

# WQD7007 Big Data Management

## Lab Test (Hive)

Instructions: Answer the following question by using command line interface or Ambari interface. Explain adequately how you get the answers that can include codes used in command line interfaces, codes used in Ambari, a number of print screens and related files. You can create a readme file (either in .txt or .docx) to explain your answer. Zip all the files and submit to the spectrum “Lab test” submission page.

### Part 1: (5 marks)

1. Create a file named: “lab\_(YourInitial).txt”. E.g. lab\_hwl.csv
2. Insert the following content into the file created.

This is a dummy text file for WQD7007.

This file contains 4 lines.

Please remember to save this file.

```
hdfs dfs -mkdir /user/hdfs/lab_test
hdfs dfs -put lab_cyq.txt /user/hdfs/lab_test/lab_cyq.txt
hdfs dfs -ls /user/hdfs/lab_test
hdfs dfs -cat /user/hdfs/lab_test/lab_cyq.txt
```

3. Upload the file to HDFS in directory “/user/hdfs/lab\_test”.

```
> create table car (car string, MPG double, Cylinders int, displacement double, horsepower double, weight double, acceleration double, model int, origin string)
> row format delimited
> fields terminated by ',';
> insert into table car values ('Chevrolet Chevelle Malibu',18.0,8,307.0,130.0,3504.,12.0,70,'US'), ('Buick Skylark 320',15.0,8,350.0,165.0,3693.,11.5,70,'US'), ('Plymouth Satellite',18.0,8,318.0,150.0,3436.,11.0,70,'US'), ('AMC Rebel SST',16.0,8,304.0,150.0,3433.,12.0,70,'US'), ('Ford Torino',17.0,8,302.0,140.0,3449.,10.5,70,'UK');
> select * from car;
```

### Part 2: (7 marks)

1. Create a Hive table using the following data. Insert the following data to the table created. Display the table after all the data is inserted.

```
Chevrolet Chevelle Malibu;18.0;8;307.0;130.0;3504.;12.0;70;US
Buick Skylark 320;15.0;8;350.0;165.0;3693.;11.5;70;US
Plymouth Satellite;18.0;8;318.0;150.0;3436.;11.0;70;US
AMC Rebel SST;16.0;8;304.0;150.0;3433.;12.0;70;US
Ford Torino;17.0;8;302.0;140.0;3449.;10.5;70;UK
```

Attributes used:

Car (String); Miles Per Gallon, MPG (double), Cylinders (int), Displacement (double); Horsepower (double); Weight (double); Acceleration (double); Model (int); Origin (String).

2. Update the record that have horsepower = 150 to 120.
3. Delete the record that have weight = 3433. 

```
> select * from car where displacement < 320 and origin = 'US';
```
4. Display the record that displacement smaller than 320 and origin from US.

### Part 3: (3 marks)

1. Download “cars.csv” from spectrum. Then, upload the file to Hive database table. Hint: you will need to change some settings and edit the csv file manually in order for you can upload to upload the content successfully.
2. Select maximum number of cylinders for each car model. Display the result.

```
hdfs dfs -put /home/yukichen/Downloads/cars.csv /user/yukichen
hive
> CREATE EXTERNAL TABLE IF NOT EXISTS cars(car string, MPG double, Cylinders int, displacement double, horsepower double, weight double, acceleration double, model int, origin string)
> row format delimited
> fields terminated by ',';
> tblproperties("skip.header.line.count"="2");
> load data inpath 'cars.csv' into table cars;
> select * from cars;
```

```
> select car,max(cylinders) from cars group by car;
```