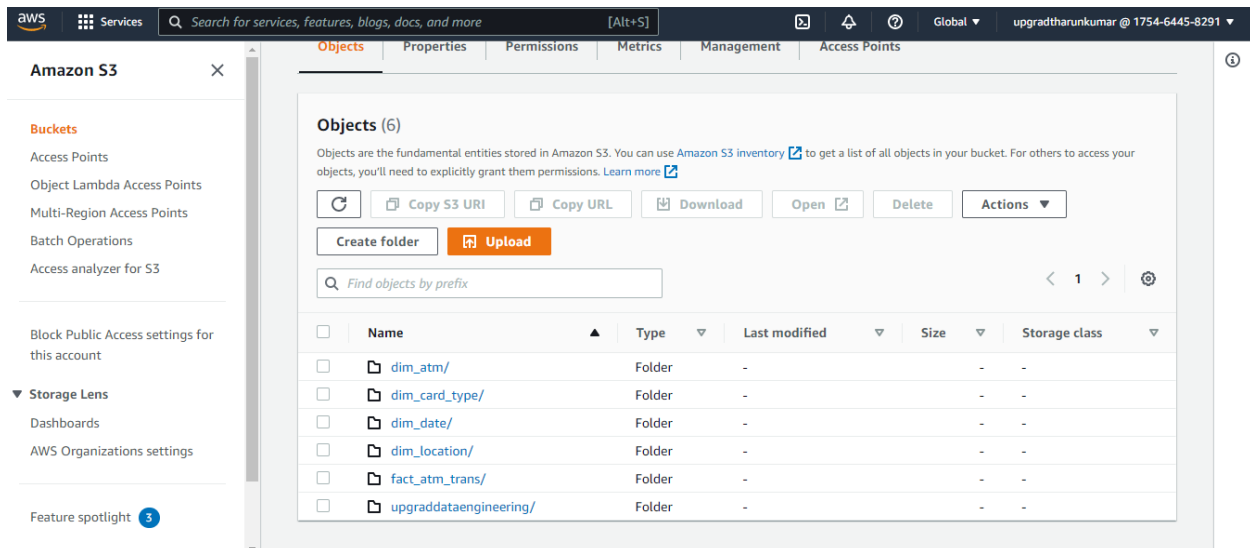
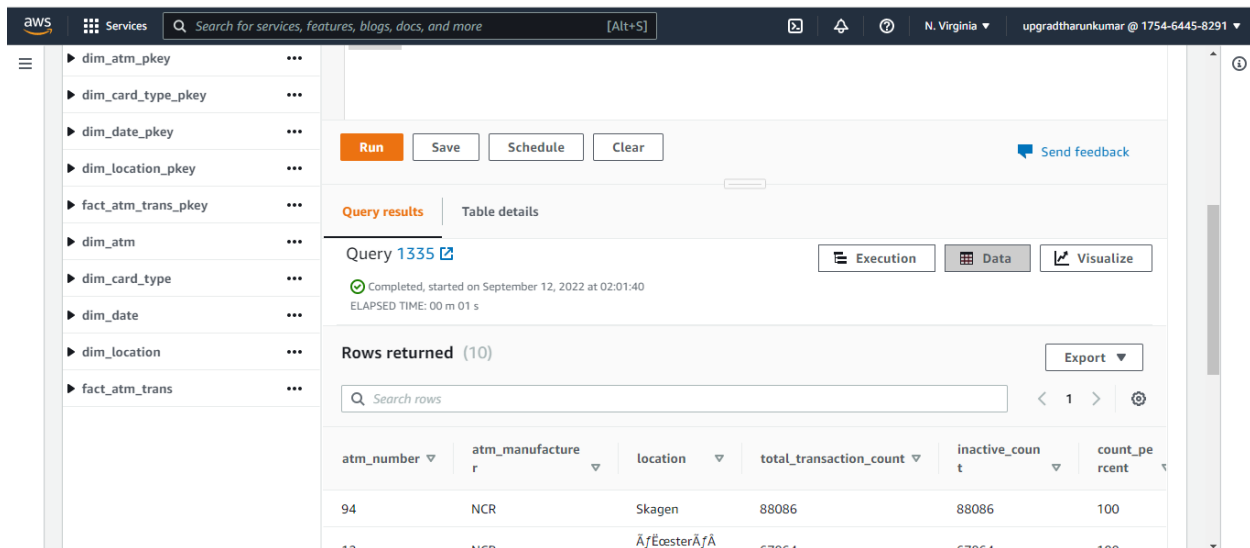


From Pyspark data from csv file made into a tabular form with a schema, then they are exported to the s3 buckets(as shown in the image). Data is stored in the form of CSV files.



From S3 buckets data is transferred to amazon redshift data base called dev which is configured while launching amazon redshift.



Creation of a Redshift Cluster

These following images are the screenshots of the configurations made while we launch the amazon redshift cluster named as upgrad.

