

Creation of a Redshift Cluster

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

The cluster created is etl-project-purvi and there are two nodes used

✔ etl-project-purvi has been successfully created.
✕

Amazon Redshift > Clusters > etl-project-purvi

etl-project-purvi
Actions ▼
Edit
Add partner integration
Query data ▼

General information ↻

Cluster identifier etl-project-purvi	Status ✔ Available	Node type dc2.large	Endpoint etl-project-purvi.c9nka65kpik.us-east-...
Cluster namespace d8c239cd-fb68-4b66-bd54-b4777948fbc4	Date created May 31, 2022, 01:25 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://etl-project-purvi.c9nka65kpik.us-east-...
Storage used -	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=...	

The configurations used are –

Create cluster Info

Cluster configuration

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

etl-project-purvi

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 ▼

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)

Configuration summary [Info](#)

dc2.large | 2 nodes

\$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.

Enter a login ID for the admin user of your DB instance.

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 this cluster.

<input type="checkbox"/>	IAM roles 🔗	▼	Status	▼	Role type	▼
<input type="checkbox"/>	redshift_s3_fullaccess		Not applied		--	

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-0d9c51cda1b21b0fb

[📘](#) 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

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-0226bd8f56b333210

Cluster subnet group

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

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

Database name

Specify a database name to create an additional database.

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.

The port must be numeric (1150-65535).

Parameter groups

Defines database parameter and query queues for all the databases.

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)

Automated snapshot retention period

Specify how many days to retain automated snapshots.

The retention period must be 0-35 days. Choose zero (0) to not create automated snapshots.

Manual snapshot retention period


Specify how long do you want to retain your snapshot.

The retention period must be 1-3653 days.

Configure cross-region snapshot

- ☒ Disabled
- ☐ Enabled

Cluster relocation

Enable the ability to relocate your cluster in another Availability Zone. After you enable relocation, you use the VPC endpoint of the cluster to determine the cluster IP address, instead of the leader node IP address. You can find the VPC endpoint in the Network and security section of the cluster details page. [Learn more](#) 

- ☒ No
- ☐ Enable

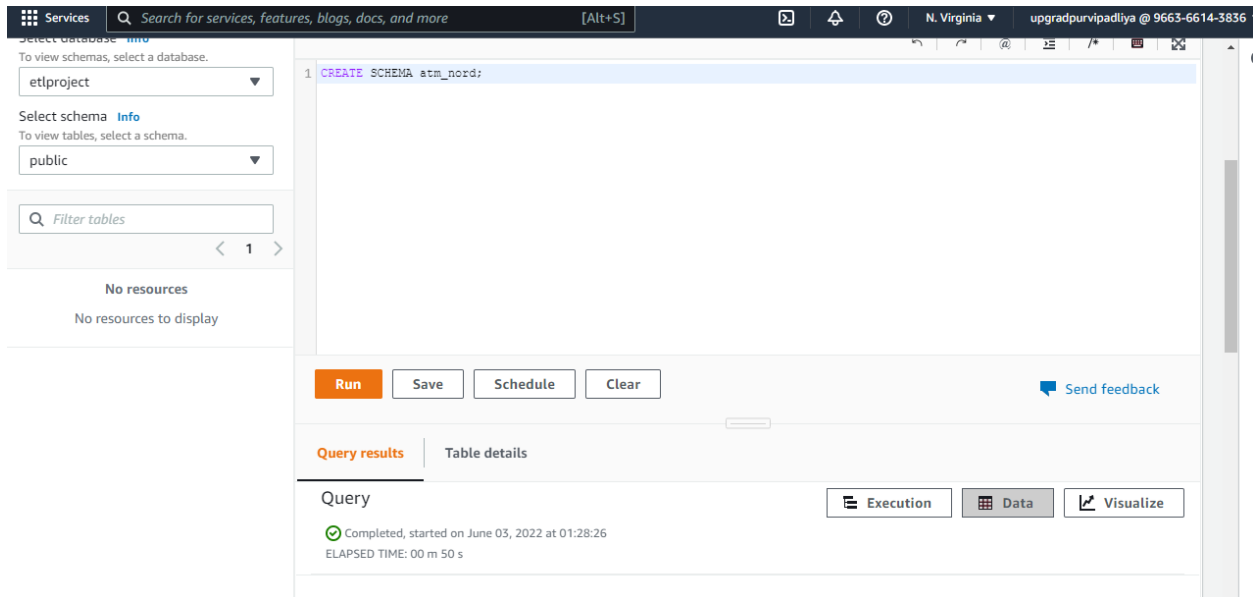
Cancel

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:

