

Mata Kuliah :: Data Warehousing
Komposisi oleh Budi Prayitno
2014
Batam

Data Warehouse :: Overview

Data Warehouse

Data Warehouse, dalam konteks Teknologi Informasi, adalah basis data yang digunakan untuk pelaporan dan analisis data

Bahan dasar data yang digunakan untuk Data Warehouse berasal dari sistem transaksional

E :: T :: L

E T L merupakan kependekan dari Extract, Transform and Load

Seiring dengan ETL terdapat pula terminologi lain terkait tahapan/*layer* proses:

- Staging Layer
- Integration Layer
- Access Layer

Extract

Proses ekstraksi data dari sistem sumber [sistem transaksional] untuk pengolahan lebih lanjut

Metode ekstraksi:

- Full
Mengekstrak keseluruhan data
- Incremental
Mengekstrak hanya perubahan data

Transform

Proses transformasi data dari format sistem sumber ke format sistem tujuan

Yang harus diperhatikan terkait data:

- Struktur
Struktur data sistem tujuan
- Type
Type data pada database sistem tujuan
- Unit of Measurement
Satuan data

Load

Proses memasukkan data yang telah ditransformasikan ke dalam database sistem tujuan

Yang harus diperhatikan:

- Constraints: Keys, Nullity, Check

Performance

- Metode INSERT
- Indexes

Staging Layer

Staging Layer menyimpan data mentah yang diekstraksi dari berbagai sistem database yang saling lepas

Integration Layer

Integration Layer mengintegrasikan berbagai set data yang saling lepas dengan mentransformasikannya dari Staging Layer dan menyimpannya pada Operational Data Store (ODS)

Data yang sudah terintegrasi kemudian disimpan ke Data Warehouse dalam struktur hirarki yang disebut Dimension dan Fact

Gabungan antara Dimension dan Facts dikenal dengan Star Schema

Access Layer

Access Layer adalah tahap dimana data dapat diambil oleh pengguna sesuai kebutuhan

Biasanya dikenal dengan Business Intelligence atau Decision Support System

Kelebihan

Penggunaan arsitektur data warehouse memberikan kelebihan:

- Disimpannya salinan informasi dari sistem sumber akan mempertahankan *history* data
- Penggabungan beberapa sistem sumber data memberikan sudut pandang menyeluruh terhadap perusahaan
- Memperbaiki kualitas data dengan menandai dan memperbaiki data rusak
- Secara konsisten menampilkan informasi terkini perusahaan
- Merestrukturisasi data menjadi informasi yang berguna bagi pengguna
- Merestrukturisasi data untuk mempercepat performa *query*, terutama untuk *query* analisis yang kompleks
- Semua dilakukan tanpa mengganggu sistem transaksional

Sejarah

Konsep Data Warehouse bermula pada akhir era 1980-an ketika peneliti IBM Barry Devlin dan Paul Murphy mengembangkan Business Data Warehouse

Dimensional v.s. Normalized

Terdapat dua pendekatan umum dalam penyimpanan data di Data Warehouse:

- Pendekatan Dimensional
- Pendekatan Normalized

Pendekatan Dimensional

Pendekatan Dimensional, yang praktisinya dikenal sebagai kaum *Kimballites*, menggunakan pendekatan Ralph Kimball yang menyebutkan bahwa model data warehouse harus menggunakan model dimensional/Star Schema

Pendekatan Dimensional...

Pada pendekatan Dimensional, data transaksi dibedakan menjadi Fact yang pada umumnya merupakan data numerik dan Dimension yang merupakan informasi yang memberi konteks terhadap Fact

Sebagai contoh: data penjualan/Sales dapat dibedakan menjadi Fact seperti jumlah produk yang dijual dan harga produk, kemudian menjadi Dimension seperti tanggal penjualan, nama konsumen, kode produk, alamat tujuan, alamat tagihan serta nama wiraniaga yang menangani penjualan

Pendekatan Dimensional...