The screenshot displays the AWS Management Console interface for an Amazon Redshift cluster named 'upgrad'. The top navigation bar includes the AWS logo, 'Services' link, a search bar, and user information for 'upgradtharunkumar' in the 'N. Virginia' region. The breadcrumb trail shows 'Amazon Redshift > Clusters > upgrad'. The cluster name 'upgrad' is prominently displayed with action buttons: 'Actions', 'Edit', 'Add partner integration', and 'Query data'.

General information

Cluster identifier upgrad	Status Available	Node type dc2.large	Endpoint upgrad.cstb06gyjnso.us-east-1.redshift...
Cluster namespace a1ed7bad-435b-4e98-89a3-520b7fcfdda	Date created September 12, 2022, 00:48 (UTC-04:00)	Number of nodes 2	JDBC URL jdbc:redshift://upgrad.cstb06gyjnso.us-...
	Storage used 0.14% (0.44 of 320 GB used)	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server...

Database configurations

Database name dev	Parameter group default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5439	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQ...	AWS KMS key ID -	

Network and security settings

Virtual private cloud (VPC) vpc-00e2cee8eff90b333	Availability Zone us-east-1a	VPC security group sg-09e930132af321c9a	Publicly accessible Disabled
Subnet cluster-subnet-group-upgrad	Enhanced VPC routing Disabled		
Endpoint URL -			

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) [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.

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

< 1 >

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	s3_fullaccess_tharun	in-sync	--

Integrations (0) [Delete](#) [Add partner integration](#)

Choose an AWS Partner to edit the integration on the partner's website. [Learn more](#)

Partner name	Status	Database	Last successful connection
No connections			

Node IP addresses (3)

Node role	Public IP address	Private IP address
Compute-0	34.236.137.255	10.0.7.110
Compute-1	54.157.86.149	10.0.1.148
Leader	3.220.44.38	10.0.7.39

Tags (0) [Manage tags](#)

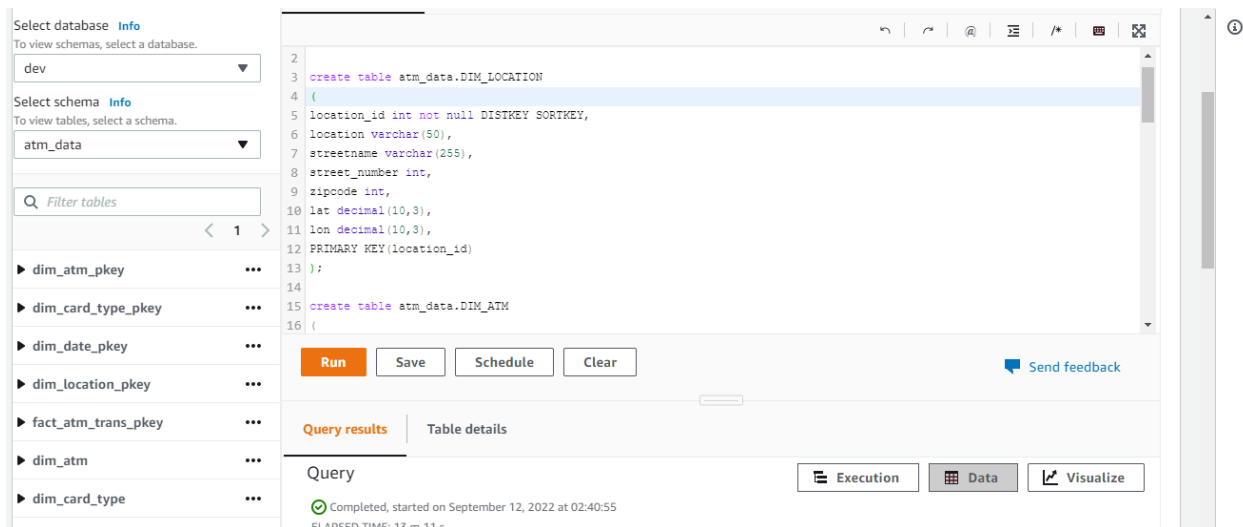
Key	Value
No tags	
No tags are associated with this cluster.	

[Add tags](#)

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

I created a atm_data schema and then I made the table using the following sql queries

```
create schema atm_data;
```



```
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)
);
```

```

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)
);

```

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)
);

```

15 create table atm_data.DIM_ATM
16 (
17 atm_id int not null DISTKEY SORTKEY,
18 atm_number varchar(20),
19 atm_manufacturer varchar(50),
20 atm_location_id int,
21 PRIMARY KEY(atm_id),
22 FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
23 );
24

```

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)
);

```
25 create table atm_data.DIM_DATE
26 (
27 date_id int not null DISTKEY SORTKEY,
28 full_date_time timestamp,
29 year int,
30 month varchar(20),
31 day int,
32 hour int,
33 weekday varchar(20),
34 PRIMARY KEY(date_id)
35 );
36
```

