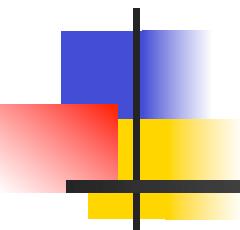
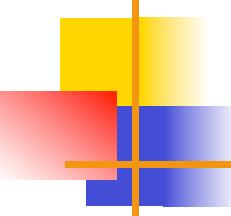


Business Intelligence and Decision Support Systems

(9th Ed., Prentice Hall)

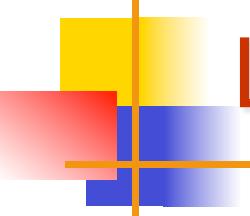


Chapter 8: Data Warehousing



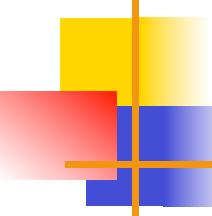
Learning Objectives

- Understand the basic definitions and concepts of data warehouses
- Learn different types of data warehousing architectures; their comparative advantages and disadvantages
- Describe the processes used in developing and managing data warehouses
- Explain data warehousing operations
- Explain the role of data warehouses in decision support



Learning Objectives

- Explain data integration and the extraction, transformation, and load (ETL) processes
- Describe real-time (a.k.a. right-time and/or active) data warehousing
- Understand data warehouse administration and security issues

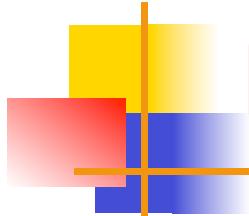


Opening Vignette:

“DirecTV Thrives with Active Data Warehousing”

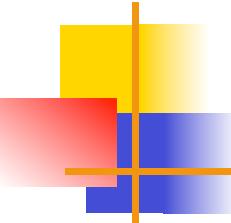
- Company background
- Problem description
- Proposed solution
- Results
- Answer & discuss the case questions.





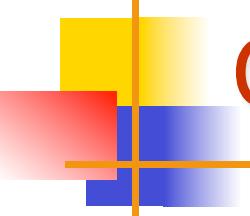
Main Data Warehousing (DW) Topics

- DW definitions
- Characteristics of DW
- Data Marts
- ODS, EDW, Metadata
- DW Framework
- DW Architecture & ETL Process
- DW Development
- DW Issues



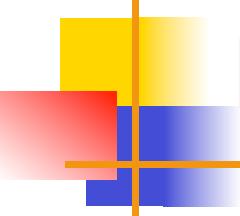
Data Warehouse Defined

- A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
- “The data warehouse is a collection of integrated, subject-oriented databases design to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time”



Characteristics of DW

- Subject oriented
- Integrated
- Time-variant (time series)
- Nonvolatile
- Summarized
- Not normalized
- Metadata
- Web based, relational/multi-dimensional
- Client/server
- Real-time and/or right-time (active)



Data Mart

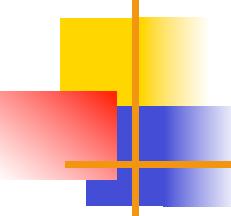
A departmental data warehouse that stores only relevant data

- **Dependent data mart**

A subset that is created directly from a data warehouse

- **Independent data mart**

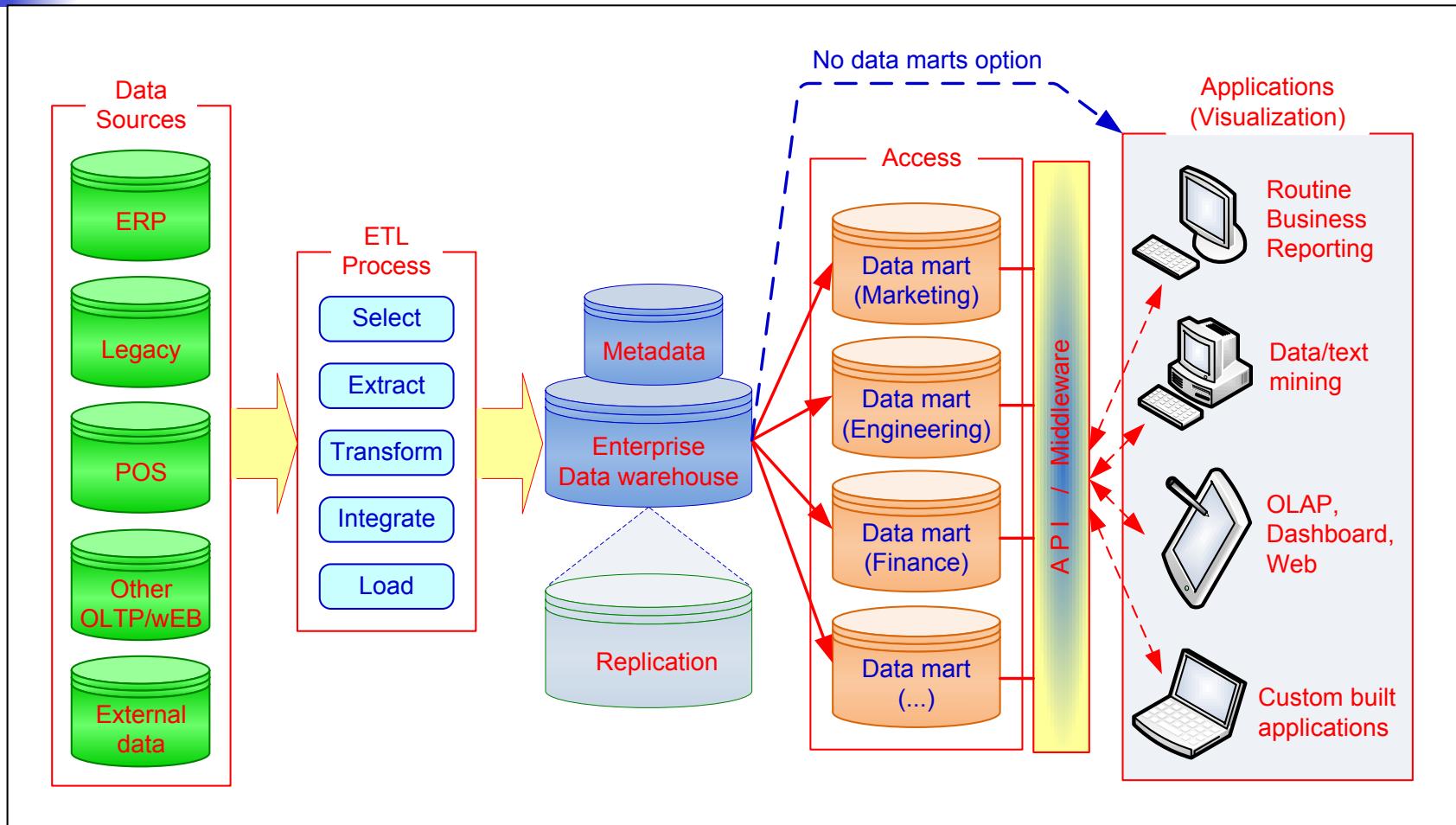
A small data warehouse designed for a strategic business unit or a department

