LECTURE 1: INTRODUCTION TO DATA WAREHOUSE

Dr. Alshimaa Hamdy

BUSINESS INTELLIGENT

- ✓Refers to a set of tools and techniques that enable a company to transform its business data into timely and accurate information for the decisional process, to be made available to the right persons in the most suitable form.

BI IS DIFFERENT FROM ARTIFICIAL INTELLIGENCE (AI)

- ✓AI systems make decisions for the users.
- ✓BI systems help the users make the right decisions, based on available data.

Combination of technologies

- ✓Data Warehousing (DW)
- ✓On-Line Analytical Processing (OLAP)
- ✓Data Mining (DM)

DEFINITIONS

- **Data Warehouse**
  - A subject-oriented, integrated, time-variant, non-updatable collection of data used in support of management decision-making processes
    - ***Subject-oriented:*** e.g. customers, patients, students, products
    - ***Integrated:*** consistent naming conventions, formats, encoding structures; from multiple data sources
    - ***Time-variant:*** can study trends and changes
    - ***Non-updatable:*** read-only, periodically refreshed
- **Data Mart**
  - A data warehouse that is limited in scope

HISTORY LEADING TO DATA WAREHOUSING

- Improvement in database technologies, especially relational DBMSs
- Advances in computer hardware, including mass storage and parallel architectures
- Emergence of end-user computing with powerful interfaces and tools
- Advances in middleware, enabling heterogeneous database connectivity
- Recognition of difference between operational and informational systems

APPLICATIONS THAT DATA WAREHOUSE SUPPORTS

- **OLAP** (Online Analytical Processing) is a term used to describe the analysis of complex data from the data warehouse.
- **DSS** (Decision Support Systems) also known as EIS (Executive Information Systems) supports organization's leading decision makers for making complex and important decisions.
- **Data Mining** is used for knowledge discovery, the process of searching data for unanticipated new knowledge.

NEED FOR DATA WAREHOUSING

- **Integrated, company-wide view of high-quality information (from disparate databases)**

- **for improved performance *operational*, *informational* systems and data.**

ISSUES WITH COMPANY-WIDE VIEW

- ✓ Inconsistent key structures
- ✓ Inconsistent data values
- ✓ Missing data

ORGANIZATIONAL TRENDS MOTIVATING DATA WAREHOUSES

- ✓ No single system of records
- ✓ Multiple systems not synchronized
- ✓ Organizational need to analyze activities in a balanced way
- ✓ Customer relationship management
- ✓ Supplier relationship management

SEPARATING OPERATIONAL AND INFORMATIONAL SYSTEMS

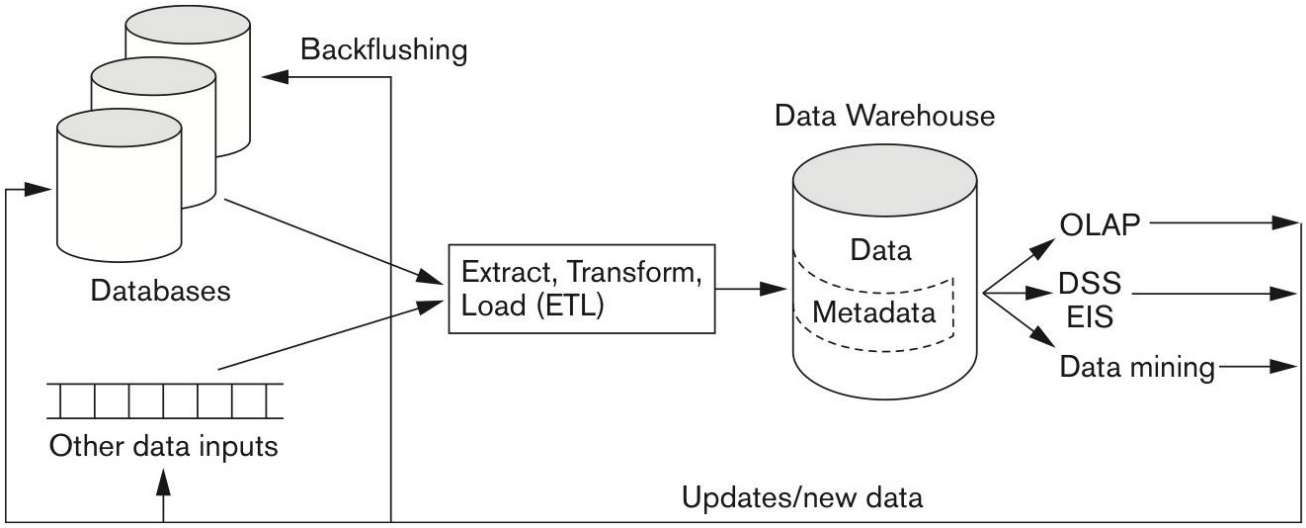
- Operational system – a system that is used to run a business in real time, based on current data; also called a system of record

