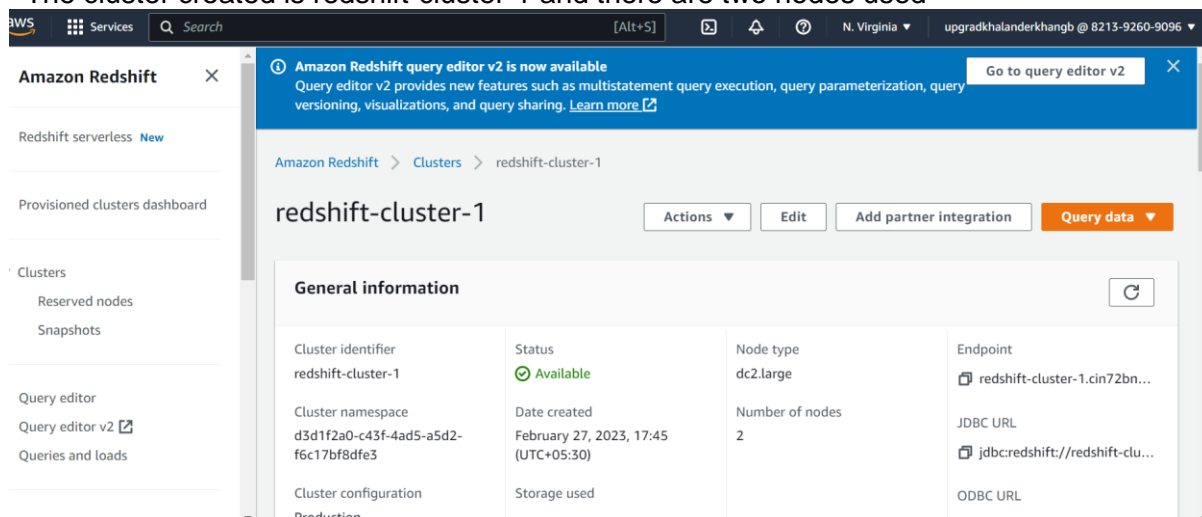


## Creation of a Redshift Cluster

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

The cluster created is redshift-cluster-1 and there are two nodes used



**Amazon Redshift**

Redshift serverless **New**

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

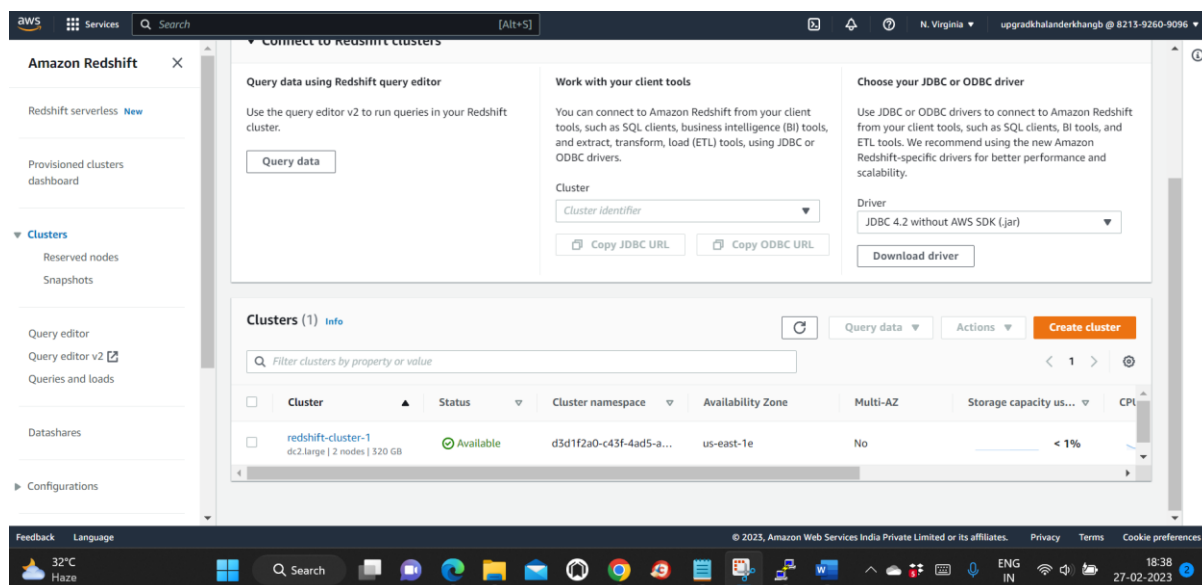
Queries and loads

**redshift-cluster-1**

Actions Edit Add partner integration Query data

**General information**

Cluster identifier redshift-cluster-1	Status Available	Node type dc2.large	Endpoint redshift-cluster-1.cin72bn...
Cluster namespace d3d1f2a0-c43f-4ad5-a5d2-f6c17bf8dfe3	Date created February 27, 2023, 17:45 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-clu...
Cluster configuration Production	Storage used -		ODBC URL -



**Connect to Redshift clusters**

Query data using Redshift query editor

Use the query editor v2 to run queries in your Redshift cluster.

