

1. PENDAHULUAN

Dalam beberapa kurun waktu kebelakang ini salah satu aspek terpenting pada dunia teknologi adalah *Big Data*. Istilah *Big Data* itu sendiri mengacu pada pertumbuhan data yang sangat besar yang sehingga susah untuk diolah oleh data *management* sederhana. Dan untuk menangani masalah ini maka dibutuhkan sebuah data managemet yang lebih baik, salah satu contoh nya yaitu Hadoop. Hadoop merupakan *framework software* berbasis *java* dan *open source* yang berfungsi untuk mengolah data yang besar secara terdistribusi dan berjalan di atas *cluster* yang terdiri atas beberapa komputer yang terhubung [1]. Arsitektur Hadoop terdiri dari dua layer yaitu layer *MapReduce* dan layer *Hadoop Distributed File System (HDFS)* [2]. *Mapreduce* merupakan *framework* dari aplikasi yang terdistribusi dan *Hadoop Distributed File System (HDFS)* merupakan data yang terdistribusi. Algoritma *job scheduler* yang terdapat dalam *MapReduce* berguna untuk mengatur setiap *job* yang berjalan pada sistem Hadoop. Tiap *job* yang diolah pada *job scheduler* memiliki sifat yang berbeda [3]. Dalam membantu tugasnya menjalankan pengolahan data yang sangat besar Hadoop di bantu oleh *job scheduler*. *Job scheduler* merupakan suatu algortima yang berjalan pada Hadoop untuk membantu Hadoop menjalankan tugasnya secara efisien. Dalam Hadoop memiliki banyak *job scheduler*, dan *job scheduler* utama dari Hadoop ialah algoritma *Fair Scheduler*, algoritma *Capasitive Scheduler* dan algoritma *FIFO Scheduler*. Seiring berjalannya waktu ketiga *scheduler* tersebut sudah banyak di tinggalkan karena tidak dapat menunjang performansi dari Hadoop tersebut, maka dikembangkan lah banyak *scheduler* dari ketiga *scheduler* tersebut. Salah satu *scheduler* yang banyak digunakan saat ini ialah algoritma *Delay Scheduling* karena algoritma *Delay Scheduling* dapat menguraikan *job* berdasarkan waktu tunggu. Namun, seiring berkembangnya hadoop algoritma *Delay Scheduling* juga memiliki kekurangan, maka dikembangkan lah algoritma *Delay Scheduling* yaitu algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)*.

Pada paper ini dilakukan pengujian terhadap algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)* dan *Delay Scheduling*. Algoritma *Task Classification Dynamic Delay Scheduling (TCDDS)* ini bekerja dengan membagi *job* kedalam tiga *priority* yaitu *high priority task*, *normal priority task* dan *low priority task*. Sedangkan algoritma *Delay Scheduling* bekerja dengan metode menunda jalannya jobs selanjutnya untuk memperbaiki data lokalitas sebelumnya. Pada pengujian ini dilakukan pada dua tipe *job* yaitu *job* homogen (satu jenis *job*) dan *job* heterogen (beberapa jenis *job* yang dijalankan bersamaan). Dan parameter pada pengujian ini yaitu *average completion time*, *job throughput*, dan *task fail rate*.