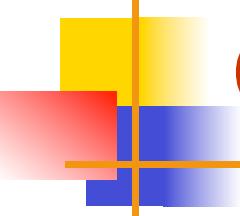


Data Warehousing Definitions

- **Operational data stores (ODS)**
 - A type of database often used as an interim area for a data warehouse
- **Oper marts**
 - An operational data mart.
- **Enterprise data warehouse (EDW)**
 - A data warehouse for the enterprise.
- **Metadata**
 - Data about data. In a data warehouse, metadata describe the contents of a data warehouse and the manner of its acquisition and use

A Conceptual Framework for DW





Generic DW Architectures

■ Three-tier architecture

1. Data acquisition software (back-end)
2. The data warehouse that contains the data & software
3. Client (front-end) software that allows users to access and analyze data from the warehouse

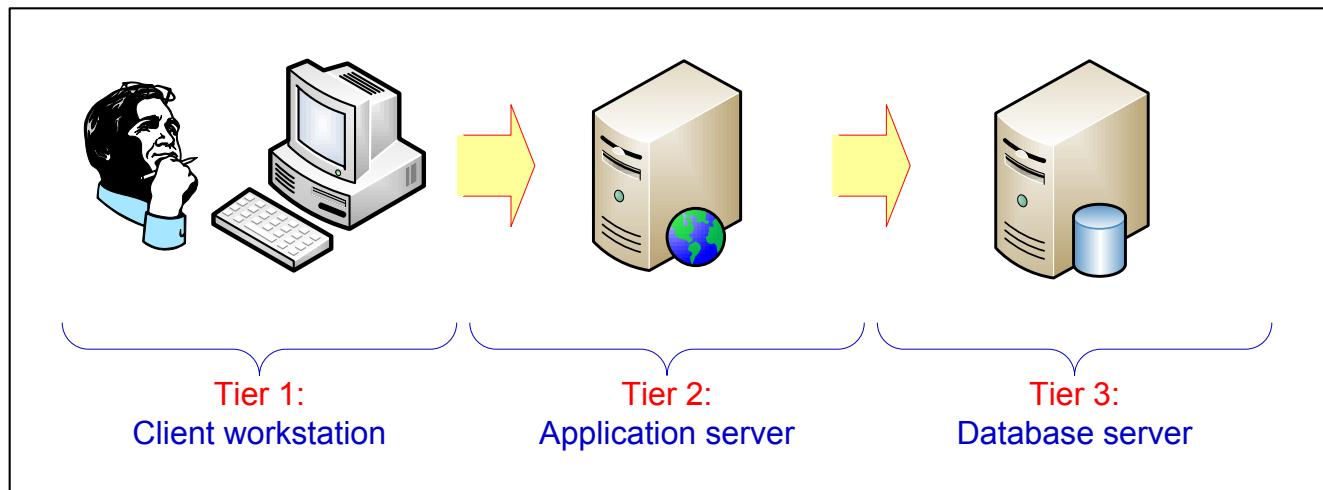
■ Two-tier architecture

First 2 tiers in three-tier architecture is combined into one

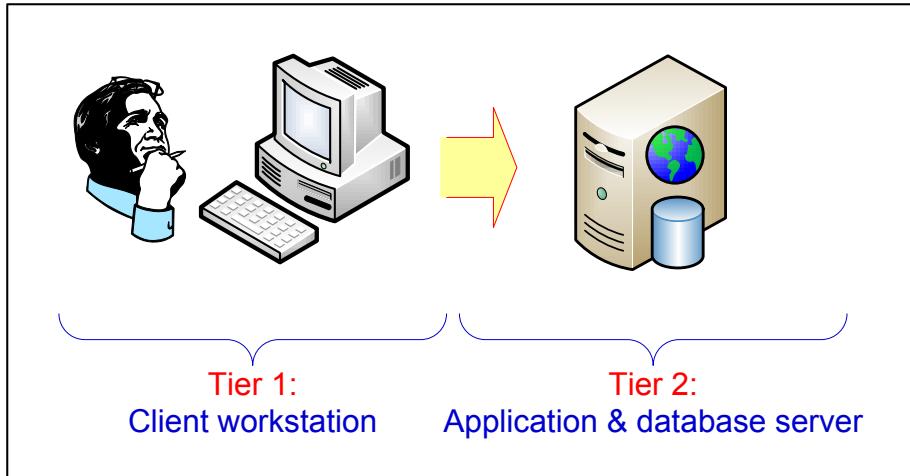
... sometime there is only one tier?

Generic DW Architectures

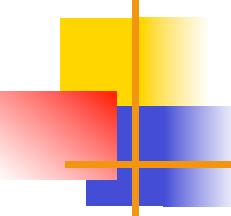
**3-tier
architecture**



**2-tier
architecture**



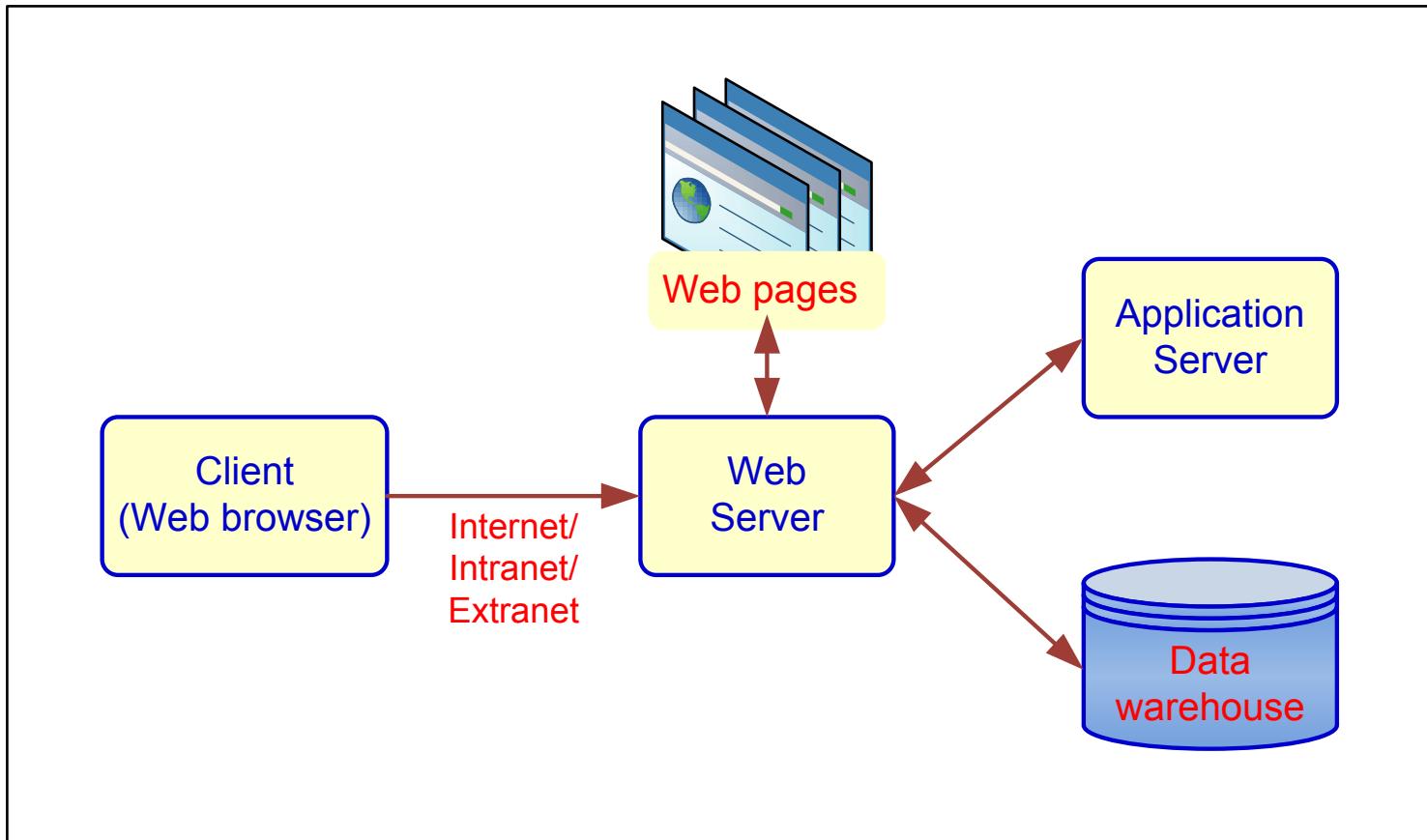
**1-tier
Architecture
?**



DW Architecture Considerations

- 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?

