MSc GFIS

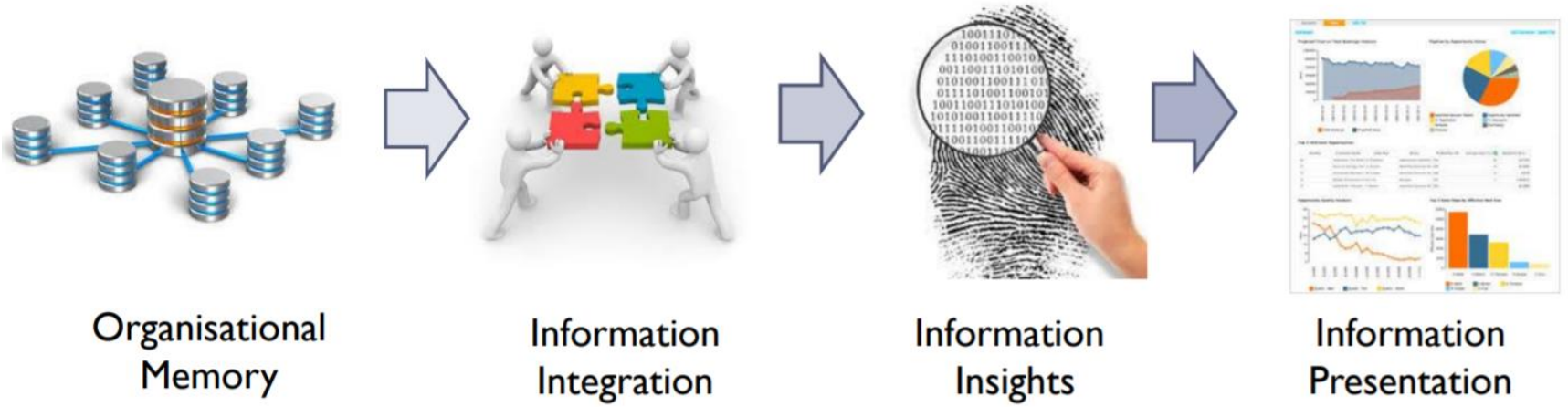
BI – Organisation Memory

Ruth Barry, rbarry@wit.ie

Learning Objectives:

- Understand the capabilities of Business Intelligence
- Explain Organisation memory
- Understand the basic definitions and characteristics of data warehousing
- Describe the processes used in developing and managing data warehouses
- Understand data warehousing architectures
- Explain data warehousing operations
- Explain the role of data warehouses in decision support
- Explain data integration and the extraction, transformation, and load (ETL) processes

Business Intelligence



BI Capabilities

- Organisational Memory: The ability to store information and knowledge.
- Information Integration: The ability to link structured and unstructured data from a variety of sources.
- Insight Creation: The ability to develop new insights and use them in the short-term or long-term to make better decisions.
- Presentation: The ability to use appropriate reporting and balanced scorecards tools, and thereby make BI more valuable to users.