Query data

Work with your client tools

You can connect to Amazon Redshift from your client tools, such as SQL clients, business intelligence (BI) tools, and extract, transform, load (ETL) tools, using JDBC or ODBC drivers.

Cluster

Cluster identifier

Copy JDBC URL Copy ODBC URL

Choose your JDBC or ODBC driver

Use JDBC or ODBC drivers to connect to Amazon Redshift from your client tools, such as SQL clients, BI tools, and ETL tools. We recommend using the new Amazon Redshift-specific drivers for better performance and scalability.

Driver

JDBC 4.2 without AWS SDK (jar)

Download driver

**Clusters (1)**

Filter clusters by property or value

Cluster	Status	Cluster namespace	Availability Zone	Multi-AZ	Storage capacity us...	CPU
redshift-cluster-1 dc2.large   2 nodes   320 GB	Available	d3d1f2a0-c43f-4ad5-a...	us-east-1e	No	< 1%	

The configurations used are –

aws

Services

Search

[Alt+S]

N. Virginia

upgradkhalanderkhangb @ 8213-9260-9096

Amazon Redshift

Redshift serverless New

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Create cluster

Cluster configuration

Cluster identifier

This is the unique key that identifies a cluster.

redshift-cluster-1

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

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

Estimated on-demand compute

320 GB

Total compressed storage

aws

Services

Search

[Alt+S]

N. Virginia

upgradkhalanderkhangb @ 8213-9260-9096

Amazon Redshift

Redshift serverless **New**

Provisioned clusters dashboard

▼ Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

☐ 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

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 "@".

Redshift123

☒ Show password

aws

Services

Search

[Alt+S]

N. Virginia

upgradkhalanderkhamb @ 8213-9260-9096

Amazon Redshift

X

Redshift serverless New

Provisioned clusters dashboard

Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Databases

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 (1/1)

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.

Set default

Manage IAM roles

Search for associated IAM role by name, status, or role type

< 1 >

<input checked="" type="checkbox"/>	IAM roles	Status	Role type
<input checked="" type="checkbox"/>	S3RedshiftFullaccess	Not applied	--

Step 2

Add permissions

Step 3

Name, review, and create

### Use case

Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Common use cases

☐ EC2  
Allows EC2 instances to call AWS services on your behalf.

☐ Lambda  
Allows Lambda functions to call AWS services on your behalf.

Use cases for other AWS services:

Choose a service to view use case

### Permissions policies (815)

Choose one or more policies to attach to your new role.

Filter policies by property or policy name and press enter. 1 match

"s3" "s3full" Clear filters

<input type="checkbox"/>	Policy name	Type	Description
<input type="checkbox"/>	AmazonS3FullAccess	AWS m...	Provides full access to all buckets via the AWS Management...

► Set permissions boundary - optional

Set a permissions boundary to control the maximum permissions this role can have. This is not a common setting, but you can use it to delegate permission management to others.

Services

Search

[Alt+S]

N. Virginia

upgradkhalanderkhangb @ 8213-9260-9096

Amazon Redshift

Redshift serverless

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

### 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

Using default VPC (vpc-d4007ea9) and default subnet.

#### Security

Using default (sg-2451ba35) cluster security group.

#### Backup

Automated snapshots are created about every eight hours or following every 5 GB per node of data changes, whichever comes first.

#### Configuration

Using default.redshift-1.0 parameter group with no database encryption.

#### Maintenance

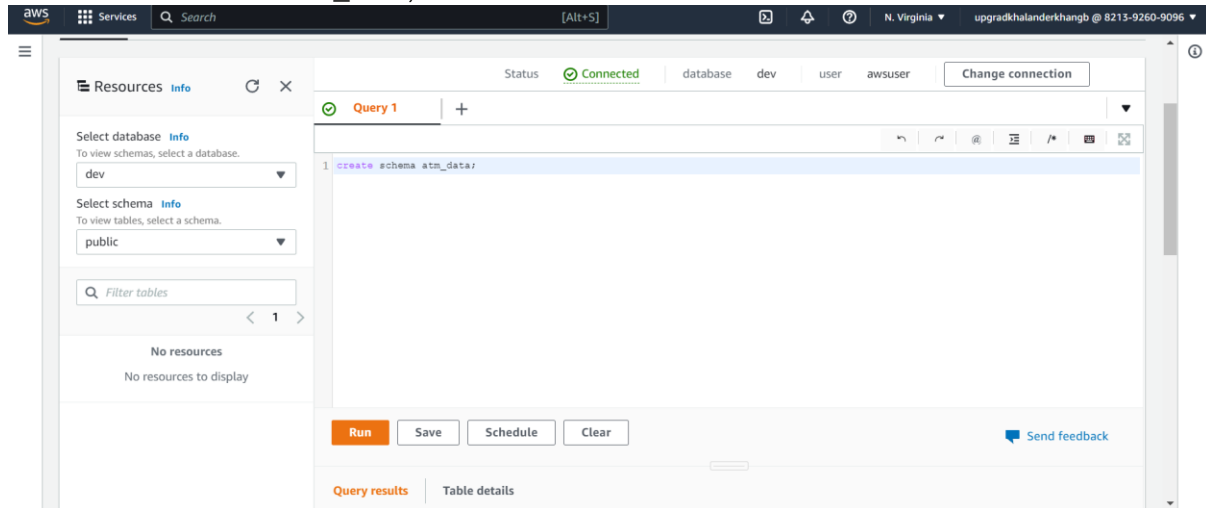
Using current maintenance track.

