

2.Buku ajar Data Warehouse 2020

by Buku 2

Submission date: 05-Mar-2023 01:59AM (UTC+0700)

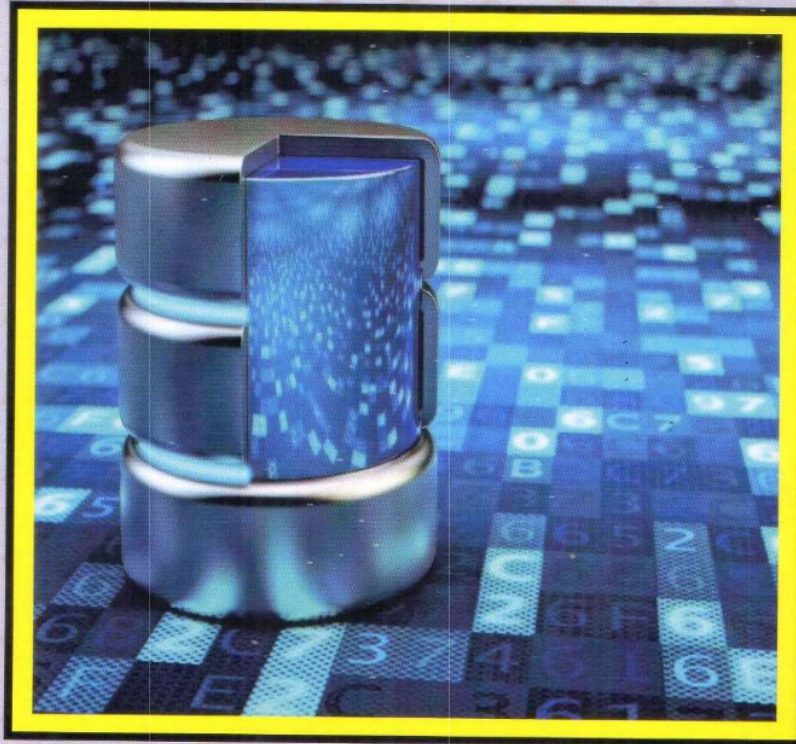
Submission ID: 2028806590

File name: 2.Buku_ajar_Data_Warehouse_2020.pdf (4.98M)

Word count: 10299

Character count: 66979

BUKU AJAR DATA WAREHOUSE



Oleh:

I GL. Putra Eka Prisma, S.Kom., M.Kom
Dedy Rahman Prehanto, S.Kom., M.Kom
Ginanjari Setyo Permadi S.Kom., M.Kom



Penerbit BINTANG Surabaya

BUKU AJAR DATA WAREHOUSE

Oleh:

I GL. Putra Eka Prisma, S.Kom., M.Kom
Dedy Rahman Prehanto, S.Kom., M.Kom
Ginanjari Setyo Permadi S.Kom., M.Kom



Penerbit **BINTANG** Surabaya

BUKU AJAR
DATA WAREHOUSE

BUKU AJAR DATA WAREHOUSE

Hak Cipta © 2020

Penulis : I Gusti Lanang Putra Eka Prisma, S.Kom., M.Kom
Dedy Rahman Prehanto, S.Kom., M.Kom
Ginanjar Setyo Permadi S.Kom., M.Kom

Design Sampul : Tim Penerbit
Layout : Tim Penerbit

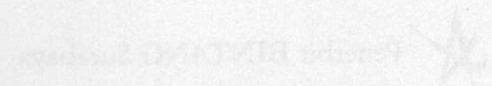
Penerbit : Penerbit Bintang Surabaya
Anggota IKAPI daerah Jawa Timur
No: 011/JTI/95

Isi : viii, 101 hal

ISBN 978-602-6534-19-4



Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi buku ini dalam bentuk apapun, baik secara elektronik maupun mekanis, termasuk memfotokopi, merekam atau dengan sistem penyimpanan lainnya, tanpa izin tertulis dari penerbit



Kata Pengantar

Alhamdulillah, ducapkan kehadiran Allah SWT atas segala limpahan rahmat, taufik serta hidayahNya sehingga dapat diselesaikannya buku ajar "Data Warehouse"

Buku ajar mencakup sistem data warehouse yang mulai menjadi kebutuhan untuk melakukan analisa data menjadi informasi terkini. Data Warehouse merupakan basisdata yang dirancang untuk mengerjakan proses query, membuat laporan dan analisa. Data yang disimpan dalam sebuah Data Warehouse adalah data histori dari sebuah organisasi/perusahaan yang mana data tersebut tidak tersimpan secara rinci/detil. Data Warehouse berbeda dengan data OLTP (*Online Transactional Processing*) yang tersimpan sampai prosesnya berlangsung secara lengkap.

Semoga Buku Ajar ini akan dapat membantu para mahasiswa untuk lebih memahami dan mampu mengaplikasikan penggunaan Data Warehouse. Akhir kata penulis sangat mengharapkan masukan maupun saran yang membangun guna perbaikan, karena penulis menyadari bahwa buku ajar ini jauh dari kata sempurna.

DAFTAR ISI

Kata Pengantar	iii
Daftar Isi	v
BAB I Pendahuluan	1
A. Latar Belakang	1
B. Deskripsi Mata Kuliah	2
C. Tujuan Mata Kuliah	3
BAB II Pengantar Data Warehouse	4
A. Tujuan Instruksional Umum	4
B. Tujuan Instruksional Khusus	4
C. Materi	4
1. Penertian Data Warehouse	4
2. Data Warehouse	8
3. Istilah Data Warehouse	11
4. Peran Data Warehouse	12
5. Kelebihan Data Warehouse	14
6. Contoh Aplikasi Data Warehouse Secara Umum	15
7. Evaluasi	16
BAB III Karakteristik Data Warehouse	17
A. Tujuan Instruksional Umum	17
B. Tujuan Instruksional Khusus	17
C. Karakteristik Data Warehouse	17
D. Tugas	26
E. Evaluasi	28
BAB IV Komponen Data Warehouse	29
A. Tujuan Instruksional Umum	29
B. Tujuan Instruksional Khusus	29
C. Komponen Sumber Data	31

D. Komponen Data Staging	94
1. Data Extraction	36
2. Data Transformation	37
3. Data Loading	39
E. Komponen Data Storage	40
F. Komponen Information Delivery	41
G. Komponen Metadata	42
H. Komponen Kontrol dan Manajemen	43
I. Evaluasi	43
 BAB V Dukungan Infrastruktur Data Warehouse	 44
A. Tujuan Instruksional Umum	44
B. Tujuan Instruksional Khusus	44
C. Materi	44
1. Infrastruktur Operasional	45
2. Infrastruktur Fisik	46
3. Sistem Operasi dan Perangkat Keras	47
D. Evaluasi	61
 BAB VI Pemodelan Data Multidimensi	 62
A. Tujuan Instruksional Umum	62
B. Tujuan Instruksional Khusus	62
C. Materi	62
1. Dimensional Modeling	62
2. Tabel Dimensional	64
3. Fact	65
4. Tabel Fact	67
5. Pemodelan Dimensional	68
6. Hirarki Dalam Dimensi	72
7. Pengimplementasian Hirarki Dimensional	76
8. Penggunaan Hirarki Dimensional	77
9. Operasi OLAP	78
10. Metodologi Perancangan Database untuk Data Warehouse	81
D. Evaluasi	84

BAB VII Extraction, Transformation dan Loading	86
A. Tujuan Instruksional Umum	86
B. Tujuan Instruksional Khusus	86
C. Materi	86
1. Extraction Data Warehouse	87
2. Loading dan Transformasi	94
D. Evaluasi	100
Daftar Rujukan	101

PENDAHULUAN

A. Latar Belakang

Bahan ajar merupakan salah satu komponen sistem pembelajaran yang memegang peranan penting dalam membantu siswa mencapai Standar Kompetensi dan Kompetensi Dasar atau tujuan pembelajaran yang telah ditentukan.

Salah satu masalah penting yang sering dihadapi oleh guru dalam kegiatan pembelajaran adalah memilih atau menentukan bahan ajar atau materi pembelajaran yang tepat dalam rangka membantu siswa mencapai kompetensi yang diharapkan. Hal ini disebabkan oleh kenyataan bahwa dalam kurikulum atau silabus, materi bahan ajar hanya dituliskan secara garis besar dalam bentuk materi pokok. Tugas guru adalah menjabarkan materi pokok tersebut sehingga menjadi bahan ajar yang dikembangkan dan mudah dipahami oleh siswa.

Dengan menerapkan bahan ajar yang telah dikembangkan tersebut, diharapkan menjadi alternatif bagi guru dalam menyampaikan suatu materi pembelajaran sehingga proses belajar mengajar akan berjalan lebih baik dan bervariasi. Pada akhirnya hasil belajar siswa juga ikut meningkat. Maka dalam makalah ini penulis akan membahas tentang bahan ajar yang merupakan bagian dari hasil perencanaan seorang guru sebelum mengajar di kelas.

B. Pengertian Data Warehouse

Berdasarkan dari definisi dari ahli yang telah dijelaskan maka dapat disimpulkan data warehouse merupakan database yang saling bereaksi dan dapat digunakan untuk query dan analisis, bersifat orientasi subjek, terintegrasi, time-variant, tidak mudah berubah yang digunakan untuk membantu oleh decision maker. Data Warehouse merupakan sistem apa saja yang mengumpulkan data dari berbagai sumber kedalam suatu tempat yang baru. Data Warehouse tersebut digunakan sebagai pusat repositori data untuk keperluan analisis dan pelaporan.

C. Tujuan Penulisan

1. Untuk mengetahui kompetensi pengajaran mata kuliah data warehouse
2. Untuk menjadikan pokok pembelajaran yang sistematis
3. Membantu mahasiswa dalam memperoleh alternatif buku pembelajaran data warehouse

BAB II

PENGANTAR DATA WAREHOUSE

A. Tujuan Instruksional Umum

Mahasiswa diharapkan dapat memahami konsep Data Warehouse dan urgensi data warehouse.

B. Tujuan Instruksional Khusus

Mahasiswa diharapkan mampu :

1. Memahami pengertian data, informasi, database
2. Memahami pengertian data warehouse
3. Memahami konsep dasar data warehouse
4. Menjelaskan urgensi dari data warehouse

Sebelum ⁴ membahas tentang data warehouse, adapun hal yang harus dimengerti terlebih dahulu antara lain pengertian dari data, informasi dan database. Kita mungkin pernah secara sadar atau tidak mengucapkan tentang data atau informasi. Namun kenyataannya banyak yang belum memahami serta membedakan apa itu data dan informasi.

Oleh karena itu akan dibahas pengertian dari data dan informasi, serta apa yang membedakan penggunaan kedua istilah tersebut. Adapun juga akan menjelaskan pengertian dari database dan contoh penggunaannya secara nyata.

Menurut Steven Alter, data merupakan nilai fakta, gambar atau suara yang mungkin atau tidak berhubungan atau berguna

bagi tugas tertentu serta dapat dipertanggung jawabkan kebenarannya. Sedangkan informasi merupakan hasil proses olah dari data sehingga memberikan arti dan nilai tambah. Adapun contoh dari data berupa nilai yang terformat, teks, citra, audio, dan video.

1. Data yang terformat, misalnya data yang menyatakan tanggal, waktu atau nilai mata uang.
2. Teks, misalnya deretan huruf dan angka serta simbol.
3. Audio, misalnya Instrumen musik, suara orang atau suara binatang, gemericik air, detak jantung, dll.
4. Video, misalnya film dan video pendek.

Data yang diperoleh suatu organisasi atau perusahaan pada umumnya dari rutinitas kegiatan operasional ataupun hasil transaksi yang dilakukan. Jadi pengertian database secara sederhana merupakan media penyimpanan data yang saling berkaitan secara logika, sehingga bisa digunakan untuk menghasilkan informasi yang diperlukan oleh suatu organisasi atau perusahaan.

Menurut Fatansyah, database merupakan kumpulan data yang saling berhubungan yang disimpan secara bersama sedemikian rupa dan tanpa pengulangan (redundansi) yang tidak perlu, untuk memenuhi berbagai kebutuhan. Adapun juga Menurut James A. O'Brien, database merupakan suatu koleksi terintegrasi dimana secara logika berhubungan dengan record dari file. Seiring banyaknya model database yang mulai dikembangkan, maka muncullah sebuah istilah baru yang dikenal dengan data warehouse.

C. ²² Data Warehouse

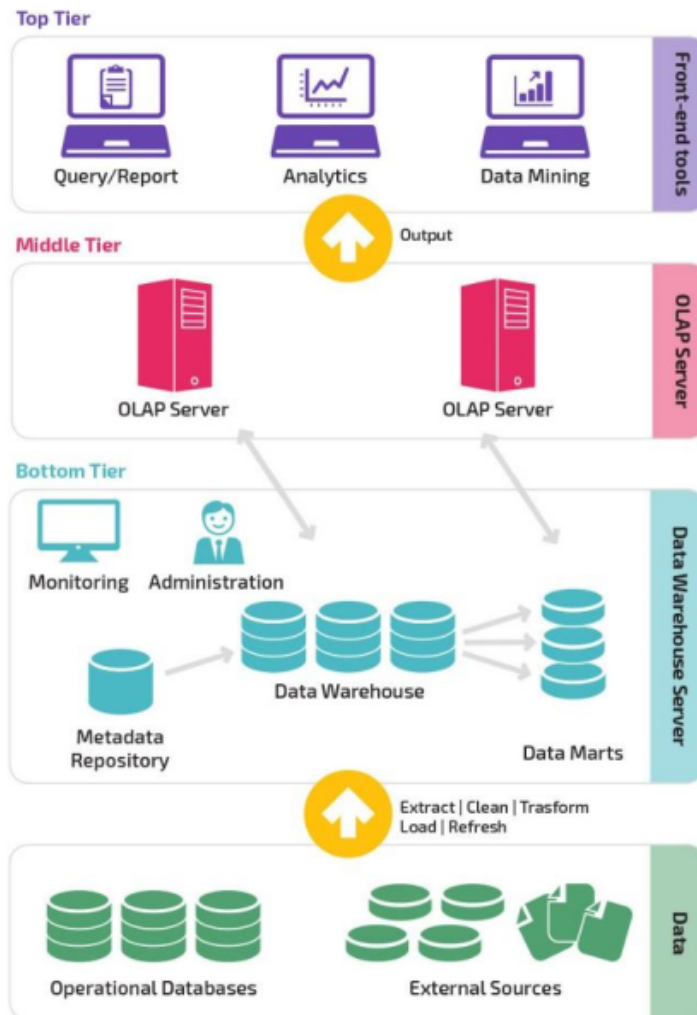
Adapun pengertian dari data warehouse menurut para ahli adalah sebagai berikut.

Menurut W.H. Inmon dan Richard D.H., data warehouse adalah koleksi data yang mempunyai sifat berorientasi pada subjek, terintegrasi²⁴ time-variant, dan bersifat tetap atau tidak mudah berubah dari koleksi data dalam mendukung proses pengambilan keputusan management.

Menurut Vidette Poe, data warehouse merupakan basisdata yang bersifat analisis dan read only yang digunakan sebagai¹⁵ pondasi dari sistem penunjang keputusan (SPK).

Menurut Paul Lane, data warehouse merupakan basisdata relasional yang didesain lebih ke query dan analisa dari pada proses transaksi, biasanya mengandung history data dari proses transaksi dan bisa juga data dari sumber lainnya.

Berdasarkan dari definisi dari ahli yang telah dijelaskan maka dapat disimpulkan data warehouse merupakan database yang saling bereaksi dan dapat digunakan untuk query dan analisis, bersifat orientasi subjek, terintegrasi, time-variant, tidak mudah be³⁵ah yang digunakan untuk membantu oleh decision maker. Data Warehouse merupakan sistem apa saja yang mengumpulkan data dari berbagai sumber kedalam suatu tempat yang baru. ³⁴ta Warehouse tersebut digunakan sebagai pusat repositori data untuk keperluan analisis dan pelaporan. Arsitektur data warehouse tradisional menggunakan struktur tingkat tiga yang terdiri dari tingkatan berikut.



Gambar 1 Arsitektur data warehouse tradisional

1. Tingkat Atas

Tingkat atas adalah lapisan klien. Tingkatan ini memegang alat yang digunakan untuk analisis data tingkat tinggi, laporan kueri, dan data mining.

2. Tingkat Tengah

Tingkat menengah menampung server OLAP, yang mengubah data menjadi struktur yang lebih cocok untuk analisis dan permintaan yang kompleks. Server OLAP dapat bekerja dalam dua cara yaitu sebagai sistem manajemen basis data relasional yang diperluas untuk memetakan operasi pada data multidimensi ke operasi relasional standar (Relational OLAP), atau menggunakan model OLAP multidimensi yang secara langsung mengimplementasikan data dan operasi multidimensi.

3. Tingkat Bawah

Tingkatan ini berisi server database yang digunakan untuk mengekstrak data dari berbagai sumber, seperti basis data transaksional yang digunakan untuk aplikasi front-end.

