

# **PERTEMUAN 2**

## **Data warehouse and Business Intelligence**

# Pokok Bahasan

Pertemuan Ke-	Pokok Bahasan
1	Business Intelligence
2	Data Warehousing
3	Business Performance Management
4	Business Performance Management Methodologies
5	Pengantar Data Mining
6	Metode Learning Algoritma Data Mining
7	Review dan Quiz
8	UTS
9	Studi Kasus
10-15	Presentasi Tugas Kelompok
16	UAS

# Rencana Pembelajaran

## Tugas Kelompok

- ✓ Buat Kelompok maximal 4 orang/kelompok.
- ✓ Pengolahan data menggunakan tools rapidminer
- ✓ Menggunakan salahsatu metode data mining
- ✓ Dataset menggunakan data public atau private, setiap kelompok beda dataset
- ✓ Kumpulkan tugas kelompok tersebut berupa : makalah, dan powerpoint pada pertemuan 10 dan bisa dipresentasikan.
- ✓ Mengumpulkan draft artikel ilmiah.
- ✓ Nilai project & presentasi akan menjadi nilai kelompok, keaktifan dan nilai penguasaan materi.

# **Business Intelligence and Decision Support**

**(9<sup>th</sup> Ed., Prentice Hall)**

## **Chapter 2: Data Warehousing**

# Pengertian Data Warehouse

- Repositori fisik di mana data relasional diorganisasikan secara khusus untuk menyediakan data yang dibersihkan perusahaan dalam format standar
- Data Warehouse adalah kumpulan desain basis data terintegrasi dan berorientasi subjek untuk mendukung fungsi DSS, di mana setiap unit data adalah non-volatile dan relevan dengan waktu tertentu”

# Karakteristik DW

- Berorientasi pada objek
- Terintegrasi
- Time-variant (time series)
- Nonvolatile
- Diringkas
- Tidak dinormalisasi
- Metadata
- Berbasis web, relational/multi-dimensi
- Client/server
- Real-time dan/atau right-time (aktif)

# Data Mart

Departmen data warehouse yang hanya menyimpan data yang relevan

- **Dependent data mart**

Subset yang dibuat langsung dari data warehouse

- **Independent data mart**

Data warehouse kecil yang dirancang untuk unit bisnis strategis atau departemen

# Pengertian Data Warehousing

- **Operational data stores (ODS)**

Suatu tipe database sering digunakan sebagai area sementara untuk suatu data warehouse

- **Oper marts**

Mart data operasional.

- **Enterprise data warehouse (EDW)**

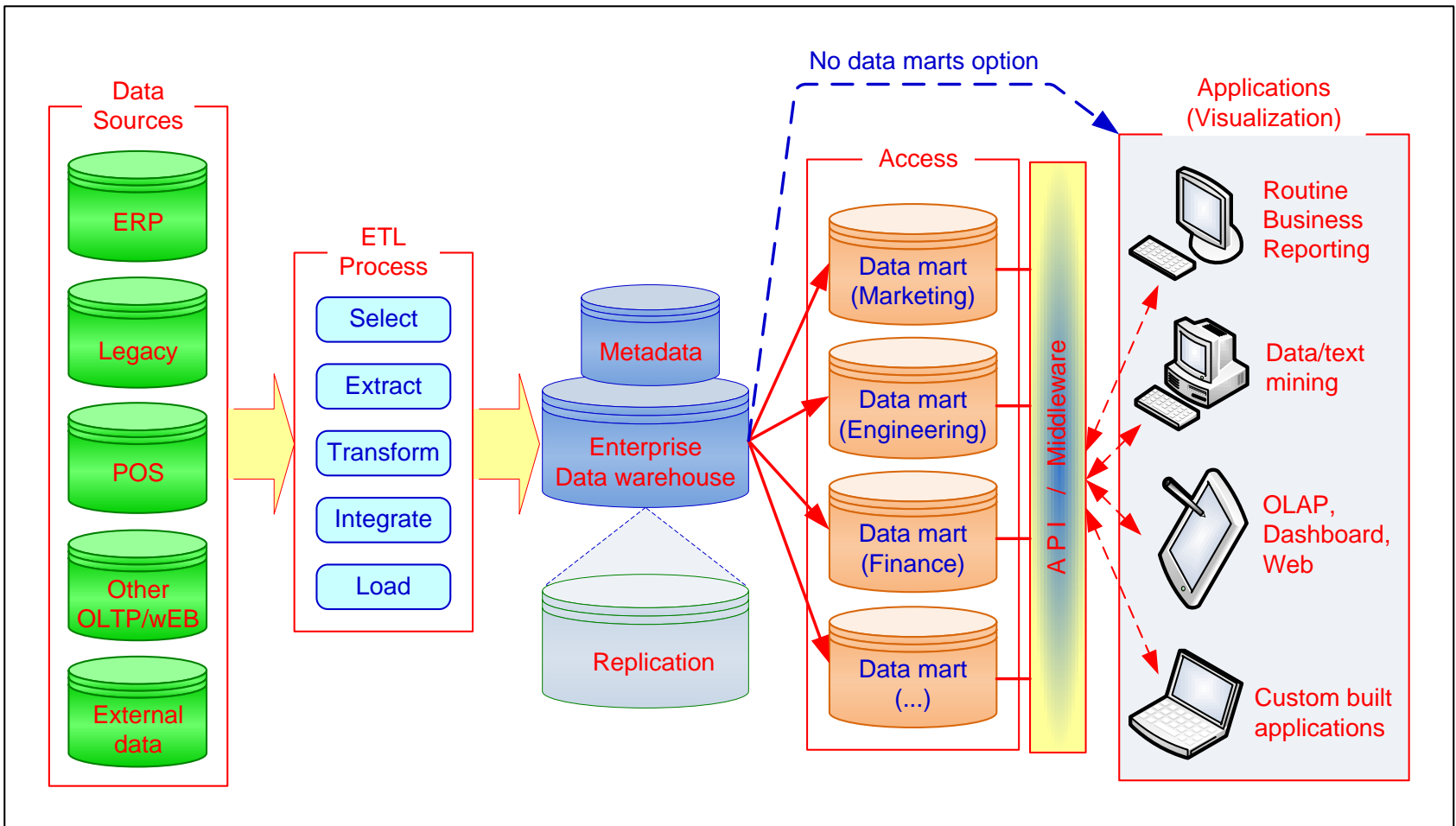
Data warehouse untuk perusahaan.