Cancel Create cluster

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

**Query to create the database's schema:**

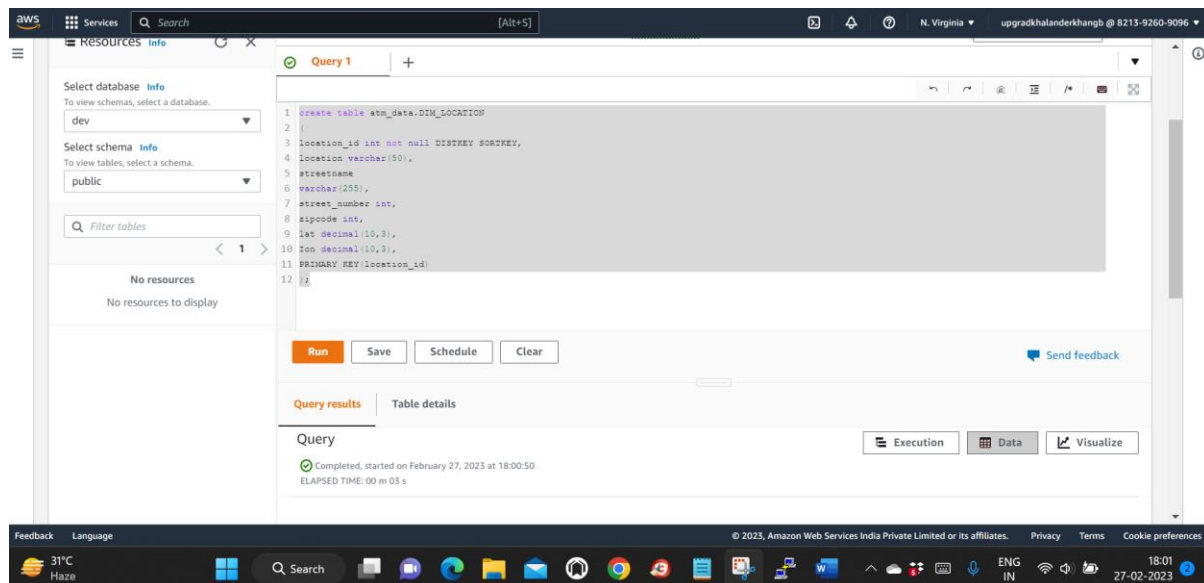
**CREATE SCHEMA atm\_data;**



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

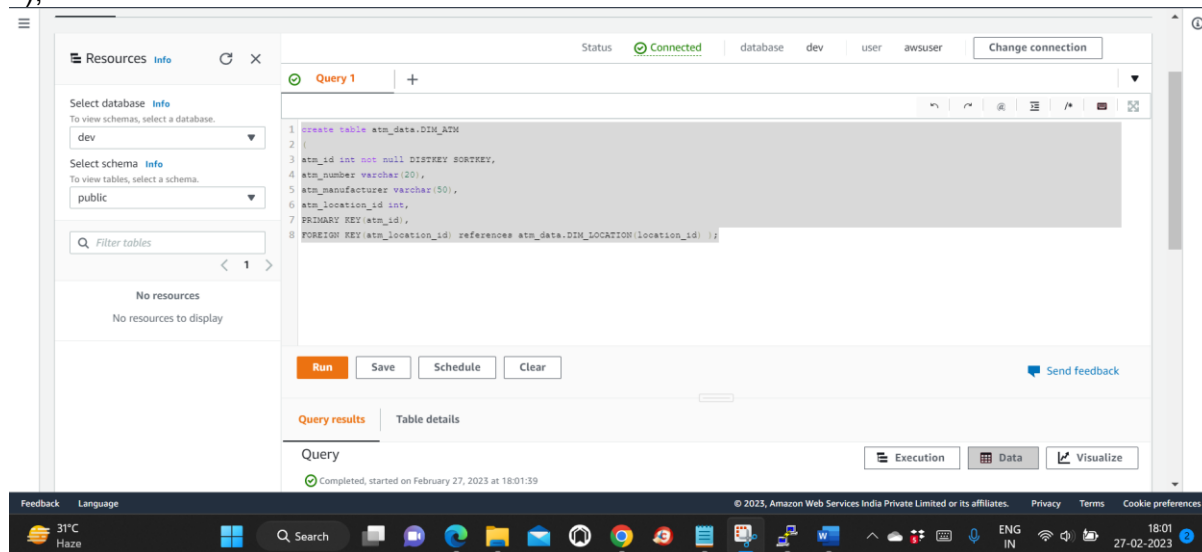
**Query to create the dimension table DIM\_LOCATION:**

```
CREATE TABLE atm_nord.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)
);
```



