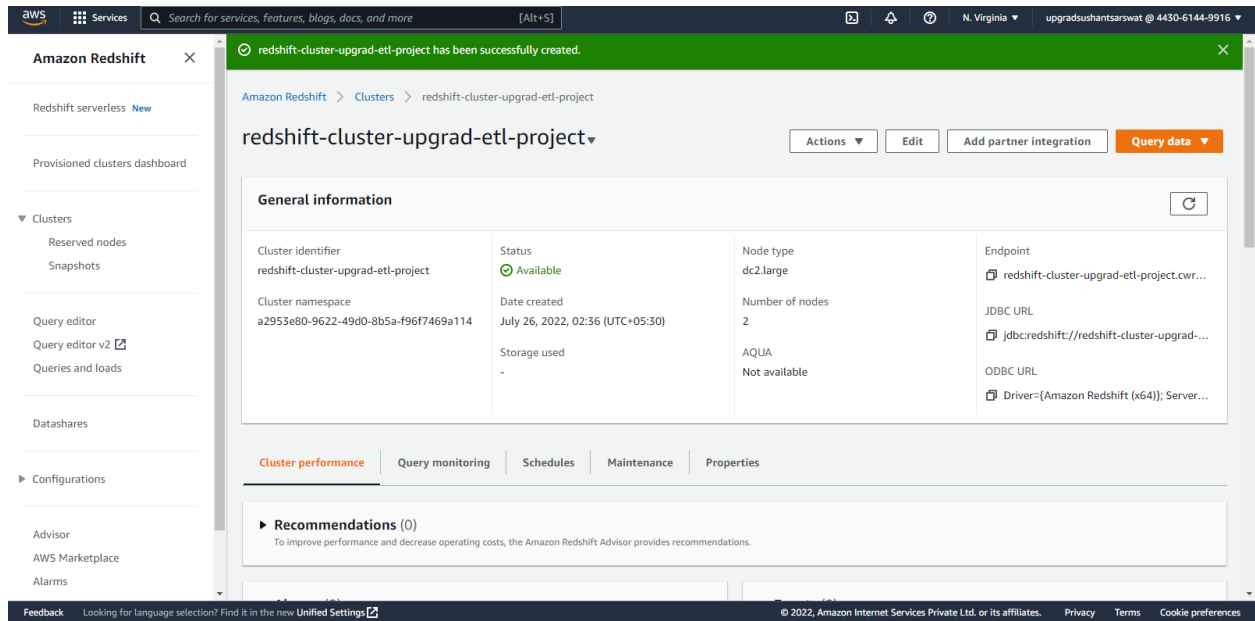


Creation of a Redshift Cluster

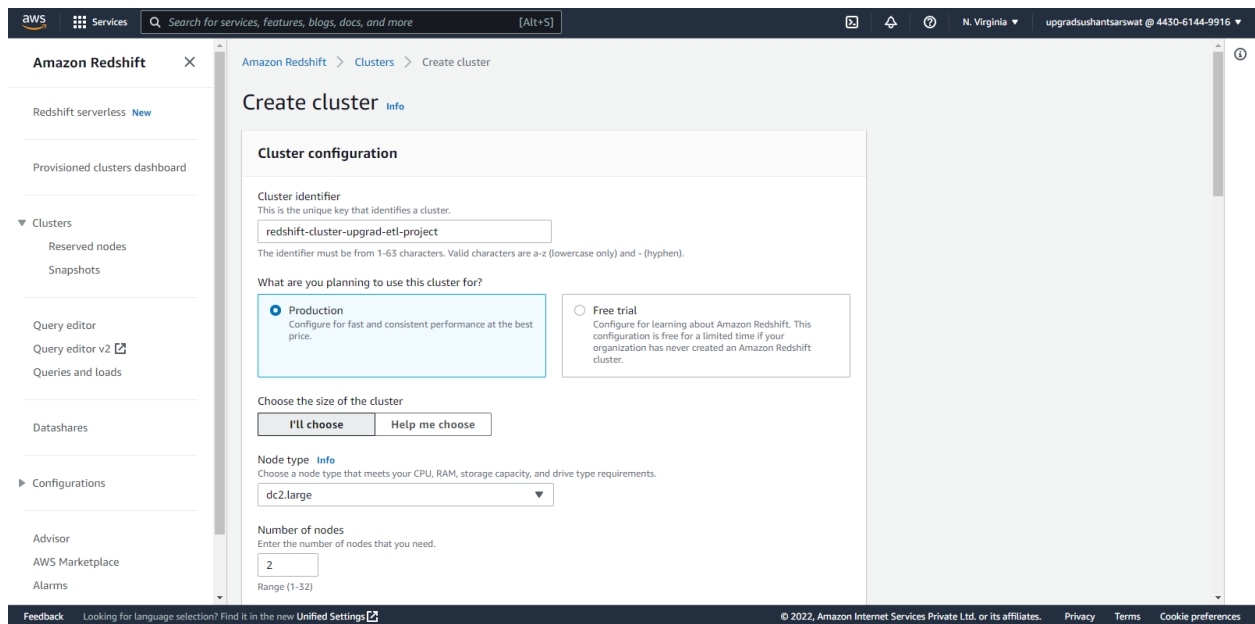
Screenshots of the configuration of the Redshift cluster that I have created:



The screenshot shows the Amazon Redshift console interface. A green notification banner at the top states: "redshift-cluster-upgrad-etl-project has been successfully created." The main content area displays the cluster details for "redshift-cluster-upgrad-etl-project".

General information			
Cluster identifier redshift-cluster-upgrad-etl-project	Status Available	Node type dc2.large	Endpoint redshift-cluster-upgrad-etl-project.cwr...
Cluster namespace a2953e80-9622-49d0-8b5a-f96f7469a114	Date created July 26, 2022, 02:36 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-upgrad-...
	Storage used -	AQUA Not available	ODBC URL Driver=(Amazon Redshift (x64)); Server=...

Below the general information, there are tabs for "Cluster performance", "Query monitoring", "Schedules", "Maintenance", and "Properties". A "Recommendations (0)" section is also visible, stating: "To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations."



