

PROSES EXTRACT, TRANSFORM DAN LOAD PADA DATA WAREHOUSE

Oktavian Abraham Lantang

ABSTRAK

Saat ini seiring dengan perkembangan teknologi informasi yang semakin pesat, ketergantungan proses bisnis suatu perusahaan terhadap pengolahan data akan terotomasi semakin besar pula, tentu saja kita bisa membayangkan apabila sebuah perusahaan besar dengan cabang perusahaan di beberapa pulau harus mengolah data secara manual dan tidak terintegrasi.

Permasalahan yang ada diatas sudah tidak asing lagi, karena hampir setiap perusahaan berskala besar juga memiliki masalah seperti ini. Banyaknya data transaksi dan desentralisasi yang menyebar diseluruh dunia menyebabkan perusahaan harus membuat atau membeli sistem yang lebih baik, terintegrasi dan mudah dikelola.

Dalam proses pengambilan kebijakan perusahaan dibidang strategi bisnis maka perusahaan memerlukan suatu data olahan yang bersifat analisa, serta dapat menjawab terhadap kebutuhan yang ditawarkan berdasarkan konsep data warehouse. Untuk menghasilkan analisis data yang berkualitas dan akurat maka konsep data warehouse mempunyai beberapa tahapan. Salah satu tahapan dari proses data warehouse adalah ETL (Extract Transform Load), dan karya tulis ini membahas tentang proses ETL.

Tujuan dari penelitian ini adalah bagaimana mengembangkan ETL yang berkualitas untuk mendukung data warehouse yang baik. Metode yang penulis pakai mengacu kepada case real yang bisa kita coba untuk membuat data warehousenya. Hasil dan kesimpulannya dari ETL ini diharapkan bisa membuat proses data warehouse lebih terkontrol dengan baik.

Kata Kunci : *Data warehouse, Extract, Transform, Load*

PENDAHULUAN

1.1 Latar Belakang

Saat ini seiring dengan perkembangan teknologi informasi yang semakin pesat, ketergantungan proses bisnis suatu perusahaan terhadap pengolahan data akan terotomasi semakin besar pula, tentu saja kita bisa membayangkan apabila sebuah perusahaan besar dengan cabang perusahaan di beberapa pulau harus mengolah data secara manual dan tidak terintegrasi.

Untuk menjawab permasalahan yang dihadapi oleh perusahaan besar seperti diatas maka perlu dipahami pentingnya penggunaan *data warehouse*. Tujuan penggunaan data warehouse yaitu untuk mengintegrasikan semua data yang ada di setiap cabang perusahaan sehingga diperoleh data akurat yang memudahkan bagi *top manager* perusahaan tersebut dalam menganalisa informasi dan kemudian dapat mengambil keputusan yang tepat dalam menentukan strategi bisnis .

Dalam membangun suatu data warehouse ada beberapa tahapan yang harus diperhatikan,

salah satu bagian dari data warehouse yang mempunyai peranan penting dalam menghasilkan suatu analisa data yang akurat adalah proses ETL (Extract, Transform, Load)

1.2 Rumusan Masalah

Masalah timbul dikarenakan adanya redundansi data yang dikelola secara terpisah di masing-masing cabang dan tidak terkait satu sama lainnya sehingga sulit bagi pihak *management* untuk menganalisis data yang menyebabkan perusahaan kesulitan dalam menentukan strategi bisnisnya.

Sering kali dalam perancangan sebuah data warehouse si perancang tidak memperhatikan proses ETL sehingga dampaknya data hasil saringan yang dihasilkan oleh data warehouse tidak *qualified*. Untuk itulah penulis merumuskan suatu rumusan masalah yaitu "Apa dan bagaimana cara membuat suatu proses *Extract, Transform dan Load* yang baik sehingga informasi yang dihasilkan oleh data warehouse mempunyai kualitas yang baik ? "

1.3 Tujuan Penulisan

Tujuan dari penulisan ini adalah untuk memberikan pemahaman mengenai konsep data warehouse, tahapan perancangan data warehouse, proses extract transform load.

