Report

The objective of this project is to chain multiple map reduce jobs and perform operations to obtain solutions to real world problems.

PROBLEM 1:

To find the increase in average enrollment in capacity over years from the given dataset.

SOLUTION:

We obtained a solution to the problem using two mappers and two reducers. In the first mapper the key is the year and the key is capacity for a particular year. The average enrollment for a particular year is computed using the first reducer.

In the second mapper we have two keys to show the year span. In the second reducer we find the increase or decrease in the average enrollment.

This operation indicates the trend in enrollment every year in the university.

PROBLEM 2:

To find the course with maximum enrollment.

SOLUTION:

We have used two mappers and two reducers. In the first mapper we have taken the course as the key and its capacity as the value. In the reducer we have computed the total capacity for the course over the years. In the second mapper the key and value were split using the delimiter. In the second reducer we have used **setup()** and **cleanup()** methods to print the maximum value. We iterated through the values and obtained the key with maximum enrollment.

PROBLEM 3:

To find the hall with minimum capacity.

SOLUTION:

The hall is used as a key and the capacity is used as the key. The total capacity is obtained for each hall. Here, the hall number is omitted as it is not considered to be of significant importance. A variable is set to a maximum value and every other value is compared to it and the hall with minimum capacity is obtained. The cleanup() and setup() methods are used.

PROBLEM 4:

To display the top 10 courses with maximum enrollment.

SOLUTION:

Two mappers and reducers were used. In the first mapper the key is course and the value is enrollment. In the reducer a TreeMap data structure is used with the capacity as the key and the course name is used as the value. The TreeMap sorts the values in the required order. In the second mapper and reducer the top 10 courses are obtained from the TreeMap by iterating through the map.

PROBLEM 5:

To obtain the number of courses offered in the university.

SOLUTION:

Two mappers and two reducers are used. In the first mapper the course is set as the key and the integer value 1 is set as the value. The courses are added and the output of the reducer is nothing but courses with a value associated with it which tells us how many times the course has appeared in the dataset. In the second reducer the key and value are split using the delimiter and they are passed to the reducer. In the reducer the values are iterated and a count variable is incremented. At the end of the iteration we have a value associated with the count variable which gives us the total count of the courses offered.S