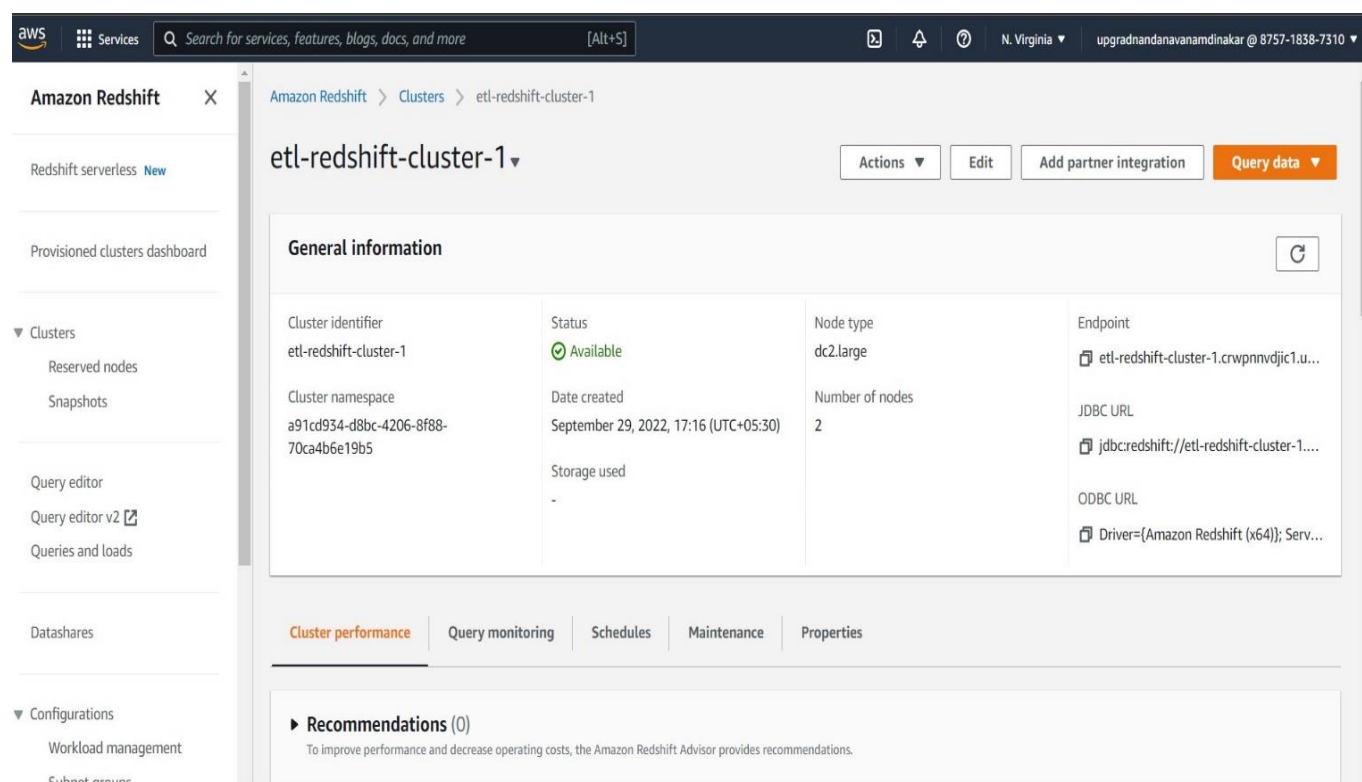


Creation of a Redshift Cluster

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

Below Screenshot shows the general information like type of node used and number of nodes used for setting up Redshift cluster.

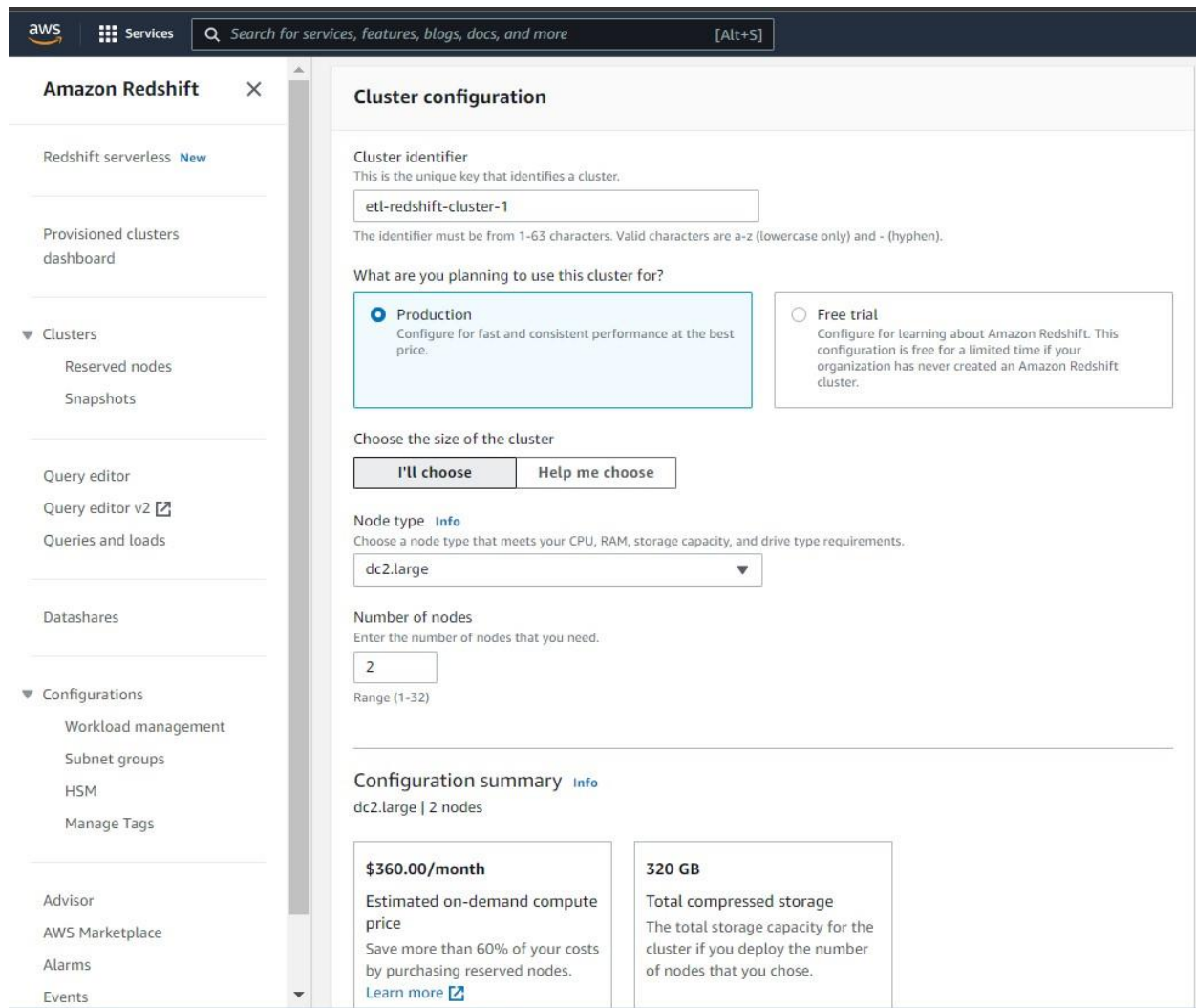


The screenshot displays the Amazon Redshift console interface. The left sidebar shows the navigation menu with options like 'Amazon Redshift', 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Workload management', and 'Subnet groups'. The main content area shows the details for the cluster 'etl-redshift-cluster-1'. The 'General information' tab is selected, displaying a table with the following data:

Cluster identifier	Status	Node type	Endpoint
etl-redshift-cluster-1	Available	dc2.large	etl-redshift-cluster-1.crwpmvjdjic1.u...
Cluster namespace	Date created	Number of nodes	JDBC URL
a91cd934-d8bc-4206-8f88-70ca4b6e19b5	September 29, 2022, 17:16 (UTC+05:30)	2	jdbcredshift://etl-redshift-cluster-1....
Storage used			ODBC URL
-			Driver={Amazon Redshift (x64)}; Serv...

