

Pengantar Datawarehouse

Muhammad Subhan

subhan@binusian.org

m@tsubhan.com

subhan_0072001@live.com

<http://subhan.blog.binusian.org> / <http://geeks.netindonesia.net/blogs/muhammadsubhan>

Lisensi Dokumen:

Copyright © 2003-2009 IlmuKomputer.Com

Seluruh dokumen di IlmuKomputer.Com dapat digunakan, dimodifikasi dan disebarkan secara bebas untuk tujuan bukan komersial (nonprofit), dengan syarat tidak menghapus atau merubah atribut penulis dan pernyataan copyright yang disertakan dalam setiap dokumen. Tidak diperbolehkan melakukan penulisan ulang, kecuali mendapatkan ijin terlebih dahulu dari IlmuKomputer.Com.

Pengertian Datawarehouse

Datawarehouse adalah kumpulan macam-macam data yang **subject oriented, integrated, time variant, dan nonvolatile**. dalam mendukung proses pembuatan keputusan. Inmon and Hackathorn (1994).

Datawarehouse sering diintegrasikan dengan berbagai sistem aplikasi untuk mendukung **proses laporan dan analisis data** dengan menyediakan data histori, yang menyediakan infrastruktur bagi EIS dan DSS.

Karakteristik Datawarehouse :

subject oriented, integrated, time variant, non volatile

Kenapa **subject oriented**?

Datawarehouse diorganisasikan pada subjek-subjek utama, seperti pelanggan, barang/ produk, dan penjualan. Berfokus pada model dan analisis pada data untuk membuat keputusan, jadi bukan pada setiap proses transaksi atau bukan pada OLTP. Menghindari data yang tidak berguna dalam mengambil suatu keputusan.

Kenapa **integrated** ?

Dibangun dengan menggabungkan/menyatukan data yang berbeda. *relational databse, flat file, dan on-line transaction record*. Menjamin konsistensi dalam penamaan, struktur pengkodean, dan struktur atribut diantara data satu sama lain.

Datawarehouse **time variant**?

Data disimpan untuk menyediakan informasi dari perspektif *historical*, data yang tahun-tahun lalu/ 4-5 thn. Waktu adalah elemen kunci dari suatu datawarehouse/ pada saat pengcapture-an.

Kenapa **Non Volatile**?

Setiap kali proses perubahan, data akan di tampung dalam tiap-tiap waktu. Jadi tidak di perbaharui terus menerus. Datawarehouse tidak memerlukan pemrosesan transaksi dan recovery. Hanya ada dua operasi **initial loading of data** dan **access of data**.

Karakteristik Datawarehouse :

subject oriented, integrated, time variant, non volatile

Datawarehouse bukan hanya tempat penyimpanan data, Datawarehouse adalah *Business Intelligence tools, tools to extract*, merubah (*transform*) dan menerima data (*load*) ke penyimpanan (*repository*) serta mengelola dan menerima metadata.

Sejarah / Evolution

- 1960, Dunia komputerisasi membuat aplikasi individu yang digunakan pada file utama. General mill mulai mengembangkan istilah ***dimensi dan fakta***.
- 1970, IRI menyediakan database dimensi untuk pembeli eceran, tahun untuk memperbaiki, mengembangkan dan mencocokkan dengan hardware yang dimiliki.
- 1983, DBMS diperkenalkan untuk mengambil keputusan.
- 1988, Barry dan Paul mempublikasikan karyanya tentang Arsitektur Bisnis dan Sistem Informasi
- 1990, memperkenalkan tool DBMS sebagai alat untuk datawarehouse.
- 1990-sekarang, banyak bermunculan buku-buku datawarehouse dan aplikasi-aplikasi datawarehouse.

Arsitektur Datawarehouse

Maksudnya adalah bagaimana datawarehouse dibangun, arsitektur tidak ada yang benar dan salah tetapi suatu arsitektur dibangun tergantung situasi dan kondisi. Arsitektur datawarehouse akan berpengaruh pada penggunaan dan pemeliharaan.

