

# BAIT 3013

# Business Intelligence

## Lecture 5

## FUNCTIONAL COMPONENTS OF BUSINESS INTELLIGENCE SYSTEMS

# Data Warehousing Definitions and Concepts

- Data warehouse
  - A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format

# Data Warehousing Definitions and Concepts

- Characteristics of data warehousing
  - Subject oriented
  - Integrated
  - Time variant (time series)
  - Nonvolatile

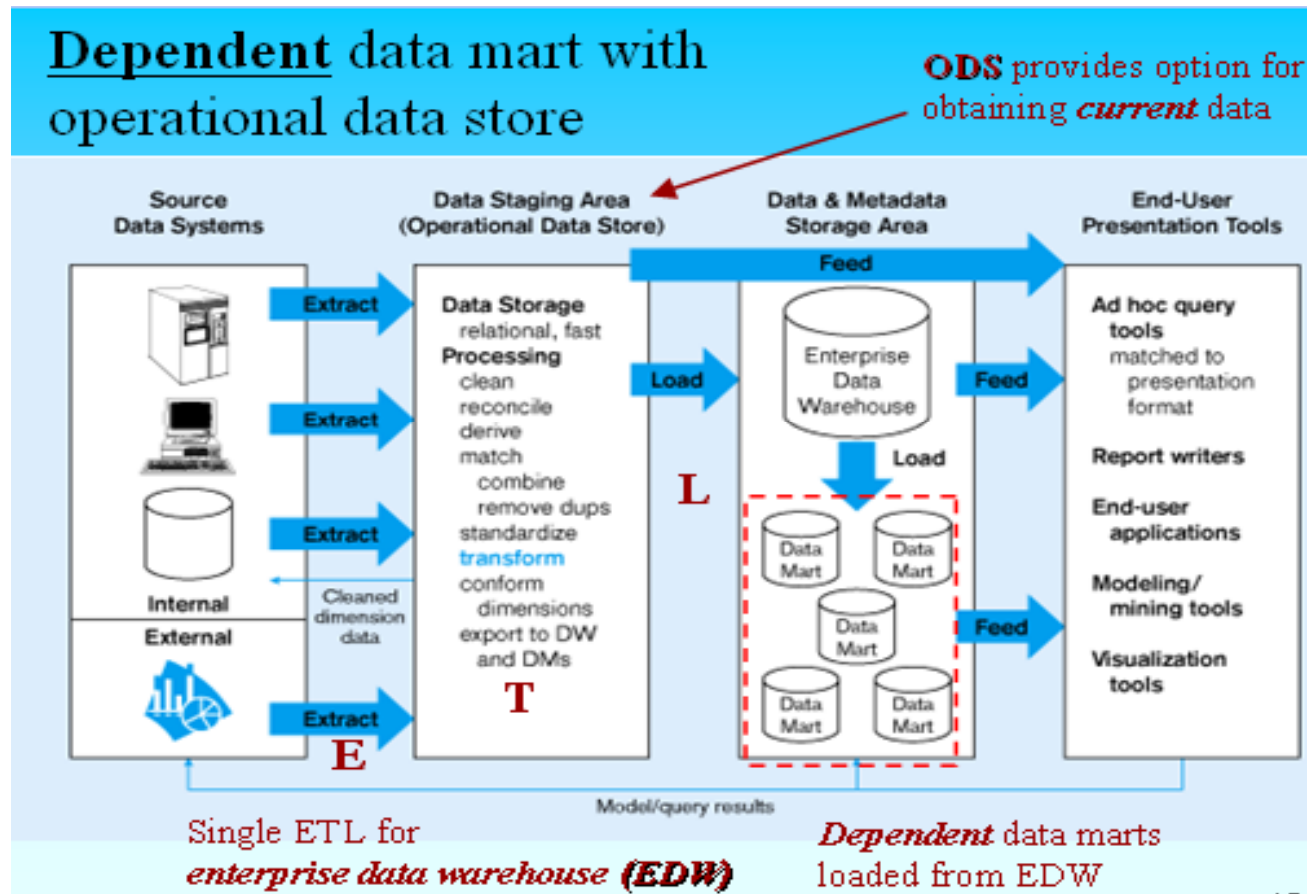
# Data Warehousing Definitions and Concepts

- Other Characteristics of data warehousing
  - Web based
  - Relational/multidimensional
  - Client/server
  - Real-time
  - Include metadata

# Data Warehousing : Definitions and Concepts

- Data mart
  - A departmental data warehouse that stores only relevant data
- Dependent data mart
  - A subset that is created directly from a data warehouse
- Independent data mart
  - A small data warehouse designed for a strategic business unit or a department

