Milap Patel
Kevin Fortier

Project 2 Problems Build/Evacuation directions

Package is as follows: package project2.problem_set.(java program file name)
Points file should be named: points.txt //parameter file for problems 1, 2, 4
K centroids file should be named: kCentroids.txt //extra parameter file for problem 2

To run the problems in hadoop, please make the following directories in hdfs using following commands: (the "/user/hadoop" part can be different for you machine)
hadoop fs -mkdir /user/hadoop/inputPoints //File directory where the points file will go
hadoop fs -mkdir /user/hadoop/inputKcentroids //File directory where the K centroids file with go

Please upload the files to these directories using the following commands: (the "/user/hadoop" part can be different for you machine)
hadoop fs -put /home/hadoop/Desktop/Proj2/points.txt /user/hadoop/inputPoints
hadoop fs -put /home/hadoop/Desktop/Proj2/kCentroids.txt /user/hadoop/inputKcentroids

Problem 1:

1) Generate Data by running the GenData class. The main method of this class takes 4 parameters (path to points output file, number of points to generate, path to rectangles output, number of rectangles to generate).
   In order to generate two 100 mb files, you will need to create 11 million points and 7 million rectangles.

2) ProblemOne class contains the job for the spatial join. I have noticed that this job gets really slow on the larger data.
   This class takes up to 6 arguments and at least 2. The first two arguments are required (Input data path, output data path). The other 4 requirements are if there is a window (bottom_left_x, bottom_left_y, height, width).

To run the program for problem 2, execute the following commands:
cd to directory to the appropriate directory where the java files are located
mkdir prob2_classes //make folder to compile java file
javac -classpath /usr/share/hadoop/hadoop-core-1.2.1.jar -d prob2_classes ./Proj2Prob2.java //Compile java files

jar -cvf ./Proj2Prob2.jar -C prob2_classes/ . //Make jar file
hadoop jar ./Proj2Prob2.jar hadoop.proj2.Proj2Prob2 /user/hadoop/inputPoints /user/hadoop/outputProb2/output /user/hadoop/inputKcentroids/kCentroids.txt //Run Job
output for the job will be located in the outputProb2 folder in HDFS

Problem 3:

1) Problem three takes 2 arguments (path to json data, path to output directory).

To run the program for problem 4, execute the following commands:
cd to directory to the appropriate directory where the java files are located
mkdir prob4_classes //make folder to compile java file

javac -classpath /usr/share/hadoop/hadoop-core-1.2.1.jar -d prob4_classes ./Proj2Prob4.java //Compile java files

jar -cvf ./Proj2Prob4.jar -C prob4_classes/ . //Make jar file
hadoop jar ./Proj2Prob4.jar hadoop.proj2.Proj2Prob4 /user/hadoop/inputPoints /user/hadoop/outputProb4/output "Radius" "K#" //Run Job
output for the job will be located in the outputProb4 folder in HDFS