The screenshot shows the "Create cluster" configuration page in the Amazon Redshift console. The cluster identifier is "redshift-cluster-upgrad-etl-project".

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.
redshift-cluster-upgrad-etl-project
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 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)

Services

Search for services, features, blogs, docs, and more

[Alt+S]

N. Virginia

upgradsushantsarawat @ 4430-5144-9916

Amazon Redshift

Redshift serverless

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Configurations

Advisor

AWS Marketplace

Alarms

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.

Associated IAM roles (0)

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.

Feedback

Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms

Cookie preferences

Amazon Redshift

Redshift serverless

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Configurations

Advisor

AWS Marketplace

Alarms

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-0a6d06b6499c67233

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-066e5d3a90b0e784a

Cluster subnet group

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

cluster-subnet-group-redshift

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

Feedback

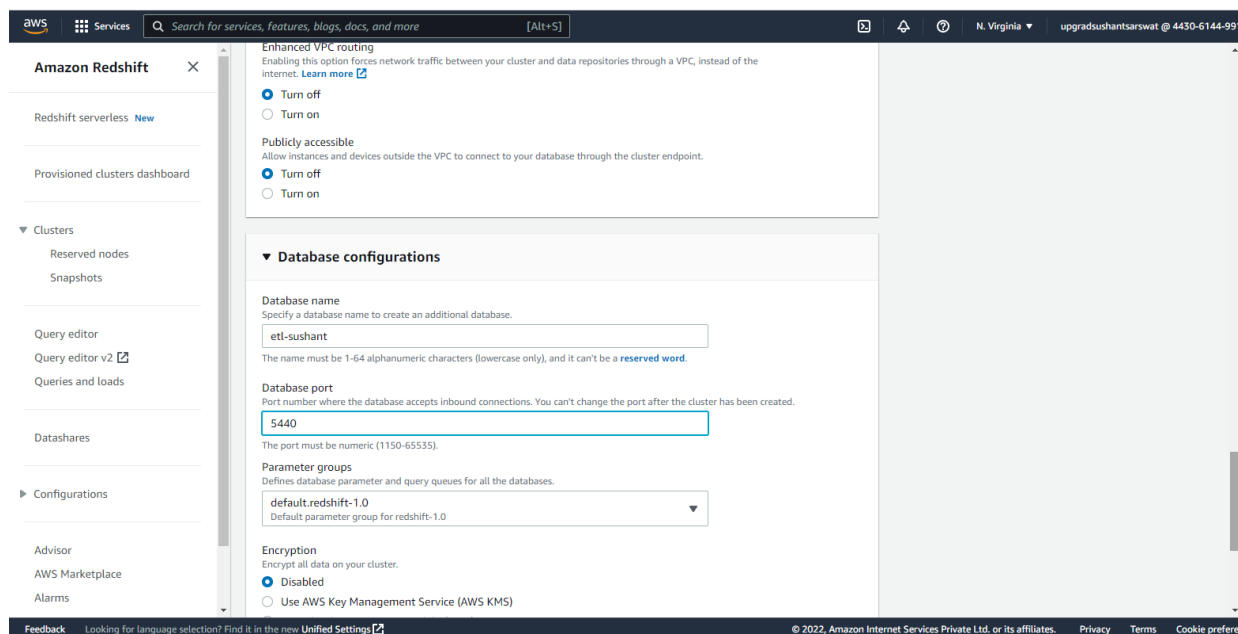
Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates.

Privacy

Terms

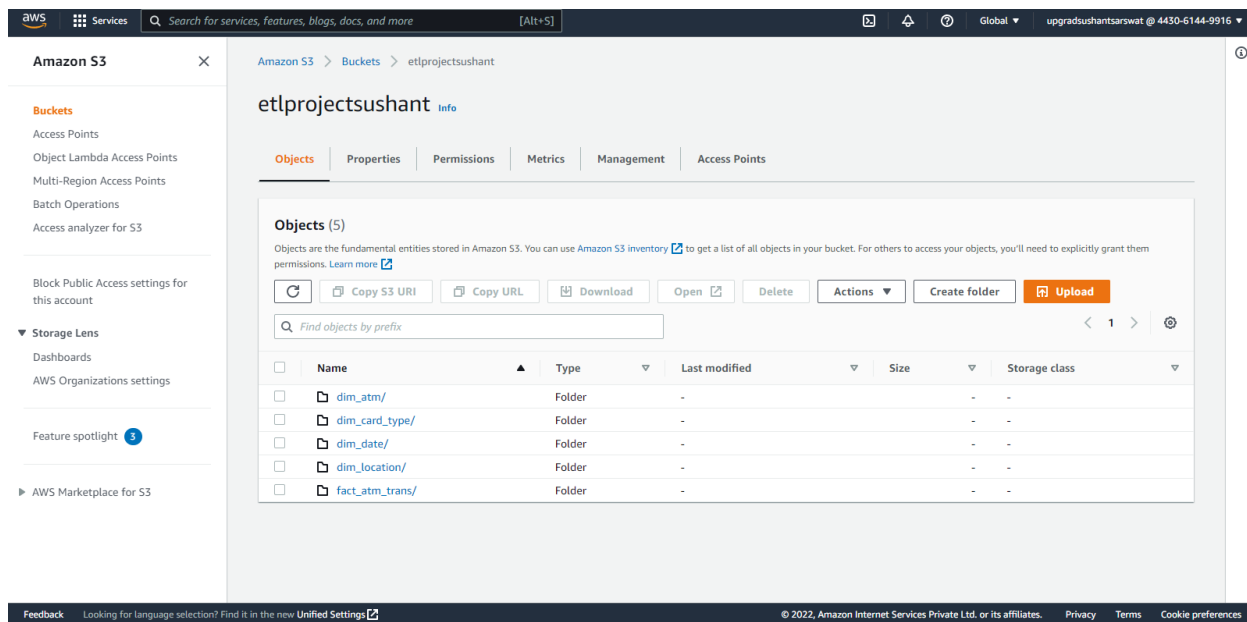
Cookie preferences



The screenshot shows the AWS Redshift console interface. The left sidebar contains navigation options like 'Amazon Redshift', 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Query editor v2', 'Queries and loads', 'Databases', 'Configurations', 'Advisor', 'AWS Marketplace', and 'Alarms'. The main content area is titled 'Database configurations' and includes sections for 'Enhanced VPC routing', 'Publicly accessible', 'Database name' (with input 'etl-sushant'), 'Database port' (with input '5440'), 'Parameter groups' (with dropdown 'default.redshift-1.0'), and 'Encryption' (with radio buttons for 'Disabled' and 'Use AWS Key Management Service (AWS KMS)').

