

Abstrak

Pada era sekarang ini, *Big Data* memiliki peran penting pada bidang teknologi. Dalam menangani data yang berukuran kecil biasa digunakan data *management* yang sederhana, tetapi untuk menangani data yang berukuran besar di butuhkan sebuah data *management* yang lebih baik, salah satu contoh nya yaitu Hadoop. Hadoop terdiri dari dua layer yaitu *MapReduce* dan *Hadoop Distributed File System (HDFS)*. *MapReduce* terdiri dari *map* dan *reduce* yang digunakan untuk pemrosesan data, sementara *Hadoop Distributed File System (HDFS)* adalah tempat direktori dimana data hadoop disimpan.

Untuk menjalankan sebuah *job* pada Hadoop di perlukan sebuah algoritma untuk menyelesaikan suatu *job*. Dalam paper ini diusulkan algoritma *Delay Scheduling* dan algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)*. Algoritma *Delay Scheduling* bekerja dengan menunda jalannya *job* selanjutnya untuk memperbaiki data lokalitas sebelumnya, sedangkan algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)* adalah sebuah algoritma yang bekerja dengan membagi waktu tunggu menjadi 3 *priority* utama yaitu *Low Priority Task*, *Medium Priority Task*, dan *High Priority Task*.

Pada hasil pengujian menunjukan algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)* mempunyai waktu total penyelesaian *job* lebih cepat 94,63 menit untuk parameter *average completion time* dan mempunyai waktu total pengerjaan *job* per menit lebih baik sebesar 10,6 untuk parameter *job througput* dari pada algoritma *Delay Scheduling*, sedangkan pada parameter *fail task rate* lebih unggul algoritma *Delay Scheduling* sebesar 0,73% dari pada algoritma *Task Classification Dyanmic Delay Scheduling (TCDDS)*.

Kata Kunci : Hadoop, Task Classification Dynamic Delay Scheduling (TCDDS), Delay Scheduling