- Informational system – a system designed to support decision making based on historical point-in-time and prediction data for complex queries or data-mining applications

CONCEPTUAL STRUCTURE OF DATA WAREHOUSE

Data Warehouse processing involves

- Cleaning and reformatting of data
- ETL (Extract, Transform, Load)
- OLAP – Data Analytics
- Data Mining

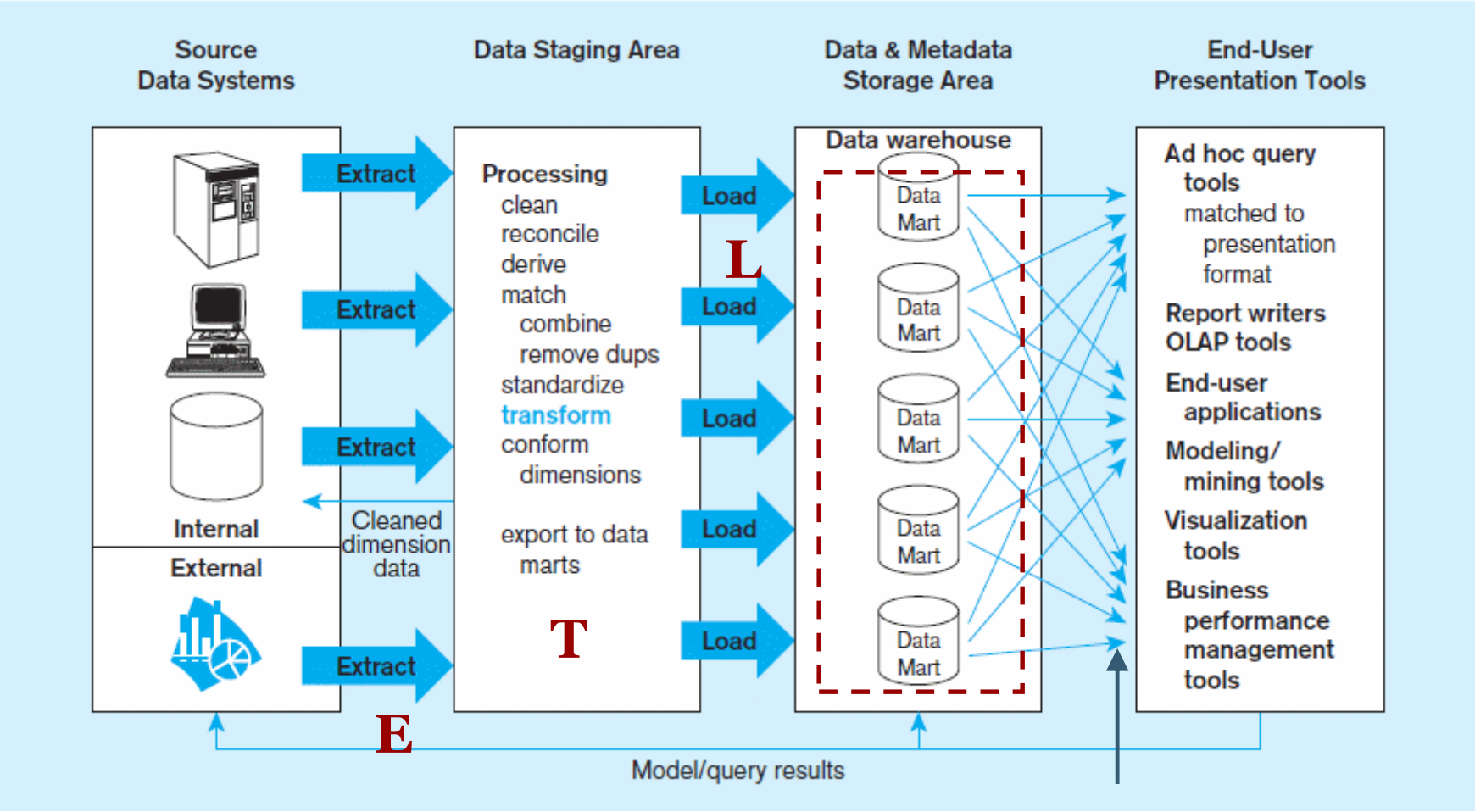


DATA WAREHOUSE ARCHITECTURES

- Independent Data Mart
- Dependent Data Mart and Operational Data Store
- Logical Data Mart and Real-Time Data Warehouse
- Three-Layer architecture