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

Viewing all data in Amazon S3 Bucket

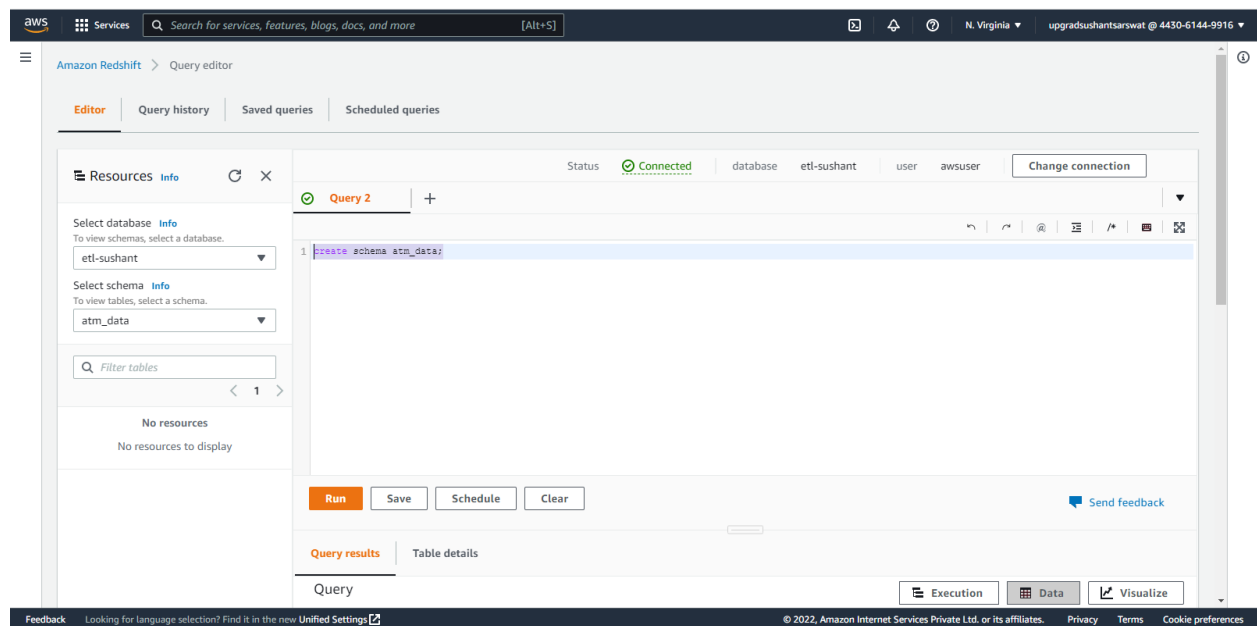


The screenshot shows the AWS S3 console interface. The left sidebar contains navigation options like 'Amazon S3', 'Buckets', 'Access Points', 'Object Lambda Access Points', 'Multi-Region Access Points', 'Batch Operations', 'Access analyzer for S3', 'Block Public Access settings for this account', 'Storage Lens', 'Dashboards', 'AWS Organizations settings', 'Feature spotlight', and 'AWS Marketplace for S3'. The main content area is titled 'etlprojectsushant' and shows a list of objects in the bucket. The objects are listed in a table with columns: Name, Type, Last modified, Size, and Storage class.

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 dimensions and fact tables:

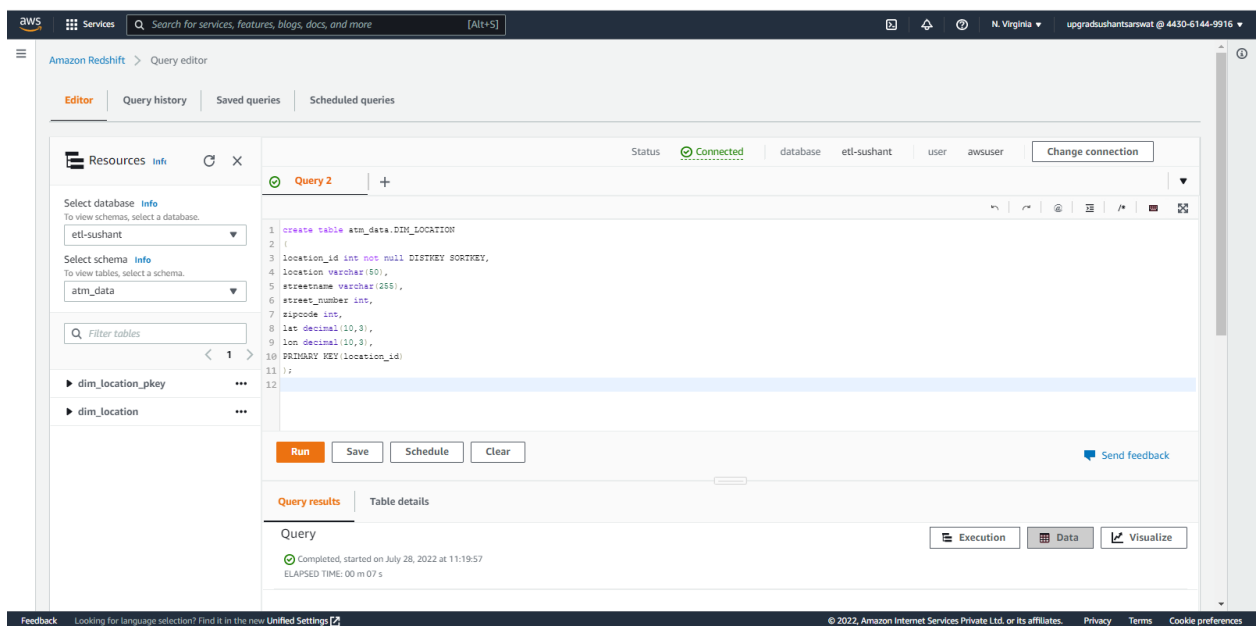
create schema atm_data;