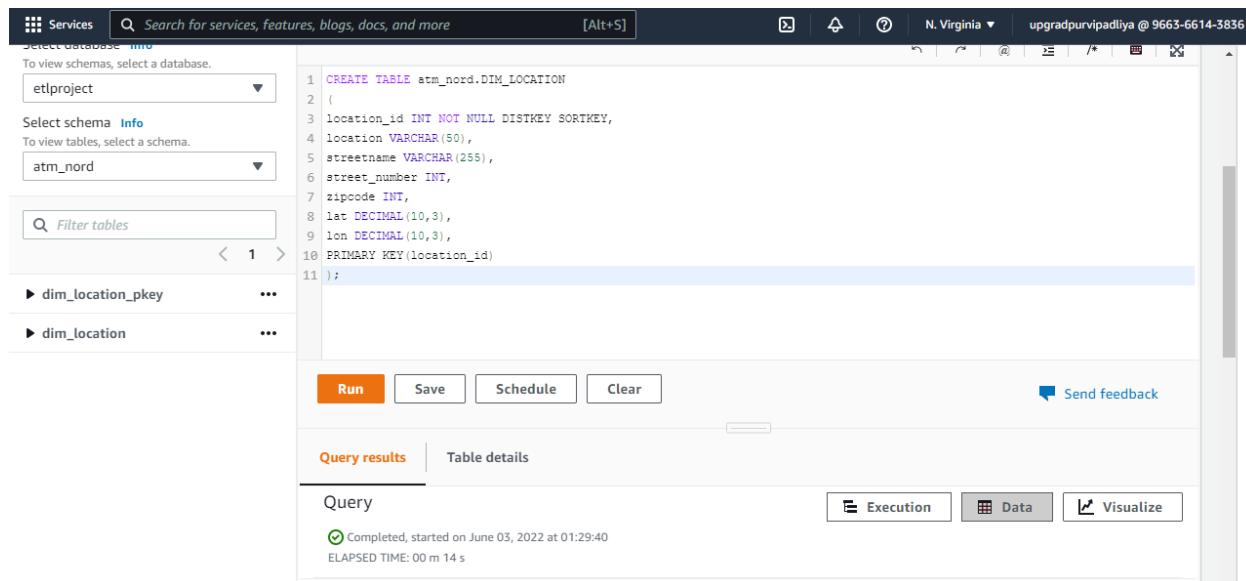
CREATE SCHEMA atm_nord;



