

Abstrak

Hadoop merupakan *framework* software berbasis *java* dan *open-source* yang berfungsi untuk mengolah data yang besar secara terdistribusi dan berjalan diatas *cluster* yang terdiri dari beberapa gabungan komputer. Arsitektur Hadoop terdiri dari 2 *layer* pokok, yaitu *layer MapReduce* dan *layer Hadoop Distributed File System (HDFS)*. Map Reduce adalah komponen service kunci yang berfungsi untuk melakukan proses komputasi *Big Data* secara paralel dan terdistribusi dan (HDFS) berfungsi untuk menyediakan *bandwidth* sangat tinggi yang di agregasi ke semua *cluster (node)*.

Dalam *MapReduce* terdapat terdapat *job scheduler* yang berfungsi untuk memetakan antrian *job* yang masuk. *Job scheduler* default dari Hadoop adalah FIFO dan Hadoop mengizinkan penggantian *job scheduler* default dengan *custom job scheduler*. *Capacity Scheduling* merupakan *job scheduler* pada Hadoop yang berkarakteristik memberikan *capacity guarantee* kepada antrian yang masuk pada *queue* yang telah disediakan sehingga bisa ditekannya nilai Fail Rate. Tetapi karena *resource* harus dibagi menjadi beberapa bagian maka performansi *Response Time* dan *Job Throughput* menurun. Pada algoritma FIFO nilai maksimal *Job Fail rate* yaitu 10%, sedangkan pada *Capacity Scheduling* nilai *Job Fail Rate* maksimal adalah 4,3%

Kata kunci : Hadoop, data, *multi-user*, *Capacity Scheduling*, FIFO