Lapisan-lapisan arsitektur datawarehouse :

1. Operational database layer / Lapisan basis data operasional
 - a. Sumber data (source) untuk datawarehouse
 - b. Data lengkap, Data hari ke hari
 - c. Mempunyai nilai saat ini/ data berarti.
 - d. Tingkat kemungkinan data besar.
2. Data Access Layer/ Lapisan Akses Data
 - a. Tools untuk mengekstrak, mengubah dan mengambil(load) data.
 - b. Meliputi karakteristik datawarehouse.
3. Metadata Layer / Lapisan Metadata
 - a. File data tersimpan / Direktori
 - b. Lebih detil dari direktori data sistem, maksudnya lebih mendalam dari file data yang tersedia sebelumnya.
 - c. Ada petunjuk untuk keseluruhan warehouse dan ada petunjuk data yang dapat diakses report khusus untuk di analisis.
4. Informational access layer (lapisan akses informasi)
 - a. Akses data dan juga tool untuk laporan dan analisis.
 - b. Tools Business Intelligence masuk ke tahap ini.

Kenapa dan Untuk apa Datawarehouse?

Disini saya akan memberikan ilustrasi untuk datawarehouse. Datawarehouse itu dimiliki oleh perusahaan yang sudah besar, yang memiliki banyak cabang, data yang banyak dan tentunya struktur organisasi yang kompleks.

Mari bayangkan sebuah perusahaan yang memiliki banyak transaksi, yang memiliki banyak cabang. Tentu data-data tersebut tersebar dilokasi yang berbeda, sistem operasi yang berbeda, bahkan di basisdata (database) yang berbeda. **Nah, Lalu bagaimana seorang pimpinan/ manager mengambil sebuah keputusan?** Tentu bagi sang pembuat keputusan hanya membutuhkan akses ke semua sumber data tersebut. Kalau melakukan query di setiap masing-masing cabang tentu tidak efisien dan tidak praktis. Atau bahkan data yang dimiliki oleh perusahaan adalah data-data terbaru, bukan data-data terdahulu dari perusahaan tersebut. Dari permasalahan ini, Datawarehouse hadir sebagai solusinya.

Jika mengandalkan database OLTP untuk dilakukan query terlalu besar. Datawarehouse dibuat agar prosesnya lebih efisien. Dan selalu berkompetitif, maksudnya di zaman saat ini perusahaan sudah mengandalkan teknologi datawarehouse untuk pengambilan keputusan di perusahaan.

Tujuan Akhir menggunakan Datawarehouse

- Menyediakan data organisasi yang mudah diakses oleh manager.
- Data yang berada di datawarehouse bersifat konsisten, dan merupakan kebenaran.
- Datawarehouse merupakan tempat, dimana data yang telah digunakan di publikasikan.
- Kualitas data di datawarehouse dapat diandalkan.
-

OLAP [On-Line Analitical Process]

OLAP adalah operasi basis data (database) untuk mendapatkan dalam bentuk kesimpulan dengan menggunakan agregasi sebagai mekanisme utama. Mekanisme berupa analisis dan pengambilan keputusan.

OLTP [On-Line Transaction Process]

Proses transaksi di suatu proses bisnis seperti penjualan, pemesanan dan pembayaran. Proses yang dilakukan rutin dari hari ke hari.

Karakteristiknya dengan input/ data entry, update dan delete. Berfokus pada satu area bisnis saja, entah itu 'persediaan', 'penjualan' atau bahkan 'sumber daya'.

Tugas utamanya terlihat dari hubungan tradisional ke DBMS.

Saya akan memberikan contoh **perbedaan fitur yang mendasar** antara OLTP dan OLAP.

