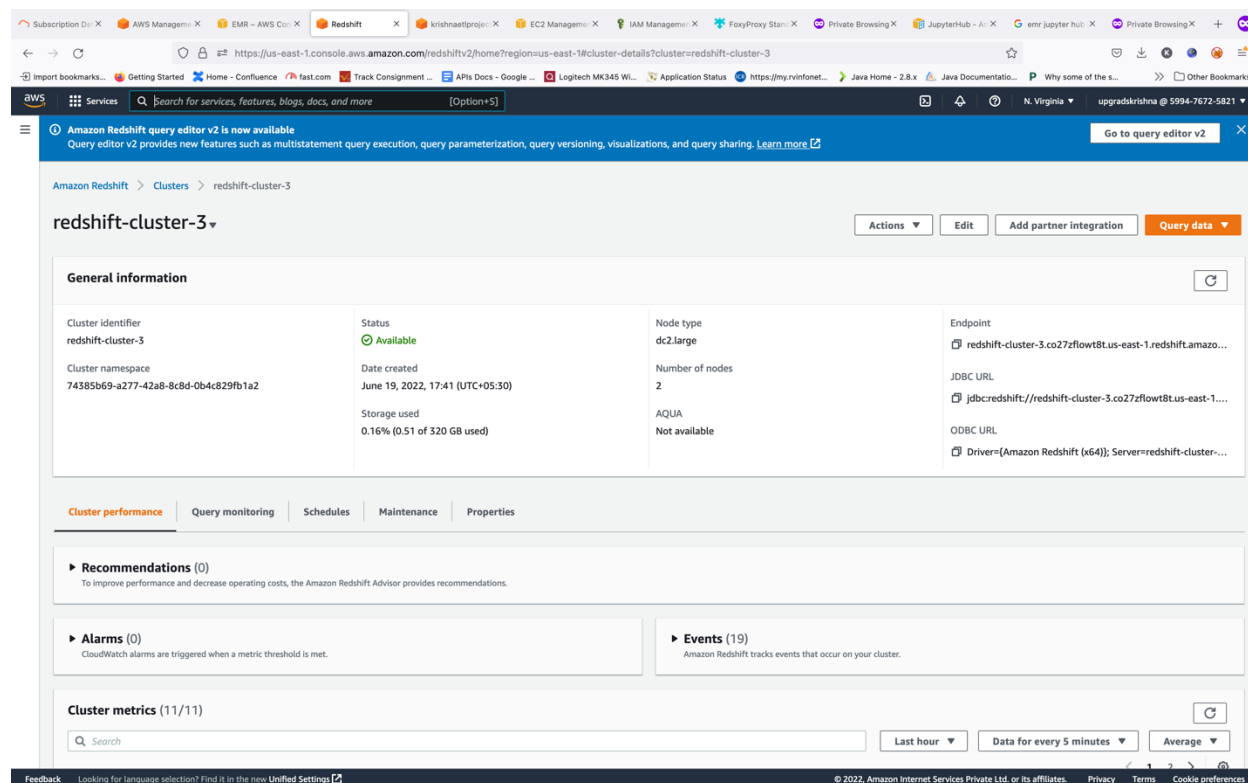


Creation of a Redshift Cluster

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

Screenshot of the type of machine used along with number of nodes:



The screenshot shows the Amazon Redshift console interface. The browser address bar indicates the URL: <https://us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#clusters:cluster=redshift-cluster-3>. The console displays the details for the cluster 'redshift-cluster-3'.

General information

Cluster identifier redshift-cluster-3	Status Available	Node type dc2.large	Endpoint redshift-cluster-3.co2zflowt8t.us-east-1.redshift.amazo...
Cluster namespace 74385b69-a277-42a8-8c8d-0b4c829fb1a2	Date created June 19, 2022, 17:41 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-3.co2zflowt8t.us-east-1....
	Storage used 0.16% (0.51 of 320 GB used)	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=redshift-cluster-...

Cluster performance | Query monitoring | Schedules | Maintenance | Properties

Recommendations (0)
To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

Alarms (0)
CloudWatch alarms are triggered when a metric threshold is met.

