ANALISIS PERBANDINGAN KINERJA ALGORITME *FAIR SHARE SCHEDULING* DENGAN *CAPACITY SCHEDULING* TERHADAP PENGIRIMAN *JOB* PADA HADOOP *CLUSTER* MULTI*NODE*

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Friska Anggia Nurlita Sari
NIM: 135150201111035

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

ANALISIS PERBANDINGAN KINERJA ALGORITME *FAIR SHARE SCHEDULING* DENGAN *CAPACITY SCHEDULING* TERHADAP PENGIRIMAN *JOB* PADA HADOOP *CLUSTER* MULTINODE

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Friska Anggia Nurlita Sari
NIM: 135150201111035

Skripsi ini telah diuji dan dinyatakan lulus pada
17 Januari 2018
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I	Dosen Pembimbing II
Adhitya Bhawiyuga, S.Kom, M.S	Achmad Basuki, S.T, M.MG, Ph.D
NIK : 201405 890720 1 001	NIP : 19741118 200312 1 002

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP : 19710518 200312 1 001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 17 Januari 2018

Friska Anggia Nurlita Sari
NIM: 135150201111035

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi ini. Semoga rahmat dan hidayah-Nya selalu dilimpahkan kepada kita semua. Tidak lupa shalawat serta salam selalu penulis haturkan kepada Rasulullah, Nabi Muhammad SAW.

Skripsi yang berjudul "*Analisis Perbandingan Kinerja Algoritme Fair Share Scheduling dengan Capacity Scheduling terhadap Pengiriman Job pada* Hadoop *Cluster Multinode"* ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) di Fakultas Ilmu Komputer, Jurusan Informatika, Universitas Brawijaya. Penulis menyadari bahwa skripsi ini dapat diselesaikan berkat dukungan dan bimbingan dari beberapa pihak. Oleh karena itu dengan segala kerendahan hati penulis menyampaikan terima kasih kepada:

1. Untuk orang tuaku tercinta, Bapak Sayono, S.E. dan Ibu Utari Margawati, S.E. Terima kasih atas dukungan moral maupun materi, motivasi, doa yang selalu diberikan kepada penulis untuk dapat menyelesaikan skripsi ini.
2. Bapak Adhitya Bhawiyuga, S.Kom, M.S selaku Dosen Pembimbing I dan Bapak Achmad Basuki, S.T, M.MG, Ph.D selaku Dosen Pembimbing II, atas kesabaran dalam membimbing penulis, memberikan masukan, motivasi dan ilmu yang berharga untuk penulis dalam menyelesaikan skripsi ini.
3. Untuk saudaraku tercinta, Fighi Fridyawati dan Wahyu Aji Bimantoro yang selalu memberikan motivasi, semangat, dukungan dan doa kepada penulis dalam menyelesaikan skripsi ini.
4. Untuk teman-teman SMA, Risna Dwi Kartika, Lailatul Rohmania, Wiwit, Sarah Christiawan, Fadhilatus Solicha, Gita Larasati, dan Hilda Sukma Puspita yang selalu memberikan dukungan, semangat dan motivasi kepada penulis.
5. Untuk teman-teman, Fidia Rosianti, Rinindya Nurtiara Puteri, Pratomo Adinegoro, dan Labib Imam Hidayat yang selalu bersama dari awal menutnut ilmu dikampus ini, suka, duka, semuanya telah kita rasakan bersama. Terima Kasih untuk dukungan, semangat, motivasi, nasihat kalian selama ini.
6. Untuk rekan-rekan asisten praktikum Basis Data Laboraturium Pengajaran atas dukungan dorongan serta doa kepada penulis untuk segera menyelesaikan skripsi ini dengan lancar dan dapat diberi kemudahan.
7. Untuk rekan-rekan Eksekutif Mahasiswa Informatika periode 2016-2017 yang selalu memberikan masukan, dukungan semangat, dan motivasi kepada penulis dalam menyelesaikan skripsi ini.
8. Untuk teman-teman Kos Kertosariro 59 yang selalu memberikan motivasi dan semangat kepada penulis dalam menyelesaikan skripsi ini.
9. Untuk seluruh teman-teman Jurusan Informatika angkatan 2013 yang telah membantu dan selalu memberikan motivasi kepada penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih belum sempurna, oleh karena itu saran dan kritik sangat diperlukan untuk dapat membantu dalam pembuatan

penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan dapat menjadi acuan dan sebagai masukan untuk penelitian selanjutnya.

