

## RDS

- **Command to load csv files to EMR cluster.**

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-01.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv)

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-02.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv)

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-03.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-03.csv)

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-04.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-04.csv)

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-05.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-05.csv)

wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-06.csv](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-06.csv)

```
hadoop@ip-172-31-89-117:~
Saving to: 'yellow_tripdata_2017-03.csv'

100%[=====] 969,809,025 31.3MB/s in 30s

2023-01-09 09:31:30 (31.3 MB/s) - 'yellow_tripdata_2017-03.csv' saved [969809025/969809025]

[hadoop@ip-172-31-89-117 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-04.csv
--2023-01-09 09:31:37-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-04.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 54.231.233.121, 52.217.164.89, 52.216.161.123, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|54.231.233.121|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 946349441 (903M) [text/csv]
Saving to: 'yellow_tripdata_2017-04.csv'

100%[=====] 946,349,441 30.1MB/s in 30s

2023-01-09 09:32:07 (30.3 MB/s) - 'yellow_tripdata_2017-04.csv' saved [946349441/946349441]

[hadoop@ip-172-31-89-117 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-05.csv
--2023-01-09 09:32:14-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-05.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.216.39.57, 52.217.162.33, 3.5.1.147, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.216.39.57|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 951965526 (908M) [text/csv]
Saving to: 'yellow_tripdata_2017-05.csv'

100%[=====] 951,965,526 32.2MB/s in 29s

2023-01-09 09:32:44 (31.1 MB/s) - 'yellow_tripdata_2017-05.csv' saved [951965526/951965526]

[hadoop@ip-172-31-89-117 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-06.csv
--2023-01-09 09:32:49-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-06.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.216.154.188, 52.216.94.83, 52.216.40.145, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.216.154.188|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 910028408 (868M) [text/csv]
Saving to: 'yellow_tripdata_2017-06.csv'

100%[=====] 910,028,408 34.3MB/s in 26s

2023-01-09 09:33:16 (32.9 MB/s) - 'yellow_tripdata_2017-06.csv' saved [910028408/910028408]

[hadoop@ip-172-31-89-117 ~]$ ls -l
total 5425468
-rw-rw-r-- 1 hadoop hadoop 914029540 Nov 25 15:05 yellow_tripdata_2017-01.csv
-rw-rw-r-- 1 hadoop hadoop 863487050 Nov 25 15:07 yellow_tripdata_2017-02.csv
-rw-rw-r-- 1 hadoop hadoop 969809025 Nov 25 15:07 yellow_tripdata_2017-03.csv
-rw-rw-r-- 1 hadoop hadoop 946349441 Nov 25 15:08 yellow_tripdata_2017-04.csv
-rw-rw-r-- 1 hadoop hadoop 951965526 Nov 25 15:08 yellow_tripdata_2017-05.csv
-rw-rw-r-- 1 hadoop hadoop 910028408 Nov 25 15:09 yellow_tripdata_2017-06.csv
[hadoop@ip-172-31-89-117 ~]$
```

- **Command to create directory in HDFS**

hadoop fs -mkdir /user/hbase/csv

- **Command to put files into HDFS**

hadoop fs -put yellow\_tripdata\_2017-01.csv /user/hbase/csv

hadoop fs -put yellow\_tripdata\_2017-02.csv /user/hbase/csv

```
hadoop fs -put yellow_tripdata_2017-03.csv /user/hbase/csv
```

