CSE587: DATA INTENSIVE COMPUTING
PROJECT 2 STAGE 2
REPORT

In this stage, we used multiple Map reduce stages to solve 5 complex problems. The problems were deduced from the classroom schedule dataset.

For problem 3 alone the dataset used was “bina_classschedule2.csv”. For other problems, “bina_classschedule.csv” was used.

PROBLEM 1:

Find the most popular course in every semester based on the number of students enrolled for that course each semester.

SOLUTION:
The solution involves finding the total number of students enrolled for each course in a particular semester using the first map-reduce stage. The key the first reducer is the semester name_course name. The value is the current capacity of students.

The second map-reduce stage calculates the most popular course for a particular year by finding the maximum number of enrolled students. The key to the second reducer is the year and the value is a combination of the total capacities and the corresponding course names.

The focus of this problem is to find which course has the maximum number of enrolled students and which is popular among students.

PROBLEM 2:

Find the topmost used locations in each semester, based on the total number of days in a week for which it is used.

SOLUTION:
The solution involves finding the topmost locations which are used by calculating the number of days (in a total of seven days in a week) for which it is used. The first map-reduce stage calculates the total number of unique days in a week for which that location is used in a semester. The key to the first reducer is the semester name_location name and the value is a string consisting of the unique days in a semester.

Then in the second stage, the key is the year and the value from the second reducer gives the topmost locations arranged in descending order in the format “no of days used_location name”. We could see that only at most 10 locations are used for all 6 working days.

This problem focuses on the trend in how each location is used based on the number of working days.

PROBLEM 3:

Which departments are allocated the least number of rooms every semester?

SOLUTION:
The solution to this problem involves finding the unlucky departments, which are allocated the least number of rooms each semester. In the first stage, the unique rooms assigned to each department for a particular semester is calculated. The key to the reducer is the semester name_department name and the value produced is a string consisting of unique rooms assigned to that department in that semester.

In the second stage, the total of unique rooms assigned to each department is calculated and the department with the least number of rooms allocated is found. The key is the year and the value is in the format "unique room count_department name". It can be seen from the output that many departments even have been assigned just one unique room in a year.

PROBLEM 4:

Find the number of halls with total of maximum capacity ranging from 100 to 150.

SOLUTION:
The solution is to find the total capacity for each hall during every semester and then finding the number of halls that fall within the given range of 100 to 150. The first part is done in the first map-reduce stage. The key to the first stage is the combination of hall name and semester name and the value is the corresponding total capacity for each hall in each semester.

Then in the second stage, the number of halls, which fall in the mentioned range in that year, is calculated by taking year as key and the value is the count of no of halls in that range.

This problem focuses to evaluate the capacity of every building by providing a particular range.

PROBLEM 5:

Find the top 3 busiest day of the week for every semester based on the number of times that day is working for any department in a particular year.

SOLUTION:
This problem aims to find the top 3 busy days of the week based on how many times that day is allocated during a particular semester. The key to the first stage is the semester name_day and the value is the total count. Then in the second stage the key is the semester name and the top 3 days is calculated.

This problem focuses to find the busiest days in a semester so that the most capacity can be allocated during that period.