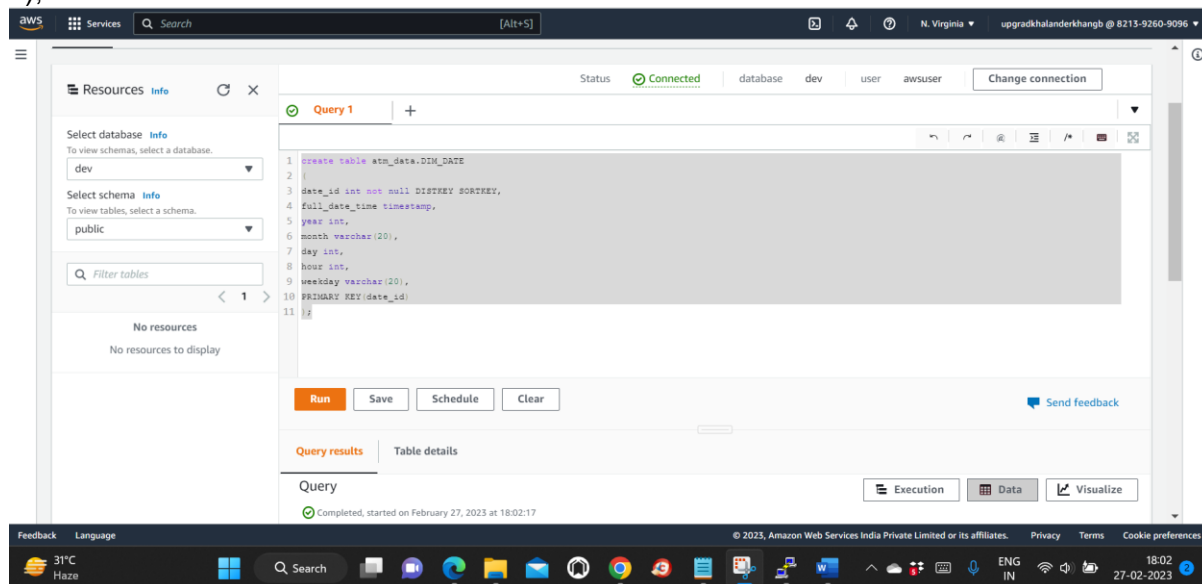
### Query to create the dimension table DIM\_ATM:

```
CREATE TABLE atm_nord.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_nord.DIM_LOCATION(location_id)
);
```



