


Name:	
Enrolment No:	
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES End Semester Examination, July 2020	
Course: Data Mgmt. & Analytics in E & P	Semester: VI
Course Code: CSOG 3004	Time: 3 hours
Programme: OGI	Max. Marks: 70 = 30 (viva)
Instructions: Nil	

Q1. Solve the following riddle to get your question. - 15 Marks

(Note: there are 72 unique questions and only 55 students. Enjoy)

, your roll number%3 = 1

Part 1: Solve to select a category, () = { , your roll number%3 = 0

, %3=2

- A. You work at Nescafe and need to report the sale of following to the head office every hour. Nescafe operates globally and head office wants sale to country-wise. List of items:
- a. Classic coffee
 - b. Chicken Burgers
 - c. Veg Burgers
 - d. Cappuccino
 - e. Ice Tea
- B. You work at PUBG head office and you need to report the number of following weapons distributed over Sanhok and Miramar every 30 mins to the C.E.O. of the company. List of weapons:
- a. M416.
 - b. AKM.
 - c. M16A4.
 - d. SCAR-L.
 - e. Vector.
 - f. P18C.

g. Kar98k.

h. Mini 14.

C. You are appointed as a data analyst in the health ministry at Delhi and the minister wants number of people who have entered any particular state. Here we are only counting entry (if same person enters multiple times, he is counted multiple times). Present a border wise list to the minister.

Part 2: if your roll number is even, then design the mapper.

If your roll number is odd, then design the reducer.

Part 3: Additional Constrains:

A. The data needs to be read from and stored to the following formats:

0. HDFS
1. HBASE
2. MySQL

To determine the database for your problem use the formula.
()%3

B. The language options are:

0. Pig
1. Java/Python
2. JAQL
3. Hive

To determine the language for your problem use the formula.
()%4

Answer format for question 1: (please use the table)

Part 1 output	Part 2 output	Part 3.A - output	Part 3.B – output
			Note – Mapper/Reducer has to be in the specified languages. Logic can be written in language of choice.

Q2. Write a code for the question generated in Q1.* -55 Marks.

Please submit only code here. Also, assumptions about schema, if any.

C. Assumptions provided:

1. Part 1 –C:

Only provide number of entries to the state. In practical life, states may share numerous borders. But for this problem, a state shares only one border post with its neighboring state.

You may use a code to determine the neighboring states to a particular state or can perform that activity manually. But stick to geographical map of India.

The data is located at the capital of each state for reducer part.

The data is to be displayed to minister in Delhi, for the reducer part.

2. Part 1 – A,B:

The data is distributed across as:

S. No.	Area	Location	Server Name
1	SAARC countries	Delhi	Delhi@1
2	Europe	Geneva	Geneva@2
3	Africa	Cape Town	Cape@21
4	Asia	Tokyo	Gohul@Ken
5	North America	Vancouver	Lucifer@6
6	South America	Rio De Jenerio	Rio@De

3. You are free to design the schema. Schema needs to be provided with the solution.

Marking Scheme:

- a. All complete (Either Mapper or Reducer as per the algorithm) - Full marks.
- b. If you cannot write the code using the solution generated in Q1. Then, you may replace a particular part (all 4 can be replaced) but for each replacement you lose 5 marks. If you change any part of question 1 while answering question 2. Please specify that clearly in question 2.

Note: Extra data is used for padding purpose.