Homework 3

Please use mrjob to create a K-means (e.g., K=3) mapreduce class. You just need to implement the first iteration of Kmeans algorithm.

Please submit the code as well as the screenshot to show that you successfully run the job using “-r hadoop”.

You will need two input files:

1. Points with coordinates (x and y). You could write a program to generate these points, e.g., 100 points
2. Coordinates for the initial centroids of the three clusters (K, x and y)

Output:

Coordinates for the new centroids after the iteration