Malang, 17 Januari 2018

Penulis

friskanggi@gmail.com

ABSTRAK

Hadoop merupakan sebuah *framework* berbasis *open-source* digunakan untuk pengolahan data dengan skala besar dan tersimpan dalam sekelompok komputer yang saling terhubung dalam suatu jaringan secara terdistribusi. Hadoop terdapat beberapa algoritme penjadwalan untuk mengatur antrian *job* yang masuk, *job* yang telah di-*submit* oleh pengguna akan saling berkompetisi untuk merebutkan suatu *resource* yang tersedia. Sehingga perlu adanya algoritma penjadwalan yang bertugas mengatur jalannya *job* pada sebuah *resource* serta mengelola pemrosesan data agar hasil yang dikeluarkan sesuai dengan apa yang diharapkan. Penelitian ini dilakukan untuk mengetahui perbandingan dari kinerja algoritme *Fair Share Scheduling* dan *Capacity Scheduling* pada pengiriman *job* pada Hadoop *multinode* untuk mengetahui nilai dari parameter *job fail rate, latency,* dan *throughput*. *Capacity Scheduling* mendukung antrian secara hirarki yaitu dapat membagi resource yang tersedia pada *cluster* ke beberapa antrian. *Fair Share Scheduling* adalah metode alokasi pembagian resource pada antrian secara adil untuk seluruh *job* yang masuk pada antrian. Berdasarkan hasil pengujian yang telah dilakukan, memiliki nilai minimal *Job Fail rate* yang lebih baik dibandingkan *Fair Share Scheduling* sebesar 0,99%. *Fair Share Scheduling* memiliki nilai minimal *Latency* yang lebih baik sebesar 19,25 menit. Sedangkan nilai Throughput pada *Fair Share Scheduling* lebih cepat sebesar 0,47 Mbit/s.

Kata kunci: Hadoop, Hadoop Multi-*node*, *Fair Share Scheduling, Capacity Scheduling, Job Fail rate, Latency,* dan *Throughput.*

ABSTRACT

Hadoop is a framework based on an open-source which is used for a data processing in a large scale and stored in a group of interconnected computers that are connected in a network which are distributed. Hadoop there are several scheduling algorithms to set the incoming job queue, jobs that have been submitted by the user will compete for a resource available. So the need for a scheduling algorithm in charge of managing the job on a resource and managing the data processing so that output issued in accordance with what is expected. This research was conducted to find out the comparison of performance of Fair Share Scheduling and Capacity Scheduling algorithm on job delivery on Hadoop multinode to know the value of job fail rate, latency, and throughput parameters. Capacity Scheduling supports the queue in a hierarchy that is able to divide the available resources in the cluster into multiple queues. Fair Share Scheduling is a fair allocation of resource allocation methods for queues for all jobs that enter the queue. Based on the results of tests that have been done, has a minimum value of Job Fail rate is better than Fair Share Scheduling of 0.99%. Fair Share Scheduling has a minimum value of better Latency of 19.25 minutes. While the value of Throughput on Fair Share Scheduling faster by 0.47 Mbit/s.