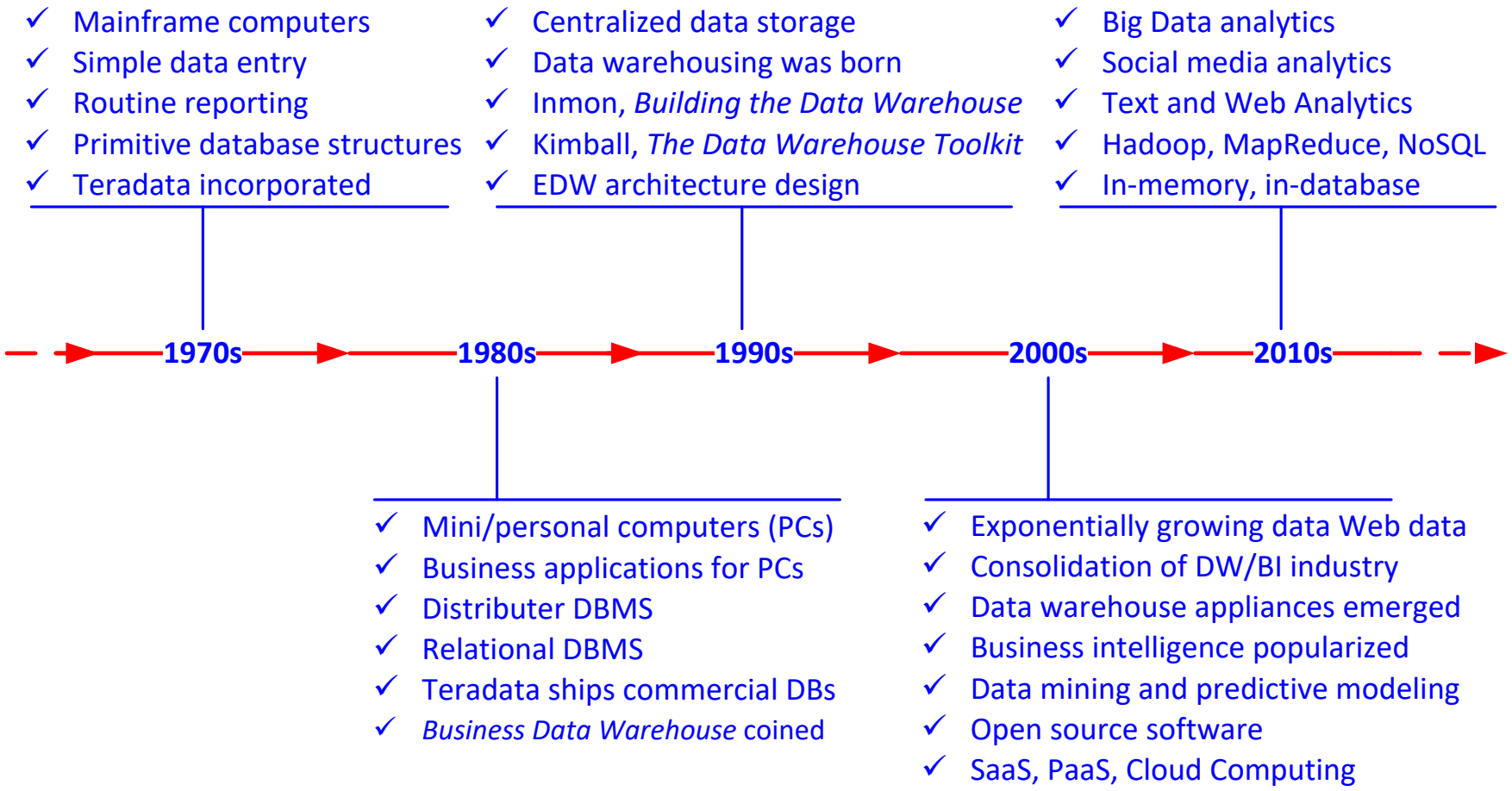
Organisational Memory

- Organizational Memory
  - Storage of information in such a form that it can be later accessed and used for BI
  - Also termed as corporate memory, or institutional memory
  - Key to current data skills required for business intelligence careers is a knowledge of data warehousing