Beberapa bidang yang berhasil mengaplikasikan teknologi data warehouse :

1. Penjualan : analisis penjualan dan klaim, pengiriman dan kontrol inventori, layanan konsumen, dan layanan masyarakat.
2. Keahlian : kontrol biaya produksi, pemasok dan pendukung penjualan
3. Jasa Keuangan : analisis resiko dan kartu kredit, pendeteksi penipuan
4. Transportasi industri : manajemen kendaraan
5. Layanan Telekomunikasi : menganalisa analisa panggilan dan profil pelanggan
6. Layanan Perawatan Panggilan : penerimaan pasien dan analisis debit dan pembukuan di departemen akun.

D. Istilah Data Warehouse

Adapun Istilah-istilah yang berkaitan dengan data warehouse seperti,

1. Data Mart

suatu bagian pada data warehouse yang mendukung pembuatan laporan dan analisa data pada suatu unit, bagian atau operasi pada suatu perusahaan.

2. On-Line Analytical Processing(OLAP)

suatu pemrosesan database yang menggunakan tabel fakta dan dimensi untuk dapat menampilkan berbagai macam bentuk laporan, analisis, query dari data yang berukuran besar.

3. On-Line Transaction Processing(OLTP)

Merupakan suatu pemrosesan yang menyimpan data mengenai kegiatan operasional transaksi sehari-hari.

4. Dimension Table

Tabel yang berisikan kategori dengan ringkasan data detail yang dapat dilaporkan. Seperti laporan laba pada tabel fakta dapat dilaporkan sebagai dimensi waktu(yang berupa perbulan, perkwartal dan pertahun).

5. Fact Table

Merupakan tabel yang umumnya mengandung angka dan data history dimana key (kunci) yang dihasilkan sangat unik, karena key tersebut terdiri dari foreign key(kunci asing) yang merupakan primary key (kunci utama) dari beberapa dimension table yang berhubungan.

6. DSS

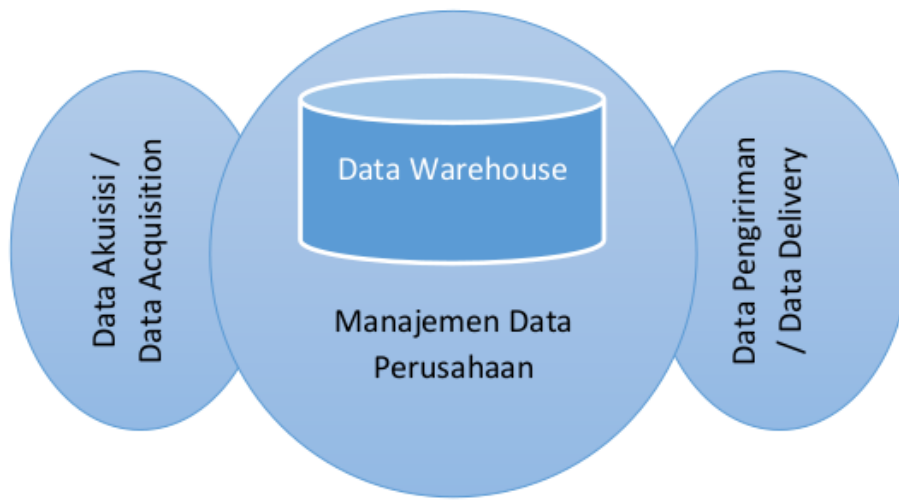
Merupakan sistem yang menyediakan informasi kepada pengguna yang menjelaskan bagaimana sistem ini dapat menganalisa situasi dan mendukung suatu keputusan yang baik.

E. Peran Data Warehouse

Setelah dijelaskan pengertian serta penggunaan data warehouse pada pembahasan sebelumnya, maka adapun fungsi utama dari penggunaan data warehouse sebagai berikut.

- ✚ Dapat mencerminkan secara langsung tentang aturan dan kebijakan sebuah perusahaan.
- ✚ Dapat mengandung informasi-informasi yang strategis.
- ✚ Berisi informasi di masa lalu yang dapat digunakan untuk pengembangan.
- ✚ Sebagai sumber informasi yang akan dikirimkan ke data mart.
- ✚ Sebagai sumber data yang menggambarkan bagaimana sebuah proses bisnis mengalami perubahan.

Pada Gambar menjelaskan posisi dari penggunaan data warehouse dalam perusahaan atau organisasi.



Gambar 2 Data Warehouse dalam perusahaan

F. Kelebihan Data Warehouse

Setelah mempelajari seputar data warehouse tentu kita bertanya apa yang dapat dilakukan atau keuntungan oleh penggunaan data warehouse tersebut. Berikut adalah salah satu manfaat penggunaan data warehouse :

- ✚ Memberikan informasi secara cepat
- ✚ Data terintegrasi secara menyeluruh, bahkan hingga luar organisasi.
- ✚ Dapat memberikan gambaran / visi di masa mendatang dengan melihat data-data (trend) di masa lalu.
- ✚ Sebagai cara baru untuk melihat data.
- ✚ Tidak terikat dengan bagian Sistem Informasi (Tidak memerlukan programmer khusus untuk memanfaatkan data warehouse)
- ✚ Kemampuan untuk mengakses data yang besar

- ✚ Kemampuan untuk memiliki data yang konsisten
- ✚ Kemampuan kinerja analisa yang cepat
- ✚ Mengetahui adanya hasil yang berulang-ulang
- ✚ Menemukan adanya celah pada business knowledge atau business process.

G. Contoh Aplikasi Data Warehouse Secara Umum

Adapun beberapa contoh dari pengaplikasian data warehouse yang sering kita temui.

1. Analisis Penjualan

- a. Penjualan produk secara real-time untuk menentukan harga dan kebijakan-kebijakannya.
- b. Analisis penjualan produk di masa lalu untuk melakukan pengembangan atau pencegahan kegagalan.
- c. Evaluasi produk-produk yang berhasil dan menentukan faktor-faktor kunci keberhasilan.
- d. Identifikasi kelompok pelanggan berdasarkan pemasukan dan pengeluaran.
- e. Pemisahan secara cepat pelanggan-pelanggan yang sudah lama tidak datang membeli.
- f. Identifikasi secara harian produk-produk yang tersedia di gudang dan jalur distribusi.

2. Analisis Keuangan

- a. Membandingkan biaya-biaya secara tahunan, bulanan dan harian.
- b. Melakukan review tren cash flow di masa lalu dan prediksi kebutuhan di masa depan.

3. Analisis SDM

- a. Evaluasi tren dalam penggunaan program-program yang menguntungkan.
- b. Mengidentifikasi keuntungan-keuntungan biaya untuk menentukan skala perluasan perusahaan.

4. Area Lain

Warehouse juga diterapkan pada berbagai bidang, antara lain: logistics, inventory, purchasing, detailed transaction analysis dan load balancing.

H. Evaluasi

1. Jelaskan apakah perbedaan data dan informasi.
2. Apakah yang anda ketahui tentang data warehouse?
3. Tujuan apa yang mendasari untuk memilih menggunakan data warehouse?
4. Jelaskan struktur tingkatan dari data warehouse
5. Sebutkan dan jelaskan contoh bidang yang menerapkan data warehouse.

BAB III

KARAKTERISTIK DATA WAREHOUSE

A. Tujuan Instruksional Umum

Mahasiswa diharapkan dapat memahami karakteristik dan tugas Data Warehouse.

B. Tujuan Instruksional Khusus

Mahasiswa diharapkan mampu :

1. Memahami karakteristik data warehouse
2. Memahami tugas data warehouse

44

C. Karakteristik Data Warehouse

Menurut Inmon karakteristik data warehouse sebagai berikut ini.

1. Subject Oriented (Berorientasi Subject)

Data warehouse berorientasi subject artinya data warehouse didesain untuk menganalisa data berdasarkan subject-subject tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu. Data warehouse diorganisasikan disekitar subjek-subjek utama dari perusahaan (customers, products dan sales) dan tidak diorganisasikan pada area-area aplikasi utama (customer invoicing, stock control dan product sales). Hal ini dikarenakan kebutuhan dari data warehouse untuk menyimpan data-data yang bersifat sebagai penunjang suatu keputusan, dari pada aplikasi yang berorientasi terhadap data.

Adapun area khas subjek seperti :

- ✚ Customer
- ✚ Product
- ✚ Transaction or activity
- ✚ Policy
- ✚ Claim
- 5 ✚ Account

Jadi dengan kata lain, data yang disimpan adalah berorientasi kepada subjek bukan terhadap proses. Secara garis besar perbedaan antara data operasional dan data warehouse seperti pada Tabel 1 berikut ini.

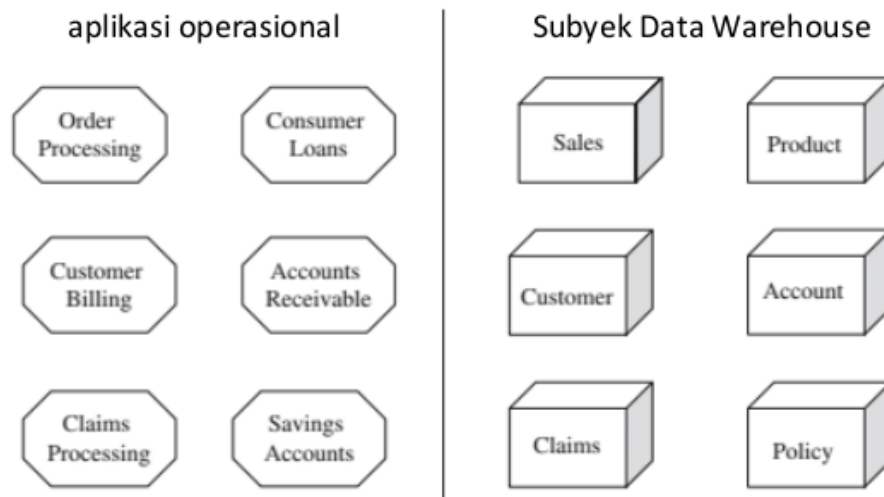
3
Tabel 1 Perbedaan Data Operasional dan Data Warehouse

Data Operasional	Data Warehouse
1. Dirancang berorientasi hanya pada aplikasi dan fungsi tertentu.	1. Dirancang berdasar pada subjek-subjek tertentu(utama)
2. Fokusnya pada desain database dan proses	2. Fokusnya pada pemodelan data dan desain data
3. Berisi rincian atau detail data	3. Berisi data history yang dipakai proses analisis
4. Relasi antar table berdasar aturan terkini (selalu mengikuti rule (aturan terbaru))	4. Banyak aturan bisnis dapat tersaji antara tabel-tabel

40
Pada Gambar membedakan antara bagaimana data disimpan dalam sistem operasional dan dalam data warehouse. Dalam sistem operasional, data untuk setiap aplikasi disusun secara terpisah dengan aplikasi yang ditampilkan (pemrosesan pesanan, pinjaman konsumen, tagihan pelanggan, piutang,

pemrosesan klaim, dan rekening tabungan). Misalkan klaim sangat penting untuk subjek bisnis perusahaan asuransi. Klaim berdasarkan polis asuransi mobil diproses dalam aplikasi asuransi mobil. Klaim data untuk asuransi mobil terorganisir dalam satu aplikasi. Demikian juga, data klaim untuk asuransi kompensasi pekerja diatur dalam aplikasi asuransi khusus pekerja. Tetapi di data warehouse untuk perusahaan asuransi, data tentang klaim diatur oleh subjek klaim dan tidak oleh masing-masing aplikasi asuransi mobil ataupun pekerja. Di data warehouse, tidak ada rasa aplikasi.

Dalam data warehouse, data tidak disimpan secara operasional aplikasi, melainkan oleh subyek bisnis.



Gambar 3 Data Warehouse Berorientasi Objek

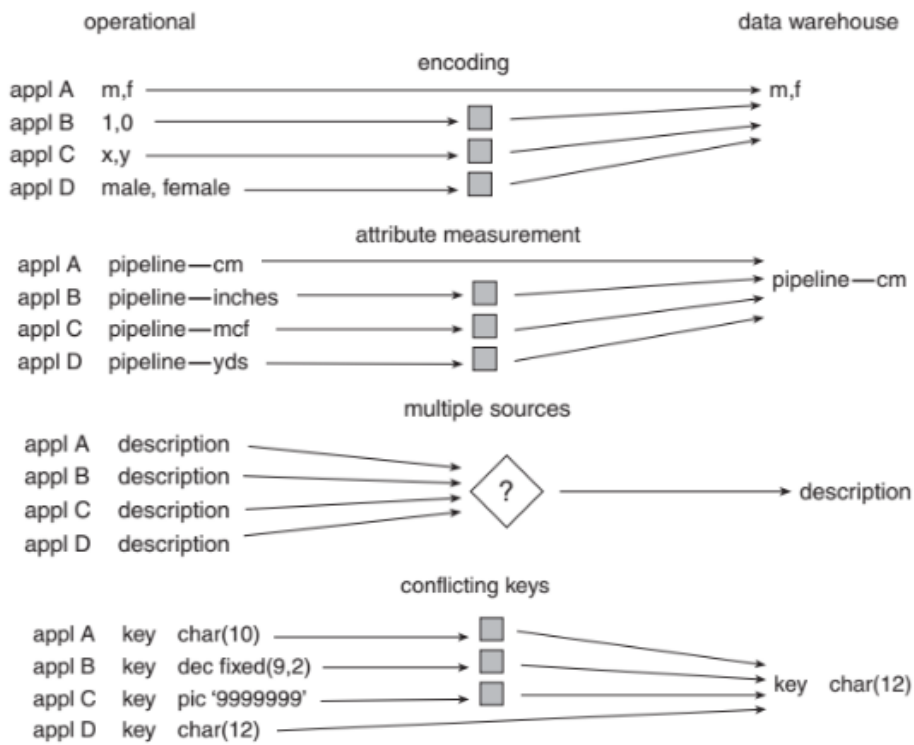
3

2. Integrated (Terintegrasi)

Data warehousing dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah kedalam suatu format yang konsisten dan saling terintegrasi satu dengan lainnya. Dengan demikian data tidak bisa dipecahpecah karena data yang ada merupakan suatu kesatuan yang menunjang keseluruhan konsep data warehousing itu sendiri.

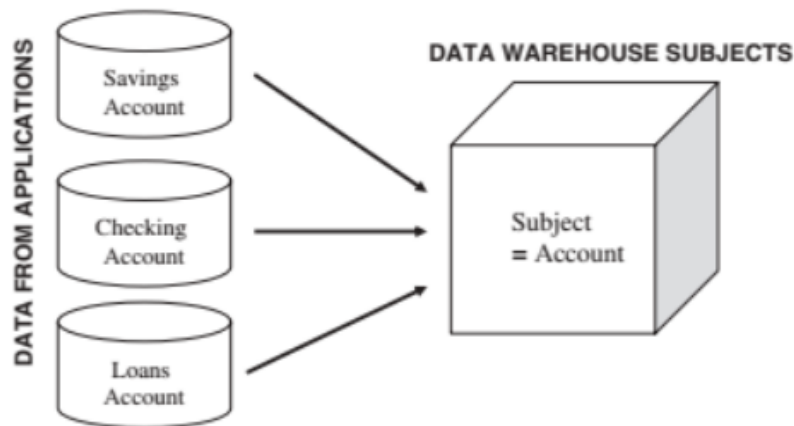
Syarat integrasi sumber data dapat dipenuhi dengan berbagai cara seperti konsisten dalam penamaan variable, konsisten dalam ukuran variable, konsisten dalam struktur pengkodean dan konsisten dalam atribut fisik dari data.

Contoh pada lingkungan operasional terdapat berbagai macam aplikasi yang mungkin pula dibuat oleh developer yang berbeda. Oleh karena itu, mungkin dalam aplikasi-aplikasi tersebut ada variable yang memiliki maksud yang sama tetapi nama dan formatnya berbeda. Variable tersebut harus dikonversi menjadi nama yang sama dan format yang disepakati bersama. Dengan demikian tidak ada lagi kerancuan karena perbedaan nama, format dan lain sebagainya. Barulah data tersebut bisa dikategorikan sebagai data yang terintegrasi karena kekonsistennya seperti pada Gambar.



Gambar 4 Integrasi Data Warehouse

Berikut contoh gambaran proses integrasi data yang sederhana dari perbankan.



Gambar 5 Integrasi Data Warehouse Perbankan

46

Pada Gambar data yang dimasukkan ke dalam subjek data warehouse berasal dari tiga aplikasi operasional yang berbeda. Dalam tiga aplikasi tersebut mungkin terdapat beberapa variasi penamaan yang berbeda seperti atribut untuk item data bisa berbeda. Nomor rekening dalam aplikasi saving account bisa panjang delapan byte, tetapi hanya enam byte dalam aplikasi checking account.