# Data Warehousing: Definitions and Concepts

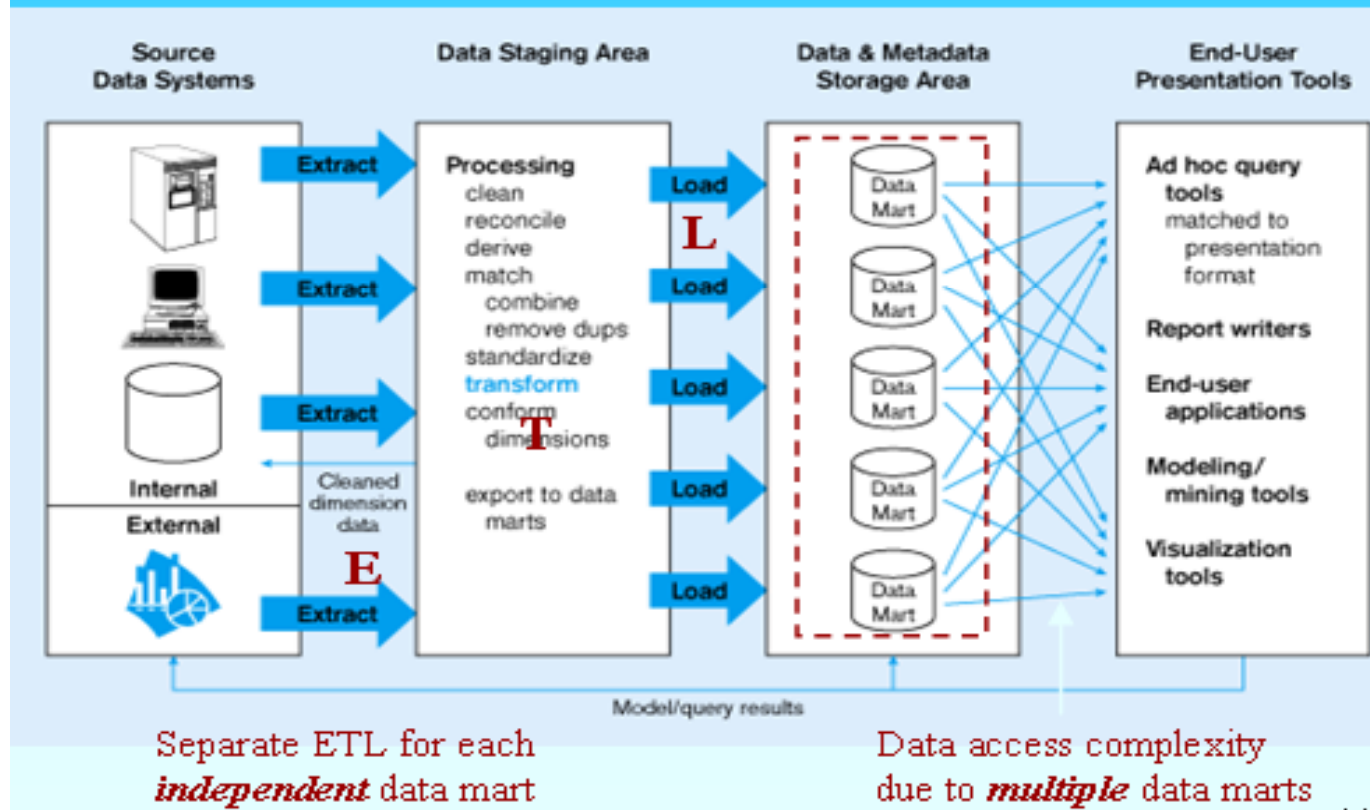


# Data Warehousing: Definitions and Concepts

## Independent data mart

### **Data marts:**

Mini-warehouses, limited in scope



# Data Warehousing: Definitions and Concepts

**Table 11-2 Data Warehouse Versus Data Mart**

<i>Data Warehouse</i>	<i>Data Mart</i>
<i>Scope</i> <ul style="list-style-type: none"> <li>• Application independent</li> <li>• Centralized, possibly enterprise-wide</li> <li>• Planned</li> </ul>	<i>Scope</i> <ul style="list-style-type: none"> <li>• Specific DSS application</li> <li>• Decentralized by user area</li> <li>• Organic, possibly not planned</li> </ul>
<i>Data</i> <ul style="list-style-type: none"> <li>• Historical, detailed, and summarized</li> <li>• Lightly denormalized</li> </ul>	<i>Data</i> <ul style="list-style-type: none"> <li>• Some history, detailed, and summarized</li> <li>• Highly denormalized</li> </ul>
<i>Subjects</i> <ul style="list-style-type: none"> <li>• Multiple subjects</li> </ul>	<i>Subjects</i> <ul style="list-style-type: none"> <li>• One central subject of concern to users</li> </ul>
<i>Sources</i> <ul style="list-style-type: none"> <li>• Many internal and external sources</li> </ul>	<i>Sources</i> <ul style="list-style-type: none"> <li>• Few internal and external sources</li> </ul>
<i>Other Characteristics</i> <ul style="list-style-type: none"> <li>• Flexible</li> <li>• Data-oriented</li> <li>• Long life</li> <li>• Large</li> <li>• Single complex structure</li> </ul>	<i>Other Characteristics</i> <ul style="list-style-type: none"> <li>• Restrictive</li> <li>• Project-oriented</li> <li>• Short life</li> <li>• Start small, becomes large</li> <li>• Multi, semi-complex structures, together complex</li> </ul>

Adapted from Strange (1997)



# Data Warehousing : Definitions and Concepts

- Operational data stores (ODS)
  - A type of database often used as an interim area for a data warehouse, especially for customer information files
- Operational marts
  - An operational data mart. An oper mart is a small-scale data mart typically used by a single department or functional area in an organization

# Data Warehousing : Definitions and Concepts

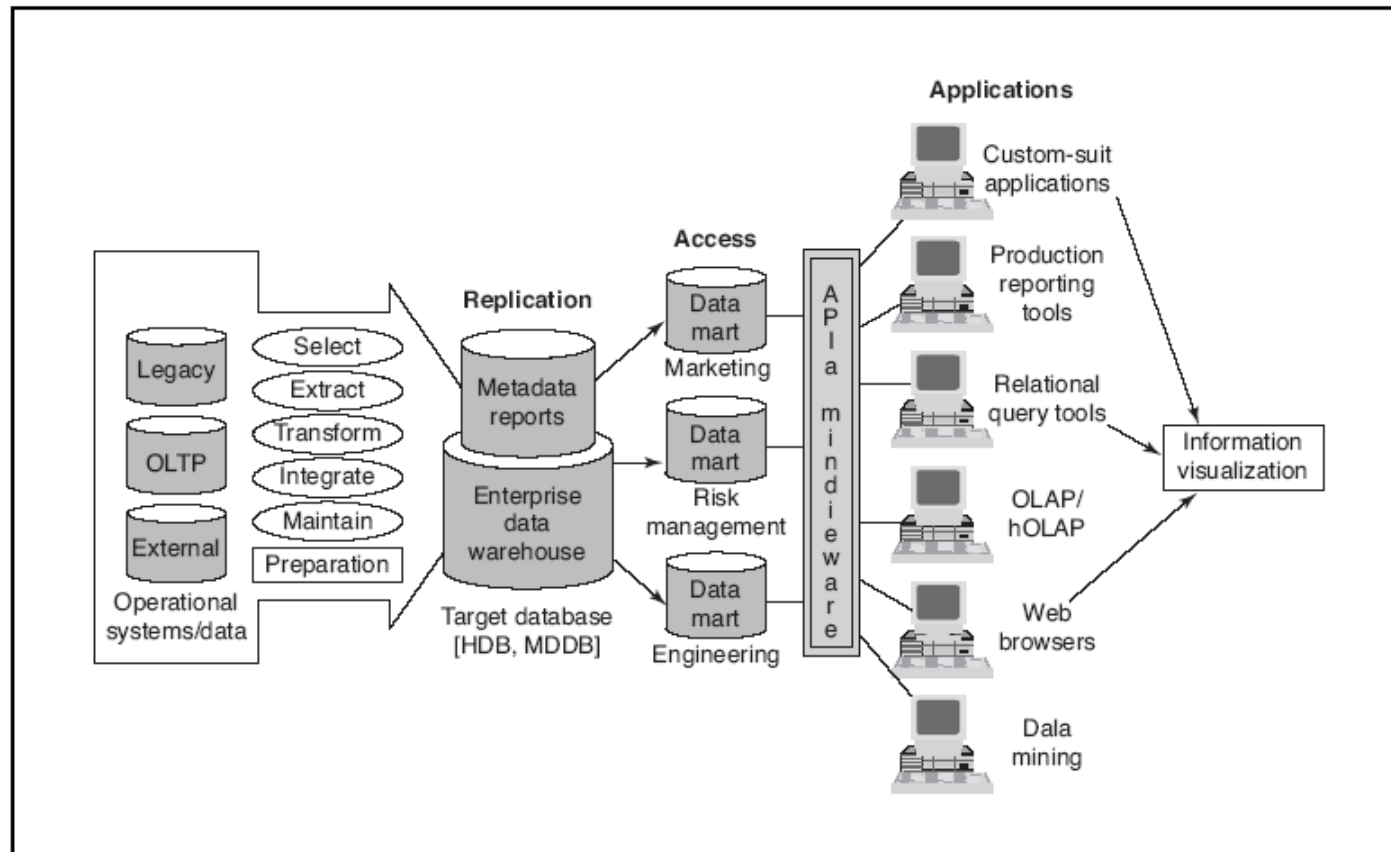
- Enterprise data warehouse (EDW)
  - A technology that provides a vehicle for pushing data from source systems into a data warehouse
- Metadata
  - Data about data. In a data warehouse, metadata describe the contents of a data warehouse and the manner of its use

# Data Warehousing : Process Overview

- The major components of a data warehousing process
  - Data sources
  - Data extraction
  - Data loading
  - Comprehensive database
  - Metadata
  - Middleware tools