- User dan Orientasi Sistem : customer vs market.
- Data contents : sekarang, detil vs historical
- Design Database : ER Application vs Star Subject
- View : Sekarang, Lokal vs Perubahan dari setiap data, data terintegrasi.
- Pola Akses : Update vs hanya bisa membaca tetapi dengan bermacam sudut pandang.

	OLTP	OLAP
users	data entry	manager perusahaan
function	transaksi setiap hari	Pendukung keputusan
db design	aplikasi oriented	subjek oriented
data	sekarang, terbaru, lengkap detail	historical, ringkas, multidimensi terintegrasi
usage	repetitive	ad-hoc
access	membaca, menulis dan merubah	membaca dengan rinci <kompleks>
unit of work	pendek, transaksi yang ringkas	query kompleks
db size	mega byte	Terra byte
size_record_access	10	jutaan
size_user	ribuan	ratusan

Cara **mengembangkan datawarehouse**, melalui beberapa **metode pendekatan**.

- Pendekatan Ralph Kimball

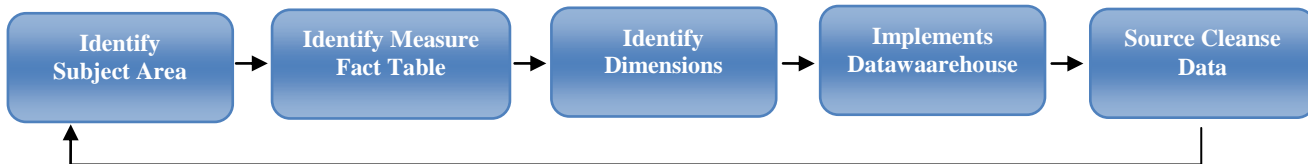
Seorang pakar dibidang datawarehouse yang membuat metode perancangan datawarehouse yang disebut dengan perancangan bottom-up. Disebut **bottom-up** karena pertama-tama data pada kantor cabang diidentifikasi dan dibuat terlebih dahulu, untuk menyediakan kapabilitas laporan dan analisis untuk proses bisnis tertentu. Data cabang ini dapat dikombinasikan untuk membuat sebuah datawarehouse.

Kelebihannya :

- Nilai Bisnis dapat dikembalikan secepat data cabang yang pertama dibuat.
- Model yang berdimensi
- Mudah dimengerti.

Kekurangan :

- Integrasi
- Mapping dari pemodelan berdimensi ke sistem yang sudah ada.
- Sulit untuk memastikan kekonsisten dimensi dari semua data cabang.



- Pendekatan Bill Inmonn

Bill Inmonn adalah orang yang menggagas pertama kali datawarehouse sebagai wadah/tempat (repository) untuk seluruh data perusahaan. Pendekatan perancangan Inmonn disebut dengan top-down, dimana datawarehouse dirancang menggunakan data model yang telah dinormalisasi.

Keuntungan :

- Perancangan top-down menghasilkan dimensional view yang konsisten untuk semua data, Karena semua data di-load dari repository terpusat/ terintegrasi.
- Top down sanggup menghadapi perubahan bisnis, membuat data dimensional cabang yang baru menjadi tugas yang mudah.

Kekurangan :

