Section 1: Technical Questions

Question 1

Consider the following steps that a user executes on Hive CLI.

Step 1. Upload query-hive.csv

Step 2. Select substring from names, from first letter to fourth.

Query : …………………………………………………………………………

Output : …………………………………………………………………………

Question 2

What command should you try if DBMS\_OUTPUT.PUT\_LINE is not doing what it is supposed to do?

1. SET ECHO ON
2. SET DISPLAY ON
3. SET TERMOUT ON
4. SET SERVEROUTPUT ON

Question 3

Which of the following reduce-side tuning properties allows the maximum number of streams to merge at once when sorting files, and is also used in the map?

1. mapred.reduce.parallel.copies
2. mapred.reduce.copy.backoff
3. io.sortfactor
4. mapred.job.shuffle.input.buffer.copy

Question 4

The basic programming unit of a PL/SQL code is a

1. Procedure
2. sub-program
3. module
4. block

Section 2: Descriptive Questions

Question 5

Please explain approaches when confronted with challenges from multi-modal data storage system for use cases.

Question 6

Please explain the impetus of each, its difference and similarity between MapReduce, Spark and CUDA programming.

Section 3: Practical Task

Create a VM, running elasticsearch. In this instance, upload dataset from:

<https://www.kaggle.com/anderas/car-consume/data>

With the correct meta-data representation. After this is done, show how 2 user sessions is being passed to elasticsearch that searches;

User 1: All index term-name that within temp\_outside field , between 5 and 10

User 2: All index term-name that within speed , between 30 and 50

from Django.