1.4 Batasan Penulisan

LANDASAN TEORI

2.1 Teori Database

Secara umum, database adalah sekumpulan data atau informasi yang disimpan dalam bentuk digital. Bahasa yang digunakan untuk merepresentasikan database adalah Structured Query Language (SQL). SQL terdiri dari dua macam, yaitu *Data Definition Language* (DDL) dan *Data Manipulation Language* (DML). Implementasi DDL dan DML berbeda untuk tiap sistem manajemen basis data (DBMS). Namun secara umum implementasi tiap bahasa ini memiliki bentuk standar yang ditetapkan ANSI. Artikel ini akan menggunakan bentuk paling umum yang dapat digunakan pada kebanyakan DBMS.

2.1.1 Data Definition Language

DDL digunakan untuk mendefinisikan, mengubah, serta menghapus basis data dan objek-objek yang diperlukan dalam basis data, misalnya tabel, *view*, *user*, dan sebagainya. Secara umum, DDL yang digunakan adalah *CREATE* untuk membuat objek baru, *USE* untuk menggunakan objek, *ALTER* untuk mengubah objek yang sudah ada, dan *DROP* untuk menghapus objek. DDL biasanya digunakan oleh administrator basis data dalam pembuatan sebuah aplikasi basis data.

2.1.2 Data Manipulation Language

DML digunakan untuk memanipulasi data yang ada dalam suatu tabel. Perintah yang umum dilakukan adalah:

- *SELECT* untuk menampilkan data.
- *INSERT* untuk menambahkan data baru.
- *UPDATE* untuk mengubah data yang sudah ada.
- *DELETE* untuk menghapus data.

2.2.1 Arsitektur Data Warehouse

Seperti halnya perancangan perangkat lunak, maka arsitektur merupakan bagian penting dalam membangun suatu *Data Warehouse*. Pada

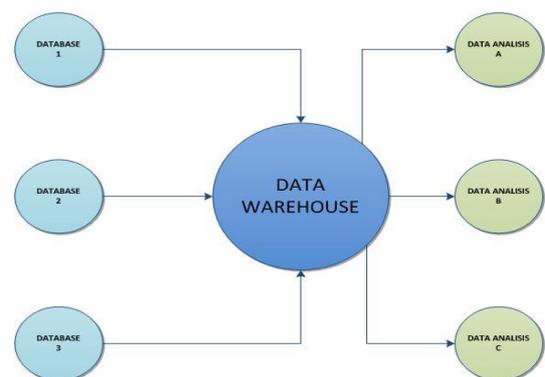
Batasan dari penulisan karya tulis ini adalah hanya pada konsep data warehouse dasar data warehouse, tahapan apa saja yang harus diperhatikan dalam membangun data warehouse dan pembahasan proses ETL yang baik dalam data warehouse.

2.2 Data Warehouse

Data warehouse adalah kumpulan dari banyak data yang dapat diekstrak menjadi informasi digital dalam bentuk laporan dan analisa data untuk membantu perencanaan strategi perusahaan. Berikut ini adalah komponen utama dari datawarehouse, diantaranya adalah:

- Data akuisisi, adalah proses mengekstraksi data dari sebuah database ke wilayah staging dan mempersiapkan data untuk dikirim ke sistem repositori.
- Penyimpanan data, adalah wilayah repository yang digunakan untuk menyimpan data di data warehouse.
- Pengiriman informasi, adalah proses dimana informasi dikirimkan dalam bentuk laporan dan analisa perusahaan, melalui query-query. Pengguna melakukan analisa multidimensi cube dari sebuah data warehouse.

Dalam proses datawarehouse, data dinormalisasikan dan diatur agar konsumen atau klien mendapatkan analisa data yang akurat dan grafik-grafik yang dapat menunjang kinerja perusahaan. Untuk lebih mudah memahami konsep data warehouse dapat dilihat pada diagram dibawah ini.



Gbr. 2.1 : Diagram konsep data warehouse

sub bagian ini akan dijelaskan teori mengenai arsitektur *Data Warehouse*. Arsitektur *Data Warehouse*, terdiri atas beberapa komponen,

yaitu: *Data Source*, *Data Staging*, *Data Storage*, *Metadata*, dan *Information Delivery*. Model dasar arsitektur *data warehouse* ini biasa disebut dengan arsitektur *two-level* [HOF05]. Model arsitektur *two-level* ini memiliki tiga sub model yaitu *independent data mart*, *dependent data mart* dan terakhir *active data warehouse*. Pada model arsitektur *two-level* ini data diekstrak dari berbagai macam *database* dan arsip yang bersumber dari internal maupun eksternal. Data-data tersebut ditransformasi dan diintegrasikan sebelum dimuatkan pada *Data Warehouse*. Data-data diklasifikasikan baik berupa ringkasan data maupun rinci untuk mendukung pengambilan keputusan. Pengguna kemudian mengakses *Data Warehouse* tersebut dengan berbagai *query* maupun *analytical tools*.