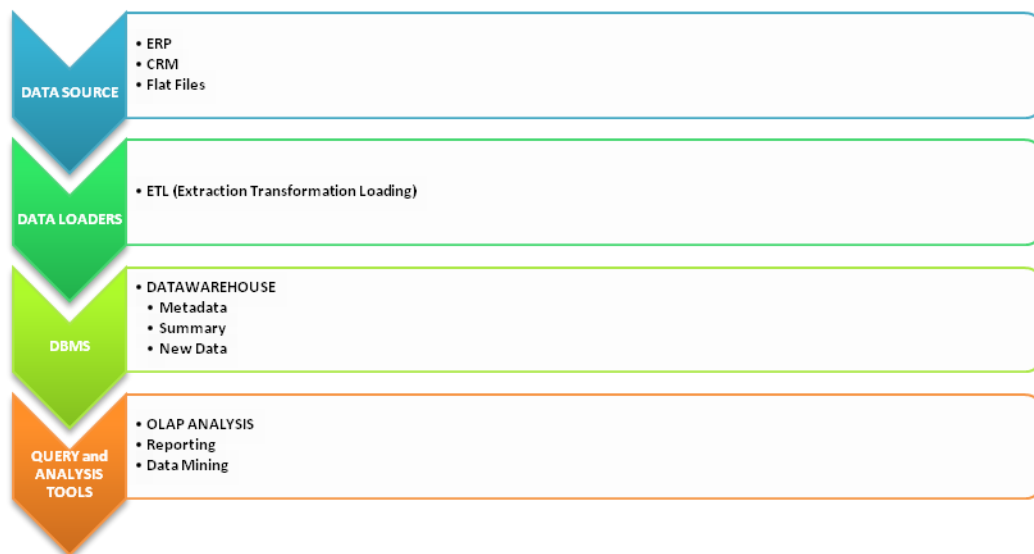
- Kesulitan dan biaya mendisain model data perusahaan.

- Feedback dari pengguna.
- Durasi waktu yang cukup lama saat dimulainya projek hingga manfaat dari datawarehouse itu sendiri.
- Tidak fleksibel terhadap perubahan kebutuhan pada saat implementasi.



- Pendekatan Hybrid

Suatu pendekatan yang menggunakan campuran dari pendekatan top down dan pendekatan bottom up.



Penyimpanan data dalam datawarehouse?

- **Dimensional**

Data transaksi dipecah menjadi fakta terutama data transaksi yang numeric atau di rubah menjadi dimensi.

Contohnya : *Data pembayaran pada EF Course, faktanya seberapa banyak anggota yang melakukan pembayaran tepat waktu. Dimensinya, seperti kode anggota, tanggal, dan jenis kursus.*

Kelebihannya :

- Datawarehouse lebih mudah dimengerti oleh pengguna.
- Data yang dicari / ditampilkan dengan lebih cepat.

Kekurangan :

- Kurang mempertahankan integritas dari fakta dan dimensi jika sistem operasi yang digunakan berbeda-beda.
- Kesulitan mengubah struktur datawarehouse, jika proses bisnis yang dilakukan perusahaan berubah.

- **Normalisasi**

Penyimpanan yang dilakukan berdasarkan aturan normalisasi data, dikelompokkan berdasarkan kategori (pelanggan, produk, keuangan,dll).

Keuntungan : Mudah untuk menambah informasi ke database.

Kekurangan : Karena banyak table yang terlibat maka akan sulit bagi pengguna untuk menggabungkan data dari sumber yang berbeda.

Kedua penyimpanan diatas dimensional dan normalisasi tidaklah berpisah secara utuh, melainkan satu sama lain saling berhubungan.

***Normalisasi** adalah upaya untuk menghindari duplikasi data dalam sebuah tabel dengan melakukan pemisahan informasi dari sebuah entitas ke dalam banyak tabel, dan menghubungkan tabel-tabel tersebut dengan sebuah key yang biasa disebut dengan foreign key.*

***Denormalisasi** adalah informasi dari sebuah entitaas yang diletakan dalam sebuah table sehingga table tersebut memiliki banyak duplikasi data.*

Evolusi perusahaan yang menggunakan Datawarehouse

Pada awalnya perusahaan menerapkan datawarehouse yang sangat sederhana. Saat ini perkembangan datawarehouse semakin canggih dan kebutuhan perusahaan pun semakin meningkat. Tingkat- tingkat kebutuhan datawarehouse:

- Database operasional offline

Tahap ini secara sederhana datawarehouse dibangun dengan menduplikasi data dari sistem operasional ke server lain. Sehingga saat membutuhkan data untuk membuat laporan, tidak akan berimbas pada sistem operasional.

