

1. Pendahuluan

1.1 Latar Belakang

Semakin meningkatnya suatu data akan mempengaruhi dalam pengaksesan data terutama dalam hal pendistribusiannya. Besarnya suatu data atau biasa dikenal dengan *Big Data* akan mempengaruhi dalam hal kecepatan dan ketepatan pengaksesan serta pengolahan data itu sendiri. Kecepatan dan ketepatan pengolahan data yang besar akan sangat dibutuhkan oleh pengguna *Big Data*. Hadoop merupakan sebuah *framework* pemrograman berbasis Java yang mendukung penempatan dan pemrosesan suatu data set yang besar dalam pendistribusian berbasis komputasi dan ini sangat sesuai dengan data yang memiliki volume yang tinggi [4]. Dalam pengaplikasiannya Hadoop memiliki kelebihan dimana dari sisi ekonomisnya, hal ini dikarenakan Hadoop merupakan suatu aplikasi *freeware* atau tidak berbayar, selain itu Hadoop dapat diimplementasikan pada berbagai perangkat yang tidak memakai spesifikasi terlalu tinggi. Dalam hadoop terdapat MapReduce dimana MapReduce merupakan suatu model pemrograman dan merupakan suatu penggabungan implementasi untuk pemrosesan dan pengerjaan suatu data set yang besar [2].

Dalam MapReduce ini terdapat beberapa algoritma *Job Scheduler* yang mana digunakan untuk pemetaan suatu data. Algoritma *job scheduler* ini merupakan suatu fungsi yang hadoop miliki dimana kegunaannya untuk meningkatkan kualitas *job* pada sistem [3]. Terdapat beberapa keragaman terhadap *job* pada MapReduce seperti peningkatan kualitas *job* berdasarkan perbedaan waktu eksekusi dan *resource* data yang berjalan pada *jobs*. Atas beberapa keberagaman terhadap *job* pada MapReduce ini maka dilakukan beberapa penyesuaian mengenai *job scheduler* dengan melihat *scheduler* apa yang dipakai sesuai karakteristik *job* pada suatu sistem yang akan dibangun.

FIFO (First In First Out) dan *SARS (Self Adaptive Reduce Scheduling)* merupakan *job scheduler* pada sistem hadoop serta terdapat beberapa ciri tersendiri dalam hal menangani antrian untuk *job scheduling* [4]. *FIFO* dalam penerapannya berdasarkan atas sistem antrian *job* yang datang pertama kali

sehingga proses penyelesaian akan terlihat lama. Sedangkan pada SARS menerapkan sistem memperlambat proses *reduce* dan mengurangi waktu proses penyelesaian *job* yang bekerja serta mengurangi *response time* terhadap *job* [4].

Pada tugas akhir ini akan dibahas penggunaan *FIFO* dan *SARS Scheduler* sebagai algoritma penjadwalan pada sistem hadoop. Penggunaan algoritma ini akan dibandingkan dengan pengimplementasian pada berbagai macam karakteristik job dengan parameter *Job Fail Rate*, *Job Throughput*, dan *Response Time* sebagai acuan performansi sistem. Serta juga dilakukan klasifikasi dan analisis *job scheduler* berdasarkan performansi sistem terhadap karakteristik *job* pada setiap *job scheduler*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat beberapa perumusan masalah yang muncul pada tugas akhir ini, yaitu sebagai berikut:

- a. Bagaimana kualitas performansi layanan algoritma *FIFO* dan algoritma *SARS Scheduling* pada Hadoop dalam berbagai macam karakteristik job pada sistem Hadoop.
- b. Bagaimana klasifikasi *job scheduler FIFO* dan *SARS Scheduling* terhadap berbagai macam karakteristik job pada sistem hadoop.

1.3 Batasan Masalah

Adapun beberapa batasan masalah yang digunakan pada tugas akhir ini, yaitu sebagai berikut:

- a. *Server* berjumlah dua buah yang menjadi *server* utama pada Hadoop (*Head* dan *Slave*) dan berbasis Linux.
- b. *User client* yang berjumlah satu buah yang akan mengakses *job*.
- c. Penggunaan algoritma penjadwalan hanya menggunakan algoritma *FIFO* dan algoritma *SARS Scheduling*.
- d. Parameter yang menjadi pembanding yaitu *Job Fail Rate*, *Job Throughput*, dan *Response Time*.
- e. Menggunakan tiga jenis job yaitu *job wordcount*, *job grep*, dan *job randomtextwriter*.

