Abstract

Multiple Sequence Alignment (MSA) is an important process in the analysis of biological sequences by making comparisons of a number of biological sequences. In some MSA algorithms (such as CLUSTALW, for example), the formation of phylogenetic trees as a guideline in the alignment process has an important role in determining the accuracy of the final alignment results. From the whole MSA process, the formation of phylogenetic tree computational time increased as the number of sequences increases. Computational score similarity for all combinations of sequence pairs carried out sequentially becomes a problem at computational time. This Final Project examines the potential efficiency of computational phylogenetic trees in parallel and is distributed to the Hadoop environment using MapReduce. The results showed that phylogenetic tree can be generated using MapReduce computation.

Keywords: Biology sequence, Multipe Sequence Alignment, Phylogenetic Tree, Hadoop, MapReduce.