Sebelum data dari berbagai sumber yang berbeda dapat disimpan dalam suatu data warehouse, langkah awal harus menghapus ketidakkonsistenan. Kita harus membuat aturan baku sendiri dari berbagai elemen data dan memastikan makna nama data dalam setiap aplikasi sumber. Sebelum memindahkan data ke dalam data warehouse harus melalui proses transformasi, konsolidasi, dan integrasi sumber data.

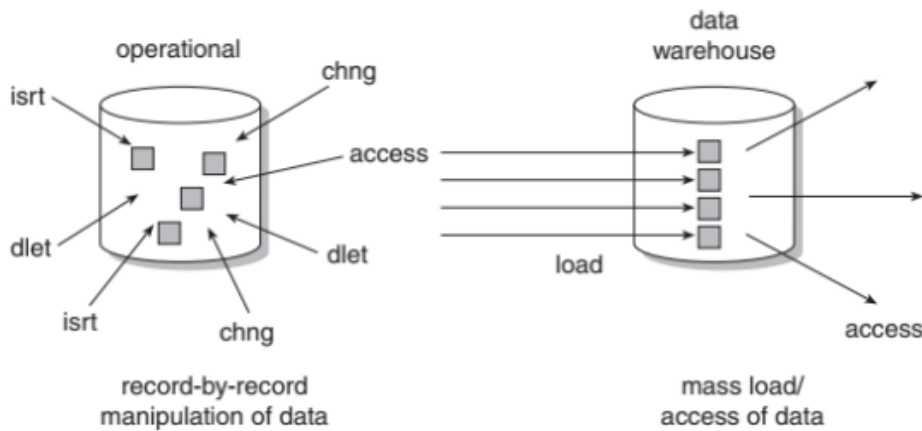
Berikut adalah beberapa item yang perlu distandarisasi dan dibuat konsisten, yaitu :

- ✚ Naming conventions

- ✚ Codes
- ✚ Data attributes
- ✚ Measurements

3. Non-volatile

Karakteristik penting selanjutnya dari data warehouse adalah data tersebut tidak mudah berubah-ubah dengan artian data tidak di-update secara real time tetapi di refresh dari sistem operasional secara reguler. Berbeda dengan sistem database operasional yang dapat melakukan update, insert dan delete terhadap data untuk keperluan mengubah isi dari database sedangkan pada data warehouse hanya dapat melakukan dua kegiatan memanipulasi data seperti pada Gambar yaitu loading data (mengambil data) dan akses data (mengakses data warehouse seperti melakukan query atau menampilkan laporan yang dibutuhkan, tidak ada kegiatan updating data).



Gambar 6 Non-Volatile Data Warehouse

4. Rentang Waktu (Time Variant Data)

6

Seluruh data pada data warehouse dapat dikatakan akurat atau valid pada rentang waktu tertentu. Untuk melihat interval waktu yang digunakan dalam mengukur keakuratan suatu data warehouse, kita dapat menggunakan cara antara lain :

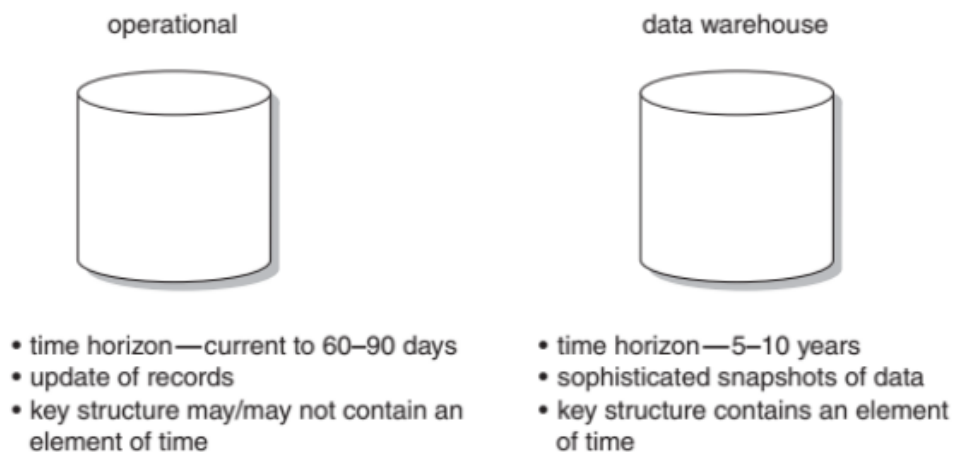
- ✚ Cara yang paling sederhana adalah menyajikan data warehouse pada rentang waktu tertentu, misalnya antara 5 sampai 10 tahun ke depan.
- ✚ Cara yang kedua, dengan menggunakan variasi/perbedaan waktu yang disajikan dalam data warehouse baik implicit maupun explicit secara explicit dengan unsur waktu dalam hari, minggu, bulan dsb. Secara implicit misalnya pada saat data tersebut diduplikasi pada setiap akhir bulan, atau per tiga bulan. Unsur waktu akan tetap ada secara implisit didalam data tersebut.
- ✚ Cara yang ketiga, variasi waktu yang disajikan data warehouse melalui serangkaian snapshot yang panjang. Snapshot merupakan tampilan dari sebagian data tertentu sesuai keinginan pemakai dari keseluruhan data yang ada bersifat read-only.

Sebagai contoh, di data warehouse yang berisi unit penjualan, jumlah disimpan di setiap file catatan atau baris tabel

berhubungan dengan elemen waktu tertentu. Tergantung pada tingkat detail di data warehouse, jumlah penjualan dalam catatan mungkin berhubungan dengan tanggal, minggu, bulan tertentu. Sifat varian waktu dari data dalam data warehouse untuk :

- ✚ Memungkinkan analisis masa lalu
- ✚ Mengaitkan informasi dengan saat ini
- ✚ Memungkinkan perkiraan untuk masa depan

Adapun perbandingan operasional dengan data warehouse pada Gambar 7 berikut.



Gambar 7 Time variant

- D. Tugas Data Warehouse**
Ada empat tugas yang bisa dilakukan dengan adanya data warehouse menurut Williams, keempat tugas tersebut yaitu:
1. Pembuatan laporan

Pembuatan laporan merupakan salah satu kegunaan data warehouse yang paling umum dilakukan. Dengan menggunakan query sederhana didapatkan laporan perhari,perbulan, pertahun atau jangka waktu kapanpun yang diinginkan.

2. On-Line Analytical Processing (OLAP)

Dengan adanya data warehouse, semua informasi baik detail maupun hasil summary yang dibutuhkan dalam proses analisa mudah didapat.

3. Data mining

Data mining merupakan proses untuk menggali(mining) pengetahuan dan informasi baru dari data yang berjumlah banyak pada data warehouse, dengan menggunakan kecerdasan buatan (Artificial Intelligence), statistik dan matematika. Data mining merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pemakainya.

4. Proses informasi executive

Data warehouse dapat membuat ringkasan informasi yang penting dengan tujuan membuat keputusan bisnis, tanpa harus menjelajahi keseluruhan data. Dengan menggunakan data warehouse segala laporan telah diringkas dan dapat pula mengetahui segala rinciannya secara lengkap, sehingga mempermudah proses pengambilan keputusan. Informasi dan data pada laporan data warehouse menjadi target informative bagi user.

E. Evaluasi

1. Apa pemahaman Anda tentang arsitektur data warehouse?
2. Sebutkan empat karakteristik yang membedakan arsitektur data warehouse
3. Jelaskan apa perbedaan operasional dengan data warehouse menurut data warehouse sebagai subject oriented.
4. Berikan contoh data yang telah dilakukan integrasi

BAB IV

KOMPONEN DATA WAREHOUSE

A. Tujuan Instruksional Umum

Mahasiswa diharapkan dapat memahami komponen Data Warehouse.

B. Tujuan Instruksional Khusus

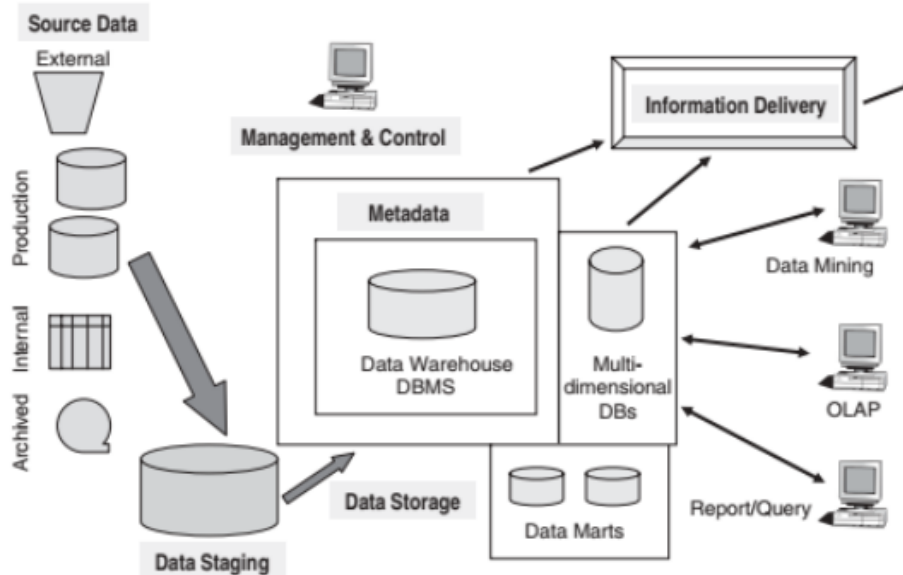
Mahasiswa diharapkan mampu :

1. Memahami komponen sumber data
2. Memahami komponen data staging
3. Memahami komponen data storage
4. Menjelaskan komponen information delivery
5. Memahami komponen metadata
6. Memahami komponen kontrol dan manajemen

Ketika membangun sebuah sistem operasional perlu memasukkan beberapa komponen untuk membentuk sebuah sistem. Komponen front-end terdiri dari GUI (Graphical User Interface) untuk menghubungkan dengan pengguna bagi proses input data. Komponen penyimpanan data memasukkan sistem manajemen database seperti Oracle, informix atau Microsoft SQL Server. Komponen display adalah tampilan di layar dan laporan bagi pengguna. Interface data dan perangkat lunak

jaringan membentuk komponen jaringan. Bergantung pada kebutuhan informasi dan framework organisasi kita menyusun komponen-komponen ini dengan seoptimal mungkin.

Pada Gambar ditunjukkan komponen dasar Data Warehouse. Pada gambar komponen source data ditunjukkan pada bagian sebelah kiri, kemudian pada bagian Building Block berikutnya komponen Data Staging. Di tengah, terdapat komponen Data Storage yang mengatur data Data Warehouse. Komponen ini tidak hanya menyimpan dan mengatur data, tapi juga menjaga jalur data dengan menggunakan tempat penyimpanan data. Komponen Information delivery ditampilkan pada sebelah kanan terdiri dari seluruh cara-cara yang berbeda untuk menjadikan informasi Data Warehouse tersedia bagi pengguna.



Gambar 8 Arsitektur Komponen Data Warehouse

C. Komponen Sumber Data

Komponen sumber data (Data Source) dibagi menjadi 4 kategori utama, yakni :

1. Data Produksi (Production Data)

Kategori data ini berasal dari beberapa sistem operasional enterprise. Berdasarkan kebutuhan informasi dalam Data Warehouse, anda memilih segmen data dari sistem operasional yang berbeda. Ketika berurusan dengan data ini, anda akan mendapatkan banyak variasi format data. Hal lain yang perlu diperhatikan adalah platform perangkat keras yang berbeda. Data juga didukung sistem operasi dan sistem database yang berbeda.

Pada sistem operasional, kueri informasi bersifat sempit. Anda melakukan kueri sistem operasional mengenai informasi mengenai instan spesifik dari suatu objek bisnis. Misalkan anda membutuhkan untuk hanya mengetahui nama dan alamat pelanggan tunggal, atau melihat sebuah invoice dan tagihan item-item pada invoice tersebut.

Dalam sistem operasional anda tidak melakukan kueri yang luas. Seluruh kueri dapat diprediksi, karena tidak membutuhkan kueri yang berjalan pada sistem operasional yang berbeda. Karakteristik yang pasti dan cukup merepotkan pada data produksi adalah perbedaan data. Tantangan anda adalah menstandarisasi dan mentransformasi data yang berbeda dari berjenis sistem produksi, mengkonversi data, dan mengintegrasikan bagian-bagiannya menjadi data yang berguna bagi penyimpanan dalam Data Warehouse.

2. Data Internal (Internal Data)

Dalam setiap organisasi, pengguna memiliki lembar kerja yang bersifat “rahasia” begitu juga dokumen, profil pelanggan dan juga database per departemen. Ini adalah data internal, bagian yang boleh jadi berguna dalam Data Warehouse. Jika organisasi melakukan bisnis ke pelanggan yang bersifat satu-ke-satu dan kontribusi masing-masing pelanggan bersifat berarti. Profil pelanggan individual menjadi sangat penting untuk dipertimbangkan.

Anda tidak dapat mengabaikan data internal yang disimpan dalam file rahasia dalam organisasi anda. Departemen IT harus bekerja dengan departemen user untuk mengumpulkan data internal.

Internal Data menambah kompleksitas ke dalam proses transformasi dan integrasi data sebelum disimpan ke dalam Data Warehouse. Anda harus menentukan strategi pengumpulan data dari lembar kerja, menemukan cara mengambil data dari dokumen tekstual dan mengaitkannya menjadi database per departemen untuk memperoleh data yang berhubungan dari sumber-sumber tersebut.

3. Data Arsip (Archived Data)

Sistem operasional ditujukan untuk menjalankan bisnis saat ini. Dalam setiap sistem operasional, anda secara periodik mengambil data lama dan menyimpannya dalam file arsip. Sikon organisasi yang menentukan seberapa sering dan seberapa banyak data operasional yang diarsipkan untuk disimpan. Beberapa data diarsipkan setelah satu tahun. Terkadang data ditinggalkan dalam sistem database operasional selama 5 tahun.

Ada banyak metode pengarsipan yang berbeda yang ada hingga saat ini. Ada metode pengarsipan bertingkat, pada tingkat pertama data yang paling akhir diarsipkan ke arsip database yang terpisah yang mungkin masih online. Pada tingkat kedua, data yang lebih lama diarsipkan ke bentuk flat file pada media penyimpanan berupa disk. Pada tahapan berikutnya data yang paling lama diarsipkan ke media penyimpanan data atau microfilm.

Organisasi pasti membutuhkan data historis untuk analisis setiap waktu. Untuk memperoleh informasi historis, dapat diperoleh dari data set yang diarsipkan. Tergantung pada kebutuhan Data Warehouse, organisasi harus memasukkan data historis yang mencukupi. Tipe data ini berguna dalam melihat pattern atau analisis trend.

4. Data Eksternal (External Data)

Kebanyakan eksekutif bergantung pada data dari sumber eksternal untuk persentase informasi tinggi yang dibutuhkan. Mereka menggunakan statistik untuk industri yang dihasilkan agen eksternal. Mereka juga menggunakan share data pesaing. Dan juga indikator keuangan bagi bisnis mereka untuk mengukur kinerja mereka.

Contoh Data Warehouse rental mobil berisikan data jadwal produksi saat ini dari perusahaan automobil ternama. Data Eksternal di dalam Data Warehouse membantu perusahaan rental mobil merencanakan manajemen armada mereka. Tujuan yang dilayani sumber data eksternal tidak dapat dipenuhi ketersediaan data didalam organisasi. Umumnya data yang bersumber dari luar tidak sesuai dengan format yang ada di organisasi. Karena itu dibutuhkan konversi data menjadi format

internal dan juga tipe datanya. Anda harus melakukan pengorganisasian transmisi data dari sumber eksternal. Beberapa sumber menyediakan informasi reguler maupun dengan interval.

D. Komponen Data Staging

Setelah mengekstrak data dari beberapa sistem operasional dan sumber eksternal, kemudian data disiapkan untuk disimpan dalam Data Warehouse. Data yang diekstrak berasal dari beberapa sumber yang terpisah yang perlu diubah, dikonversikan dan dibuat siap menjadi sebuah format yang sesuai untuk pengkuerian dan analisis.

Tiga fungsi utama yang perlu dijalankan untuk menjadi data siap yakni mengekstrak data, mentransformasi data dan me-load data ke dalam media penyimpanan Data Warehouse. Tiga fungsi utama ini yakni extraction, transformation dan persiapan loading data dilakukan dalam staging area (wilayah yang berbeda-beda). Data Staging area menyediakan sebuah ruang dan area dengan seperangkat fungsi untuk membersihkan, merubah, mengkombinasikan, mengkonversi, menduplikasi dan mempersiapkan source data untuk penyimpanan dan penggunaan di dalam Data Warehouse.

1. Data Extraction

Fungsi ini berhubungan dengan sejumlah Data Sources. Anda harus menggunakan teknik yang cocok untuk masing-masing data source. Data Source bisa jadi berasal dari sumber-sumber mesin yang berbeda-beda dengan format yang beragam. Beberapa bagian data bisa jadi berasal dari DBMS atau dari data hierarkis dan data jaringan atau

mungkin beberapa flat file, juga mungkin dimasukkan data dari spreadsheet ataupun data set departemen. Data Extraction dalam hal ini terbilang cukup kompleks.