Data Warehousing

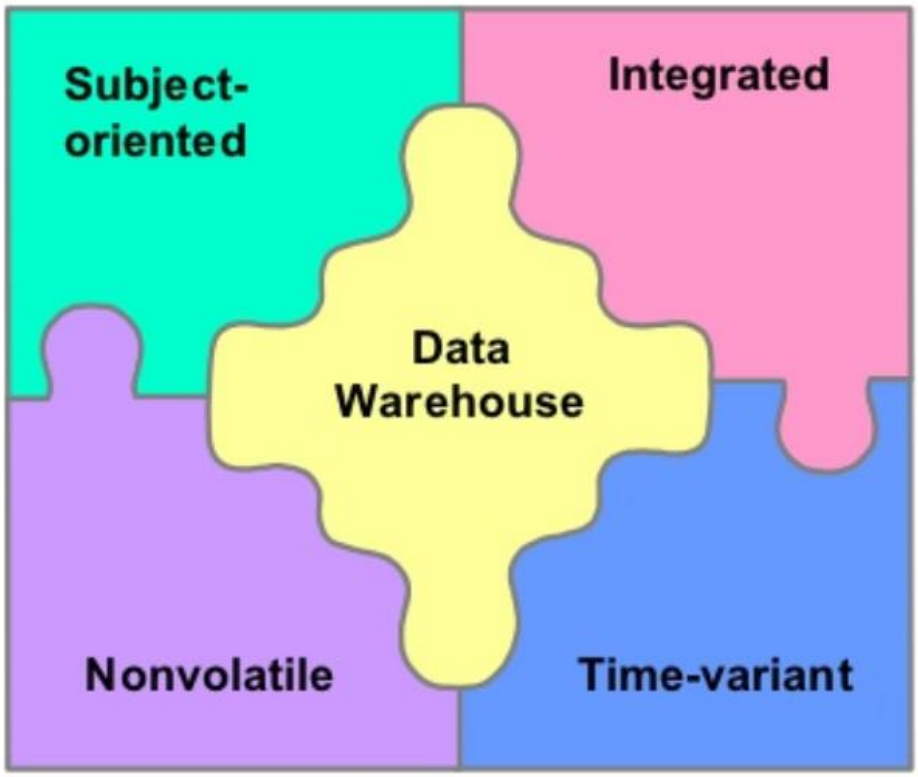
- What is a data warehouse?
  - A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
  - Holds a copy of transactional data, structured for querying and reporting.
  - "The data warehouse is a collection of integrated, 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



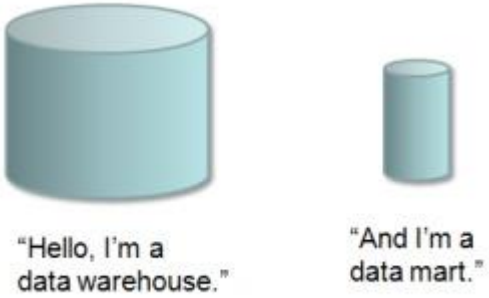
Characteristics of DW (Inmon, 2005)

- Subject oriented
- Integrated
- Nonvolatile
- Time-variant (time series)