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)
);
```



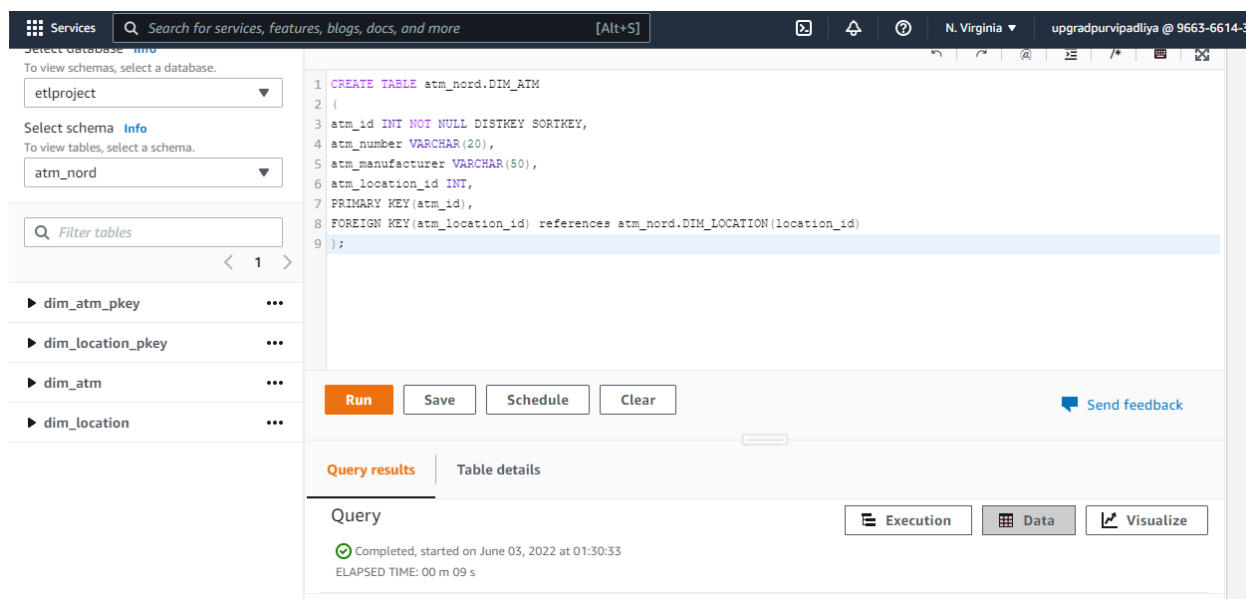
The screenshot shows the upGrad SQL editor interface. On the left, the 'Services' panel is open, showing the 'etlproject' database and the 'atm_nord' schema. Below the schema, a list of tables is shown, including 'dim_location_pkey' and 'dim_location'. The main editor area contains the following SQL query:

```
1 CREATE TABLE atm_nord.DIM_LOCATION
2 (
3   location_id INT NOT NULL DISTKEY SORTKEY,
4   location VARCHAR(50),
5   streetname VARCHAR(255),
6   street_number INT,
7   zipoode INT,
8   lat DECIMAL(10,3),
9   lon DECIMAL(10,3),
10  PRIMARY KEY(location_id)
11 );
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing the query execution status: 'Completed, started on June 03, 2022 at 01:29:40' and 'ELAPSED TIME: 00 m 14 s'.

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)
);
```



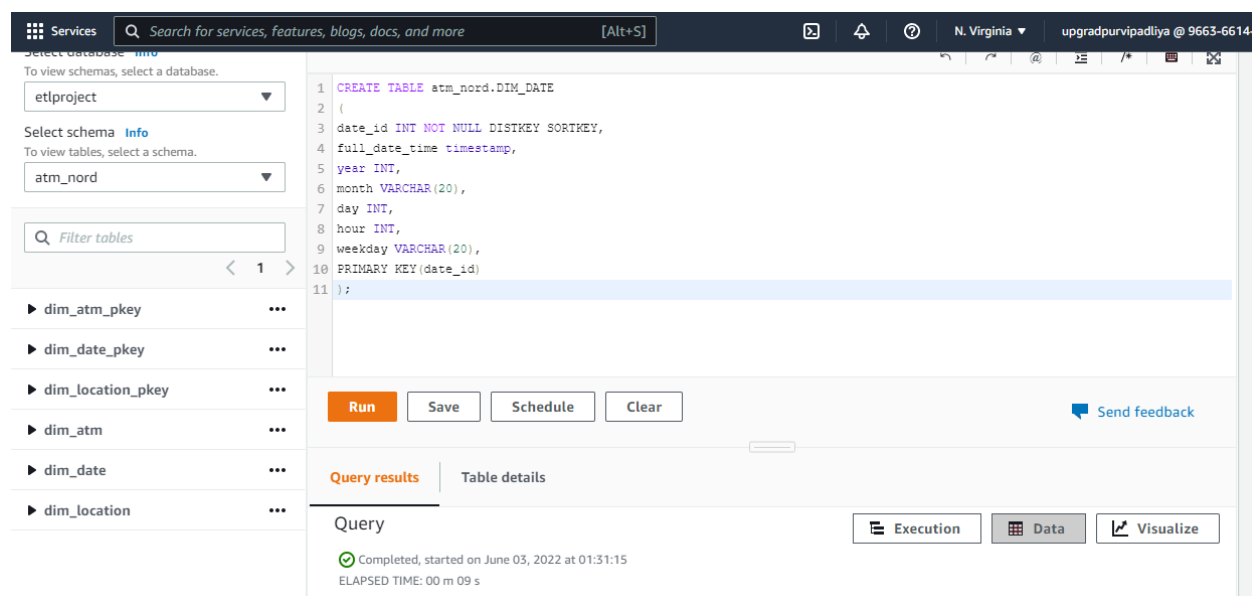
The screenshot shows the upGrad SQL editor interface. On the left, the 'Services' panel is open, showing the 'etlproject' database and the 'atm_nord' schema. Below the schema, a list of tables is shown, including 'dim_atm_pkey', 'dim_location_pkey', 'dim_atm', and 'dim_location'. The main editor area contains the following SQL query:

