CSE 587: DATA INTENSIVE COMPUTING

REPORT: PROJECT 2 PART 2

Group Members:

1. Garima Pradeep Chaturvedi Person#:50169189

2. Namita Balkrishna Marathe Person#:50170066

5 Questions Implementing Two Mappers:

1. **Find the number of departments that teach in a particular classroom for a particular semester**.
   Contained in folder: DC
   Dataset Used: bina_classschedule2.csv

   This question aims to find how many different departments, for ex: MATH, EE, PHY etc., conduct classes in a particular classroom. Through this we aim to show which classrooms are the most utilized among all departments.
   In the solution: Mapper 1 takes semester, classroom and department and marks a count of 1.
   In Mapper 2 we take the output from MR1. For each pair of Classroom and Semester we count the number of departments teaching there.

2. **Find the number of days in a week a particular time slot is occupied for a classroom in a given semester.**
   Contained in folder: TS
   Dataset Used: bina_classschedule.csv

   This question aims to show for each classroom the timings occupied throughout a day and how many days of the week are these time slots occupied. Through this we try to analyze the combination of classrooms and time slots that are least occupied throughout the week.
   In the solution: Mapper 1 takes Semester, Timing, Days of the Week and Classroom to give a count of 1 for each entry.
   Mapper 2 takes the output from MR1. For the combination of Semester, Timing and Classroom we count the number of days in a week it was occupied.

3. **Find the number of classes that are over utilized, underutilized or utilized properly for a given semester with respect to courses.**

   Contained in folder: U
   Dataset Used: bina_classschedule.csv

   This question aims to show for every semester how well the classrooms are utilized with respect to the difference between the number of students enrolled and max enrollment and the capacity of the classroom assigned for the course. Through this we analyze frequently underutilized or over utilized classrooms.
   In the solution: Mapper 1 takes Semester, classroom, course and the difference between number of students enrolled and max enrollment based on which we categorize each entry into each entry as “underutilized”, “over utilized”, “properly utilized”.
   Mapper 2 takes the output of MR1. It then counts the number of times in a given semester a classroom was underutilized, over utilized, properly utilized.

4. **Find for each day in a week how many time slots of a particular classroom are occupied.**
   Contained in folder: DES
   Dataset Used: bina_classschedule.csv

   This question aims to show for each classroom and for each day of the week, ex: M, T, W, R, F, S, the number of time slots occupied. Through this we try to analyze the days with the least number of occupied time slots for a classroom.
   In the solution: Mapper 1 takes Semester, Timing, Days of the Week and Classroom to give a count of 1 for each entry.
   Mapper 2 takes the output from MR1. For the combination of Semester, Days of Week and Classroom we count the number of time slots for each day that it was occupied.

5. **Find the number of time slots a classroom is occupied on the day of examinations.**
   Contained in folder: ES
   Dataset Used: bina_examschedule.csv (converted given tsv file to csv for readability)

   This question aims to show for each day of an examination how many time slots are occupied for a given classroom. Through this we examine the available slots for a given classroom to schedule other examinations or classes.
   In the solution: Mapper 1 takes Semester, Classroom, Exam dates, capacity, Exam Start Time and End Time. It gives a count 1 for each entry.
   Mapper 2 takes the output of MR1 and gives the count of 1 for each unique combination of Exam Start Time and End Time which later gives the total number of occupied slots for each day of examination.