Independent data mart data warehousing architecture

Data marts:
Mini-warehouses, limited in scope

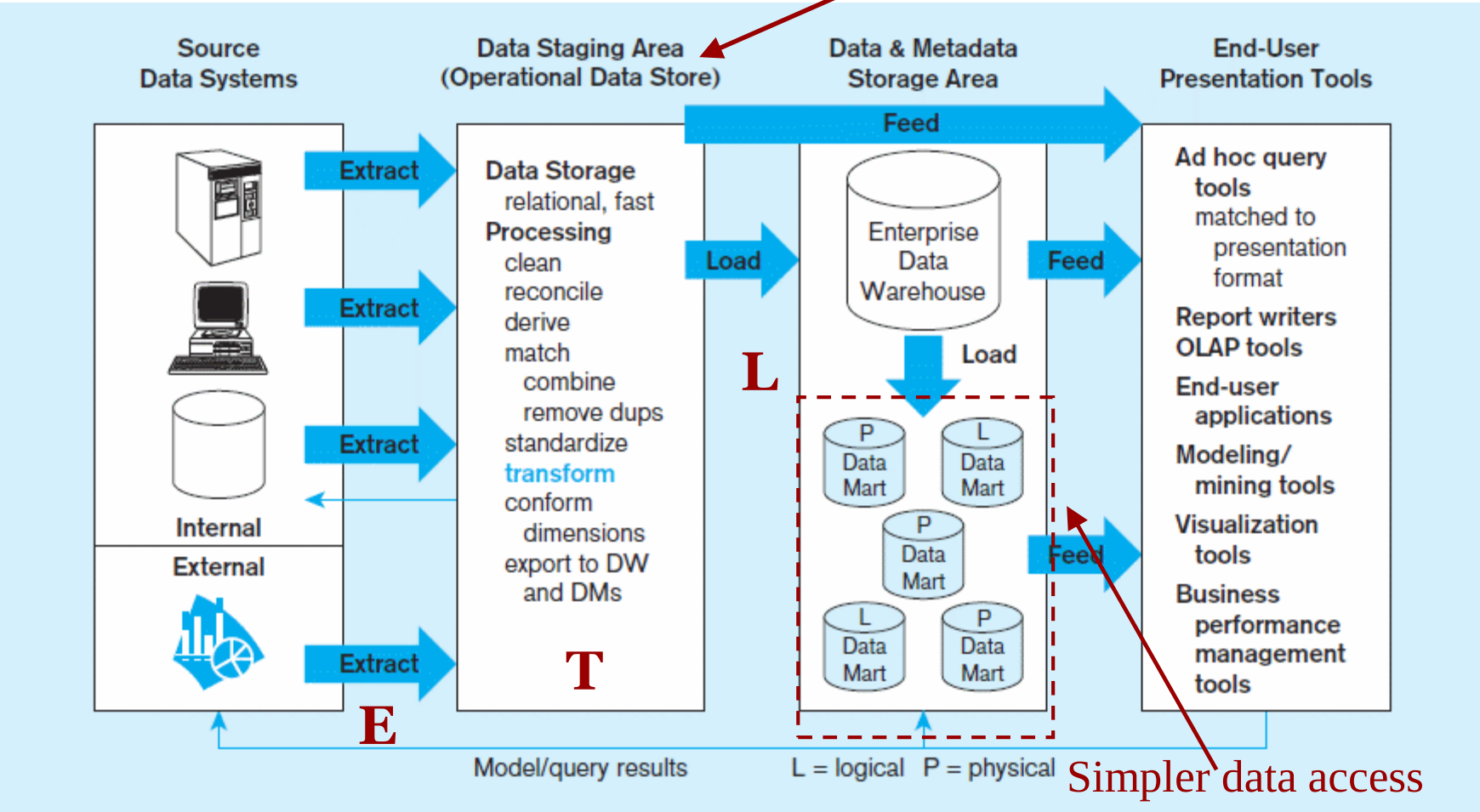


Separate ETL for each ***independent*** data mart

Data access complexity due to ***multiple*** data marts

