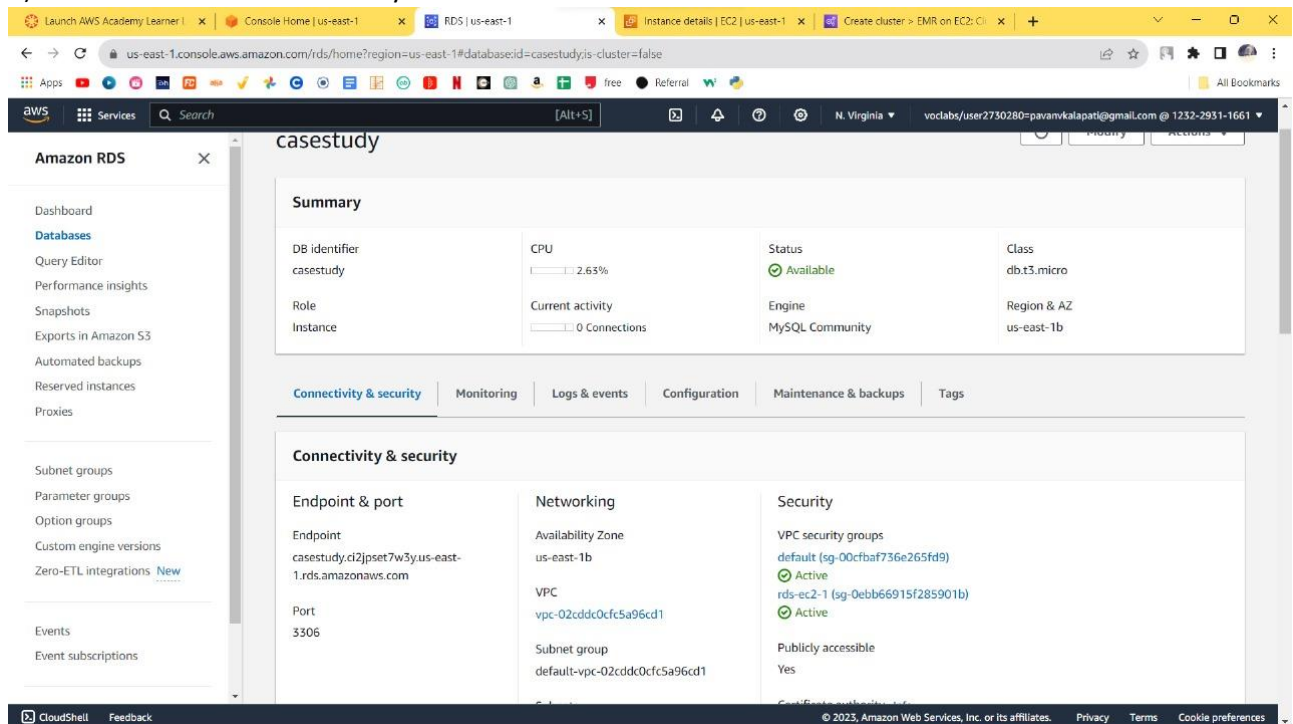


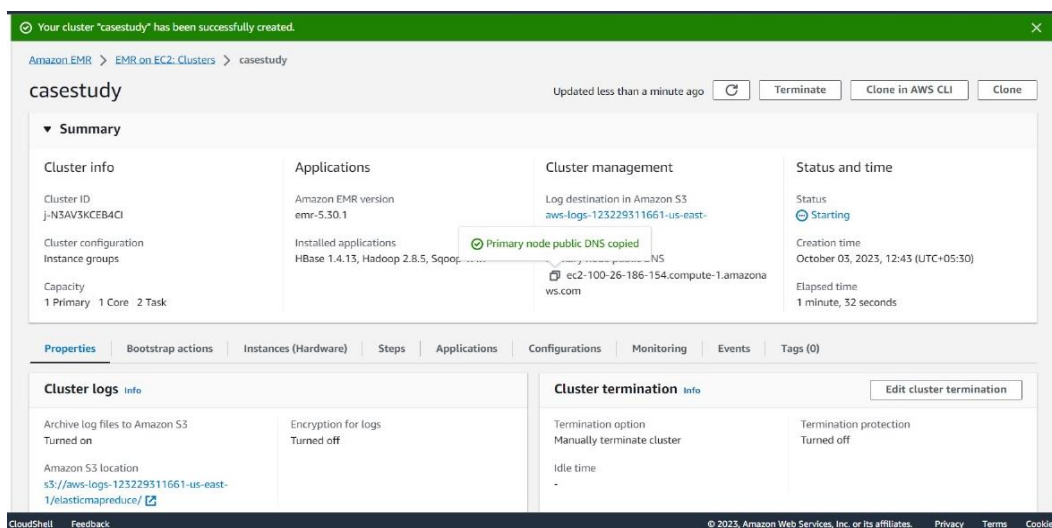
Task 1

Create an RDS instance in your AWS account and upload the data to the RDS instance