Tool-tool tersedia di pasaran untuk proses data extraction. Tim Implementasi Data Warehouse mengekstraksi Data Source ke dalam lingkungan fisik yang terpisah proses menjadikan data ke dalam Data Warehouse menjadi lebih mudah. Dalam sebuah lingkungan yang berbeda, anda perlu mengekstrak source data menjadi sekelompok file-file flat, atau data staging relational database atau kombinasi dari keduanya.

2. Data Transformation

Dalam setiap implementasi sistem, konversi data adalah sebuah fungsi penting. Sebagai contoh ketika ingin mengimplementasikan sistem operasional seperti aplikasi majalah, pertama yang harus dilakukan adalah pemopulasian database anda dengan data dari rekod sistem yang terlebih dahulu. Yang boleh jadi dikonversikan dari sistem manual, atau sistem berorientasi file menjadi sistem modern dengan dukungan tabel relational database.

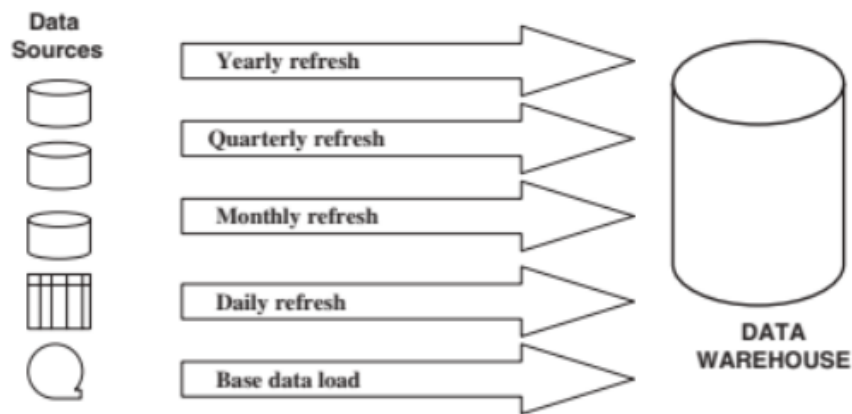
Anda juga dapat melakukan sejumlah aktivitas sebagai bagian transformasi data. Yang pertama, membersihkan data yang diekstraksi dari masing-masing sumber. Pembersihan disini dapat berupa koreksi kesalahan penulisan atau bisa jadi memasukkan pemecahan masalah antara kode wilayah dan kode pos dalam source data atau boleh jadi juga berhubungan dengan penyediaan nilai default untuk elemen-elemen data yang hilang atau menghilangkan duplikat-duplikat ketika membawa didalam data yang sama yang berasal dari source system yang berbeda. Data transformasi melibatkan banyak bentuk kombinasi data dari sumber yang berbeda-beda. Anda mengkombinasikan data dari sebuah sumber record tunggal

atau elemen-elemen data terkait yang berasal dari banyak record. Disisi lain transformasi data juga melibatkan pembersihan source data yang tidak berguna dan pemisahan rekod sumber luar ke dalam kombinasi baru. Pengurutan dan penggabungan data terjadi dalam skala luas dalam Data Staging Area.

Sistem operasional rantai grosir POS menjaga jumlah unit penjualan dan pendapatan yang dilakukan dalam transaksi tunggal pada konter masing-masing toko. Namun di dalam Data Warehouse tidak dibutuhkan menjaga data pada tingkat yang detil ini. Anda mungkin meringkas total produk pada masing-masing toko untuk hari tertentu dan mempertahankan jumlah total unit penjualan serta pendapatan dalam media penyimpana Data Warehouse. Dalam hal ini maka fungsi data transformation termasuk ringkasan yang sesuai. Setelah fungsi Data Transformation berakhir, anda memiliki sekumpulan data terintegrasi yang bersih, standard dan ringkas. Maka Data siap untuk di load ke dalam masing-masing Data Set dalam Data Warehouse.

3. Data Loading

Dua kelompok kegiatan membentuk fungsi Data Loading. Setelah melengkapi desain dan konstruksi Data Warehouse dan berjalan untuk pertama kalinya, perlu dilakukan initial loading Data Warehouse ke dalam media penyimpanan Data Warehouse. Load permulaan memindahkan data bervolume besar menggunakan jumlah waktu yang substansi²³ Pada Gambar mengilustrasikan tipe-tipe umum perpindahan data dari staging area menjadi media penyimpanan Data Warehouse.



23

Gambar 1 Perpindahan Data Dalam Data Warehouse

E. Komponen Data Storage

Penyimpanan **Data** untuk Data Warehouse adalah tempat penyimpanan yang berbeda. Penyimpanan data untuk sistem operasional hanya berisikan data saat ini. Penyimpanan data-data ini berisikan struktur yang benar-benar dalam format normal untuk pemrosesan yang cepat dan efisien. Sebaliknya, dalam penyimpanan data Data Warehouse, yang harus dilakukan adalah menjaga data dengan volume besar untuk analisis. Lebih jauh lagi data di dalam Data Warehouse harus dijaga dalam struktur yang sesuai untuk analisis dan tidak untuk penerimaan secara cepat seperangkat informasi yang dibutuhkan. Oleh karenanya, media penyimpanan untuk Data Warehouse dipisahkan dari media penyimpanan untuk sistem operasional.

Di dalam database yang mendukung sistem operasional, perbaikan terhadap data terjadi seiring terjadinya transaksi, ketika analis menggunakan data dalam Data Warehouse untuk analisis, yang harus diketahuinya adalah data bersifat stabil dan menggambarkan snapshot pada periode tertentu.

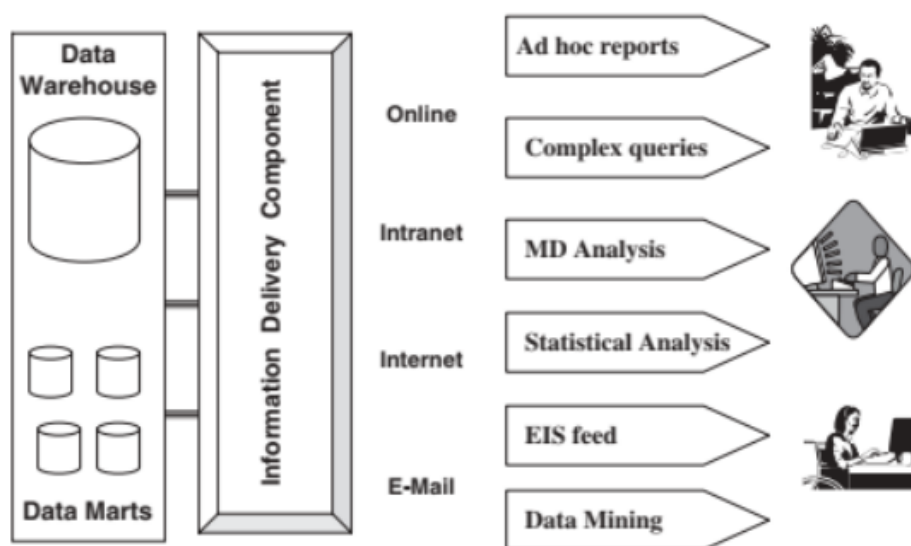
Kebanyakan Data Warehouse juga menggunakan DBMS. Data diekstrak dari media penyimpanan Data Warehouse dijumlahkan

dengan banyak macam cara dan ringkasan datanya dijaga dalam Multidimensional databases (MDDDB).

F. Komponen Information Delivery

Pengguna awam membutuhkan Data Warehouse tanpa training dan oleh karenanya membutuhkan laporan dan kueri, pengguna yang secara kebetulan hanya membutuhkan informasi satu kali tidak secara teratur. Pengguna tipe ini juga membutuhkan informasi yang dipaketkan. Analis bisnis membutuhkan kemampuan untuk melakukan analisis kompleks menggunakan informasi dalam Data Warehouse. Power user ingin dapat bernavigasi dengan Data Warehouse memperoleh data yang menarik, membentuk kuerinya, menembus data layer dan membuat kustomisasi laporan serta kueri ad hoc.

Agar dapat menyediakan informasi kepada komunitas pengguna data Warehouse komponen information delivery memiliki beberapa metode information delivery. Pada Gambar 9 menyajikan beberapa metode information delivery yang berbeda. Ad hoc report adalah laporan yang berarti bagi pengguna awam dan umum. Ketentuan kueri yang kompleks, analisis multidimensional (MD) dan analisis statistik untuk memenuhi kebutuhan analis bisnis dan power user. Informasi yang memenuhi kebutuhan Executive Information System (EIS) untuk eksekutif senior dan manajer level atas. Beberapa Data Warehouse juga menyediakan data untuk aplikasi data-mining. Aplikasi Data-mining adalah knowledge discovery system dimana algoritma mining membantu anda untuk mengetahui tren dan pattern dari penggunaan data anda.



Gambar 2 Komponen Information Delivery

G. Komponen Metadata

Metadata dalam sebuah Data Warehouse mirip dengan kamus data atau katalog data dalam sebuah DBMS. Dalam kamus data, informasi seperti struktur data dijaga, informasi mengenai file dan alamatnya, informasi mengenai indeks dan lain sebagainya. Kamus data berisikan data tentang data di dalam database. Komponen Metadata adalah data mengenai data di dalam Data Warehouse. Metadata di dalam Data Warehouse mirip dengan sebuah kamus data, namun lebih dari sekedar kamus data.

H. Komponen Kontrol dan Manajemen

Komponen ini berada pada posisi yang paling atas dalam arsitektur Data Warehouse. Komponen Kontrol dan manajemen mengkoordinasikan aktivitas dan layanan dalam data Warehouse. Komponen ini mengontrol transformasi data dan data ditransfer kedalam media penyimpanan Data Warehouse. Disisi lain, mencukupi untuk information delivery ke pengguna, bekerja

dengan DBMS dan memungkinkan data disimpan secara benar dalam tempat penyimpanannya. Juga memonitor pergerakan data ke dalam staging area dan dari situ ke dalam media penyimpanan Data Warehouse itu sendiri.

I. Evaluasi

1. Sebutkan setidaknya enam karakteristik atau fitur data warehouse.
2. Mengapa integrasi data diperlukan dalam data warehouse, lebih dari pada operasional aplikasi?
3. Setiap struktur data di data warehouse berisi elemen waktu. Mengapa?
4. Sebutkan sumber data untuk data warehouse
5. Sebutkan enam metode berbeda untuk pengiriman informasi.

BAB V
DUKUNGAN INFRASTRUKTUR
DATA WAREHOUSE

A. Tujuan Instruksional Umum

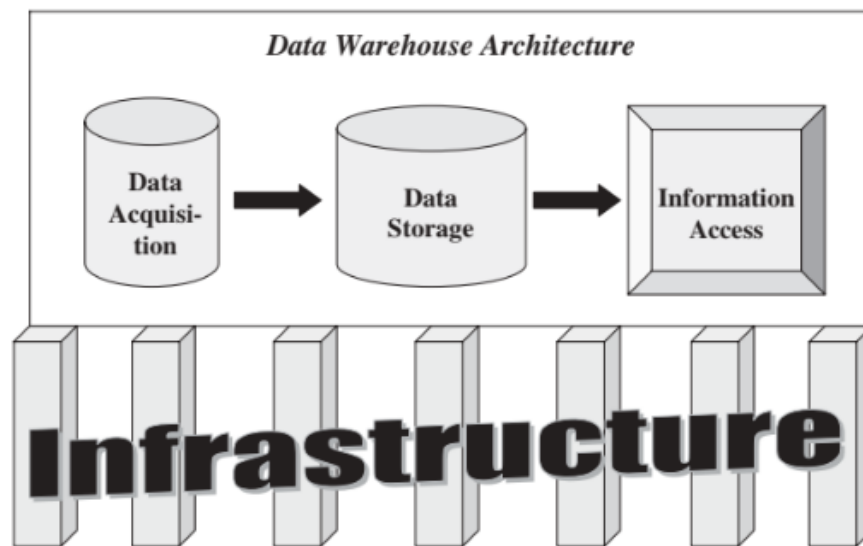
Mahasiswa diharapkan dapat memahami dukungan infrastruktur Data Warehouse.

B. Tujuan Instruksional Khusus

Mahasiswa diharapkan mampu :

1. Memahami infrastruktur operasional
2. Memahami infrastruktur fisik data warehouse
3. Memahami sistem operasi dan perangkat keras data warehouse
4. Menjelaskan urgensi dari data warehouse

Infrastruktur Data Warehouse memasukkan semua elemen dasar yang memungkinkan arsitektur untuk diimplementasikan. Sehingga infrastruktur memasukkan beberapa elemen seperti perangkat keras server, sistem operasi, perangkat lunak jaringan, perangkat lunak database, LAN dan WAN, tool vendor untuk setiap komponen arsitektur, orang, prosedur dan pelatihan. Pada Gambar berikut menjelaskan betapa pentingnya dukungan infrastruktur dalam data warehouse.



Gambar 3 Dukungan Infrastruktur Data Warehouse

Elemen-elemen infrastruktur Data Warehouse dapat dikelompokkan menjadi dua kategori yakni : infrastruktur operasional dan infrastruktur fisik. Perbedaan ini penting karena elemen-elemen pada masing-masing kategori memiliki fitur yang berbeda jika dibandingkan dengan kategori lainnya.

C. Infrastruktur Operasional

Salah satu komponen infrastruktur merujuk pada perangkat ³⁶ perangkat keras komputer dan perangkat lunak terkait. Anda membutuhkan ³¹ perangkat keras dan perangkat lunak untuk menjalankan fungsi data staging dan memberikan servis yang sesuai. Anda juga membutuhkan tool perangkat lunak untuk menjalankan transformasi data, anda membutuhkan perangkat lunak untuk membuat file output, anda membutuhkan perangkat keras disk untuk menempatkan data dalam file staging area.

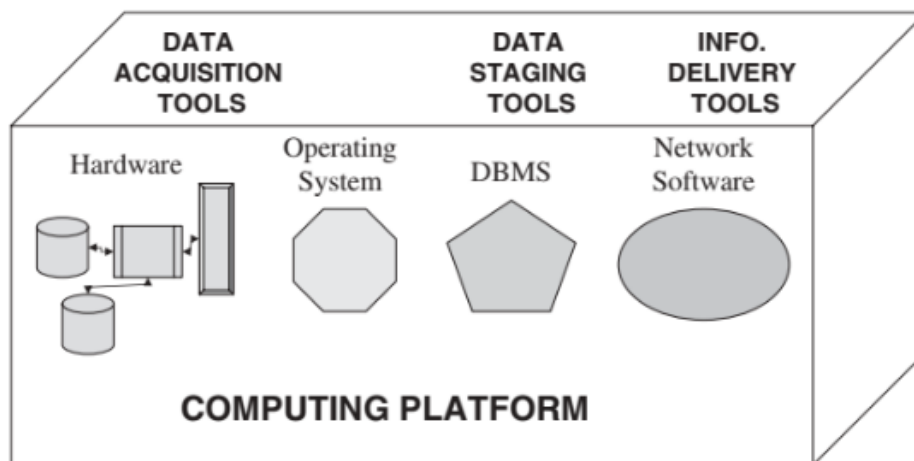
Infrastruktur Operasional dalam mendukung masing-masing komponen arsitektur terdiri dari :

- ✚ Orang
- ✚ Prosedur
- ✚ Pelatihan
- ✚ Perangkat lunak manajemen

Orang dan prosedur disini bukan untuk pengembangan Data Warehouse, namun agar Data Warehouse dapat tetap berjalan, elemen-elemen ini sama pentingnya dengan perangkat lunak dan perangkat keras yang menjaga agar Data Warehouse tetap berjalan. Mereka mendukung manajemen Data Warehouse dan memelihara efisiensinya.

D. Infrastruktur Fisik

Pada Gambar menampilkan elemen-elemen utama infrastruktur fisik. Setiap sistem termasuk Data Warehouse harus memiliki keseluruhan platform dimana akan ditempatkan. Secara esensial, platform terdiri dari komponen perangkat keras dasar, sistem operasi disertai dengan software utiliti, jaringan dan software jaringan.



Gambar 4 Infrastruktur Fisik

Keputusan menentukan perangkat keras tidak mudah. Harus mempertimbangkan banyak faktor dan memastikan bahwa perangkat keras yang dipilih akan mendukung seluruh arsitektur data warehouse.

52

E. Sistem Operasi dan Perangkat Keras

Sistem operasi dan perangkat keras menciptakan lingkungan komputer untuk operasional Data Warehouse. Seluruh kegiatan ekstraksi data, transformasi, integrasi dan staging berjalan pada perangkat keras terpilih dengan sistem operasinya. Ketika mentransportasikan data yang telah dikonsolidasikan dan diintegrasikan dari staging area ke tempat penyimpanan Data Warehouse, maka menggunakan perangkat lunak sistem operasi dan perangkat keras server.

