

ABSTRACT

PERFORMANCE ANALYSIS OF FIFO AND FAIR JOB SCHEDULING ALGORITHM ON *SPARK* CLUSTER BY USING *SPARK-BENCH* BENCHMARK

Irvin Habibie Iskandar

16/392769/PA/17073

In recent years, the needs of *Big Data* increase rapidly due to maintaining business data needed. Big Data framework that has goals to analyze the data and later for visualizing the data. By using Apache Spark we can overcome the solutions, along with cluster computer. *Spark* itself does have two *job scheduling* algorithms, *FAIR* scheduling algorithm where each job will get equal distribution of resources per each job, and *FIFO* is a job scheduling algorithm that every time a job is approaching, the first job who sends it first will have priority compared to latter jobs submitted.

By using benchmarking, Spark-Bench, we could analyze the performance of job scheduling in Apache Spark. This research consisted of 3 scenarios, the first scenario comparing the performance of Fair and FIFO job scheduling, scenario 2 were comparing each parameter configuration performance along with correlation with the same job scheduling, FIFO and Fair. Scenario 3 is where a combination of several parameters in job scheduling algorithms with correlation. The evaluation metrics used in this research are Total runtime and execution runtime.

Based on scenario 1, Fair outperform FIFO , 1.6184 and 1.6589 seconds respectively. In scenario 2, overall, Fair outperform FIFO. Parameter configurations that have highest Pearson correlations both in Fair job scheduling and FIFO job scheduling is 0.1606. In the last scenario, Fair is performed better than FIFO. Be careful with parameter configurations, some of them will slow down the job scheduling.