- **Metadata**

Suatu data tentang data. Dalam data warehouse, metadata mendeskripsikan konten datawarehouse dan cara akuisisi dan penggunaannya



# Kerangka Konseptual DW



# Arsitektur Umum DW

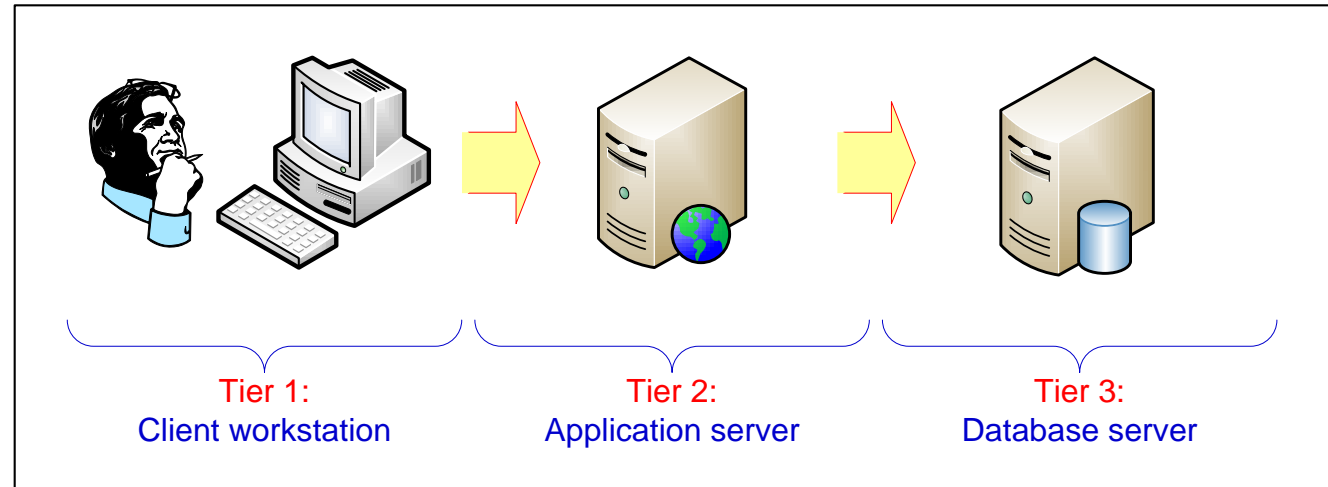
- **Arsitektur Tiga Tingkat (Three-tier architecture)**
  1. Perangkat lunak akuisisi data (back-end)
  2. Data warehouse yang berisi data & perangkat lunak
  3. Software Client (front-end) yang memungkinkan pengguna untuk mengakses dan menganalisis data warehouse
- **Arsitektur Dua Tingkat (Two-tier architecture)**

2 tingkatan pertama dalam arsitektur tiga tingkat digabungkan menjadi satu

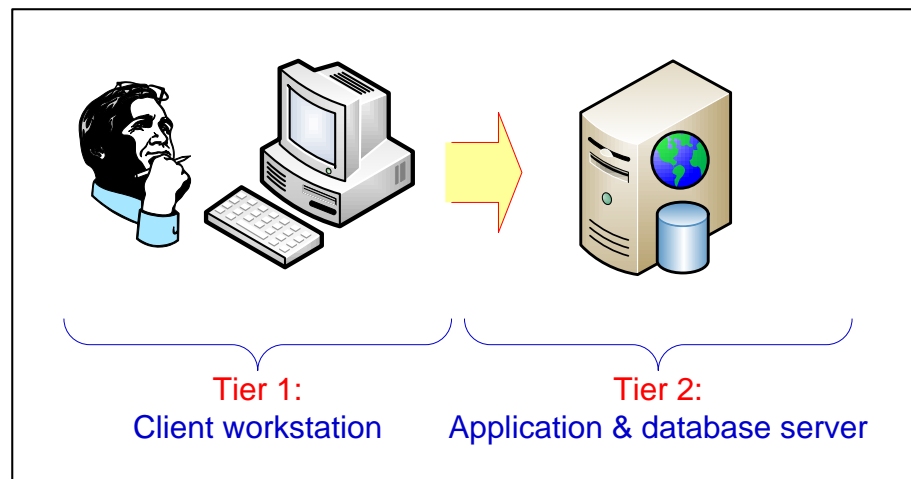
... kadang hanya ada satu tingkat?