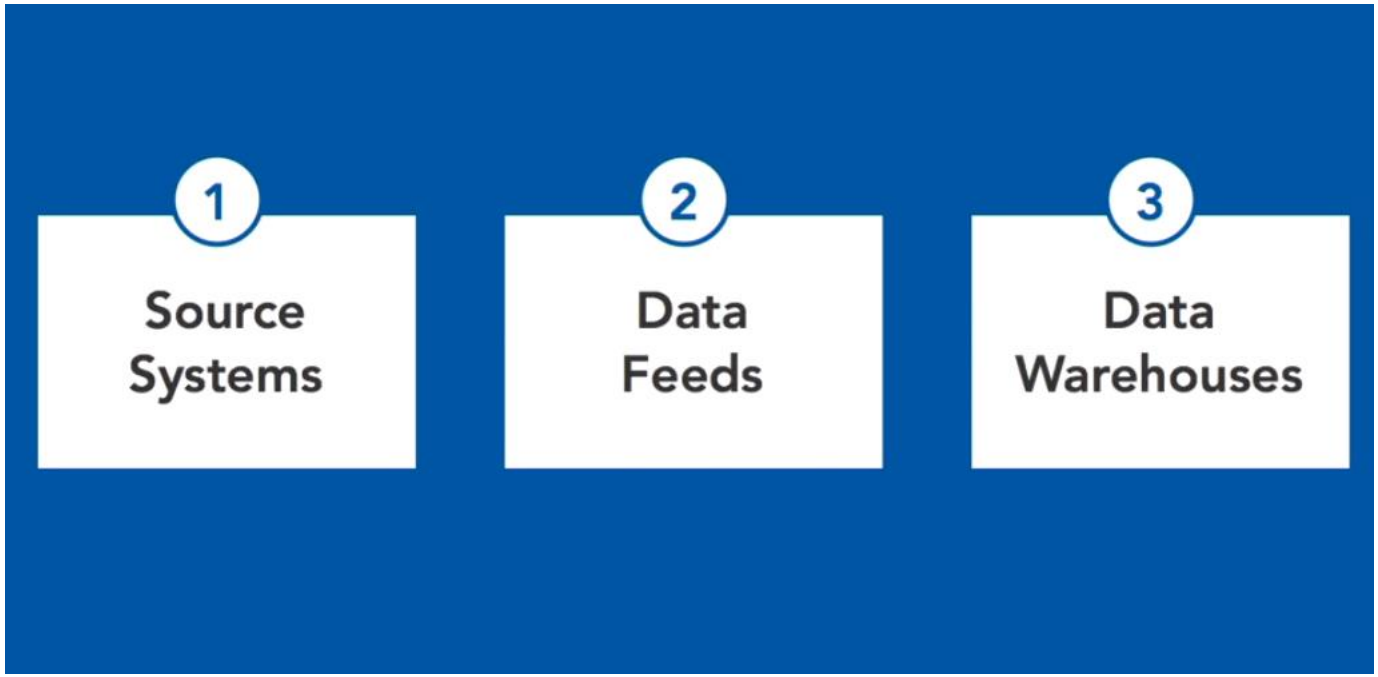
DW definitions

- 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

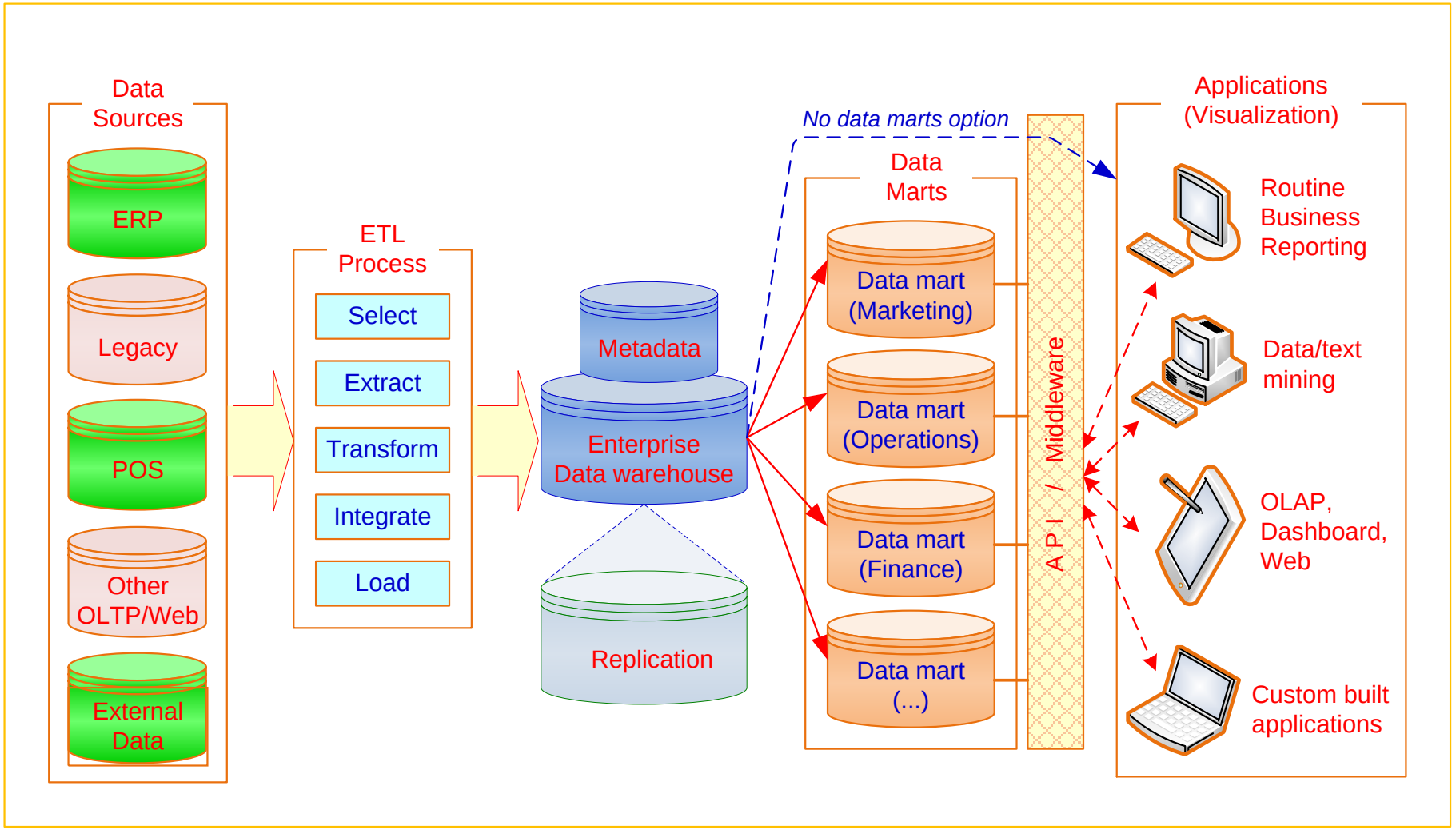


- Operational data stores (ODS): a type of database often used as an interim area for a data warehouse
- Enterprise data warehouse (EDW): a data warehouse for the enterprise
- Metadata: data about data. In a data warehouse, metadata describes the contents of a data warehouse and the manner of its acquisition and use.

Data Warehouse Architecture



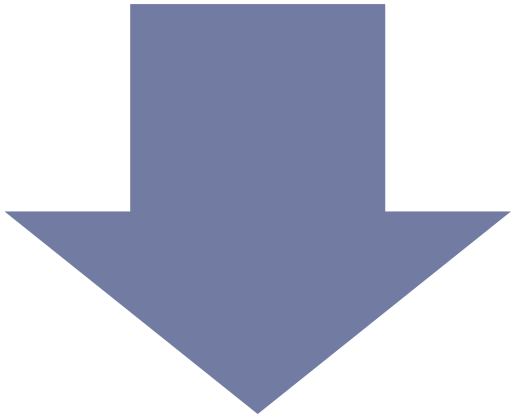
A Generic DW Framework



Data Warehouse Process:

- Data sources
- Data extraction and transformation
- Data loading
- Comprehensive database – EDW
- Metadata
- Middleware tools