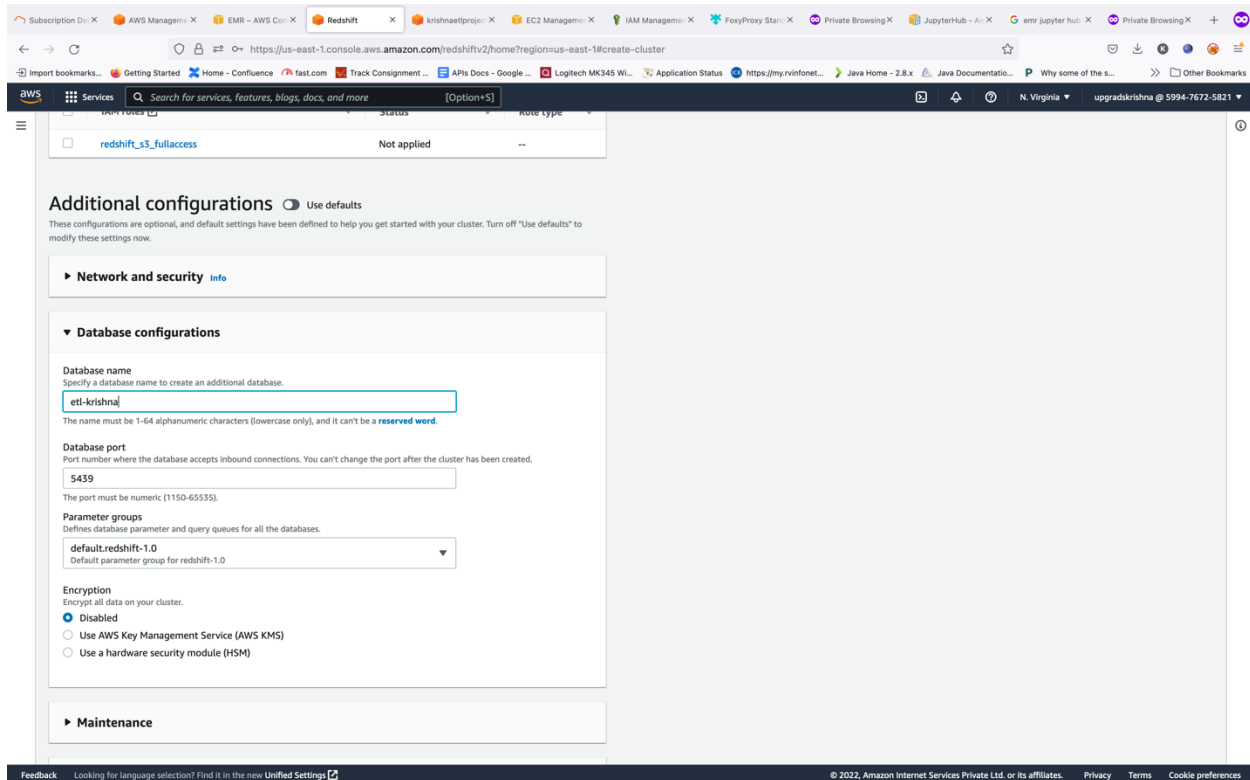
Events (19)
Amazon Redshift tracks events that occur on your cluster.

Cluster metrics (11/11)
Search: [] Last hour | Data for every 5 minutes | Average

Screenshot of steps taken to create the Redshift cluster:

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

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



The screenshot shows the AWS Redshift console interface. At the top, there's a navigation bar with the AWS logo and a search bar. Below it, a sidebar lists various services. The main content area is titled 'Additional configurations' and includes sections for 'Network and security', 'Database configurations', and 'Maintenance'. The 'Database configurations' section is expanded, showing fields for 'Database name' (eti-krishna), 'Database port' (5439), 'Parameter groups' (default.redshift-1.0), and 'Encryption' (Disabled).

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 [Info](#)

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)