```
1 CREATE TABLE atm_nord.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_nord.DIM_LOCATION(location_id)
9 );
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing the query execution status: 'Completed, started on June 03, 2022 at 01:30:33' and 'ELAPSED TIME: 00 m 09 s'.

Query to create the dimension table DIM_DATE:

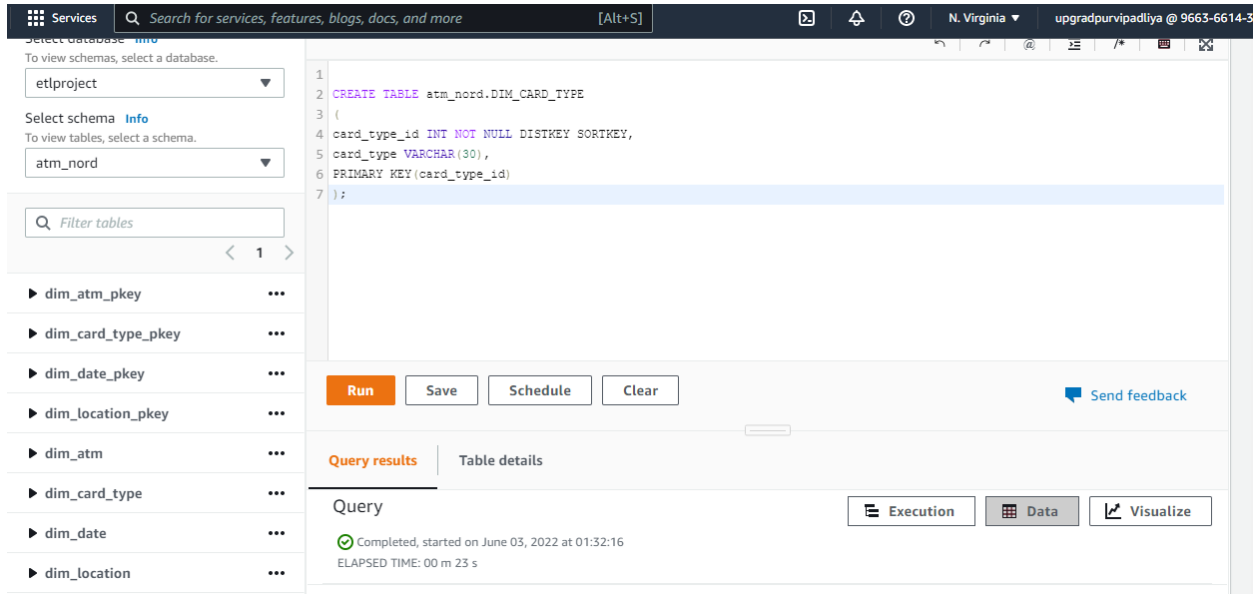
```
CREATE TABLE atm_nord.DIM_DATE
(
date_id INT NOT NULL DISTKEY SORTKEY,
full_date_time timestamp,
year INT,
month VARCHAR(20),
day INT,
hour INT,
weekday VARCHAR(20),
PRIMARY KEY(date_id)
);
```