A Web-based DW Architecture

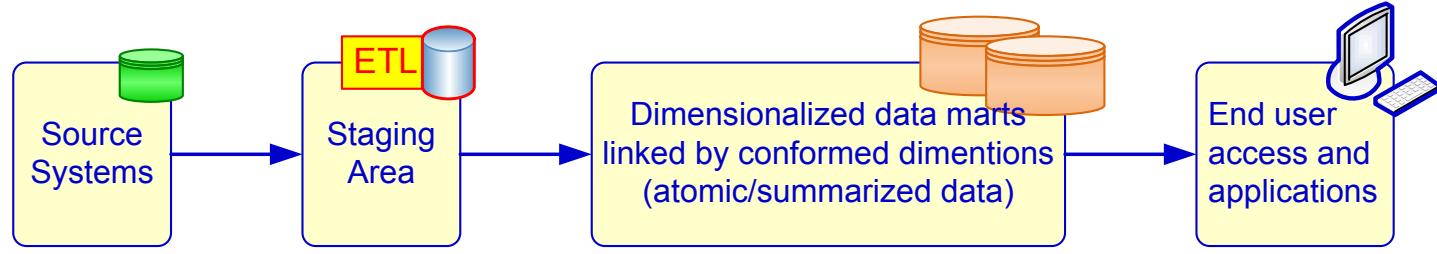


Alternative DW Architectures

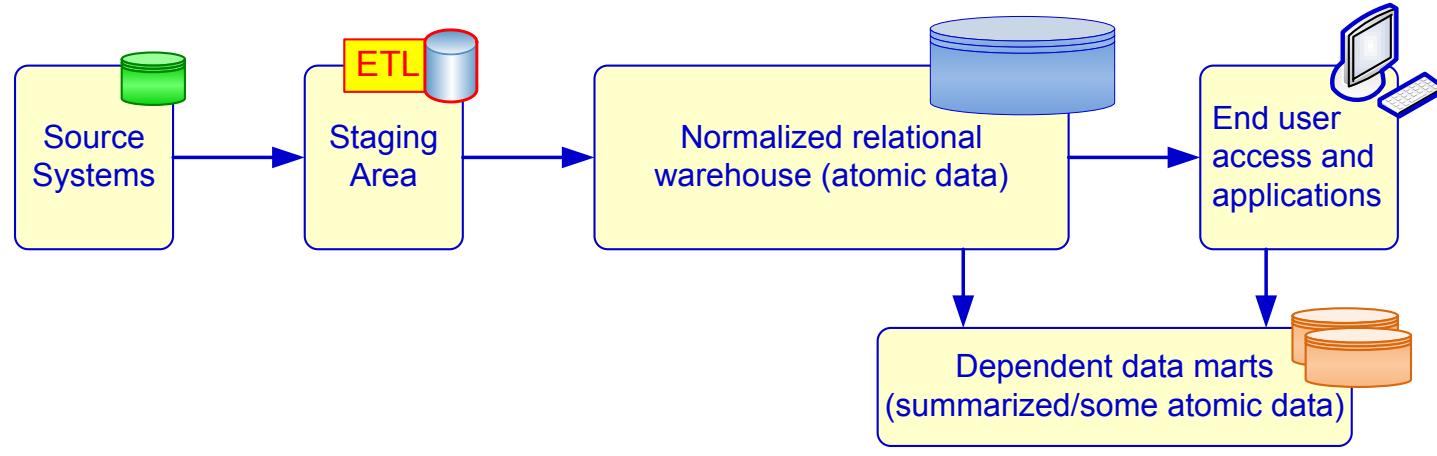
(a) Independent Data Marts Architecture



(b) Data Mart Bus Architecture with Linked Dimensional Data Marts

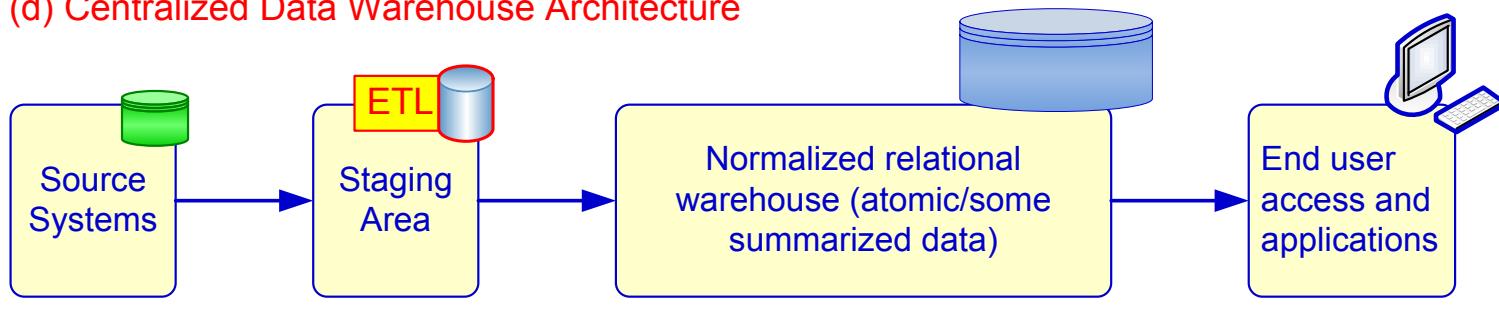


(c) Hub and Spoke Architecture (Corporate Information Factory)