Keywords : Hadoop, Hadoop *Multi-node, Fair Share Scheduling, Capacity Scheduling Job Fail rate, Latency,* dan *Throughput.*

DAFTAR ISI

PENGESAHAN ii
PERNYATAAN ORISINALITAS ii
KATA PENGANTAR iv
ABSTRAK vi
ABSTRACT vii
DAFTAR ISI viii
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
DAFTAR LAMPIRAN xv
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Tujuan 2
1.4 Manfaat 3
1.5 Batasan Masalah 3
1.6 Sistematika Penulisan 4
BAB 2 LANDASAN KEPUSTAKAAN 5
2.1 Kajian Pustaka 5
2.2 Konsep Big Data 6
2.3 Hadoop 7
2.3.1 HDFS (Hadoop *Distributed File System)* 8
2.3.2 MapReduce 9
2.3.3 YARN *(Yet Another Resource Negotiator)* 10
2.4 Hadoop *Multinode* 11
2.5 Algoritme S*cheduling* 12
2.5.1 Algoritme *Fair Share Scheduling* 12
2.5.2 Algoritme *Capacity Scheduling* 14
2.6 Parameter Pengujian 15
2.6.1 Job Fail Rate 15
2.6.2 Latency 16
2.6.3 Throughput 16
BAB 3 METODOLOGI 17

3.1 Studi Literatur 17
3.2 Lingkungan Penelitian 18
3.3 Implementasi 19
3.4 Pengujian 19
3.5 Hasil dan Analisis 21
3.6 Kesimpulan dan Saran 21
BAB 4 IMPLEMENTASI DAN PENGUJIAN 22
4.1 Implementasi .....**Error! Bookmark not defined.**
4.2 Konfigurasi Jaringan 23
4.3 Instalasi Hadoop 25
4.3.1 Instalasi Hadoop *M*ulti*node Cluster* 26
4.3.2 Konfigurasi Komputer *Master* 29
4.3.3 Konfigurasi Komputer *Slave* 29
4.3.4 Menjalankan Perintah SSH 30
4.3.5 Menjalankan dan Menghentikan Hadoop 32
4.4 Konfigurasi *Fair Share Scheduler* 34
4.4.1 Instalasi *Fair Share Scheduling* 35
4.4.2 Membuat Daftar Antrian 36
4.4.3 Konfigurasi Alokasi *File* 36
4.4.4 Melihat Daftar Antrian 37
4.5 Konfigurasi *Capacity Scheduling* 37
4.5.1 Instalasi *Capacity Scheduling* 38
4.5.2 Membuat Daftar Antrian 39
4.5.3 Konfigurasi Alokasi *File* 39
4.5.4 Konfigurasi Access Control List (ACL) 40
4.5.5 Melihat Daftar Antrian 40
4.6 Konfigurasi *Job Wordcount* 41
4.6.1 Melakukan *Compile* Kode program 41
4.6.2 Membuat Direktori HDFS 42
4.6.3 Menjalankan *Job Wordcount* 43
4.7 Pengujian 45
4.7.1 Alat Pengujian 45
4.7.2 Pengujian Variasi Ukuran Data 46
4.7.2.1 *Fair Share Scheduling* 46

4.7.2.2 *Capacity Scheduling* 48
4.7.3 Pengujian variasi jumlah *job* 50
4.7.3.1 *Fair Share Scheduling* 51
4.7.3.2 *Capacity Scheduling* 53
4.7.4 Pengujian variasi jenis *job* 56
4.7.4.1 *Fair Share Scheduling* 57
4.7.4.2 *Capacity Scheduling* 58
BAB 5 HASIL DAN ANALISIS 61
5.1 Hasil 61
5.1.1 Variasi Ukuran Data 63
5.1.1.1 *Job Fail Rate* 63
5.1.1.2 *Latency* 63
5.1.1.3 *Throughput* 64
5.1.2 Variasi Jumlah *Job* 64
5.1.2.1 *Job Fail Rate* 65
5.1.2.2 *Latency* 65
5.1.2.3 *Throughput* 65
5.1.3 Variasi Jenis *Job* 66
5.1.3.1 *Job Fail Rate* 66
5.1.3.2 *Latency* 67
5.1.3.3 *Throughput* 67
5.2 Analisis 68
5.2.1 Nilai *Job Fail Rate* pada Skenario Variasi Ukuran Data 68
5.2.2 Nilai *Latency* pada Skenario Variasi Ukuran Data 69
5.2.3 Nilai *Throughput* pada Skenario Variasi Ukuran Data 70
5.2.4 Nilai *Job Fail Rate* pada Skenario Variasi Jumlah *Job* 71
5.2.5 *Nilai Latency* pada Skenario Variasi Jumlah *Job* 72
5.2.6 Nilai t*hroughput* pada skenario Variasi Jumlah *Job* 73
5.2.7 Nilai *Job Fail Rate* pada Skenario Variasi Jenis *Job* 74
5.2.8 Nilai *Latency* pada Skenario Variasi Jenis *Job* 75
5.2.9 Nilai *Throughput* pada Skenario Variasi Jenis *Job* 76
BAB 6 PENUTUP 77
6.1 Kesimpulan 77
6.2 Saran 78