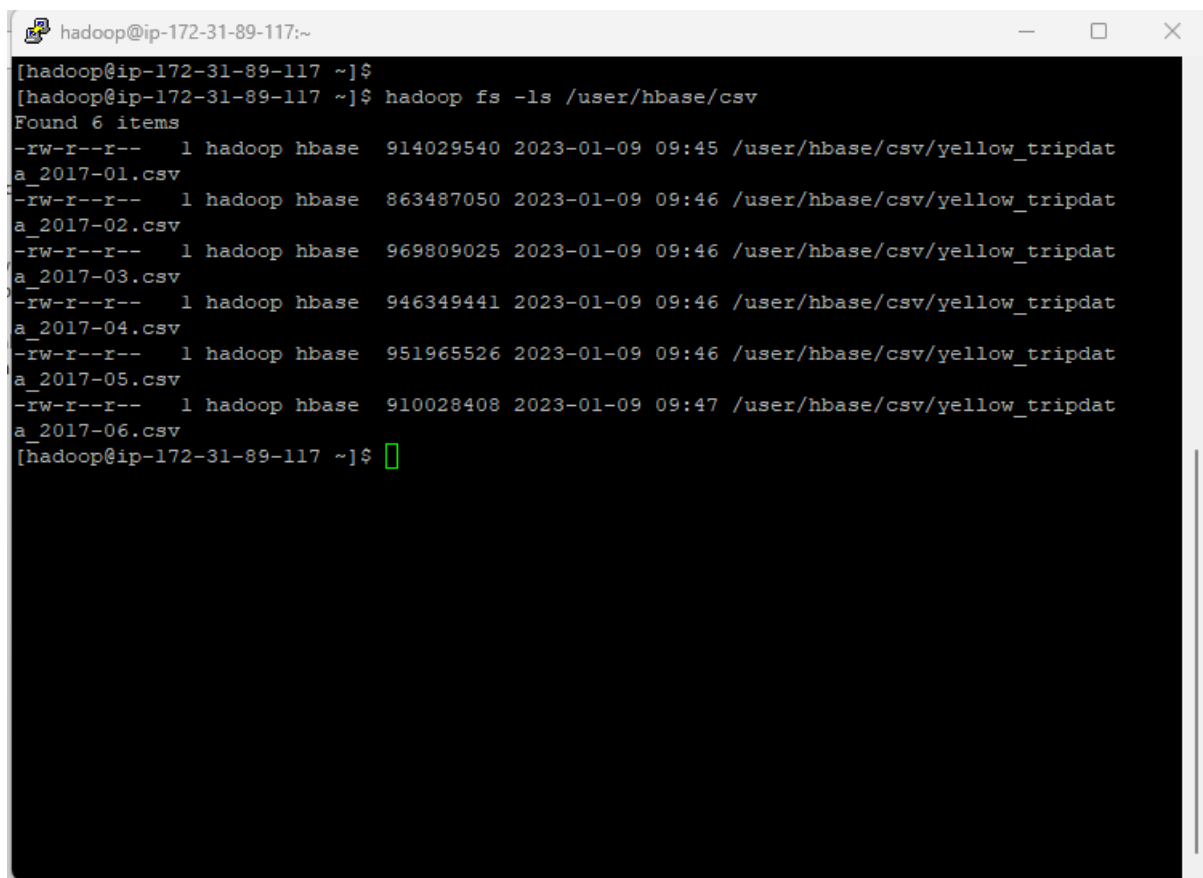
```
hadoop fs -put yellow_tripdata_2017-04.csv /user/hbase/csv
```

```
hadoop fs -put yellow_tripdata_2017-05.csv /user/hbase/csv
```

```
hadoop fs -put yellow_tripdata_2017-06.csv /user/hbase/csv
```

- **Command to list files in HDFS**

```
hadoop fs -ls /user/hbase/csv
```

A terminal window titled 'hadoop@ip-172-31-89-117:~' with standard window controls. The terminal shows the command 'hadoop fs -ls /user/hbase/csv' being executed. The output lists 6 items with their permissions, owner, group, size, modification time, and path. The paths are truncated in the screenshot.

```
hadoop@ip-172-31-89-117:~$  
[hadoop@ip-172-31-89-117 ~]$  
[hadoop@ip-172-31-89-117 ~]$ hadoop fs -ls /user/hbase/csv  
Found 6 items  
-rw-r--r--  1 hadoop hbase  914029540 2023-01-09 09:45 /user/hbase/csv/yellow_tripdat  
a_2017-01.csv  
-rw-r--r--  1 hadoop hbase  863487050 2023-01-09 09:46 /user/hbase/csv/yellow_tripdat  
a_2017-02.csv  
-rw-r--r--  1 hadoop hbase  969809025 2023-01-09 09:46 /user/hbase/csv/yellow_tripdat  
a_2017-03.csv  
-rw-r--r--  1 hadoop hbase  946349441 2023-01-09 09:46 /user/hbase/csv/yellow_tripdat  
a_2017-04.csv  
-rw-r--r--  1 hadoop hbase  951965526 2023-01-09 09:46 /user/hbase/csv/yellow_tripdat  
a_2017-05.csv  
-rw-r--r--  1 hadoop hbase  910028408 2023-01-09 09:47 /user/hbase/csv/yellow_tripdat  
a_2017-06.csv  
[hadoop@ip-172-31-89-117 ~]$
```

