

DAFTAR ISI

HALAMAN PENGESAHAN	ii
PERNYATAAN BEBAS PLAGIASI	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiv
DAFTAR KODE	xvi
DAFTAR NOTASI	xvii
Intisari	xviii
<i>Abstract</i>	xix
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
II TINJAUAN PUSTAKA	5
III DASAR TEORI	7
3.1 Apache Hadoop	7
3.2 YARN	7
3.3 Container	11

3.4	MapReduce	12
3.5	Wordcount	13
3.6	Jobhistory dan Rumen	13
3.7	HiBench	14
IV	RANCANGAN PENELITIAN	15
4.1	Deskripsi Penelitian	15
4.2	Konfigurasi awal	16
4.2.1	HDFS Replication	17
4.2.2	Scheduler	17
4.2.3	Reduce Slowstart	17
4.2.4	Maximum CPU Resource Allocation	18
4.2.5	Minimum Memory Resource Allocation	18
4.2.6	Maximum Memory Resource Allocation	18
4.3	Data pengujian	18
4.4	Struktur Pengujian	19
4.4.1	Skenario Pertama	21
4.4.2	Skenario Kedua	22
4.5	Pengambilan Data Performa	23
4.5.1	JobHistory dan Rumen	24
4.5.2	Script Termodifikasi dari HiBench	24
4.6	Rancangan Analisis Hasil Pengujian	26
V	IMPLEMENTASI	27
5.1	Alat dan Bahan	27
5.2	Pengaturan Hadoop	28
5.2.1	Penyeragaman Waktu Kluster	28
5.2.2	Mematikan Firewall	28
5.2.3	Memulai Hadoop	29
5.2.4	Menghentikan Hadoop	29
5.3	Konfigurasi Variabel Tetap	30
5.3.1	hdfs-site.xml	30
5.3.2	mapred-site.xml	31
5.3.3	yarn-site.xml	31
5.4	Pembangkitan Data	33
5.4.1	Kamus kata	33

5.4.2	Kode MapReduce Pembangkitan Data Random Word	34
5.4.3	Eksekusi Pembangkitan Data	35
5.5	Wordcount	36
5.6	Pengambilan Data Performa	37
5.6.1	Modifikasi HiBench	37
5.6.2	Rumen	39
5.7	Template Konfigurasi Variabel Percobaan	39
5.7.1	mapred-site.xml	39
5.7.2	yarn-site.xml	41
5.8	Duplikasi Konfigurasi ke Semua Node	41
5.9	Script Eksekusi Keseluruhan Konfigurasi dan Job Wordcount	42
5.10	Implementasi Skenario Pertama	43
5.11	Implementasi Skenario Kedua	44
5.12	Pengolahan Data Hasil Pengujian	45
5.12.1	Import Data JSON Rumen	45
5.12.2	Import Data CSV Monitor HiBench Modifikasi	45
5.12.3	Waktu $Job (T_{Job})$	46
5.12.4	Plotting Penggunaan CPU	46
5.12.5	Plotting Penggunaan Memory	47
5.12.6	Rata-rata CPU_{Use}	48
5.12.7	Rata-rata Mem_{Use}	48
5.12.8	Rata-rata N_{Proc}	48
5.12.9	Waktu Keseluruhan Map (T_{MapAll})	48
5.12.10	Plotting Jumlah Container yang Berjalan Selama $MapAll$	49
5.12.11	Waktu Keseluruhan Reduce (T_{RedAll})	50
5.12.12	Gantt Chart Timeline Eksekusi MapReduce	51
5.12.13	Waktu Ketidakseimbangan Kerja (T_{Unb})	52

VI HASIL DAN PEMBAHASAN 54

6.1	Analisis Waktu, Penggunaan Resource CPU, dan Ketidakseimbangan Kerja	54
6.1.1	Analisis Skenario Pertama	54
6.1.2	Analisis Skenario Kedua	82

6.2	Analisis Batas Maksimum CPN Paralel, Akumulasi $T_{MapTask}$, Akumulasi $T_{RedTask}$, Penggunaan Resource Memori, dan Jumlah <i>Process</i> CPU	103
6.2.1	Analisis Skenario Pertama	103
6.2.2	Analisis Skenario Kedua	110
VII KESIMPULAN DAN SARAN		117
7.1	Kesimpulan	117
7.2	Saran	119
DAFTAR PUSTAKA		121