

LECTURE NOTES

7023T Advanced Database System

Session 01

Introduction to Data Warehouse

LEARNING OUTCOMES

1. Peserta diharapkan mampu memahami perbedaan antara database transaksional dengan data warehouse
2. Peserta diharapkan mampu mengidentifikasi karakteristik data warehouse
3. Peserta diharapkan mengerti business intelligence sebagai salah satu aplikasi dari data warehouse untuk mendukung proses pengambilan keputusan.
4. Peserta diharapkan menjelaskan perbedaan pendekatan Inmon dan Kimball dalam pengembangan data warehouse.

OUTLINE MATERI (Sub-Topic):

1. Transactional database and Data warehouse
2. Data warehouse characteristics
3. Business intelligence
4. Inmon and Kimball approach for data warehouse development

Background

Selama satu dekade terakhir, perkembangan teknologi yang signifikan dalam hal kemampuan penyimpanan dan pengolahan komputer, jaringan komputer, serta internet, memungkinkan bisnis dan organisasi pemerintahan mengumpulkan data dalam jumlah yang sangat besar. Data historis dan data transaksi ditambah dengan kemampuan analisis yang tepat dapat menjadi aset penting bagi pengambilan keputusan dan proses bisnis lainnya yang pada akhirnya akan meningkatkan untuk keuntungan mereka.

Pengelolaan data berukuran besar menimbulkan tantangan baru yang terkait dengan data, manusia, dan proses analisa. Tantangan dalam hal data adalah volume yang besar dan jenis data yang semakin bervariasi. Mengintegrasikan data yang berasal dari departemen internal, konsumen, pemerintah, dan mitra bisnis tidaklah mudah karena data tersebut akan memiliki format yang berbeda, baik terstruktur seperti database maupun tidak terstruktur seperti dokumen teks, email, dll. Tantangan yang terkait dengan manusia diakibatkan dari globalisasi informasi yang memungkinkan interaksi antar personal menembus batasan geografis. Hal ini memunculkan potensi permasalahan yang diakibatkan oleh perbedaan sistem ekonomi, budaya, dan letak geografis. Volume data yang sangat besar dan tingkat variasi yang tinggi diistilahkan sebagai “island of information” atau “information silos”. Analisis terhadap data semacam itu tidaklah mudah, karena seringkali melibatkan banyak personil dan pengolahan data yang berasal dari banyak sumber berbeda.

Untuk mengatasi tantangan tersebut, komunitas teknologi informasi mengembangkan sistem baru yang dapat mengintegrasikan data yang berasal dari sumber berbeda. Sistem itu disebut Data Warehouse (DW). Data hasil integrasi pada data warehouse memudahkan proses analisis untuk menemukan pola dan tren, membuat laporan, dan dapat dimanfaatkan untuk mendukung proses pengambilan keputusan dengan cara memberikan informasi yang tepat kepada orang yang tepat pada waktu yang tepat. Hal tersebut pada akhirnya akan meningkatkan kalangan bisnis dalam menjawab kebutuhan konsumen yang semakin kompleks serta bersaing dalam dunia bisnis yang semakin kompetitif.

Tujuan dari data warehouse adalah meningkatkan kualitas dan akurasi informasi bisnis dan menyampaikannya kepada pengguna dalam bentuk yang mudah diinterpretasi dan dapat diakses dengan mudah. Sifat data warehouse bertolak belakang dengan data transaksional yang lebih berorientasi kepada proses atau *On-Line Transactional Processing* (OLTP), data warehouse lebih berorientasi kepada analisis atau *On-Line Analytical*

Processing (OLAP). Kebutuhan informasi dari setiap level manajemen di suatu organisasi seperti digambarkan oleh gambar 1, tidaklah sama. Contoh, pada sebuah usaha restoran cepat saji maka level *Strategic Management* akan memerlukan informasi seperti bagaimana pola belanja konsumennya, apakah terdapat kecenderungan konsumen yang membeli produk X akan membeli juga produk Y. Pihak yang berada pada level *Tactical Management* akan membutuhkan informasi seperti bagaimana trend penjualan produk X pada tahun lalu. Informasi yang dibutuhkan oleh pihak *Operational Management* mungkin hanya berupa berapa jumlah produk X dan Y yang terjual pada minggu terakhir.



