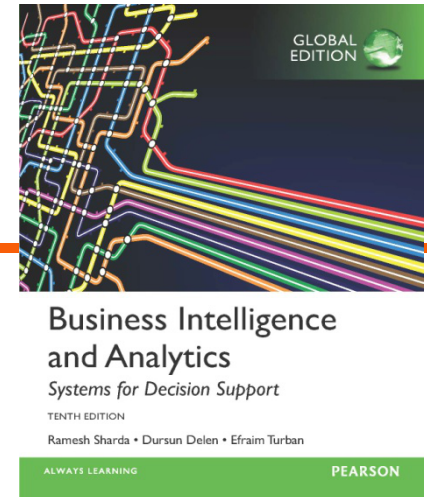


Business Intelligence and Analytics: Systems for Decision Support

Global Edition
(10th Edition)



Chapter 3: 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

(Continued...)

- Explain data warehousing operations



Learning Objectives

- Explain the role of data warehouses in decision support
- 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...

“Isle of Capri Casinos Is Winning with Enterprise Data Warehouse”

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



Main Data Warehousing Topics

- DW definition
- Characteristics of DW
- Data Marts
- ODS, EDW, Metadata
- DW Framework
- DW Architecture & ETL Process
- DW Development
- DW Issues
- **From** <https://www.youtube.com/watch?v=JcRhduFurCU>



What is a Data Warehouse?

- 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 designed to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time”



A Historical Perspective to Data Warehousing



Characteristics of DWs

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



Data Mart

A departmental small-scale “DW” that stores only limited/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



Other DW Components

- **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



Application Case 3.1

A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry

Questions for Discussion

1. What are the main challenges for TELCOs?
2. How can data warehousing and data analytics help TELCOs in overcoming their challenges?
3. Why do you think TELCOs are well suited to take full advantage of data analytics?



A Generic DW Framework



Application Case 3.2

Data Warehousing Helps MultiCare Save More Lives

Questions for Discussion

1. What do you think is the role of data warehousing in healthcare systems?
2. How did MultiCare use data warehousing to improve health outcomes?



DW Architecture

■ 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 two tiers in three-tier architecture is combined into one

... sometimes there is only one tier?



DW Architectures

**3-tier
architecture**

**2-tier
architecture**

**1-tier
Architecture
?**



Data Warehousing Architectures

- **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?



