

Data Warehouse

FT – UPI YAI

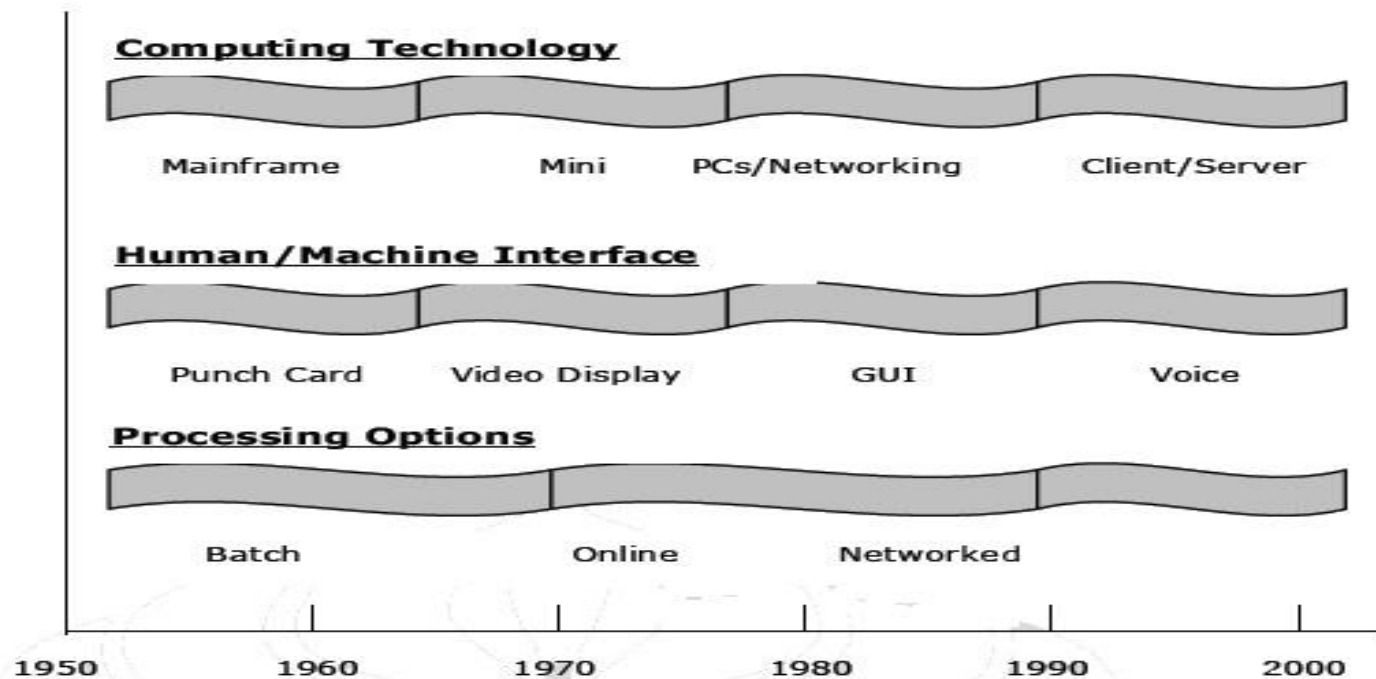
Jesa Ariawan, MTI

Latar Belakang

- Diawali dengan penggunaan komputer untuk kegiatan bisnis pada tahun 1960 an.
- Sekitar tahun 1990, penggunaan komputer untuk kegiatan bisnis semakin berkembang pesat hingga tuntutan untuk analisa pengambilan keputusan.
- Dibutuhkan *strategic information* sebagai informasi penting bagi kelangsungan hidup perusahaan. Karakteristik *Strategic Information* tersebut adalah :
 - *Integrated* : Harus memiliki suatu pandangan yang menyeluruh dari perusahaan.
 - *Data Integrity* : Informasi yang disajikan harus akurat dan memenuhi aturan serta proses bisnis perusahaan.
 - *Accessible* : Informasi harus dapat diakses dengan mudah dan dapat dianalisa.
 - *Credible* : Setiap faktor yang terkait dengan kegiatan bisnis harus memiliki nilai.
 - *Timely* : Informasi harus selalu tersedia dalam *time frame* yang telah ditentukan.