The screenshot shows a SQL query editor interface. On the left, there's a sidebar with a search bar and a list of tables: dim_atm_pkey, dim_date_pkey, dim_location_pkey, dim_atm, dim_date, and dim_location. The main area displays the SQL query for creating the DIM_DATE table. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows the query execution status: 'Completed, started on June 03, 2022 at 01:31:15' and 'ELAPSED TIME: 00 m 09 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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)
);
```



The screenshot shows the upGrad SQL editor interface. On the left, there's a sidebar with a search bar and a list of tables under the 'atm_nord' schema. The main area displays a SQL query to create a table named 'DIM_CARD_TYPE' in the 'atm_nord' schema. The query is as follows:

```

1
2 CREATE TABLE atm_nord.DIM_CARD_TYPE
3 (
4   card_type_id INT NOT NULL DISTKEY SORTKEY,
5   card_type VARCHAR(30),
6   PRIMARY KEY(card_type_id)
7 );

```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of these buttons is a 'Send feedback' link. Below the buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a green checkmark icon and the text 'Completed, started on June 03, 2022 at 01:32:16' and 'ELAPSED TIME: 00 m 23 s'. To the right of the results, there are buttons for 'Execution', 'Data', and 'Visualize'.

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)
);

```

Services Search for services, features, blogs, docs, and more [Alt+S]

Select database [Info](#)
To view schemas, select a database.
etlproject

Select schema [Info](#)
To view tables, select a schema.
atm_nord

Filter tables

1 2

- ▶ dim_atm_pkey
- ▶ dim_card_type_pkey
- ▶ dim_date_pkey
- ▶ dim_location_pkey
- ▶ fact_atm_trans_pkey
- ▶ dim_atm
- ▶ dim_card_type
- ▶ dim_date

```

1 CREATE TABLE atm_nord.FACT_ATM_TRANS
2 (
3   trans_id BIGINT NOT NULL DISTKEY SORTKEY,
4   atm_id INT,
5   weather_loc_id INT,
6   date_id INT,
7   card_type_id INT,
8   atm_status VARCHAR(20),
9   currency VARCHAR(10),
10  service VARCHAR(20),
11  transaction_amount INT,
12  message_code VARCHAR(255),
13  message_text VARCHAR(255),
14  rain_3h DECIMAL(10,3),
15  clouds_all INT,

```

Run Save Schedule Clear

Send feedback

Query results Table details

Query

Execution Data Visualize

Completed, started on June 03, 2022 at 01:34:03
ELAPSED TIME: 00 m 35 s

Services Search for services, features, blogs, docs, and more [Alt+S]

Select database [Info](#)
To view schemas, select a database.
etlproject

Select schema [Info](#)
To view tables, select a schema.
atm_nord

Filter tables

1 2

- ▶ dim_atm_pkey
- ▶ dim_card_type_pkey
- ▶ dim_date_pkey
- ▶ dim_location_pkey
- ▶ fact_atm_trans_pkey
- ▶ dim_atm
- ▶ dim_card_type
- ▶ dim_date

```

12 message_code VARCHAR(255),
13 message_text VARCHAR(255),
14 rain_3h DECIMAL(10,3),
15 clouds_all INT,
16 weather_id INT,
17 weather_main VARCHAR(50),
18 weather_description VARCHAR(255),
19 PRIMARY KEY(trans_id),
20 FOREIGN KEY(weather_loc_id) references atm_nord.DIM_LOCATION(location_id),
21 FOREIGN KEY(atm_id) references atm_nord.DIM_ATM(atm_id),
22 FOREIGN KEY(date_id) references atm_nord.DIM_DATE(date_id),
23 FOREIGN KEY(card_type_id) references atm_nord.DIM_CARD_TYPE(card_type_id)
24 );

```

Run Save Schedule Clear

Send feedback

Query results Table details

Query

Execution Data Visualize

Completed, started on June 03, 2022 at 01:34:44
ELAPSED TIME: 01 m 16 s

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

Amazon S3
> Buckets
> etlprojectpurvi

etlprojectpurvi
Info

Objects
Properties
Permissions
Metrics
Management
Access Points

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](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dim_atm/	Folder	-	-	-
<input type="checkbox"/>	dim_card_type/	Folder	-	-	-
<input type="checkbox"/>	dim_date/	Folder	-	-	-
<input type="checkbox"/>	dim_location/	Folder	-	-	-
<input type="checkbox"/>	fact_atm_trans/	Folder	-	-	-

