

**ANALISIS FAKTOR OPTIMASI UNTUK DATA
WAREHOUSE DENGAN DATA TABUNGAN PADA BANK
XYZ**

Skripsi



oleh

ALOYSIUS ADHYATMA HERFANGSYAH

71150012

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

**ANALISIS FAKTOR OPTIMASI UNTUK DATA
WAREHOUSE DENGAN DATA TABUNGAN PADA BANK
XYZ**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

**ALOYSIUS ADHYATMA HERFANGSYAH
71150012**

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

ANALISIS FAKTOR OPTIMASI UNTUK DATA WAREHOUSE DENGAN DATA TABUNGAN PADA BANK XYZ

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 22 Oktober 2019



ALOYSIUS ADHYATMA
HERFANGSYAH
71150012

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Judul Skripsi : ANALISIS FAKTOR OPTIMASI UNTUK DATA
WAREHOUSE DENGAN DATA TABUNGAN
PADA BANK XYZ
Nama Mahasiswa : ALOYSIUS ADHYATMA HERFANGSYAH
NIM : 71150012
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2019/2020

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 16 Oktober 2019

Dosen Pembimbing I


Willy Sudiarto Raharjo, S.Kom., M.Cs.

Dosen Pembimbing II


Antonius Rachmat C., S.Kom., M.Cs.

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

ANALISIS FAKTOR OPTIMASI UNTUK DATA WAREHOUSE DENGAN DATA TABUNGAN PADA BANK XYZ

Oleh: ALOYSIUS ADHYATMA HERFANGSYAH / 71150012

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 11 Oktober 2019

Yogyakarta, 21 Oktober 2019
Mengesahkan,

Dewan Penguji:

1. Willy Sudiarto Raharjo, S.Kom., M.Cs.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Aditya Wikan Mahastama, S.Kom., M.Cs.
4. Danny Sebastian, S.Kom., M.M., M.T.




Dekan
(Restyandito, S.Kom., MSIS., Ph.D.)


Ketua Program Studi
(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dalam penyusunan skripsi ini tidak terlepas dari dukungan berbagai pihak. Penulis secara khusus mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua serta kakak-kakak tercinta yang selama ini telah membantu penulis dalam bentuk perhatian, kasih sayang, semangat, serta doa demi kelancaran dan kesuksesan penulis dalam menjalani pendidikan.
2. Bapak Willy Sudiarto Raharjo, S.Kom., M.Cs., selaku dosen pembimbing I dan dosen wali yang selalu memberikan bimbingan, arahan, dorongan, dan semangat kepada penulis selama masa perkuliahan serta saat pengerjaan tugas akhir, sehingga tugas akhir ini dapat selesai.
3. Bapak Antonius Rachmat C., S.Kom., M.Cs., selaku dosen pembimbing II yang selalu memberikan dukungan, semangat, dan timbal balik baik pada saat seminar maupun selama pengerjaan tugas akhir, sehingga tugas akhir ini dapat selesai.
4. Seluruh Staff Divisi Data Warehouse Bank BCA, khususnya Yashinta Manus selaku mentor yang selalu memberikan ilmu, arahan dan semangat pada saat penulis melakukan magang.
5. Teman-teman GTS yang menemani penulis untuk berkeluh kesah dan menjadi tempat bercerita, sehingga penulis dapat menyelesaikan tugas akhir ini.

Yogyakarta, 21 Oktober 2019

Aloysius Adhyatma Herfangsyah

KATA PENGANTAR

Penulis mengucapkan banyak terima kasih kepada pihak – pihak yang terlibat dalam pembuatan “Analisis Faktor Optimasi untuk Data Warehouse dengan Data Tabungan pada Bank XYZ”, penulisan tugas akhir ini tidak lepas dari pihak – pihak yang telah mendukung penulis dalam menyelesaikan tugas akhir ini.

Tugas Akhir merupakan salah satu syarat wajib dalam memperoleh gelar sarjana di Universitas Kristen Duta Wacana, dan tidak lupa penulis mengucapkan banyak terima kasih kepada seluruh pihak dari Universitas Kristen Duta Wacana yang telah membantu penulis menyelesaikan tugas akhir ini.

Penulis mengharapkan kritik dan saran yang membangun dari pembaca apabila terdapat kekliruan dalam penulisan laporan ini. Sekian dan terima kasih.