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)
);
```



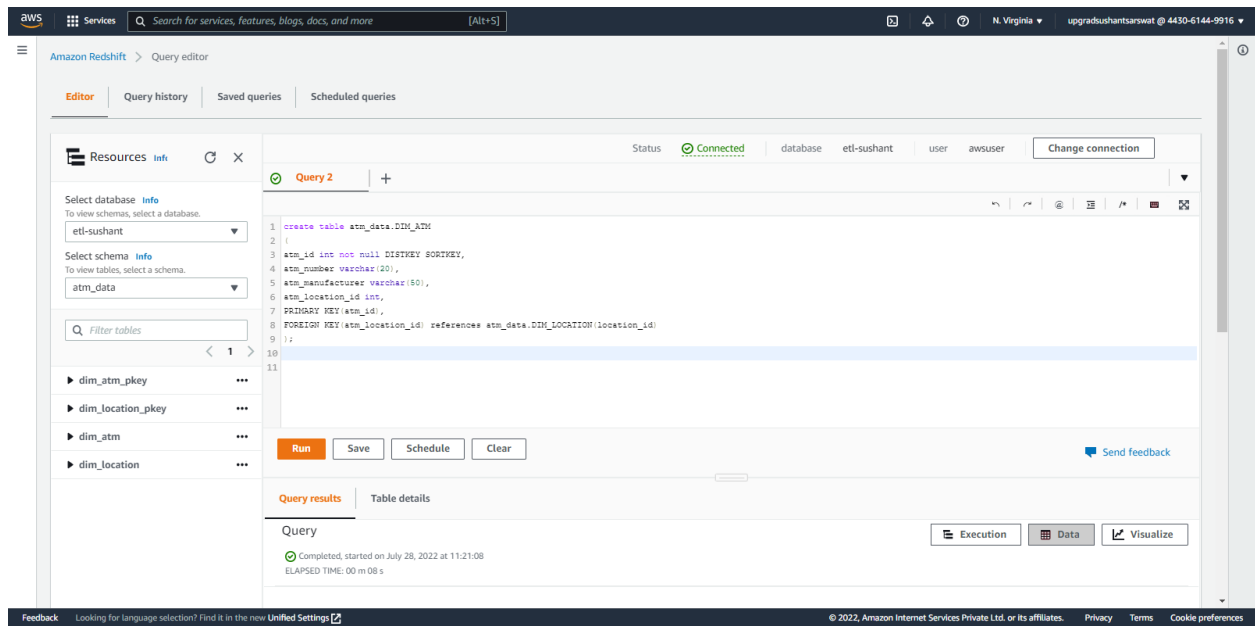
The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel displays the database 'eti-sushant' and schema 'atm_data'. The main editor area shows a SQL query to create a table named 'DIM_LOCATION' in the 'atm_data' schema. The query is as follows:

```
1 create table atm_data.DIM_LOCATION
2 (
3 location_id int not null DISTKEY SORTKEY,
4 location varchar(50),
5 streetname varchar(255),
6 street_number int,
7 zipcode int,
8 lat decimal(10,3),
9 lon decimal(10,3),
10 PRIMARY KEY(location_id)
11 );
12
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' tab is active, showing a status of 'Completed, started on July 28, 2022 at 11:19:57' and 'ELAPSED TIME: 00 m 07 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- 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)
);
```



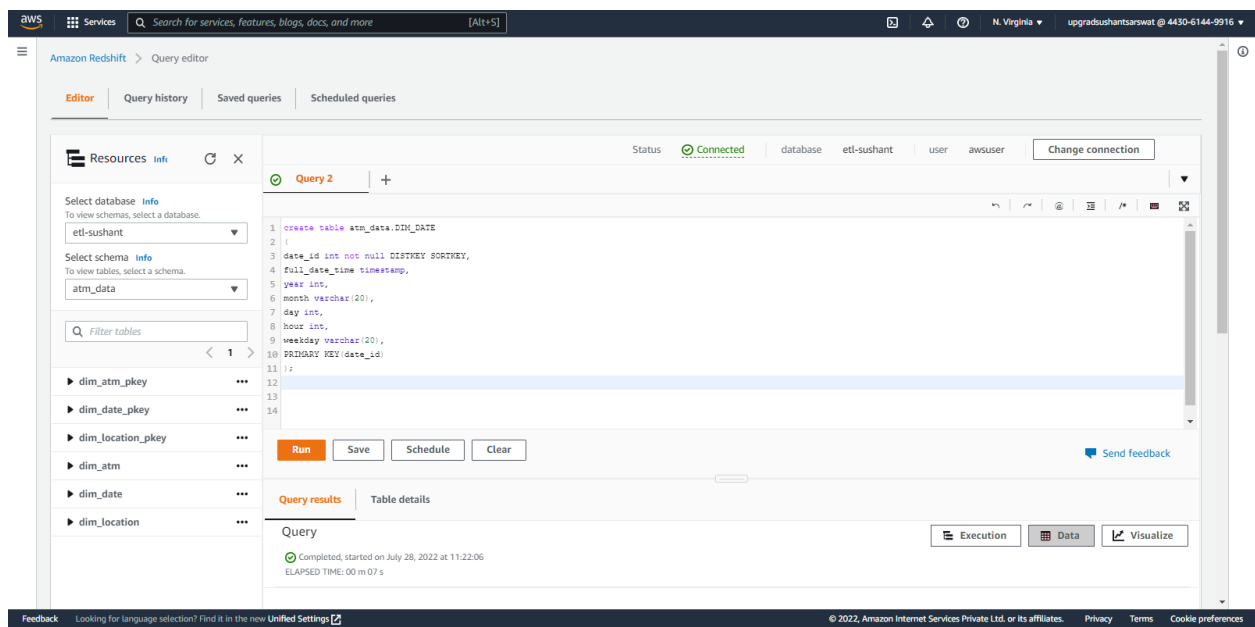
The screenshot displays the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel shows the database 'eti-sushant' and schema 'atm_data'. The main editor area contains a SQL query to create a table 'DIM_ATM' with columns 'atm_id', 'atm_number', 'atm_manufacturer', and 'atm_location_id'. The query is as follows:

```
1 create table atm_data.DIM_ATM
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 )
10
11
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button has been clicked, and the 'Query results' tab is active, showing the query execution status: 'Completed, started on July 28, 2022 at 11:21:08' and 'ELAPSED TIME: 00 m 08 s'.