Latar Belakang

- Sampai pada era tahun 2000an, perkembangan Teknologi Informasi telah menunjukkan pengolahan data informasi yang semakin mudah dan semakin cepat. Menurut Poe (1998), perkembangan Teknologi Informasi dibagi dalam 3 bagian yaitu *computing technology*, *human/machine interface*, dan *processing option*.



Latar Belakang

- Banyak perusahaan yang telah mengumpulkan data dalam jumlah yang sangat besar berkaitan dengan operasional bisnisnya tetapi data tersebut menjadi kuburan data (*data tombs*) karena tidak dapat ditransformasikan ke dalam informasi yang lebih mudah dianalisa untuk pengambilan keputusan. Kondisi tersebut menjadikan krisis informasi.
- Dibutuhkan konsep gudang data untuk memenuhi *strategic information* perusahaan dalam melakukan analisa bisnis dalam pengambilan keputusan.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

1. *Ad Hoc Reports*

Ad hoc reports merupakan tahapan paling awal dalam sejarah *Decision Support System*. *User* yang membutuhkan suatu laporan akan mengirimkan permintaan ke departemen *Information Technology* (IT). Selanjutnya departemen IT akan membuat suatu program khusus yang dapat menghasilkan *ad hoc reports*. Setiap program hanya dapat menghasilkan satu jenis laporan, sehingga jika ada permintaan laporan yang berbeda, departemen IT akan membuat program yang lain lagi.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

2. *Special Extract Programs*

Tahapan ini merupakan usaha departemen IT untuk mengantisipasi jenis laporan yang diperkirakan akan diminta secara berkala. Departemen IT akan membuat program yang sesuai dan menjalankannya secara periodik untuk mengekstrak data dari berbagai sumber aplikasi. Selanjutnya departemen IT akan membuat dan menyimpan file data yang telah diekstrak tersebut untuk memenuhi kebutuhan permintaan laporan. Jika ada permintaan laporan yang tidak sesuai, maka departemen IT akan membuat program lagi untuk memenuhi kebutuhan tersebut.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

3. *Small Applications*

Pada tahapan ini, departemen IT mengadopsi proses ekstraksi data. Departemen IT akan membuat aplikasi sederhana berdasarkan file data yang telah diekstrak. User dapat menetapkan parameter untuk setiap laporan. Program akan mencetak informasi kedalam bentuk laporan berdasarkan parameter yang telah ditentukan tersebut. Beberapa aplikasi bahkan memperkenalkan user untuk melihat informasi melalui layar monitor secara *online*.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

4. *Information Centers*

Pada awal tahun 1970an, beberapa perusahaan besar membuat pusat informasi (*information center*). Pusat informasi ini merupakan tempat dimana user dapat mengirimkan permintaan ad hoc reports atau menampilkan informasi melalui monitor. Departemen IT menyediakan pusat informasi ini untuk membantu user memenuhi kebutuhannya akan informasi tertentu.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

