

Abstract

Multiple sequence alignment (MSA) is a process of aligning several sequence into a new consensus of sequence. The alignment process carried out by MSA became the initial stage in bioinformatics to produce more biological information compared to *pairwise alignment*. The heuristic approach is an approach that is often used in MSA with a familiar application used, namely clustal. Clustal can do protein chain alignment as much as N-pairs sequentially so that it is consuming quite time when the alignment is processed. Hadoop as a large data computing environment has the ability to process large-sized data faster because of the solving of computing tasks and processing data into small pieces that are processed simultaneously by several clients. By looking at the advantages that Hadoop has, we transform the MSA computing model that is processed sequentially into a new MSA computing model that can be run within the Hadoop environment. From the results of the study, we get the results that average velocity for mapping process is 33,86 char/second, average velocity for reducing process is 12,8 char/second, average velocity for the entire clustalW process in hadoop is 5,49 char/second, and the velocity of entire clustalW process in the hadoop environment is slower 722441,035 times compared to the velocity of clustalW process without using the hadoop environment,

Keywords: *Multiple Sequence Alignment (MSA), Clustal, Hadoop.*