Kelebihan pendekatan Dimensional untuk implementasi Data Warehouse:

- Data lebih mudah dimengerti dan digunakan oleh pengguna
- Pendekatan Dimensional lebih mudah dimengerti karena dibagi menjadi Fact/nilai dan Dimension/konteks terkait proses bisnis pada sistem transaksional
- Pengambilan data dari data warehouse cenderung lebih cepat
- Database untuk data warehouse yang diimplementasikan dengan pendekatan dimensional menggunakan struktur OLAP yang memberikan kelebihan performa untuk data retrieval

Pendekatan Dimensional...

Kekurangan pendekatan Dimensional untuk implementasi Data Warehouse:

- Untuk menjaga integritas Fact dan Dimension, mengisi data dari berbagai sistem transaksional yang berbeda sangat kompleks
- Perubahan struktur data warehouse, yang biasa terjadi karena perubahan proses bisnis, menjadi sangat rumit

Pendekatan Normalized

Pendekatan Normalized atau biasa disebut model 3NF, yang praktisinya disebut kaum *Inmonites*, mengikuti pendekatan yang dibuat oleh Bill Inmon yang menyatakan bahwa Data Warehouse harus dimodelkan dengan model E-R/model Normalized

Pendekatan Normalized...

- Pada pendekatan Normalized, data disimpan dalam data warehouse mengikuti aturan normalisasi
- Tabel dikelompokkan ke dalam kategori tertentu, misal: konsumen, produk dan keuangan
- Struktur yang ternormalisasi membagi data menjadi entitas yang menghasilkan tabel yang membentuk basis data relasional
- Pada organisasi yang besar, pendekatan ini menghasilkan banyak sekali tabel yang saling terhubung melalui jaringan JOIN karena setiap entitas diimplementasikan menjadi tabel fisik

Pendekatan Normalized...

Kelebihan pendekatan Normalized dalam implementasi Data Warehouse, antara lain:

- Penambahan dan perubahan data dapat dilakukan ke dalam basis data seperti sistem transaksional biasa

Pendekatan Normalized...

Kekurangan pendekatan Normalized dalam implementasi Data Warehouse, antara lain:

- Jumlah tabel menjadi sangat banyak
- Sulit melakukan JOIN terhadap tabel yang berasal dari berbagai sumber untuk memperoleh informasi yang bermakna
- Akses informasi memerlukan pemahaman menyeluruh terhadap data warehouse

Dimensional v.s. Normalized

Persamaan antara pendekatan Dimensional dan pendekatan Normalized adalah bahwa keduanya dapat direpresentasikan dalam Entity Relationship Diagram

Perbedaan antara pendekatan Dimensional dan pendekatan Normalized adalah tingkat normalisasi yang digunakan

Pendekatan Dimensional menghasilkan ukuran fisik basis data yang relatif lebih besar daripada pendekatan Normalized untuk jumlah data yang sama

Data Warehouse v.s. Sistem Transaksional

- Sistem transaksional dibuat untuk menjaga integritas data serta performa dalam melakukan pencatatan transaksi bisnis, dengan memanfaatkan proses normalisasi dan penggunaan model E-R
- Normalisasi pada sistem transaksional dilakukan menggunakan aturan Codd, yang jika diimplementasikan secara lengkap hingga 5 level, akan menghasilkan lusinan bahkan ratusan tabel fisik
- Sistem transaksional memiliki performa sangat baik untuk *data modification* karena perubahan data hanya terkait dengan sebagian kecil data pada setiap tabel tersebut
- Data pada sistem transaksional mengalami siklus *backup* dan *purge* secara berkala

Pemutakhiran Data

Berdasarkan konteks siklus dan tingkat pemutakhiran data, Data Warehouse dapat dibedakan menjadi:

- Offline Data Warehouse
- Online Data Warehouse