



Data Warehousing

excellence

Bab 1 : Pendahuluan

Referensi : Digital Press - Oracle9iR2 Data Warehousing –
2003, by Lilian Hobbs, Susan Hillson and Shilpa Lawande



Apa itu data warehouse?

- Data warehouse adalah database yang berisi data dari beberapa system operasional yang terintegrasi, teragreggrasi dan terstruktur sehingga dapat digunakan untuk mendukung analisa dan proses pengambilan keputusan dalam bisnis



Mengapa kita perlu Data Warehouse?

- Apakah anda memiliki data detail histori untuk tujuan analisa ?
- Berapa bulan dari data histori yang diperlukan untuk di-online-kan ?
- Apakah anda menyimpan level detail yang benar ?
- Apakah anda dapat menganalisa penjualan untuk tiap produk dari setiap daerah sebelum dan sesudah struktur pelaporan penjualan di-reorganisasi ?
- Data warehouse dibangun dengan tujuan memecahkan semua persoalan tersebut.



Tinjauan Historis

- tahun 1970-an aplikasi komersial pertama dibangun => dibangun pada computer mainframe
- tahun 70-an dan awal 1980 => minicomputers seperti Digital's PDP-11 dan VAX 11/780 membuat harga hardware turun drastis
 - munculnya database relasional
 - tahun 1979, Oracle sebagai software komersial pertama
 - kemunculan Personal Computer (PC), komputasi beralih dari mainframes ke system client/server
- Tahun 1980-an
 - system OLTP dibangun dengan menggunakan system relasional



Era kemunculan data warehouse

- Dan pada akhir tahun 1980-an dan awal tahun 1990-an, mulai dilakukan tinjauan bisnis secara menyeluruh, sehingga muncullah enterprise data warehouses untuk pertama kalinya
- Istilah datawarehouse dikenalkan pertama kali oleh Bill Inmon



Data warehouse melihat ke dalam data untuk menemukan informasi

- Aktivitas utamanya adalah query, atau membaca data.
- Aktivitas perubahan (update) hanya terjadi pada saat data di-load ke dalam warehouse.
- Materialized views ditambahkan pula sejak Oracle 8i untuk meningkatkan performansi pada saat melakukan agregasi dan summary pada data.



Analisa yang diperoleh dari Data Warehouse

- Tidak cukup hanya dengan mengetahui keuntungan yang diperoleh tahun ini, seorang analis juga perlu mengetahui keuntungan untuk setiap periode waktu tertentu untuk setiap produk untuk setiap area geografis tertentu.
- Seorang analis perlu membandingkan penjualan bulan ini dengan bulan yang sama pada tahun sebelumnya untuk setiap toko.



Data warehouse membutuhkan desain database yang berbeda

- Data warehouses didesain untuk kita bisa melakukan query secara cepat.
- Informasi diturunkan dari data lain, dilakukan rolling up data untuk dijadikan ringkasan (summaries), dilakukan operasi drilling down untuk mendapatkan informasi lebih detail, atau melihat pola yang menarik atau melihat trend (kecenderungan).
- Desain yang sesuai adalah dengan menggunakan skema star yang diperkenalkan mula pertama oleh Ralph Kimball,



Pendekatan dimensional pada skema Star

- Mengorganisasi data ke dalam table fakta dan table-tabel dimensi.
- Dimensi-dimensi dalam skema star digunakan untuk melakukan analisa dan mengorganisasi data.
- Desain logika ditampilkan dalam bentuk representasi fisik yang akan mengoptimalkan performansi dan manajemen.
- Dalam hal ini, Tabel, constraints, indexes, dan partisi didefinisikan.



Data warehouses sebagai system terpisah

- Mengapa kita perlu mengkopi data dari satu system ke system lain untuk membangun warehouse?
- Ini dikarenakan system operasional tidak berisi data histori (data time series – atau data yang mengandung dimensi waktu), sehingga data tidak available untuk dianalisa. Juga skema data tidak didesain untuk keperluan queri kecerdasan bisnis (business intelligence queries).



