

Abstract

In this modern era, big data have an important role in the part of technology. To handling small data usually just using a simple data management, but to handling a big amount of data required a better data management, one of them is Hadoop. Hadoop itself consists of two layer, MapReduce and Hadoop Distributed File System (HDFS). MapReduce composed by map and reduce that are use to process a data, while Hadoop Distributed File System (HDFS) is the directory where hadoop data is saved.

To run a job on Hadoop An algorithm is needed to complete that job. In this paper the Delay Scheduling algorithm and Task Classification Dynamic Delay Scheduling (TCDDS) algorithm are proposed. Delay Scheduling works by delaying the next job to improve the previous locality data, while Task Classification Dyanmic Delay Scheduling (TCDDS) algorithm is an algorithm that works by dividing the delay time into 3 main priorities namely Low Priority Task, Medium Priority Task, and High Priority Task.

The test results show that the Task Classification Dyanmic Delay Scheduling (TCDDS) algorithm has 94,63 minutes faster job completion time for the average completion time parameter and has a better total job time per minute of 10,6 for the job througput parameter than the Delay Scheduling algorithm, but for the fail task rate parameter is better the delay scheduling algorithm by 0.15% than the Task Classification Dynamic Delay Scheduling (TCDDS) algorithm .

Keyword: Hadoop, Task Classification Dynamic Delay Scheduling (TCDDS), Delay Scheduling