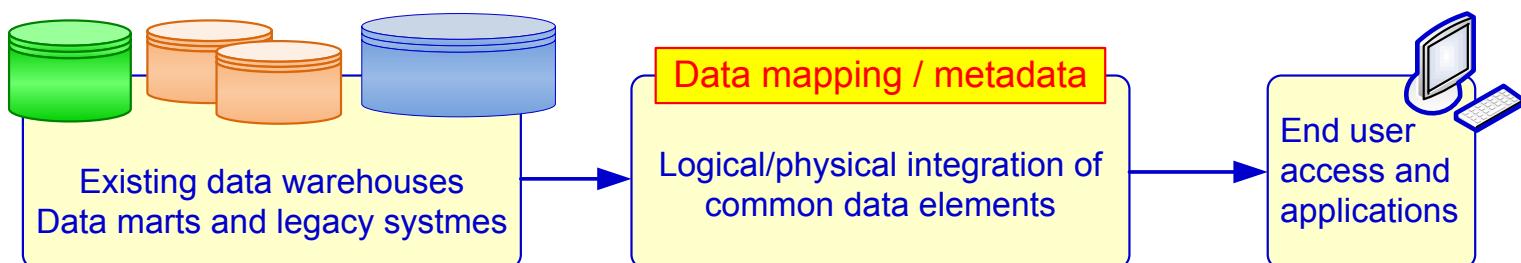


Alternative DW Architectures

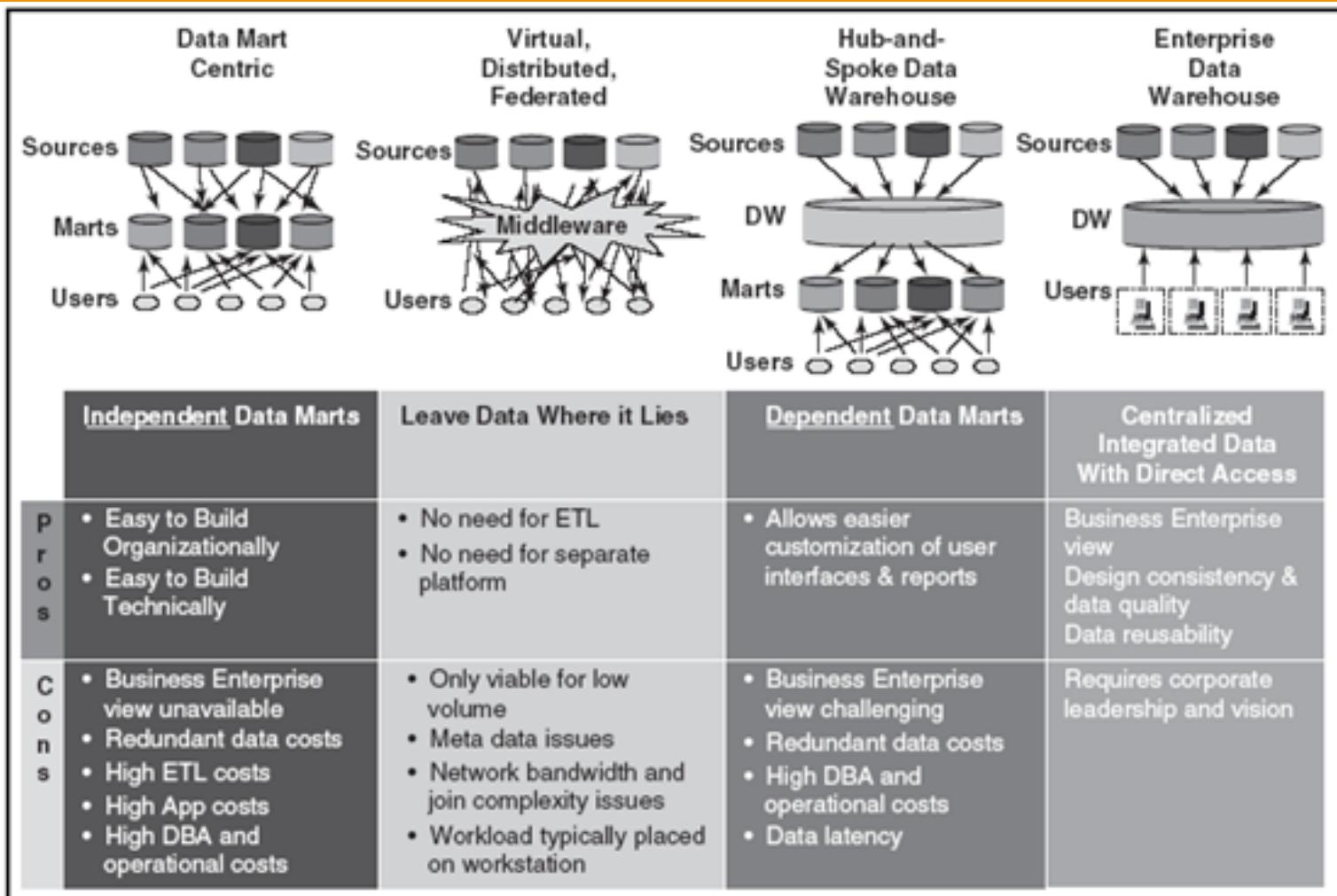
(d) Centralized Data Warehouse Architecture

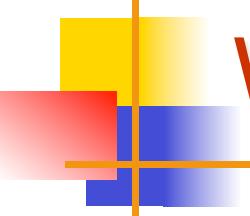


(e) Federated Architecture



Alternative DW Architectures



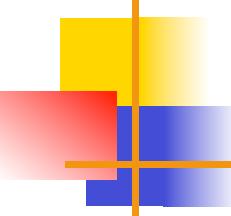


Which Architecture is the Best?

- Bill Inmon versus Ralph Kimball
 - Enterprise DW versus 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)

