

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Persembahan	v
Halaman Motto	vi
PRAKATA	vii
INTISARI	xvii
ABSTRACT	xviii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Metode Penelitian	5
1.7 Sistematika Penulisan	6
II TINJAUAN PUSTAKA	7
III DASAR TEORI	11
3.1 Data Warehouse	11
3.1.1 Karakteristik Data Warehouse	11
3.1.2 Arsitektur Data Warehouse	12
3.1.3 ETL (<i>Extract, Transform, Load</i>)	14
3.1.4 Skema Data Warehouse	14
3.2 Analisis Sosial Media	16
3.3 Twitter	16

3.4	Basis Data Relasional	17
3.4.1	MySQL	17
3.4.2	PostgreSQL	17
3.5	Apache Cassandra	18
3.5.1	Model Tabel Cassandra	20
3.5.2	Model <i>Query</i> Cassandra	21
3.5.3	Kunci Pemodelan Data pada Cassandra	22
3.6	Uji Median Mood	23
3.7	Uji Chi-Square	24
IV ANALISIS DAN PERANCANGAN PENELITIAN		25
4.1	Deskripsi Umum	25
4.2	Analisis Kebutuhan Sistem	25
4.3	Rancangan Penelitian	26
4.3.1	Streaming Twitter	27
4.3.2	Penyimpanan Sampel Arus Data	27
4.3.3	Proses ETL	28
4.3.4	Proses Percobaan (ETL dan <i>querying</i>) dan Pengujian	28
4.3.5	Visualisasi	28
4.4	Rancangan Pengujian	29
4.5	Perancangan Sistem	32
4.5.1	Rancangan Arsitektur Data Warehouse	32
4.5.2	Analisis Struktur Data Twitter	33
4.5.3	Rancangan ERD	35
4.5.4	Rancangan Skema Data Warehouse	36
V IMPLEMENTASI SISTEM		43
5.1	Spesifikasi Sistem	43
5.2	Implementasi Pengambilan Sampel Arus Data	44
5.3	Implementasi Proses ETL	45
5.3.1	Implementasi Proses ETL Menggunakan Pentaho Data Integration	46
5.3.2	Implementasi Proses ETL Menggunakan Pustaka Twitter4J	48
5.4	Implementasi Sistem Percobaan	57
5.4.1	Percobaan penyimpanan data	57
5.4.2	Percobaan pemanggilan data	58

5.5	Implementasi Visualisasi Data	60
VI	PEMBAHASAN	62
6.1	Hasil Pengambilan Sampel Arus Data	62
6.2	Hasil Percobaan	63
6.2.1	Hasil percobaan penyimpanan data	63
6.2.2	Hasil percobaan pemanggilan data	70
6.3	Hasil Visualisasi Data	73
6.4	Analisis Data percobaan	77
6.4.1	Hasil uji median Mood pada data percobaan penyimpanan data	78
6.4.2	Hasil uji median Mood pada data percobaan pemanggilan data	84
VII	KESIMPULAN DAN SARAN	86
7.1	Kesimpulan	86
7.2	Saran	86
A	DATA HASIL PERCOBAAN	93
1.1	Hasil Percobaan Penyimpanan Data	93
1.2	Hasil Percobaan Pemanggilan Data	96

DAFTAR TABEL

2.1	Tabel Perbandingan Penelitian	10
4.1	Rencana jadwal pengambilan data	27
4.2	Tabel <i>Query</i> untuk Visualisasi Data	29
4.3	Rancangan skenario pengujian	31
4.4	Deskripsi nilai pengukuran tabel fakta	37
5.1	Spesifikasi perangkat lunak	43
5.2	Spesifikasi perangkat keras	44
6.1	Jadwal perolehan sampel arus data	62
6.2	Skenario pengujian	63
6.3	Tabel hasil uji median Mood penyimpanan data pada MySQL	79
6.4	Tabel hasil uji median Mood penyimpanan data pada PostgreSQL	80
6.5	Tabel hasil uji median Mood penyimpanan data antara MySQL dan PostgreSQL dengan skenario 1 dan 3	81
6.6	Tabel hasil uji median Mood penyimpanan data pada Cassandra	82
6.7	Tabel hasil uji median Mood penyimpanan data antara Cassandra dan PostgreSQL	83
6.8	Tabel hasil uji median Mood pemanggilan data	84
1.1	Hasil percobaan penyimpanan data pada MySQL dengan perlakuan skenario 3 (detik)	93
1.2	Hasil percobaan penyimpanan data pada MySQL dengan perlakuan skenario 4 (detik)	93
1.3	Hasil percobaan penyimpanan data pada MySQL dengan perlakuan skenario 5 (detik)	94
1.4	Hasil percobaan penyimpanan data pada PostgreSQL dengan perlakuan skenario 3 (detik)	94
1.5	Hasil percobaan penyimpanan data pada PostgreSQL dengan perlakuan skenario 4 (detik)	94
1.6	Hasil percobaan penyimpanan data pada PostgreSQL dengan perlakuan skenario 5 (detik)	95

1.7 Hasil percobaan penyimpanan data pada Cassandra dengan perlakuan skenario 6 (detik)	95
1.8 Hasil percobaan penyimpanan data pada Cassandra dengan perlakuan skenario 7 (detik)	95
1.9 Hasil percobaan pemanggilan data pada MySQL (detik)	96
1.10 Hasil percobaan pemanggilan data pada PostgreSQL (detik)	96
1.11 Hasil percobaan pemanggilan data pada Cassandra (detik)	96