Yogyakarta, 21 Oktober 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	vii
INTISARI	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Metode Penelitian	3
a. Wawancara.....	3
b. Analisis Desain Skema <i>Data Warehouse</i>	3
c. Implementasi Desain <i>Data Warehouse</i>	3
d. Pembangunan Data	4
e. Desain Struktur ETL	4
f. Implementasi Struktur ETL	4
g. Pengujian Performa Tabel.....	4
1.7 Sistematika Penulisan	4
a. Bab I Pendahuluan	5
b. Bab II Landasan Teori.....	5
c. Bab III Metodologi Penelitian.....	5
d. Bab IV Hasil dan Pembahasan.....	5
e. Bab V Kesimpulan dan Saran	5
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Pustaka	6

2.2 Landasan Teori.....	7
2.2.1 Basis Data	7
2.2.2 Optimasi Basis Data.....	8
2.2.3 Data Warehouse	8
2.2.4 Extract, Transform, Load	11
2.2.5 Big Data	13
2.2.6 Apache Hadoop.....	13
2.2.7 Hadoop File System	15
2.2.8 Apache Hive.....	16
2.2.9 Kompresi Data	17
2.2.10 Apache Parquet	18
BAB III PERANCANGAN SISTEM	20
3.1 Spesifikasi Perangkat	20
3.1.1 Spesifikasi Perangkat Lunak.....	20
3.1.2 Spesifikasi Perangkat Keras.....	20
3.2 Rancangan Database	20
3.2.1 Skema Data Warehouse	21
3.3 Langkah Penelitian.....	24
3.3.1 Kerangka Penelitian.....	24
3.3.2 Flowchart Penelitian	29
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Persiapan Implementasi	30
4.1.1 Script Query Create Table.....	30
4.1.2 Script Query Analisis Optimasi Data Warehouse	30
4.2 Implementasi Optimasi Data Warehouse.....	32
4.2.1 List Tabel Implementasi.....	32
4.2.2 Pembuatan Mapping Tiap Tabel	32
4.2.3 Hasil Partisi dan Kompresi Tabel	49
4.3 Pengujian Performa Tabel.....	52
4.3.1 Performa Waktu Query Skenario 1	52
4.3.2 Performa Waktu Query Skenario 2	54
4.3.3 Performa Waktu Query Skenario 3	55
4.3.4 Performa Waktu Query Skenario 4.....	57
4.3.5 Performa Waktu Query Skenario 5.....	58

4.3.6 Performa Waktu Query Skenario 6	60
4.3.7 Performa Waktu Query Skenario 7	62
4.3.8 Performa Waktu Query Skenario 8	63
4.3.9 Performa Ruang Penyimpanan Data	65
4.4 Grafik Perbandingan Hasil Query	67
4.4.1 Rangkuman Hasil Optimasi	67
BAB V KESIMPULAN DAN SARAN.....	72
5.1 Kesimpulan	72
5.2 Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN.....	75
Lampiran 1	75
Lampiran 2	81
Form Persetujuan Revisi	90
Kartu Konsultasi	91

DAFTAR TABEL

Tabel 2.1 Perbedaan database operasional dan data warehouse	11
Tabel 3.1 List Tabel Penelitian	21
Tabel 3.2 Perlakuan Optimasi Tabel Dimensi	28
Tabel 3.3 Perlakuan Optimasi Tabel Fakta.....	28
Tabel 4.1 Tabel Pada Hive.....	30
Tabel 4.2 Hasil query 1	68
Tabel 4.3 Hasil query 2.....	68
Tabel 4.4 Hasil query 3.....	68
Tabel 4.5 Ukuran penyimpanan tiap tabel	68

