

ABSTRACT

Clustering is a technique much used for the distribution and data processing. The purpose of clustering itself is to find the fundamental structure of a data and categorized it to be a bunch of the data in that it has value for learning and analyzed further. A technique grouping and distribution of the data much used now is K-Means Clustering. K-meansCclustering mostly used because they ease in application and results from clustering is a good enough .

In the era of big data which keeps growing like today, the use of technic and analysis of data who is still in traditional or serial may not be efficient in data processing that is the sum of it measured the very large. Therefore the use of a hardware or system as Hadoop would be very helpful in the clustering process for the huge data. Hadoop can be used efficiently to processing data in large numbers because hadoop having an algorithm data processing own called MapReduce. Mapreduce is an algorithm that can be used to overcome size and the amount of data large by doing the distribution and processing data simultaneously .

In this research will be analyzed how the implementation of the use of algorithms MapReduce in K-Means Clustering by using a single node Hdoop who will be compared to processing algorithms K-Means Clustering in sequential by looking at computation or running process time.

Key words : Hadoop , *MapReduce*, *K-Means*, *Mapreduce K-Means*