- Creating date dimension table

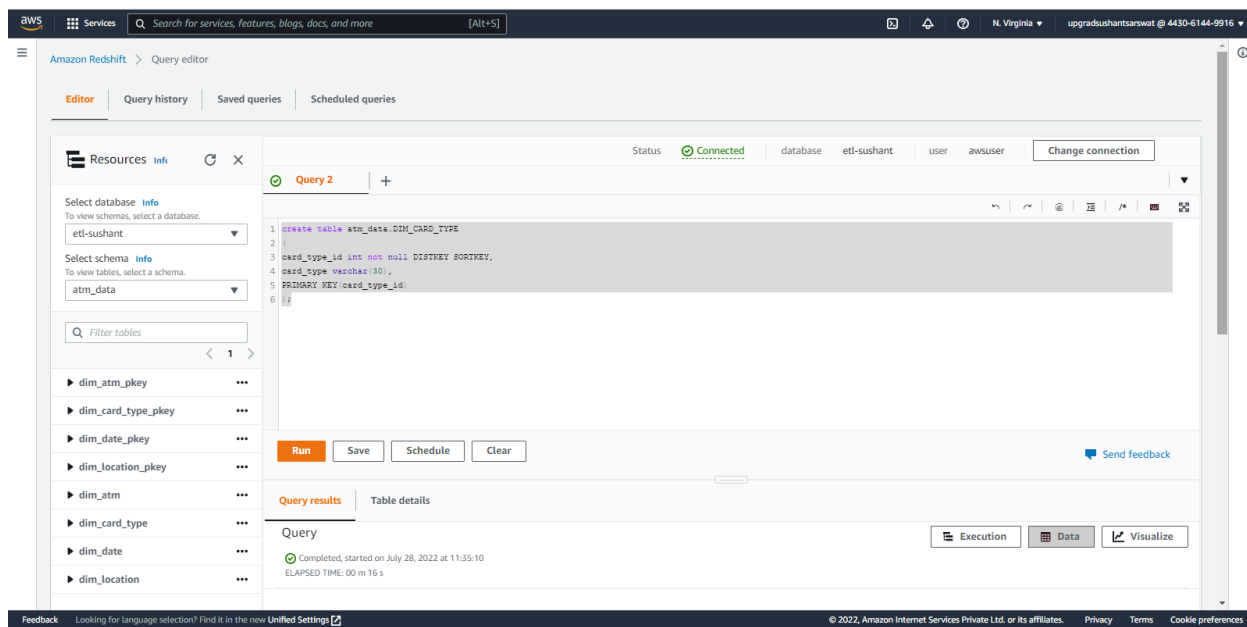
```
create table atm_data.DIM_DATE
(
date_id int not null DISTKEY SORTKEY,
full_date_time timestamp,
year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);
```



The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel displays the database 'etl-sushant' and schema 'atm_data'. The main editor area shows a SQL query for creating a table named 'DIM_DATE' in the 'atm_data' schema. The query includes columns for 'date_id' (int, not null, DISTKEY, SORTKEY), 'full_date_time' (timestamp), 'year' (int), 'month' (varchar(20)), 'day' (int), 'hour' (int), 'weekday' (varchar(20)), and a primary key on 'date_id'. Below the query editor, the 'Run' button is visible. The bottom section shows the query execution status: 'Completed, started on July 28, 2022 at 11:22:06' with an elapsed time of '00 m 07 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)  
);
```



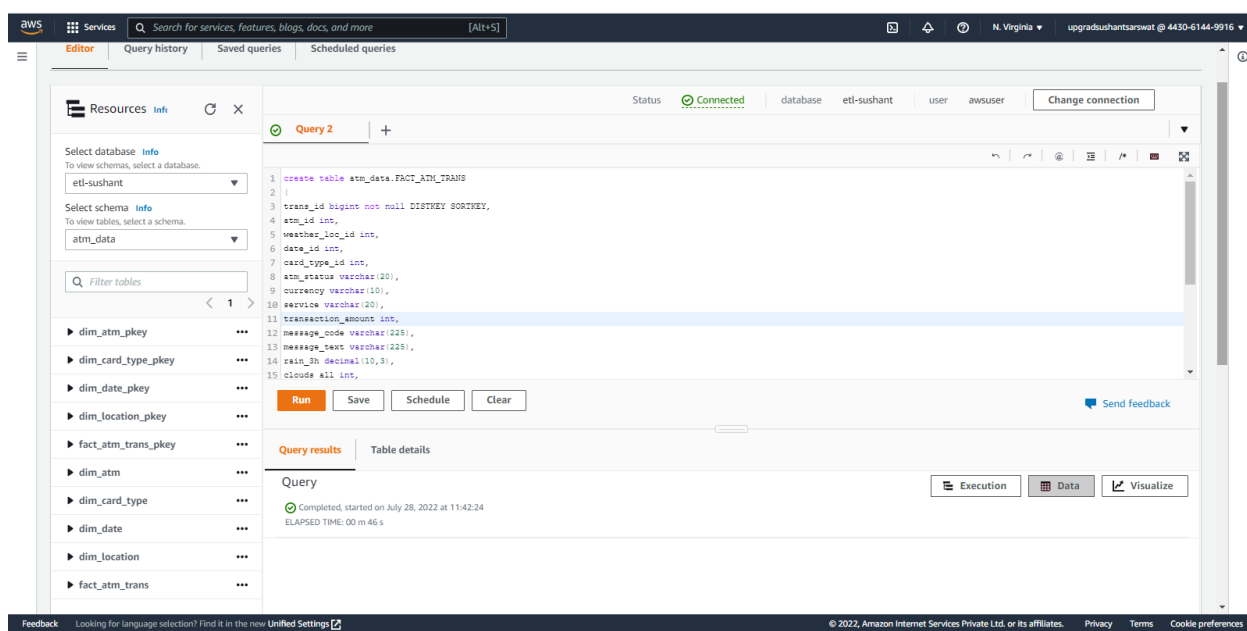
The screenshot displays the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel shows the database 'eti-sushant' and schema 'atm_data'. The main editor area contains a SQL query to create a table named 'DIM_CARD_TYPE' in the 'atm_data' schema. The query is as follows:

```
1 create table atm_data.DIM_CARD_TYPE  
2 (  
3   card_type_id int not null DISTKEY SORTKEY,  
4   card_type varchar(30)  
5   PRIMARY KEY(card_type_id)  
6 );
```

Below the query editor, the 'Run' button is highlighted, indicating the query has been executed. The 'Query results' tab is active, showing the status 'Completed, started on July 28, 2022 at 11:35:10' and 'ELAPSED TIME: 00 m 16 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_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)
);
```



The screenshot shows the AWS Glue console interface. On the left, the 'Resources' panel lists various databases and schemas, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', and 'dim_atm'. The main area displays a SQL query for creating a table named 'FACT_ATM_TRANS' in the 'atm_data' database. The query includes columns for transaction ID, ATM ID, location ID, date ID, card type ID, status, currency, service, transaction amount, message code, message text, rain, clouds, weather ID, weather main, and weather description. The query is executed successfully, as indicated by the 'Query results' tab showing 'Completed, started on July 28, 2022 at 11:42:24' and 'ELAPSED TIME: 00 m 46 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

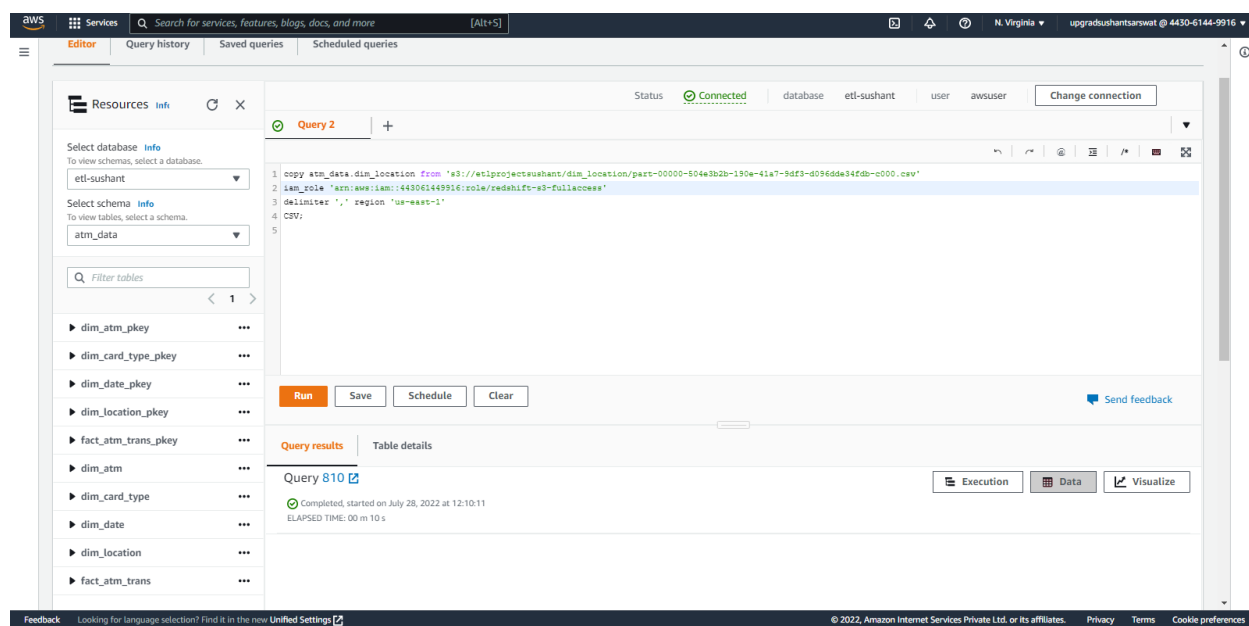
- **Copying the data to dim_location table**

```
copy atm_data.dim_location from 's3://etlprojectsushant/dim_location/part-00000-504e3b2b-190e-41a7-9df3-d096dde34fdb-c000.csv'
```

```
iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
```

```
delimiter ',' region 'us-east-1'
```