# Data Warehousing : Process Overview

FIGURE 5.1 Data Warehouse Framework and Views



# Data Warehousing Architectures

- Three tier
- Two tier

# Data Warehousing Architectures : Three tier

- Three parts of the data warehouse
  - The data warehouse that contains the data and associated software
  - Data acquisition (back-end) software that extracts data from legacy systems and external sources, consolidates and summarizes them, and loads them into the data warehouse
  - Client (front-end) software that allows users to access and analyze data from the warehouse

# Data Warehousing Architectures : Three tier

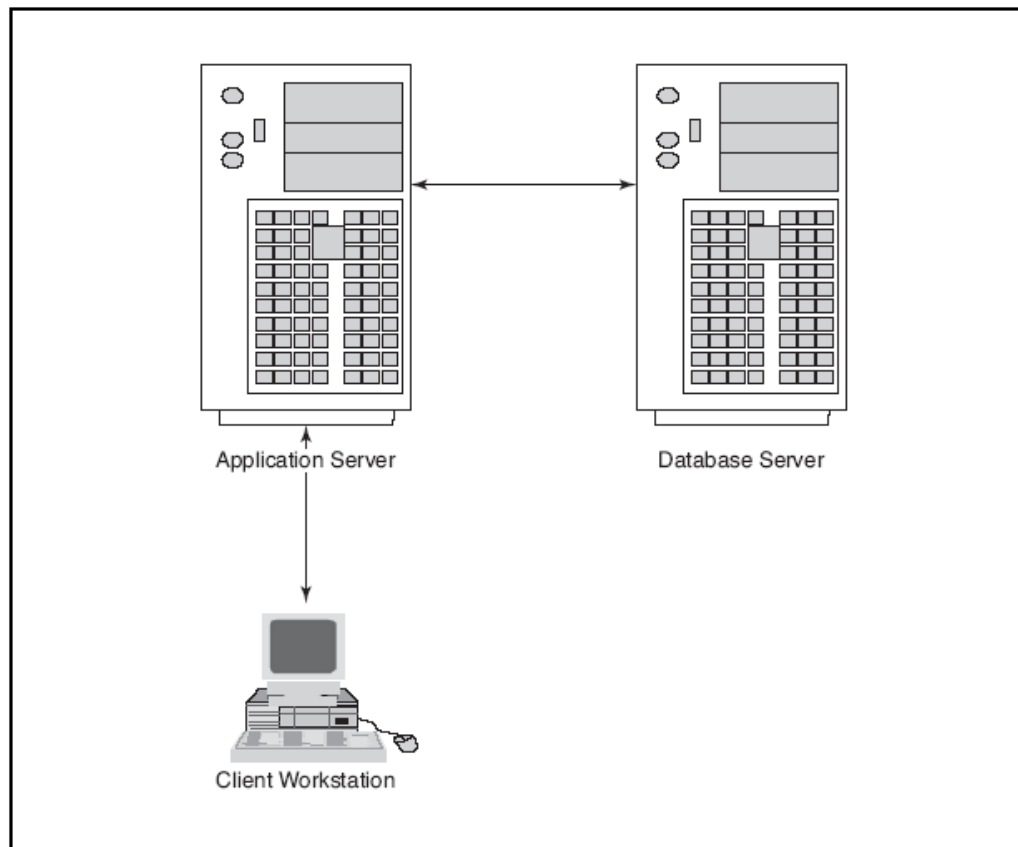


FIGURE 5.2 Architecture of a Three-Tier Data Warehouse

# Data Warehousing Architectures : Three tier

- Advantages :
  - The separate of the function of data warehouse can eliminates resource constraints and easy to create data marts



