

Abstract

Hadoop is a *Java*-based programming framework that is often used to process large amounts of data in a computationally distributed. The two main structures of Hadoop are *MapReduce* and *Hadoop Distributed File System (HDFS)*. The increase in *MapReduce* itself depends heavily on optimizing its scheduling model. Processing large amounts of data often results in waste of resources and inefficient use of time which impacts *MapReduce's* performance, therefore *job* scheduling algorithms are needed. However, the *job* scheduling algorithm does not guarantee allocation when data is processed. For this problem, *PCAD (Adaptive Delay Schedule Algorithm Based on Progress Control)* is used to maximize data location performance from the perspective of delay scheduling.

The test results show that the *Adaptive Delay Schedule (PCAD)* algorithm has a total completion time of 70,12 minutes for *average completion time* parameters and has a better total *job* time per minute of 2,06 for *job throughput* parameters than the *Delay Scheduling* algorithm, while in the *fail task rate* parameter the algorithm is superior to *Delay Scheduling* by 0,93% than the *Adaptive Delay Schedule (PCAD)* algorithm.

Keyword: Hadoop, Data Locality, Delay Scheduling, Adaptive Delay Schedule (PCAD), Progress Control