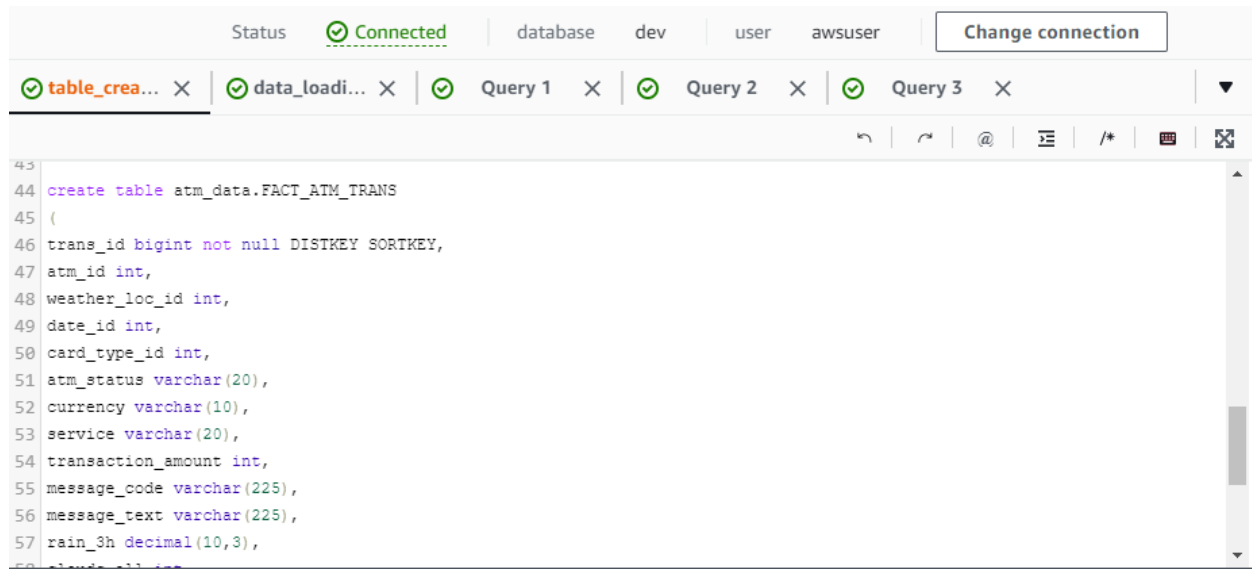
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);

```
37 create table atm_data.DIM_CARD_TYPE
38 (
39 card_type_id int not null DISTKEY SORTKEY,
40 card_type varchar(30),
41 PRIMARY KEY(card_type_id)
42 );
43
```

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 a Redshift SQL editor interface. At the top, there's a status bar indicating 'Connected' to a 'database' on 'dev' using 'awsuser'. Below this, there are tabs for 'table_crea...', 'data_loadi...', 'Query 1', 'Query 2', and 'Query 3'. The main editor area displays a SQL query starting with 'create table atm_data.FACT_ATM_TRANS' and listing various columns with their data types and constraints. The query is as follows:

```
43  
44 create table atm_data.FACT_ATM_TRANS  
45 (  
46 trans_id bigint not null DISTKEY SORTKEY,  
47 atm_id int,  
48 weather_loc_id int,  
49 date_id int,  
50 card_type_id int,  
51 atm_status varchar(20),  
52 currency varchar(10),  
53 service varchar(20),  
54 transaction_amount int,  
55 message_code varchar(225),  
56 message_text varchar(225),  
57 rain_3h decimal(10,3),  
58 ...
```

Loading data into a Redshift cluster from Amazon S3 bucket

These are the queries which I made to load the data into tables that I created as per above code.

- To copy data from S3 to dim_location table

copy atm_data.dim_location from

's3://dataengineeringtharun/dim_location/part-00000-1aee43a6-4f2e-45ef-b2d3-c17ea16d6dd0-c000.csv'

iam_role 'arn:aws:iam::175464458291:role/s3_fullaccess_tharun'

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

CSV;

- To copy data from S3 to dim_atm table

copy atm_data.dim_atm from

's3://dataengineeringtharun/dim_atm/part-00000-9885cf6d-75c6-45cc-acfc-a2e770487f9f-c000.csv'

iam_role 'arn:aws:iam::175464458291:role/s3_fullaccess_tharun'

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

CSV;

- To copy data from S3 to dim_date table

copy atm_data.dim_date from

's3://dataengineeringtharun/dim_date/part-00000-ee8df625-8c88-46ed-b3d0-038d80353f57-c000.csv'

iam_role 'arn:aws:iam::175464458291:role/s3_fullaccess_tharun'

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

CSV

TIMEFORMAT 'auto';

- To copy data from S3 to dim_card table

```
copy atm_data.dim_card_type from
's3://dataengineeringtharun/dim_card_type/part-00000-a527b9e5-
3253-4561-9d04-8689045fde03-c000.csv'

iam_role 'arn:aws:iam::175464458291:role/s3_fullaccess_tharun'
delimiter ',' region 'us-east-1'

CSV;
```

- To copy data from S3 to fact table

```
copy atm_data.fact_atm_trans from
's3://dataengineeringtharun/fact_atm_trans/part-00000-f51a20a7-
c4bc-408e-9f7f-961d8ebea103-c000.csv'

iam_role 'arn:aws:iam::175464458291:role/s3_fullaccess_tharun'
delimiter ',' region 'us-east-1'

CSV;
```