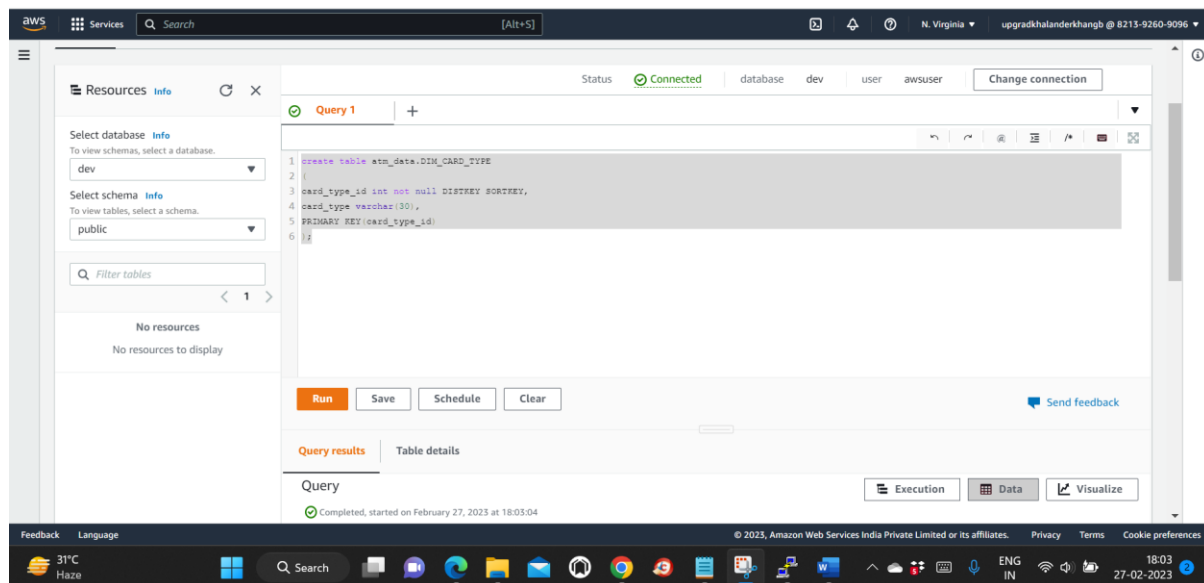
### Query to create the dimension table DIM\_DATE:

```
CREATE TABLE atm_nord.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 to create the dimension table DIM\_CARD\_TYPE:

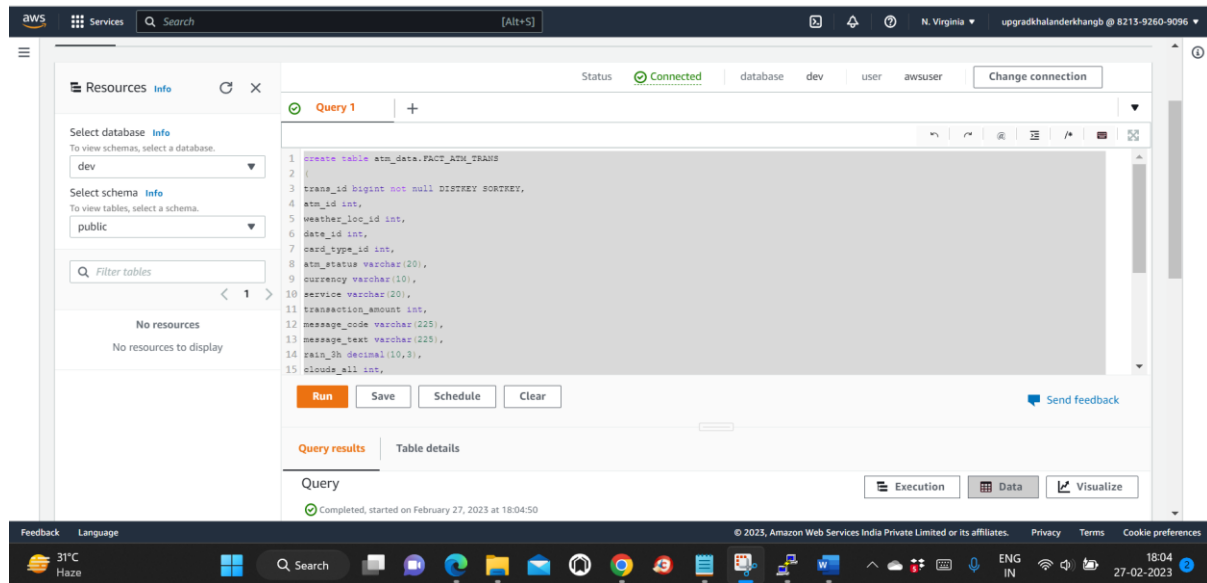
```
CREATE TABLE atm_nord.DIM_CARD_TYPE
(
card_type_id INT NOT NULL DISTKEY SORTKEY,
card_type VARCHAR(30),
PRIMARY KEY(card_type_id)
);
```



### Query to create the fact table FACT\_ATM\_TRANS:

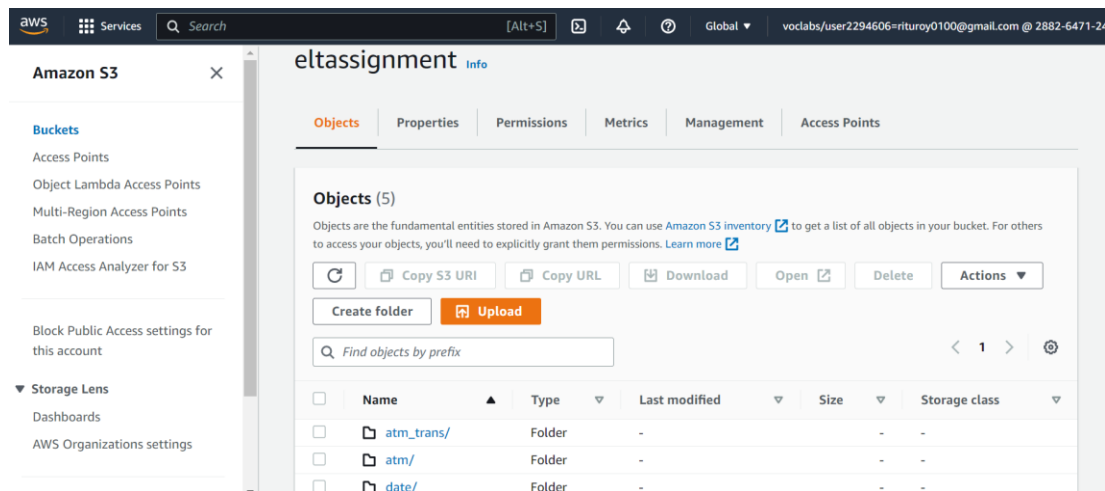
```
CREATE TABLE atm_nord.FACT_ATM_TRANS
(
trans_id BIGINT NOT NULL DISTKEY SORTKEY,
atm_id INT,
weather_loc_id INT,
date_id INT,
card_type_id INT,
atm_status VARCHAR(20),
currency VARCHAR(10),
service VARCHAR(20),
