# **BIG DATA –Spring 2019**

# **ASSIGNMENT 3**

**Due Date: 11th March 2019 till 10:00 PM on google classroom.**

Upload the Source code and the output file on Google Classroom with your roll number.

Consider the same dataset that was provided in the Assignment 2.

**QUESTION**

Write MapReduce program using **Pair’s Approach** for the following task: For each age, list the percentage of bank customers with a particular job.

**A portion of Input File**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | 30;"unemployed";"married";"primary";"no";1787;"no";"no";"cellular";19;"oct";79;1;-1;0;"unknown";"no" | | 33;"services";"married";"secondary";"no";4789;"yes";"yes";"cellular";11;"may";220;1;339;4;"failure";"no" | | 35;"management";"single";"tertiary";"no";1350;"yes";"no";"cellular";16;"apr";185;1;330;1;"failure";"no" | | 30;"management";"married";"tertiary";"no";1476;"yes";"yes";"unknown";3;"jun";199;4;-1;0;"unknown";"no" | | 35;"technician";"single";"tertiary";"no";747;"no";"no";"cellular";23;"feb";141;2;176;3;"failure";"no" | | 30;"unemployed";"married";"tertiary";"no";307;"yes";"no";"cellular";14;"may";341;1;330;2;"other";"no" | |

**Sample Output (AGE 30, Job Services -> percentage of the customer of age=30 who are doing job in Services)**

|  |
| --- |
| 30, unemployed 🡪 66%  30, management 🡪 33%  33, services 🡪 100%  35, management 🡪 50%  35, technician 🡪 50% |

**Hint:**

Percentage of customer of Age X with job J = No. of customer of age X with Job J

Total no. customer of age X

This problem is somewhat similar to relative frequency word co-occurrence problem discussed in the class. *You will have to provide* ***Mapper, Reducer, Partitioner and Comparator.***

**Hints:**

**1. C***onsult book Hadoop the definitive guide for comparator details (Chapter 5, pg 136 & onwards)*

*2. Have a look at http://codingjunkie.net/order-inversion***/**