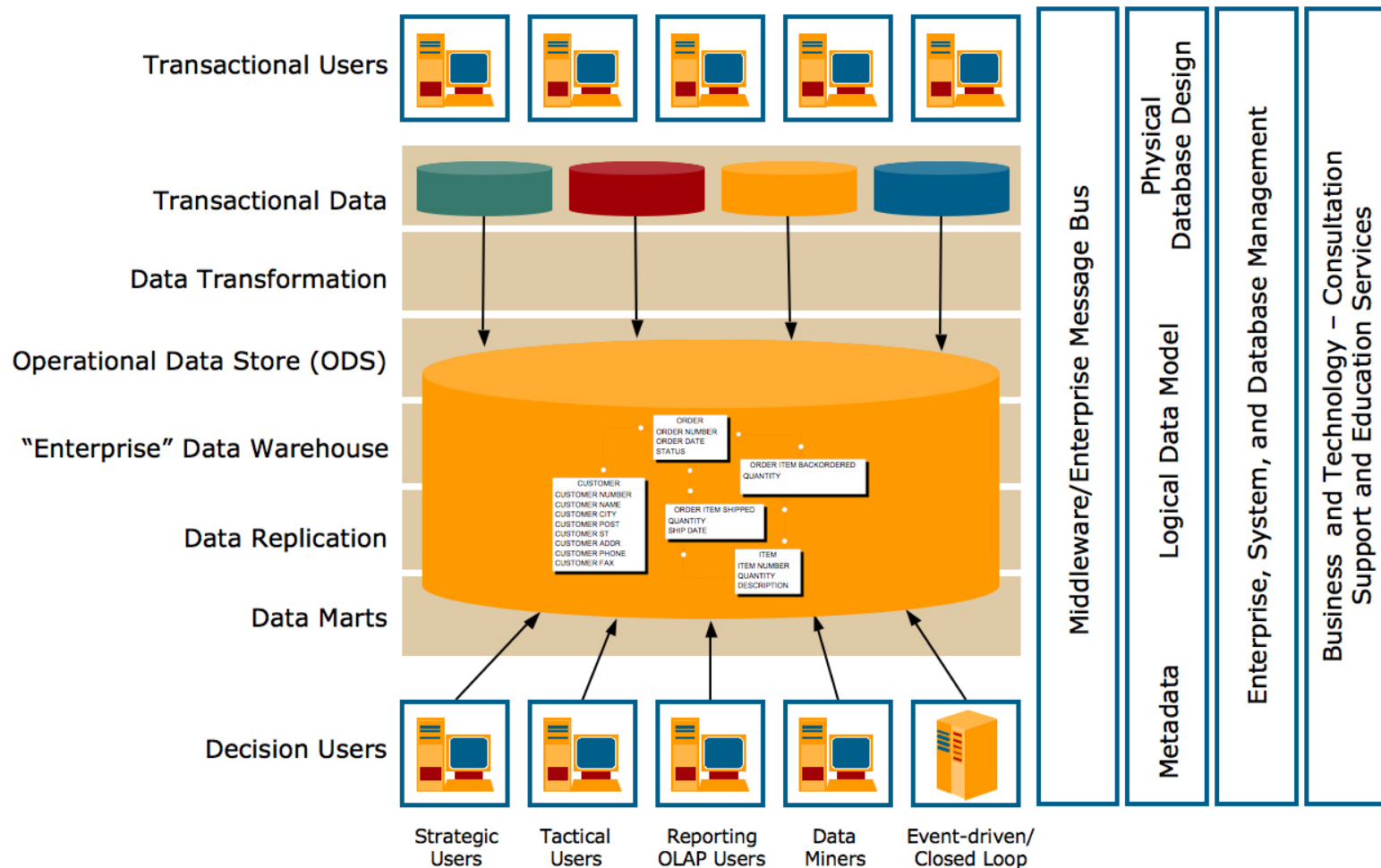
A Web-Based DW Architecture



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

Teradata Corp. DW Architecture





Data Integration and the Extraction, Transformation, and Load Process

- **ETL = Extract Transform Load**

- **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, such as relational or multidimensional databases, Web services, etc.



Data Integration and the Extraction, Transformation, and Load Process



Data Warehouse Development

Data warehouse development approaches

- **Inmon Model:** EDW approach (top-down)
- **Kimball Model:** Data mart approach (bottom-up)
- Which model is best?
- **Table 3.3** provides a comparative analysis between EDW and Data Mart approach
- **One alternative is the hosted warehouse**



Additional DW Considerations Hosted Data Warehouses

■ Benefits:

- Requires minimal investment in infrastructure
- Frees up capacity on in-house systems
- Frees up cash flow
- Makes powerful solutions affordable
- Enables solutions that provide for growth
- Offers better quality equipment and software
- Provides faster connections



Multidimensionality

The ability to organize, present, and analyze data by several dimensions, such as sales by region, by product, by salesperson, and by time (four dimensions)

■ Multidimensional presentation

- **Dimensions:** products, salespeople, market segments, business units, geographical locations, distribution channels, country, or industry
- **Measures:** money, sales volume, head count, inventory profit, actual versus forecast
- **Time:** daily, weekly, monthly, quarterly, or yearly



Analysis of Data in DW

- OLTP vs. OLAP...
- OLTP (online transaction processing)
 - Capturing and storing data from ERP, CRM, POS, ...
 - The main focus is on efficiency of routine tasks
- OLAP (Online analytical processing)
 - Converting data into information for decision support
 - Data cubes, drill-down / rollup, slice & dice, ...
 - Requesting ad hoc reports
 - Conducting statistical and other analyses
 - Developing multimedia-based applications



OLAP vs. OLTP

TABLE 3.5 A Comparison Between OLTP and OLAP

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)