Berikut adalah petunjuk pemilihan perangkat keras, yang tidak secara keseluruhan spesifik pada perangkat keras bagi Data Warehouse.

- ✚ Scalability : Ketika Data Warehouse memiliki makin banyak pengguna, maka dapat dipastikan perangkat keras pilihan anda dapat menangani jumlah kueri dan kompleksitasnya yang semakin meningkat.
- ✚ Support : Dukungan vendor bersifat krusial bagi pemeliharaan perangkat keras. Pastikan bahwa dukungan dari vendor perangkat keras berada pada peringkat tertinggi yang paling mungkin.
- ✚ Vendor Reference : Hal yang cukup penting untuk dilakukan yakni pengecekan referensi vendor dengan lainnya menggunakan perangkat keras vendor tersebut. Harus adanya antisipasi terjadinya Data Warehouse down karena malfungsi perangkat keras ketika CEO menginginkan beberapa analisis kritis yang harus diselesaikan.

- ✚ Vendor Stability : Cek stabilitas dan waktu nyala.

Berikutnya kita dapat mempertimbangkan beberapa kriteria umum untuk pemilihan sistem operasi. Yang pertama, sistem operasi harus kompatibel dengan perangkat keras. Berikut merupakan kriteria.

- ✚ Scalability : Scalability berada pada daftar pertama karena merupakan salah satu fitur utama setiap Data Warehouse. Data Warehouse tumbuh sangat cepat. Begitu juga dengan perangkat lunak dan perangkat keras database, sistem operasi harus dapat mendukung peningkatan jumlah pengguna dan aplikasi.
- ✚ Security : Ketika banyak workstation klien mengakses server, sistem operasi dapat melindungi masing-masing klien dan sumber daya terkait. Sistem operasi harus menyediakan masing-masing klien sebuah lingkungan yang aman.
- ✚ Reliability : Sistem operasi harus dapat melindungi lingkungan Data Warehouse dari malfungsi aplikasi.
- ✚ Availability : merupakan akibat langsung dari reliability. Lingkungan komputerisasi harus tetap available setelah penghentian aplikasi abnormal.
- ✚ Preemptive Multitasking : Perangkat keras server harus mampu menyeimbangkan alokasi waktu dan sumber daya antara banyaknya aktivitas yang berjalan. Sistem operasi juga harus mampu menentukan aktivitas dengan prioritas yang lebih tinggi yang mendahului atau menginterupsi aktivitas lainnya ketika dibutuhkan.
- ✚ Use multithreaded approach : Sistem operasi harus dapat melayani banyak permintaan yang terjadi secara serentak dengan pendistribusian thread ke banyak prosesor dalam konfigurasi perangkat keras multiprosesor. Fitur ini sangat penting karena konfigurasi multiprosesor adalah arsitektur pilihan dalam sebuah lingkungan Data Warehouse.

- ✚ **Memory Protection** :Dalam lingkungan Data Warehouse, ada banyak kueri dalam jumlah besar, yang berarti bahwa banyak kueri yang akan dieksekusi secara serentak. Sebuah fitur proteksi memori dalam sistem operasi mencari mencegah sebuah aktivitas dari merambah ruang memori lainnya.

Pilihan Platform, Platform komputer adalah seperangkat komponen perangkat keras, sistem operasi dan jaringan serta perangkat lunak jaringan. Baik merupakan fungsi dari sebuah sistem OLTP atau sistem pengambilan keputusan seperti Data Warehouse, fungsinya harus dijalankan dalam sebuah platform komputer.

1. Pilihan Single Platform :adalah pilihan paling sederhana dan mudah. Pada pilihan ini, seluruh fungsi dari ekstraksi data back-end hingga pemrosesan kueri front-end dijalankan pada sebuah platform komputer tunggal. Ini adalah pendekatan paling awal yang dikembangkan para developer ketika mengimplementasikan Data Warehouse pada mainframe yang sedang eksis, minikomputer atau sebuah server tunggal berbasis UNIX.
2. Legacy Platform Stretched to Capacity :Pada banyak perusahaan, lingkungan komputerisasi yang eksis boleh jadi telah ada sejak beberapa dekade dan sudah memenuhi sesuai kapasitasnya. Lingkungan komputernya boleh jadi dimana ia tidak dapat lagi diupgrade lebih jauh untuk mengakomodasi Data Warehouse.
3. Nonavailability of Tools :Tool perangkat lunak mengambil porsi yang cukup besar dari infrastruktur Data Warehouse. Kebanyakan tool disediakan oleh sejumlah vendor Data Warehouse tidak mendukung lingkungan mainframe atau minikomputer. Tanpa tool yang cocok dalam infrastruktur maka Data Warehouse akan berantakan.

4. Multiple Legacy Platform, Walaupun kita menduga bahwa warisan komputer mainframe atau lingkungan minikomputer dapat diperbesar dengan memasukkan Data Warehouse, fakta praktisnya menunjuk pada sebuah situasi yang berbeda.

Hybrid Option, Setelah menganalisis warisan sistem dan aplikasi yang lebih modern dalam perusahaan, maka berikutnya dapat memutuskan bahwa pendekatan single-platform tidak dapat bekerja untuk Data Warehouse anda. Ini menjadi alasan banyak perusahaan tidak menjalankannya, disisi lain, jika perusahaan anda harus memilih kategori dimana platform warisan dapat mengakomodasi Data Warehouse maka pendekatan single platform menjadi solusi. Untuk lebih jelas, kita akan melihat tahapan data flow dan memeriksa pilihan platform.

- ✚ Data Extraction : Pada Data Warehouse yang terbaik adalah menjalankan fungsi ekstraksi data dari masing-masing sistem sumber pada platform komputer itu sendiri.
- ✚ Initial Reformatting and Merging : Setelah pembuatan data mentah yang diekstrak dari beberapa sumber, file-file yang diekstrak dari masing-masing sumber di format kembali dan digabungkan menjadi sejumlah file ekstraksi yang lebih kecil. Verifikasi Data ekstrak terhadap laporan source system dan rekonsiliasi jumlah rekod input maupun output ada pada tahapan ini. Seperti tahapan ekstraksi, yang terbaik adalah menjalankan tahapan awal penggabungan masing-masing perangkat sumber ekstrak pada platform sumber itu sendiri.
- ✚ Preliminary Data Cleansing. Pada tahapan ini, anda memverifikasi data yang telah diekstrak dari masing-masing source data untuk nilai data yang hilang pada field field tunggal, memberikan nilai default dan menjalankan edit

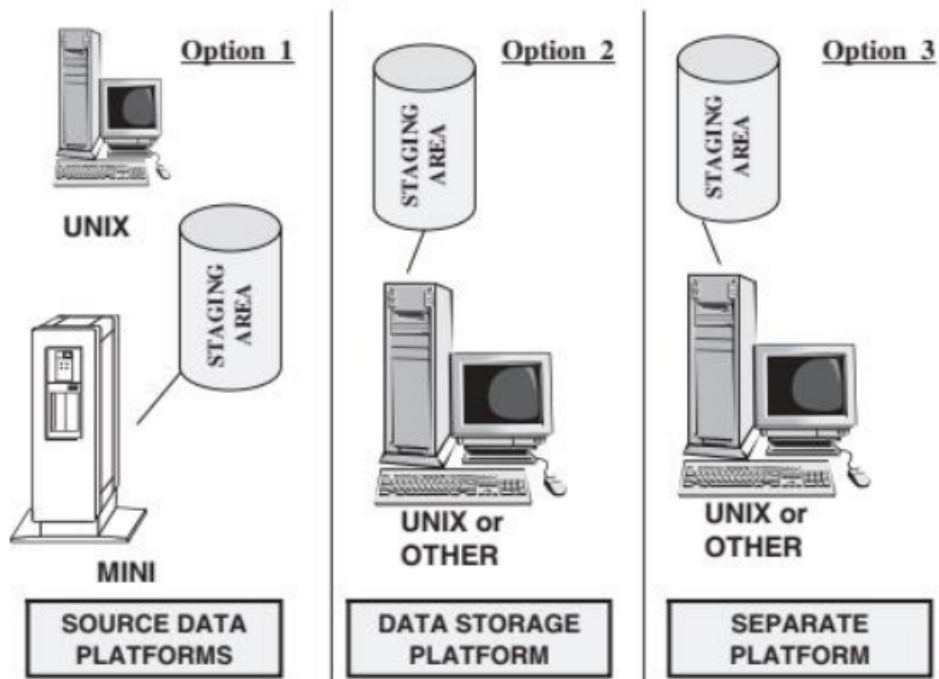
dasar. Ini adalah tahapan lainnya pada platform komputer dari source system itu sendiri. Bagaimanapun, dalam beberapa Data Warehouse tipe pembersihan data ini terjadi setelah data dari semua sumber direkonsiliasikan dan dikonsolidasikan. Pada masing-masing kasus, fitur-fitur dan kondisi-kondisi data dari source system anda mendiktekan kapan dan dimana tahapan ini harus dijalankan bagi Data Warehouse anda.

- ✚ Transformation and Consolidation : Tahapan ini berisikan seluruh transformasi data utama dan fungsi integrasi. Anda akan menggunakan tool perangkat lunak transformasi untuk tujuan ini.
- ✚ Validation and Final Quality Check : Tahapan validasi final dan pemeriksaan kualitas adalah kandidat kuat staging area. Anda akan menyusun tahapan ini agar terjadipada platform tersebut.
- ✚ Creation of Load Images : Tahapan ini menciptakan load images untuk file-file database tunggal dari media penyimpanan Data Warehouse. Tahapan ini hampir selalu terjadi dalam staging area dan oleh karenanya pada platform dimana staging area ditempatkan.

Option for Staging Area, tempat dimana seluruh data untuk Data Warehouse dikumpulkan dan disiapkan. Platform paling cocok untuk staging area tergantung status platform sumbernya, untuk lebih jelasnya mari kita eksplorasi pilihan penempatan staging area.

- ✚ Dalam salah satu Legacy Platform. Jika kebanyakan warisan Data sources ada pada platform yang sama dan jika kapasitas ekstra tersedia, maka pertimbangkan penjagaan area data staging anda dalam legacy platform. Untuk pilihan ini, anda

akan menghemat waktu dan tenaga dalam memindahkan data berbeda platform ke staging area.



Gambar 5 Platform options for the staging area

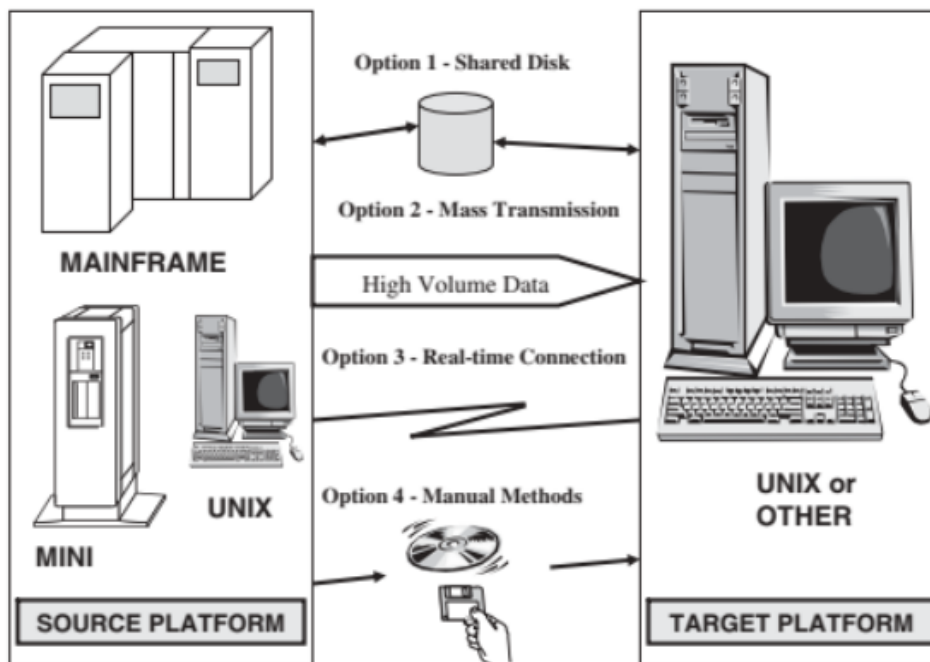
- ✚ Pada Platform Data Storage. Ini adalah platform dimana DBMS Data Warehouse berada dan berjalan.
- ✚ Pada Platform yang terpisah. Mungkin saja anda membutuhkan transformasi data kompleks. Hal yang mungkin saja terjadi dimana anda dapat bekerja dengan data anda untuk membersihkan dan mempersiapkannya bagi Data Warehouse anda. Dalam hal ini, anda membutuhkan sebuah platform terpisah untuk memilah-milah data sebelum diteruskan ke database.

Data Movement Consideration, Pada platform komputer apapun tahapan tunggal data acquisition dan data storage terjadi, data harus dipindahkan ke platform yang berbeda. Bergantung pada source platform perusahaan dan pilihan platform untuk data staging dan data storage, anda harus menyediakan transportasi data pada platform yang berbeda. Pada masing-masing pergerakan data lintas platform komputer, tentukan pilihan yang paling cocok untuk lingkungan tersebut.

- ✚ Shared Disk. Metode ini kembali ke zaman mainframe. Aplikasi berjalan pada wilayah dan partisi yang berbeda diizinkan untuk berbagi data dengan penempatan data pada suatu disk share. Anda dapat mengadaptasi metode ini dari satu tahap ke tahapan lainnya untuk data acquisition dalam Data Warehouse anda. Anda harus menentukan disk storage dan pengaturannya sehingga masing-masing dari kedua platform mengenali area disk storage nya sendiri.
- ✚ Mass Data Transmission. Dalam hal ini, transmisi data berbeda platform terjadi melalui port data. Port data adalah perangkat interplatform yang sederhana yang memungkinkan kuantitas data massal ditransportasikan dari satu platform ke yang lainnya. Masing-masing platform harus dikonfigurasi untuk menangani transfer melalui port. Pilihan ini membutuhkan komponen perangkat keras, perangkat lunak dan jaringan khusus. Juga dibutuhkan bandwidth jaringan yang cukup dalam pemrosesan data skala besar.
- ✚ Real-Time Connection. Dalam pilihan ini, dua platform menciptakan koneksi secara real time sehingga sebuah program dapat berjalan pada sebuah platform yang mampu menggunakan sumber daya platform lainnya. Sebuah program pada sebuah platform dapat melakukan penulisan terhadap disk storage lainnya. Kegiatan yang berjalan pada

sebuah platform dapat melakukan penjadwalan aktivitas dan event lainnya.

- ✚ Manual Methods, Pilihan ini sederhana dan mudah. Sebuah program pada sebuah platform melakukan penulisan terhadap sebuah media eksternal seperti tape dan disk. Program lainnya pada platform penerima membaca data dari media eksternal

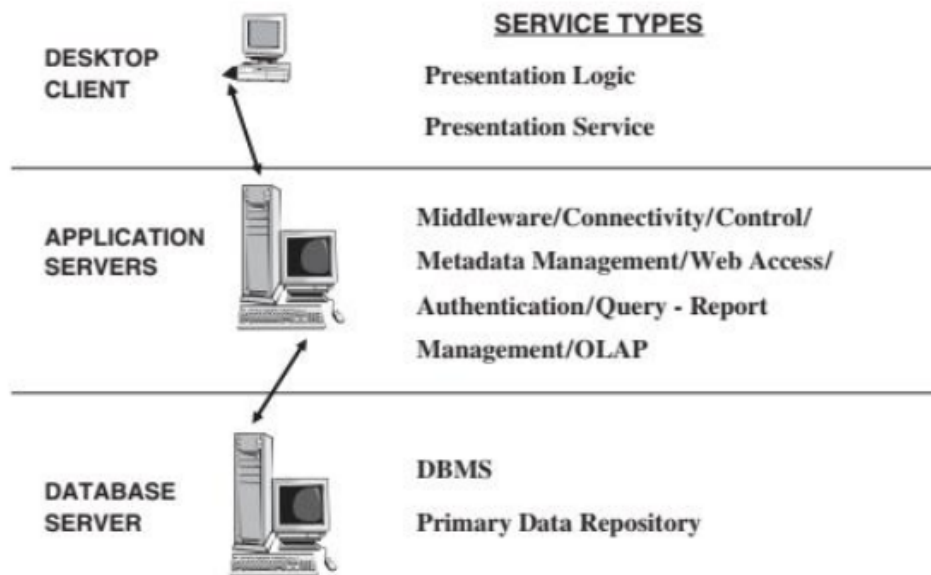


Gambar 6 Data Movement Consideration

Client-Server Architecture for the Data Warehouse.

Walaupun platform minikomputer dan mainframe lebih dahulu diimplementasikan untuk Data Warehouse, namun saat ini, ketika semakin besar, Data Warehouse dikembangkan dengan menggunakan arsitektur client-server. Kebanyakan dari ini adalah multitiered, arsitektur client/server generasi kedua. Gambar 4.4.

mengilustrasikan arsitektur client/server untuk implementasi Data Warehouse.