DAFTAR PUSTAKA.....79
LAMPIRAN A TABEL ANALISIS 81
LAMPIRAN B KONFIGURASI FILE HADOOP.....90

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Sebelumya 6
Tabel 3.1 Skenario Pengujian 19
Tabel 3.2 Jenis *Job* Hadoop 20
Tabel 3.3 *Dataset* pada HDFS 20
Tabel 4.1 Daftar Alamat IP pada Kompter *Node* 23
Tabel 5.1 Data Hasil Pengujian untuk Pengukuran Latency 62
Tabel 5.2 Hasil Nilai *Job Fail Rate* pada Skenario Variasi Ukuran Data 63
Tabel 5.3 Hasil Nilai *Latency* pada Skenario Variasi Ukuran Data 63
Tabel 5.4 Hasil Nilai *Throughput* pada Skenario Variasi Ukuran Data 64
Tabel 5.5 Hasil Nilai *Job Fail Rate* pada Variasi Jumlah *Job* 65
Tabel 5.6 Hasil Nilai Latency pada Skenario Variasi Jumlah *Job* 65
Tabel 5.7 Hasil Nilai *Throughput* pada Skenario Variasi Jumlah *Job* 66
Tabel 5.8 Hasil Nilai *Job Fail Rate* pada Skenario Variasi Jenis *Job* 66
Tabel 5.9 Hasil Nilai *Latency* pada Variasi Jenis *Job* 67
Tabel 5.10 Hasil Nilai *Throughput* pada Variasi Jenis *Job* 67

DAFTAR GAMBAR

Gambar 2. 1 Model 3V pada Big Data 7
Gambar 2.2 Arsitektur HDFS 8
Gambar 2.3 Cara Kerja MapReduce pada Pengiriman *Job* 9
Gambar 2.4 Cara Kerja pemrograman pada MapReduce 10
Gambar 2.5 Cara Kerja YARN pada Sebuah Aplikasi 11
Gambar 2.6 Topologi Hadoop *Multinode* 12
Gambar 2.7 Algoritme *Job Fair Share Scheduling* 13
Gambar 2.8 Algoritme *Job Capacity Scheduling* 14
Gambar 3.1 Diagram Blok Metodologi 17
Gambar 3.2 Arsitektur Topologi Jaringan 18
Gambar 4.1 Diagram Alir Tahap Implementasi 22
Gambar 4.2 Diagram AlirKonfigurasi Jaringan 23
Gambar 4.3 Konfigurasi Jaringan 24
Gambar 4.4 Konfigurasi *Host* 24
Gambar 4.5 Konfigurasi *Network Interfaces* eth1 pada *Master* 24
Gambar 4.6 Cek Koneksi *Master* ke Komputer *Slave* 25
Gambar 4.7 Diagram Alir Proses Instalasi Hadoop 26
Gambar 4.8 C*loning Master* Hadoop 28
Gambar 4.9 Jalankan Semua *Node* 28
Gambar 4.10 Inisialisasi *File Masters* 29
Gambar 4.11 Inisialisasi *File Slave* 30
Gambar 4.12 Melakukan *Call* SSH pada Komputer *Master* 31
Gambar 4.13 Format Name*node* 32
Gambar 4.14 Perintah untuk Menjalankan Hadoop 32
Gambar 4.15 Cek JPS pada Komputer *Master* 32
Gambar 4.16 Cek JPS pada Komputer *Slaves* 33
Gambar 4.17 Antarmuka Hadoop pada Resource Manager 33
Gambar 4.18 Antarmuka Hadoop pada Direktori HDFS 34
Gambar 4.19 Menghentikan Proses Hadoop 34
Gambar 4.20 Diagram Alir Konfigurasi *Fair Share Scheduling* 35
Gambar 4.21 Daftar Antrian *Fair Share Scheduling* pada Resource Manager 37