5. *Decision Support Systems*

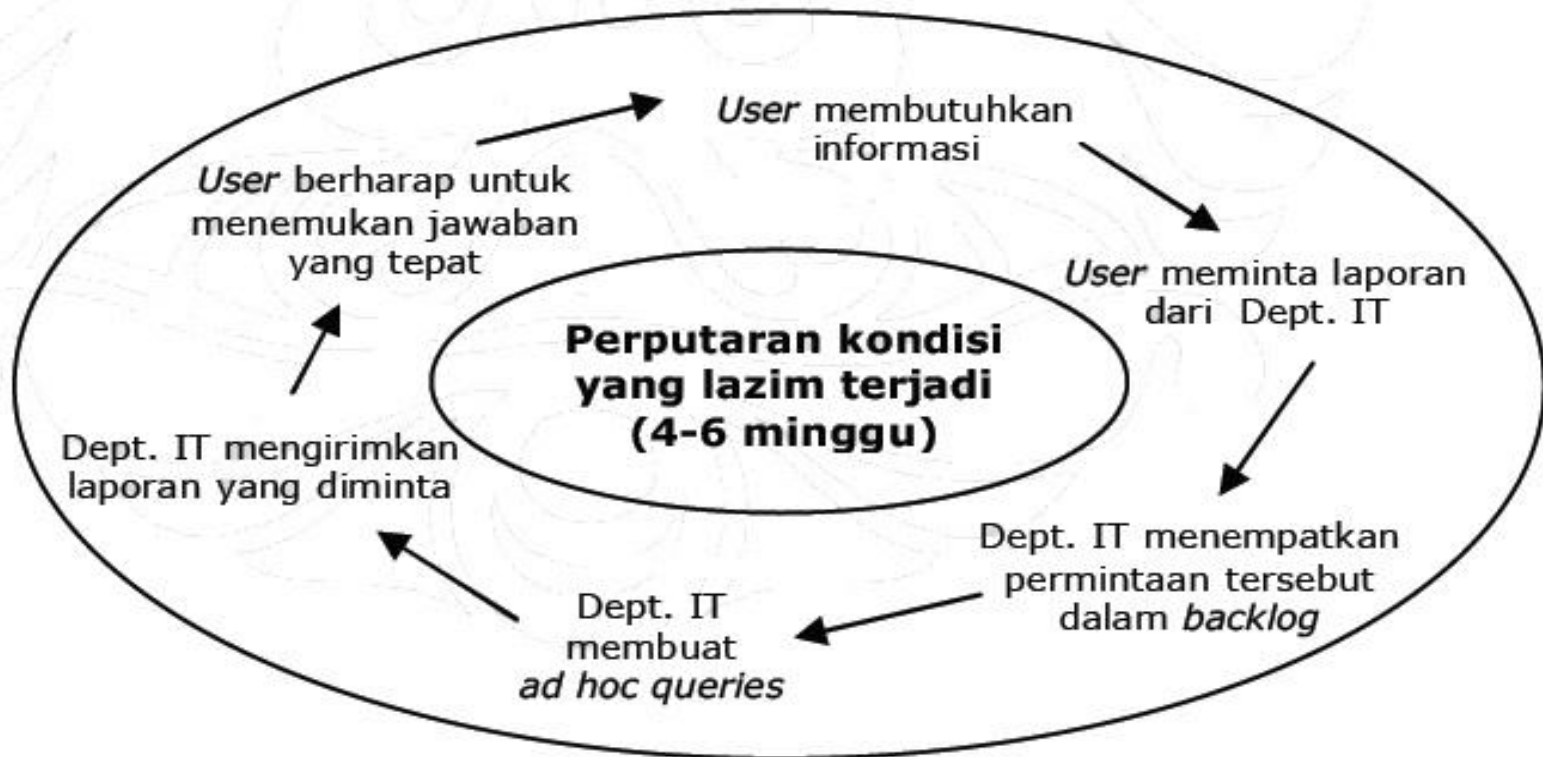
Pada tahapan ini perusahaan mulai membangun beberapa sistem yang cukup kompleks untuk memenuhi kebutuhan *strategic information*. Sama halnya dengan sistem-sistem terdahulu, sistem ini dibangun berdasarkan file data yang telah diekstrak. Sistem ini digunakan untuk analisa yang lebih fleksibel, merupakan menu-driven dan menyediakan informasi secara online serta mampu mencetak laporan.

Beberapa Tahapan menyediakan *strategic information* untuk analisa bisnis

6. *Executive Information System*

Tahapan ini merupakan usaha untuk membawa *strategic information* langsung ke komputer milik eksekutif. Sistem ini akan menampilkan informasi setiap harinya dan mampu menyediakan berbagai macam laporan.