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Data Warehouse	9
Gambar 2.2 Proses ETL	12
Gambar 2.3 Ekosistem Hadoop	14
Gambar 2.4 Arsitektur HDFS	15
Gambar 2.5 Perbandingan Snappy & Zlib dalam hal ukuran penyimpanan.....	18
Gambar 2.6 Perbandingan Snappy & Zlib dalam hal waktu eksekusi.....	18
Gambar 2.7 Contoh data berbentuk tabel.....	19
Gambar 2.8 Contoh data berbentuk Parquet	19
Gambar 3.1 Skema Tabel Fakta Saving.....	22
Gambar 3.2 Skema Tabel Fakta Checking.....	22
Gambar 3.3 Skema Tabel Fakta Deposit	23
Gambar 3.4 Skema Tabel Fakta Summary	23
Gambar 3.5 Mapping Tahap 1 Tabel Dimensi.....	25
Gambar 3.6 Mapping Tahap 2 Tabel Dimensi.....	25
Gambar 3.7 Mapping Tahap 3 Tabel Dimensi.....	26
Gambar 3.8 Mapping Final Tabel Dimensi	26
Gambar 3.9 Mapping Tahap 1 Tabel Fakta	27
Gambar 3.10 Mapping Tahap 2 Tabel Fakta	27
Gambar 3.11 Mapping Final Tabel Fakta	27
Gambar 3.12 Flowchart Penelitian.....	29
Gambar 4.1 Mapping Temporary Customer	33
Gambar 4.2 Mapping Temporary Customer 2	34
Gambar 4.3 Mapping Temporary Customer 3	34
Gambar 4.4 Mapping Temporary Customer 4	35
Gambar 4.5 Mapping Temporary Customer 5	35
Gambar 4.6 Mapping Temporary Customer 6	36
Gambar 4.7 Mapping Temporary Customer 7	36
Gambar 4.8 Mapping Temporary Customer 8	37
Gambar 4.9 Mapping Temporary Customer 9	37
Gambar 4.10 Mapping Temporary Customer 10	38
Gambar 4.11 Mapping Temporary Account 1	38
Gambar 4.12 Mapping Temporary Account 2	39

Gambar 4.13 Mapping Temporary Account 3	39
Gambar 4.14 Mapping Final Tabel Account 1.....	40
Gambar 4.15 Mapping Temporary Checking 1	41
Gambar 4.16 Mapping Temporary Checking 2	41
Gambar 4.17 Mapping Temporary Checking 3	42
Gambar 4.18 Mapping Final Checking 1.....	42
Gambar 4.19 Mapping Temporary Saving 1	43
Gambar 4.20 Mapping Temporary Saving 2	44
Gambar 4.21 Mapping Temporary Saving 3	44
Gambar 4.22 Mapping Final Saving 1	45
Gambar 4.23 Mapping Temporary Deposit 1	45
Gambar 4.24 Mapping Temporary Deposit 2	46
Gambar 4.25 Mapping Temporary Deposit 3	46
Gambar 4.26 Mapping Final Deposit.....	47
Gambar 4.27 Mapping Temporary Summary 1.....	48
Gambar 4.28 Mapping Temporary Summary 2	48
Gambar 4.29 Mapping Final Summary.....	49
Gambar 4.30 Partisi Tabel Account.....	50
Gambar 4.31 Bucketing Tabel Account.....	50
Gambar 4.32 Partisi Tabel Customer.....	50
Gambar 4.33 Bucketing Tabel Customer.....	50
Gambar 4.34 Partisi Tabel Saving	51
Gambar 4.35 Partisi Tabel Checking	51
Gambar 4.36 Partisi Tabel Deposit.....	52
Gambar 4.37 Partisi Tabel Summary	52
Gambar 4.38 Hasil Query 1 Skenario 1	53
Gambar 4.39 Hasil Query 2 Skenario 1	53
Gambar 4.40 Hasil Query 3 Skenario 1	54
Gambar 4.41 Hasil Query 1 Skenario 2	54
Gambar 4.42 Hasil Query 2 Skenario 2	55
Gambar 4.43 Hasil Query 3 Skenario 2	55
Gambar 4.44 Hasil Query 1 Skenario 3	56
Gambar 4.45 Hasil Query 2 Skenario 3	56
Gambar 4.46 Hasil Query 3 Skenario 3	57

Gambar 4.47 Hasil Query 1 Skenario 4	57
Gambar 4.48 Hasil Query 2 Skenario 4	58
Gambar 4.49 Hasil Query 3 Skenario 4	58
Gambar 4.50 Hasil Query 1 Skenario 5	59
Gambar 4.51 Hasil Query 2 Skenario 5	59
Gambar 4.52 Hasil Query 3 Skenario 5	60
Gambar 4.53 Hasil Query 1 Skenario 6	60
Gambar 4.54 Hasil Query 2 Skenario 6	61
Gambar 4.55 Hasil Query 3 Skenario 6	61
Gambar 4.56 Hasil Query 1 Skenario 7	62
Gambar 4.57 Hasil Query 2 Skenario 7	62
Gambar 4.58 Hasil Query 3 Skenario 7	63
Gambar 4.59 Hasil Query 1 Skenario 8	63
Gambar 4.60 Hasil Query 2 Skenario 8	64
Gambar 4.61 Hasil Query 3 Skenario 8	64
Gambar 4.62 Grafik Query 1	69
Gambar 4.63 Grafik Query 2	69
Gambar 4.64 Grafik Query 3	70