Data Warehouse Architecture Choices

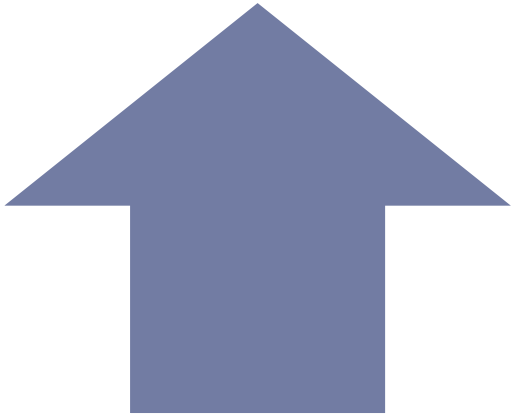


Top Down

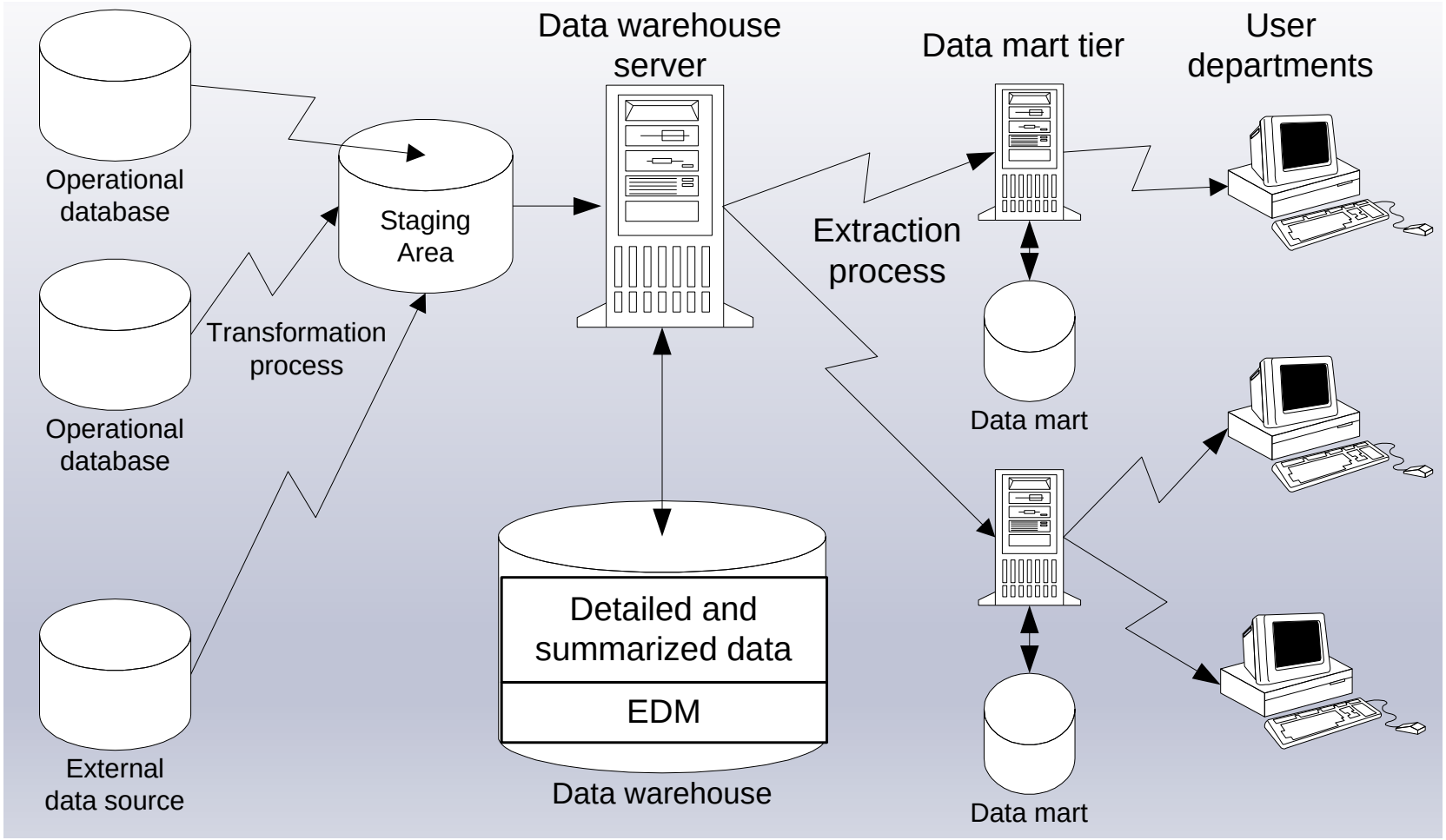
- Enterprise data warehouse
- Higher integration levels
- Centralized
- Larger project scope