Strategic Information dengan model Ad Hoc



*Strategic Information dengan model Ad Hoc
Tidak Mampu menyediakan Analisa secara Efektif*

Faktor Penyebab Ketidakmampuan penyediaan *Strategic Information* yang Efektif

- Permintaan *ad hoc reports* yang diterima oleh departemen IT **terlalu banyak**, sehingga menjadi **overload**. Dengan jumlah sumber daya yang terbatas, departemen IT menjadi tidak mampu memberikan respon dengan cepat.
- Permintaan yang ada tidak hanya banyak, tetapi juga **sering berubah** dari waktu ke waktu.
- User membutuhkan beberapa laporan tambahan untuk dapat memahami laporan sebelumnya, sehingga mereka **cenderung meminta berbagai macam kemungkinan** yang terjadi. Kondisi ini menyebabkan departemen IT menjadi *overload*.
- User menjadi **sangat bergantung** pada departemen IT untuk menyediakan informasi. Hal ini dikarenakan mereka tidak dapat mengakses informasi tersebut secara interaktif.
- Departemen IT **tidak dapat menciptakan lingkungan yang fleksibel** dan kondusif untuk proses analisa dan pengambilan keputusan.

Decision Support System (DSS)

- Untuk menyediakan *strategic information* yang efektif dibutuhkan wadah data yang dirancang untuk analisa yang lebih dikenal sebagai *Decision Support System*.

Definisi DSS :

- Menurut Finlay (1994) mendefinisikan DSS kurang lebih sebagai sebuah sistem **berbasis komputer** yang membantu dalam proses pengambilan keputusan.
- Menurut Turban (1995) mendefinisikan secara lebih spesifik dengan, sesuatu yang **interaktif, flexible dan dapat menyesuaikan diri (*adaptable*)** dari sistem informasi berdasarkan komputer, khususnya pengembangan untuk mendukung pemecahan masalah dari non-struktur management, untuk meningkatkan pengambilan keputusan. Dengan menggunakan data yang mendukung antar lini maka mudah digunakan dan memberikan wawasan untuk sang pengambil keputusan.
- Menurut Sprague dan Carlson (1982), DSS adalah sistem berdasarkan komputer **interaktif** yang membantu para pengambil keputusan menggunakan data dan model-model untuk memecahkan masalah yang tak terstruktur (unstructured problem).

Decision Support System (DSS) vs Operational System

	<i>Decision Support System</i>	<i>Operational System</i>
Isi Data	Kumpulan, perolehan, ringkasan	Nilai Terkini
Struktur Data	Optimal untuk query yang kompleks	Optimal untuk transaksi
Frekuensi Akses	Sedang sampai rendah	Tinggi
Tipe Akses	Read	Read, update, delete
Kegunaan	Ad hoc, random, heuristik	Pengulangan
Response Time	Beberapa detik sampai menit	Satu detik atau beberapa detik
Pengguna	Relatif sedikit	Berjumlah banyak

- Sedangkan *Data Warehouse* merupakan gudang data untuk analisa bisnis yang mendukung *Decision Support System* dalam pengambilan keputusan.

DEFINISI *DATA WAREHOUSE*

- Definisi DW menurut Turban

Data Warehouse (DW) adalah kumpulan data yang dihasilkan untuk **mendukung pengambilan keputusan**. *Data Warehouse* juga merupakan gudang data untuk masa kini dan masa lampau yang sangat bermanfaat bagi kalangan management sebuah organisasi.

- Definisi DW menurut Carlo Vercelis

Data Warehouse adalah gudang utama untuk **penyediaan data** bagi pengembangan *Business Intelligence* dan sistem yang **mendukung pengambilan keputusan**.

- Definisi DW menurut Barry Devlin

Data Warehouse adalah suatu tempat penyimpanan **data yang lengkap dan konsisten**, yang didapatkan dari berbagai sumber dan dibuat untuk memenuhi kebutuhan *end users* dalam konteks bisnis.

- Definisi DW menurut Bill Inmon

Data Warehouse adalah koleksi data yang bersifat *subject oriented, integrated, non-volatile* dan *time variant* yang mendukung pengambilan keputusan.

DEFINISI *DATA WAREHOUSE*

- Definisi DW menurut Sean Kelly

Data Warehouse adalah kumpulan data yang memiliki beberapa karakteristik, yaitu *separate, available, integrated, time stamped, subject oriented, non volatile* dan *accessible*.

- Definisi DW menurut Vidette Poe

Data warehouse merupakan database yang bersifat **analisis dan read only** yang digunakan sebagai **fondasi** dari sistem penunjang keputusan.