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

Query to copy the data from S3 bucket – etlprojectpurvi to dim_location table

```
copy atm_nord.dim_location from
's3://etlprojectpurvi/dim_location/part-00000-8b082841-91d1-4393-915d-278bdc37e006-
c000.csv'
iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1' CSV;
```

Services

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

[Alt+S]

N. Virginia

upgradpurvipadiya @ 9663-6614-3836

Select database [Info](#)

To view schemas, select a database.

etlproject

Select schema [Info](#)

To view tables, select a schema.

atm_nord

Filter tables

dim_atm_pkey

dim_card_type_pkey

dim_date_pkey

dim_location_pkey

fact_atm_trans_pkey

dim_atm

dim_card_type

dim_date

1 copy atm_nord.dim_location from

2 's3://etlprojectpurvi/dim_location/part-00000-8b082841-91d1-4393-915d-278bdc37e006-c000.csv'

3 iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'

4 delimiter ',' region 'us-east-1' CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

Visualize

Completed, started on June 03, 2022 at 01:39:54

ELAPSED TIME: 00 m 28 s

Query to copy the data from S3 bucket – etlprojectpurvi to dim_atm table

copy atm_nord.dim_atm from

's3://etlprojectpurvi/dim_atm/part-00000-922a116f-ad03-4301-bcc0-8ce172760ae5-c000.csv'

iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'

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

Services

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

[Alt+S]

N. Virginia

upgradpurvipadiya @ 9663-6614-3836

Select database [Info](#)

To view schemas, select a database.

etlproject

Select schema [Info](#)

To view tables, select a schema.

atm_nord

Filter tables

dim_atm_pkey

dim_card_type_pkey

dim_date_pkey

dim_location_pkey

fact_atm_trans_pkey

dim_atm

dim_card_type

dim_date

1 copy atm_nord.dim_atm from

2 's3://etlprojectpurvi/dim_atm/part-00000-922a116f-ad03-4301-bcc0-8ce172760ae5-c000.csv'

3 iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'

4 delimiter ',' region 'us-east-1' CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

