

Abstract

BLAST (Basic Local Alignment Search Tool) is the most commonly used algorithm for matching DNA chain sequences of organic specimens, NCBI is one of the institution that handle dataset storage and use the algorithm, but in general the enumeration will take a considerable amount of energy and computational time if applied in a large number of datasets. doing algorithm optimization that is universal can change the standardization in bio informatics, therefore optimization is carried out in hardware, processing and database management. In this study, NCBI BLAST was reviewed by Apache Spark platform with Apache Hadoop dataset processing management as an optimization step to optimize the processing of nucleotide protein data sequential search from organic specimens from queries. From the results of this study shows the unaffected processing time of the number of search characters, then the random and damaged sequence data can still be matched well.

Keyword: BLAST, Apache spark, Hadoop