- Definisi DW menurut Paul Lane

Data Warehouse merupakan database relasional yang didesain lebih kepada **query dan analisa** dari pada proses transaksi, biasanya mengandung *history* data dari proses transaksi dan bisa juga data dari sumber lainnya. *Data warehouse* **memisahkan beban kerja** analisis dari beban kerja transaksi dan memungkinkan organisasi **menggabung/konsolidasi data** dari berbagai macam sumber.

FUNGSI *DATA WAREHOUSE*

Fungsi *Data Warehouse* secara ringkas terdiri dari :

- Menyediakan suatu pandangan terhadap perusahaan yang **menyeluruh dan terintegrasi**.
- Menyediakan **informasi terkini dan historis** dari perusahaan untuk proses pengambilan keputusan.
- Memungkinkan proses pengambilan keputusan yang **tidak terbatas** pada *operational system*.
- Memberikan informasi secara **konsisten** kepada perusahaan.
- Menyediakan sumber *strategic information* yang **fleksibel dan interaktif**.

KEUNTUNGAN *DATA WAREHOUSE*

Menurut Ramelho, keuntungan data warehouse terdiri dari :

- Data diorganisir dengan baik untuk *query* analisis dan sebagai bahan untuk pemrosesan transaksi. (integrated)
- Perbedaan diantara struktur data yang heterogen pada beberapa sumber yang terpisah dapat diatasi. (standard)
- Aturan untuk transformasi data diterapkan untuk memvalidasi dan mengkonsolidasi data apabila data dipindahkan dari database OLTP ke *data warehouse*. (konsisten)
- Masalah keamanan dan kinerja bisa dipecahkan tanpa perlu mengubah sistem produksi. (safety)

KEUNTUNGAN *DATA WAREHOUSE*

Menurut Sean Nolan, keuntungan Data Warehouse antara lain :

- Kemampuan untuk mengakses data yang besar
- Kemampuan untuk memiliki data yang konsisten
- Kemampuan kinerja analisa yang cepat
- Mengetahui adanya hasil yang berulang-ulang
- Menemukan adanya celah pada *business knowledge* atau *business process*.
- Mengurangi biaya administrasi
- Memberi wewenang pada semua anggota dari perusahaan dengan menyediakan kepada mereka informasi yang dibutuhkan agar kinerja bisa lebih efektif.

Data Operasional vs Data Warehouse

Data Operasional	Data Warehouse
Dirancang berorientasi hanya pada aplikasi dan fungsi tertentu	Dirancang berdasar pada subjek-subjek tertentu (utama)
Focusnya pada desain database dan proses	Focusnya pada pemodelan data dan desain data
Berisi rincian atau detail data	Berisi data-data history yang akan dipakai dalam proses analisis
Relasi antar table berdasar aturan terkini (selalu mengikuti rule /aturan terbaru)	Banyak aturan bisnis dapat tersaji antara tabel-tabel

