ABSTRACT

Big data is a collection of data on a large scale, which has the characteristics of data varied, very fast growth and complex data. Complex data is unstructured data that needs to be specially processed with an infrastructure that can manage large volumes of data.

In this final project used MapReduce method to facilitate computation to be performed on a big data. MapReduce is a programming model for writing applications that can process a big data in parallel on multiple nodes. MapReduce provides analytical capabilities for analyzing large volumes of complex data. The platform used is Hadoop, Hadoop has its own MapReduce algorithm.

This final project will analyze the performance of Hadoop MapReduce and compare it with Apache Spark is a platform created to process a big data developed based on Hadoop MapReduce with improved processing performance. The scenario used is to process wordcount of a data with different data that aims to analyze the response time and hardware usage of both platforms.

Keyword : Big Data, Apache Spark, MapReduce, Hadoop