Gambar 4.22 Diagram Alir Konfigurasi *Capacity Scheduling* 38
Gambar 4.23 Daftar Antrian *Capacity Scheduling* pada Resource Manager.....41
Gambar 4.24 Kode Program *Wordcount* 42
Gambar 4.25 Hasil *Compile* Program *Wordcount* 42
Gambar 4.26 Menyalin *File* pada Direktori HDFS 43
Gambar 4.27 Tampilan Direktori HDFS.....43
Gambar 4.28 Perintah Menjalankan *Job Wordcount* 43
Gambar 4.29 Tampilan *Resourcemanager* untuk Monitoring *Job Wordcount* 44
Gambar 4.30 Antarmuka Direktori *Output* pada *Job Wordcount*.....44
Gambar 4.31 Hasil Keluaran *Job Wordcount* pada *File* adult.csv 45
Gambar 4.32 Monitoring *Job* Variasi Ukuran Data pada *Fair Share Scheduling*...48
Gambar 4.33 Monitoring *Job* Variasi Ukuran Data pada *Capacity Scheduling* 50
Gambar 4.34 Monitoring *Job* Variasi Jumlah *Job* pada *Fair Share Scheduling* 53
Gambar 4.35 Monitoring *Job* Variasi Jumlah *Job* pada *Capacity Scheduling*.....56
Gambar 4.36 Monitoring *Job* Variasi Jenis *Job* pada *Fair Share Scheduling* 58
Gambar 4.37 Monitoring *Job* Variasi Jenis *Job* pada *Fair Share Scheduling* 60
Gambar 5.1 Pengukuran *Latency* 62
Gambar 5.2 Grafik Perbandingan *Job Fail Rate* pada Skenario Variasi Ukuran Data 68
Gambar 5.3 Grafik Perbandingan *Latency* pada Skenario Variasi Ukuran Data 69
Gambar 5.3Grafik Perbandingan*Throughput*pada Skenario Variasi Ukuran Data 70
Gambar 5.5 Grafik Perbandingan *Job Fail Rate* pada Variasi Jumlah *Job*.....71
Gambar 5.6 Grafik Perbandingan *Latency* pada Variasi Jumlah *Job*.....72
Gambar 5.7 Grafik Perbandingan *Throughput* pada Variasi Jumlah *Job* 73
Gambar 5.8 Grafik Perbandingan *Job Fail Rate* pada Variasi Jumlah *Job*.....74
Gambar 5.9 Grafik Perbandingan *Latency* pada Variasi Jenis *Job* 75
Gambar 5.10 Grafik Perbandingan *Throughput* pada Variasi Jenis *Job*.....76

DAFTAR LAMPIRAN

LAMPIRAN A TABEL ANALISIS 81
A.1 Tabel Hasil pengujian Variasi Ukuran Data 81
A.2.1 *Fair Share Scheduling* 81
A.1.2 *Capacity Scheduling*.....82
A.2 Tabel Hasil Pengujian Variasi Jumlah *Job* 84
A.2.1 *Fair Share Scheduling* 84
A.2.2 *Capacity Scheduling*.....86
A.3 Tabel hasil pengujian variasi jenis *job* 88
A.3.1 *Fair Share Scheduling* 88
A.3.2 *Capacity Scheduling*.....89
LAMPIRAN B KONFIGURASI FILE HADOOP.....90