Karakteristik *Data Warehouse*

1. Subject Oriented (Berorientasi subject)

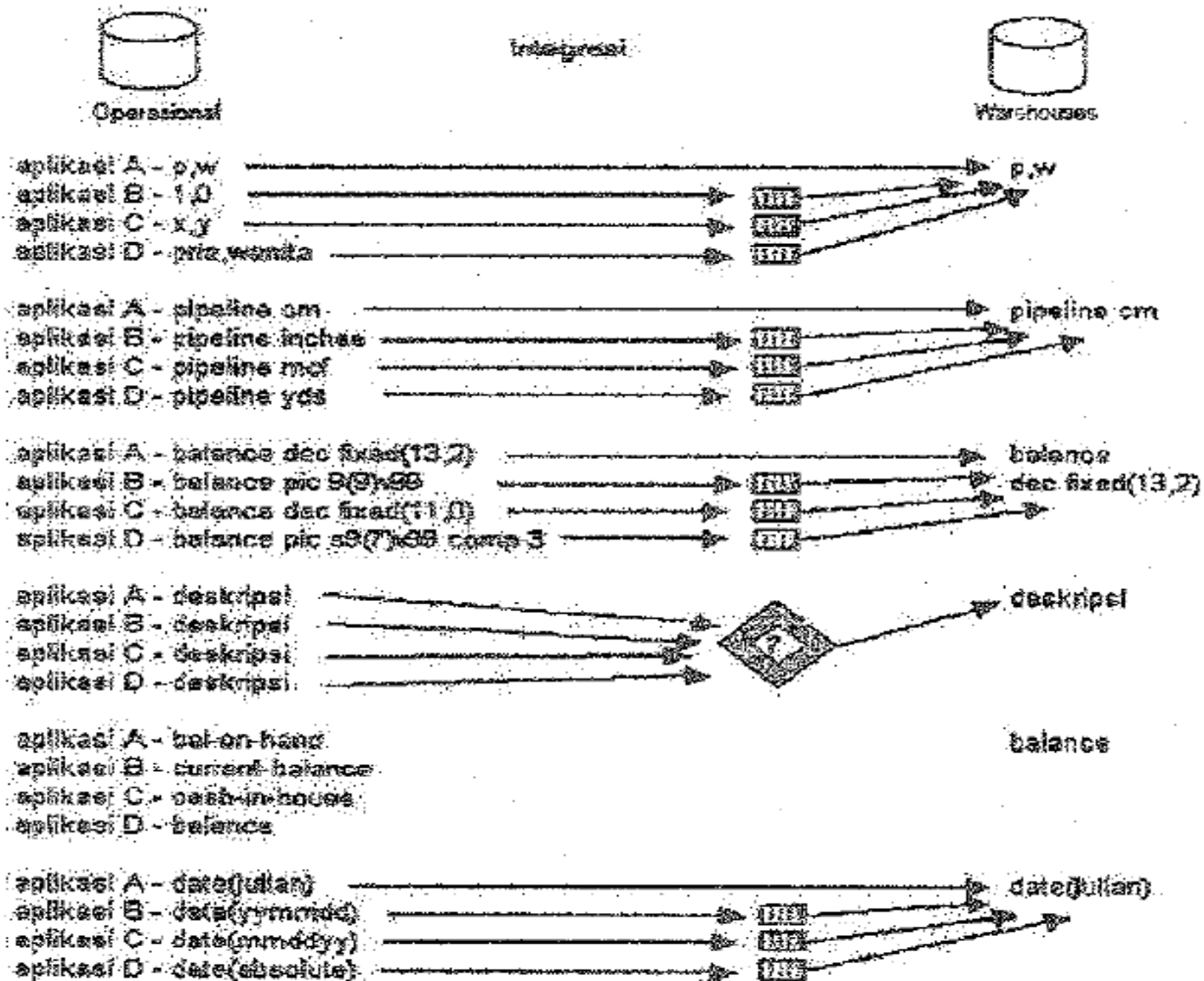
- *Data warehouse* berorientasi subject artinya data warehouse didesain untuk menganalisa data berdasarkan subject-subject tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu.
- Data warehouse diorganisasikan disekitar subjek-subjek utama dari perusahaan (customers, products dan sales) dan tidak diorganisasikan pada area-area aplikasi utama (customer invoicing, stock control dan product sales). Hal ini dikarenakan kebutuhan dari data warehouse untuk menyimpan data-data yang bersifat sebagai penunjang suatu keputusan, dari pada aplikasi yang berorientasi terhadap data.

Karakteristik *Data Warehouse*

2. *Integrated* (Terintegrasi)

- *Data Warehouse* dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah kedalam suatu format yang konsisten dan saling terintegrasi satu dengan lainnya. Dengan demikian data tidak bisa dipecah-pecah karena data yang ada merupakan suatu kesatuan yang menunjang keseluruhan konsep data warehouse itu sendiri.
- Syarat integrasi sumber data dapat dipenuhi dengan berbagai cara seperti konsisten dalam penamaan variable, konsisten dalam ukuran variable, konsisten dalam struktur pengkodean dan konsisten dalam atribut fisik dari data.