- Datawarehouse Offline

Tahapan ini dimana data dari sistem operasional ke suatu bentuk umum dan data dari datawarehouse di simpan dengan struktur yang dirancang untuk membuat laporan.

- Datawarehouse Realtime

Tahap ini setiap kali melakukan proses bisnis, akan selalu terupdate di datawarehouse.

- Datawarehouse Terintegrasi

Tahapan ini hampir sama seperti tahapan realtime, hanya saja setiap kali datawarehouse melakukan update maka secara otomatis akan memberikan data baru ke data sistem operasional.

KOMPONEN DATAWAREHOUSE

- **Data Source**

Untuk membangun suatu datawarehouse yang baik data yang didapatkan harus teralokasi dengan baik. Ini melibatkan OLTP saat ini dimana informasi 'dari hari ke hari' tentang bisnis yang berjalan, tentunya dengan data historis periode sebelumnya, yang mungkin telah dikumpulkan dalam beberapa bentuk sistem lain. Sering kali data yang terbentuk bukan terbentuk database relasional, sehingga membutuhkan banyak upaya untuk mengambil data yang diinginkan.

- **Design Datawarehouse**

Proses perancangan datawarehouse sangat berhati-hati untuk memilih jenis query yang digunakan dalam datawarehouse. Tahapan ini sangat memerlukan pemahaman yang baik tentang skema database yang ingin dibuat, dan haruslah selalu aktif untuk berkomunikasi dengan pengguna. Desain adalah proses yang tidak dilakukan satu kali, melainkan berulang-ulang agar model yang dimiliki stabil. Tahap ini harus dilakukan secara berhati-hati karena model akan diisi dengan data yang jumlahnya sangat banyak, yang salah satunya dari beberapa model adalah model yang tak dapat diubah.

- **Akuisi data**

Proses perpindahan data dari sumbernya (source) ke datawarehouse. Proses ini proses yang menggunakan banyak waktu dalam proyek datawarehouse, dan dilakukan dengan software yang dikenal dengan ETL (extract,transform,load) Tools. Sekarang sudah hamper lebih kurang 60 tool yang tersebar diranah maya.

Waktu yang dibutuhkan untuk akuisisi data bisa mencapai berbulan-bulan bahkan bertahun-tahun.

- **Perubahan data tangkapan**

Pembaharuan data periodik datawarehouse dari sistem transaksi menjadi rumit karena harus diidentifikasi dari sumber data yang selalu up to date. Ini disebut dengan '*perubahan data capture*'. Ini merupakan tahapan khusus, dan sudah cukup banyak software yang beredar untuk mengatasinya. Seperti Replication Server, Publish/Subscribe, Triggers and Stored Procedures, dan Database log Analysis.

- **Pembersihan Data**

Ini biasanya dilakukan dengan akuisisi data, kalau dalam ETL ini terdapat pada bagian 'T'. Datawarehouse yang berisi data tidak benar tidak hanya tak berguna, tapi juga menyesatkan. Ide dibalik pembuatan datawarehouse adalah untuk memudahkan pengambilan keputusan, jika keputusan yang besar yang ditunjang oleh data yang tidak valid maka perusahaan mengalami resiko yang amat besar pula.

Pembersihan data suatu proses rumit yang memvalidasi dan bila perlu mengoreksi data sebelum masuk kedalam datawarehouse. Pembersihan data dapat juga disebut sebagai "*data scrubbing*" atau "*penjamin kualitas data*", proses ini harus dilakukan secara berhati-hati dan dilakukan keseluruhan terutama datawarehouse yang diambil dari perangkat yang sudah tua.

