Abstrak

Hadoop merupakan sebuah framework pemrograman berbasis *Java* yang seringkali digunakan untuk mengolah data dalam jumlah yang besar secara terdistribusi berbasis komputasi. Dua struktur utama dari Hadoop adalah *MapReduce* dan *Hadoop Distributed File System* (HDFS). Peningkatan pada *MapReduce* sendiri sangat bergantung pada optimalisasi model penjadwalannya. Pengolahan data dalam jumlah besar seringkali menimbulkan pemborosan sumber daya serta penggunaan waktu yang tidak efisien sehingga berdampak pada kinerja *MapReduce*, oleh karena itu dibutuhkan algoritma *job* scheduling. Namun algoritma *job* scheduling tidak menjamin pengakolokasian saat data diproses. Untuk masalah ini, digunakan *PCAD (Adaptive Delay Schedule Algorithm Based on Progress Control)* kemajuan untuk memaksimalkan kinerja lokasi data dari sudut pandang optimasi delay scheduling.

Pada hasil pengujian menunjukan algoritma *Adaptive Delay Schedule* (PCAD) mempunyai waktu total penyelesaian *job* lebih cepat 70,12 menit untuk parameter *average completion time* dan mempunyai waktu total pengerjaan *job* per menit lebih baik sebesar 2,06 untuk parameter *job througput* dari pada algoritma *Delay Scheduling*, sedangkan pada parameter *fail task rate* lebih unggul algortima *Delay Scheduling* sebesar 0,93% dari pada algoritma *Adaptive Delay Schedule* (PCAD).

Kata Kunci : **Hadoop, Data Locality, *Delay Scheduling, Adaptive Delay Schedule (PCAD),* Progress Control**