Abstract

Nowadays, Implementation of Hadoop clusters for distributed processing of data on a large scale has become a trend. The presence of cluster hadoop very helpful in the field of data processing, some companies are implementing hadoop cluster such as facebook, yahoo and amazon. It is based on the hadoop excess which can have high performance by using simple hardware.

The aim of this research is to implement hadoop cluster by using Wordcount benchmarks as tools to determine the level of performance of this text file type regard to Random Access Memory (RAM) capacity. The types of text files that will be tested is the doc, pdf, csv, xlsx and .txt.

Time trials resulting from the types of the text file showing the best performance level sequence starting from the csv, txt, xlsx, pdf, and doc. The performance enhancement time of all kinds of text files are not proportional with the capacity of RAM, when the RAM capacity increased to 100% of the performance results of the experiment showed doc file increased by 4.58%, pdf file increased by 7.57%, csv file increased by 8.87%, xlsx file increased by 8,35% and txt file increased by 12.82%.

Keywords : Cluster, Hadoop, MapReduce, HDFS, RAM, Bandwidth