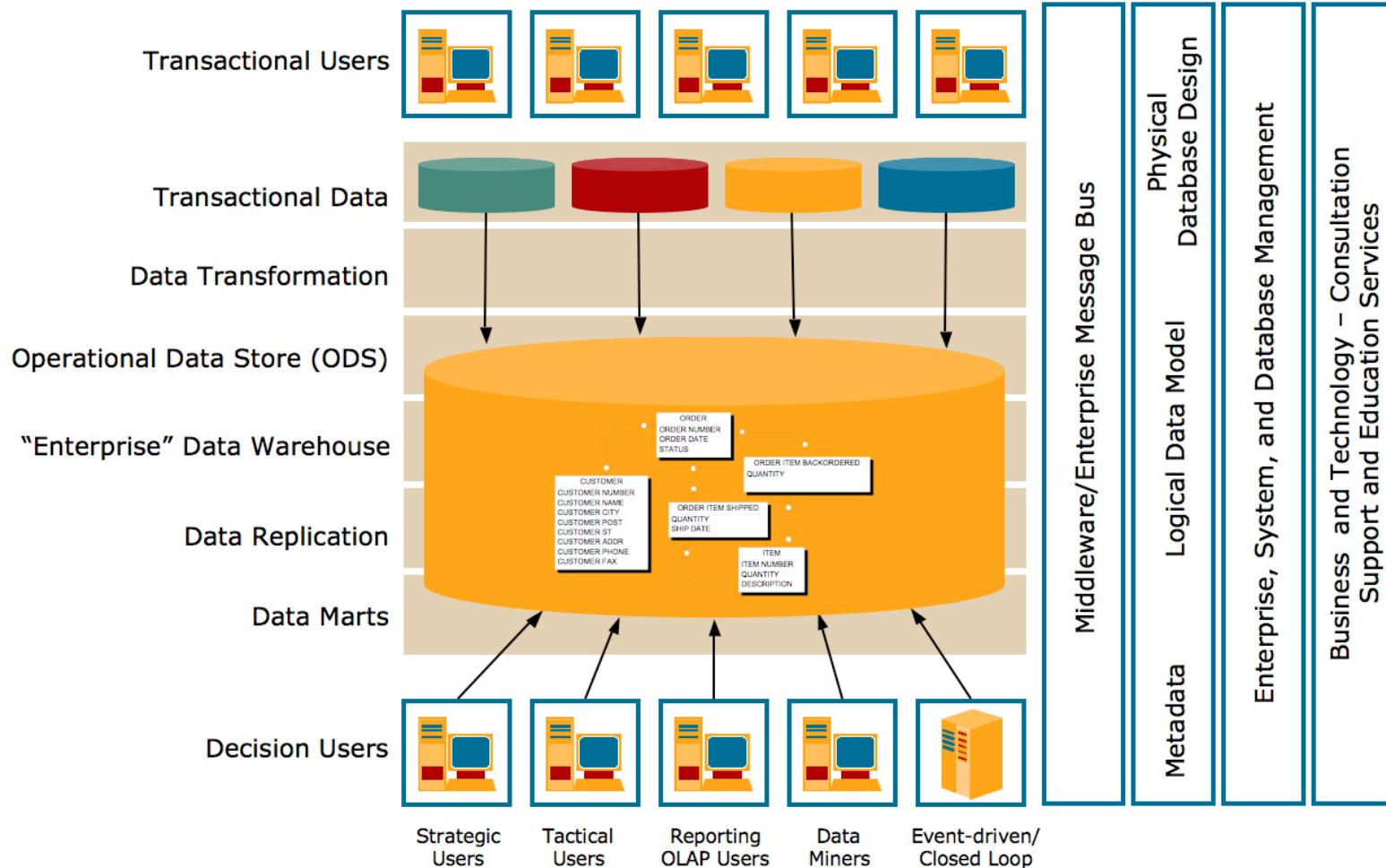


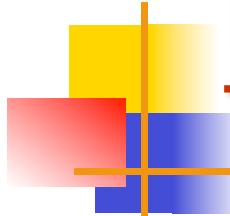
Data Warehousing Architectures

Ten factors that potentially affect the architecture selection decision:

1. Information interdependence between organizational units
2. Upper management's information needs
3. Urgency of need for a data warehouse
4. Nature of end-user tasks
5. Constraints on resources
6. Strategic view of the data warehouse prior to implementation
7. Compatibility with existing systems
8. Perceived ability of the in-house IT staff
9. Technical issues
10. Social/political factors

Enterprise Data Warehouse (by Teradata Corporation)





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.

- **Enterprise application integration (EAI)**

- A technology that provides a vehicle for pushing data from source systems into a data warehouse

- **Enterprise information integration (EII)**

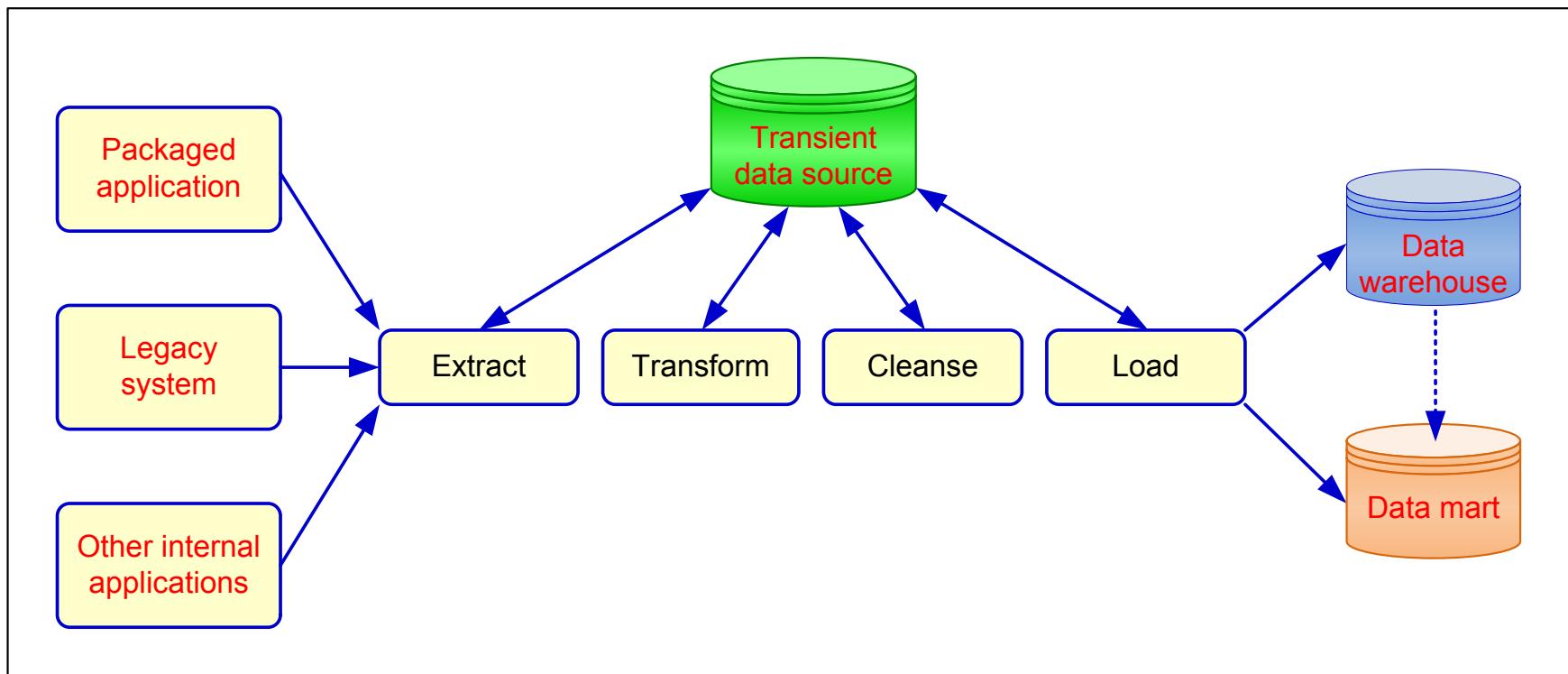
- An evolving tool space that promises real-time data integration from a variety of sources

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

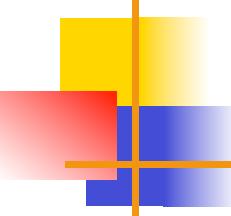
- A new way of integrating information systems

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

Extraction, transformation, and load (ETL) process

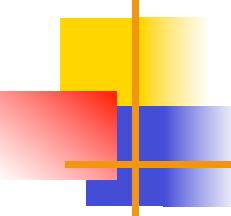


- Issues affecting the purchase of and ETL tool
 - Data transformation tools are expensive
 - Data transformation tools may have a long learning curve
- Important criteria in selecting an ETL tool
 - Ability to read from and write to an unlimited number of data sources/architectures
 - Automatic capturing and delivery of metadata
 - A history of conforming to open standards
 - An easy-to-use interface for the developer and the functional user



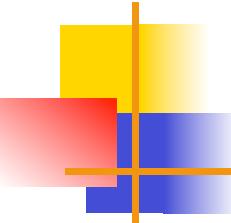
Benefits of DW

- 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
- Indirect benefits of data warehouse
 - Enhance business knowledge
 - Present competitive advantage
 - Enhance customer service and satisfaction
 - Facilitate decision making
 - Help in reforming business processes



Data Warehouse Development

- Data warehouse development approaches
 - Inmon Model: EDW approach (top-down)
 - Kimball Model: Data mart approach (bottom-up)
 - Which model is best?
 - There is no one-size-fits-all strategy to DW
 - One alternative is the hosted warehouse
- Data warehouse structure:
 - The Star Schema vs. Relational
- Real-time data warehousing?