Gambar 7 Arsitektur Client/Server Data Warehouse

Considerations for Client Workstations. Ketika siap untuk mempertimbangkan konfigurasi mesin workstation, anda akan lebih cepat menyadari bahwa anda butuh untuk memisahkan tipe pengguna. Kita hanya mempertimbangkan kebutuhan workstation terkait pengiriman informasi dari Data Warehouse. Pengguna diharapkan puas dengan sebuah mesin yang dapat berjalan pada sebuah browser web untuk mengakses laporan HTML. Analisis serius, disisi lain, membutuhkan mesin workstation yang lebih besar dan powerful. Tipe pengguna lainnya antara dua keekstriman ini membutuhkan sebuah model layanan. Gunakan ceklis dibawah ini ketika mempertimbangkan workstation :

- ✚ Sistem operasi Workstation
- ✚ Processing power
- ✚ Memori
- ✚ Disk Storage
- ✚ Transportasi Data dan jaringan
- ✚ Dukungan tool

Options as The Data Warehouse Matures. Anda mungkin dapat menyimpulkan bahwa pilihan platform telah benar dan pilihan awal ditentukan. Hal yang menarik untuk dicatat bahwa Data Warehouse pada masing-masing enterprise yang telah mapan, maka susunan platformnya juga berevolusi. Data Staging dan Data Storage dapat berada pada platform komputer yang sama. Seiring waktu berjalan dan lebih banyak pengguna mulai bergantung pada Data Warehouse anda untuk pengambilan keputusan strategis, anda akan menemukan bahwa pilihan platform dapat dievolusikan.

F. Evaluasi

1. Apa komposisi infrastruktur operasional dari data warehouse?
2. Kenapa infrastruktur operasional sama pentingnya dengan infrastruktur fisik?
3. Sebutkan komponen utama infrastruktur fisik. Tulis dua atau tiga kalimat jelaskan setiap komponen.
4. Jelaskan secara singkat enam kriteria yang akan Anda gunakan untuk memilih sistem operasi untuk data warehouse.

BAB VI

PEMODELAN DATA MULTIDIMENSI

A. Tujuan Instruksional Umum

Mahasiswa diharapkan dapat memahami pemodelan data multidimensi Data Warehouse.

B. Tujuan Instruksional Khusus

Mahasiswa diharapkan mampu :

1. Memahami dimensional modeling
2. Memahami tabel dimensional
3. Memahami konsep dasar fact
4. Menjelaskan tabel fact
5. Memahami pemodelan dimensional
6. Memahami hirarki dalam dimensi
7. Mampu mengimplementasikan Hierarki Dimensional
8. Memahami Penggunaan Hierarki Dimensional
9. Memahami operasi OLAP

C. Dimensional Modeling

Istilah atau nama Dimensional Modeling berasal dari dimensi bisnis yang kita perlu tuangkan ke dalam model data logical. Pemodelan ini adalah teknik desain untuk menstrukturkan dimensi bisnis dan matrik yang dianalisis bersama dengan dimensi-dimensi . Model juga memberikan kinerja tinggi untuk analisis dan kueri.

Bisnis bagi manajer adalah terminologi dimensi bisnis. Contoh seorang pimpinan marketing tertarik mengetahui jumlah pendapat bulanan, di dalam sebuah divisi, berdasarkan demografi pelanggan yang dilakukan sales untuk jenis produk tertentu di masa yang sudah terjadi. Dalam hal ini dimensi bisnis adalah bulan, divisi, demografi, sales, dan jenis produk. Pendapatan adalah fact yang ingin diketahui pimpinan marketing.

Dimensional Modeling adalah teknik pemodelan favorit dalam Data Warehouse. Di dalam Dimensional Modeling, sebuah model tabel dan relasi digunakan untuk tujuan pengoptimalan kinerja kueri pengambilan keputusan dalam database relasional, relatif terhadap pengukuran atau seperangkat pengukuran proses bisnis yang akan dimodel. Sebaliknya model E-R konvensional digunakan untuk (1) menghilangkan pengulangan (redundan) dalam model data, (2) memfasilitasi perolehan rekod individual yang memiliki sejumlah identifier tertentu dan (3) mengoptimalkan kinerja OLTP.

Pada sebuah toko ritel, pengukuran yang penting atau fact adalah unit penjualan. Dimensi bisnis bisa jadi waktu, promosi, produk atau toko. Untuk suatu perusahaan asuransi, pengukurannya atau fact adalah klaim dan dimensi bisnisnya agen, kebijakan, pihak yang diasuransikan, status dan waktu.

Praktisi Dimensional Modeling menggunakan pendekatan model data logik dengan pemilihan proses bisnis yang akan dimodel dan kemudian memutuskan masing-masing rekod dalam “fact table”. Fact table adalah fokus analisis dimensional, adalah tabel dengan kueri berdimensi untuk memberikan seperangkat solusi. Kriteria untuk segmentasi ada di dalam satu atau lebih “tabel dimensi” dimana primary key tunggalnya menjadi foreign key fact table yang berelasi dalam desain dimensional model.

D. Tabel Dimensional

Ketika suatu dimensi bisnis diabstraksikan dan direpresentasikan dalam sebuah tabel database, ia dikenal dengan tabel dimensional. Sebuah dimensi dapat dipandang sebuah entitas. Sebuah tabel dimensional menyediakan deskripsi tekstual dari sebuah dimensi bisnis melalui atribut-atributnya.

Tabel dimensional cenderung secara relatif dangkal dalam hal jumlah baris-barisnya, namun dilengkapi banyak kolom-kolom. Sebuah tabel dimensional memiliki sebuah primary key tunggal dan telah dinormalisasi.

Atribut tabel Dimensional memainkan sebuah peran vital dalam pemrosesan kueri dan pelabelan laporan. Kekuatan sebuah Data Warehouse secara langsung proporsional berhubungan dengan kualitas dan kedalaman atribut-atribut dimensinya.

Tabel Dimensi Produk
Id_Produk (PK)
Deskripsi_Produk
Deskripsi_Merk
Deskripsi_Kategori
Deskripsi_Departemen

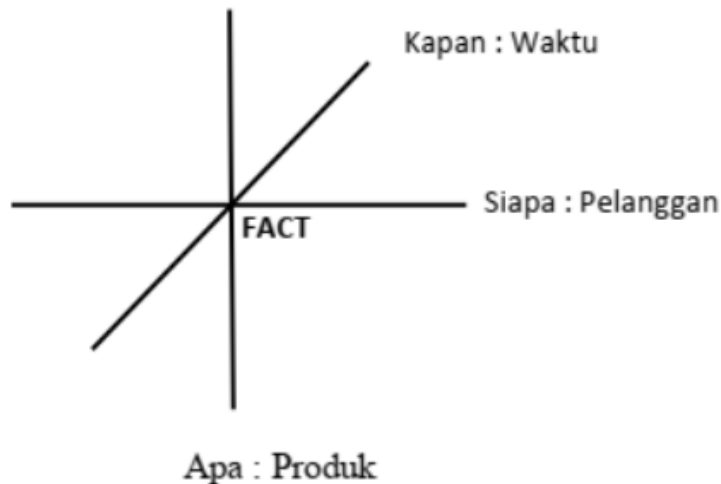
Deskripsi_Tipe
Deskripsi_Ukuran
Deskripsi_isi
Berat
Saturan_ukuran
Tipe Penyimpanan

E. Fact

Fact adalah pengukuran yang diperoleh dari kejadian yang ada dipasaran, merupakan sumber pengetahuan dan observasi. Seorang pelanggan membeli sebuah produk pada suatu lokasi tertentu pada waktu tertentu. Persilangan dari keempat dimensi yang terjadi ini menciptakan sebuah penjualan. Penjualan dapat dideskripsikan sejumlah uang yang diterima, jumlah item terjual, berat produk yang dikirimkan dan lain-lain. Fact digunakan untuk merepresentasikan sebuah ukuran bisnis.

Dalam sebuah Data Warehouse, fact didefinisikan sebuah persilangan dari dimensi-dimensi yang menggunakan entitas-entitas dasar transaksi bisnis. Untuk menunjukkan persilangan lebih dari tiga dimensi dalam sebuah diagram bukan perkara mudah, namun fact dalam sebuah Data Warehouse boleh jadi berasal dari banyak dimensi.

FACT dalam sebuah Data Warehouse



F. Tabel Fact

Sebuah tabel adalah tabel primer di dalam model dimensional dimana pengukuran kinerja secara numerik dari bisnis disimpan. Ada banyak pengukuran kinerja atau fact dalam sebuah tabel fact. Sebuah baris dalam sebuah tabel fact berhubungan dengan pengukuran. Fact yang paling berguna dalam sebuah tabel fact adalah numerik dan tambahan.

Semua tabel-tabel fact memiliki dua atau lebih foreign key yang berhubungan dengan primary key tabel dimensi. Ketika semua key di dalam tabel fact cocok dan berpasangan dengan primary key-nya secara benar dengan tabel dimensi maka dapat dikatakan tabel tersebut memenuhi referential integrity. Tabel fact diakses melalui tabel dimensi yang diikuti sertakannya.

Tabel fact sendiri secara umum memiliki primary key yang disusun seperangkat foreign key. Key ini dikenal dengan sebuah key composite atau concatenated (berentetan). Setiap tabel fact dalam sebuah model dimensional memiliki sebuah key

composite, dan sebaliknya setiap tabel yang memiliki sebuah key composite adalah sebuah tabel fact.

Cara lain untuk mengungkap hal ini adalah di dalam sebuah dimensional model, setiap tabel yang mengekspresikan relasi banyak-ke-banyak sudah pasti sebuah tabel fact. Semua tabel lainnya adalah tabel-tabel dimensi.

Tabel Fact Penjualan Harian

Key Tanggal(FK)

Key Produk(FK)

Key Toko(FK)

Jumlah terjual

Total Harga

10

G. Pemodelan Dimensional

Menurut (Connolly & Begg, 2005:1183), dimensionality modelling ialah teknik logical design yang bertujuan untuk menyajikan data dalam bentuk yang standar dan intuitif yang memungkinkan pengaksesan database dengan performance yang tinggi.

17

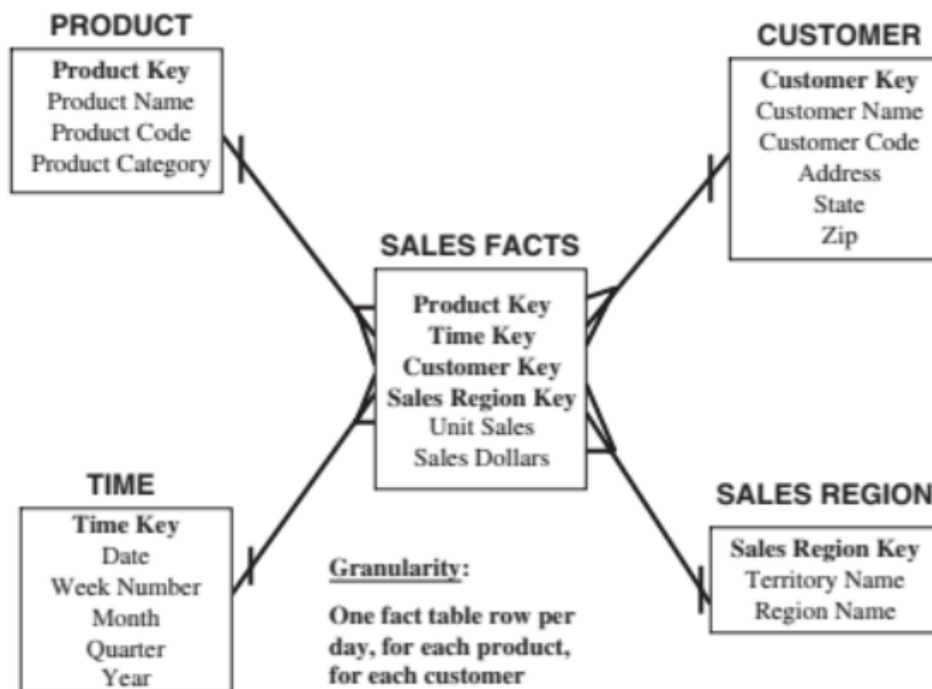
Setiap model dimensional terdiri dari sebuah tabel dengan sebuah primary key composite yang disebut dengan tabel fakta, dan sekumpulan tabel yang lebih kecil disebut dengan tabel dimensi. Setiap tabel dimensi memiliki sebuah primary key (nonkomposit) sederhana yang berkorespondensi tepat dengan satu key komposit pada tabel fakta. Dengan kata lain, primary key dari tabel fakta terbuat dari dua atau lebih foreign key.

1. Star Schema

1

Suatu model yang menggabungkan dimensi dan fact bersamaan dikenal dengan dimensional model. Di dalam model ini, tabel fact terdiri dari pengukuran numerik yang disertakan ke seperangkat tabel dimensi yang dipenuhi atribut deksriptif.

Di dalam model ini, Tabel fact berada di tengah dan tabel dimensi bergelantungan disekitarnya seperti sebuah bintang. Struktur dengan karakteristik demikian seringkali dikenal dengan star schema. Saat id pelanggan, id produk dan periode waktu digunakan untuk menentukan barisbaris manakah yang dipilih dari tabel fact, cara pengumpulan data demikian dikenal dengan istilah star schema join.



Gambar 8 Star Schema

Dimensional model sederhana dan simetris, datanya lebih mudah difahami dan dinavigasikan. Setiap dimensi ekuivalen; semua dimensi memiliki entri poin yang sama secara simetris ke dalam tabel fact. Kesederhanaan juga menguntungkan kinerja, sedikit join penting bagi pemrosesan kueri. Sebuah mesin

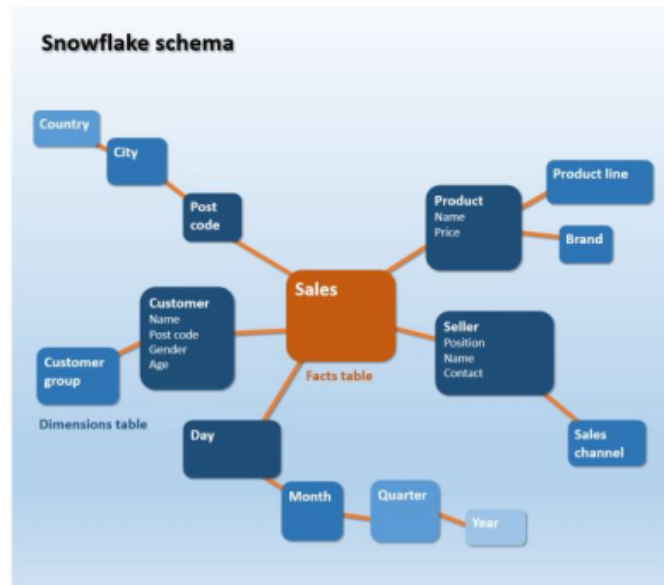
database mampu menciptakan asumsi yang kuat mengenai tabel dimensi dengan indeksinya.

Dengan dimensional model juga dapat ditambahkan secara lengkap dimensi baru ke skema bersamaan dengan nilai tunggal dimensi tersebut didefinisikan untuk masing-masing baris fact yang ada.

Cara lain memahami memahami kesederhaaan star schema adalah dengan melihat bagaimana dimensi dan fact berkontribusi terhadap laporan. Atribut tabel dimensi menawarkan pelabelan laporan, sebagaimana tabel fact menawarkan nilai numerik laporan.

2. Snowflake Schema

Snowflake schema adalah gabungan dari beberapa pemodelan data star schema yang dinormalisasi. Dalam snowflake schema, setiap tabel dimensi dapat memiliki sub-tabel dimensi lagi. Hal ini bertujuan untuk meminimalkan data yang berlebihan (redundancy data). Dimensi data inilah yang menjadi subjek informasi untuk menjadi bahan dalam pengambilan keputusan, karena pada setiap dimensi data dimungkinkan untuk dilakukan pemecahan lebih detail lagi. Dengan demikian sumber data yang bisa diolah untuk menjadi informasi bisa menjadi lebih banyak dan detail. Dibawah ini adalah gambar Snowflake Schema ditunjukkan pada Gambar 16.



Gambar 9 Snowflake Schema

Adapun kelebihan dari penggunaan snowflake schema antara lain sebagai berikut.