- 1) We've generated an RDS instance using the Learner Lab and initiated an EMR cluster.
- 2) We obtained the necessary files for this task, namely yellow_tripdata_2017-01.csv and yellow_tripdata_201702.csv.
- 3) We've established a connection from EMR to the RDS instance, created a table, and imported records from the CSV files.
- 4) The data has been successfully loaded into the RDS instance we created.



EMR Cluster:



Connecting RDS with EMR EC2 instance:

Launch AWS Academy Le...

Console Home [us-east-1]

RDS [us-east-1]

Launch an Instance [EC2]

Connect RDS database [E...

Properties > casestudy > ...

us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#RDSConnect:instanceId=i-0f3e82c28b2cc8f06

Apps

Services

Search

[Alt+S]

N. Virginia

voclabs/user2730280=pavankalapati@gmail.com @ 1232-2931-1661

Connecting an RDS database to your EC2 instance automatically creates and adds security groups to allow traffic between the database and the EC2 instance.

EC2 Instance ID
i-0f3e82c28b2cc8f06

Instance VPC ID
vpc-02cddc0cf5a96cd1

Select the RDS database to connect to your EC2 instance.

Database role

☐ Cluster
Apply security groups to all database instances within a cluster (Only regional clusters supported)

☒ Instance
Apply security groups to individual database instances that are not part of a cluster

RDS database

A security group is added to the EC2 instance. A security group is also added to the database with an inbound rule that allows the EC2 instance to access the database port.

casestudy
Engine: MySQL Community Region and AZ: us-east-1b

vpc-02cddc0cf5a96cd1

Create RDS database

You will incur data transfer fees because your EC2 instance and your RDS database are in different Availability Zones. To avoid incurring these charges, they must be in the same Availability Zone.

Cancel

Connect

CloudShell Feedback

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Launch AWS Academy Le...

Console Home [us-east-1]

RDS [us-east-1]

Launch an Instance [EC2]

Connect RDS database [E...

Properties > casestudy > ...

us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#RDSConnect:instanceId=i-0f3e82c28b2cc8f06

Apps

Services

Search

[Alt+S]

N. Virginia

voclabs/user2730280=pavankalapati@gmail.com @ 1232-2931-1661

EC2 > Instances > Connect RDS database

Success

Successfully connected instance (i-0f3e82c28b2cc8f06) to database casestudy

▶ Connection log

View instances

CloudShell Feedback

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Connecting EMR instance with PuTTY and then downloading **yellow_tripdata_2017-01.csv** & **yellow_tripdata_2017-02.csv** :

```
root@ec2-user@ip-172-31-84-154 ~# wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
--2023-10-02 12:00:06-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 3.5.21.109, 16.182.103.161, 52.216.76.132, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|3.5.21.109|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 914029540 (872M) [text/csv]
Saving to: 'yellow_tripdata_2017-01.csv'

100%[=====] 914,029,540 28.4MB/s in 32s

2023-10-02 12:00:38 (27.0 MB/s) - 'yellow_tripdata_2017-01.csv' saved [914029540/914029540]

[ec2-user@ip-172-31-84-154 ~]$
```

```
Aborted
[ec2-user@ip-172-31-84-154 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
--2023-10-02 12:00:06-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 3.5.21.109, 16.182.103.161, 52.216.76.132, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|3.5.21.109|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 914029540 (872M) [text/csv]
Saving to: 'yellow_tripdata_2017-01.csv'

100%[=====] 914,029,540 28.4MB/s in 32s

2023-10-02 12:00:38 (27.0 MB/s) - 'yellow_tripdata_2017-01.csv' saved [914029540/914029540]

[ec2-user@ip-172-31-84-154 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv
--2023-10-02 12:01:53-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.217.42.12, 52.217.107.92, 52.217.117.49, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.217.42.12|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 863487050 (823M) [text/csv]
Saving to: 'yellow_tripdata_2017-02.csv'

100%[=====] 863,487,050 24.7MB/s in 34s

2023-10-02 12:02:27 (24.5 MB/s) - 'yellow_tripdata_2017-02.csv' saved [863487050/863487050]

[ec2-user@ip-172-31-84-154 ~]$
```

