

COMS3008 - Parallel Computing

Seale Rapolai (1098005)
Mahlekenyane Ts'eole (1118248)
Meriam Elabor (1076589)
Nkosikhona Sibisi (915702)

October 10, 2017

1 Introduction

This report focuses on the common problem of clustering data. It involves classifying different data items that share similar characteristics within the same data set, into groups. By doing this various clusters of data items that share similar characteristics can be identified. One of these problems include the amount of time required to classify each data item into a certain group (cluster). This report will focus solving the performance problem related to clustering.

Since the process of clustering has been known to be quite time consuming, a proposed solution to this is to parallelise the entire process (i.e the process of clustering the input data). Parallelising this process will allow us to divide the amount of work amongst different processing items.

2 Solution technicalities

3 Results analysis

4 Conclusion