# Data Warehousing Architectures : Two tier

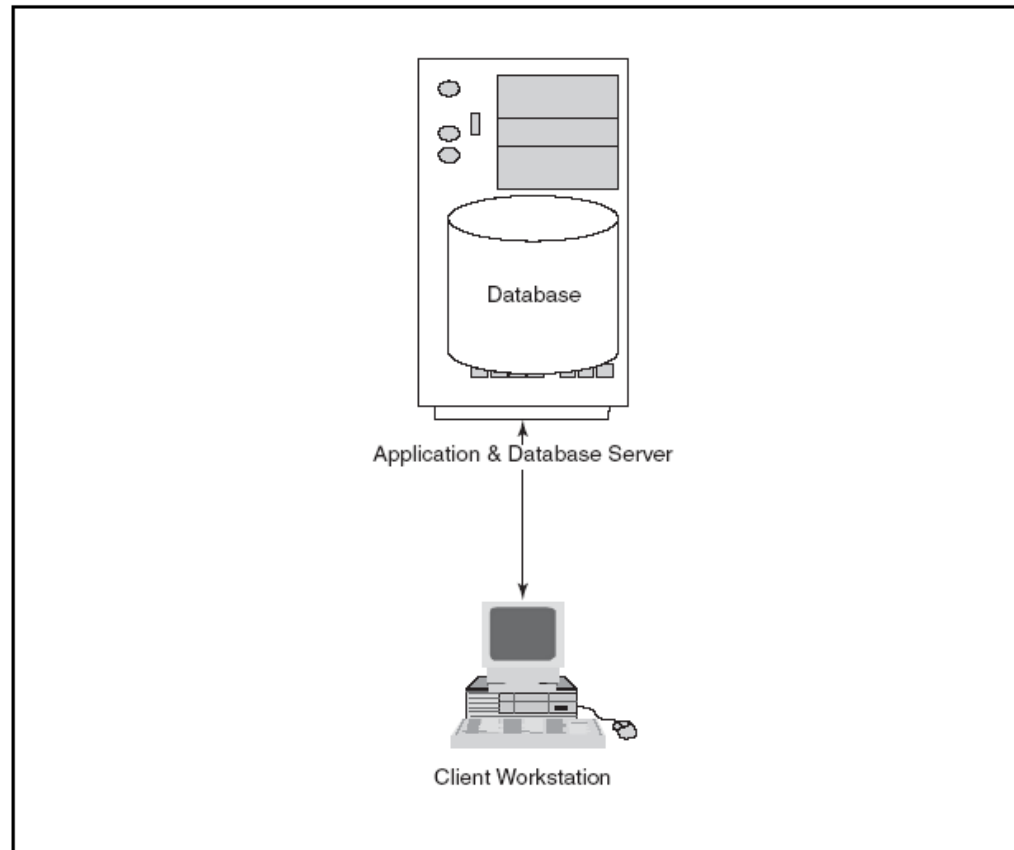


FIGURE 5.3 Architecture of a Two-Tier Data Warehouse

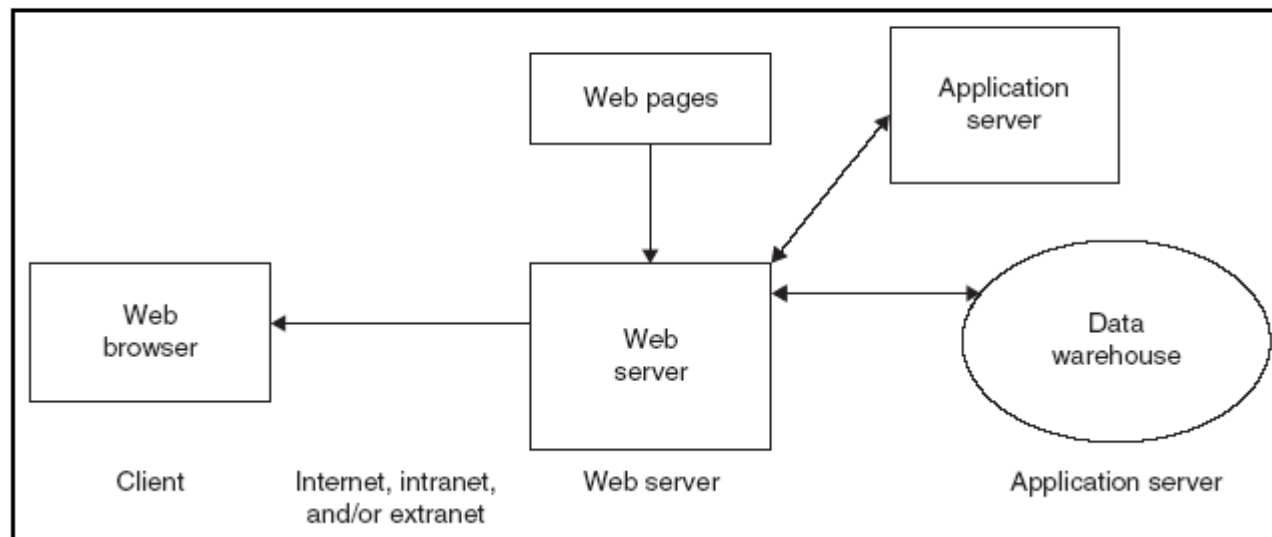
# Data Warehousing Architectures :

## Two tier

- Advantages :
  - More economic

# Web based Data Warehousing

**FIGURE 5.4** Architecture of Web-Based Data Warehousing

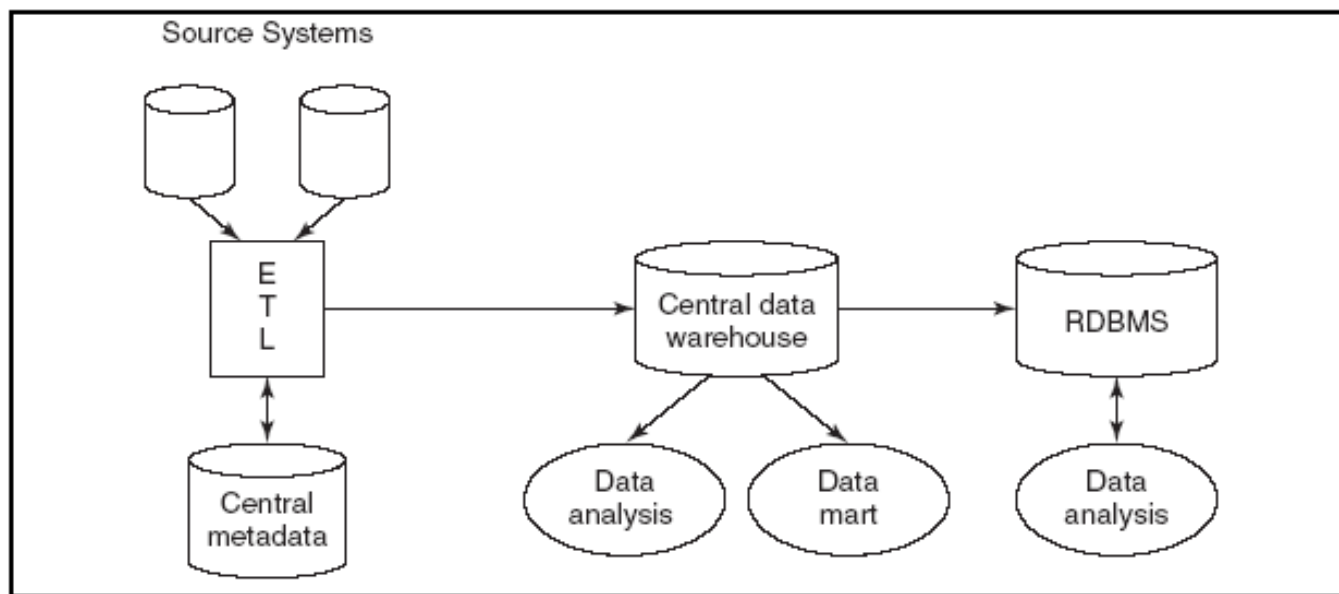


# Data Warehousing

- Issues to consider when deciding which architecture to use:
  - Which database management system (DBMS) should be used?
  - Will parallel processing and/or partitioning be used?
  - Will data migration tools be used to load the data warehouse?
  - What tools will be used to support data retrieval and analysis?

