Project Report

for

DBMS Models and Implementation

PROJECT 3 ( MAP/REDUCE IMPLEMENTATION)

**Team Members:**

|  |  |
| --- | --- |
| **Name** | **UTA ID** |
| Rahul Punde | 1001057239 |
| Rahul Raj | 1001051378 |

**OVERALL STATUS :**

The project is ***completed*** and executed successfully.

Brief description of the implementation of major components are given below:

1. Map Class:- A map class is created for which perform filtering and sorting of data. Here Master node takes the input, divides it into smaller sub-problems, and distributes them to worker nodes. A worker node may do this again in turn, leading to a multi-level tree structure. The worker node processes the smaller problem, and passes the answer back to its master node. Map takes one pair of data with a type in one data domain and return a list of pairs in different domain:

Map (k1,v1)-->list(k2,v2)

From the dataset , we return Station ID as key and temperature and windspeed as value.

2. Reduce Class: - The output of the map class is sent as input to reducer class. The master node then collects the answers to all the sub-problems and combines them in some way to form the output – the answer to the problem it was originally trying to solve. Minimum, maximum and average temperature and windspeed for each month for each station in a year is calculated and sent as output.

3. Map2 Class- Another map class is implemented where output of the first reduce class is passed as input to this map class and top 5 maximum and minimum temperature and windspeed and their location for each year is calculated. Here input to the map class is the output of previous problem and output is year as key and the min,max temp/ windspeed as the values with the station details. Here station details are included in the value part.

4. Reduce2 Class- Reduce2 class get the output from map2 class in (key,value) format and do the required computation as it is asked to solve. Reduce2 class calculate the required output and prints them in the format. Here it outputs the top 5 minimum and maximum temperature for every year and output along with the station details. An array is used here to compare the result and find out the required minimum or maximum value.

5. Main function: Input file and output file path is given in the main function. The output from reduce class would be saved in the output folder as described. The folder path has to be given correctly.

**PERFORMANCE MEASURE:-**

The project is completed and time taken for map and reduce function to complete was around 3 minutes.

**DIVISON OF LABOR –**

* Rahul Raj:

Time spent - ~35 hours

Responsibilities – Algorithm, Coding, Design, Debugging

* Rahul Punde:

Time spent - ~35 hours

Responsibilities – Report, Installation, Coding, Debugging

**LOGICAL ERRORS**

1. Key Values pairs- Due to proper identification of key and value pairs, we were facing issue initially in proceeding with the requirements. in the map 1 where station code in included in key it was concatenated with the year and month where as in the second mapper , the station was passed in the value parameter.

2. Return of values from reducer class i.e after calculation of min,max and average temperature and also for wind was a problem. Initially the values were written in two seperate lines which was causing the problem. later on the output is displayed in the single line.

3.Using array and proper looping while calculating the top 5 maximum and minimum value was one of the problem and it was causing an error.