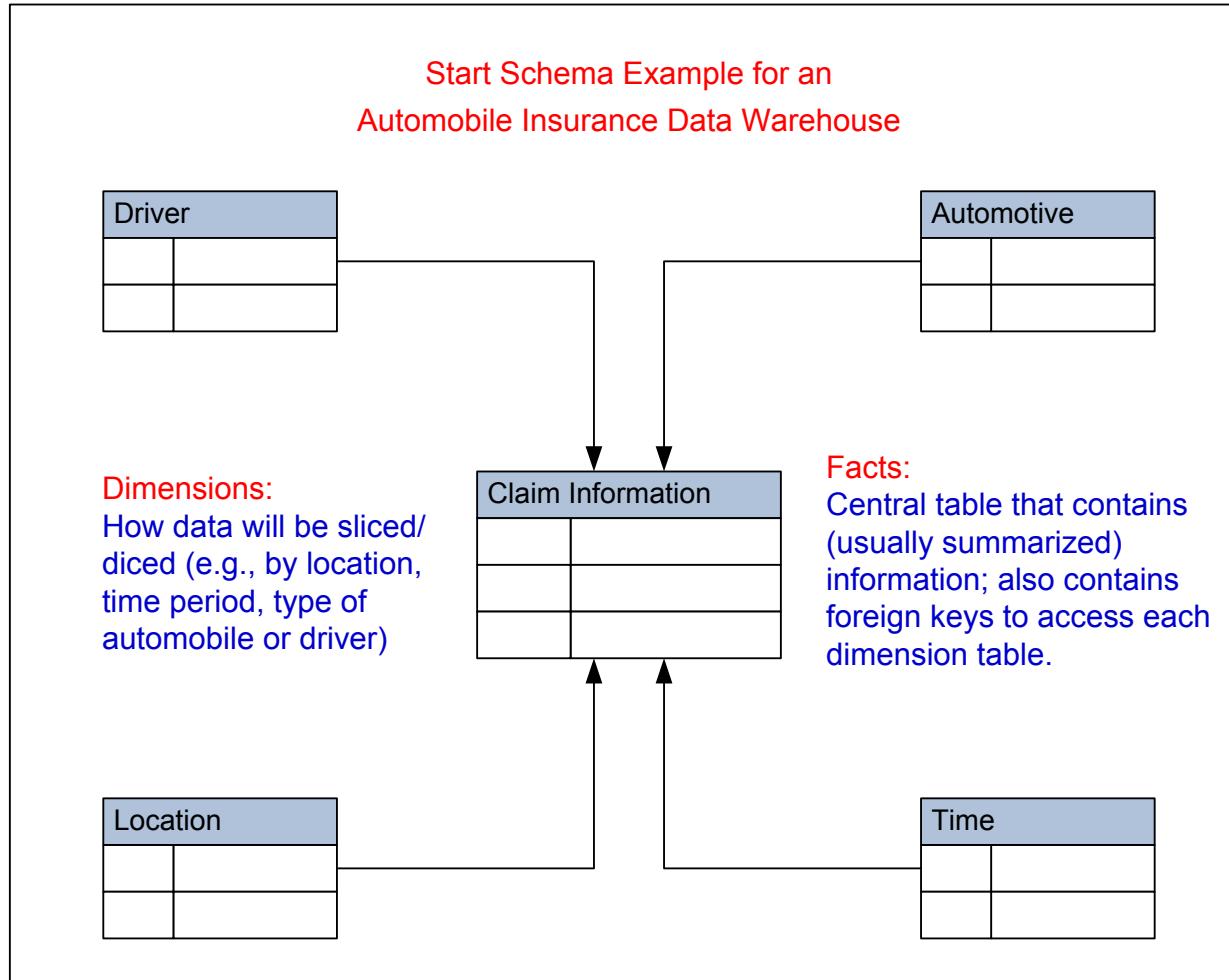


DW Development Approaches

	(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
Technology		
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)

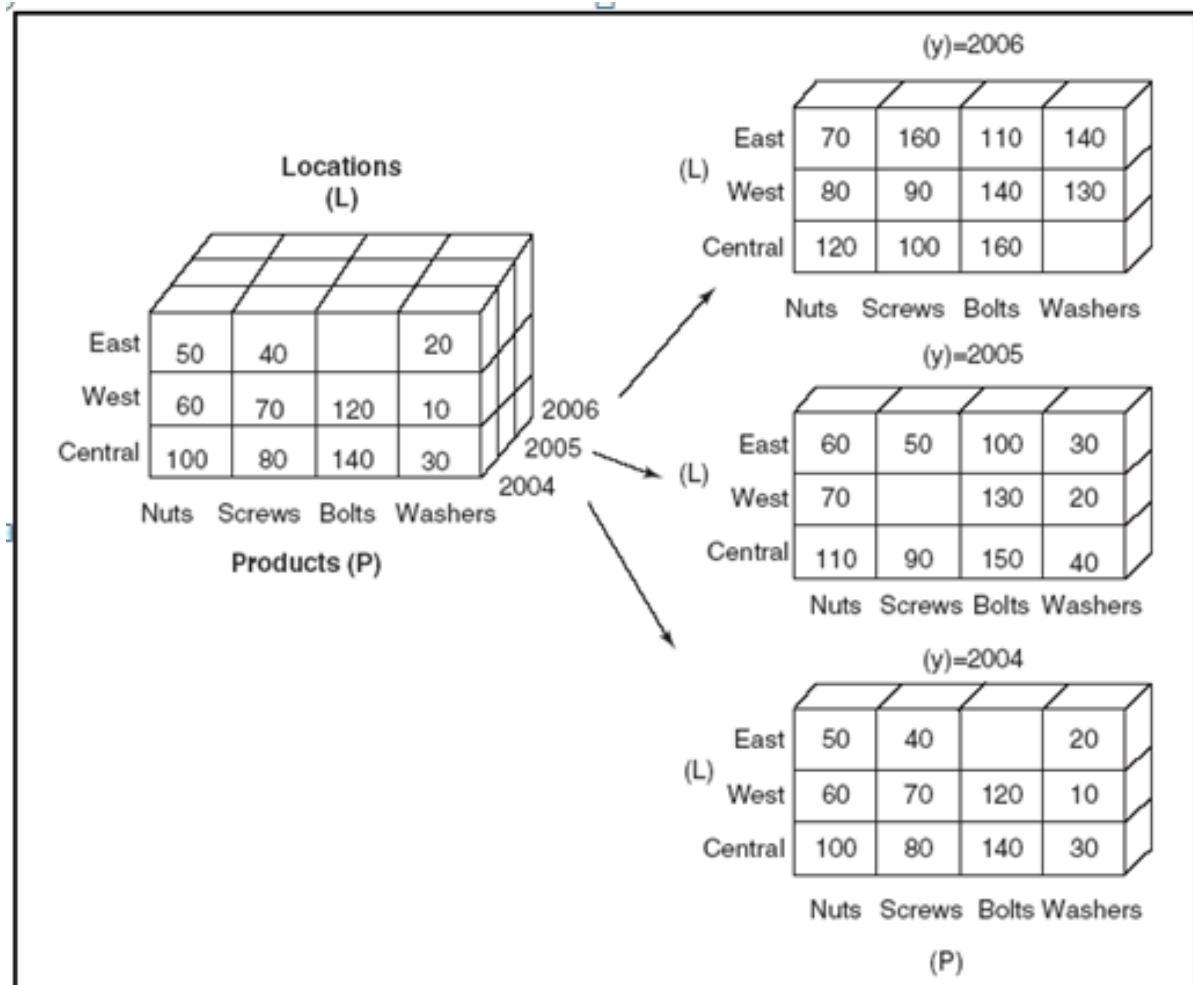


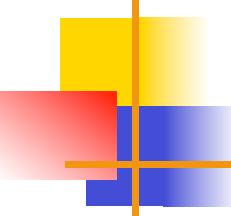
Dimensional Modeling

Data cube

A two-dimensional, three-dimensional, or higher-dimensional object in which each dimension of the data represents a measure of interest