- **Command for connecting to RDS from EMR cluster**

```
mysql -h upgradassignmentdb.cdwr877prmiq.us-east-1.rds.amazonaws.com -P 3306 -u  
admin -p
```

- **Command for creating database**

```
create database mapreduce1
```

- **Command for using database**

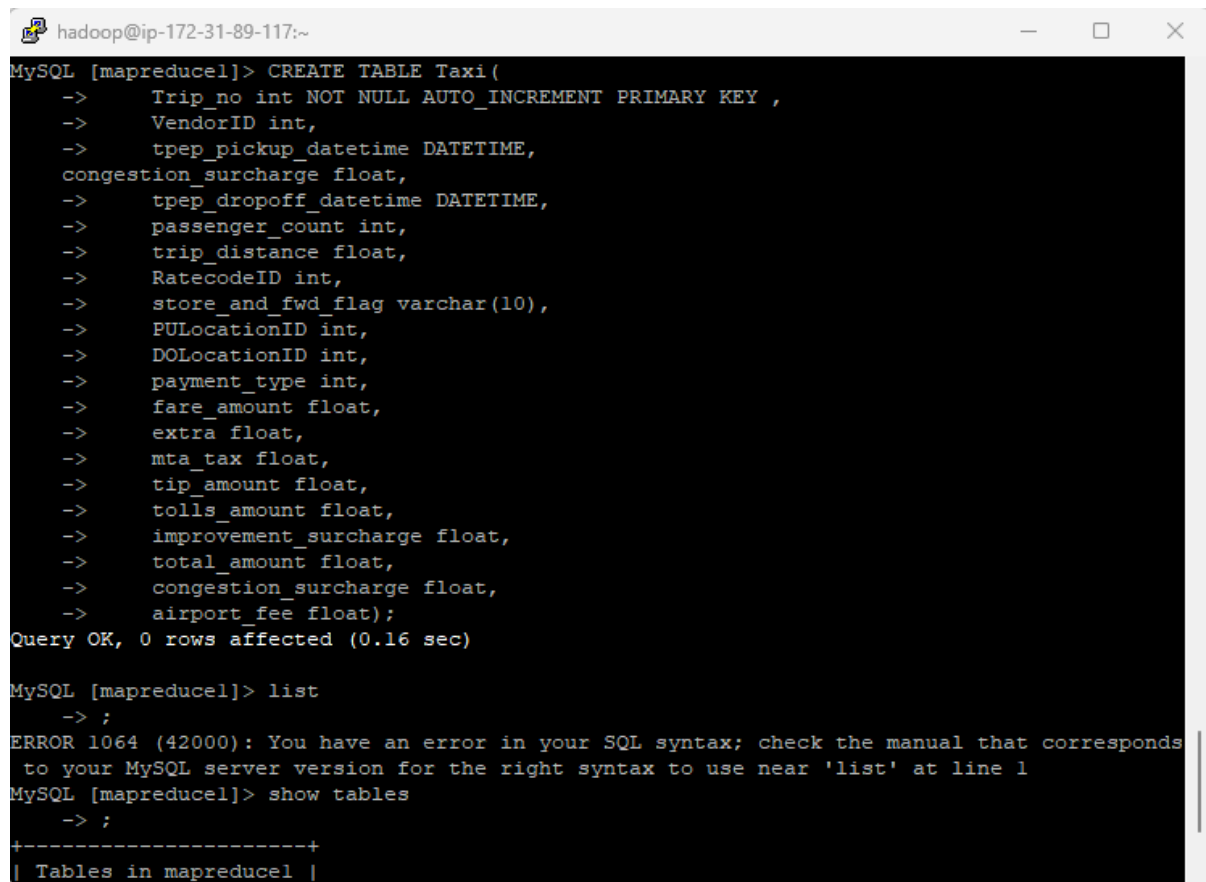
```
use mapreduce1
```

- **Command for creating table**

```

CREATE TABLE Taxi(
  Trip_no int NOT NULL AUTO_INCREMENT PRIMARY KEY ,
  VendorID int,
  tpep_pickup_datetime DATETIME,
  tpep_dropoff_datetime DATETIME,
  passenger_count int,
  trip_distance float,
  RatecodeID int,
  store_and_fwd_flag varchar(10),
  PULocationID int,
  DOLocationID int,
  payment_type int,
  fare_amount float,
  extra float,
  mta_tax float,
  tip_amount float,
  tolls_amount float,
  improvement_surcharge float,
  total_amount float,
  congestion_surcharge float,
  airport_fee float);

```



```

hadoop@ip-172-31-89-117:~
MySQL [mapreducel]> CREATE TABLE Taxi(
->   Trip_no int NOT NULL AUTO_INCREMENT PRIMARY KEY ,
->   VendorID int,
->   tpep_pickup_datetime DATETIME,
congestion_surcharge float,
->   tpep_dropoff_datetime DATETIME,
->   passenger_count int,
->   trip_distance float,
->   RatecodeID int,
->   store_and_fwd_flag varchar(10),
->   PULocationID int,
->   DOLocationID int,
->   payment_type int,
->   fare_amount float,
->   extra float,
->   mta_tax float,
->   tip_amount float,
->   tolls_amount float,
->   improvement_surcharge float,
->   total_amount float,
->   congestion_surcharge float,
->   airport_fee float);
Query OK, 0 rows affected (0.16 sec)

MySQL [mapreducel]> list
-> ;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds
to your MySQL server version for the right syntax to use near 'list' at line 1
MySQL [mapreducel]> show tables
-> ;
+-----+
| Tables_in_mapreducel |

```

- Inserting first 2 csv files data to Taxi table, following commands need to be typed inside mysql

```