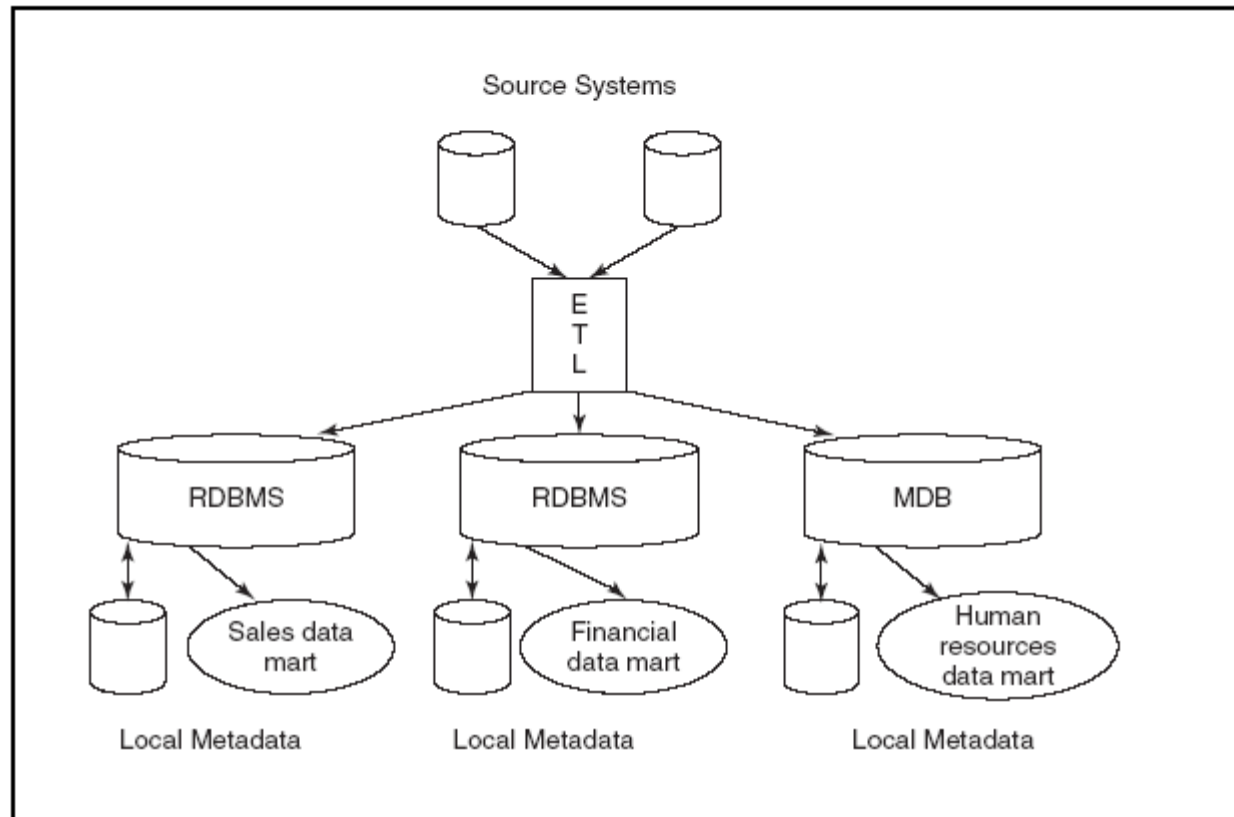
# Others Data Warehousing Architectures

FIGURE 5.5 Alternative Data Warehouse Architectures



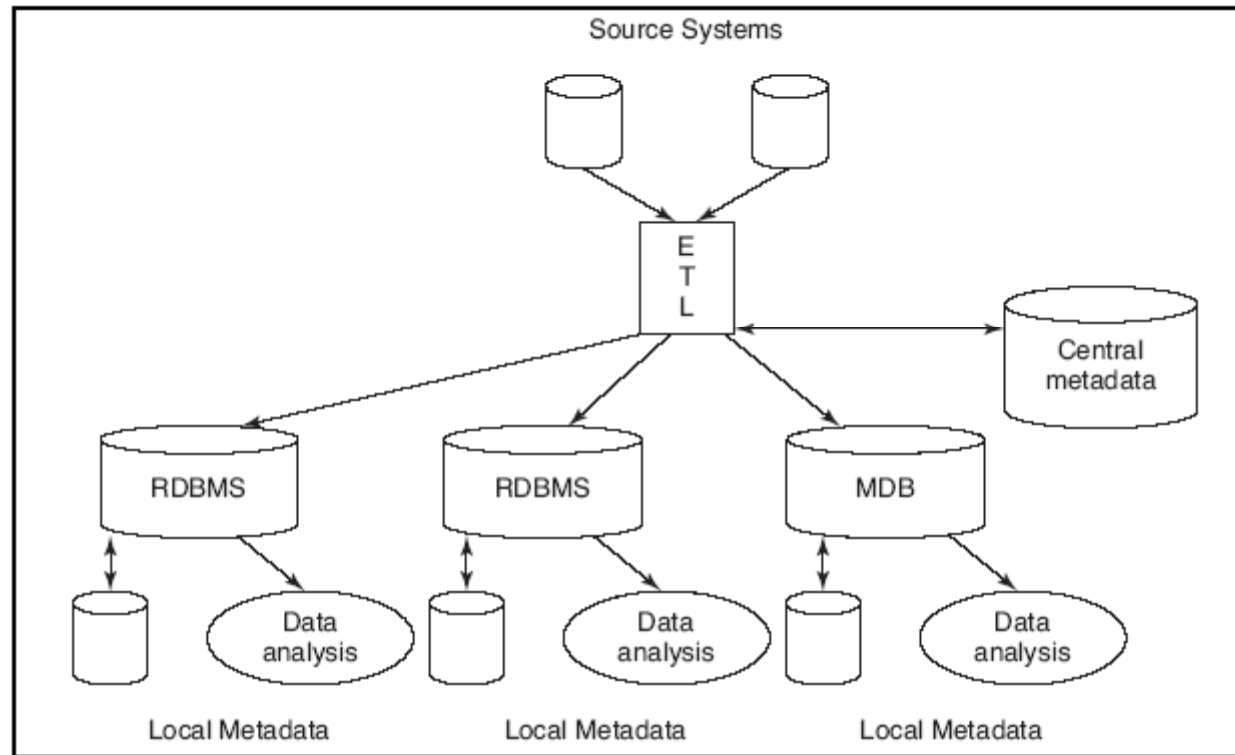
5.5a Enterprise Data Warehousing Architecture

# Others Data Warehousing Architectures



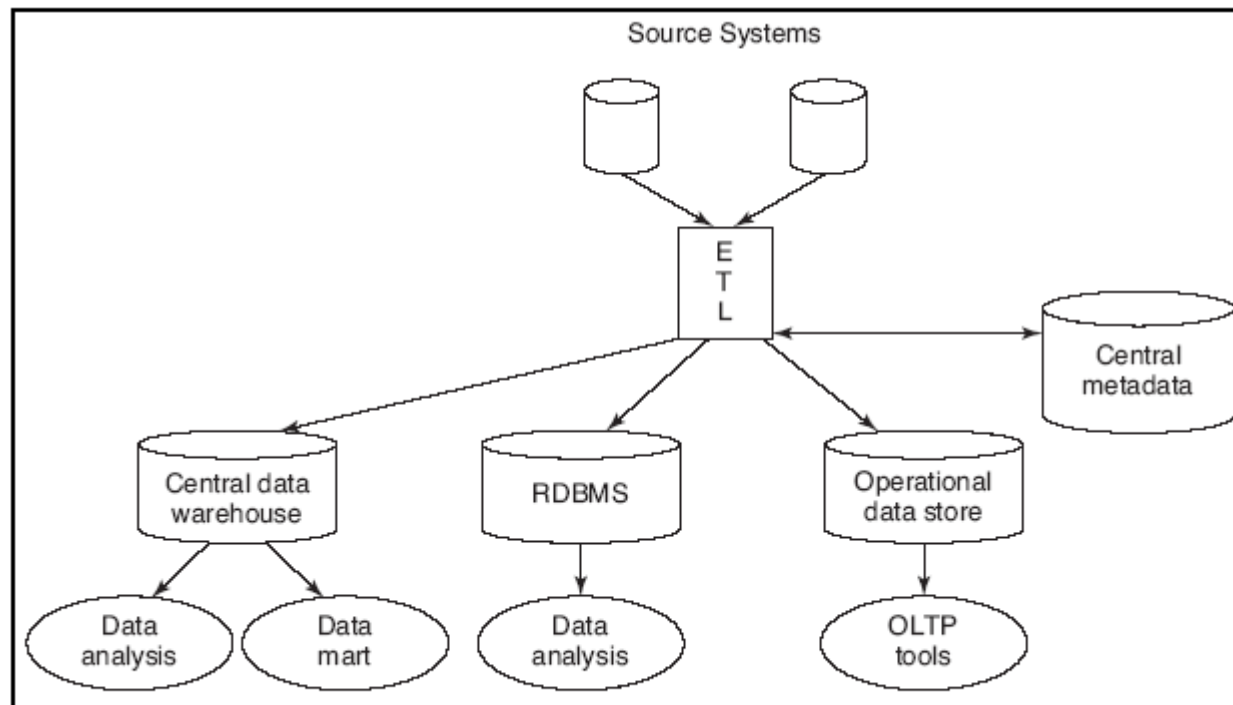
**FIGURE 5.5b** Data Mart Architecture

# Others Data Warehousing Architectures



**FIGURE 5.5c** Hub-and-Spoke Data Mart Architecture

# Others Data Warehousing Architectures



**FIGURE 5.5d** Enterprise Warehouse and Operational Data Store



# Data Warehousing Architectures

- Ten factors that potentially affect the architecture selection decision:
  - Information interdependence between organizational units
  - Upper management's information needs
  - Urgency of need for a data warehouse
  - Nature of end-user tasks
  - Constraints on resources

# Data Warehousing Architectures

- Ten factors that potentially affect the architecture selection decision:
  - Strategic view of the data warehouse prior to implementation
  - Compatibility with existing systems
  - Perceived ability of the in-house IT staff
  - Technical issues
  - Social/political factors

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

- Data integration

- Integration that comprises three major processes: data access, data federation, and change capture.
- When these three processes are correctly implemented, data can be accessed and made accessible to an array of ETL and analysis tools and data warehousing environment

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

- Enterprise application integration (EAI)
  - an integration framework composed of a collection of technologies and services which form a middleware to enable integration of systems and applications across the enterprise.