# Arsitektur Umum DW

## 3-tier architecture



## 2-tier architecture

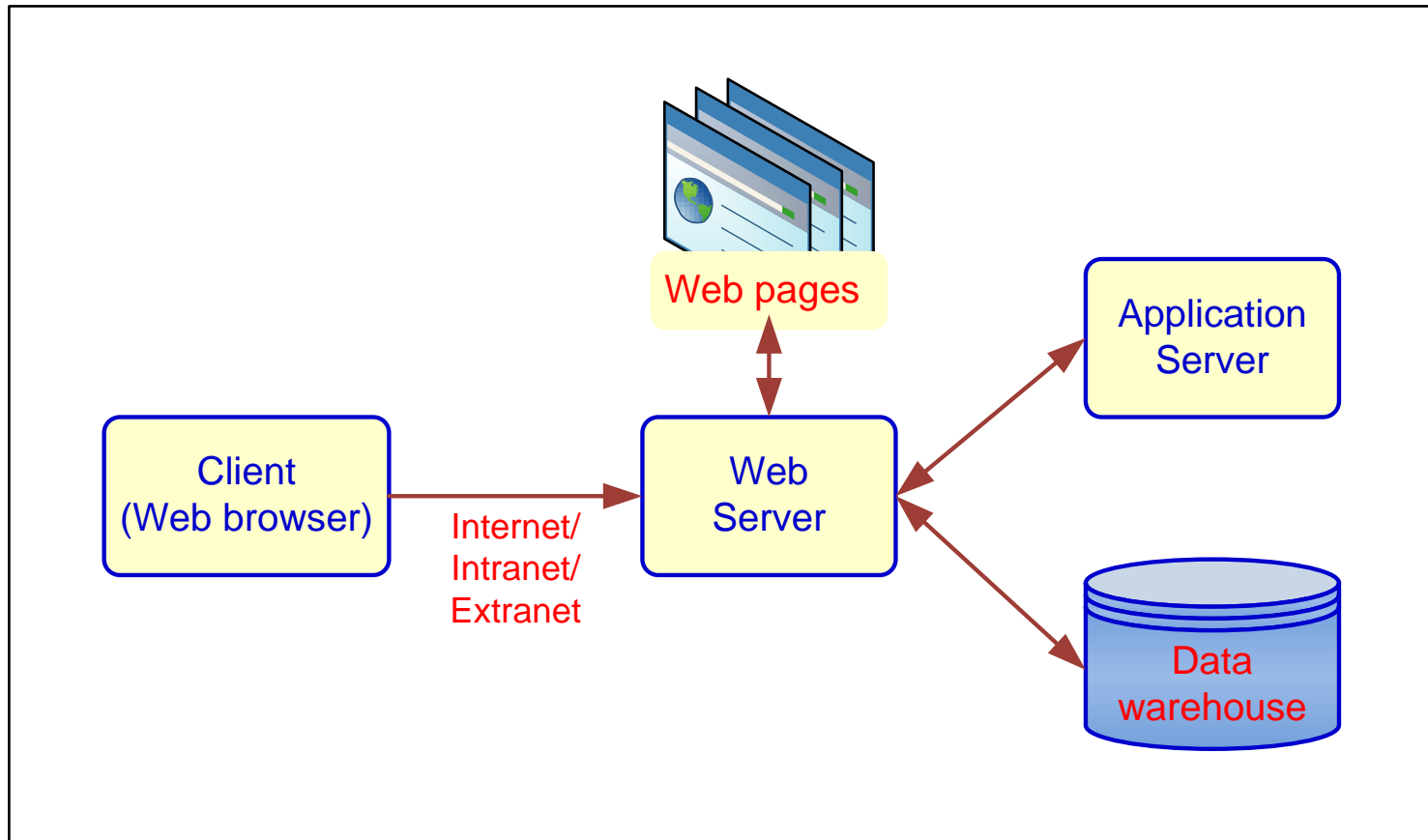


1-tier  
Architecture  
?

# Pertimbangan Arsitektur DW

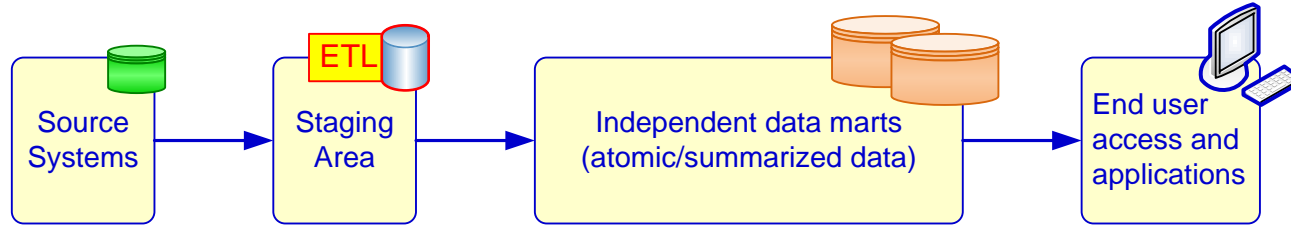
- Masalah yang perlu dipertimbangkan ketika memutuskan arsitektur mana yang akan digunakan:
  - Database Management System (DBMS) yang mana yang harus dipakai?
  - Apakah akan menggunakan proses parallel dan/atau partisi?
  - Apakah alat migrasi data akan digunakan untuk memuat data warehouse?
  - Alat apa yang akan digunakan untuk mendukung pengambilan dan analisis data?

# Arsitektur DW Berbasis Web

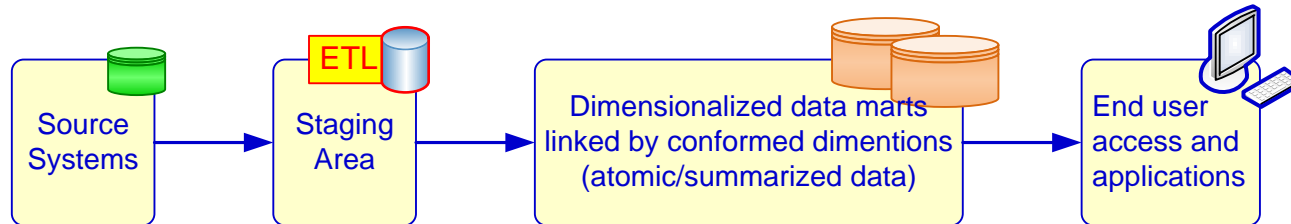


# Arsitektur Alternatif DW

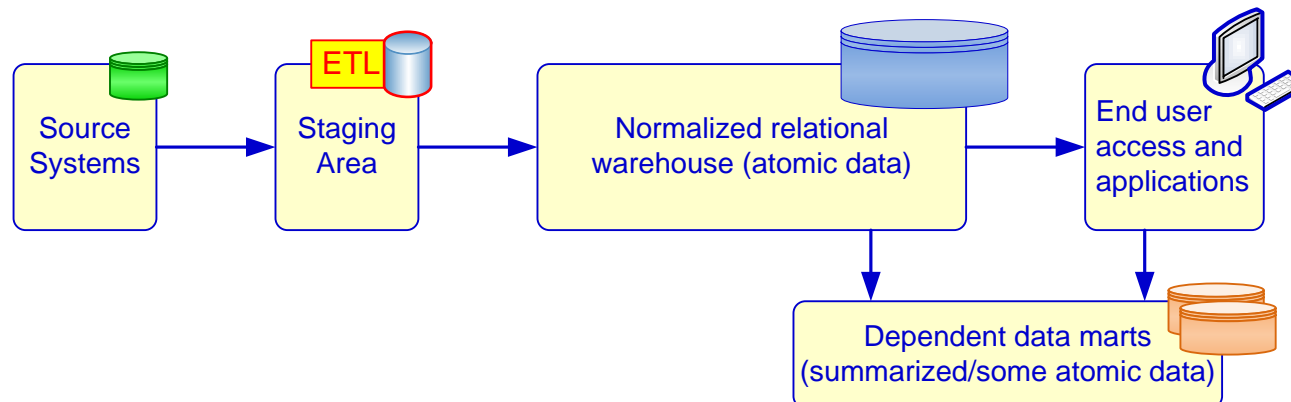
(a) Independent Data Marts Architecture



(b) Data Mart Bus Architecture with Linked Dimensional Datamarts

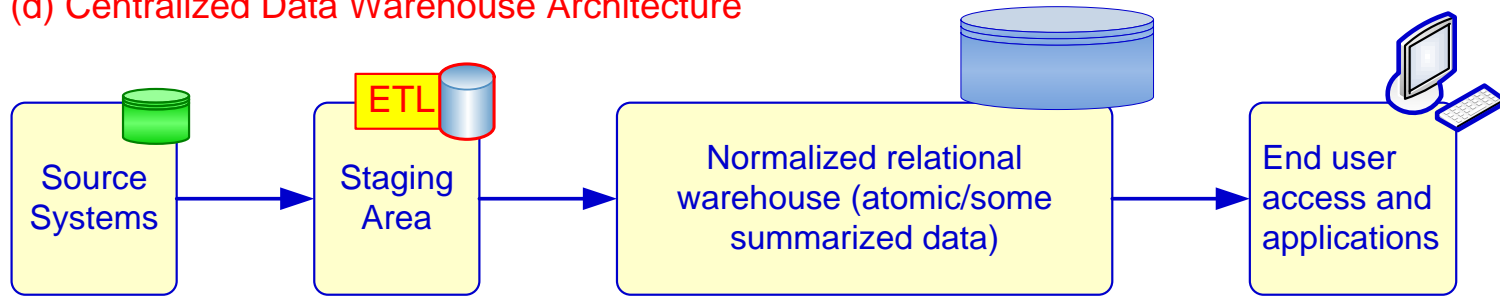


(c) Hub and Spoke Architecture (Corporate Information Factory)

