HW3 Q1 Jiawei Zhang

Part a:

1. Mapper Function: the input of map function is the csv file and the output is key-value pairs, where the key is target from the csv file and the value is the weight.

The mapper function actually loads the csv file and then split each line of the csv file and store them into a string. After that, it will go through every line of the string. The mapper function will finally select the second token and the third token from the string and write as a key-value pair as an output.

2. Reducer function: the input of reducer function is the output of the mapper function, which is the key-value pairs, and the output is also key-value pairs, but the value is the minimum value for each key.

The reduce function actually loops each each key and reduce the value by key. It will compare the minimum value of each key and keep the smallest value. Then, it will output the key with only one minimum value.

The inline example input:

100 10 3
110 10 3
200 10 1
150 130 30
110 130 67
10 101 15

The output of the Mapper Function:

10 3
10 3
10 1
130 30
130 67
101 15

The output of the Reducer Function:

10 1
101 15
130 30

Part b:

To solve the problem, it is necessary for us to use the one-to-one joins in the MapReduce Algorithms. Since for both student and department, they share a common key, Department_ID. In Mapper Function, we can implement a secondary sorting to fix the problem of unknown order of values for each key. We can tag the keys with either a “1” or “2” to determine the order of the values. For example, if we tag “1” to student name and tag

“2” to department name, the student name will be in front of the department name in the joinKey-value pairs. In Reducer function, we will take the output from Mapper as an input and loop for each key. We will output the each student name with one department name by joinKey and reduced by joinKey. The student name and the department name is actually value[1] in each value and we will format this and sort the order to output the final result.

Step 1: Loading the tsv file.

Step 2: Mapper Function:

1) Split data and create a list of of the strings

2) Set the Depatment_ID as a joinKey, tag the keys to determine the orders of value and set them as the values for the joinKey.

The input for the Mapper Function:

Student	Alice	1234
Student	Bob	1234
Department	1123	CSE
Department	1234	CS
Student	Carol	1123

The output for the Mapper Function:

1234	student	Alice
1234	student	Bob
1234	Department	CS
1123	student	Carol
1123	Department	CSE

3) Reducer Function:

Take the output from the Mapper and Reduce the them by a joinKey.

The input for the Reducer Function:

1234	student	Alice
1234	student	Bob
1234	Department	CS
1123	student	Carol
1123	Department	CSE

The output for the Reducer Function:

1234 {[(student Alice), (student Bob)], [(Department CS)]}

1123 {[(student Carol)], [(Department CSE)]}

4) Output the final result by selecting the value and sorting.

1123	Carol	CSE
1234	Bob	CS
1234	Alice	CS