OLAP

Slicing
Operations on a
Simple Tree-
Dimensional
Data Cube



Failure Factors in DW Projects

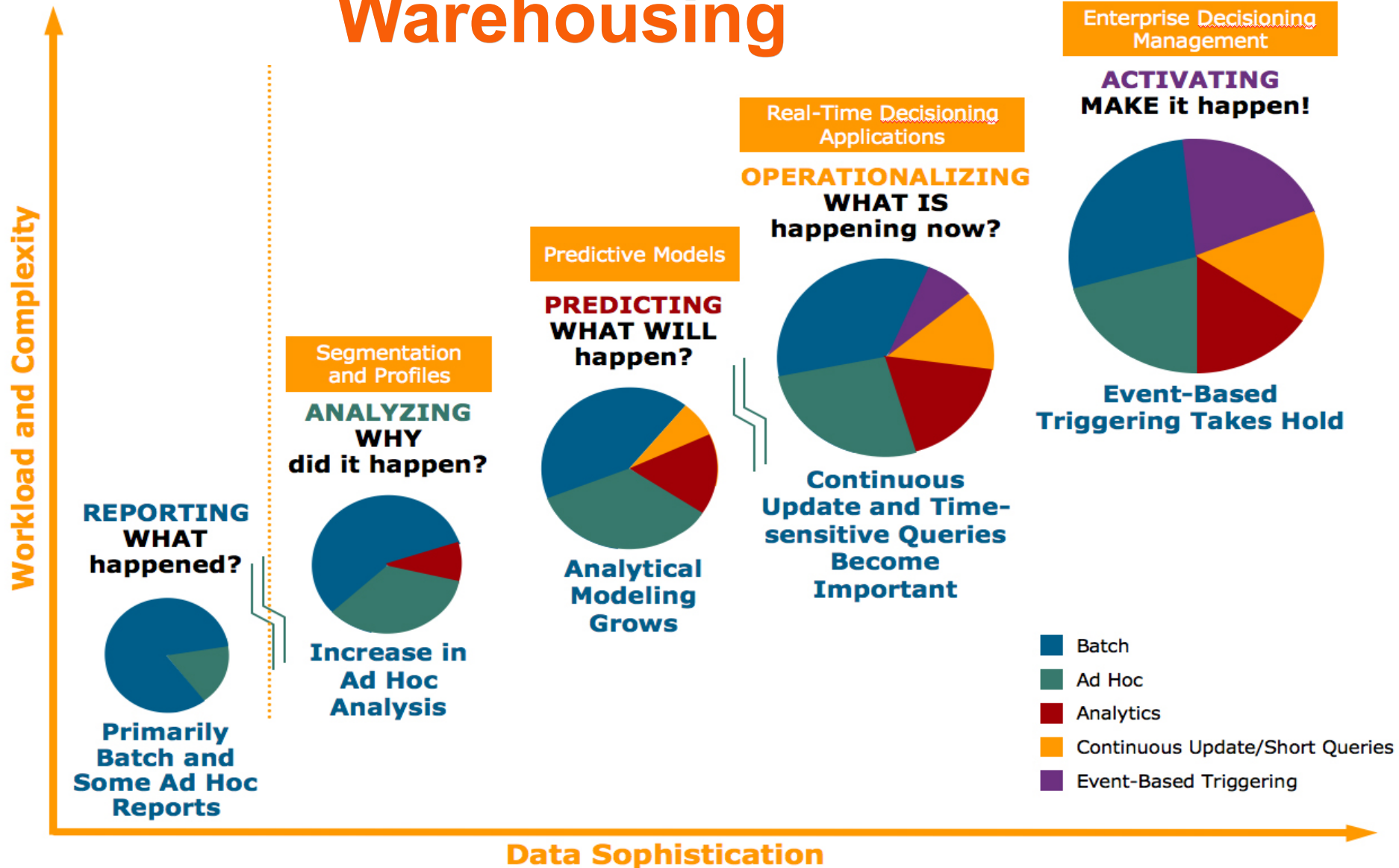
- Lack of executive sponsorship
- Unclear business objectives
- Cultural issues being ignored
 - Change management
- Unrealistic expectations
- Inappropriate architecture
- Low data quality / missing information
- Loading data just because it is available



Real-Time/Active DW/BI

- **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 and Data Warehousing



Real-Time/Active DW at Teradata

Active Access

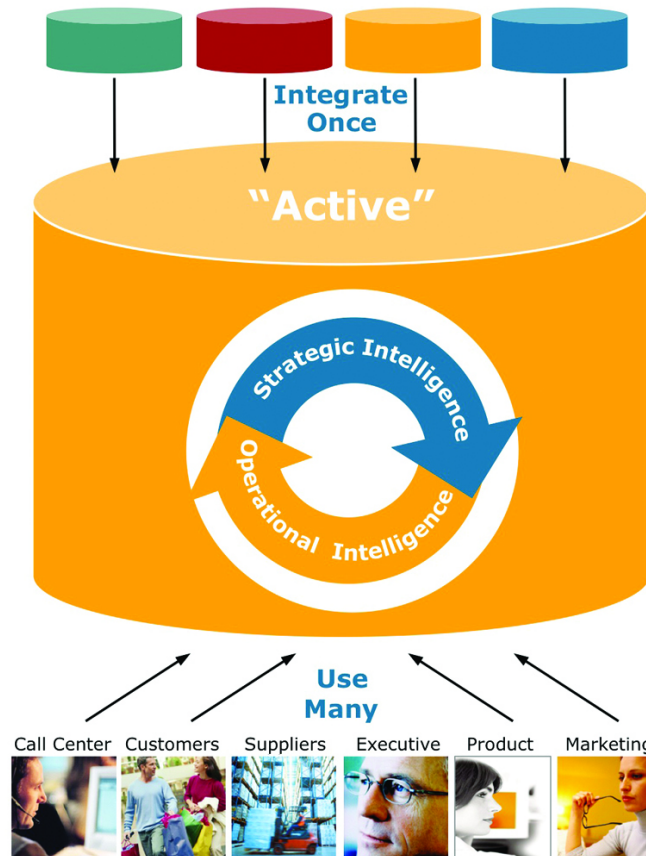
Front-Line operational decisions or services supported by near-real-time (NRT) access; Service Level Agreements of 5 seconds or less

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)