Dependent data mart with operational data store: a three-level architecture

ODS provides option for obtaining ***current*** data

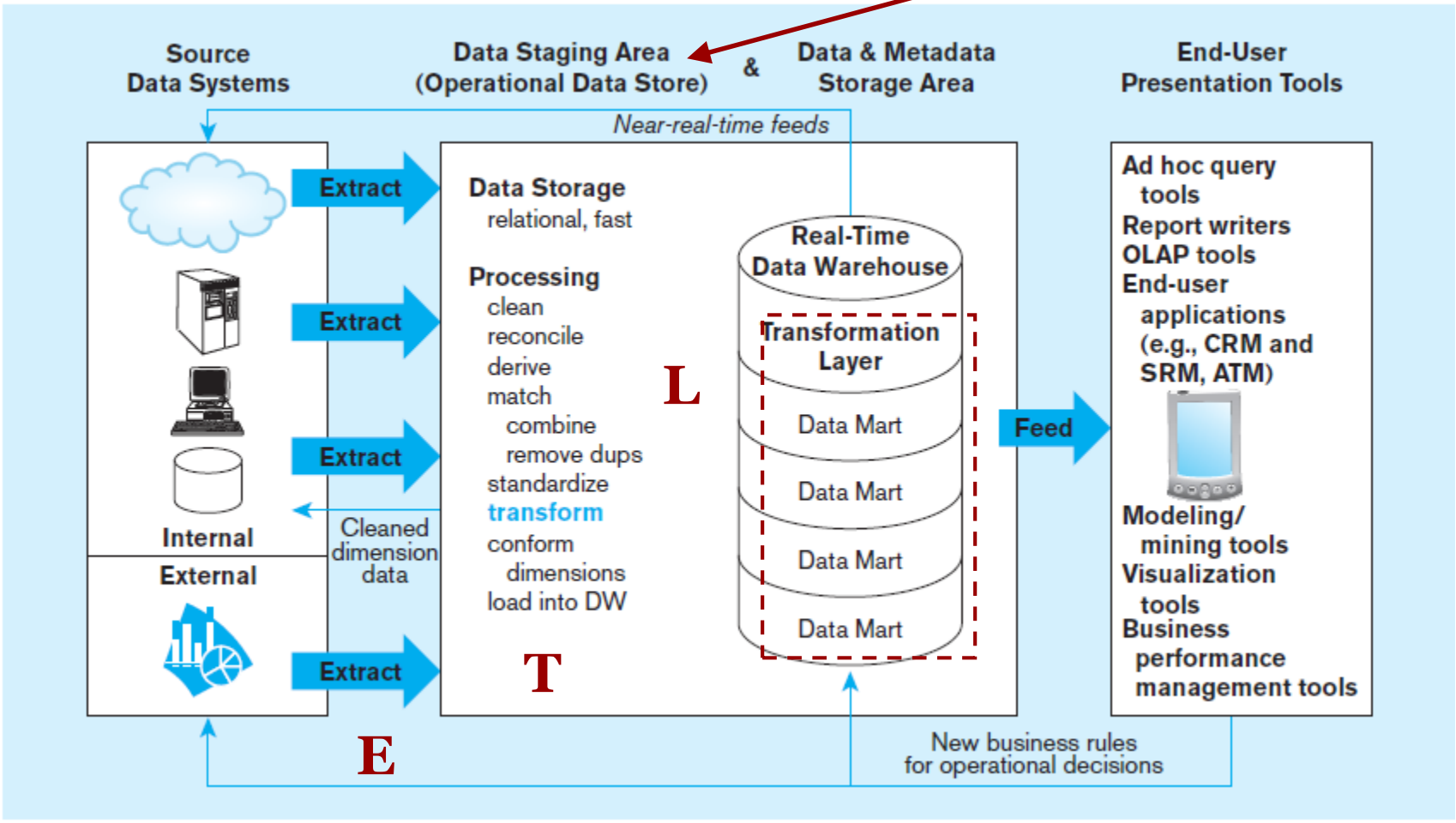


Single ETL for
Enterprise Data Warehouse (EDW)