1.4 Tujuan

Tujuan dari tugas akhir ini adalah:

- a. Dapat memaparkan kualitas layanan *FIFO* dan *SARS Scheduling* pada Hadoop dalam berbagai macam karakteristik job pada sistem Hadoop.
- b. Dapat mengklasifikasi *job scheduler FIFO* dan *SARS Scheduling* terhadap berbagai macam karakteristik *job* pada sistem Hadoop.

Hipotesa awal:

Permasalahan utama pada tugas akhir ini adalah penjadwalan *jobs* yang dilakukan pada Hadoop. Pada metode algoritma *FIFO* memiliki karakteristik yang menjalankan *job* pertama kali masuk dan akan dijalankan sampai selesai lalu dilanjutkan dengan *job* selanjutnya, untuk proses *reduce* yang dimiliki pada algoritma *FIFO* akan mulai berjalan tidak terlalu jauh dari dimulainya *run maps*. Sedangkan pada *SARS Scheduling* dapat menggunakan metoda dimana akan memperlambat durasi dari *reduce* dan akan mulai berjalan jauh setelah dimulainya *run maps*, sehingga dapat mengurangi *response time* dan meningkatkan *job throughput*.

1.5 Metodologi Penyelesaian Masalah

Pembuatan tugas akhir ini dilakukan dengan menggunakan metodologi penyelesaian masalah sebagai berikut:

1. Studi literatur

Studi literatur merupakan suatu proses dimana akan dilakukan pencarian, pengumpulan, penyaringan, pembelajaran, pendalaman, dan pemahaman literatur yang berhubungan dengan Hadoop, *job scheduling*, algoritma *SARS Scheduler*, *MapReduce*. Beberapa *literature* berupa buku dan *paper* yang mendukung teori dan konsep tersebut adalah:

- a. *A Comprehensive View of Hadoop MapReduce Scheduling Algorithms*
- b. *Recent Job Scheduling Algorithms in Hadoop Cluster Environments: A Survey*
- c. *A Self-adaptive Scheduling Algorithm for Reduce Start Time.*

- d. *Job Scheduling Optimization for Multi-User MapReduce Clusters*
- e. *MapReduce: Simplified Data Processing on Large Clusters*
- f. *Survey on Improved Scheduling in Hadoop MapReduce in Cloud Environments*
- g. *Hadoop: The Definitive Guide*
- h. *Job Scheduling for Multi-User MapReduce Clusters*

2. Analisis perancangan dan desain jaringan

Analisis perancangan dan desain jaringan dalam tugas akhir ini adalah melakukan pengujian atau menganalisis serta pembangunan jaringan.

3. Implementasi

Implementasi dalam tugas akhir ini adalah penginstalan Hadoop pada dua *server* utama dan sepuluh komputer *user* yang nantinya dapat membantu memberikan *jobs* pada *server* utama. Pada *server* utama selanjutnya akan diujikan implementasi sebuah algoritma *SARS Scheduling*.

4. Pengujian sistem

Pengujian sistem dalam tugas akhir ini ialah semua perangkat komputer *user* akan memberikan *jobs* pada *server* utama dalam waktu yang bersamaan sesuai dengan skenario yang telah dibuat. Selanjutnya melakukan konfigurasi dengan menggunakan algoritma *default* Hadoop (*FIFO*) dengan algoritma *SARS Scheduling* pada *server* utama.

5. Analisis hasil pengujian

Analisis hasil pengujian dalam tugas akhir ini adalah menganalisis hasil pengujian sistem yang telah dilaksanakan sebelumnya dengan penggunaan parameter yang telah ditentukan.

6. Pembuatan laporan

Pembuatan dalam tugas akhir ini yaitu dengan melakukan dokumentasi dari hasil pengujian sistem dengan penulisan yang benar dan telah sesuai dengan ketentuan-ketentuan yang telah ditetapkan secara umum.