Karakteristik Data Warehouse

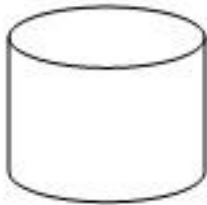


Karakteristik *Data Warehouse*

3. *Time-variant* (Rentang Waktu)

- Seluruh data pada data warehouse dapat dikatakan akurat atau valid pada rentang waktu tertentu. Untuk melihat interval waktu yang digunakan dalam mengukur keakuratan suatu data warehouse, kita dapat menggunakan cara antara lain :
- Cara yang paling sederhana adalah menyajikan data warehouse pada rentang waktu tertentu, misalnya antara 5 sampai 10 tahun ke depan.
- Cara yang kedua, dengan menggunakan variasi/perbedaan waktu yang disajikan dalam data warehouse baik implicit maupun explicit secara explicit dengan unsur waktu dalam hari, minggu, bulan dsb. Secara implicit misalnya pada saat data tersebut diduplikasi pada setiap akhir bulan, atau per tiga bulan. Unsur waktu akan tetap ada secara implisit didalam data tersebut.
- Cara yang ketiga, variasi waktu yang disajikan data warehouse melalui serangkaian snapshot yang panjang. Snapshot merupakan tampilan dari sebagian data tertentu sesuai keinginan pemakai dari keseluruhan data yang ada bersifat read-only.

Karakteristik *Data Warehouse*

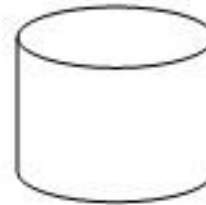


Operasional

Current value data:

- time horizon :60-90 days
- key may or may not have an element of time
- data can be update

Time Variancy



Data Warehouse

Snapshot data:

- time horizon :5-10 years
- key contain an element of time
- once snapshot is created, record cannot be update

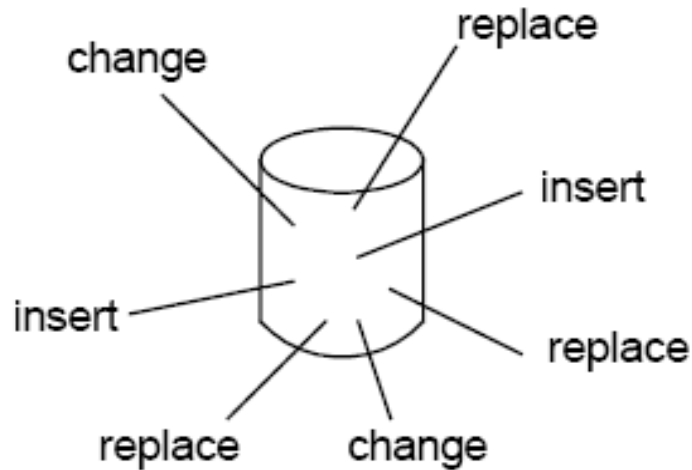
Time Variance Data Warehouse

Karakteristik *Data Warehouse*

4. Non-Volatile

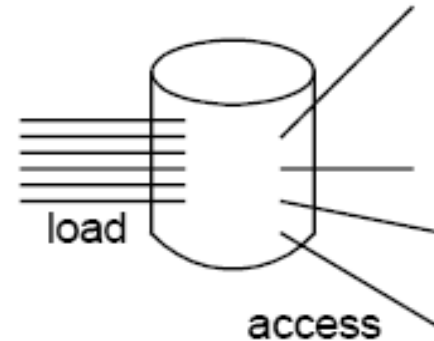
- Data pada data warehouse tidak di-*update* secara *real time* tetapi di *refresh* dari sistem operasional secara reguler. Data yang baru selalu ditambahkan sebagai suplemen bagi database itu sendiri dari pada sebagai sebuah perubahan. Database tersebut secara kontinyu menyerap data baru ini, kemudian secara incremental disatukan dengan data sebelumnya.
- Berbeda dengan database operasional yang dapat melakukan update, insert dan delete terhadap data yang mengubah isi dari database sedangkan pada data warehouse hanya ada dua kegiatan memanipulasi data yaitu loading data (mengambil data) dan akses data (mengakses data warehouse seperti melakukan query atau menampilkan laporan yang dibutuhkan, tidak ada kegiatan updating data).

Karakteristik *Data Warehouse*



operational

data is updated on a record-by-record basis regularly



data warehouse

data is loaded into the warehouse and is accessed there, but once snapshot of data is made, the data in the warehouse does not change

Non Volatile Data Warehouse