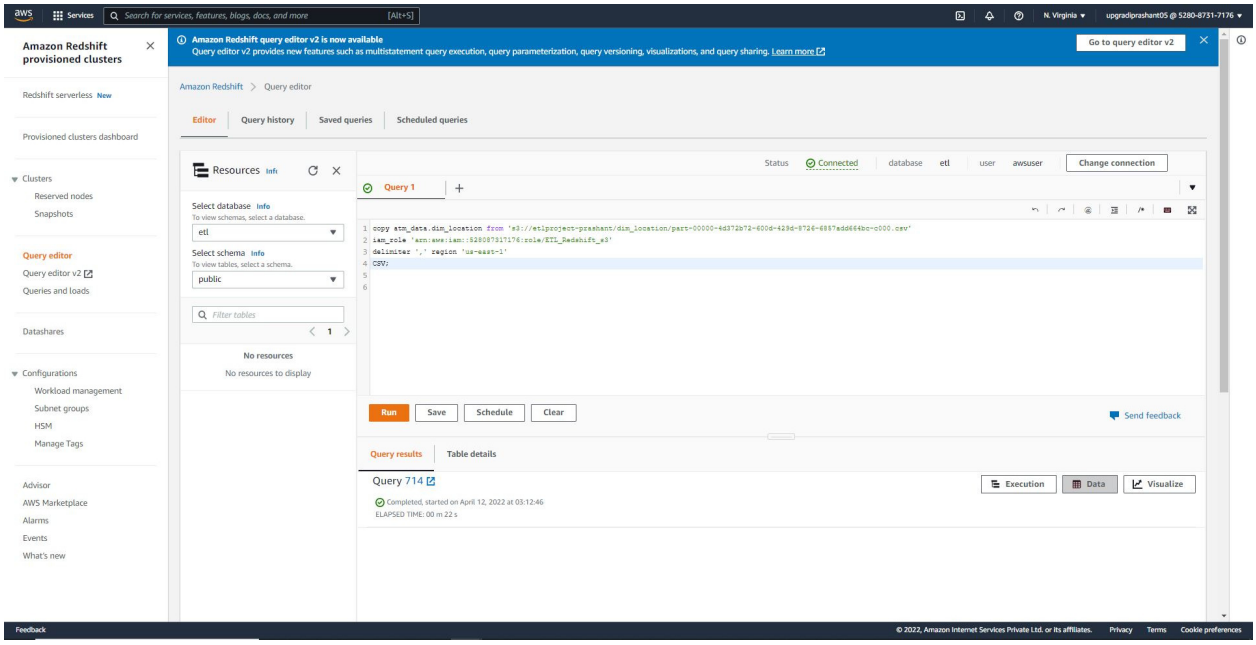


Loading data into a Redshift cluster from Amazon S3 bucket

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

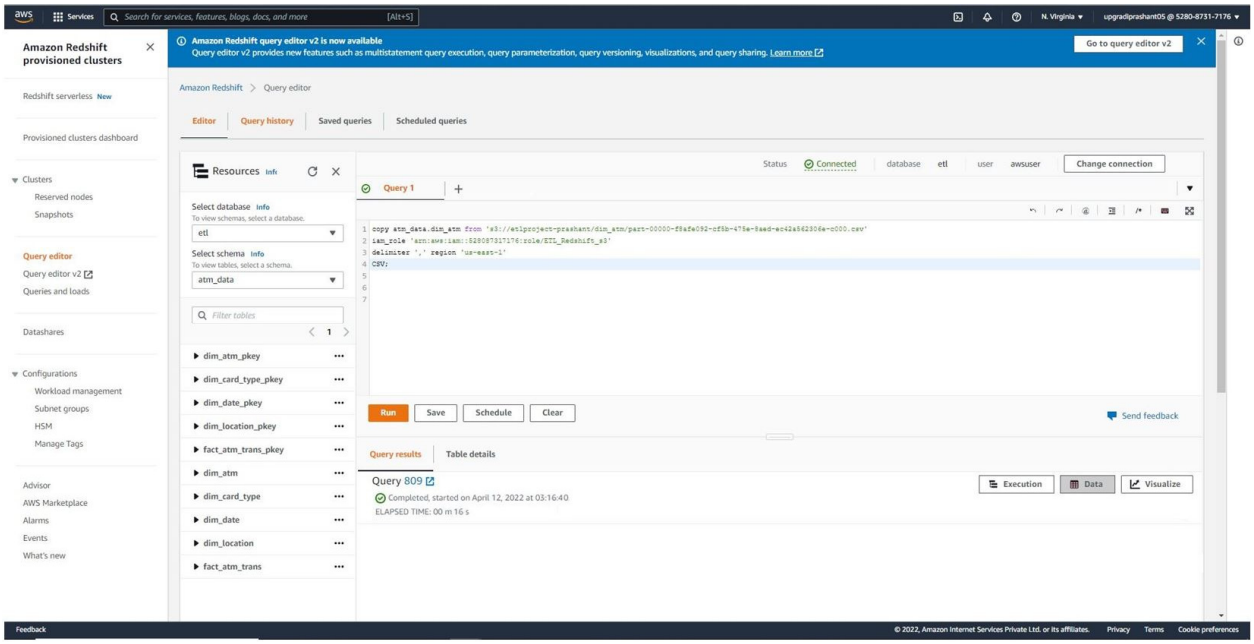
- **Copying the data to dim_location table**

