

Abstract

Hadoop is an open source software framework that allows data processing in a distributed from a collection of large data sets on clusters of computer by using a particular programming model. Hadoop consists of 2 core components in solving the problems of Big Data, there are MapReduce and HDFS (Hadoop Distributed File System).

These two core components, the HDFS is an important role in managing File System. When MasterNode which has Namenode and Jobtracker roles got a failure, then the File System can not serve user client. Therefore, there is a solution for it by making a cloning of MasterNode on another server so that when MasterNode experiencing failure then it could be replaced by its cloning called StandbyNode.

This research applying Zookeeper and QJM method on Hadoop cluster to be able to improve on Hadoop High Availability with failover downtime ranges from 4 seconds to 63 seconds. When compared with DRBD and Heartbeat method then Zookeeper and QJM method used today is better to use in a long term. Also by using Zookeeper and QJM, downtime failover relatively stable without increasing the downtime towards the number data blocks.

Key Words : Hadoop, High Availability, NameNode, Failover, Downtime