

Abstrak

Hadoop merupakan sebuah *framework software* yang bersifat *open source* dan berbasis *java* yang banyak digunakan oleh perusahaan untuk pemrosesan data yang sangat besar (*big data*). Hadoop terdiri atas dua arsitektur utama, yaitu *MapReduce* dan *Hadoop Distributed File System* (HDFS). *MapReduce* terdiri atas *Map* dan *Reduce* yang digunakan untuk pemrosesan data, sementara *Hadoop Distributed File System* (HDFS) adalah sebuah tempat atau direktori dimana data hadoop dapat disimpan. Hadoop memiliki bermacam-macam *job scheduler* untuk mengatur *job* yang sesuai dengan karakteristiknya. Salah satunya algoritma *Self Adaptive Reduce Scheduling* (*SARS*).

SARS menggunakan sebuah metode yang akan memperlambat proses *reduce* sehingga tidak mengikuti dari awal proses *jobs* akan masuk, proses menunda waktu *reduce* bertujuan untuk mengurangi *average completion time* pada berbagai *jobs* yang akan masuk. Algoritma *scheduling Bipartite Graph Oriented Locality Aware Scheduling* merupakan *scheduler* yang menggunakan prinsip *graph matching problem* pada logika *Kuhn munhkers* untuk dapat menemukan pencocokan yang optimal terhadap *resources* yang dimiliki dengan mengelompokkan data sesuai dengan karakteristik *job* dengan bertujuan dapat mengurangi *average completion time*.

Performansi algoritma *BOLAS* terlihat lebih efektif daripada algoritma *SARS* pada jenis *job wordcount* dengan nilai *fail* sebesar 1,30%, lama *average completion time* 265,98 menit mengakibatkan nilai *job throughput* sebesar 5,83 *job*/menit lebih baik 2% dibandingkan *SARS*. Hal ini disebabkan karena algoritma *BOLAS* memiliki keunggulan dengan menggunakan Teknik *bipartite matching problem* pada algoritma *Kuhn munhkers* untuk membagi *job* sesuai dengan karakteristik *job* yang kemudian di proses dalam *Map Reduce*.

Kata kunci: Hadoop, *scheduler*, *BOLAS*, *SARS*, *job*