Bottom Up

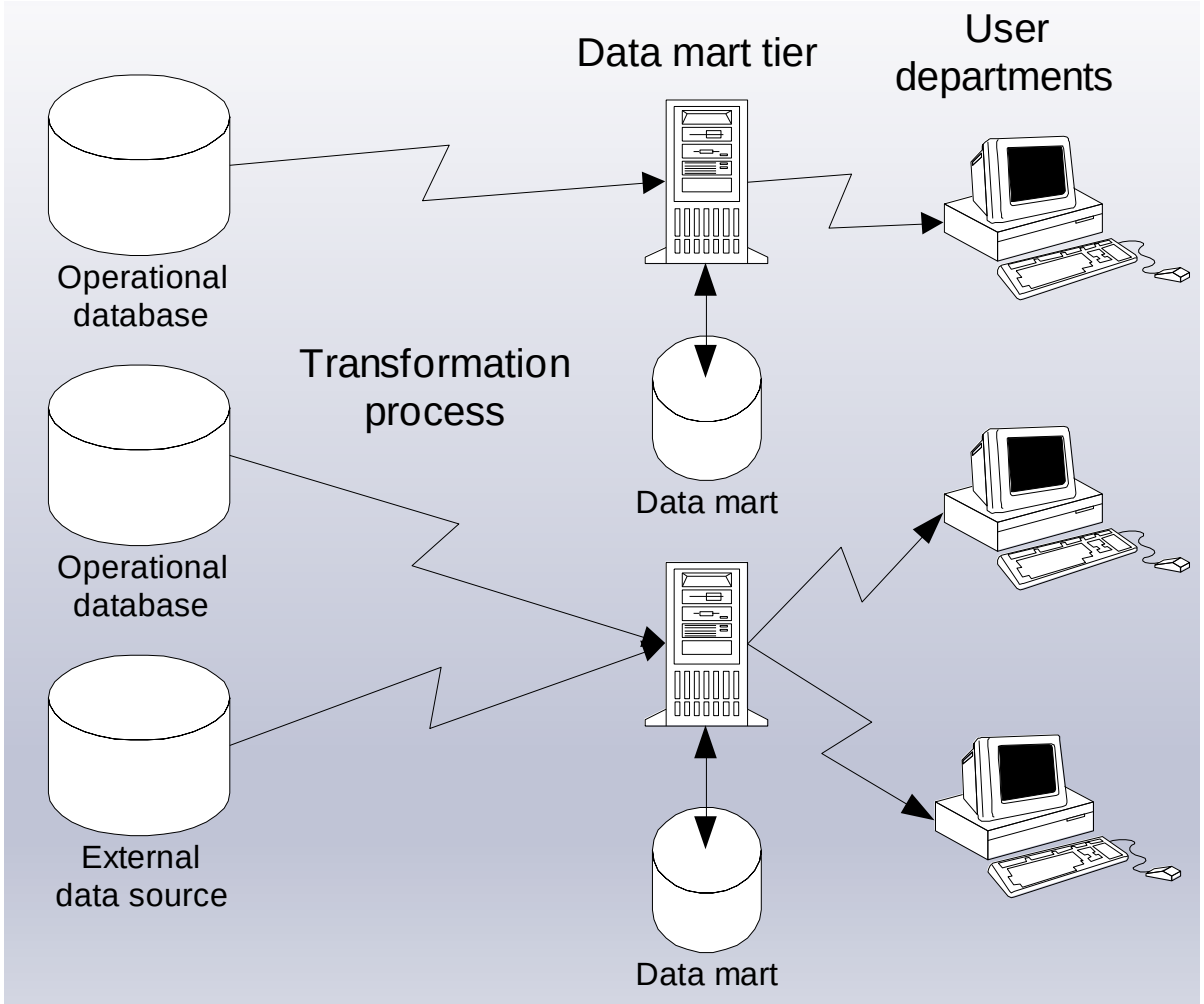
- Independent data marts
- Lower integration levels
- Decentralized
- Smaller project scope