Gambar 1. Level manajemen pada organisasi (Kimball, 2008)

Transactional Database

Online Transactional Processing (OLTP) atau seringkali disebut sebagai *Operational System* adalah sistem yang digunakan untuk menjalankan bisnis inti perusahaan sehari-hari. Sistem dibuat agar roda bisnis berputar, OLTP mendukung proses dasar bisnis perusahaan. Sistem ini biasanya mendapatkan data transaksional dan menyimpannya ke dalam database. Setiap transaksi memproses informasi tentang sebuah entitas seperti sebuah perintah, satu faktur, atau satu pelanggan. Di sisi lain, *Online Analytical Processing (OLAP)* dirancang dan dibangun untuk sistem pendukung pengambilan keputusan bisnis yang tidak dimaksudkan untuk menjalankan proses bisnis inti. OLAP digunakan untuk melihat bagaimana bisnis berjalan, dan kemudian membuat keputusan strategis untuk meningkatkan bisnis. Sistem pendukung keputusan yang dikembangkan untuk mendapatkan informasi strategis dari database, yang berbeda terhadap sistem OLTP yang dirancang untuk menempatkan data ke dalam database. Sistem pendukung keputusan dikembangkan untuk menyediakan informasi strategis agar dapat memonitor bahwa roda bisnis tetap berputar sesuai yang diharapkan.

Data Warehouse Concepts

Penggunaan data warehouse secara real-time serta sistem pengambilan keputusan atau *Decision Support System* (DSS) dan *Business Intelligence* (BI) merupakan salah satu pendekatan yang dapat membantu proses pengambilan keputusan bisnis yang kritis secara cepat. Informasi yang tersimpan di data warehouse dapat disajikan dalam bentuk yang mudah diinterpretasi memungkinkan perusahaan dapat melihat keadaan bisnis terkini dengan cepat dan membantu identifikasi permasalahan yang ada. Hal tersebut merupakan langkah awal yang penting dalam merumuskan alternatif solusi dari permasalahan tersebut. Selain itu pelanggan dapat memperoleh informasi mengenai produk, jasa, dan informasi terkait lainnya. Terbukanya akses kepada pelanggan dapat memberikan keunggulan kompetitif yang cukup signifikan terhadap kompetitor.

Data warehouse dapat didefinisikan sebagai kumpulan data yang dihasilkan untuk mendukung proses pengambilan keputusan. Data warehouse juga merupakan penyimpanan data saat ini dan data historis yang sangat diperlukan oleh para manajer di dalam organisasi. Sebuah data warehouse adalah kumpulan data berorientasi subyek, terintegrasi, varian terhadap waktu, serta merupakan kumpulan data *non-volatile* (tidak berubah) untuk mendukung proses pengambilan keputusan para manajemen. Secara fungsional data warehouse dapat didefinisikan sebagai berikut:

- Merupakan lingkungan informasi yang memberikan pandangan terpadu dan total suatu perusahaan
- Menjadikan informasi saat ini dan histori mudah diperoleh dan selalu tersedia untuk pengambilan keputusan
- Melakukan transaksi pada sistem dukungan keputusan tanpa mengganggu sistem operasional
- Menjadikan informasi organisasi yang lebih konsisten di semua unit organisasi Menyajikan secara fleksibel dan interaktif dari berbagai sumber data bagi informasi strategis

Data Warehouse Characteristics

Terdapat empat karakteristik utama dari data warehouse, yaitu:

- a. **Berorientasi subjek:** data disusun menurut subyek, berisi informasi yang penting untuk proses pengambilan keputusan. Informasi yang disimpan dalam database biasanya dikelompokkan berdasarkan subyek tertentu, seperti pelanggan, produk, gudang, dan lain-lain. Data berorientasi subjek memungkinkan pengguna untuk menentukan tidak saja bagaimana mereka menjalankan bisnis tapi juga mengapa.
- b. **Terintegrasi:** Integrasi sangat berhubungan erat dengan pengertian orientasi subjek dari data warehouse yang menempatkan data dari berbagai sumber yang berbeda ke dalam format yang konsisten. Untuk melakukannya, kita harus berurusan dengan penamaan data yang berbeda dan kesenjangan antar satuan ukuran. Sebuah data warehouse harus dibuat sedemikian rupa sehingga benar-benar terintegrasi.
- c. **Varian terhadap waktu (time variant):** Pada data warehouse, data yang tersimpan adalah data historis dalam kurun waktu tertentu, bukan data terkini. Oleh karena itu data yang tersimpan mengandung keterangan waktu, misalnya tanggal, minggu, bulan, catur wulan, dan sebagainya.
- d. **Non-volatile:** Berbeda dengan database operasional yang banyak melakukan operasi pembaharuan, penambahan, atau penghapusan data yang pada akhirnya akan mengubah isi dari database sedangkan pada data warehouse hanya terdapat dua kegiatan manipulasi data yaitu loading (mengambil data) dan akses data.