Connecting RDS with EMR Instance:

Hostname: **casestudy.ci2jpset7w3y.us-east-1.rds.amazonaws.com**

Command:

mysql -h case-study-dbb.ck4jzoqb1yn7.us-east-1.rds.amazonaws.com -P 3306 -u root -p

```

E:::EEEEEEEEEE M:::MM M:::MM M:::MM R::RRRRR:::R
E:::E M:::MM M:::MM M:::MM R:::R R:::R
E:::E EEEEE M:::MM MM M:::MM R:::R R:::R
RE:::EEEEEE:::E M:::MM M:::MM R:::R R:::R
M:::MM:::MM M:::MM M:::MM RR:::R R:::R
EEEEEEEEEEEEEEEE MMMMMM MMMMMM RRRRRR RRRRRR

[ec2-user@ip-172-31-84-154 ~]$ mysql -h casestudy.ci2jpset7w3y.us-east-1.rds.amazonaws.com -P 3306 -u root -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MySQL connection id is 16
Server version: 8.0.33 Source distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MySQL [(none)]> create database yellow_taxi;
Query OK, 1 row affected (0.00 sec)

MySQL [(none)]> use yellow_taxi;
Database changed
MySQL [yellow_taxi]> CREATE TABLE taxi
-> (
-> vendorID INT,trip_pickup_datetime DATETIME, tpep_dropoff_datetime DATETIME, passenger_count INT, trip_distance DOUBLE, puLocationID INT, doLocationID INT, rateCodeID INT, store_and_fwd_flag VARCHAR(255), payment_type INT, fare_amount DOUBLE, extra DOUBLE, mta_tax DOUBLE, improvement_surcharge DOUBLE, tip_amount DOUBLE, tolls_amount DOUBLE, total_amount DOUBLE, congestion_surcharge DOUBLE, airport_fee DOUBLE );
Query OK, 0 rows affected (0.04 sec)
```

```
create database
```

```
yellow_taxi;
```

```
use yellow_taxi;
```

Creating table—

```
CREATE TABLE taxi
```

```
(
```

```
vendorID INT,
```

```
tpep_pickup_datetime DATETIME,
```

```
tpep_dropoff_datetime DATETIME,
```

```
passenger_count INT, trip_distance
```

```
DOUBLE, puLocationID INT,
```

```
doLocationID INT, rateCodeID INT,
```

```
store_and_fwd_flag VARCHAR(255),
```

```
payment_type INT, fare_amount
```

```
DOUBLE, extra DOUBLE, mta_tax
```

```
DOUBLE, improvement_surcharge
```

```
DOUBLE, tip_amount DOUBLE,
```

```
tolls_amount DOUBLE, total_amount
```

```
DOUBLE, congestion_Surcharge
```

```
DOUBLE, airport_fee DOUBLE
```

```
);
```

```

o your MySQL server version for the right syntax to use near 'describe table taxi' at line 1
MySQL [yellow_taxi]> describe taxi;
+-----+-----+-----+-----+-----+-----+
| Field                | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| vendorID              | int           | YES  |     | NULL    |       |
| tpep_pickup_datetime  | datetime      | YES  |     | NULL    |       |
| tpep_dropoff_datetime | datetime      | YES  |     | NULL    |       |
| passenger_count       | int           | YES  |     | NULL    |       |
| trip_distance         | double        | YES  |     | NULL    |       |
| puLocationID          | int           | YES  |     | NULL    |       |
| doLocationID          | int           | YES  |     | NULL    |       |
| rateCodeID            | int           | YES  |     | NULL    |       |
| store_and_fwd_flag    | varchar(255)  | YES  |     | NULL    |       |
| payment_type          | int           | YES  |     | NULL    |       |
| fare_amount           | double        | YES  |     | NULL    |       |
| extra                 | double        | YES  |     | NULL    |       |
| mta_tax               | double        | YES  |     | NULL    |       |
| improvement_surcharge | double        | YES  |     | NULL    |       |
| tip_amount            | double        | YES  |     | NULL    |       |
| tolls_amount          | double        | YES  |     | NULL    |       |
| total_amount          | double        | YES  |     | NULL    |       |
| congestion_surcharge  | double        | YES  |     | NULL    |       |
| airport_fee           | double        | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
19 rows in set (0.00 sec)

```

Load Data into Above table

```

LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
INTO TABLE taxi
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

```

```

LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv'
INTO TABLE taxi
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

```