- + Ukuran data lebih kecil di dalam tempat penyimpanan
- + Lebih mudah dilakukan maintenance dan update
- + Proses query lebih cepat

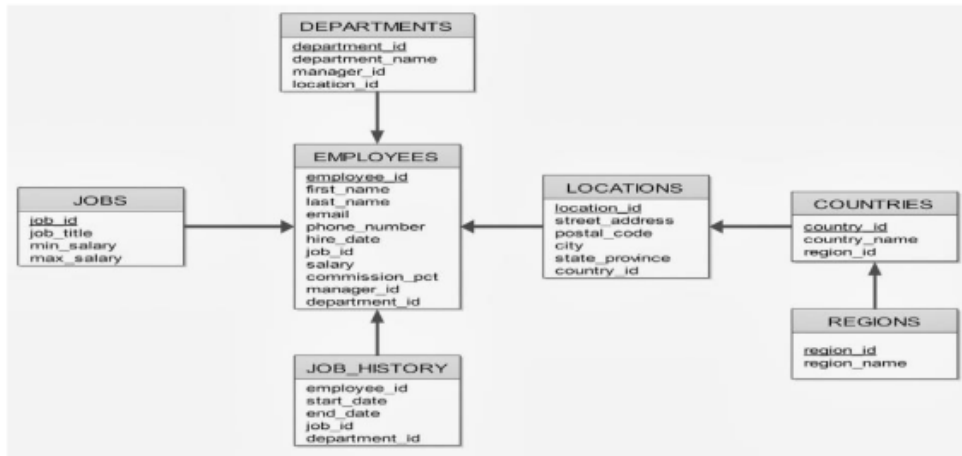
Selain kelebihan dari snowflake schema diatas, Snowflake Schema juga mempunyai kekurangan sebagai berikut.

- + Cenderung lebih sulit dipahami karena kompleksitasnya
- + Sulit mencari isi karena melihat strukturnya yang kompleks dan bercabang-cabang

3. Starflake Schema

Starflake schema adalah model data dimensional yang memiliki sebuah tabel fakta sebagai pusatnya, dikelilingi oleh tabel-tabel dimensi yang ternormalisasi dan terdenormalisasi. Beberapa pemodelan dimensional menggunakan campuran

dari star schema yang terdenormalisasi dan snowflake schema yang ternormalisasi. Sehingga starflake schema merupakan perpaduan dari star schema dan snowflake schema seperti pada Gambar berikut.



Gambar 10 Starflake Schema

H. Hirarki Dalam Dimensi

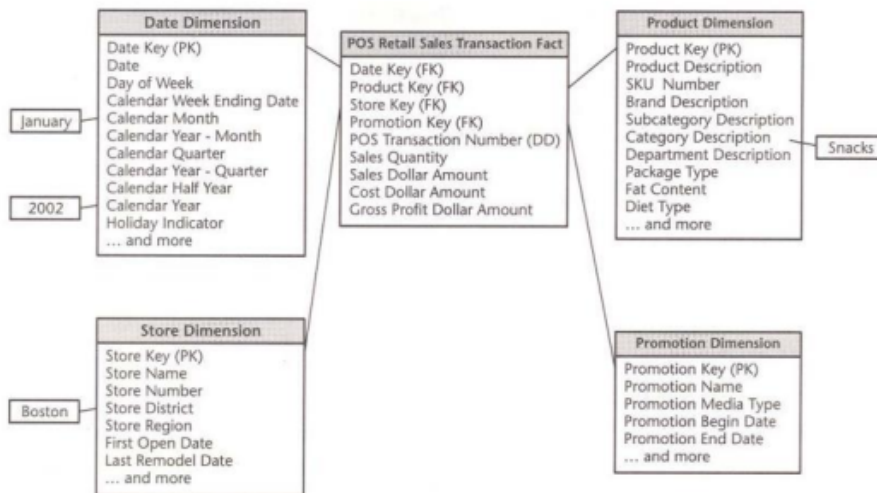
Dalam sebuah Data Warehouse atau Data Mart, Pengukuran disimpan dalam tabel fact dengan tingkat ketelitian sedemikian rupa sehingga pengguna dapat melacak ringkasan berdasarkan levelnya. Hal ini dikenal dengan agregasi. Sebagai contoh, jika data penjualan dalam sebuah toko grosirdijaga dalam level sebuah pelanggan tunggal yang membeli item tertentu pada hari tertentu dalam toko tertentu, sehingga kita dapat meringkas atau menjumlahkan data untuk hari, minggu, bulan, kuartal dan tahun tertentu; dan semua hal ini adalah untuk sebuah toko, zona, wilayah dan negara sebagaimana berdasarkan produk, kelompok produk, departemen dan seterusnya.

Hanya data penjualan pada level terendah yang dijaga dalam tabel fact, namun deskripsi dari berbagai level data di jaga dalam tabel dimensi, sehingga tool yang sesuai dapat digunakan untuk meringkas data dalam level yang bervariasi. Sebuah hierarki mendefinisikan sebuah urutan pemetaan dari seperangkat konsep low-level hingga level yang lebih tinggi, konsep level yang lebih umum. Bayangkan sebuah hierarki dimensi Lokasi. Jika Kota adalah level terendah dari hierarki, maka semua kota dapat dipetakan ke level yang lebih tinggi lagi, yakni provinsi, dan semua provinsi dapat dipetakan ke level yang lebih tinggi lagi yakni negara. Dst.

Level dimensional membentuk struktur seperti sebuah pohon dan anggota level terendah dari hierarki dikenal dengan leaf. Hanya ada satu anggota pada level paling atas. Sebuah dimensi tidak dapat eksis tanpa anggota leafnya, namun hal yang mungkin untuk memiliki sebuah dimensi tanpa anggota leaf – yakni yang hanya memiliki sebuah level.

I. Pengimplementasian Hierarki Dimensional

Hierarki dimensional disimpan sebagai atribut dalam tabel dimensi, dan seluruh hierarki terkait disimpan dalam sebuah tabel dimensi tunggal. Deskripsi untuk masing-masing level hierarki dijaga dalam metadata multidimensional. Sebagai contoh, tanggal, hari, bulan, dan tahun disimpan dalam sebuah dimensi tanggal; sementara produk, merek, kategori dan departemen disimpan dalam dimensi produk. Contoh pada Gambar dibawah ini mengilustrasikan skema database toko ritel, dan tanggal yang terkait maupun tabel dimensi produk,



Gambar 11 Ilustrasi skema database toko ritel

J. Penggunaan Hierarki Dimensional

Hierarki dalam dimensi digunakan untuk pemilihan dan agregasi data dengan level yang ditentukan sesuai kebutuhan. Tabel fact berisikan data hanya level paling rendah dalam hierarki. Data pada level yang lebih tinggi diperoleh melalui agregasi data fact level paling rendah untuk instan yang sama dari sebuah atribut level dimensional. Pada contoh yang telah diberikan, jika kita ingin menemukan total Jumlah Penjualan serta Total Harga Penjualan untuk masing dari dua departemen, Bakery dan Frozen Food, Kita pertama kali memilih Bakery dan Frozen Food dari tabel Dimensi Produk dan kemudian menambahkan seluruh nilai kuantitas penjualan serta Total Harga dari tabel fact yang berhubungan dengan kedua produk tersebut. Hal ini membutuhkan penambahan secara terpisah, nilai fact untuk key produk = 1,2,3 dan 4 serta key produk = 5,6,7,8 dan 9 bagi seluruh nilai yang mungkin dari key lainnya dalam tabel fact.

Deskripsi Departemen	Kuantitas Penjualan	Total Harga Penjualan
----------------------	---------------------	-----------------------

Bakery	5,088	\$12,331
Frozen Food	15,565	\$31,776

Selain agregasi berdasarkan deskripsi produk, jika kita ingin melihat lebih rinci dekripsi merk dari produk, kita rancang Deskripsi Produk dan Deskripsi Merk dari Dimensi Produk dan kemudian memilih seluruh kuantitas penjualan serta jumlah Total Harga Penjualan dari tabel fact dan menambahkannya.

K. Operasi OLAP

Dalam model multidimensional, data diorganisasikan menjadi dimensi banyak, dan masing-masing dimensi berisikan level abstraksi yang banyak yang didefinisikan oleh hierarki-hierarki. Organisasi ini menyediakan bagi pengguna kemampuan untuk melihat data dari perspektif yang berbeda. Sejumlah operasi data cube eksis untuk terwujudnya sudut pandang yang berbeda, memungkinkan pnegkuerian dan analisis data yang interaktif. Berikut adalah beberapa operasi OLAP untuk Data Multidimensional.

Dalam penjelasan ini akan digunakan contoh sebuah kubus (cube) yang berisikan dimensi lokasi, waktu dan item dimana lokasi diagregasikan dengan respek terhadap data kota, waktu diagregasikan dengan respek terhadap kuartal dan item diagregasikan dengan respek terhadap tipe.

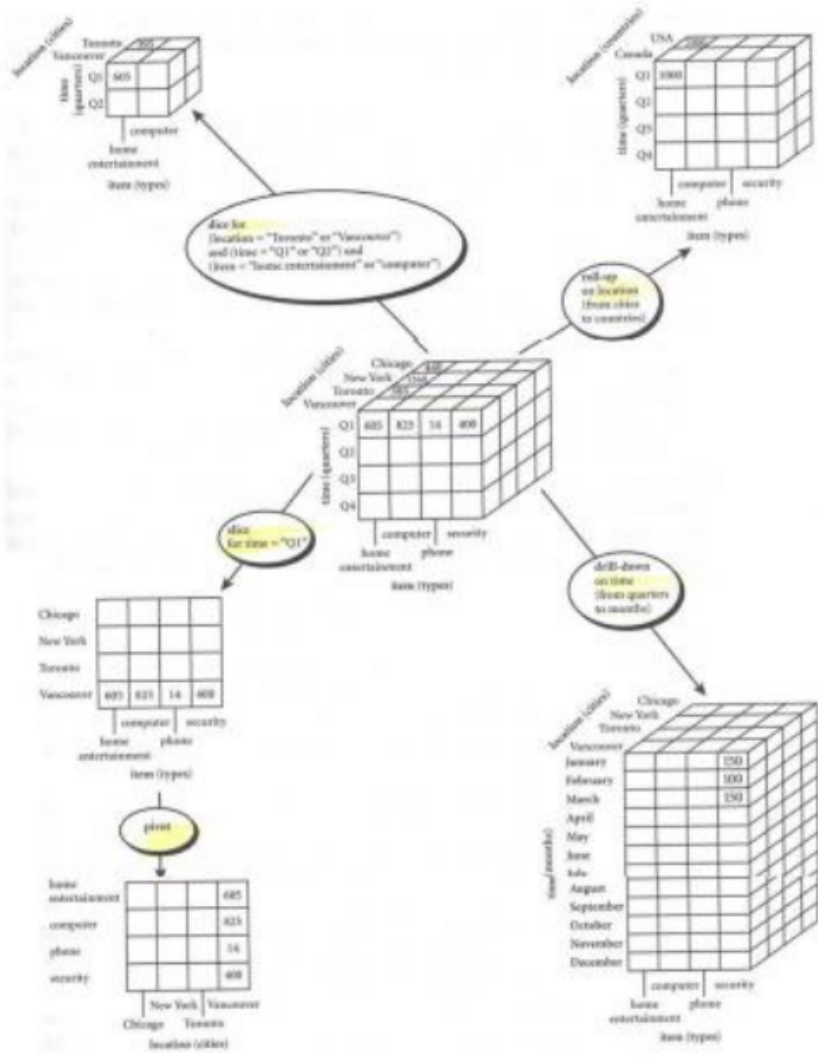
Roll-Up : Operasi roll-up atau drill-up menjalankan agrehasi pada sebuah data cube, baik dengan menaiki sebuah hierarki data untuk sebuah dimensi atau dengan pengurangan dimensi. Roll-up dengan pengurangan dimensi berarti bahwa agregasi dijalankan hingga level atas dimensi. Sebagai contoh, jika hierarki lokasi berisikan tiga level, kota → provinsi → negara, maka reduksi dari

dimensi lokasi berarti hasil data fact diringkaskan terhadap kotanya dan kemudian provinsinya.

Drill-Down: kebalikan roll-up. Drill down menavigasikan dari data rinci yang sedikit hingga yang lebih detil, hal ini dapat dilakukan dengan menuruni hierarki untuk sebuah dimensi atau memperkenalkan dimensi tambahan. Penambahan sebuah dimensi baru berarti tabel fact pasti berisikan (atau ditambahkan) data di dalam dimensi tersebut.

Slide and Dice : Operasi slice menjalankan seperangkat seleksi pada sebuah dimensi dari cube tertentu yang menghasilkan sebuah subcube. Sebagai contoh, kita dapat memilih seluruh data penjualan untuk seluruh kota dan item pada kuartar tertentu = Q1 Operasi dice mendefinisikan sebuah subcube dengan menjalankan sebuah seleksi pada dua atau lebih dimensi. Sebagai contoh, kita dapat melakukan slice pertama kali pada waktu untuk memasukkan penjualan pada beberapa kuartal. Dan kemudian pada location untuk memasukkan penjualan pada beberapa kota.

Pivot(Rotate) : Pivot adalah sebuah operasi visualisasi yang merotasikan data axis untuk memberikan sebuah alternatif presentasi data.



L. Evaluasi

1. Apa itu hierarki dan kategori yang berlaku untuk tabel dimensi?
2. Jelaskan komposisi kunci utama untuk tabel dimensi dan fakta.
3. Dari berbagai schema dari data warehouse yang ada, bagaimana cara menentukan schema mana yang sesuai.
4. Sebutkan dan jelaskan beberapa operasi OLAP untuk Data Multidimensional.

BAB VII

Extraction, Transformation dan Loading

A. Tujuan Instruksional Umum

Mahasiswa diharapkan dapat memahami ETL.

B. Tujuan Instruksional Khusus

Mahasiswa diharapkan mampu :

1. Memahami extraction data warehouse
2. Memahami loading dan transformasi

Data Warehouse dibutuhkan selalu untuk melayani tujuan dan kepentingan analisis bisnis. Untuk melakukan hal ini maka data yang berasal dari satu atau beberapa sistem operasi harus diekstrak dan dikopi ke dalam Data Warehouse. Tantangan dalam lingkungan Data Warehouse adalah untuk mengintegrasikan, menyusun kembali dan mengkonsolidasikan volume data yang cukup besar melalui banyak sistem, yang oleh karenanya memberikan satu kesatuan informasi baru yang menjadi basis bagi business intelligence.

Proses Ekstraksi data dari beberapa sumber data dan memasukkannya ke dalam Data Warehouse dikenal dengan sebutan ETL, yang merupakan singkatan dari Extraction, Transformation dan Loading. Dimana ETL menunjukkan sebuah proses luas yang terdiri dari tiga tahapan. Akronim ETL nampak terlihat sederhana karena menghilangkan fase transportasi dan masing-masing fase proses lainnya yang berbeda-beda. Walaupun, keseluruhan proses tersebut dikenal dengan sebutan ETL.

Metodologi dan kegiatan ETL dikenal dengan baik selama beberapa tahun, dan tidak hanya bersifat unik digunakan dalam lingkungan Data Warehouse: sistem database dan aplikasi berhak cipta yang menjadi backbone proses IT skala Enterprise. Data telah dishare antara aplikasi atau sistem, mencoba untuk mengintegrasikannya, memberikan paling tidak dua aplikasi gambaran yang sama mengenai dunia. Data Sharing demikian paling sering dikenal dengan mekanisme yang disebut ETL.

C. Extraction Data Warehouse

Ekstraksi (Extraction) adalah operas ekstraksi data dari sebuah sistem sumber untuk digunakan lebih jauh dalam lingkungan Data Warehouse. Tahapan ini adalah yang paling pertama dalam proses ETL. Setelah Ekstraksi, data ini akan ditransformasikan dan di-load ke dalam Data Warehouse.

Pendesainan dan Pembuatan proses Ekstraksi adalah satu kegiatan yang paling sering menyita waktu di dalam proses ETL dan dalam keseluruhan proses Data Warehouse. Sistem sumber sangat proses kompleks serta didokumentasikan secara buruk, sehingga menentukan data yang dibutuhkan untuk diekstrak menjadi sulit. Data diekstrak tidak hanya sekali namun beberapa

kali dalam suatu periode untuk mensuplai data ke dalam Data Warehouse dan menjaga agar up-to-date. Lebih jauh lagi, Sistem sumber tidak dapat dimodifikasi atau bahkan kinerja dan ketersediaannya tidak dapat diatur untuk mengakomodasi kebutuhan proses ekstraksi Data Warehouse.

Metode Ekstraksi dalam Data Warehouse adalah beberapa bentuk dan benar-benar tergantung pada Sistem sumber juga kebutuhan bisnis dalam lingkungan target Data Warehouse. Tidak ada kemungkinan untuk menambahkan logika tambahan ke dalam sistem sumber dalam proses ekstraksi data yang bersifat inkremental sehubungan dengan kinerja atau beban kerja sistem yang meningkat. Ada dua bentuk Metode Ekstraksi logical.

19

1. Ekstraksi Full(Full Extraction)

Data diekstrak secara lengkap dari sistem sumber. Ekstraksi ini melibatkan seluruh data yang sedang tersedia dalam sistem sumber. Data sumber disediakan dan tidak dibutuhkan logika informasi tambahan (seperti timestamp) yang dibutuhkan pada situs sumber. Sebuah contoh ekstraksi penuh adalah ekspor file dari sebuah tabel yang berbeda atau kueri remote SQL yang membaca sumber data lengkap.