2.2.2 Data Mart

Sebelum menjelaskan arsitektur *independent data mart*, akan diulas terlebih dahulu pengertian *data mart*. *Data mart* adalah sebuah *Data Warehouse* atau data hasil dari proses ekstraksi, transformasi dan *loading*, yang dibatasi oleh ruang lingkungannya untuk digunakan pada kelompok *end-user* tertentu. *Data mart* dirancang untuk mengoptimalkan pemakaian *query* yang telah didefinisikan dengan baik dan telah dapat diprediksi sebelumnya.

Arsitektur *independent data mart* yaitu suatu *data mart* yang diisi oleh data hasil ekstraksi dari operasional tanpa memanfaatkan *Data Warehouse*.

2.3 ETL (Extract, Transform and Load)

Suatu proses yang berhubungan dengan ETL dalam data warehouse sebagian besar adalah masalah yang rumit dan membutuhkan kinerja manusia yang besar. ETL berfungsi untuk mengubah data dari sumber data menjadi informasi yang dapat disimpan ke dalam sistem data warehouse.

2.3.1 Ekstraksi Data

Dalam memproses data, setiap profesional pasti pernah melakukan ekstraksi data dan konversi seperti yang dilakukan pada sistem basis data. Ketika kita mengkoneksikan sistem database dengan sistem operasional peranti lunak. Dua factor penting yang membedakan

ekstraksi data dalam sistem operasi dengan ekstraksi data dalam data warehouse, dalam data warehouse data yang diekstraksi dapat dilakukan perbagian, sedangkan dalam sistem operasi data dikonversi dan diekstrak dalam sekali waktu.

2.3.2 Transformasi Data

Banyak fungsi transformasi yang dibutuhkan sebelum data di-mapping dan disimpan ke dalam repository data warehouse. Fungsi ini meliputi seleksi input, pemisahan struktur input, normalisasi dan denormalisasi struktur data, agregasi, konversi, menemukan nilai yang hilang, dan konversi nama dan alamat. Transformasi data diantaranya adalah mengubah dari satu sumber ke format target sumber.

2.3.3 Loading Data

Dalam proses memindahkan data ke repository data warehouse dapat dilakukan dengan berbagai macam cara. Dalam hal ini ada 3 tahapan yang dilakukan dalam me-load data, yaitu:

- **Initial Load**

mempopulasikan seluruh table data warehouse pada pertama kalinya. Dalam Initial Load, ada 2 metode yang dapat digunakan untuk merefresh data yaitu : update (melakukan perubahan secara incremental) dan refresh(me-load kembali dengan data yang baru).

- **Incremental Load**

Melakukan proses perubahan yang dibutuhkan untuk strategi planningnya.

- **Full Refresh**

Melakukan penghapusan isi dari table dan memasukkan kembali dengan data yang baru.

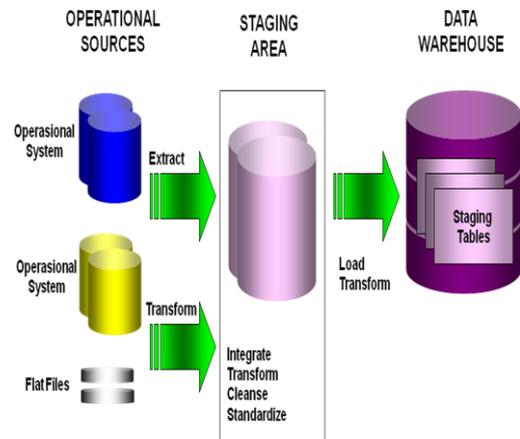
PEMBAHASAN

3.1 Alur Proses ETL

Langkah pertama adalah mengidentifikasi tabel-tabel pada data sumber yang memuat data-data yang dibutuhkan untuk metrik dan dimensi yang telah ditetapkan sesuai kebutuhan user (*user requirements*). Data-data tersebut kemudian dimasukkan ke dalam *staging area*, hal ini dimaksudkan agar proses ETL yang akan dilakukan terhadap data-data tersebut tidak mengganggu proses transaksi OLTP yang sedang berlangsung. Untuk lebih memahami proses ETL penulis coba menjelaskan melalui tahapan dibawah ini :