Below the table, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'. The 'Cluster performance' tab is currently selected. At the bottom, there is a section for 'Recommendations (0)' with a note: 'To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.'

Screenshot of configurations made on cluster creation:

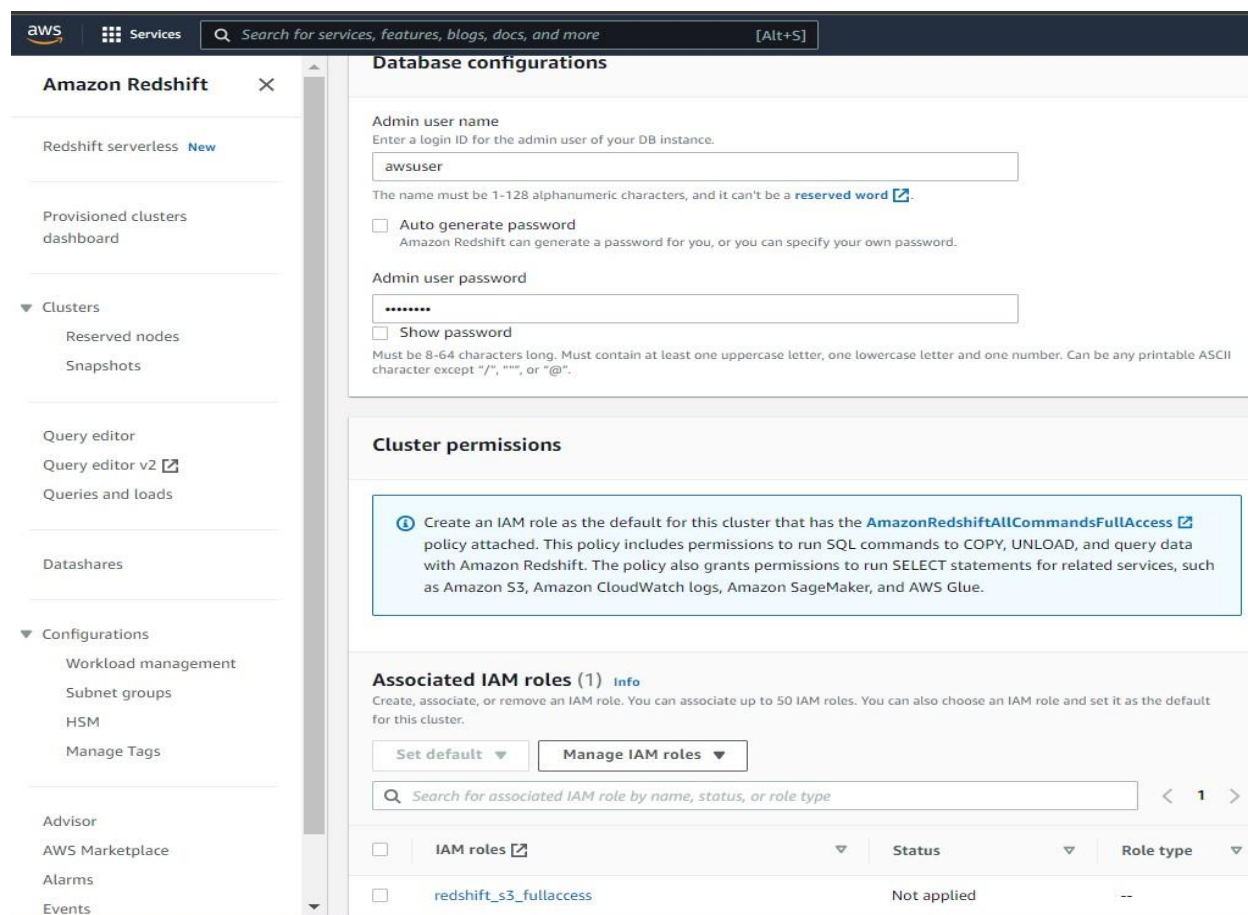


The screenshot shows the Amazon Redshift console interface for creating a new cluster. The left sidebar contains navigation links for Redshift serverless, Provisioned clusters dashboard, Clusters (Reserved nodes, Snapshots), Query editor, Query editor v2, Queries and loads, Datashares, Configurations (Workload management, Subnet groups, HSM, Manage Tags), Advisor, AWS Marketplace, Alarms, and Events.

The main content area is titled "Cluster configuration" and includes the following sections:

- Cluster identifier:** A text input field containing "etl-redshift-cluster-1". Below it, a note states: "The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen)."
 - ☒ **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:** Two buttons: "I'll choose" (selected) and "Help me choose".
- Node type:** A dropdown menu showing "dc2.large". A note below says: "Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements."
- Number of nodes:** A text input field containing "2". Below it, a note says: "Range (1-32)".
- Configuration summary:** A section titled "dc2.large | 2 nodes" with two summary boxes:
 - \$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.

Screenshot showing Database configurations and IAM roles associated with the cluster:



The screenshot displays the AWS Redshift console interface. On the left, a navigation pane shows the 'Amazon Redshift' section with options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', and 'Configurations'. The main content area is titled 'Database configurations' and includes sections for 'Admin user name', 'Admin user password', 'Cluster permissions', and 'Associated IAM roles (1)'. The 'Admin user name' field contains 'awsuser'. The 'Admin user password' field is masked with dots. The 'Cluster permissions' section contains a note about creating an IAM role with the 'AmazonRedshiftAllCommandsFullAccess' policy. The 'Associated IAM roles (1)' section shows a table with one role, 'redshift_s3_fullaccess', which is not applied.

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 "@".