2. Ekstraksi Inkremental (Incremental Extraction)

11

Pada poin waktu tertentu, hanya data yang memiliki histori data akan diekstrak. Event ini adalah proses ekstraksi yang dilakukan paling akhir atau sebagai contoh sebuah event bisnis yang kompleks seperti hari booking terakhir dari suatu periode fiskal. Informasi ini juga dapat disediakan oleh data sumber itu sendiri seperti sebuah kolom aplikasi, merefleksikan timestamp yang paling akhir berubah atau sebuah tabel yang berubah

dimana sebuah mekanisme tambahan yang sesuai menjaga track perubahan selain transaksi yang permulaan. Dalam banyak hal, menggunakan metode ini berarti menambah logika ekstraksi ke dalam sistem sumber.

Kebanyakan Data Warehouse tidak menggunakan teknik menangkap perubahan sebagai bagian proses ekstraksi. Sebaiknya seluruh tabel dari sistem sumber diekstrak ke Data Warehouse atau staging area, dan tabel-tabel ini dibandingkan dengan yang telah diekstrak sebelumnya dari sistem sumber untuk mengidentifikasi data yang berubah. Pendekatan ini boleh jadi tidak memiliki pengaruh yang signifikan terhadap sistem sumber, namun secara jelas dapat menciptakan sebuah beban yang dapat dipertimbangkan pada proses Data Warehouse, terutama sekali jika volume data sangat besar.

Bergantung pada metode ekstraksi logikal yang dipilih dan kapabilitasnya serta restriksi pada sumber data, data yang diekstrak secara fisik diekstrak dengan dua mekanisme. Data dapat diekstrak online dari sistem sumber atau dari sebuah struktur offline. Struktur offline demikian boleh jadi ada atau dapat dihasilkan oleh sebuah fungsi ekstraksi. Ada dua metode ekstraksi fisik (physical extraction)

1. Online Extraction

Data diekstrak langsung dari sistem sumber itu sendiri. Proses ekstraksi dapat berhubungan secara langsung dengan sistem sumber untuk mengakses tabel sumber atau ke sebuah sistem perantara yang menyimpan data dengan sebuah cara yang dikonfigurasi terlebih dahulu (sebagai contoh log atau tabel perubahan). Dengan catatan bahwa sistem perantara secara fisik tidak berbeda dari sistem sumber.

2. Offline Extraction

Data tidak diekstrak secara langsung dari sistem sumber namun dibatasi secara eksplisit diluar sistem sumber orisinal. Data telah memiliki struktur atau telah dibuat melalui prosedur ekstraksi.

Beberapa struktur yang harus dipertimbangkan antara lain :

- ✚ Flat file
- ✚ Dump File, informasi mengenai objek yang dimasukkan atau tidak dimasukkan, bergantung pada utility yang dipilih.
- ✚ Log Archive dan Redo
- ✚ Transportable Tablespaces, cara ekstrak dan memindahkan data bervolume besar antar Database.

Beberapa contoh Ekstraksi Data Warehouse, antara lain :

1. Ekstraksi menggunakan Data File

Kebanyakan Sistem database menyediakan mekanisme pengeksporan atau pelimpahan data dari format database internal menjadi file flat. Ekstraksi dari sistem mainframe seringkali menggunakan program COBOL, namun kebanyakan database, sebagaimana vendor perangkat lunak pihak ketiga, menyediakan ekspor atau pelimpahan utility.

Ekstraksi Data tidak berarti bahwa seluruh struktur database dilimpahkan dalam flat file. Dalam banyak hal, boleh jadi hal ini sesuai atau cocok untuk melimpahkan seluruh tabel database atau objek. Dalam hal lain, bisa jadi yang cocok hanya melimpahkan seperangkat tabel yang diberikan seiring perubahan pada sistem sumber. Jika dalam sistem Database oracle atau SQL

Server, ada beberapa alternatif yang tersedia untuk mengekstrak Data menjadi File antara lain :

a. Ekstraksi menjadi Flat File menggunakan SQL*Plus

Adalah teknik dasar dalam ekstraksi data yang paling banyak digunakan dan langsung menciptakan hasil kueri ke dalam bentuk file. Sebagai contoh mengekstraksi flat file country_city.log dengan menggunakan delimiter piping '|' antar koom data,yang berisikan daftar kota di US dalam tabel countries dan customers, maka kueri yang dapat dijalankan adalah sbb:

```
2
SET echo off SET pagesize 0 SPOOL country_city.log
SELECT distinct t1.country_name ||'|'|| t2.cust_city
FROM countries t1, customers t2 WHERE t1.country_id = t2.country_id
AND t1.country_name= 'United States of America';
SPOOL off
```

b. Ekstraksi ke File Ekspor menggunakan Utility Export

Utility Export memungkinkan Tabel (termasuk data) untuk diekspor menjadi file Ekspor Database. Yang menerangkan ekstraksi dari hasil kueri SQL. Ekspor berbeda dengan metode sebelumnya dengan perbedaan sbb:

- ✚ File Export berisikan metadata seperti data. File Export berisikan tidak hanya Data mentah dari sebuah tabel, melainkan juga informasi bagaimana membuat tabel kembali, secara potensial termasuk indeks, constraint, grant dan atribut lainnya yang berhubungan dengan tabel.
- ✚ Sebuah File export tunggal boleh jadi berisikan seperangkat objek tunggal, banyak objek database atau bahkan keseluruhan skema.

- ✚ Export tidak dapat secara langsung digunakan untuk mengekspor hasil kueri kompleks. Export dapat digunakan hanya untuk mengekstrak seperangkat objek database yang berbeda.
- ✚ Hasil dari utility export dapat diproses menggunakan utility import.

c. Ekstraksi ke File Eksport menggunakan Tabel Eksternal

Selain menggunakan utility Export, eksternal tabel juga dapat digunakan untuk mengekstrak hasil menggunakan operasi SELECT. Data disimpan dalam platform bebas. Contoh ekstraksi hasil operasi join yang diparalelkan menjadi empat file. Jika menggunakan Oracle, maka format yang digunakan ORACLE_DATAPUMP

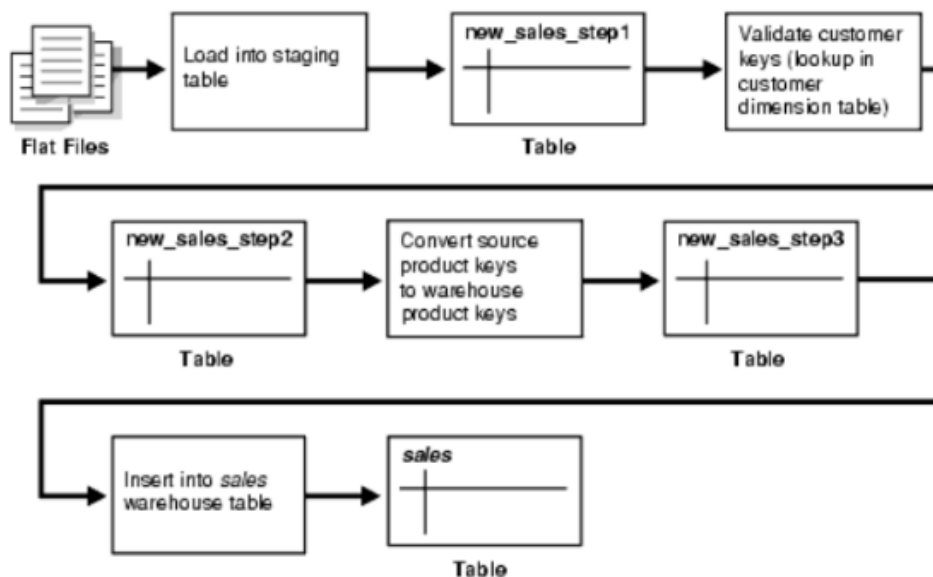
```
CREATE DIRECTORY def_dir AS
/net/dlsun48/private/hbaer/WORK/FEATURES/et';
DROP TABLE extract_cust;
CREATE TABLE extract_cust
ORGANIZATION EXTERNAL
(TYPE ORACLE_DATAPUMP DEFAULT DIRECTORY def_dir ACCESS
PARAMETERS
(NOBADFILE NOLOGFILE)
LOCATION ('extract_cust1.exp', 'extract_cust2.exp', 'extract_cust3.exp',
'extract_cust4.exp'))
PARALLEL 4 REJECT LIMIT UNLIMITED AS
SELECT c.*, co.country_name, co.country_subregion, co.country_region
FROM customers c, countries co where co.country_id=c.country_id;
```

D. Loading dan Transformasi

Transformasi Data seringkali sangat kompleks, dalam hal waktu proses, bagian proses ekstraksi, transformasi dan loading yang paling membutuhkan banyak biaya. Proses ini boleh jadi merentang dari konversi data sederhana hingga teknik pengumpulan data kompleks yang ekstrim. Dari perspektif arsitektural, Data dapat ditransformasikan dengan dua cara :

1. Multistage Data Transformation

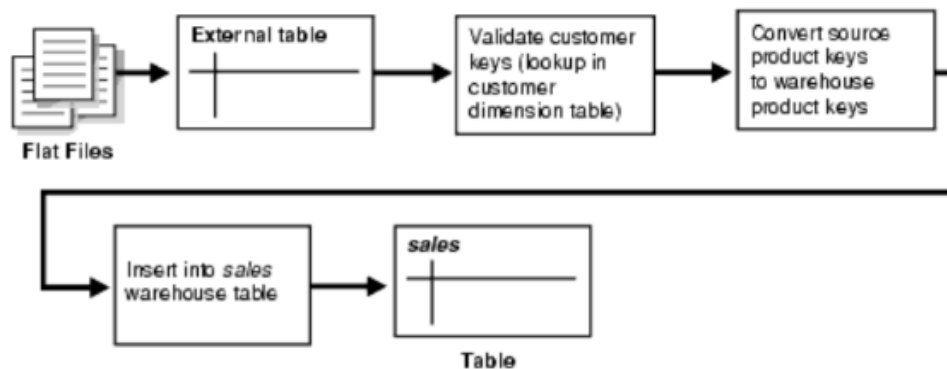
Logika transformasi data bagi kebanyakan Data Warehouse terdiri dari beberapa tahapan. Sebagai contoh, dalam transformasi rekod baru yang dimasukkan ke dalam sebuah tabel penjualan (sales), boleh jadi terdapat tahapan transformasi logik yang terpisah untuk memvalidasi masing-masing key dimensi. Gambaran secara grafis dari proses transformation logic adalah sbb :



Gambar 12 Transformasi Data Multistage

2. Pipelined Data Transformation

Arus proses ETL dapat diubah secara dramatis dan database menjadi sebuah bagian integral solusi ETL. Fungsionalitas barunya melukiskan beberapa pembentukan tahapan proses penting yang kuno ketika beberapa yang lainnya dapat dimodel kembali untuk menambah arus data dan transformasi data menjadi lebih dapat diukur. Kegiatannya bergeser dari transformasi serial hingga proses load (dengan kebanyakan kegiatan dilakukan diluar database) atau load-kemudian proses transformasi untuk meningkatkan transformasi selagi loading.



Gambar 13 Transformasi Pipelined Data

E. Evaluasi

1. Berikan tiga alasan mengapa fungsi ETL paling menantang di lingkungan data warehouse.
2. Sebutkan lima jenis kegiatan yang merupakan bagian dari proses ETL.
3. Sebutkan lima jenis tugas transformasi utama. Berikan contoh untuk masing-masing.

4. Sebutkan dua kategori umum dari data yang disimpan dalam sistem operasional ? Memberikan dua contoh untuk masing-masing.
5. Berikan contoh Ekstraksi Data Warehouse

DAFTAR RUJUKAN

- Anahory, Murray, *Datawarehousing in The Real World*, Addison Wesley.
- Alex Berson and Stephen J. Smith, 1997, *Datawarehousing, DataMining and OLAP*, McGraw Hill,
- Carl Rabeler , 2003 , "*Microsoft SQL Server 2000 DTS Step by Step*", Microsoft Press
- Connoly, Thomas; Begg, Carolyn; Strachan, Anne 2003; *Database Systems,; A Practical Approach to Design, Implementation and Management*, 3rd edition, Addison Wesley.
- Elmasri, Ramez; Navathe, Shamkant B.; *Fundamentals of Database Systems*, The Benjamin/Cummings Publishing Company, Inc., California.
- Korth, H, 2002. *Database System Concept*, Mc Graw Hill, 4th edition, New York, 2002.
- Paulraj Ponniah, 2001, "*Data Warehousing Fundamentals*", John Wiley & Sons, Inc
- Raphl Kimbal, 2002, "*The Data Warehouse Toolkit*", Willey
- Subrahmanian, V.S, 1998, *Principles of Multimedia Database System*, Morgan Kaufmann. Inc.
- Won Kim, 2001, *Modern Database System : The Ricardi*, Greg, *Principles of Database Systems With Internet and Java Applications*, Addison Wesley, Boston.
- W. H. Inmon, 2002, "*Building the Data Warehouse Third edition*", John Wiley & Sons, Inc

978-602-6534-19-4



Penerbit BINTANG

Jl. Potro Agung III 41 C, Surabaya 60135
Telp. 031-3770687, Fax. 031-3715941
E-mail: bintangsb@yahoo.co.id

2.Buku ajar Data Warehouse 2020

ORIGINALITY REPORT

20%
SIMILARITY INDEX

20%
INTERNET SOURCES

5%
PUBLICATIONS

0%
STUDENT PAPERS

PRIMARY SOURCES

1	vdocuments.site Internet Source	1%
2	pt.scribd.com Internet Source	1%
3	journal.lppmunindra.ac.id Internet Source	1%
4	docplayer.info Internet Source	1%
5	widuri.raharja.info Internet Source	1%
6	epratikna.wordpress.com Internet Source	1%
7	eprints.umg.ac.id Internet Source	1%
8	informasidatawarehouse.wordpress.com Internet Source	1%
9	ojs.uho.ac.id Internet Source	1%

10	library.binus.ac.id Internet Source	1 %
11	123dok.com Internet Source	1 %
12	ronallaruku.blogspot.com Internet Source	1 %
13	eprints.ums.ac.id Internet Source	1 %
14	sumberkumakalah.blogspot.com Internet Source	1 %
15	repository.unej.ac.id Internet Source	1 %
16	xinez.site90.com Internet Source	<1 %
17	dspace.uui.ac.id Internet Source	<1 %
18	volontegenerale.nl Internet Source	<1 %
19	prosiding.seminar-id.com Internet Source	<1 %
20	blog.umy.ac.id Internet Source	<1 %
21	pakdosen.pengajar.co.id Internet Source	<1 %

22 pinandhitahasbullahmustaqim.wordpress.com <1 %
Internet Source

23 ojs.uajy.ac.id <1 %
Internet Source

24 ihatsolehat14.wordpress.com <1 %
Internet Source

25 youdhee.wordpress.com <1 %
Internet Source

26 nurromantis201501084.blogspot.com <1 %
Internet Source

27 ajunj22.blogspot.com <1 %
Internet Source

28 ADJAT SUDRADJAT. "PENGEMBANGAN DATA WAREHOUSE CALL CENTER DENGAN METODOLOGI KIMBALL NINE-STEP PADA BINA SARANA INFORMATIKA", Jurnal Informatika, 2016
Publication

29 pengajar.co.id <1 %
Internet Source

30 blog.ub.ac.id <1 %
Internet Source

31 text-id.123dok.com <1 %
Internet Source

es.scribd.com

32

Internet Source

<1 %

33

I Putu Agus Eka Pratama, Ni Putu Nirmala Dewi Widhiasi. "Perancangan Data Warehouse Untuk Prediksi Penjualan Pada Orba Express Menggunakan Pentaho", JUSS (Jurnal Sains dan Sistem Informasi), 2020

Publication

<1 %

34

e-journals.unmul.ac.id

Internet Source

<1 %

35

fififaristia.blogspot.com

Internet Source

<1 %

36

johannessimatupang.wordpress.com

Internet Source

<1 %

37

rajatrepik.com

Internet Source

<1 %

38

docplayer.net

Internet Source

<1 %

39

sisdal0823.wordpress.com

Internet Source

<1 %

40

www.scribd.com

Internet Source

<1 %

41

m-zulkifli.blogspot.com

Internet Source

<1 %

khouwrio.blogspot.com

42

Internet Source

<1 %

43

slideplayer.info

Internet Source

<1 %

44

www.kajianpustaka.com

Internet Source

<1 %

45

ejournal.ukrida.ac.id

Internet Source

<1 %

46

jurnal.dcc.ac.id

Internet Source

<1 %

47

pdfcoffee.com

Internet Source

<1 %

48

www.bigdatadba.com

Internet Source

<1 %

49

www.ingroup.co.uk

Internet Source

<1 %

50

adalah.top

Internet Source

<1 %

51

repository.usd.ac.id

Internet Source

<1 %

52

kavinadinugraha.wordpress.com

Internet Source

<1 %

53

rifdadenita.blogspot.com

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On