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini banyak perusahaan yang memanfaatkan teknologi informasi untuk meningkatkan kinerja bisnisnya dalam menghadapi persaingan bisnis yang ketat dan meraih pasar lebih luas, serta memperoleh keuntungan lebih besar, yaitu dengan cara mengoptimalkan penyajian informasi secara terintegrasi dan konsisten. Penerapan sistem *data warehouse* yang selama ini diterapkan oleh Bank XYZ, nyatanya memberikan dampak positif bagi perusahaan, diantaranya proses analisis ataupun pengelolaan informasi berdasarkan data historis yang terdapat di perusahaan akan menjadi lebih mudah, sehingga proses pengambilan keputusan strategis dapat dilakukan dengan lebih cepat berdasarkan proses penganalisan yang akurat karena didasarkan pada data historis yang telah terjadi selama ini.

Data warehouse merupakan salah satu bagian dari teknologi informasi yang dapat menunjang proses analisis perusahaan, dengan membantu *user* bisnis perusahaan dalam membantu menentukan kebijakan – kebijakan perusahaan agar dapat menghasilkan keputusan yang cepat dan tepat berdasarkan hasil analisis dari *data mining* dengan menggunakan data dan fakta yang telah diproses oleh *data warehouse*, dengan cara melakukan proses ETL, yaitu proses yang bertujuan untuk mengumpulkan sumber data yang berasal dari berbagai macam data operasional agar dapat terkumpul ke dalam sebuah media penyimpanan berupa basis data.

Seiring dengan penambahan jumlah dan kompleksitas data yang cukup pesat, pemrosesan data menggunakan *data warehouse* dirasa kurang optimal (Costa, 2019). Hal ini dipengaruhi oleh beberapa faktor, yaitu *data warehouse* yang tidak bisa memproses data yang tidak terstruktur, serta tidak mendukung pemrosesan data secara *near real time*. Maka pendekatan *data warehouse* menggunakan konsep *big data* menjadi solusi atas permasalahan tersebut.

Karena pada *big data*, proses yang dilakukan akan memakan waktu lebih cepat karena menggunakan proses paralelisasi dan *map reduce*, dan dapat menerima data yang tidak terstruktur seperti gambar, video dan suara.

Kinerja basis data merupakan isu yang tidak kalah penting saat data bertambah besar dan kompleks. Joseph McKendrick (2011) melakukan penelitian dengan kesimpulan bahwa jumlah data yang disimpan di piringan magnetik (disket, *harddisk*, dll) meningkat 100% pertahun, artinya setiap perusahaan di dunia yang menggunakan komputer sebagai tempat menyimpan data yang dimilikinya akan meningkat dua kali lipat per tahun. Oleh sebab itu perlu suatu cara untuk menjaga kinerja sistem basis data agar tetap optimal seiring dengan bertambahnya jumlah dan kompleksitas data.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

Seberapa besar peningkatan performa dari *data warehouse* setelah di optimasi dengan kompresi, partisi, *bucketing* pada basis data?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

- a. Optimasi yang dilakukan pada basis data adalah kompresi, partisi, dan *bucketing*.
- b. Performa yang diukur adalah waktu *query* pada basis data.
- c. Data yang digunakan adalah data pada server development yang telah *dimasking* atau dirubah.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan dari tugas akhir ini adalah sebagai berikut:

- a. Mengoptimalkan proses *data warehouse* pada Bank XYZ.
- b. Menganalisis pemrosesan *data warehouse* dengan desain yang telah di optimasi dan membandingkannya dengan desain yang belum di optimasi dalam hal kecepatan proses *query* pada tabel.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh setelah penelitian ini selesai yaitu:

- a. Pihak Bank XYZ dapat melakukan analisis terhadap data yang telah diproses dengan lebih cepat.
- b. Penelitian ini dapat menjadi referensi bagi penelitian lain, khususnya terkait optimasi *data warehouse*.