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv' INTO
TABLE Taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n'
IGNORE 1 LINES (VendorID ,tpep_pickup_datetime, tpep_dropoff_datetime ,
passenger_count , trip_distance, RatecodeID, store_and_fwd_flag ,
PULocationID , DOLocationID, payment_type , fare_amount, extra, mta_tax,
tip_amount, tolls_amount, improvement_surcharge , total_amount,
congestion_surcharge, airport_fee);
```

```
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv' INTO
TABLE Taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n'
IGNORE 1 LINES (VendorID ,tpep_pickup_datetime, tpep_dropoff_datetime ,
passenger_count , trip_distance, RatecodeID, store_and_fwd_flag ,
PULocationID , DOLocationID, payment_type , fare_amount, extra, mta_tax,
tip_amount, tolls_amount, improvement_surcharge , total_amount,
congestion_surcharge, airport_fee);
```

```
hadoop@ip-172-31-90-95:~
, total_amount, congestion_surcharge, airport_fee);
ERROR 2 (HY000): File '/home/Hadoop/yellow_tripdata_2017-02.csv' not found (Errcode:
2)
MySQL [mapreduser]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv'
INTO TABLE Taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES
(VendorID ,tpep_pickup_datetime, tpep_dropoff_datetime , passenger_count , trip_d
istance, RatecodeID, store_and_fwd_flag , PULocationID , DOLocationID, payment_ty
pe , fare_amount, extra, mta_tax, tip_amount, tolls_amount, improvement_surcharge
, total_amount, congestion_surcharge, airport_fee);
Query OK, 9169775 rows affected, 65535 warnings (1 min 48.70 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 18339550

MySQL [mapreduser]> select count (*) from Taxi
-> ;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corre
sponds to your MySQL server version for the right syntax to use near '*) from Taxi' a
t line 1
MySQL [mapreduser]> select count(*) from Taxi;
+-----+
| count(*) |
+-----+
| 18880597 |
+-----+
1 row in set (42.29 sec)

MySQL [mapreduser]> 
```