- **Data Aggregation**

Termasuk proses transformasi, datawarehouse dirancang untuk menyimpan yang amat **detil** dari tiap transaksi, untuk beberapa tingkat **aggregate** (ringkasan). Keuntungan jika data diringkas yaitu query khusus dalam datawarehouse berjalan lebih cepat. Kekurangannya adalah informasi yang didapat kurang, karena ringkasnya data yang ada pada datawarehouse. Ini harus berhati-hati karena keputusan tidak dapat dibatalkan tanpa membangun kembali datawarehouse dan mencocokkan dengan datawarehouse lain (atau sumber data lain). Paling aman digunakan oleh perusahaan yang amat besar, yang mampu membangun datawarehouse tingkat detail yang tinggi dengan biaya yang besar pula.

Metadata adalah istilah dari proses pengidentifikasian suatu atribut dan struktur dari sebuah data atau informasi. Metadata menjelaskan sebuah data itu sendiri

3 Model Datawarehouse

- **Enterprise Datawarehouse**

Semua informasi yang dikumpulkan berupa subjek, yang mencakup seluruh organisasi perusahaan.

- **Data Mart**

Sebagian data dari bagian perusahaan yang mempunyai nilai bagi pengguna. Ruang lingkupnya lebih spesifik, seperti data Penjualan atau marketing saja.

- **Virtual Warehouse**

Memantau melalui data operasional pada database. Suatu ringkasan dari data yang fleksibel, mengurangi biaya untuk pengguna yang membutuhkan. Karena tersedianya data yang siap disajikan tidak hanya untuk beberapa pengguna didalam perusahaan, akan tetapi perusahaan lain yang membutuhkan data tersebut dapat mudah untuk memperolehnya.

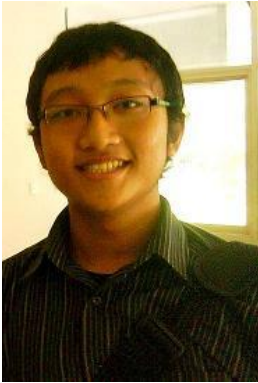
Keuntungan Datawarehouse

- Datawarehouse menyediakan model data yang bervariasi, dan tidak bergantung pada satu sumber data saja. Hal ini memudahkan pimpinan perusahaan/manager membuat laporan dan menganalisa.
- Saat me-load data ke dalam datawarehouse, data yang tidak konsisten akan diketahui dan secepatnya dirubah. Mendukung proses pembuatan laporan, agar keputusan yang diambil adalah keputusan yang benar sesuai data.
- Keamanan informasi didalam datawarehouse terjamin, karena datawarehouse selalu digunakan dan dimonitor oleh pengguna datawarehouse tersebut.
- Dalam membuat laporan tidak membuat proses transaksi yang ada menjadi lambat, karena datawarehouse terpisah dengan database operasional.
- Datawarehouse menyediakan berbagai macam bentuk laporan yang terbaru.

Kerugian Datawarehouse

- Datawarehouse tidak cocok untuk data yang tidak struktur.
- Data perlu di extract, diubah, dan di load ke datawarehouse, sehingga membutuhkan waktu (**delay**) kerja untuk datawarehouse yang belum terbentuk.
- Semakin lama masa hidup bisnis yang menggunakan datawarehouse, maka semakin banyak biaya yang dikeluarkan oleh perusahaan untuk memodifikasi teknologi datawarehouse atau perawatan berjalan datawarehouse.
- Jika data yang diambil lambat, maka data yang dimiliki di datawarehouse tidak berkualitas/ sehingga laporan tidak optimal.

Biografi Penulis



Muhammad Subhan,

I am Student in Binus University, Faculty Computer Science Major Information System, Spezialitation of Business Integece. SQL Server 05/08 my tools for developing my project.

YM : subhan_0072001@yahoo.com

Facebook : <http://www.facebook.com/muhammadsubhan>

Email : subhan@binusian.org

<http://www.subhan.blog.binusian.org>

<http://geeks.netindonesia.net/blogs/muhammadsubhan>