

DAFTAR ISI

DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	x
INTISARI.....	xi
ABSTRACT.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI.....	31
3.1 Hadoop	31
3.2 <i>MapReduce Framework</i>	31
3.2.1 MapReduce	31
3.2.2 Model Pemrograman.....	32
3.2.3 Alur Eksekusi	33
3.2.4 Struktur Data Master	35
3.2.5 Toleransi Terhadap Kesalahan.....	36
3.2.6 Lokalitas Pemrosesan.....	37
3.3 HADOOP dan <i>MapReduce</i>	37
3.4 Hadoop <i>Cluster</i>	40
3.5 Pengujian TestDFSIO pada HADOOP	42
3.6 <i>Multicore System</i>	43
3.7 Pengindeksan Dokumen	44
3.8 Pembobotan Dokumen	47
BAB IV ANALISIS DAN PERANCANGAN	49

4.1.	Deskripsi sistem	49
4.2.	Analisis Kebutuhan	49
4.2.1	Kebutuhan fungsional sistem	49
4.2.2	Kebutuhan non fungsional sistem	50
4.3.	Kebutuhan data	50
4.4.	Arsitektur Sistem	50
4.5.	Rancangan Sistem	51
4.6.	Rancangan TF-IDF – <i>Weighting</i> – <i>Map Reduce</i> Job Data Flow	52
4.7.	Perancangan Perangkat Lunak pada Tfidf – <i>Weighting</i>	54
BAB V IMPLEMENTASI		56
5.1.	Konfigurasi <i>MapReduce</i>	56
5.2.	Penghitungan Tfidf	57
5.3.	Dataset	57
5.4.	Implementasi Pembobotan Dokumen Menggunakan Metode tfidf	58
BAB VI HASIL DAN PEMBAHASAN		72
6.1.	Hasil Pengujian	72
6.2.	Pengujian Tanpa Pembatasan Penggunaan <i>Core</i>	72
6.2.1	Pengujian Dengan Pembatasan Penggunaan <i>Core</i>	74
6.2.2	Pengujian Menggunakan Resource Maksimal	74
6.2.3	Pengujian Menggunakan Jumlah Core 2 dan 4.	75
6.3.	Analisis Hasil Pengujian	77
BAB VII KESIMPULAN DAN SARAN		82
7.1.	Kesimpulan	82
7.2.	Saran	82
DAFTAR PUSTAKA		83

DAFTAR GAMBAR

Gambar 3.1 Alur eksekusi (Dean dan Ghemawat, 2004).....	34
Gambar 3.2 Aliran data <i>MapReduce</i> pada Hadoop (Lam, 2011).....	38
Gambar 3.3 Hadoop <i>Cluster</i>	40
Gambar 3.4. Skema Prosesor <i>Multicore</i> (Ogi, 2010).....	43
Gambar 4.1 Arsitektur Sistem.....	51
Gambar 4.2 Skema Perangkat Lunak.....	51
Gambar 4.3. <i>Map Reduce</i> Job Data Flow pada pembobotan dokumen menggunakan Tfidf	52
Gambar 4.4. <i>Map Reduce</i> Job Data Flow pada Job WordCount	53
Gambar 4.5. Map Reduce Job Data Flow pada Job WordCount Per Doc	53
Gambar 4.6. <i>Map Reduce</i> Job Data Flow pada Job docCount per Word.....	54
Gambar 4.7 Flow chart menghitung frekuensi kemunculan suatu <i>term</i> per dokumen.....	54
Gambar 4.8 Flow chart menghitung frekuensi kemunculan <i>term</i> relatif	55
Gambar 4.9 Flow chart menghitung tfidf.....	56
Gambar 5.1 Susunan Direktori Hadoop pada Ubuntu	56
Gambar 5.2 Seting Java_Home dan Hadoop_Prefix.....	56
Gambar 5.3 Potongan Kode Program WordFrequencyInDoc.java	59
Gambar 5.4 Potongan source code WordFrequencyInDocMapper.java	61
Gambar 5.5 Potongan source code WordFrequencyInDocReducer.java	63
Gambar 5.6 <i>Output</i> hasil WordFrequencyInDocReducer	64
Gambar 5.7 Potongan source code WordCountsForDocsMapper.java.....	65
Gambar 5.8 Potongan source code WordCountsForDocsReducer.java.....	66
Gambar 5.9 <i>Output</i> hasil WordCountsForDocsReducer	67
Gambar 5.10 Potongan program WordsInCorpusTFIDFMapper.java	68
Gambar 5.11 Potongan program WordsInCorpusTFIDFReducer.java.....	70
Gambar 5.12 Contoh <i>output</i> hasil WordsInCorpusTFIDFReducer	70
Gambar 6.1 Rata-rata waktu komputasi tanpa pembatasan penggunaan core.....	73
Gambar 6.2. Konfigurasi <i>file mapred-site.xml</i> untuk <i>core</i> 8	74
Gambar 6.3. Konfigurasi <i>file yarn-site.xml</i> untuk <i>core</i> 8	74
Gambar 6.4. Konfigurasi <i>file mapred-site.xml</i> untuk <i>core</i> 4.....	75
Gambar 6.5. Konfigurasi <i>file yarn-site.xml</i> untuk jumlah <i>core</i> 4	75
Gambar 6.6. Konfigurasi <i>file mapred-site.xml</i> untuk <i>core</i> 2.....	76
Gambar 6.7. Konfigurasi <i>file yarn-site.xml</i> untuk jumlah <i>core</i> 2	76
Gambar 6.8 <i>Cluster</i> Hadoop pada <i>Core</i> 8.....	77
Gambar 6.9 <i>Output terminal</i> untuk pengujian <i>core</i> 8.....	77



Gambar 6.10 <i>Output terminal</i> untuk pengujian core 2 dan 4	78
Gambar 6.11. <i>Cluster</i> hadoop untuk core 2	78
Gambar 6.12. <i>Cluster</i> hadoop untuk core 4	78
Gambar 6.13 Grafik pengujian dengan pembatasan penggunaan core.	79

DAFTAR TABEL

Tabel 2.1 Perbandingan dengan penelitian terdahulu	9
Tabel 3.1 Contoh penggunaan <i>MapReduce</i>	33
Tabel 5.1 Ukuran <i>Corpus</i> dan Komposisi dataset.....	58
Tabel 6.1 Pengujian tanpa pembatasan penggunaan core	72
Tabel 6.2 Rata-rata hasil pengujian tanpa pembatasan penggunaan <i>core</i>	73
Tabel 6.2 Hasil pengujian dengan pembatasan penggunaan core.....	79