copy atm_data.dim_location from 's3:// etlproject-prashant/dim_location/part-00000-4d372b72-600d-429d-8726-6857add664bc-c000.csv'
iam_role ' arn:aws:iam::528087317176:role/ETL_Redshift_s3'
delimiter ',' region 'us-east-1'
CSV
IGNOREHEADER 1;



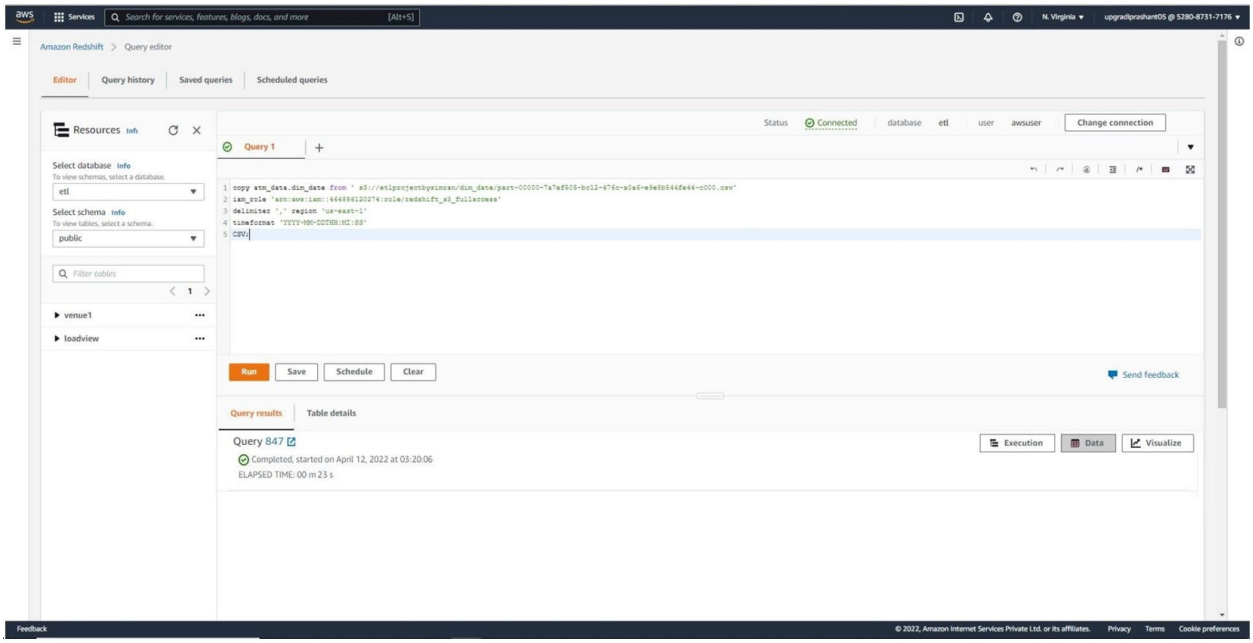
- **Copying the data to dim_atm table**

copy atm_data.dim_atm from 's3:// etlproject-prashant /dim_atm/part-00000-24973d5b-d4be-47f8-acdd-9d4f60720cb9-c000.csv '
iam_role ' arn:aws:iam::528087317176:role/ETL_Redshift_s3'
delimiter ',' region 'us-east-1'
CSV
IGNOREHEADER 1;