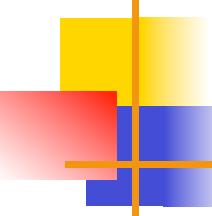
- Grain
- Drill-down
- Slicing





Best Practices for Implementing DW

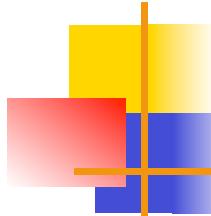
- The project must fit with corporate strategy
- There must be complete buy-in to the project
- It is important to manage user expectations
- The data warehouse must be built incrementally
- Adaptability must be built in from the start
- The project must be managed by both IT and business professionals (a business–supplier relationship must be developed)
- Only load data that have been cleansed/high quality
- Do not overlook training requirements
- Be politically aware.



Risks in Implementing DW

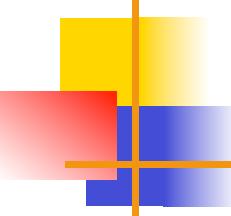
- No mission or objective
- Quality of source data unknown
- Skills not in place
- Inadequate budget
- Lack of supporting software
- Source data not understood
- Weak sponsor
- Users not computer literate
- Political problems or turf wars
- Unrealistic user expectations

(Continued ...)



Risks in Implementing DW – Cont.

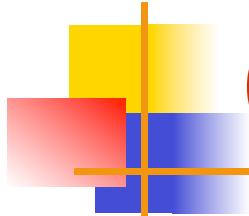
- Architectural and design risks
- Scope creep and changing requirements
- Vendors out of control
- Multiple platforms
- Key people leaving the project
- Loss of the sponsor
- Too much new technology
- Having to fix an operational system
- Geographically distributed environment
- Team geography and language culture



Things to Avoid for Successful Implementation of DW

- Starting with the wrong sponsorship chain
- Setting expectations that you cannot meet
- Engaging in politically naive behavior
- Loading the warehouse with information just because it is available
- Believing that data warehousing database design is the same as transactional DB design
- Choosing a data warehouse manager who is technology oriented rather than user oriented

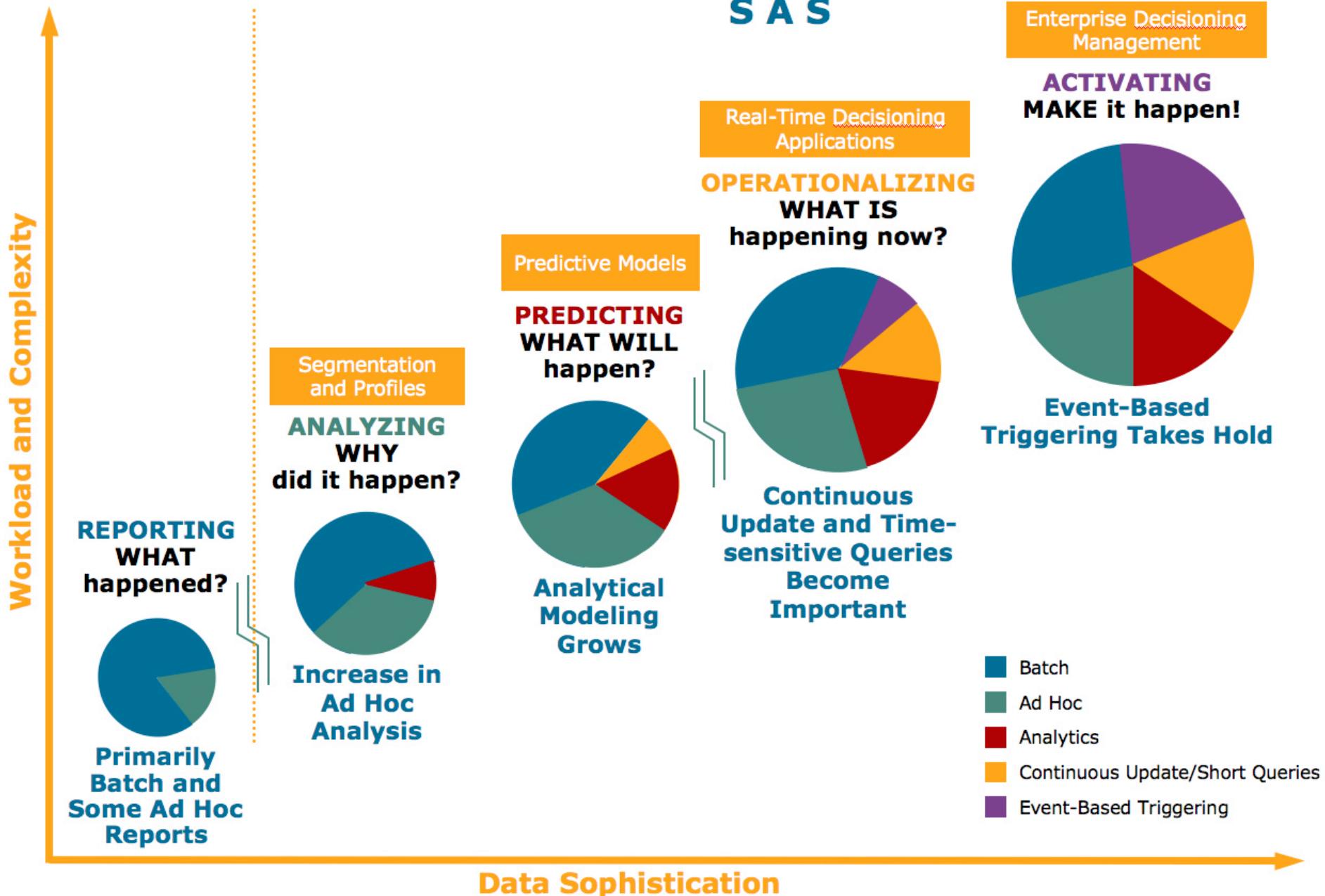
(...see more on page 356)



Real-time DW (a.k.a. Active Data Warehousing)

- Enabling real-time data updates for real-time analysis and real-time decision making is growing rapidly
 - Push vs. Pull (of data)
- Concerns about real-time BI
 - Not all data should be updated continuously
 - Mismatch of reports generated minutes apart
 - May be cost prohibitive
 - May also be infeasible

Evolution of DSS & DW



Active Data Warehousing (by Teradata Corporation)