Dependent data marts loaded from EDW

Logical data mart and real time warehouse architecture

ODS and **data warehouse** are one and the same

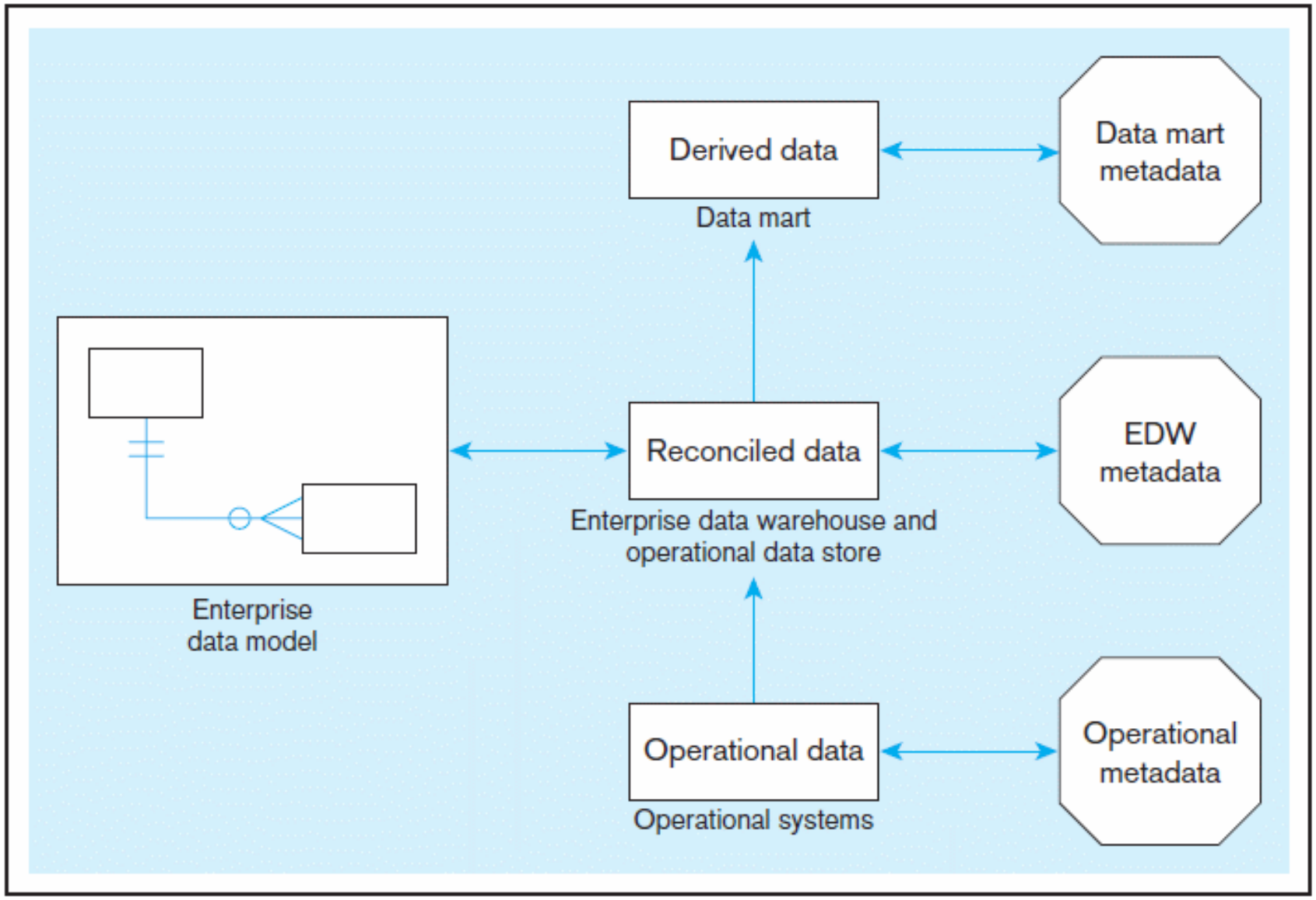


Near real-time ETL for ***Data Warehouse***

Data marts are NOT separate databases, but logical ***views*** of the data warehouse

➔ Easier to create new data marts

Three-layer data architecture for a data warehouse




Thank
You