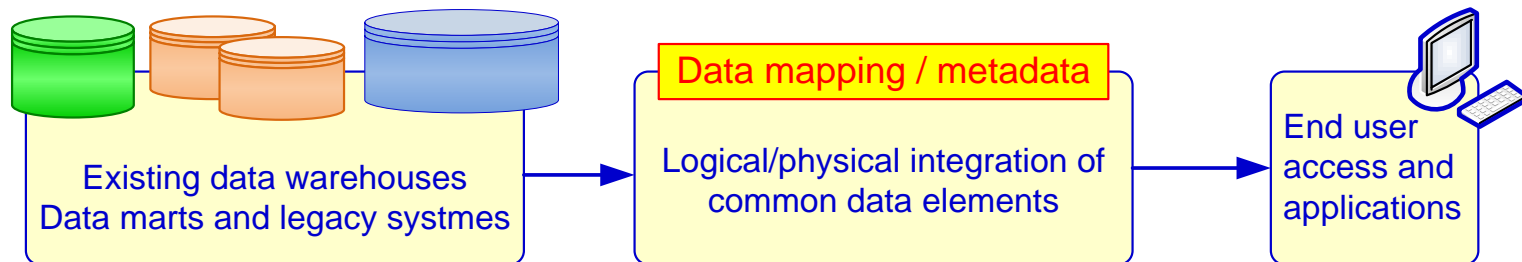


# Arsitektur Alternatif DW

(d) Centralized Data Warehouse Architecture



(e) Federated Architecture



# Arsitektur Mana yang Paling Baik?

- Bill Inmon Vs. Ralph Kimball
- Enterprise DW Vs. Data Marts approach

	Independent Data Marts	Bus Architecture	Hub-and- Spoke Architecture	Centralized Architecture (No Dependent Data Marts)	Federated Architecture
Information Quality	4.42	5.16	5.35	5.23	4.73
System Quality	4.59	5.60	5.56	5.41	4.69
Individual Impacts	5.08	5.80	5.62	5.64	5.15
Organizational Impacts	4.66	5.34	5.24	5.30	4.77

*Empirical study by Ariyachandra and Watson (2006)*

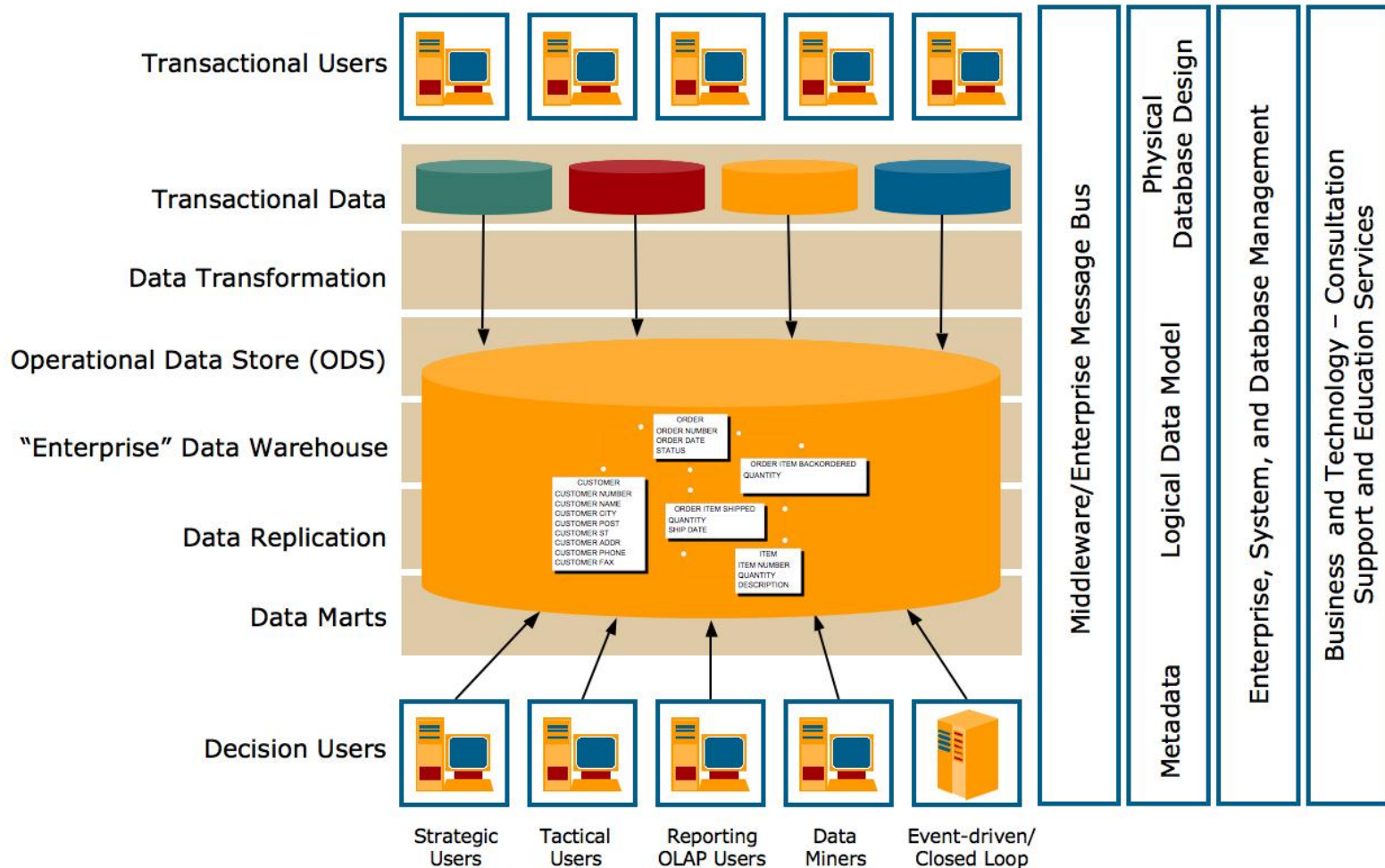


# Arsitektur Data Warehousing

Sepuluh faktor yang berpotensi mempengaruhi keputusan pemilihan arsitektur:

1. Saling ketergantungan informasi antar unit organisasi
2. Kebutuhan informasi manajemen tingkat atas
3. Urgensi kebutuhan akan data warehouse
4. Sifat tugas end-user
5. Kendala pada sumber daya
6. Pandangan strategis dari data warehouse sebelum implementasi
7. Kompatibilitas dengan sistem yang ada
8. Kemampuan yang dirasakan dari staf TI in-house
9. Masalah teknis
10. Faktor Sosial/politik

# Enterprise Data Warehouse (by Teradata Corporation)



# Data Integration and the Extraction, Transformation, and Load (ETL) Process

- **Integrasi Data (Data integration)**

Integrasi yang terdiri dari tiga proses utama: akses data, federasi data, dan perubahan tangkapan.

- **Enterprise application integration (EAI)**

Sebuah teknologi yang menyediakan kendaraan untuk mendorong data dari sumber sistem ke dalam data warehouse

- **Enterprise information integration (EII)**

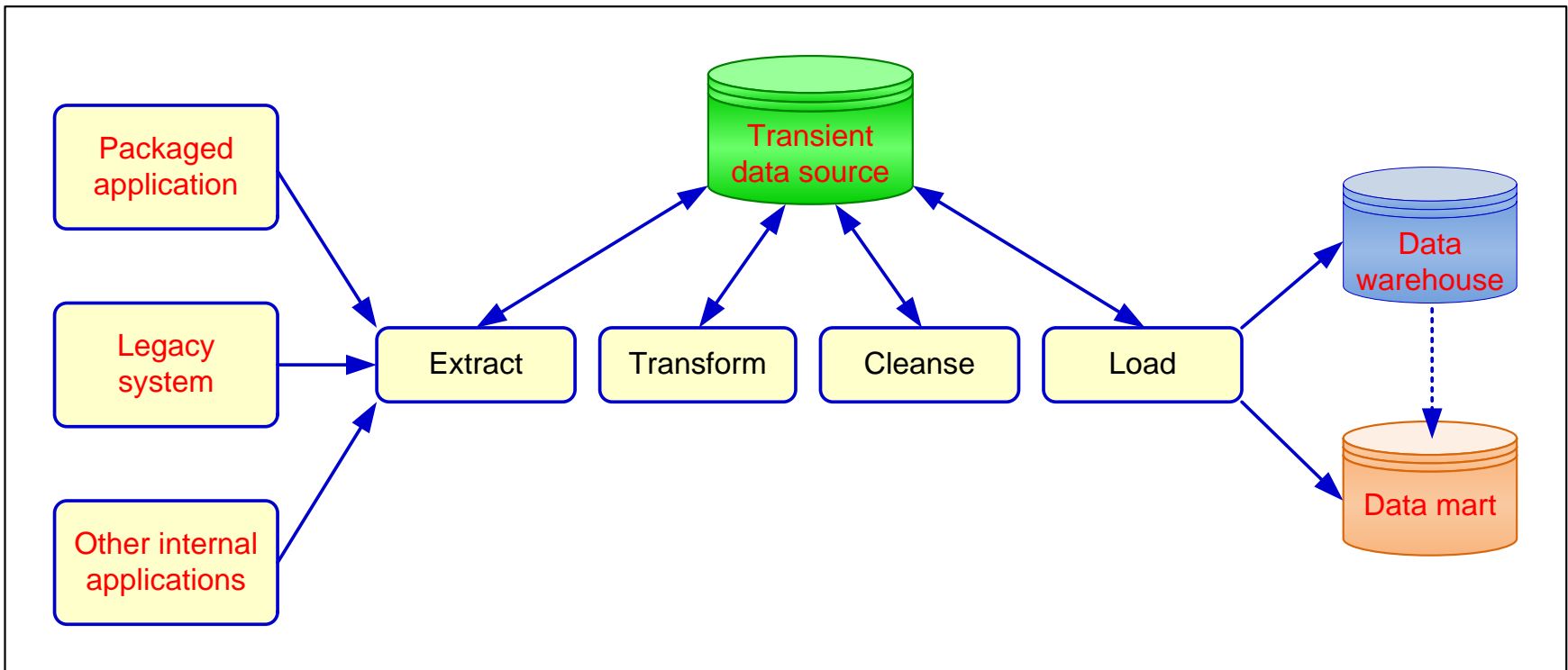
evolving tool space yang menjanjikan integrasi data real-time dari berbagai sumber

- **Service-oriented architecture (SOA)**

Cara baru untuk mengintegrasikan sistem informasi

# Data Integration and the Extraction, Transformation, and Load (ETL) Process

Extraction, transformation, and load (ETL) process



# ETL

- Masalah yang mempengaruhi pembelian alat ETL
  - Alat transformasi data mahal
  - Alat transformasi data mungkin memiliki kurva belajar yang panjang
- Kriteria penting dalam memilih alat ETL
  - Kemampuan untuk membaca dan menulis ke sejumlah sumber data / arsitektur yang tidak terbatas
  - Pengambilan dan pengiriman metadata secara otomatis
  - Sejarah kesesuaian dengan standar terbuka
  - Antarmuka yang mudah digunakan untuk pengembang dan fungsional user

# Manfaat DW

- Manfaat langsung dari data warehouse
  - Mengizinkan pengguna akhir melakukan analisis ekstensif
  - Mengizinkan tampilan konsolidasi data perusahaan
  - Informasi yang lebih baik dan lebih tepat waktu
  - Peningkatan kinerja sistem
  - Penyederhanaan akses data
- Manfaat tidak langsung dari data warehouse
  - Meningkatkan pengetahuan bisnis
  - Menghadirkan keunggulan kompetitif
  - Meningkatkan layanan dan kepuasan pelanggan
  - Memfasilitasi pengambilan keputusan
  - Membantu dalam mereformasi proses bisnis

# Pengembangan Data Warehouse

- Pendekatan pengembangan data warehouse
  - Model Inmon : Pendekatan EDW (top-down)
  - Model Kimball : Pendekatan Data mart (bottom-up)
  - Model mana yang terbaik?
    - There is no one-size-fits-all strategy to DW
  - Salah satu alternative ada pada hosted warehouse
- Struktur Data warehouse:
  - The Star Schema vs. Relational
- Real-time pada datawarehouse?

