



Ramakrishna Mission

Vivekananda Educational and Research Institute
Department of Computer Science
MSc in BDA ETBDC

Total Marks: 100 Time: Three Hours

Saturday 15th April, 2023

1 Application Development: Marks 90

Programme output needs to be shown during the exam and the same needs to be submitted along with source files. One question needs to be answered from this section as per instruction in exam hall.

1.1 Distribution of Employees

Develop a single hadoop application to perform the following.

- Calculate the number of employees associated with each designation.
- Calculate the number of employees associated with each department.

Data: teachers.csv

1.2 Distribution of the Classes taken by Teachers

Develop a hadoop application to perform the following.

- Calculate the number of classes taken by each Teacher.
- Use custom counter to display number of headers in input data.

Data: classes.csv

1.3 Distribution of the Students

Develop a hadoop application to perform the following.

- Calculate the number of students in each Programme.
- Use custom counter to display number of headers in input data.

Data: students.csv

2 Analyze Hadoop Framework: Marks 10

Answer any one of the following. Reference: [https://github.com/facebookarchive/hadoop-20/blob/master/src/mapred/org/ap](https://github.com/facebookarchive/hadoop-20/blob/master/src/mapred/org/apache/hadoop/mapred/JobTracker.java)

2.1 Modification of framework for use of streaming

Suggest modifications to be made in Hadoop Framework so that streaming characteristics can be implemented in Hadoop application. Example: Moving average of temperature in every 5 minutes.

2.2 Improvement of Hadoop Framework for 'no code' or 'less code' solution

Multiple applications use 'groupby' operation on Big Data. Suggest some scope of improvements on Hadoop framework so that code size gets reduced for all such applications.