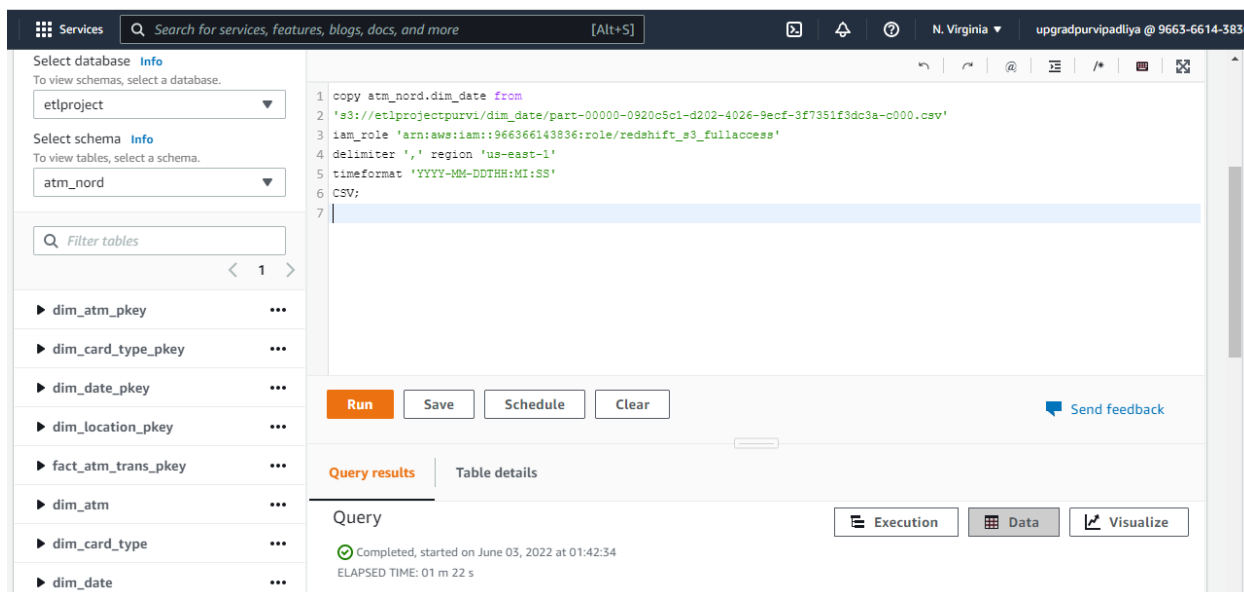
Visualize

Completed, started on June 03, 2022 at 01:41:24

ELAPSED TIME: 00 m 12 s

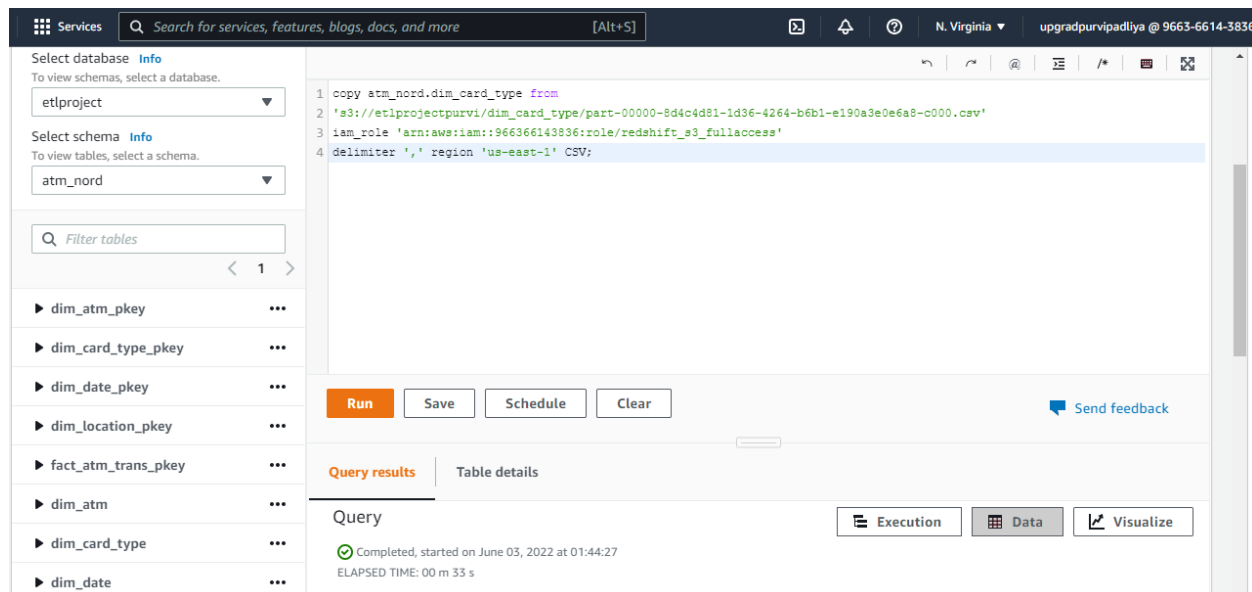
Query to copy the data from S3 bucket – etlprojectpurvi to dim_date table

```
copy atm_nord.dim_date from
's3://etlprojectpurvi/dim_date/part-00000-0920c5c1-d202-4026-9ecf-3f7351f3dc3a-c000.csv'
iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



Query to copy the data from S3 bucket – etlprojectpurvi to dim_card_type table

