Edmond Wu
Vincent Xie
CS417

Hadoop Project Phase 2 Readme

How we solved this problem:

We solved Phase 2 by using Hadoop's ChainReducer and 2 maps. The first map counts all the instances, as well as counting the total number of results, while the second map uses the results from the first map to calculate the ratios. The reduce function went unchanged.

Challenges:

This assignment took us a few hours more than Phase 1 did, because it took a while to figure out how to properly chain jobs together. Counting the number of wins/draws was fairly simple, but calculating the ratios became fairly challenging as we didn't know how to keep track of the grand total early enough to calculate the ratios. We ended up using the fact that Hadoop outputs the keys in a sorted order to get the total early enough to get the correct ratios.