transaction_amount INT,
message_code VARCHAR(255),
message_text VARCHAR(255),
rain_3h 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_nord.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm_nord.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm_nord.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_nord.DIM_CARD_TYPE(card_type_id)
);
```





Loading data into a Redshift cluster from Amazon S3 bucket

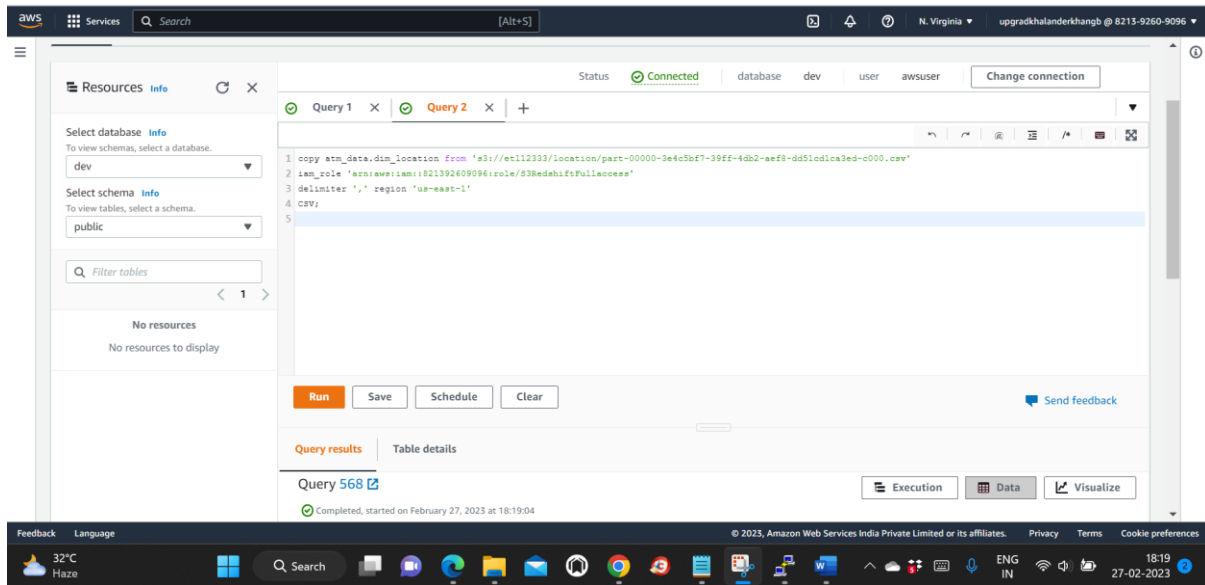
The different folders created by us for writing the data in csv format for dimension and fact table



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

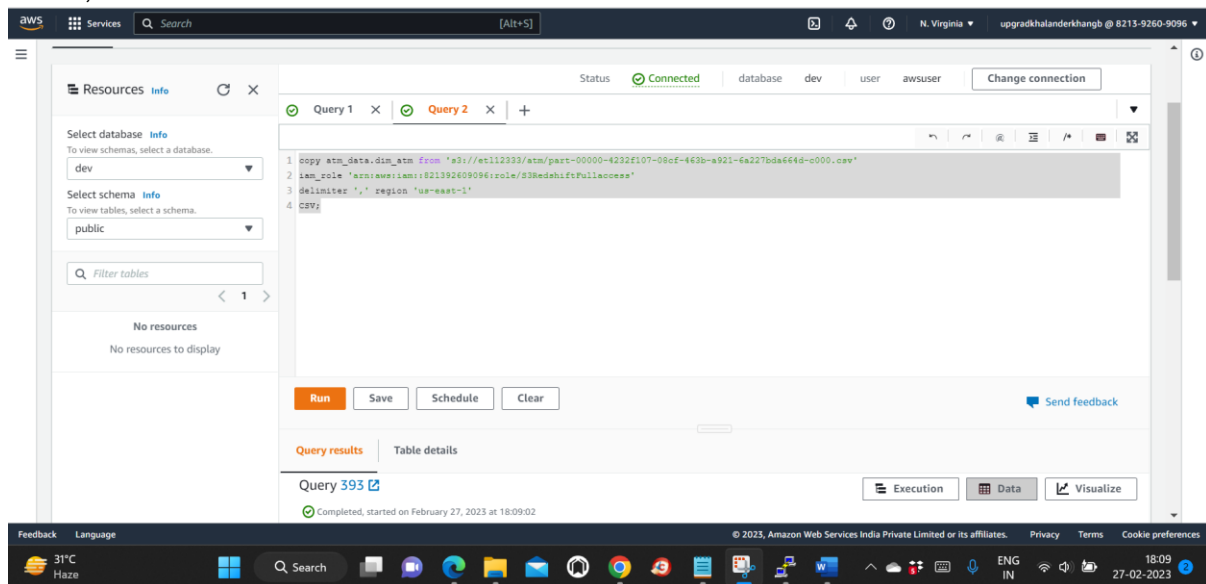
**Query to copy the data from S3 bucket – dev to dim\_location table**