```
copy atm_nord.dim_card_type from
's3://etlprojectpurvi/dim_card_type/part-00000-8d4c4d81-1d36-4264-b6b1-e190a3e0e6a8-c000.csv'
iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1' CSV;
```



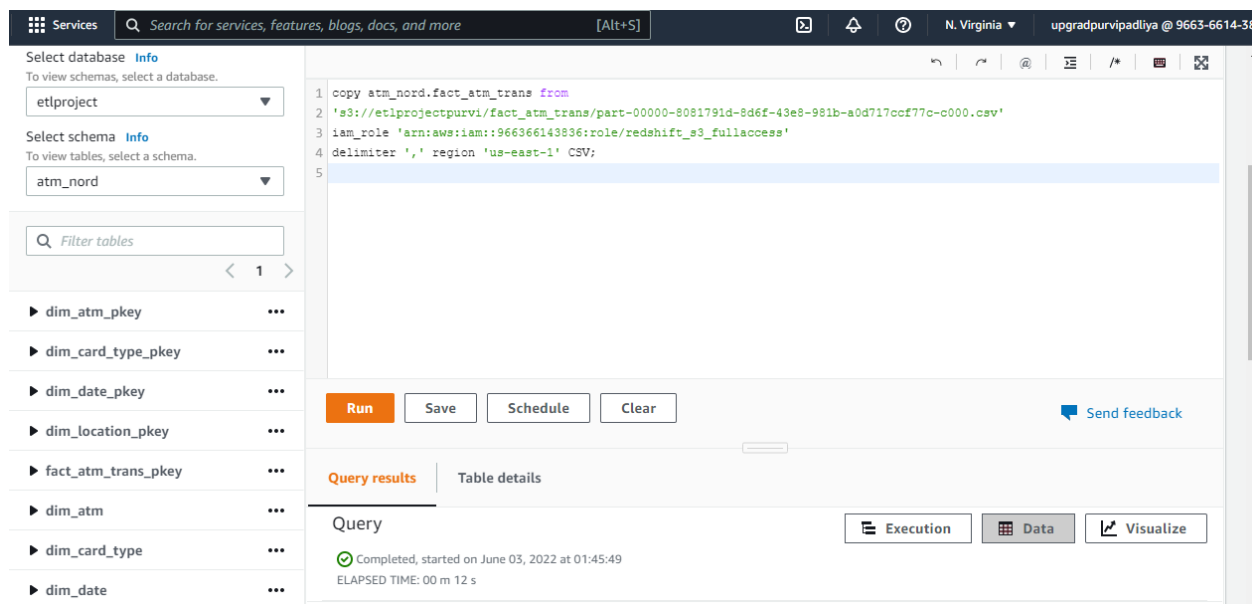
The screenshot shows the AWS Redshift console interface. On the left, the 'Select database' dropdown is set to 'etlproject' and the 'Select schema' dropdown is set to 'atm_nord'. Below these, a list of tables is visible, including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', and 'dim_date'. The main query editor contains the following SQL code:

```
1 copy atm_nord.dim_card_type from
2 's3://etlprojectpurvi/dim_card_type/part-00000-8d4c4d81-1d36-4264-b6b1-e190a3e0e6a8-c000.csv'
3 iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
4 delimiter ',' region 'us-east-1' CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link. Below the buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating the query was completed. The status bar shows 'Completed, started on June 03, 2022 at 01:44:27' and 'ELAPSED TIME: 00 m 33 s'. On the right side of the 'Query results' tab, there are buttons for 'Execution', 'Data', and 'Visualize'.

Query to copy the data from S3 bucket – etlprojectpurvi to fact_atm_trans table

```
copy atm_nord.fact_atm_trans from
's3://etlprojectpurvi/fact_atm_trans/part-00000-8081791d-8d6f-43e8-981b-a0d717ccf77c-c000.csv'
iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1' CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, the 'Select database' dropdown is set to 'etlproject' and the 'Select schema' dropdown is set to 'atm_nord'. 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', and 'dim_date'. The main query editor contains the following SQL code:

```
1 copy atm_nord.fact_atm_trans from
2 's3://etlprojectpurvi/fact_atm_trans/part-00000-8081791d-8d6f-43e8-981b-a0d717ccf77c-c000.csv'
3 iam_role 'arn:aws:iam::966366143836:role/redshift_s3_fullaccess'
4 delimiter ',' region 'us-east-1' CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link. Below the buttons, there are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating the query was completed. The status bar shows 'Completed, started on June 03, 2022 at 01:45:49' and 'ELAPSED TIME: 00 m 12 s'. On the right side of the 'Query results' tab, there are buttons for 'Execution', 'Data', and 'Visualize'.