

ABSTRACT

Finds the region of similarity in DNA, RNA or protein sequence on Bioinformatics is used to find structural, functional and evolutionary relationships between organisms. BLAST is a biological sequence similarity analysis tool that compares one sequence to a collection of sequences in the database with computations are performed in pairs for all sequences. Sequence collection enhancement in the database can extend the similarity search process. Hadoop Mapreduce is used as a computational framework that can improve BLAST computing performance because in principle the pairwise comparison operation is independent so that can be paralleled. This final project measures the potential for BLAST computational efficiency by utilizing the hadoop framework. The results showed that the Basic Local Alignment Search Tool (BLAST) built on was speedup and the Hadoop cluster with 3 nodes was 33 times faster than without using Hadoop.

Keywords: Bioinformatics, BLAST, Sequence Alignment, Hadoop, Mapreduce