```
copy atm_data.dim_location from 's3://etl12333/location/part-00000-3e4c5bf7-39ff-4db2-aef8-
dd51cd1ca3ed-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



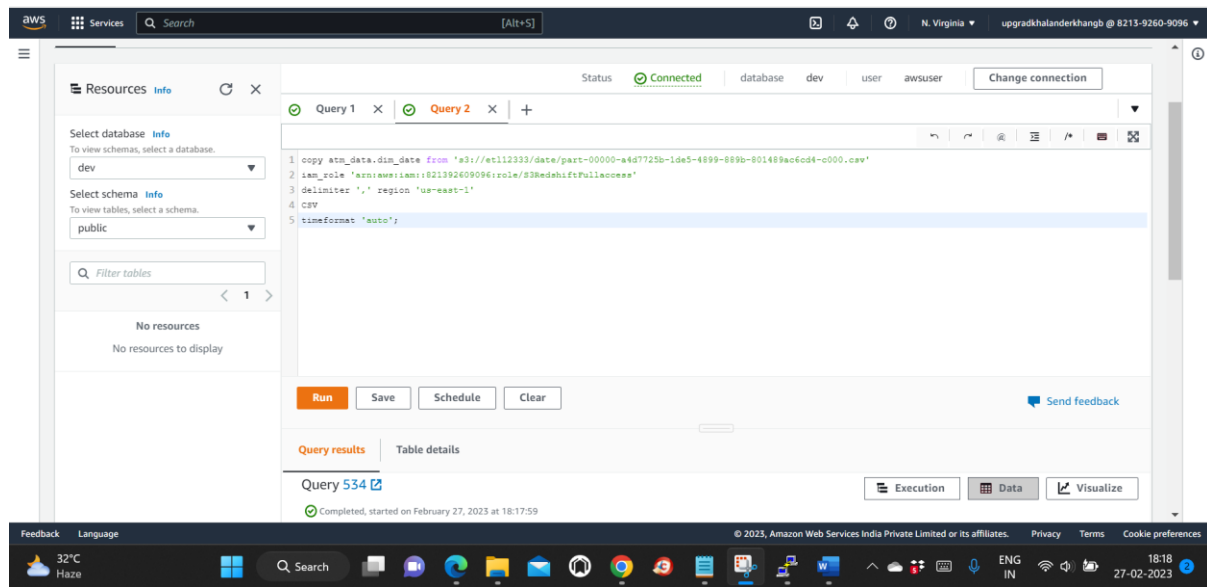
## Query to copy the data from S3 bucket – dev to dim\_atm table

```
copy atm_data.dim_atm from 's3://etl12333/atm/part-00000-4232f107-08cf-463b-a921-6a227bda664d-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



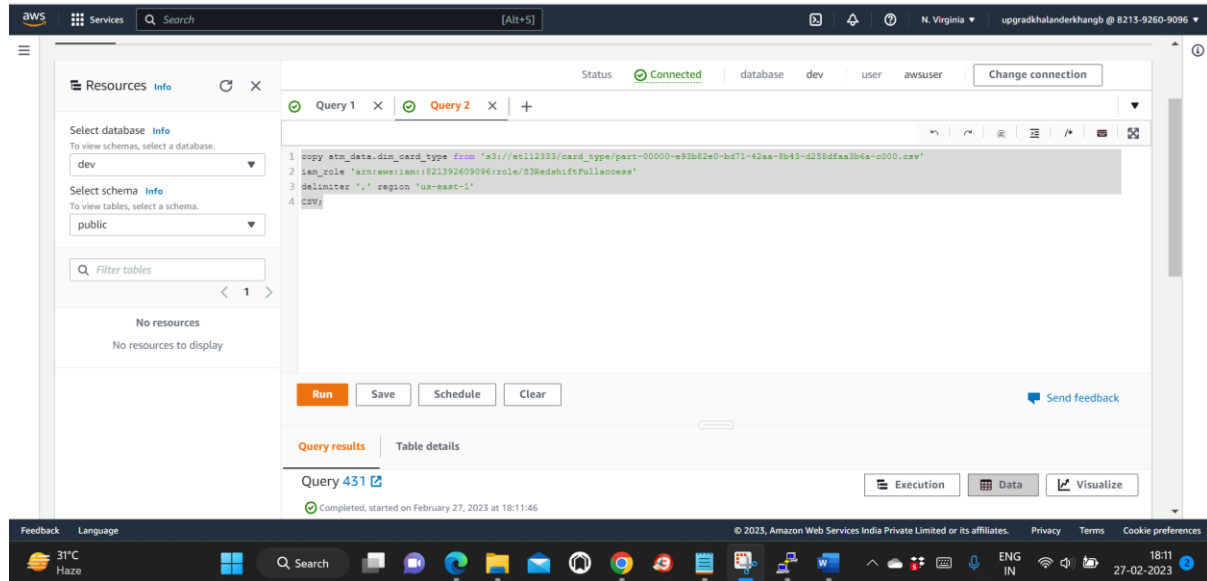
## Query to copy the data from S3 bucket – dev to dim\_date table