Data warehouse dibangun dari system operasional

- Membangun warehouse melibatkan ekstraksi data dari system operasional, dikombinasikan dengan informasi tambahan dari pihak ketiga, mentransformasinya ke dalam format yang seragam dan di-load ke dalam database.
- 80% waktu digunakan untuk proses extraction/transformation/load (ETL) : menempatkan data, menulis program untuk ekstraksi; memfilter, dan membersihkan data; mentransformasi data ke dalam skema pengkodean standart; kemudian me-load data ke dalam data warehouse.



Contoh format yang tidak sama

- Produk dengan kode "1234" disebut dengan "roti" sementara di tempat lain "1234" disebut "kue"
- kode produk direpresentasikan dengan karakter yang dipisahkan dengan tanda '-'(xxx-xx-xxx) sedangkan ditempat lain direpresentasikan dengan karakter yang dipisahkan dengan tanda spasi (xxx xx xxx).
- untuk atribut gender pada table customer digunakan nilai "0" atau "1." Sedangkan di system lain, digunakan "M" atau "F."
- Perbedaan mata uang yang tidak sama untuk data penjualan, digunakan rupiah, dollar atau euro.



Data mart

- Data marts adalah subyek spesifik atau aplikasi spesifik dari data warehouses yang berisi data untuk satu line bisnis tertentu semisal penjualan atau pemasaran.
- Ruang lingkup data mart lebih kecil, data didapatkan dari sumber yang lebih sedikit, dan waktu implementasinya juga lebih pendek.