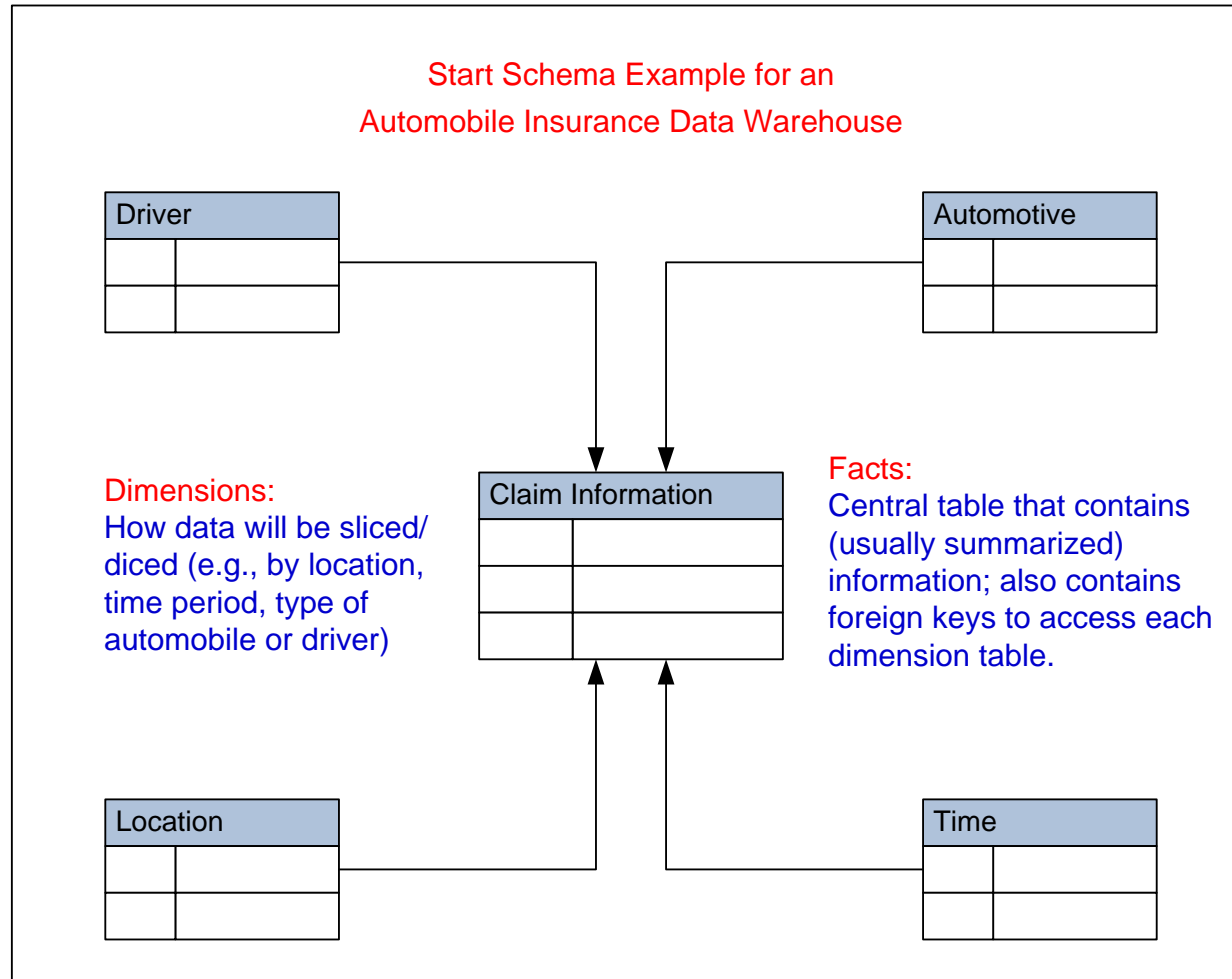
# DW Development Approaches

	(Inmon Approach)	(Kimball Approach)
Effort	Data Mart Approach	EDW Approach
Scope	One subject area	Several subject areas
Development time	Months	Years
Development cost	\$10,000 to \$100,000+	\$1,000,000+
Development difficulty	Low to medium	High
Data prerequisite for sharing	Common (within business area)	Common (across enterprise)
Sources	Only some operational and external systems	Many operational and external systems
Size	Megabytes to several gigabytes	Gigabytes to petabytes
Time horizon	Near-current and historical data	Historical data
Data transformations	Low to medium	High
Update frequency	Hourly, daily, weekly	Weekly, monthly
<i>Technology</i>		
Hardware	Workstations and departmental servers	Enterprise servers and mainframe computers
Operating system	Windows and Linux	Unix, Z/OS, OS/390
Databases	Workgroup or standard database servers	Enterprise database servers