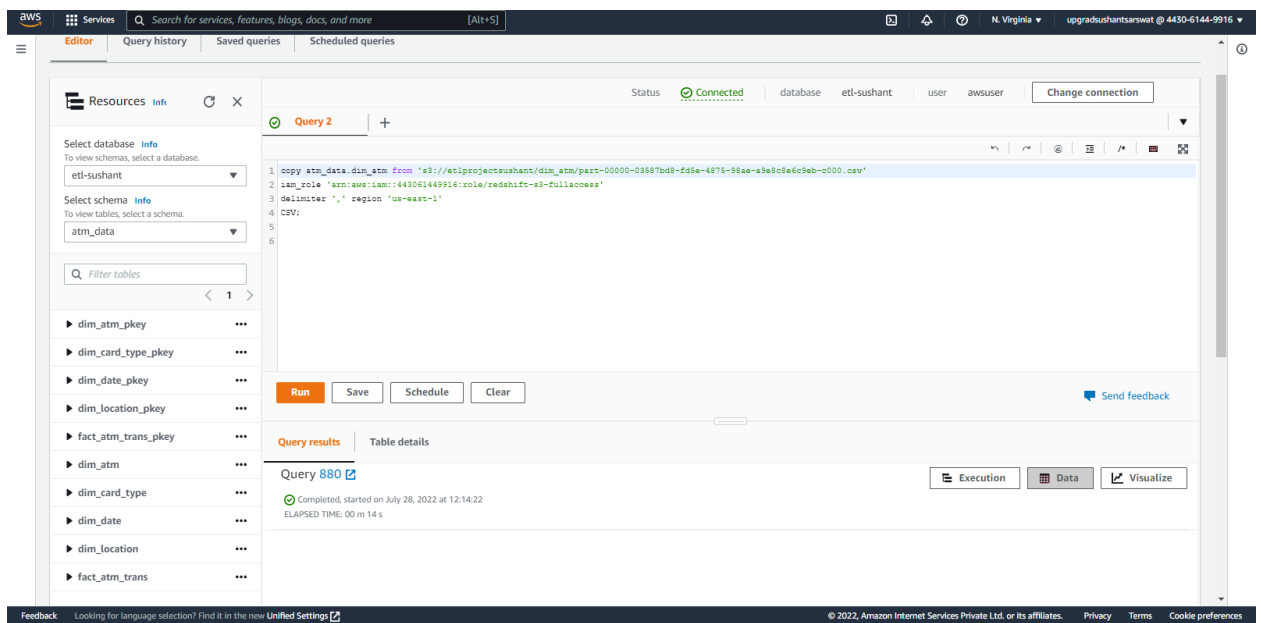
```
CSV;
```



The screenshot displays the AWS Redshift console interface. On the left, the 'Resources' panel shows the selected database 'etl-sushant' and schema 'atm_data'. The main editor area contains a SQL query (Query 2) for copying data from an S3 bucket to the 'dim_location' table. The query is highlighted in blue. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' section shows the query status as 'Completed' with a timestamp and elapsed time. The bottom of the console features a footer with copyright information and links for feedback, privacy, and terms.

- Copying the data to dim_atm table

```
copy atm_data.dim_atm from 's3://etlprojectsushant/dim_atm/part-00000-03587bd8-
fd5e-4875-98ae-a9e8c8e6c9eb-c000.csv'
iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



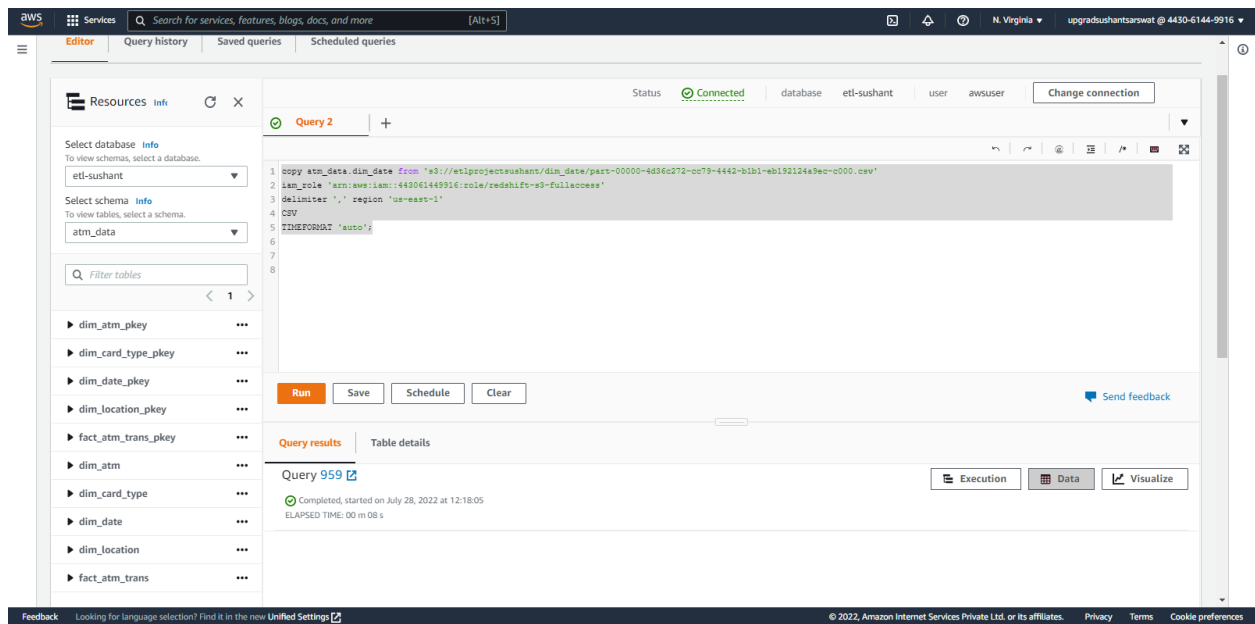
The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel is open, showing the 'Select database' dropdown set to 'eti-sushant' and the 'Select schema' dropdown set to 'atm_data'. Below these, a list of tables is visible, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main panel displays 'Query 2' with the following SQL code:

```
1 copy atm_data.dim_atm from 's3://etlprojectsushant/dim_atm/part-00000-03587bd8-fd5e-4875-98ae-a9e8c8e6c9eb-c000.csv'
2 iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
5
6
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of the 'Run' button is a 'Send feedback' link. Below the query editor, the 'Query results' tab is active, showing 'Query 880' with a status of 'Completed, started on July 28, 2022 at 12:14:22' and an 'ELAPSED TIME: 00 m 14 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to dim_date table

```
copy atm_data.dim_date from 's3://etlprojectsushant/dim_date/part-00000-4d36c272-cc79-4442-b1b1-eb192124a9ec-c000.csv'
iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



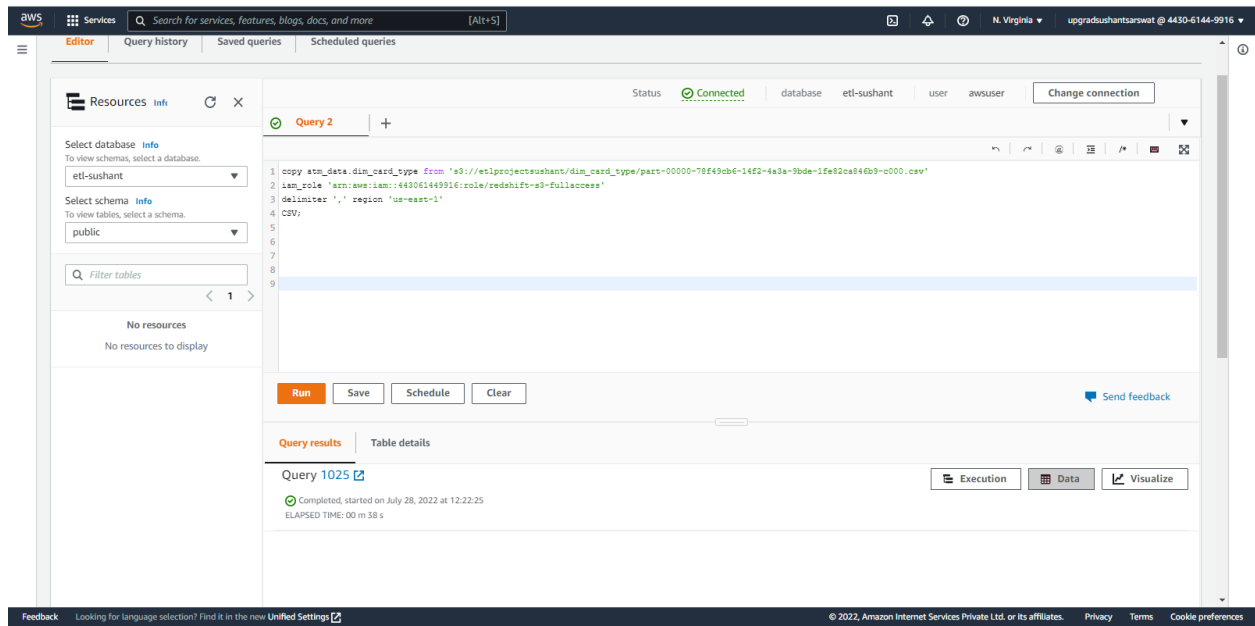
The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and a list of tables including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main area displays a SQL query in the 'Query 2' editor. The query is as follows:

```
1 copy atm_data.dim_date from 's3://etlprojectsushant/dim_date/part-00000-4d36c272-cc79-4442-b1b1-eb192124a9ec-c000.csv'
2 iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' section shows 'Query 959' with a status of 'Completed, started on July 28, 2022 at 12:18:05' and an 'ELAPSED TIME: 00 m 08 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

- Copying the data to dim_card_type table

```
copy atm_data.dim_card_type from 's3://etlprojectsushant/dim_card_type/part-00000-78f49cb6-14f2-4a3a-9bde-1fe82ca846b9-c000.csv'
iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



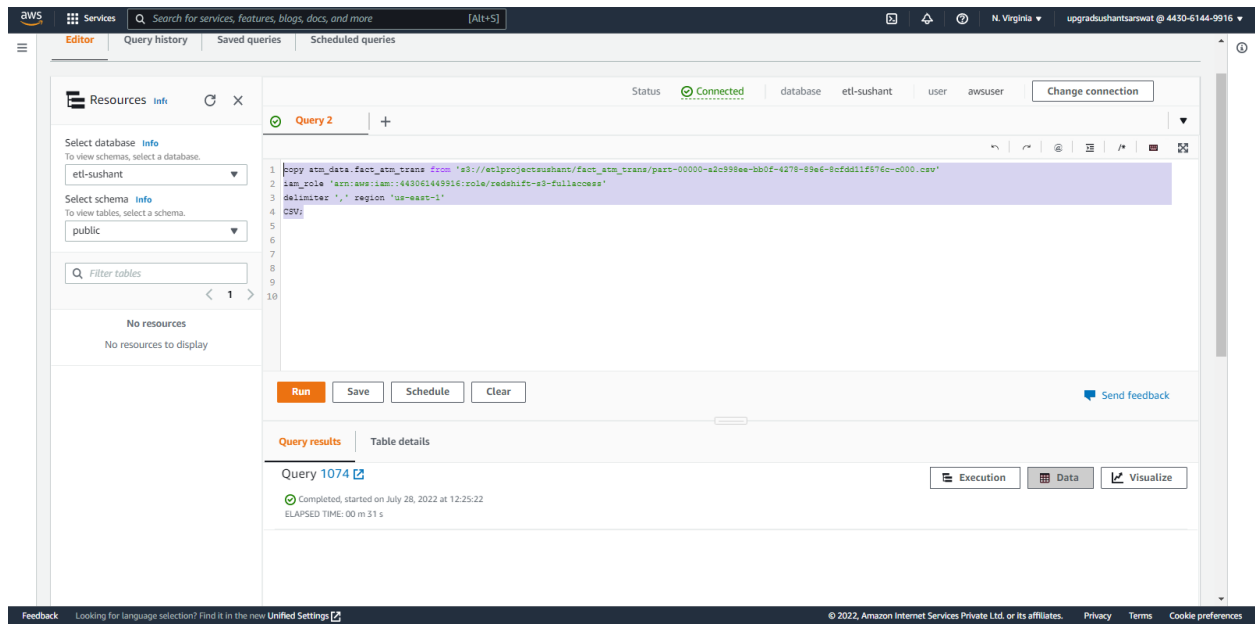
The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Query history'. The main area displays 'Query 2' with the following SQL code:

```
1 copy atm_data.dim_card_type from 's3://etlprojectsushant/dim_card_type/part-00000-78f49cb6-14f2-4a3a-9bde-1fe82ca846b9-c000.csv'
2 iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 1025' with a status of 'Completed, started on July 28, 2022 at 12:22:25' and 'ELAPSED TIME: 00 m 38 s'.

- Copying the data to fact_atm_trans table

```
copy atm_data.fact_atm_trans from 's3://etlprojectsushant/fact_atm_trans/part-00000-
a2c998ee-bb0f-4278-89e6-8cfdd11f576c-c000.csv'
iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. The main area displays a SQL query in the 'Query 2' editor. The query is as follows:

```
1 copy atm_data.fact_atm_trans from 's3://etlprojectsushant/fact_atm_trans/part-00000-a2c998ee-bb0f-4278-89e6-8cfdd11f576c-c000.csv'
2 iam_role 'arn:aws:iam::443061449916:role/redshift-s3-fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows 'Query 1074' with a status of 'Completed, started on July 28, 2022 at 12:25:22' and an 'ELAPSED TIME: 00 m 31 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.