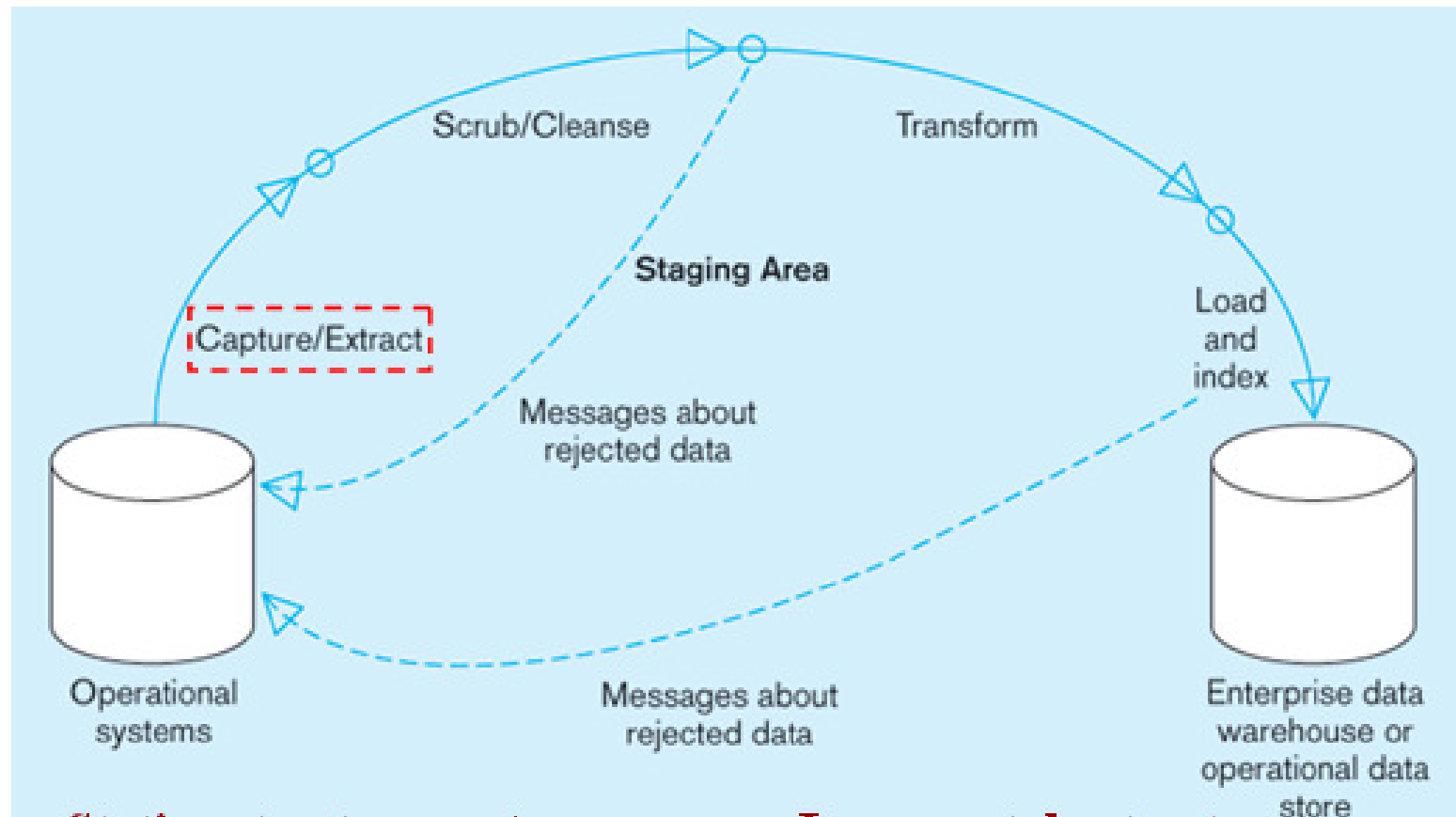
# Data Integration and the Extraction, Transformation, and Load (ETL) Process

- Enterprise information integration (EII)
  - A process of information integration, using data abstraction to provide a single interface for viewing all the data within an organization, and a single set of structures and naming conventions to represent this data

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

- Extraction, transformation, and load (ETL)
  - A data warehousing process that consists of extraction (i.e., reading data from a database)
  - transformation (i.e., converting the extracted data from its previous form into the form in which it needs to be so that it can be placed into a data warehouse or simply another database)
  - load (i.e., putting the data into the data warehouse)

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



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

- Issues affect whether an organization will purchase data transformation tools or build the transformation process itself
  - Data transformation tools are expensive
  - Data transformation tools may have a long learning curve
  - It is difficult to measure how the IT organization is doing until it has learned to use the data transformation tools



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

- Important criteria in selecting an ETL tool
  - Ability to read from and write to an unlimited number of data source architectures
  - Automatic capturing and delivery of metadata
  - A history of conforming to open standards
  - An easy-to-use interface for the developer and the functional user

# Benefits of a data warehouse

- Direct benefits of a data warehouse
  - Allows end users to perform extensive analysis
  - Allows a consolidated view of corporate data
  - Better and more timely information
  - Enhanced system performance
  - Simplification of data access

# Benefits of a data warehouse

- Indirect benefits result from end users using these direct benefits
  - Enhance business knowledge
  - Present competitive advantage
  - Enhance customer service and satisfaction
  - Facilitate decision making
  - Help in reforming business processes

# Data Warehouse Development

- Data warehouse vendors
  - Six guidelines to considered when developing a vendor list:
    - Financial strength
    - ERP linkages
    - Qualified consultants
    - Market share
    - Industry experience
    - Established partnerships

# Data Warehouse Development

- Data warehouse development approaches
  - Inmon Model: EDW approach
  - Kimball Model: Data mart approach

# Inmon Model

- Recommends to start with building a large centralized enterprise-wide data warehouse, followed by several satellite databases to serve the analytical needs of departments (later known as “data marts”). Hence, his approach has received the “Top Down” title.
- stated that the data warehouse should be modeled using an E-R model/normalized model.

# Kimball Model

- Recommends to start with building several data marts that serve the analytical needs of departments, followed by “virtually” integrating these data marts for consistency through an Information Bus. Hence, his approach received the “Bottom Up” title.
- stated that the data warehouse should be modeled using a Dimensional Model/star schema.

