Hw2

Submitted by Chengle Zhang and Yihao Zhou

yzhou2@wpi.edu

Zcl8213208@gmail.com

Problem 1:

The code to generate 100MB points and rectangles is written in Step1.java in problem 1 directory. The code to compute spatial join is written in Step2.java.

Compile and run Step1.java, you will get two files with size of 100MB. P.data contains all the points with format of <id,x,y>. R.data contains all the rectangles with format of <id,x,y,w,h>.

To compile and run the Step2.java. Do the following commands:

mkdir problem1\_class/

javac -classpath /usr/share/hadoop/hadoop-core-1.2.1.jar -d problem1\_class Step2.java

jar -cvf problem1.jar -C problem1\_class/ .

hadoop jar problem1.jar problem1.Step2 input/P.data input/R.data output/prob1dir [xmin,xmax,ymin,ymax]

Note in the last command, the window size is optional and should be within the range of 0 - 10,000. If the window is not specified, it will use default window size which is 0 to 10,000 in both x and y.

Problem 2:

To run it, follow the commands below. We also hard coded the split size so that we can have 5 mappers.

mkdir flag\_classes/

javac -classpath /usr/share/hadoop/hadoop-core-1.2.1.jar -d flag\_classes/ ./flag.java

jar -cvf ./flag.jar -C flag\_classes/ .

sudo hadoop jar ./flag.jar org.apache.hadoop.examples.flag /user/hadoop/airfield.txt /user/hadoop/output

Problem 3:

The scripts to generate all points is written in data.java. The scripts to generate k initial points is written in kpoints.java. The default k is 8. If you need modifications, please modify the k value in code. Then just compile and run them. You will get data files k.txt and P.txt.

To compile and run the kmeans.java. Do the following. Note for the last command, you need to specify k number as a parameter, s.t. 8.

mkdir kmeans\_classes/

javac -classpath /usr/share/hadoop/hadoop-core-1.2.1.jar -d kmeans\_classes/ ./kmeans.java

jar -cvf ./kmeans.jar -C kmeans\_classes/ .

sudo hadoop jar ./kmeans.jar org.apache.hadoop.examples.kmeans /user/hadoop/k.txt /user/hadoop/P.txt /user/hadoop/output 8

Please let us know if you encounter any problems.