Active Access

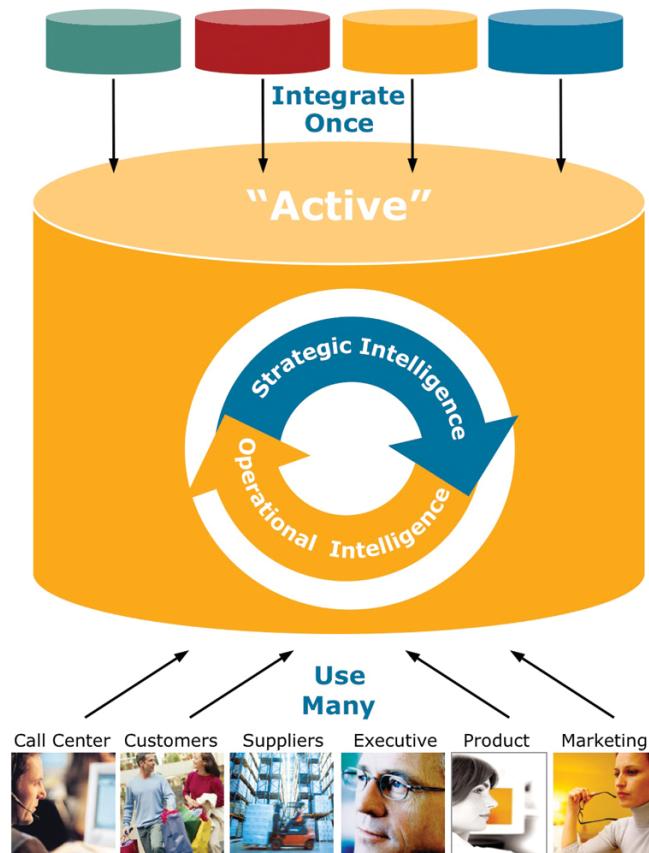
Front-Line operational decisions or services supported by near-real-time (NRT) access; Service Level Agreements of 5 seconds or less

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

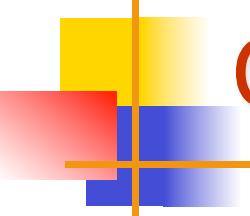
Dynamically manage system resources for optimum performance and resource utilization supporting a mixed-workload environment

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)



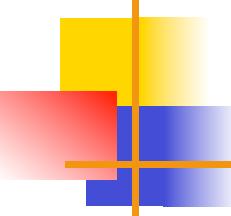
Comparing Traditional and Active DW

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

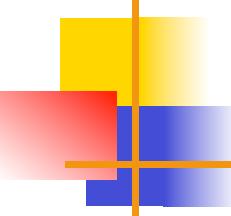
Active Data Warehouse Environment

- 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

- Due to its **huge size** and its intrinsic nature, a DW requires especially strong monitoring in order to sustain its efficiency, productivity and security.
- The successful administration and management of a data warehouse entails skills and proficiency that go past what is required of a traditional database administrator.
 - Requires expertise in high-performance software, hardware, and networking technologies

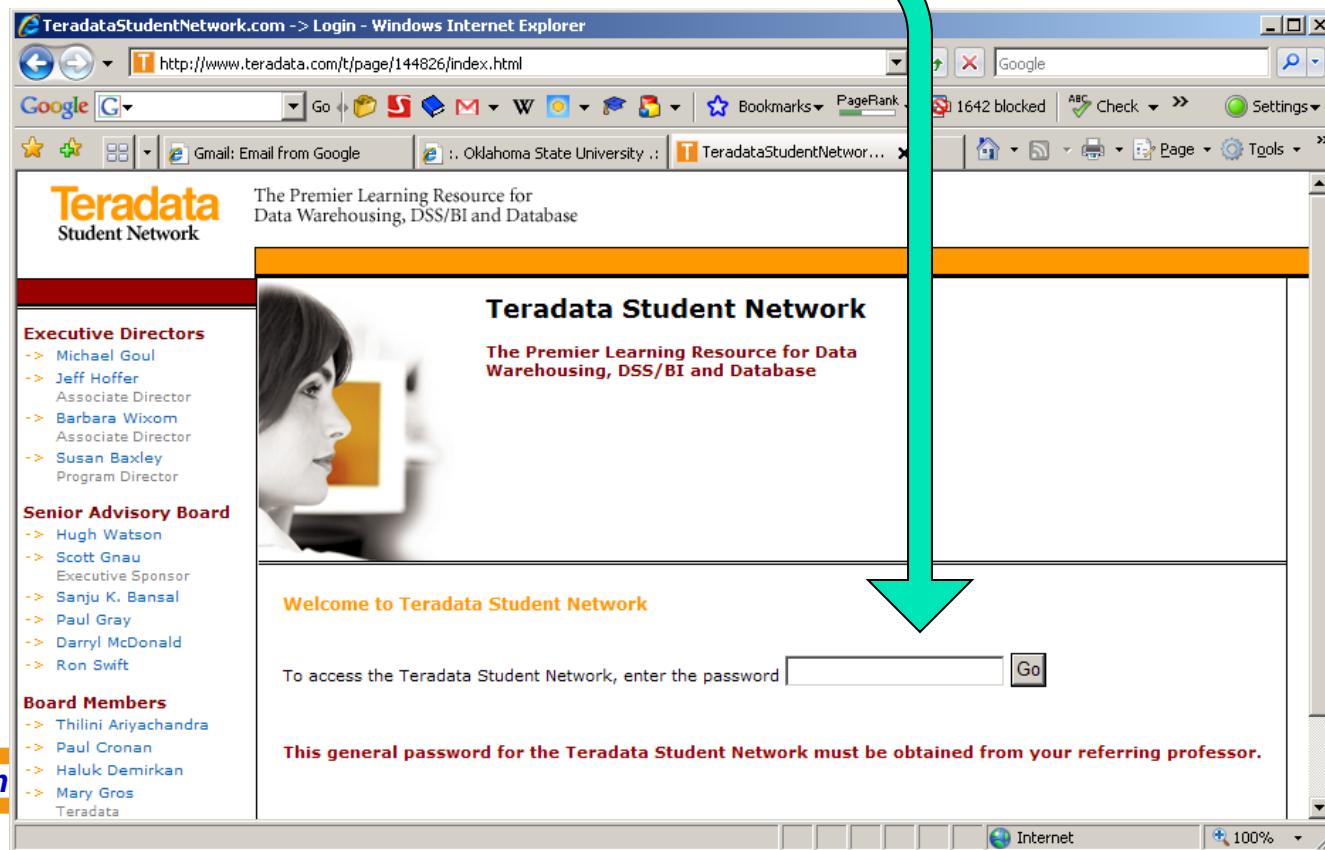


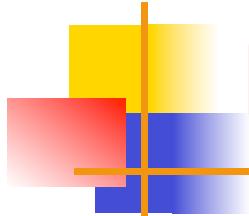
DW Scalability and Security

- Scalability
 - The main issues pertaining to scalability:
 - The amount of data in the warehouse
 - How quickly the warehouse is expected to grow
 - The number of concurrent users
 - The complexity of user queries
 - Good scalability means that queries and other data-access functions will grow linearly with the size of the warehouse
- Security
 - Emphasis on security and privacy

BI / OLAP Portal for Learning

- MicroStrategy, and much more...
- www.TeradataStudentNetwork.com
- Password: [**Keyword**]





End of the Chapter

- Questions, comments