Bussiness Intelligence

Kemampuan analitik dari sistem data warehouse diistilahkan sebagai *Business Intelligence* atau BI. Berikut beberapa definisi dari *business intelligence* menurut pandangan beberapa ahli:

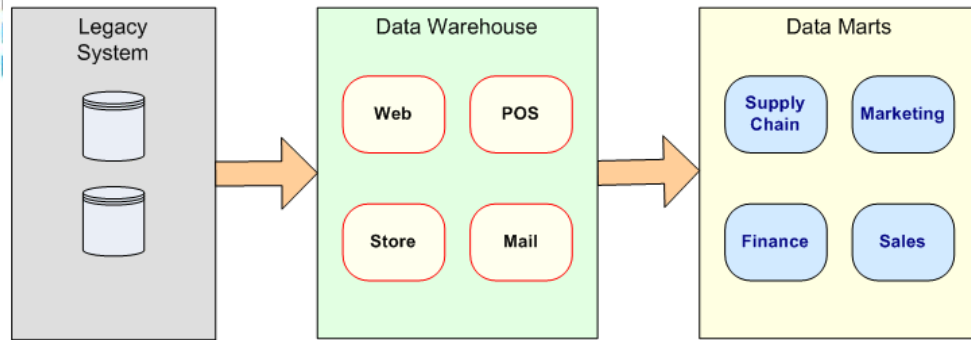
- a. Business intelligence adalah sebuah konsep atau metode tentang bagaimana cara untuk meningkatkan kualitas pengambilan keputusan bisnis yang didasari oleh sistem berbasis data. Business intelligence sering disamakan dengan *briefing books, report and query tools*, serta sistem informasi eksekutif. Business intelligence adalah sistem pendukung pengambilan keputusan yang berbasis data (DJ Powers, 2002).

- b. Business intelligence merupakan salah satu cara menyimpan, mengumpulkan, mengorganisasi, meringkas data dan membentuk ulang serta menyediakan sebuah informasi, mulai dari bentuk data aktifitas bisnis internal maupun eksternal perusahaan terutama aktifitas para pesaing yang mudah dianalisis untuk berbagai keperluan manajemen (Edward David, 2000)
- c. Business intelligence adalah alat untuk menganalisis yang akan digunakan sebagai konsolidasi data, analisa data, menyimpan dan mengakses data dengan tujuan membantu memberikan keputusan seperti perangkat lunak untuk query database dan pelaporan, serta alat untuk menganalisis data multidimensi dan data mining (C. Loudon, 2007).

Berdasarkan pengertian business intelligence tersebut dapat ditarik kesimpulan umum bahwa business intelligence adalah proses melakukan ekstraksi data operasional perusahaan yang kemudian dikumpulkan ke dalam data warehouse. Data pada data warehouse selanjutnya diproses untuk digunakan dalam proses analisis dan mining. Hasil dari proses analisis dan mining dapat memberikan kecenderungan atau trend dari data yang dapat digunakan dalam mendukung pengambilan keputusan bisnis.

Pendekatan dalam Pengembangan Data Warehouse

Terdapat dua pendekatan dalam pembuatan data warehouse, yaitu pendekatan *top-down* dan pendekatan *bottom-up* (Ponniah, 2010:p.29). Bill Inmon adalah salah satu pendukung terdepan dari pendekatan *top-down* seperti yang diilustrasikan pada gambar 2. Inmon mendefinisikan data warehouse sebagai repositori terpusat untuk seluruh organisasi. Pada pendekatan ini, data warehouse menyimpan data pada tingkat terendah dari *granularity* yang didasarkan pada model data dinormalisasi. Dalam visi Inmon, data warehouse adalah *Corporate Information Factory* (CIF) yang menyediakan kerangka logis yang memungkinkan pemanfaatan *Business Intelligence* (BI) untuk organisasi. Operasi bisnis menyediakan data untuk mendorong CIF. Data warehouse terpusat akan menyediakan kebutuhan untuk *dependent data mart* yang mungkin dirancang berdasarkan model data multidimensi.