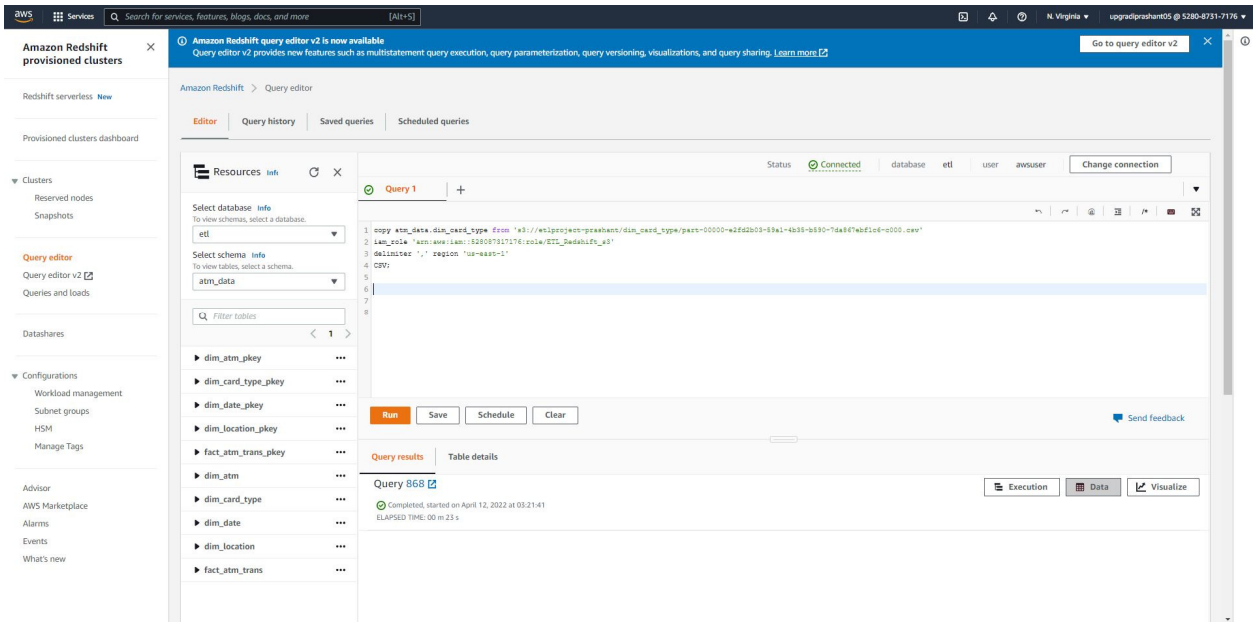
- **Copying the data to dim_date table**

copy atm_data.dim_date from 's3:// etlproject-prashant /dim_date/ part-00000-da39da18-7577-4471-9d06-0ca9e6e705d0-c000.csv'
iam_role ' arn:aws:iam::528087317176:role/ETL_Redshift_s3'
delimiter ',' region 'us-east-1'
CSV
IGNOREHEADER 1
TIMEFORMAT ‘auto’;



- **Copying the data to dim_card_type table**

copy atm_data.dim_card_type from 's3:// etlproject-prashant / dim_card_type / part-00000-e2fd2b03-59a1-4b35-b590-7da867ebf1c6-c000.csv '
iam_role ' arn:aws:iam::528087317176:role/ETL_Redshift_s3'
delimiter ',' region 'us-east-1'
CSV
IGNOREHEADER 1;



- **Copying the data to fact_atm_trans table**

copy atm_data. fact_atm_trans from 's3:// etlproject-prashant / fact_atm_trans / part-00000-3ad2dd7c-c18f-4ca1-ab99-48f2a1871bef-c000.csv '
iam_role ' arn:aws:iam::528087317176:role/ETL_Redshift_s3'
delimiter ',' region 'us-east-1'
CSV
IGNOREHEADER 1;