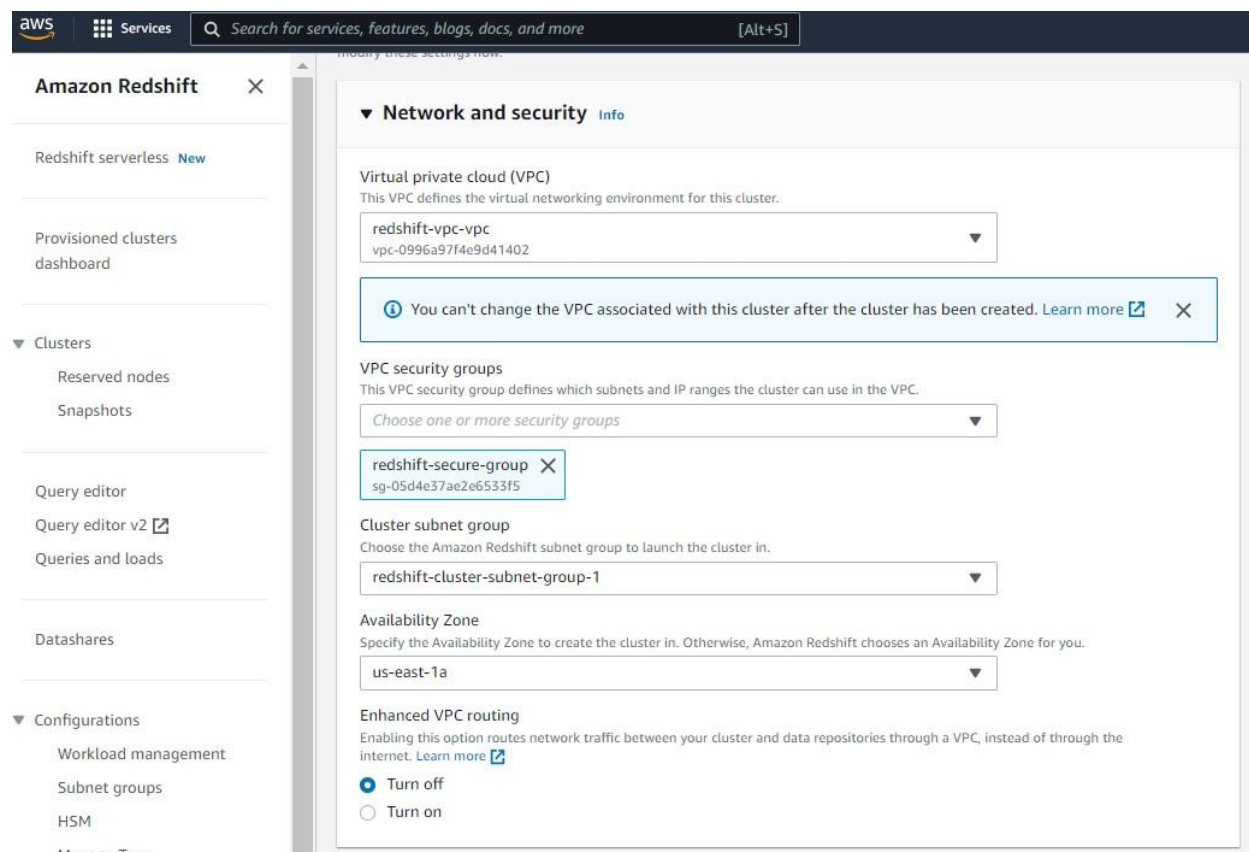
Cluster permissions

Associated IAM roles (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.

< 1 >

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


The following screenshots shows the other configurations made on cluster:



The screenshot displays the Amazon Redshift console interface. On the left, the navigation pane shows the 'Amazon Redshift' service with options for 'Redshift serverless' (marked as 'New'), 'Provisioned clusters dashboard', 'Clusters' (expanded), 'Reserved nodes', 'Snapshots', 'Query editor', 'Query editor v2', 'Queries and loads', 'Datashares', and 'Configurations' (expanded). Under 'Configurations', 'Workload management', 'Subnet groups', 'HSM', and 'Manage Tans' are listed. The main content area is titled 'Network and security' and contains the following settings:

- Virtual private cloud (VPC):** This VPC defines the virtual networking environment for this cluster. The selected VPC is 'redshift-vpc-vpc-0996a97f4e9d41402'.
- VPC security groups:** This VPC security group defines which subnets and IP ranges the cluster can use in the VPC. The selected security group is 'redshift-secure-group sg-05d4e37ae2e6533f5'.
- Cluster subnet group:** Choose the Amazon Redshift subnet group to launch the cluster in. The selected subnet group is 'redshift-cluster-subnet-group-1'.
- Availability Zone:** Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you. The selected zone is 'us-east-1a'.
- Enhanced VPC routing:** Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the internet. The option is currently set to 'Turn off'.

A warning message is displayed: 'You can't change the VPC associated with this cluster after the cluster has been created. Learn more'.



Services

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

[Alt+S]

N. Virginia

upgradnandanavanamdinakar @ 8757-1838-7310

Amazon Redshift

Redshift serverless

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Configurations

Workload management

Subnet groups

Database configurations

Change admin user password

Rotate encryption keys


Edit

Database name dev	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled AWS KMS key ID -	Audit logging Disabled
Port 5439	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAAD...		
Admin user name awsuser			

Network and security settings

Edit

Virtual private cloud (VPC) vpc-0996a97f4e9d41402	Availability Zone us-east-1a	VPC security group Specify which instances and devices can connect to the cluster. sg-05d4e37ae2e6533f5	Publicly accessible Allow instances and devices outside the VPC to connect to the database. Disabled
Subnet redshift-cluster-subnet-group-1	Enhanced VPC routing Disabled		
Endpoint URL -			



Services

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

[Alt+S]

N. Virginia

upgradnandanavanamdinakar @ 8757-1838-7310

Amazon Redshift

Redshift serverless

Provisioned clusters dashboard

Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Configurations

Workload management

Subnet groups

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)

Set default

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.

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

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	redshift_s3_fullaccess	in-sync	--

Granted accounts (0)

Edit

Revoke

Grant access

VPCs in other accounts that are allowed to access this cluster. [Learn more](#)