Gambar 2. Pengembangan data warehouse berdasarkan pendekatan Inmon

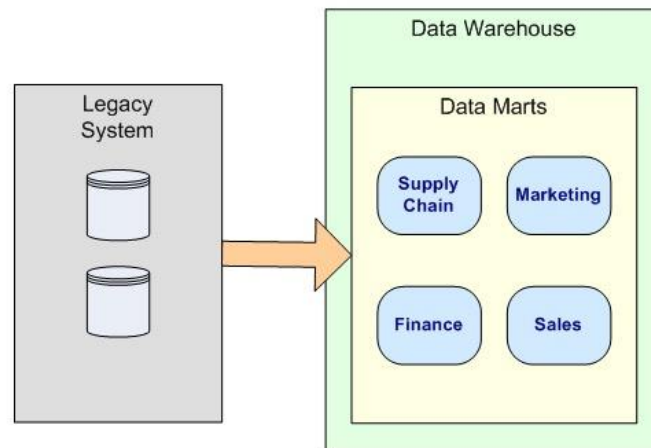
Keuntungan dari pendekatan *top-down* adalah:

- Bukan penggabungan dari *data mart-data mart* yang berbeda.
- Tempat penyimpanan data hanya satu, terpusat.
- Aturan dan kontrol dilakukan secara terpusat.
- Dapat melihat hasil cepat jika diimplementasikan secara iteratif.

Sedangkan kekurangan dari pendekatan *top-down* adalah:

- Membutuhkan waktu pengembangan yang lebih lama walaupun dengan menggunakan metode iteratif.
- Memiliki resiko kegagalan yang sangat tinggi.
- Membutuhkan keterampilan lintas fungsional yang sangat tinggi.
- Pengeluaran akan besar jika tidak terdapat pembuktian dari konsep.

Ralp Kimball, merupakan salah satu pendukung terdepan untuk pendekatan *Bottom-Up* seperti yang diilustrasikan pada gambar 3. Dalam pendekatan ini *data mart* dibuat pertama kali untuk memberikan analisis dan kemampuan pelaporan untuk subjek bisnis yang spesifik berdasarkan pada model data multidimensi. *Data mart* berisi data pada tingkat terendah dari *granularity* dan juga sebagai ringkasan, tergantung pada kebutuhan untuk analisis. *Data mart* ini kemudian akan digabungkan menjadi suatu data warehouse.



Gambar 3. Pengembangan data warehouse berdasarkan pendekatan Kimball

Keuntungan dari pendekatan *bottom-up*:

- Cepat dan mudah untuk diimplementasikan.
- Dapat memberikan keuntungan atas investasi dengan suatu konsep yang dapat dibuktikan.
- Resiko kegagalan kecil.
- Dapat melakukan penjadwalan supaya *data mart* yang penting dibuat terlebih dahulu.

Kerugian dari pendekatan *bottom-up*:

- Setiap *data mart* memiliki pandangan yang berbeda akan data.
- Dapat terjadi redundansi data pada setiap *data mart*
- Data tidak konsisten.

SIMPULAN

- Sistem transaksional tidak sesuai untuk kebutuhan analisis dan pembuatan laporan
- Data warehouse adalah koleksi data operasional yang bersifat *subject-oriented, time-variant, non-volatile*.
- *Data mart* mendukung kebutuhan pengambilan keputusan yang dibutuhkan oleh grup atau departemen pada suatu organisasi
- *Business intelligence* adalah pemanfaatan informasi untuk mendukung pengambilan keputusan.

DAFTAR PUSTAKA

1. Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit*. John Wiley & Sons.
2. Kimball, R., & Ross, M. (2011). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons.
3. Inmon, W. H. (2005). *Building the Data Warehouse*. John Wiley & Sons.