

INTISARI

ANALISIS KINERJA ALGORITMA JOB SCHEDULING PADA YARN CLUSTER

Oleh

Ilham Sulaksono
13/347352/PA/15198

Algoritma Job Scheduling pada *cluster* sangat mempengaruhi kinerja dari suatu sistem *cluster*. Dalam *Hadoop cluster*, framework yang berfungsi untuk *job scheduling* terdapat dalam *YARN*. *YARN* memiliki algoritma job scheduling berupa *fair* dan *capacity scheduling*. *Fair scheduling* melakukan pembagian *resource* secara merata untuk tiap *job*. Sedangkan *capacity scheduling* melakukan pembagian *resource* berdasarkan pada kapasitas *queue* yang dikonfigurasi.

Penelitian ini melakukan analisa pada kinerja dari *job scheduling* yang terdapat dalam *YARN*. Analisa dilakukan dengan mengobservasi hasil *benchmarking* dengan menggunakan *HiBench*. Parameter yang dijadikan sebagai penilaian kinerja adalah *waiting time*, *execution time*, serta *CPU* dan *memory fairness*. Pengujian dilakukan 3 kali skenario, dimana skenario 1 dengan menggunakan tipe dan ukuran *workload* yang berbeda serta skenario 2 dan 3 menggunakan konfigurasi yang berbeda.

Penelitian ini menghasilkan kinerja dan hubungan antara konfigurasi dengan kinerja yang dihasilkan untuk tiap *job scheduling* yang dianalisis. Penelitian ini mendapatkan hasil bahwa kinerja dari *fair scheduling* lebih optimal daripada *capacity scheduling* dengan berdasarkan pada *execution time* rata-rata 231 detik untuk *fair scheduling* dan 252 detik untuk *capacity* pada skenario 1 menggunakan “*Huge*” *benchmark*. Dengan menggunakan 3 *workload* bersamaan, parameter konfigurasi dari *fair scheduling* yang berhubungan dengan kinerja secara signifikan adalah *assignmultiple*. Konfigurasi *assignmultiple* memiliki nilai korelasi koefisien dengan *execution time* sebesar 0,847782. *Capacity scheduling* memiliki konfigurasi dengan nilai korelasi yang signifikan, yaitu *capacity* dan *maxcapacity*. Nilai korelasi untuk *execution time* dengan *capacity queue* “a” adalah -0,9742. Sedangkan, untuk *queue maxcapacity* adalah -0,43275.

Kata Kunci: *Hadoop*, *YARN*, *Job Scheduling*, *Cluster Computer*, Analisa kinerja

ABSTRACT

PERFORMANCE ANALYSIS OF JOB SCHEDULING ALGORITHM ON *YARN CLUSTER*

By

Ilham Sulaksono
13/347352/PA/15198

Algorithm of Job Scheduling on cluster greatly affects the performance of a cluster system. In Hadoop cluster, framework for job scheduling is contained in the YARN. YARN has a job scheduling algorithm, fair and capacity scheduling. Fair scheduling algorithm allocating cluster resource evenly to each job. While capacity scheduling algorithm allocating cluster resource based on the capacity of the queue that is configured.

This research analyzing the performance of the job scheduling that contained in YARN. The analysis is done by observing the results of benchmarking by using *HiBench*. The parameters that serve as a metrics is waiting time, execution time, CPU and memory fairness. The testing was done with 3 scenario, where scenario 1 using different type and size of *workload* and scenario 2 and 3 with different configuration.

This research resulted in performance and the relationship between configurations and performance for each job scheduling. This research have result that performance from fair scheduling more optimal than capacity scheduling with average of waiting time 231 second for *fair scheduling* and 252 second for *capacity* using “Huge” benchmark first scenario. The configuration of the fair scheduling that affect performance in a significant way is assignmultiple with correlation coefficient value for *waiting time* is 0.847782. While capacity scheduling is queue capacity and maxcapacity. The value of correlation coefficient for execution time with capacity queue “a” -0.9742. And, correlation for queue maxcapacity with execution time is -0.43275.

Keywords: *Hadoop*, *YARN*, Job Scheduling, Cluster Computer, performance Analysis