1.6 Metode Penelitian

a. Wawancara

Tahap wawancara ini dilakukan untuk mendapatkan informasi tentang proses bisnis yang terjadi, serta mendapatkan informasi tentang lingkungan untuk mengelola *data warehouse* pada Bank XYZ. Tujuan dari wawancara ini adalah sebagai bahan rujukan untuk pembangunan lingkungan basis data yang akan digunakan pada penelitian ini.

b. Analisis Desain Skema *Data Warehouse*

Proses analisis dilakukan untuk menarik kesimpulan dari hasil wawancara yang telah dilakukan sebelumnya. Pada proses wawancara sebelumnya yang menghasilkan gambaran lingkungan basis data yang sebenarnya, akan dilakukan proses implementasi basis data pada penelitian ini. Proses analisis ini dilakukan untuk perancangan struktur *data warehouse*, seperti menentukan tabel – tabel apa saja yang akan digunakan pada penelitian ini dan mengoptimasi performa tabel, langkah selanjutnya adalah menentukan skema *data warehouse* yang akan digunakan, skema yang didapat dari hasil analisis ini ada 4, yaitu skema untuk tabel rekening, pelanggan, transaksi, dan saldo.

c. Implementasi Desain *Data Warehouse*

Implementasi desain pada tahap ini adalah untuk mengimplementasikan hasil rancangan atau desain *data warehouse* yang telah dilakukan pada tahap analisis, setelah itu dilakukan juga implementasi

skema *data warehouse* untuk menghasilkan tabel-tabel yang akan digunakan dalam penelitian ini.

d. Pembangunan Data

Pada tahap ini, dilakukan proses pembangunan data, yaitu membangun data untuk dimasukkan ke dalam tabel-tabel pada skema basis data, yang selanjutnya akan digunakan untuk diolah dan digunakan sebagai bahan sumber proses ETL, data sumber tersebut akan disimpan dalam bentuk *textfile*, setelah pembangunan data selesai, maka data tersebut akan dimasukkan ke dalam tabel-tabel pada basis data dengan format *parquet*, yaitu tipe data yang telah dikompresi dalam tabel basis data.

e. Desain Struktur ETL

Desain struktur ETL adalah proses yang digunakan untuk merancang aliran data yang akan diproses dalam proses ETL, tahap ini dilakukan agar dapat mengetahui aliran data yang akan terjadi pada tiap skema tabel.

f. Implementasi Struktur ETL

Pada tahap implementasi struktur ETL ini dilakukan penerapan desain struktur ETL yang sebelumnya telah dirancang, implementasi ini akan menghasilkan struktur ETL yang akan digunakan.

g. Pengujian Performa Tabel

Tujuan dari pengujian ini adalah untuk menilai hasil performa implementasi tabel dengan desain yang telah dioptimasi sebelumnya. Performa tabel yang diuji pada penelitian ini yaitu membandingkan kecepatan dari segi jalan nya *query* antar tabel yang telah dioptimasi dan tidak dioptimasi, sehingga dari pengujian ini akan menghasilkan grafik perbedaan dengan ukuran waktu.

1.7 Sistematika Penulisan

Dalam penulisan skripsi ini, dibagi menjadi lima bab dengan beberapa sub pokok batasan. Adapun sistematika dari skripsi ini adalah sebagai berikut:

a. Bab I Pendahuluan

Dalam bab ini akan diuraikan tentang latar belakang penulisan, ruang lingkup atau batasan penelitian ini, tujuan serta manfaat dari penelitian ini, dan metodologi yang digunakan dalam penelitian ini secara sistematis.

b. Bab II Landasan Teori

Dalam bab ini akan diuraikan mengenai teori – teori yang digunakan dalam pembahasan penulisan skripsi ini dan sumber dari teori – teori tersebut.

c. Bab III Metodologi Penelitian

Dalam bab ini, akan dijelaskan tentang tata cara penelitian berdasarkan metodologi penelitian yang dipilih dan metode yang digunakan. Dimana pada metode pengumpulan data, penulis menggunakan metode observasi, wawancara, dan studi pustaka.

d. Bab IV Hasil dan Pembahasan

Bab ini akan dibahas mengenai gambaran umum perusahaan, analisis proses yang sedang berjalan, analisis dari data – data yang digunakan dalam perusahaan, serta akan dibahas juga cara optimasi pada desain tabel yang akan dilakukan.

e. Bab V Kesimpulan dan Saran

Bab ini merupakan akhir penulisan skripsi, di mana berdasarkan uraian – uraian yang telah dibahas akan dituangkan ke dalam suatu bentuk kesimpulan akhir serta saran – saran.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian tentang analisis optimasi yang telah dilakukan, penulis memperoleh hasil kesimpulan sebagai berikut:

- a. Optimasi yang dilakukan dengan cara analisis penggunaan partisi, *bucketing*, dan kompresi pada basis data Hive telah berhasil.
- b. Perlakuan partisi dan *bucketing* pada tabel akan memberikan waktu *query* yang lebih cepat, hal ini dibuktikan dari hasil skenario 7 (skenario dengan partisi dan *bucketing*) yang dibandingkan dengan skenario 1 (tanpa partisi, tanpa *bucketing*), pada *query* 1, waktu yang diperlukan untuk menyelesaikan *query* pada skenario 7 menjadi lebih cepat 38 detik atau sekitar 20.7% jika dibandingkan dengan skenario 1, begitu juga pada *query* 2, waktu yang diperlukan bertambah cepat sebesar 16 detik atau sekitar 6.5%, sedangkan pada *query* 3, waktu yang diperlukan bertambah cepat sebesar 158 detik atau sekitar 58%.
- c. Kompresi data akan mengecilkan ukuran data pada ruang penyimpanan, jika skenario 4 (skenario dengan kompresi) dibandingkan dengan skenario awal, terdapat penyusutan ukuran sebesar 8 kali hingga 30 kali, tetapi ada kelemahan dari kompresi ini, saat data di kompresi, rata-rata waktu *query* yang dibutuhkan akan menjadi lebih lama, hal ini dibuktikan dari membandingkan skenario 1 (skenario awal), dengan skenario 4 (skenario berupa kompresi saja), pada *query* 1, waktu yang diperlukan oleh skenario 4 untuk menyelesaikan *query* ini melambat sebesar 86% jika dibandingkan dengan skenario 1, sementara pada *query* 2, waktu yang diperlukan untuk menyelesaikan *query* ini melambat sebesar 122%, sedangkan untuk *query* 3, waktu yang diperlukan melambat sebesar 34%.

5.2 Saran

Dari hasil penelitian yang telah dilakukan, berikut saran yang dihasilkan:

Basis data yang digunakan pada penelitian ini berfokus pada Hive, yaitu basis data yang ditujukan untuk melakukan proses analisis data, terdapat contoh lain basis data pada ekosistem Hadoop, contohnya HBase, HBase berfokus pada *real time querying*, yang ideal digunakan jika memerlukan data yang acak saat *write* atau *read*, sehingga perlu dilakukan penelitian faktor optimasi lebih lanjut tentang basis data HBase atau basis data lain yang terintegrasi dengan Hadoop, sehingga pada penelitian lain, dapat dipilih basis data yang sesuai dan dibuat secara optimal untuk keperluan pemrosesan data secara efisien dan cepat.

DAFTAR PUSTAKA

- Abdel, A. (2018). Optimizing Join in HIVE Star Schema Using Key/Facts Indexing. *IETE Technical Review*, 132-144.
- Ahmad, A. (2018). Multilevel Data Processing Using Parallel Algorithms for Analyzing Big Data in High-Performance Computing. *Int J Parallel Prog.*
- Basanta-Val, P. (2017). An Efficient Industrial Big-Data Engine. *IEEE Transactions on Industrial Informatics*, 1361 - 1369.
- Costa, E. (2019). Evaluating partitioning and bucketing strategies for Hive-based Big Data Warehousing systems. *Journal of Big Data*.
- Dean, DJ. (2017). Efficient Data Retrieval in Big-Data Processing Systems. *International Business Machines Corporation*.
- Hashem, AT. (2018). Multi-objective scheduling of MapReduce jobs in big data processing. *Multimed Tools Appl.*
- Kaur, R., & Chadha, R. (2017). Comparative Analysis of Various File Formats in HIVE. *International Journal of Technology and Computing*, Volume 3 Issue 6.
- Li, W. (2019). Data mining optimization model for financial management information system based on improved genetic algorithm. *Inf Syst E-Bus Manage.*
- Rathore, MM. (2018). Real-Time Big Data Stream Processing Using GPU with Spark Over Hadoop Ecosystem. *Int J Parallel Prog.*
- Saavedra, MZN., & Yu, S. (2017). A Comparison between Text, Parquet, and PCAP Formats for Use in Distributed Network Flow Analysis on Hadoop. *Journal of Advances in Computer Networks*, Volume 5.
- Sachin, S. (2019). Nuts and Bolts of ETL in Data Warehouse. *Emerging Trends in Expert Applications and Security. Advances in Intelligent Systems and Computing*, vol. 841.
- Wang, L. (2018). High performance cloud computing for remote sensing big data management and processing. *Future Generation Computer Systems*, 353-368.
- White, M. (2018). Big Data vs. Data Warehousing. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 3 Issue 5.