*See Table 8.3 for details*



# DW Structure: Star Schema (a.k.a. Dimensional Modeling)

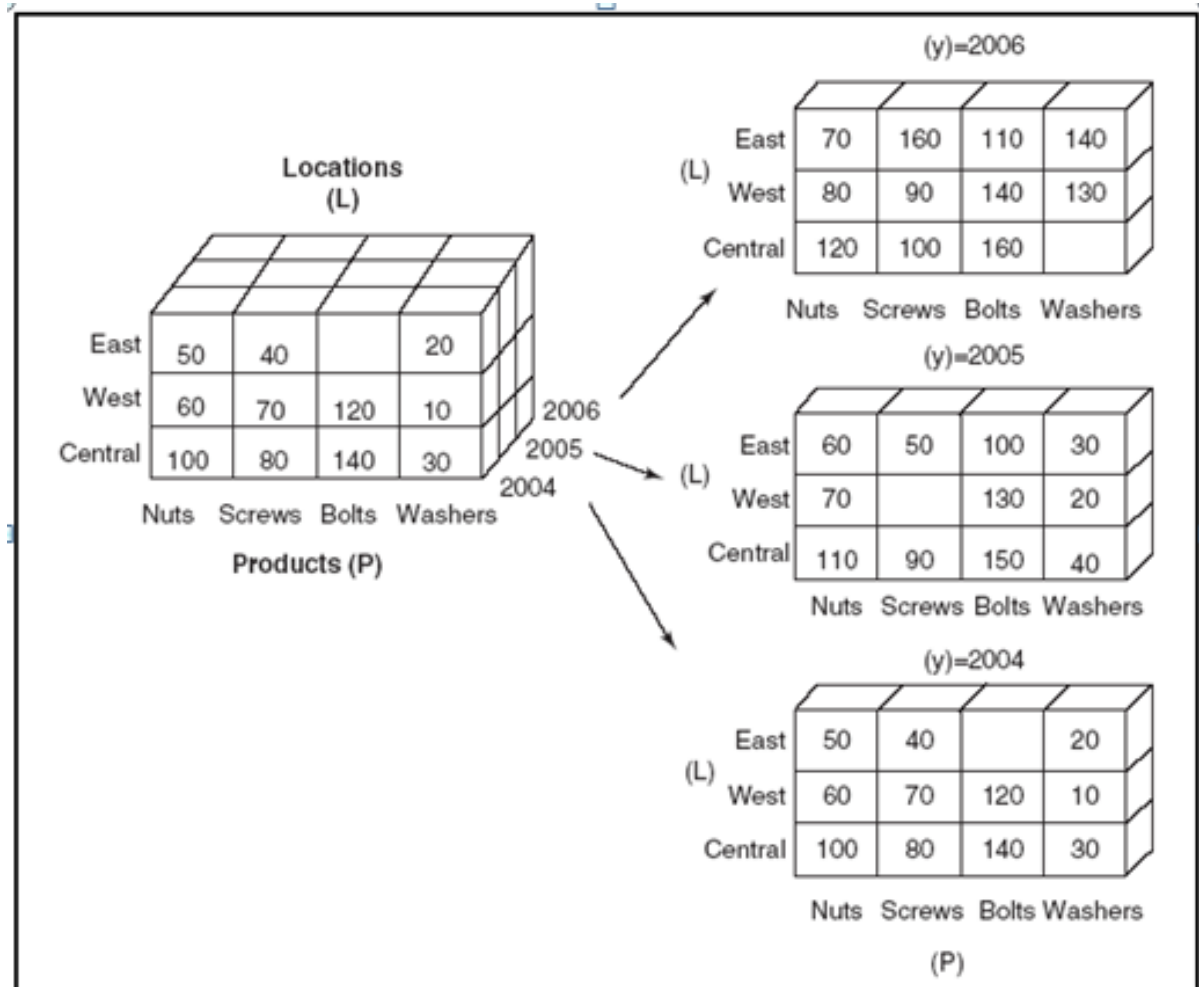


# Pemodelan Dimensi

## Kubus Data

Objek dua dimensi, tiga dimensi, atau lebih tinggi di mana setiap dimensi data mewakili ukuran yang menarik

- Grain
- Drill-down
- Slicing



# Praktik Terbaik untuk Mengimplementasikan DW

- Proyek harus sesuai dengan strategi perusahaan
- Harus ada persetujuan penuh untuk proyek
- Sangat penting untuk mengelola ekspektasi user
- Data warehouse harus dibangun secara bertahap
- Kemampuan beradaptasi harus dibangun sejak awal
- Proyek harus dikelola oleh profesional IT dan bisnis (hubungan bisnis-pemasok harus dikembangkan)
- Hanya memuat data yang telah dibersihkan / berkualitas tinggi
- Jangan mengabaikan persyaratan pelatihan
- Sadar secara politis.

# Resiko Implementasi DW

- Tidak ada misi atau tujuan
- Kualitas data sumber tidak diketahui
- Keterampilan tidak pada tempatnya
- Anggaran tidak memadai
- Kurangnya perangkat lunak pendukung
- Sumber data tidak dipahami
- Sponsor yang lemah
- Pengguna tidak melek komputer
- Masalah politik atau perang wilayah
- Ekspektasi pengguna yang tidak realistis

(Continued ...)

# Resiko Implementasi DW

- Risiko arsitektur dan desain
- Cakupan creep dan perubahan persyaratan
- Vendor di luar kendali
- Berbagai platform
- Key people meninggalkan proyek
- Kehilangan sponsor
- Terlalu banyak teknologi baru
- Harus memperbaiki sistem operasional
- Lingkungan terdistribusi secara geografis
- Tim geografi dan budaya bahasa

# Hal yang harus dihindari untuk keberhasilan Implementasi DW

- Dimulai dengan rantai sponsor yang salah
- Menetapkan harapan yang tidak dapat Anda penuhi
- Terlibat dalam perilaku naif secara politis
- Memuat warehouse dengan informasi hanya karena ketersediaan
- mempercayai bahwa desain database data warehouse sama dengan desain DB transaksional
- Memilih manajer data warehouse yang berorientasi teknologi daripada berorientasi pengguna