# Inmon Model VS. Kimball Model

**TABLE 2.4** Essential Differences Between Inmon and Kimball's Approaches

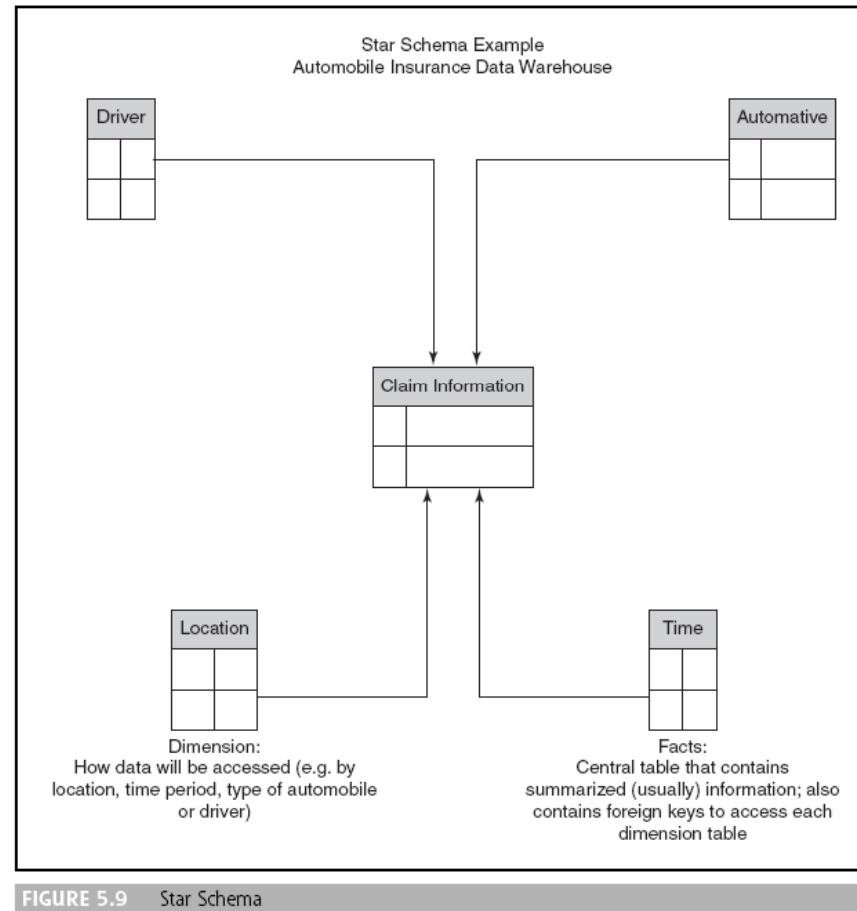
Characteristic	Inmon	Kimball
<i>Methodology and Architecture</i>		
Overall approach	Top-down	Bottom-up
Architecture structure	Enterprise-wide (atomic) data warehouse "feeds" departmental databases	Data marts model a single business process, and enterprise consistency is achieved through a data bus and conformed dimensions
Complexity of the method	Quite complex	Fairly simple
Comparison with established development methodologies	Derived from the spiral methodology	Four-step process; a departure from RDBMS methods
Discussion of physical design	Fairly thorough	Fairly light
<i>Data Modeling</i>		
Data orientation	Subject or data driven	Process oriented
Tools	Traditional (ERD, data flow diagrams)	Dimensional modeling; a departure from relational modeling



# Data Warehouse Development

- Data warehouse structure: The Star Schema
  - Dimensional modeling
    - A retrieval-based system that supports high-volume query access
  - Dimension tables
    - A table that address *how* data will be analyzed

# Data Warehouse Development



# Data Warehouse Development

- Grain
  - A definition of the highest level of detail that is supported in a data warehouse
- Drill-down
  - The process of probing beyond a summarized value to investigate each of the detail transactions that comprise the summary

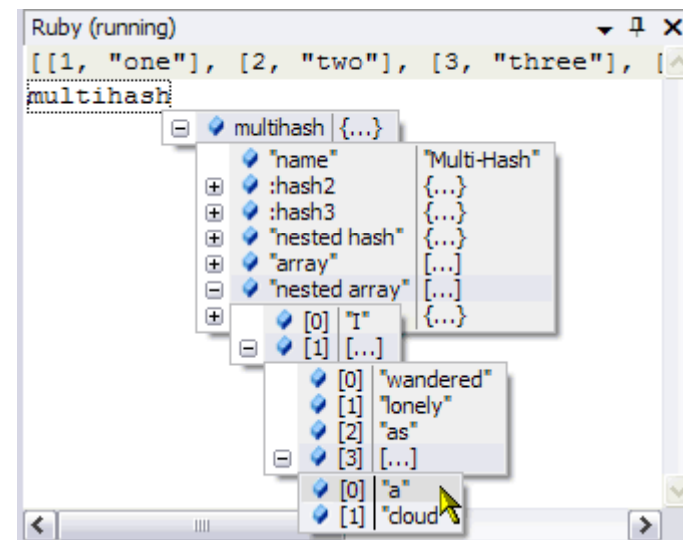
# Drill-down

Summary report

Brand	Package size	Sales
SofTowel	2-pack	\$75
SofTowel	3-pack	\$100
SofTowel	6-pack	\$50

Drill-down with color added

Brand	Package size	Color	Sales
SofTowel	2-pack	White	\$30
SofTowel	2-pack	Yellow	\$25
SofTowel	2-pack	Pink	\$20
SofTowel	3-pack	White	\$50
SofTowel	3-pack	Green	\$25
SofTowel	3-pack	Yellow	\$25
SofTowel	6-pack	White	\$30
SofTowel	6-pack	Yellow	\$20



# Data warehousing implementation issues

- Some best practices for implementing a data warehouse (Weir, 2002):
  - Project must fit with corporate strategy and business objectives
  - There must be complete buy-in to the project by executives, managers, and users
  - It is important to manage user expectations about the completed project
  - The data warehouse must be built incrementally
  - Build in adaptability

# Data warehousing implementation issues

- Some best practices for implementing a data warehouse (Weir, 2002):
  - The project must be managed by both IT and business professionals
  - Develop a business/supplier relationship
  - Only load data that have been cleansed and are of a quality understood by the organization
  - Do not overlook training requirements
  - Be politically aware

# Data warehousing implementation issues

- Failure factors in data warehouse projects:

- Cultural issues being ignored
- Inappropriate architecture
- Unclear business objectives
- Missing information
- Unrealistic expectations
- Low levels of data summarization
- Low data quality

# Data warehousing implementation issues

- Failure factors in data warehouse projects:
  - Starting with the wrong sponsorship chain
  - Setting expectations that you cannot meet and frustrating executives at the moment of truth
  - Loading the warehouse with information just because it is available



# Data warehousing implementation issues

- Failure factors in data warehouse projects:

- Believing that data warehousing database design is the same as transactional database design
- Choosing a data warehouse manager who is technology oriented rather than user oriented
- Focusing on traditional internal record-oriented data and ignoring the value of external data and of text, images, and, perhaps, sound and video

# Data warehousing implementation issues

- Failure factors in data warehouse projects:

- Delivering data with overlapping and confusing definitions
- Believing promises of performance, capacity, and scalability
- Believing that your problems are over when the data warehouse is up and running
- Focusing on ad hoc data mining and periodic reporting instead of alerts

