ABSTRACT

The current server technology continues to evolve, so does the computer cluster technology for various purposes in data processing. By this, author makes the computer clusters and study advantage cluster computer to execute a machine learning algorithm. Thus the presence of this cluster computer will offer services to make the process faster computing.

To create a cluster computer takes several steps ranging from the installation of Apache Hadoop to divide the nodes on each computer. Proceed with the installation of Apache Mahout as software that provides machine learning algorithms for execution on a computer cluster. Zabbix monitoring application that is used to determine the extent to which performance of the CPU load and network traffic ethernet port that is used for machine learning algorithm execution process takes place.

From the results of this final project, the dataset 20 Newsgroups with file size 36 MB, looks average length of time of execution with 1 node reached 22 minutes 12 seconds, 2 nodes reached 20 minutes 42 seconds and 3 nodes reached 18 minutes 3 seconds. While the Wikipedia XML dataset with a file size of 1 GB, seen the average length of time of execution with 1 node reached 22 minutes 6 seconds, 2 nodes reached 18 minutes 39 seconds and 3 nodes reached 15 minutes 6 seconds. The test results show, there are significant time ratio reached an average of 3 minutes and 12 minutes faster when more nodes are added, so three nodes faster than two nodes and one node.

Keywords: Komputer kluster, Apache Hadoop, Apache Mahout, Machine learning, Zabbix