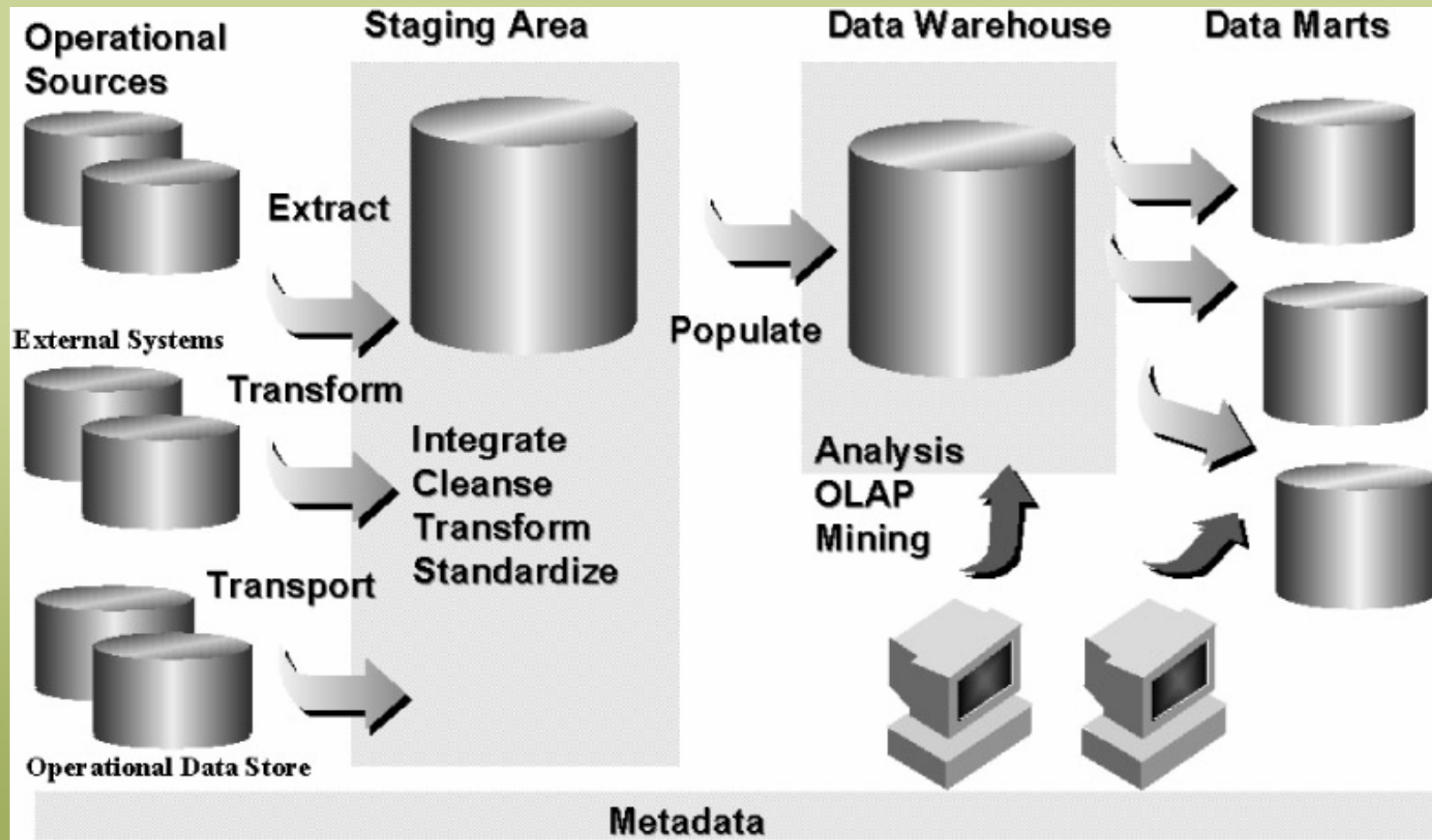


Tipe data marts

- Data marts bisa bersifat dependent atau independent, tergantung pada sumber informasinya.
- Sumber informasi untuk data mart yang bersifat dependent adalah data warehouse itu sendiri.
- Sedangkan data mart bersifat independent jika tidak terdapat data warehouse, dan data diekstrak secara langsung dari system operasional.



Enterprise data warehouse dengan dependent data marts





Reporting, query, dan tools untuk analisa

- Di tahun 1998 Oracle merilis aplikasi yang dapat dijalankan di Web
- Kemampuan untuk mempublish report di Web dapat membuat informasi selalu tersedia dimanapun tanpa perlu diinstall terlebih dahulu.
- Ini membuat pegawai, rekanan, dan customer secara realtime dapat mengakses informasi penting



Fitur-fitur Data warehousing dalam Oracle

- Bitmap join indexes
- Partisi
- proses ETL (Extract-Transform-Load):
 - mekanisme untuk mengidentifikasi perubahan baris dalam Oracle database
 - Tabel eksternal
 - Perintah SQL baru yaitu Merge
 - eksekusi parallel
- Materialized views



Tantangan dalam membangun data warehouse

- Manage- warehouse
 - Warehouse perlu di-tuning untuk mendapatkan performansi yang lebih baik
- Perlunya metadata
- Volume data yang meningkat
- Ketersediaan (availability) yang tinggi
- Lebih banyak users
- Tipe-tipe aplikasi yang baru



Data warehouse di masa depan

■ Real-time data warehouses

- Data warehouse tidak perlu di-update secara periodic dalam system batch, karena jika suatu transaksi sudah komit di system OLTP, maka pada saat itu juga akan langsung tersedia dalam data warehouse

■ Bukan merupakan system terpisah

- Suatu hari nanti hanya diperlukan satu database tunggal untuk OLTP dan data ware-housing. Tidak perlu lagi memisahkan database untuk OLTP, ODS, data warehouse, dan data marts.



Ringkasan

- Pada bab ini kita telah mempelajari latar belakang dari kemunculan data warehouses dan data marts, memahami secara garis besar fungsionalitas dari Oracle 9i, beberapa tantangan yang dihadapi oleh pengembang warehouse di masa depan.
- Dan tibalah saatnya, sekarang kita akan menggunakan semua teknologi ini untuk membangun dan mengakses data warehouse.