- Data-data yang ada dalam tabel harus diekstraksi, yaitu dengan cara memilih field-field yang dibutuhkan yang sesuai dengan metrik dan dimensi pada user requirements, sedangkan field-field yang tidak dibutuhkan diabaikan atau dibuang. Inti dari tahapan ini adalah memfilter data yang mempunyai relevansi dengan kebutuhan user yang akan diproses ketahapan selanjutnya agar menghasilkan data yang berkualitas
- Dilakukan proses transformasi terhadap data-data terpilih itu, data-data tersebut diproses menjadi bentuk yang diinginkan untuk kemudian dimasukkan ke Data Mart, proses transformasi sendiri meliputi beberapa tahapan yaitu :
 - a) Proses transformasi yang pertama adalah memformat data, yaitu menentukan tipe data pada setiap field.
 - b) Selanjutnya adalah proses joining antara tabel dan tabel agar dapat diketahui yang masih aktif dan ya sudah tidak aktif di perusahaan.
 - c) Membuat tabel-tabel fakta dan tabel-tabel dimensi sesuai dengan user requirement yang telah ditentukan
 - d) Kemudian pada tabel-tabel fakta dan dimensi tersebut ditambahkan beberapa field turunan, yaitu field-field yang diturunkan (dihasilkan) dari perhitungan field-field lain.
 - Setelah ditransform data-data tersebut diisikan ke dalam tabel tabel fakta dan dimensi yang telah dibuat.
 - Data-data yang telah tersimpan dalam tabel fakta dan dimensi itu kemudian diubah menjadi bentuk data yang multidimensi yang dapat dilihat dari

segala sudut sehingga dapat diakses oleh client untuk keperluan analisis. Proses ini menggunakan query-query SQL dan divisualisasikan menjadi aplikasi sehingga memudahkan user dalam melihat data-data di data mart.



Gambar 3.1 : Proses ETL

PENUTUP

4.1 Kesimpulan

Dari pembahasan pada bagian sebelumnya maka dapat ditarik beberapa kesimpulan yaitu :

- Implementasi Data Warehouse dapat menjadi salah satu jawaban bagi kebutuhan sebuah perusahaan akan data yang bersifat analisis.
- Untuk mencapai suatu data warehouse yang berkualitas maka dibutuhkan proses Extract Transform dan Load (ETL).
- Dalam Pengimplementasian proses ekstrasi, tranformasi dan loading tahapan yang sudah dibahas pada bab pembahasan haruslah dilakukan untuk menghasilkan data analitical yang berkualitas

4.2 Saran

- Apabila pembaca berniat membangun suatu data warehouse perancang sebaiknya memberikan perhatian lebih pada proses ETL karena proses ini sangat mempengaruhi kualitas dari informasi yang dihasilkan oleh data warehouse yang sudah ada.
- Perancang harus memperhatikan aspek pendekatan yang sudah baku dan benar dalam implementasinya merancang data

warehouse, karena bila tidak melaksanakan pendekatan yang benar

maka data analitical yang dihasilkan nanti diragukan kualitasnya.

DAFTAR PUSTAKA

Ferdinansyah, Alex. (2006). *Perancangan Data Mart Untuk Pengelolaan Sumber Daya Manusia*. Tesis. Fakultas Ilmu Komputer Universitas Indonesia. Jakarta.

Ladjamudin, Al Bahra Bin. (2005). *Analisa dan Desain System Informasi*. Edisi pertama. Graha Ilmu. Jakarta.

Connolly Thomas, Carolyn Begg. (2002). *Database system (A practical approach to design, implementation and management*. Third Edition.

Melaniawati. (2001). *Perancangan Data Warehouse Sistem Pembelian dan Penjualan Pada PD Raya Makmur Motor*. Skripsi Fakultas Ilmu Komputer Universitas Gunadarma. Jakarta.

Paulraj Ponniah. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. Copyright John Wiley & Sons, Inc.

Riyanto, Tikno. (2006). *Pengembangan Data Warehouse Untuk Mendukung Pengambilan Keputusan Dalam Mengelola Sumber Daya Manusia (SDM) : Sebuah studi kasus TNI Angkatan Udara*. Tesis. Fakultas Ilmu Komputer Universitas Indonesia. Jakarta