1. Which InputFormat did you use in the MapReduce program?

I define the X, Y of query point sent by user, and they are define as "XQuery", "YQuery". I also define the K number as "K". The path of input file and output file are also given by the command line.

2. What is the input and output format of the map function?

The input format has two inputs. The first one is the sets of points including the ID, x coordinates and y coordinates of each point. The second one is the query point including the x coordinate and the y coordinate.

The output format are the Euclidean distance between the query point and the points, and the information of the points.

3. What is the logic of the map function?

First, get the x and y coordinate of the points.

Second, get the x and y coordinate of the query point.

Third, calculate the Euclidean distance between the point and the query point.

Fourth, map the distance and the information of the points.

4. If a combiner function is used, what is the signature of the combiner function (input and output) and what is its logic?

No combiner function is used in my program.

5. If a reduce function is used, what is the signature of the reduce function (input and output) and what is its logic?

The input of the reduce function is the key-value pairs containing the distance and the information of points. The output of the reduce function is using the format like "The distance is distance The point is ID, x, y" to show the information of k points of P which are closest to the query point.

The logic is reading the number of K by configuration, and using TreeMap to find the smallest key which is the distance. Every time the reduce output a point, the value of K will minus one. If the value of K is no big than zero, then end the program.

6. How many mappers and reducers are needed for your program?

The size of the file is about 445 MB and the size of each block is 128 MB. So the number of the mappers is about 3 or 4. The number of the reducer is 1.

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

The number of the records which are shuffled just like the number of the input records since each record of the input file will be the input of reduces.