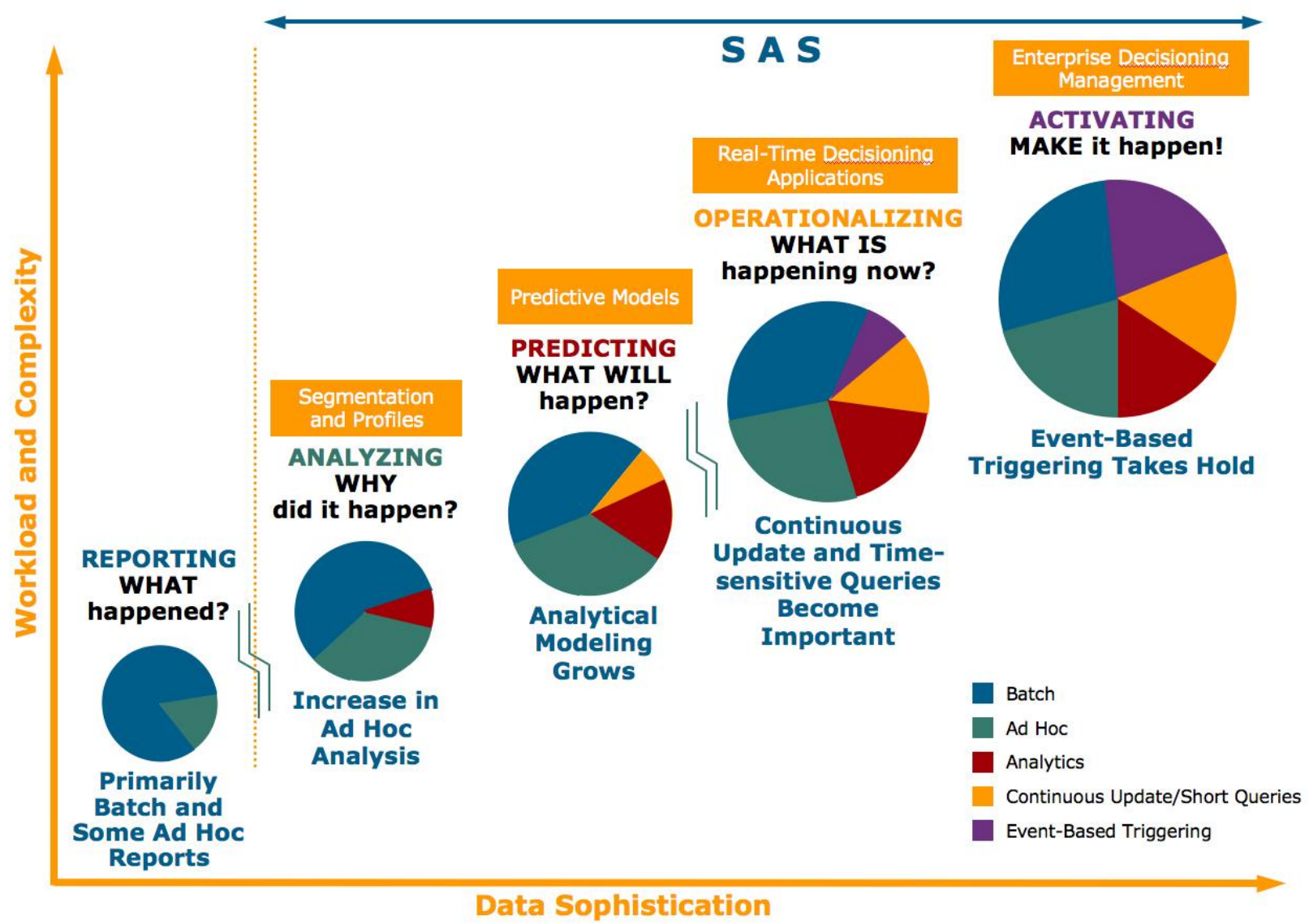
*(...see more on page 356)*

# Real-time DW

## (a.k.a. Active Data Warehousing)

- Memungkinkan pembaruan data real-time untuk analisis real-time dan pengambilan keputusan secara real-time pula sedang berkembang sangat pesat
  - Push vs. Pull (of data)
- Memperhatikan real-time BI
  - Tidak semua data harus diperbarui terus menerus
  - Ketidakcocokan laporan menghasilkan laporan yg terpisah
  - Mungkin mahal biaya
  - Mungkin juga tidak layak

# Evolution of DSS & DW





# Active Data Warehousing (by Teradata Corporation)

## Active Access

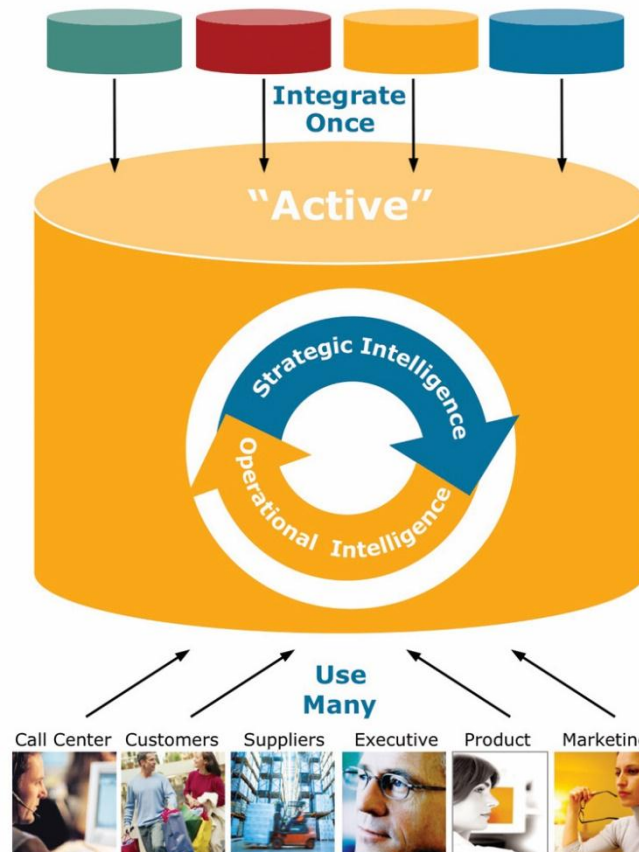
Front-Line operational decisions or services supported by near-real-time (NRT) access; Service Level Agreements of 5 seconds or less

## Active Load

Intra-day data acquisition; Mini-batch to NRT trickle data feeds measured in minutes or seconds

## Active Events

Proactive monitoring of business activity initiating intelligent actions based on rules and context; to systems or users supporting an operational business process



## Active Workload Management

Dynamically manage system resources for optimum performance and resource utilization supporting a mixed-workload environment

## Active Enterprise Integration

Integration into the Enterprise Architecture for delivery of intelligent decisioning services

## Active Availability

Business Continuity to support the requirements of the business (up to 7X24)

# Comparing Traditional and Active DW

---

## Traditional Data Warehouse Environment

---

Strategic decisions only

Results sometimes hard to measure

Daily, weekly, monthly data currency acceptable; summaries often appropriate

Moderate user concurrency

Highly restrictive reporting used to confirm or check existing processes and patterns; often uses predeveloped summary tables or data marts

Power users, knowledge workers, internal users

---

## Active Data Warehouse Environment

---

Strategic and tactical decisions

Results measured with operations

Only comprehensive detailed data available within minutes is acceptable

High number (1,000 or more) of users accessing and querying the system simultaneously

Flexible ad hoc reporting, as well as machine-assisted modeling (e.g., data mining) to discover new hypotheses and relationships

Operational staffs, call centers, external users

---

# Data Warehouse Administration

- Karena ukurannya yang besar dan sifatnya yang intrinsik, DW membutuhkan pemantauan yang kuat untuk menjaga efisiensi, produktivitas, dan keamanannya.
- Administrasi dan manajemen gudang data yang berhasil memerlukan keterampilan dan kemahiran yang melampaui apa yang diperlukan oleh administrator basis data tradisional.
  - Membutuhkan keahlian dalam perangkat lunak yang tinggi, perangkat keras, dan teknologi jaringan berkinerja tinggi

# DW Scalability and Security

- Scalability
  - Masalah utama yang berkaitan dengan skalabilitas:
    - Jumlah data di warehouse
    - Seberapa cepat warehouse diharapkan tumbuh
    - Jumlah pengguna bersamaan
    - Kompleksitas kueri pengguna
  - GSkalabilitas yang baik berarti bahwa kueri dan fungsi akses data lainnya akan tumbuh secara linear dengan ukuran warehouse
- Security
  - Penekanan pada keamanan dan privasi