Project 4: OpenMPI

Michael Wang - mhw1
William Maynes - wmaynes

Section 1 - Parallel Design

Introduction

When looking at how to parallel any algorithm, one of the most direct approaches is to split the work and distribute it to multiple workers. This was the fundamental idea behind the Map-Reduce framework and is the main reason why those frameworks are useful.

When you examine the K-Means algorithm, it is simple to see that the vast majority of the work is done in the last two steps of the algorithm. In one of these steps each data value is associated with one of the means, and in the other the means are recalculated based on the points they are associated with. To parallelize the algorithm, we split up the work in each of these two steps differently.

For Data Associating:

Master:

Section 2 - Experimentation and Analysis