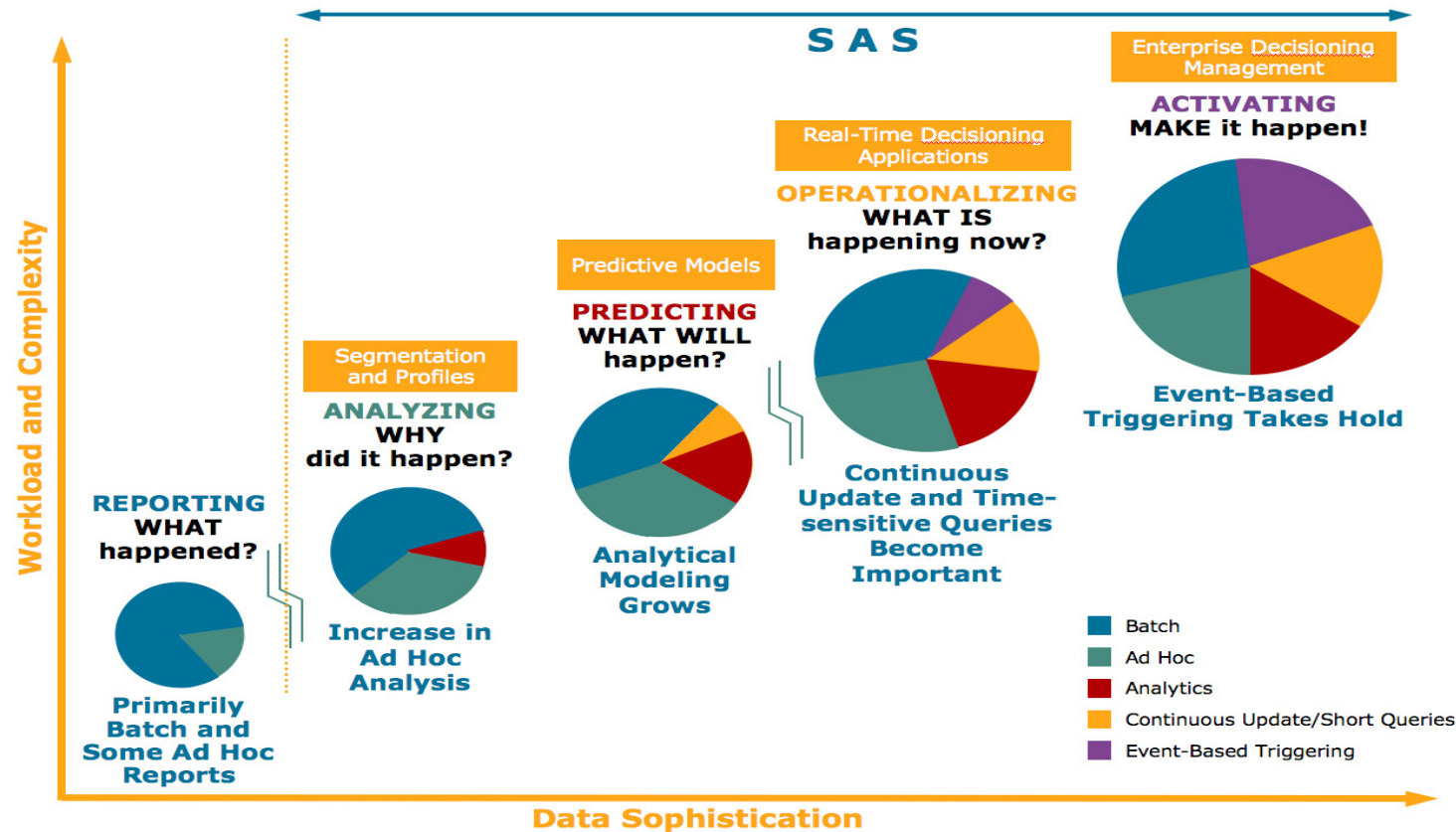
# Real Time Data Warehousing

- Real-time (active) data warehousing
  - The process of loading and providing data via a data warehouse as they become available
  - Enabling real-time data updates for real-time analysis and real-time decision making is growing rapidly

# Real Time Data Warehousing

- Levels of data warehouses:
  - Reports what happened
  - Some analysis occurs
  - Provides prediction capabilities,
  - Operationalization
  - Becomes capable of making events happen

# Real Time Data Warehousing



# Real Time Data Warehousing

## Active Access

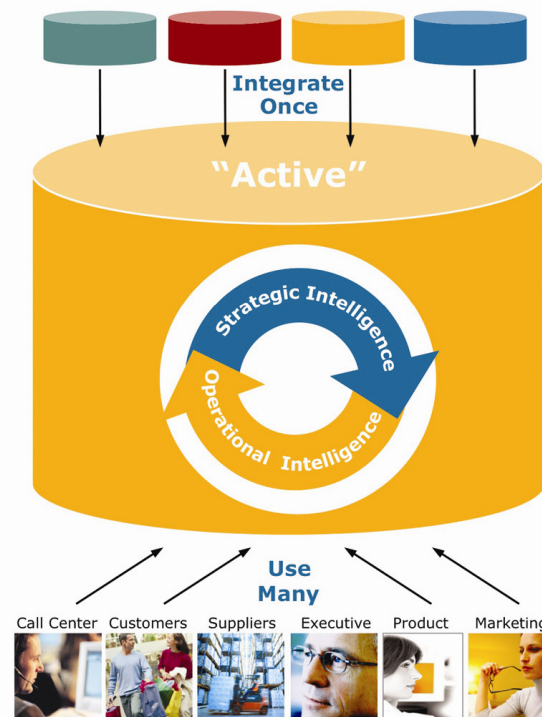
Front-Line operational decisions or services supported by near-real-time (NRT) access; Service Level Agreements of 5 seconds or less

## 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)

# Traditional DW VS. Active DW

<b>Traditional Data Warehousing Environment</b>	<b>Active Data Warehousing Environment</b>
Strategic decisions only	Strategic and tactical decisions
Results sometimes hard to measure	Results measured with operations
Daily, weekly, monthly data currency is acceptable; summaries are often appropriate	Only comprehensive detailed data available within minutes is acceptable
Moderate user concurrency	High number (1,000 or more) of users accessing and querying the system simultaneously
Highly restrictive reporting used to confirm or check existing processes and patterns; often uses predeveloped summary tables or data marts	Flexible ad hoc reporting as well as machine-assisted modeling (e.g., data mining) to discover new hypotheses and relationships
Power users, knowledge workers, internal users	Operational staffs, call centers, external users

# Real Time Data Warehousing

- Concerns about real-time DW
  - Not all data should be updated continuously
  - Mismatch of reports generated minutes apart
  - May be cost prohibitive
  - May also be infeasible



# Data Warehouse Administration and Security Issues

- Data warehouse administrator (DWA)
  - a person responsible for the administration and management of a data warehouse
  - have the knowledge of high-performance software, hardware and networking technologies.
  - possess solid business knowledge and insight.
  - be familiar with the decision-making processes so as to suitably design/maintain the data warehouse structure.
  - possess excellent communications skills.

# Data Warehouse Administration and Security Issues

- Effective security in a data warehouse should focus on four main areas:
  - Establishing effective corporate and security policies and procedures
  - Implementing logical security procedures and techniques to restrict access
  - Limiting physical access to the data center environment
  - Establishing an effective internal control review process with an emphasis on security and privacy

# References

- Sharda, R., Delen, D., Turban, E., 2018. Business Intelligence, Analytics, and Data Science : A Managerial Perspective. Pearson.
- Samaddar, S. and Nargundkar, S.. 2019. Data Analytics : Effective Methods for Presenting Results. CRC Press.
- Verbeke, W., Baesens, B. and Bravo, C.. 2018. Profit Drive Business Analytics : A Practitioner's Guide to Transforming Big Data into Added Value. John Wiley & Sons.