```
copy atm_data.dim_date from
's3://etl12333/date/part-00000-
a4d7725b-1de5-4899-889b-
801489ac6cd4-c000.csv'
iam_role
'arn:aws:iam::821392609096:role/S3
RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV
timeformat 'auto';
```



## Query to copy the data from S3 bucket – dev to dim\_card\_type table

```
copy atm_data.dim_card_type from 's3://etl12333/card_type/part-00000-e93b82e0-bd71-42aa-
8b43-d258dfaa3b6a-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel is visible with 'dev' selected as the database and 'public' as the schema. The main area displays a SQL query for 'Query 2':

```
1 copy atm_data.dim_card_type from 's3://etl12333/card_type/part-00000-e93b62e0-bd71-42aa-8b43-d258dfa3b6a-c000.csv'
2 iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
3 delimiter ',' region 'us-east-1'
4 csv;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 431' as 'Completed, started on February 27, 2023 at 18:11:46'. The bottom of the screen shows a Windows taskbar with the date 27-02-2023 and time 18:11.

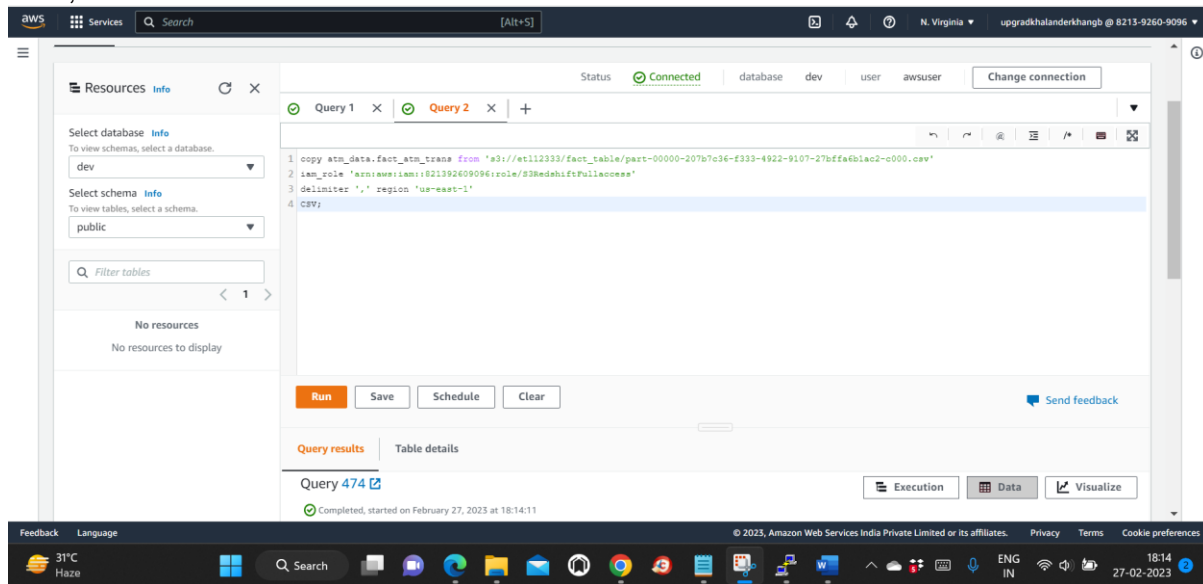
## Query to copy the data from S3 bucket – dev to fact\_atm\_trans table

copy atm\_data.fact\_atm\_trans from 's3://etl12333/fact\_table/part-00000-207b7c36-f333-4922-9107-27bffa6b1ac2-c000.csv'

iam\_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'

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

CSV;



This screenshot is similar to the first one, showing the AWS Redshift console. The 'Resources' panel on the left still shows 'dev' and 'public'. The main area displays a SQL query for 'Query 2' with the following code:

```
1 copy atm_data.fact_atm_trans from 's3://etl12333/fact_table/part-00000-207b7c36-f333-4922-9107-27bffa6b1ac2-c000.csv'
2 iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

The 'Run' button is highlighted. Below the query, the 'Query results' tab shows 'Query 474' as 'Completed, started on February 27, 2023 at 18:14:11'. The bottom of the screen shows a Windows taskbar with the date 27-02-2023 and time 18:14.