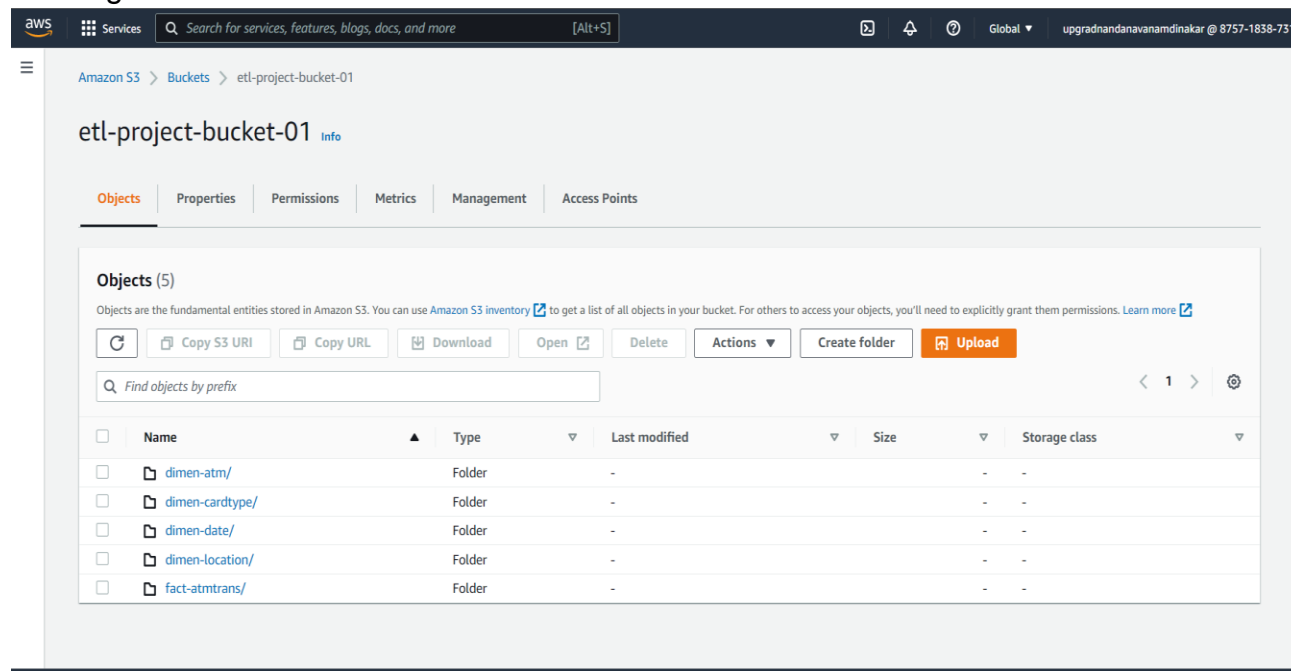
Find accounts or VPCs

Account ID	VPC	Endpoints created
------------	-----	-------------------

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

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:



Amazon S3 > Buckets > etl-project-bucket-01

etl-project-bucket-01 [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](#)

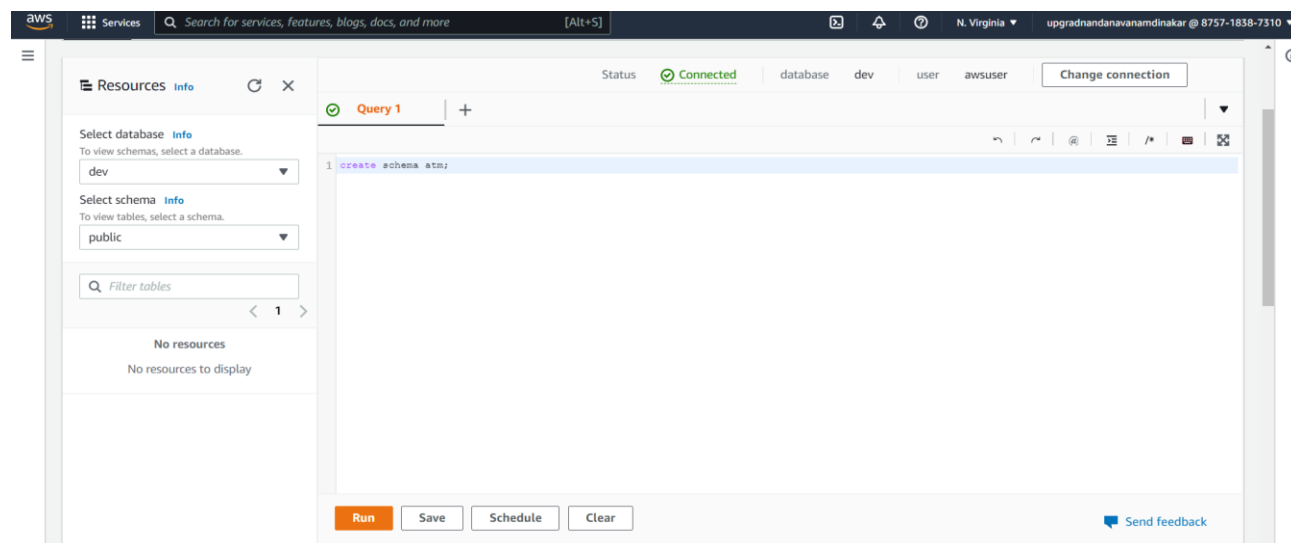
[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dimen-atm/	Folder	-	-	-
<input type="checkbox"/>	dimen-cardtype/	Folder	-	-	-
<input type="checkbox"/>	dimen-date/	Folder	-	-	-
<input type="checkbox"/>	dimen-location/	Folder	-	-	-
<input type="checkbox"/>	fact-atmtrans/	Folder	-	-	-

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

Query to create a schema for the dimension and fact tables:

create schema atm;



Resources [Info](#) [Refresh](#) [Close](#)

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

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

< 1 >

No resources
No resources to display

Status [Connected](#) database dev user awsuser [Change connection](#)

Query 1 +

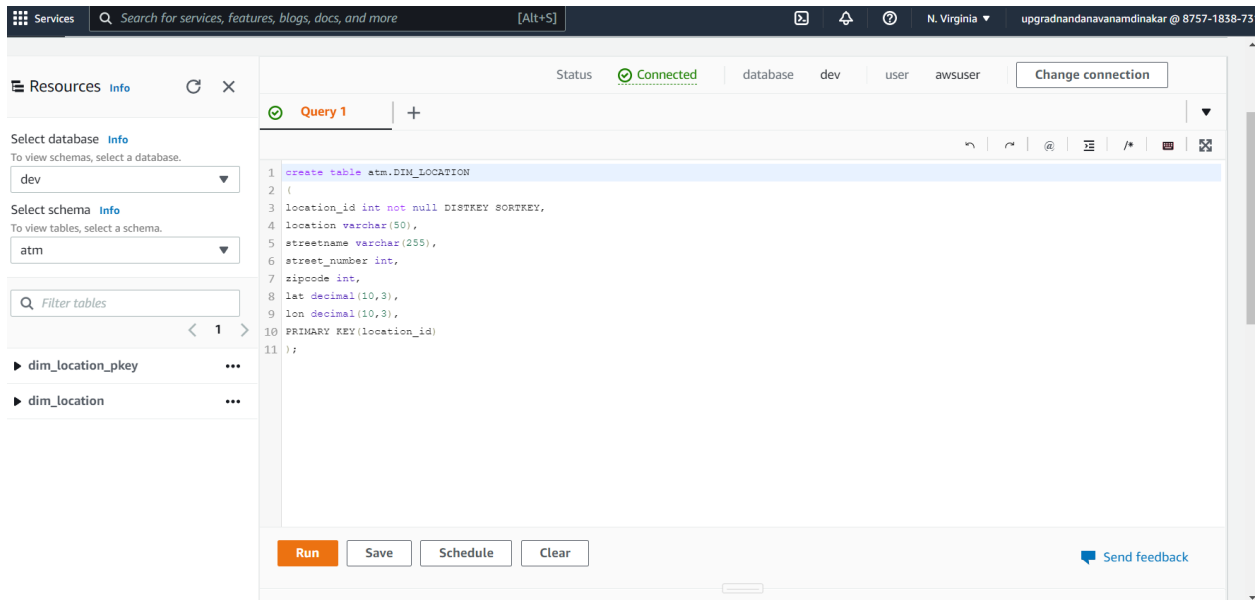
```
1 create schema atm;
```

[Run](#) [Save](#) [Schedule](#) [Clear](#)

[Send feedback](#)

