BIG DATA MANAGEMENT ASSIGNMENT

Submitted by: Sudip Bala(862188812)

1. Which InputFormat did you use in the MapReduce program?
Ans: FileInputFormat is being used in the program.

2. What is the input and output format of the map function?
Ans: The map function takes input in a text format and reads it line by line. The output is the distance between query points and the points in <key, value> format.

3. What is the logic of the map function?
Ans: In order to measure the proximity of the data and query points, we calculate the distance. The distance is calculated using the Euclidean distance formula. Data points and distance is returned in output along with the ID value.

4. If a combiner function is used, what is the signature of the combiner function (input and output) and what is its logic?
Ans: If a combiner function is used, then it combines all the top 'k' values from different mappers and then the values are sent to the Reducer.

5. If a reduce function is used, what is the signature of the reduce function (input and output) and what is its logic?
Ans: If a reduce function is used, it gets all top k distances and then writes it to the context needed.
Input : Distance(DoubleWritable)Points(Text)
Output : Distance(DoubleWritable)Points(Text)

6. How many mappers and reducers are needed for your program?
Ans: The number of mappers is equal to the number of splits. Here, 4 mappers are used and only 1 reducer is needed.

7. How many records are shuffled between the mappers and reducers?
Ans:

```
10521231 number of records are shuffled by the input of the
combiner.
29142 number of records is shuffled by the input of the reducer.
```