CSE 5331

Database Management Systems

Models and Implementation

(002)

**Project 3:**

**Map Reduce**

Prepared By: Meet Brahmbhatt (1001119131)

Yash Shah (1001111714)

Instructor: Prof. Sharma Chakravarthy

1. **Overall Status:**  
   Project status: Completed.  
     
   First of all, we installed Hadoop in our computers. To complete this project we used Eclipse IDE. Due to that, we installed plugin for Hadoop in eclipse. To solve both problems, we have implemented two separate files called ProblemOne.java and ProblemTwo.java. As we have made two different files for both problems, we have done hadoop configuration in main method of each files. For each problem, we have implemented mapper and reducer method in each class.  
     
   We are unable to complete the bonus problem. We have understood the problem but we are unable to implement the logic of problem 3. What we have understood are we had to first calculate number of persons for each age ranges from 0 to 99. After that, we have to calculate average of these values. After finding average, we have to use merge and split methods to forming and reforming buckets. Using this merge and split method iteratively, we can get equi-depth histogram for varying bucket range.
2. **Performance Measure:**Method List: Map method of Mapper class and Reduce method of Reducer method.
3. **File Description:**
   1. **ProblemOne.java**
      1. Map method:  
         creates (key,value) pair of (year\_age-range,number of students)
      2. Reduce method (main method for function):  
         It counts the number of persons in specific age group ranges for specific year.
   2. **ProblemTwo.java**
      1. Map method:  
         creates (key,value) pair of (state\_SEX, average value of salary)
      2. Reduce method(main method for function):  
         It calculates the average of salary for particular state\_sex group.
4. **Division of Labor:** Yash did initial setup as he has already done project in Hadoop. After that we have distributed the two problems. Yash did the first problem. Meet did the second problem.
5. **Logical errors and how you handle them:**
   1. As we are using plugin for this, the parameters of one of the method were different. So, we have to change the parameters as per our requirements.
   2. We made mistakes in condition checking due to that we were getting incorrect output. We change the condition checking which gives proper output.
   3. When we were reading files, code was reading header line also due to that code was giving number format exception. We have changed code in such way that it will ignore first line.