Creating location dimension table:

```
create table atm.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.DIM_ATM

(

atm_id int not null DISTKEY SORTKEY,

atm_number varchar(20),

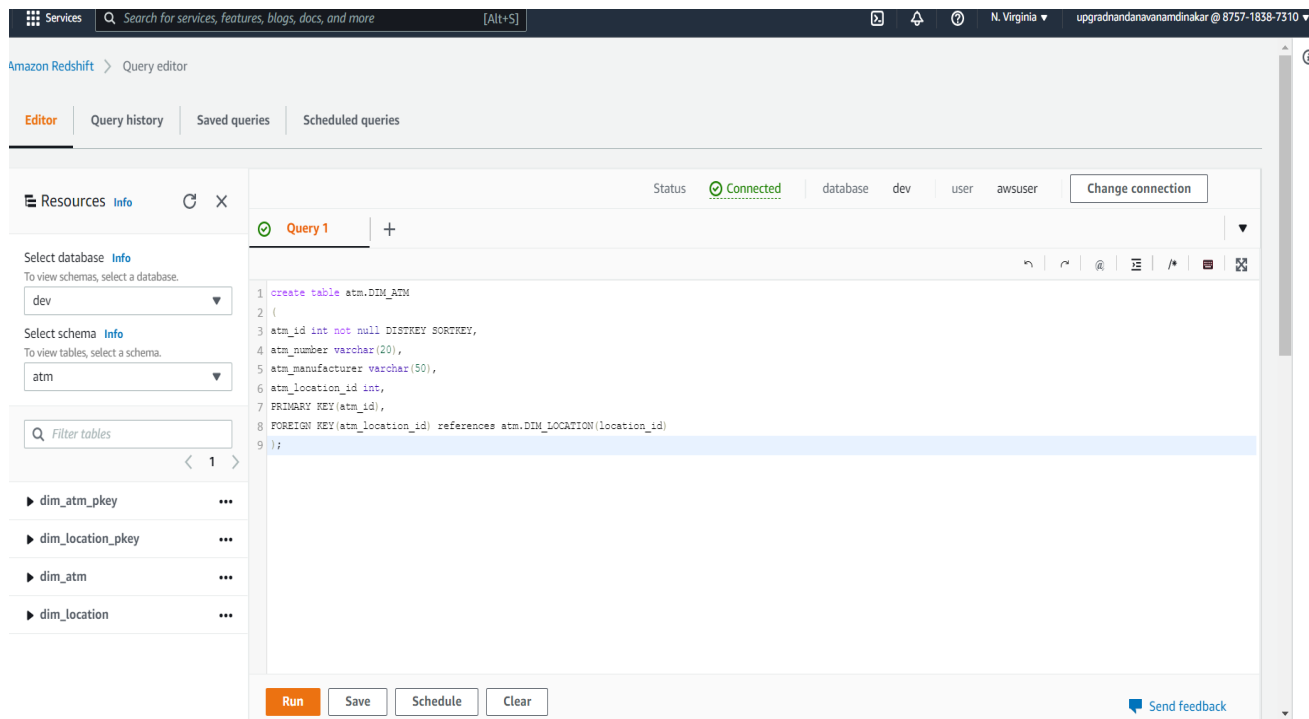
atm_manufacturer varchar(50),

atm_location_id int,

PRIMARY KEY(atm_id),

FOREIGN KEY(atm_location_id) references atm.DIM_LOCATION(location_id)

);



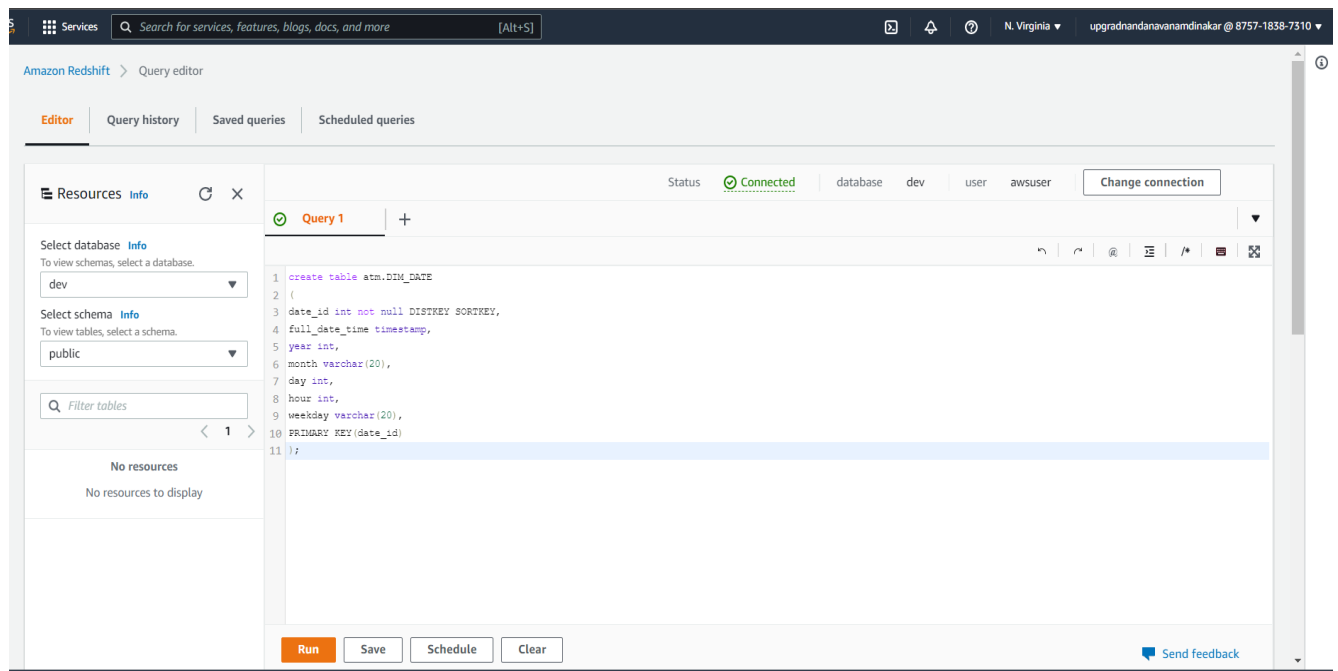
The screenshot displays the Amazon Redshift Query Editor interface. The top navigation bar includes a search bar and user information. The main workspace is titled "Query editor" and shows a SQL query for creating a table named "atm.DIM_ATM". The query is as follows:

```
1 create table atm.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.DIM_LOCATION(location_id)
9 );
```

The interface also shows a left sidebar with a "Resources" panel containing a list of databases and schemas. The "dev" database is selected, and the "atm" schema is chosen. A list of tables is visible, including "dim_atm_pkey", "dim_location_pkey", "dim_atm", and "dim_location". At the bottom of the query editor, there are buttons for "Run", "Save", "Schedule", and "Clear".

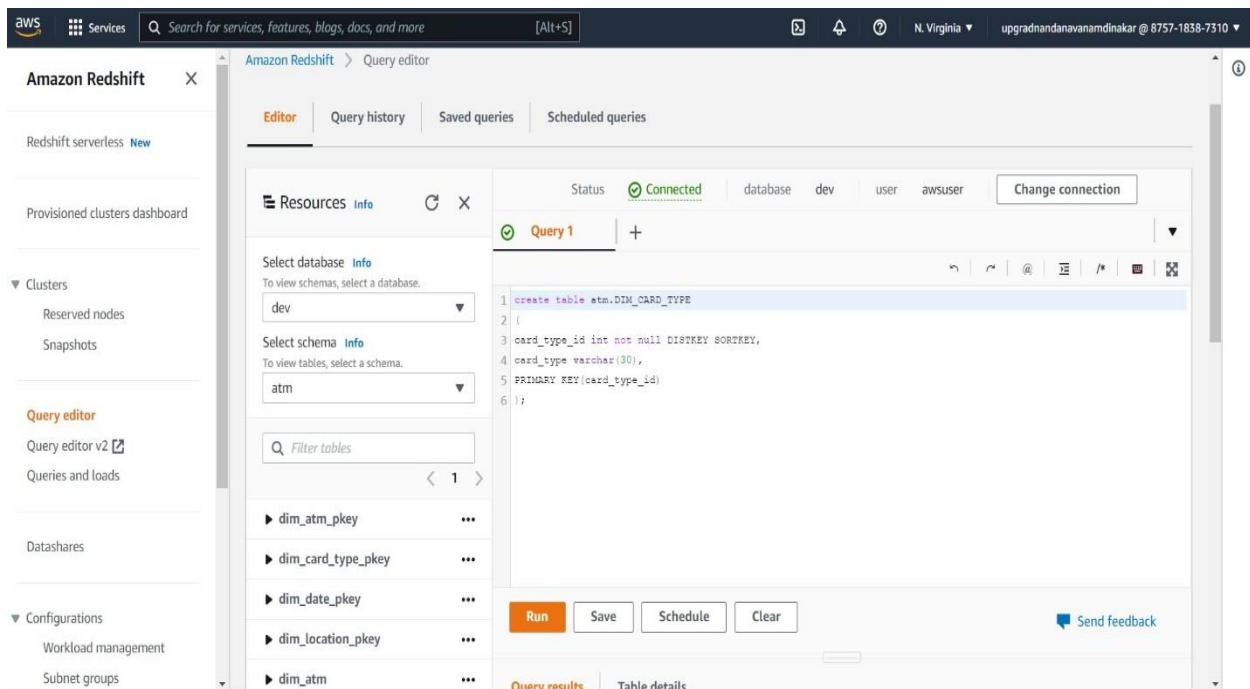
Creating date dimension table:

```
create table atm.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.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 navigation pane shows the 'Query editor' section with options like 'Query editor v2' and 'Queries and loads'. The main workspace is divided into several panels. The 'Resources' panel on the left shows the selected database 'dev' and schema 'atm'. The 'Query editor' panel on the right contains the SQL code for creating the 'DIM_CARD_TYPE' table. The code is as follows:

```
1 create table atm.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 code editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. The interface also shows a 'Status' bar at the top indicating 'Connected' and a 'Change connection' button.

Creating atm transactions fact table:

```
create table atm.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.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm.DIM_CARD_TYPE(card_type_id)
);
```

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

N. Virginia
upgradnandanavanamdinakar@8757-1838-7310

Amazon Redshift > Query editor

Editor
Query history
Saved queries
Scheduled queries

Resources
Info

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

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

Filter tables

1

dim_atm_pkey
dim_card_type_pkey
dim_date_pkey
dim_location_pkey
fact_atm_trans_pkey
dim_atm

Status
Connected
database
dev
user
awsuser
Change connection

Query 1
+

```

1 create table atm.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(225),
13  message_text varchar(225),
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.DIM_LOCATION(location_id),
21  FOREIGN KEY(atm_id) references atm.DIM_ATM(atm_id)

```

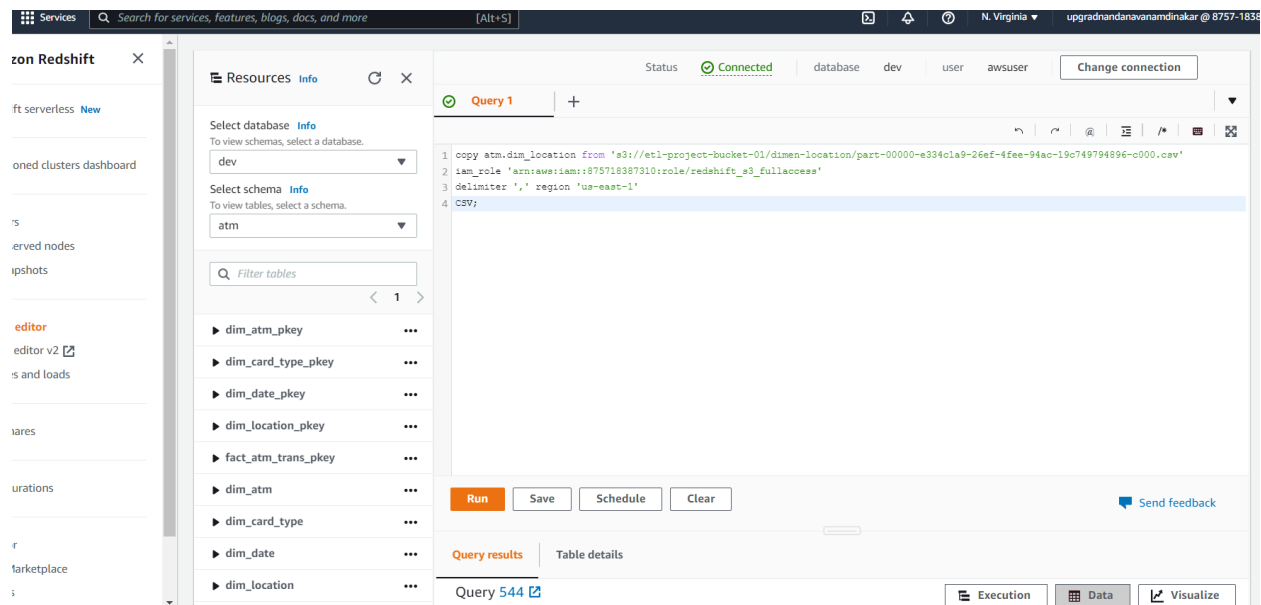
Run
Save
Schedule
Clear

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 from dimen-location folder on S3 bucket to dim_location table:

```
copy atm.dim_location from 's3://etl-project-bucket-01/dimen-location/part-00000-
e334c1a9-26ef-4fee-94ac-19c749794896-c000.csv'
iam_role 'arn:aws:iam::875718387310:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with navigation options like 'Resources', 'Info', 'dev', 'atm', 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', and 'dim_location'. The main area displays a SQL query in the 'Query 1' editor. The query is:
 1 copy atm.dim_location from 's3://etl-project-bucket-01/dimen-location/part-00000-e334c1a9-26ef-4fee-94ac-19c749794896-c000.csv'
 2 iam_role 'arn:aws:iam::875718387310: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'. At the bottom, there are tabs for 'Query results' and 'Table details', and a 'Query 544' link. The status bar at the bottom shows 'Execution', 'Data', and 'Visualize' options.