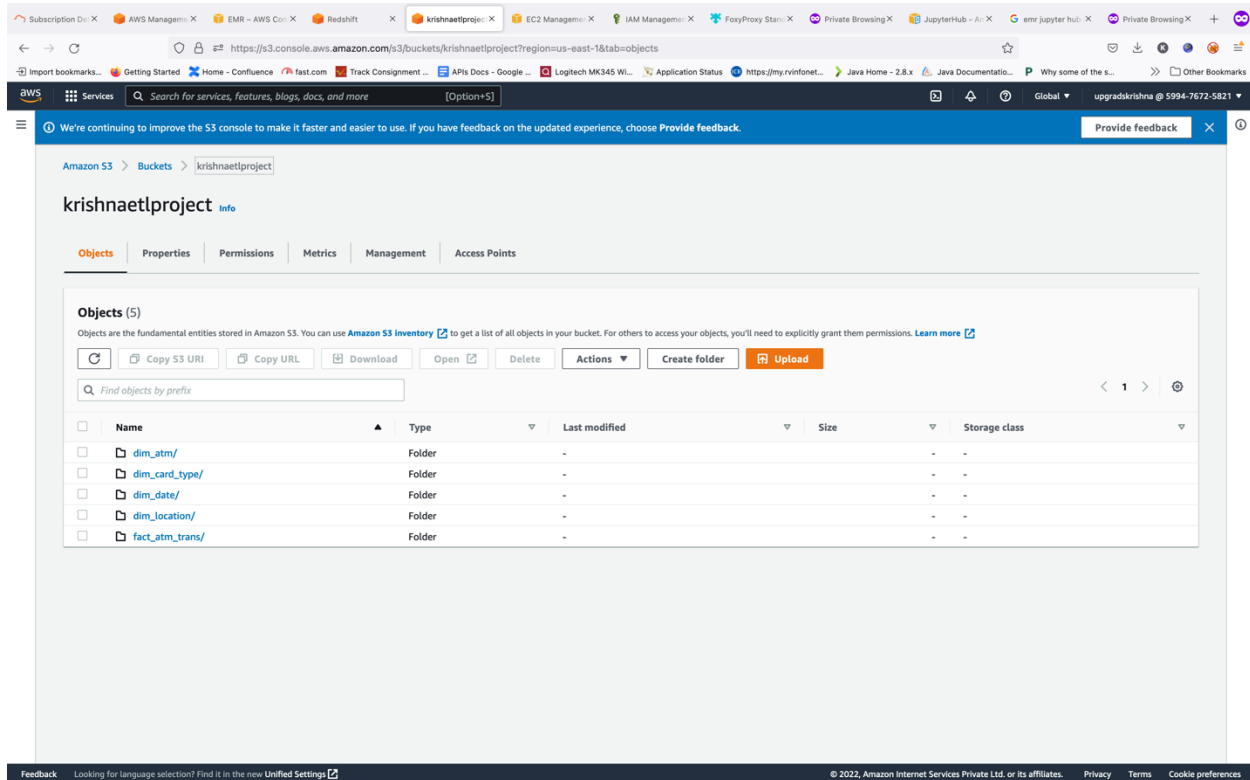
Maintenance

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](#)

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

Viewing data in s3 bucket:



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

Pre-requisites:

```
create schema atm_data;

drop table atm_data.dim_atm;
drop table atm_data.dim_card_type;
drop table atm_data.dim_date;
drop table atm_data.dim_location;
drop table atm_data.fact_atm_trans;
```

Queries to create the tables:

To create DIM_LOCATION:

```
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)
);
```

To create DIM_ATM:

```
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)
);
```

To create DIM_DATE:

```
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)
);
```

To create DIM_CARD_TYPE:

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

To create FACT_ATM_TRANS:

```
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)
);
```

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

To copy data to the appropriate tables from s3, I used the following queries:

To copy data from s3 to DIM_LOCATION table:

```
copy atm_data.dim_location from 's3://krishnaetlproject/dim_location/part-00000-
a5d22bc1-c41e-4feb-ae90-c7043bd03103-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

To copy data from s3 to DIM_ATM table:

```
copy atm_data.dim_atm from 's3://krishnaetlproject/dim_atm/part-00000-a1acba1d-585d-403c-b6a4-5cfc0704ac4b-c000.csv'  
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1'  
CSV;
```

To copy data from s3 to DIM_DATE table:

```
copy atm_data.dim_date from 's3://krishnaetlproject/dim_date/part-00000-d6700597-cd65-4555-b8ef-735e8deb375a-c000.csv'  
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1'  
TIMEFORMAT AS 'auto'  
CSV;
```

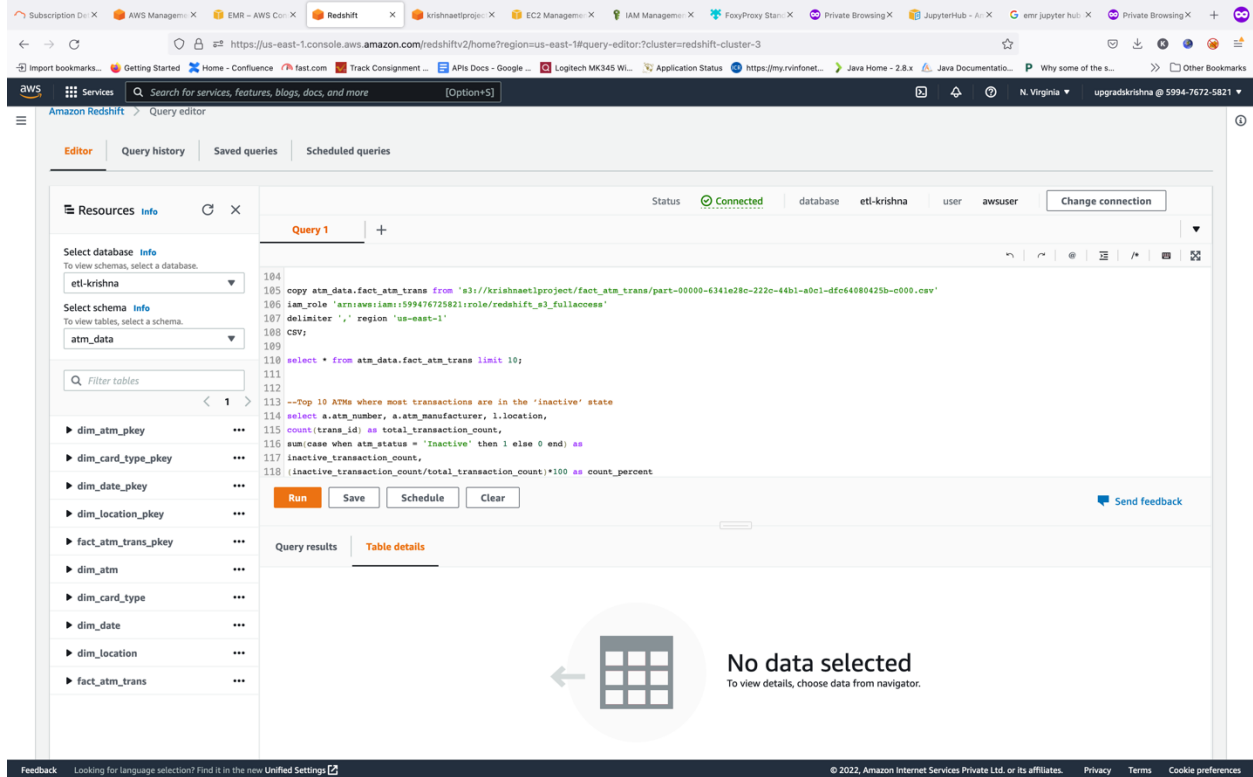
To copy data from s3 to DIM_CARD_TYPE table:

```
copy atm_data.dim_card_type from 's3://krishnaetlproject/dim_card_type/part-00000-f0097118-fead-46ed-b0fd-cc5606086622-c000.csv'  
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1'  
CSV;
```

To copy data from s3 to FACT_ATM_TRANS table:

```
copy atm_data.fact_atm_trans from 's3://krishnaetlproject/fact_atm_trans/part-00000-6341e28c-222c-44b1-a0c1-dfc64080425b-c000.csv'  
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'  
delimiter ',' region 'us-east-1'  
CSV;
```


Screenshot of created tables list:



The screenshot shows the Amazon Redshift Query Editor interface. On the left, there is a sidebar with a 'Resources' section. Under 'Select database', 'etl-krishna' is selected. Under 'Select schema', 'atm_data' is selected. Below this, 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 area shows a SQL query in the 'Query 1' tab. The query is as follows:

```
104 copy atm_data.fact_atm_trans from 's3://krishnaetlproject/fact_atm_trans/part-00000-6341e28c-222c-44b1-a0c1-dfc64080425b-c000.csv'
105 iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
106 delimiter ',' region 'us-east-1'
107 CSV;
108
109
110 select * from atm_data.fact_atm_trans limit 10;
111
112
113 --Top 10 ATMs where most transactions are in the 'inactive' state
114 select a.atm_number, a.atm_manufacturer, l.location,
115 count(trans_id) as total_transaction_count,
116 sum(case when atm_status = 'Inactive' then 1 else 0 end) as
117 inactive_transaction_count,
118 (inactive_transaction_count/total_transaction_count)*100 as count_percent
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of the query, there is a 'Status' section showing 'Connected' and a 'Change connection' button. Below the query, there is a 'Query results' section with a 'Table details' tab. The 'Table details' tab is active, and it shows a message: 'No data selected. To view details, choose data from navigator.'