DAFTAR GAMBAR

3.1	Arsitektur data warehouse (Laudon dan Laudon, 2007)	13
3.2	Skema bintang (Serhal, 2009)	15
3.3	Skema butir salju (Serhal, 2009)	15
3.4	Struktur tabel dengan kunci partisi tunggal (Johnson, 2013)	20
3.5	Struktur tabel dengan kunci partisi ganda (Johnson, 2013)	21
3.6	Struktur tabel dua arah untuk uji median Mood (Utss dan Heckard, 2011)	23
4.1	Alur penelitian	26
4.2	Arsitektur data warehouse	33
4.3	Contoh struktur data Twitter	34
4.4	Diagram relasi entitas (ERD)	35
4.5	Skema data warehouse pada basis data relasional	36
4.6	Skema data warehouse pada basis data Cassandra	39
4.7	Struktur data mart pada data warehouse Cassandra	39
4.8	Diagram alur kerja <i>query</i> antar tabel pada data warehouse Cassandra .	40
4.9	Skema data warehouse pada basis data relasional tanpa dimensi entitas untuk percobaan penyimpanan data skenario 4	40
4.10	Skema data warehouse <i>flat</i> pada basis data relasional tanpa dimensi entitas untuk percobaan penyimpanan data skenario 3	41
4.11	Skema data warehouse pada basis data Cassandra tanpa dimensi entitas untuk percobaan penyimpanan data skenario 6	42
5.1	Code program pengambilan dan penyimpanan arus data Twitter	45
5.2	Tampilan file transformasi untuk basis data SQL pada PDI <i>spoon</i>	47
5.3	Tampilan file transformasi untuk basis data Cassandra pada PDI <i>spoon</i>	47
5.4	Code program implementasi file transformasi dari PDI	48
5.5	Cuplikan kode program implementasi percobaan proses ETL menggunakan pustaka Twitter4j untuk basis data SQL	49
5.6	Cuplikan kode implementasi sistem penyimpanan langsung arus data Twitter pada metode <i>onStatus</i>	50
5.7	Cuplikan kode implementasi sistem penyimpanan langsung arus data Twitter pada metode <i>onStatus</i>	51

5.8	Cuplikan kode program implementasi pembuatan tabel dimensi dan tabel fakta pada basis data SQL	52
5.9	Cuplikan kode program implementasi penyimpanan tabel dimensi . .	52
5.10	Cuplikan kode program implementasi penyimpanan pada Cassandra .	53
5.11	Cuplikan kode program implementasi penyimpanan tabel pada Cassandra	55
5.12	Cuplikan kode program implementasi penyimpanan data entitas . . .	56
5.13	Tampilan menu program	57
5.14	Tampilan program Java ketika hasil percobaan didapatkan	58
5.15	Tampilan program klien mysql pada saat menampilkan waktu proses <i>query</i>	59
5.16	Tampilan program klien psql pada saat menampilkan waktu proses <i>query</i>	59
5.17	Tampilan program klien cqlsh pada saat menampilkan waktu proses <i>query</i>	59
5.18	Cuplikan kode CQL pada visualisasi data untuk Q8	60
5.19	Cuplikan kode SQL pada visualisasi data untuk Q10	61
6.1	Grafik Perbandingan Hasil Percobaan Skenario 1, 2, dan 3 pada MySQL untuk data 10-100	64
6.2	Grafik Perbandingan Hasil Percobaan Skenario 1, 2, dan 3 pada MySQL untuk data 100-500	64
6.3	Grafik Perbandingan Hasil Percobaan Skenario 1, 2, dan 3 pada PostgreSQL untuk data 10-100	65
6.4	Grafik Perbandingan Hasil Percobaan Skenario 1, 2, dan 3 pada PostgreSQL untuk data 100-500	65
6.5	Grafik Perbandingan Hasil Percobaan Skenario 4 dan 5 untuk data 10-100	66
6.6	Grafik Perbandingan Hasil Percobaan Skenario 4 dan 5 untuk data 100-500	66
6.7	Grafik Perbandingan Hasil Percobaan Skenario 1 dan 4 untuk data 10-100	67
6.8	Grafik Perbandingan Hasil Percobaan Skenario 1 dan 4 untuk data 100-500	67

6.9	Grafik Perbandingan Hasil Percobaan Skenario 3 dan 5 untuk data 10-100	68
6.10	Grafik Perbandingan Hasil Percobaan Skenario 3 dan 5 untuk data 100-500	68
6.11	Grafik Perbandingan Hasil Percobaan Skenario 1 dan 5 untuk data 10-100	69
6.12	Grafik Perbandingan Hasil Percobaan Skenario 1 dan 5 untuk data 100-500	69
6.13	Grafik Kecepatan Pemanggilan Data pada MySQL	70
6.14	Grafik Kecepatan Pemanggilan Data pada PostgreSQL	71
6.15	Grafik Kecepatan Pemanggilan Data pada Cassandra	72
6.16	Grafik Perbandingan Kecepatan Pemanggilan Data	72
6.17	Tampilan menu utama <i>dashboard</i>	73
6.18	Tampilan halaman utama basis data dan visualisasi dari <i>query</i> Q1 pada Cassandra	74
6.19	Tampilan halaman visualisasi dari <i>query</i> Q2 pada MySQL	74
6.21	Tampilan halaman visualisasi dari <i>query</i> Q9 dan Q10 pada Cassandra	75
6.20	Tampilan halaman visualisasi dari <i>query</i> Q8 pada PostgreSQL	75
6.22	Tampilan cuplikan hasil uji median Mood pada MySQL dengan skenario 2 dan 3 pada jumlah data <i>tweet</i> 10	78