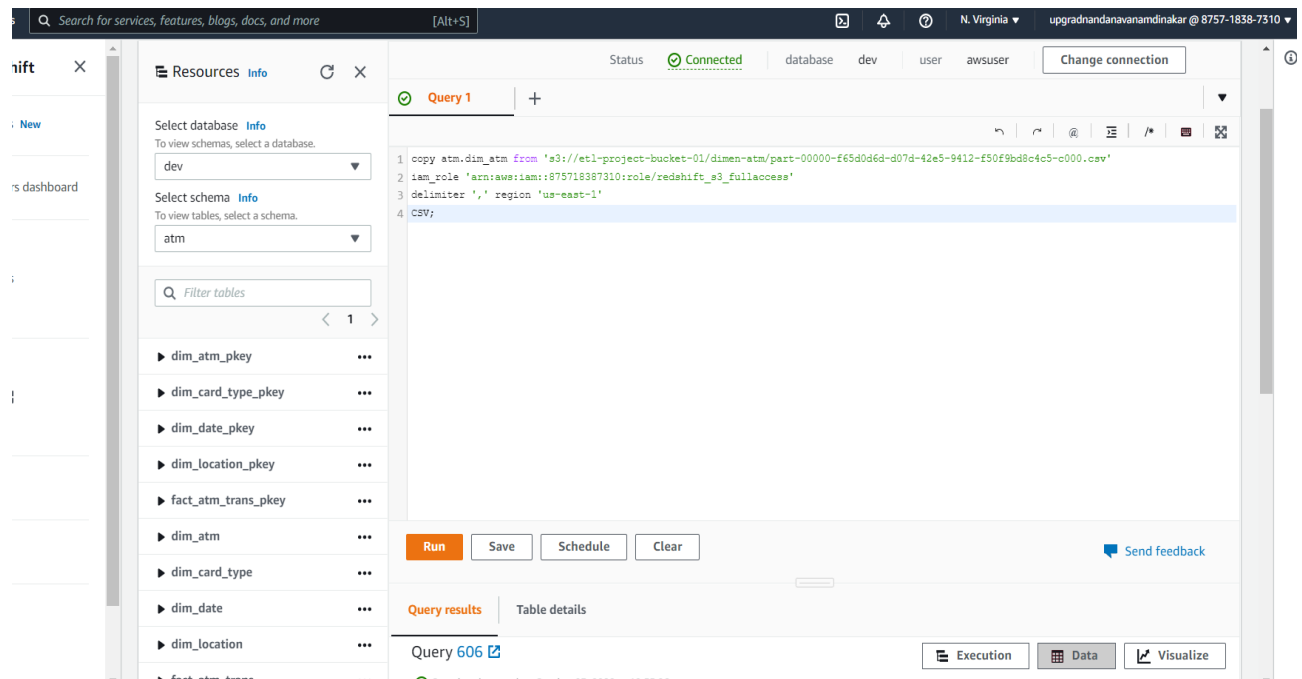
Copying the data from dimen-atm folder on S3 bucket to dim_atm table:

copy atm.dim_atm from 's3://etl-project-bucket-01/dimen-atm/part-00000-f65d0d6d-d07d-42e5-9412-f50f9bd8c4c5-c000.csv'

iam_role 'arn:aws:iam::875718387310: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 'dev' database and 'atm' schema. A list of tables is displayed, 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 shows a SQL query being executed:

```
1 copy atm.dim_atm from 's3://etl-project-bucket-01/dimen-atm/part-00000-f65d0d6d-d07d-42e5-9412-f50f9bd8c4c5-c000.csv'
2 iam_role 'arn:aws:iam::875718387310: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. At the bottom, there are tabs for 'Query results' and 'Table details', and a 'Query 606' label. The status bar at the bottom indicates 'Completed' and 'Execution' details.

Copying data from dimen-data folder from S3 bucket to dim_date table:

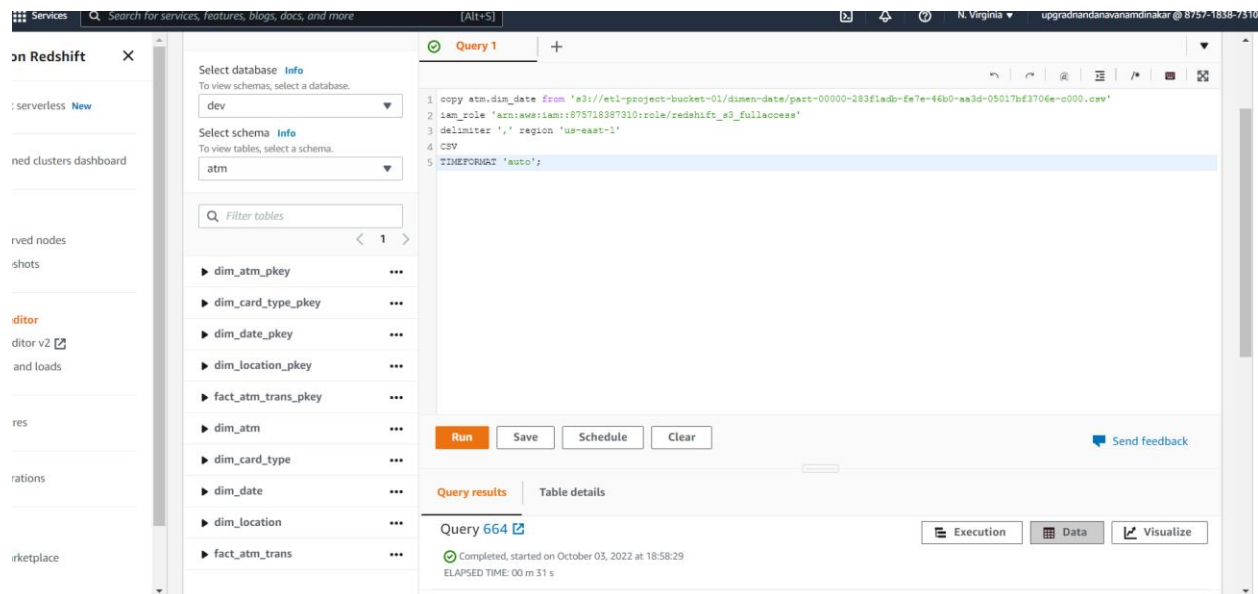
copy atm.dim_date from 's3://etl-project-bucket-01/dimen-date/part-00000-283f1adb-fe7e-46b0-aa3d-05017bf3706e-c000.csv'

iam_role 'arn:aws:iam::875718387310:role/redshift_s3_fullaccess'

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

CSV

TIMEFORMAT 'auto';



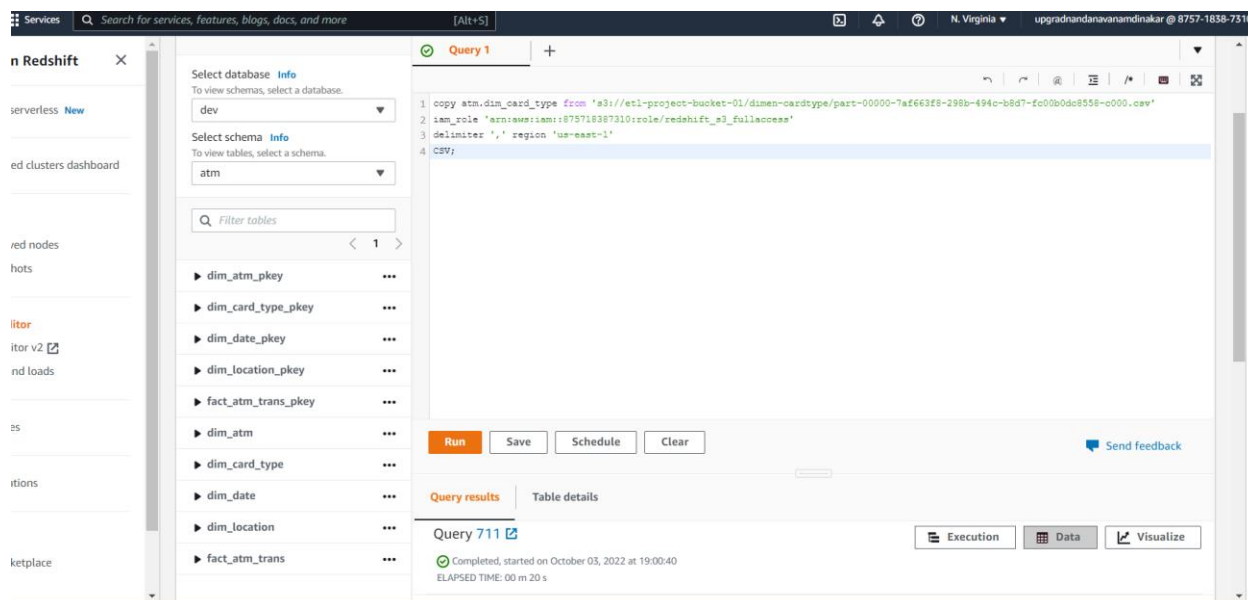
The screenshot shows the Amazon Redshift console interface. On the left, there is a sidebar with navigation options like 'serverless', 'clusters dashboard', 'nodes', 'shots', 'editor', 'editor v2', 'and loads', 'res', 'rations', and 'rketplace'. The main area is divided into two panes. The left pane shows the 'Select database' dropdown set to 'dev' and the 'Select schema' dropdown set to 'atm'. Below these, there is a list of tables with a search filter 'Filter tables'. The right pane shows a SQL query being executed. The query is as follows:

```
1 copy atm.dim_date from 's3://etl-project-bucket-01/dimen-date/part-00000-283f1adb-fe7e-46b0-aa3d-05017bf3706e-c000.csv'
2 iam_role 'arn:aws:iam::875718387310:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below the buttons, there is a section for 'Query results' and 'Table details'. The 'Query results' section shows the query ID 'Query 664' and the status 'Completed, started on October 03, 2022 at 18:58:29'. The 'ELAPSED TIME' is '00 m 31 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Copying data from dimen-cardtype folder from S3 to dim_card_type table:

```
copy atm.dim_card_type from 's3://etl-project-bucket-01/dimen-cardtype/part-00000-7af663f8-298b-494c-b8d7-fc00b0dc8558-c000.csv'
iam_role 'arn:aws:iam::875718387310:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



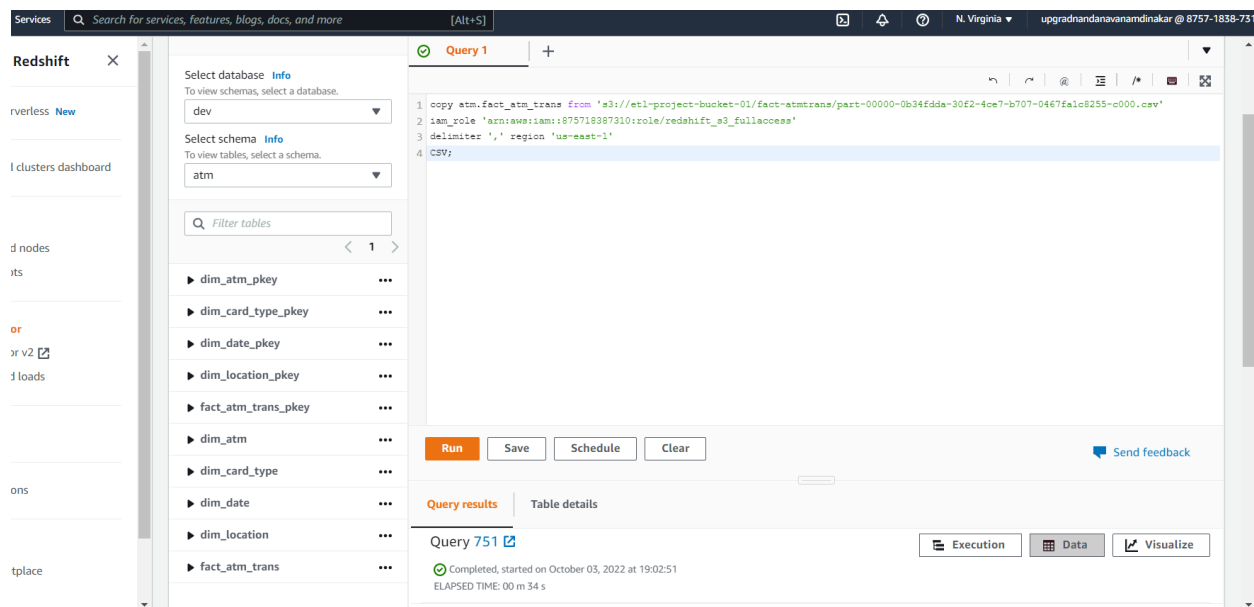
The screenshot shows the Amazon Redshift console interface. On the left, there is a sidebar with navigation options like 'serverless', 'ed clusters dashboard', 'red nodes', 'hotos', 'itor', 'itor v2', 'nd loads', 'es', 'rtions', and 'ketplace'. The main area is divided into two panels. The left panel shows the 'Select database' dropdown set to 'dev' and the 'Select schema' dropdown set to 'atm'. Below these, there is a 'Filter tables' search bar 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 right panel shows a SQL query editor with the following code:

```
1 copy atm.dim_card_type from 's3://etl-project-bucket-01/dimen-cardtype/part-00000-7af663f8-298b-494c-b8d7-fc00b0dc8558-c000.csv'
2 iam_role 'arn:aws:iam::875718387310: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 the buttons, there is a 'Query results' tab and a 'Table details' tab. The 'Query results' tab is active, showing 'Query 711' with a status of 'Completed, started on October 03, 2022 at 19:00:40' and 'ELAPSED TIME: 00 m 20 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Copying data from fact-atmtrans folder in S3 to fact_atm_trans table:

```
copy atm.fact_atm_trans from 's3://etl-project-bucket-01/fact-atmtrans/part-00000-0b34fdda-30f2-4ce7-b707-0467fa1c8255-c000.csv'
iam_role 'arn:aws:iam::875718387310:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the Amazon Redshift console interface. On the left, there is a sidebar with navigation options like 'Redshift', 'clusters dashboard', 'nodes', 'roles', 'groups', 'v2', 'loads', 'ons', and 'tplace'. The main area is divided into two panels. The left panel shows the 'Select database' dropdown set to 'dev' and the 'Select schema' dropdown set to 'atm'. Below these, there is a 'Filter tables' search bar 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 right panel shows a SQL query editor with the following query:

```
1 copy atm.fact_atm_trans from 's3://etl-project-bucket-01/fact-atmtrans/part-00000-0b34fdda-30f2-4ce7-b707-0467fa1c8255-c000.csv'
2 iam_role 'arn:aws:iam::875718387310: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 is a 'Query results' tab and a 'Table details' tab. The 'Query results' tab is active, showing the query ID 'Query 751' and the status 'Completed, started on October 03, 2022 at 19:02:51' with an 'ELAPSED TIME: 00 m 34 s'.