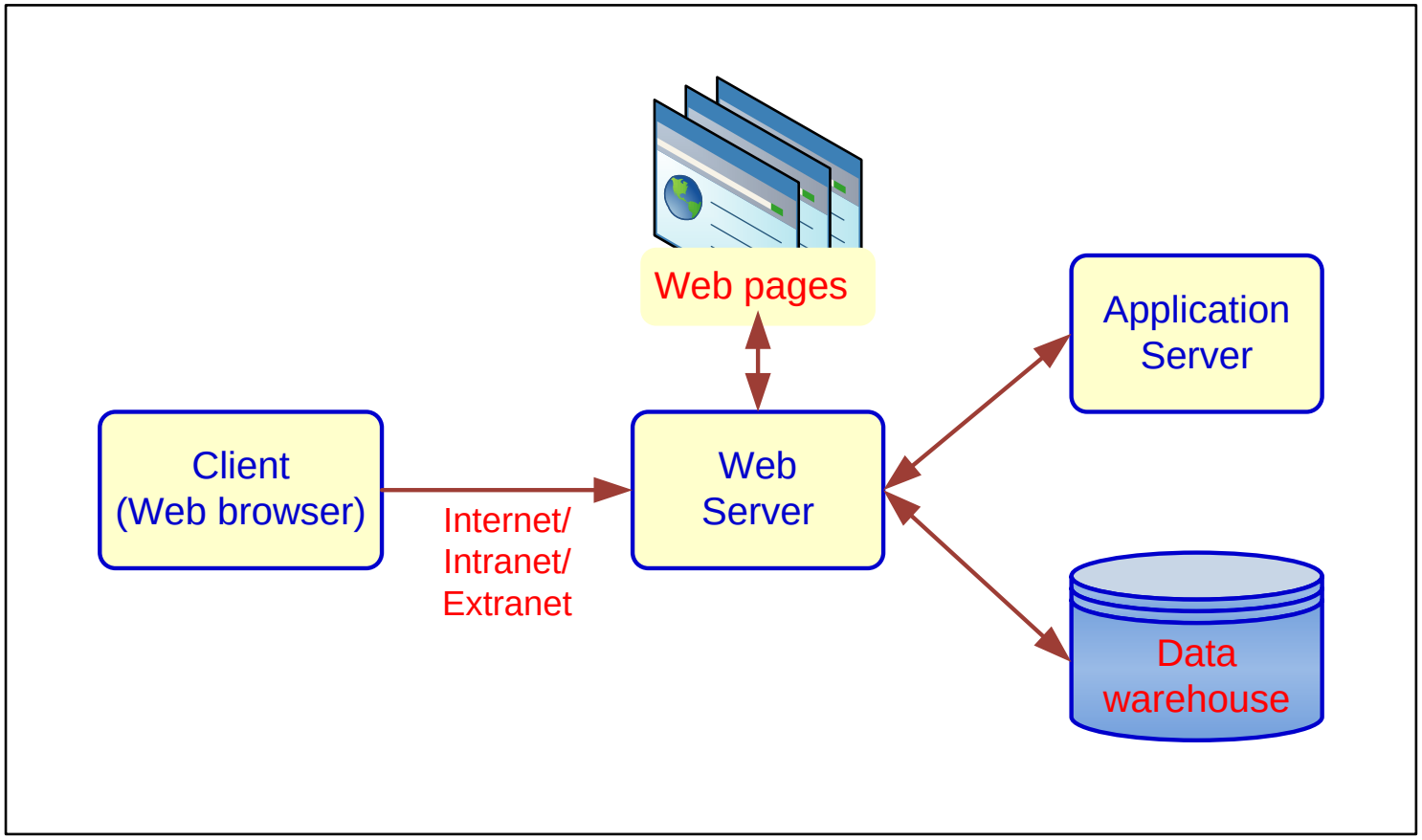
Top-Down Architecture



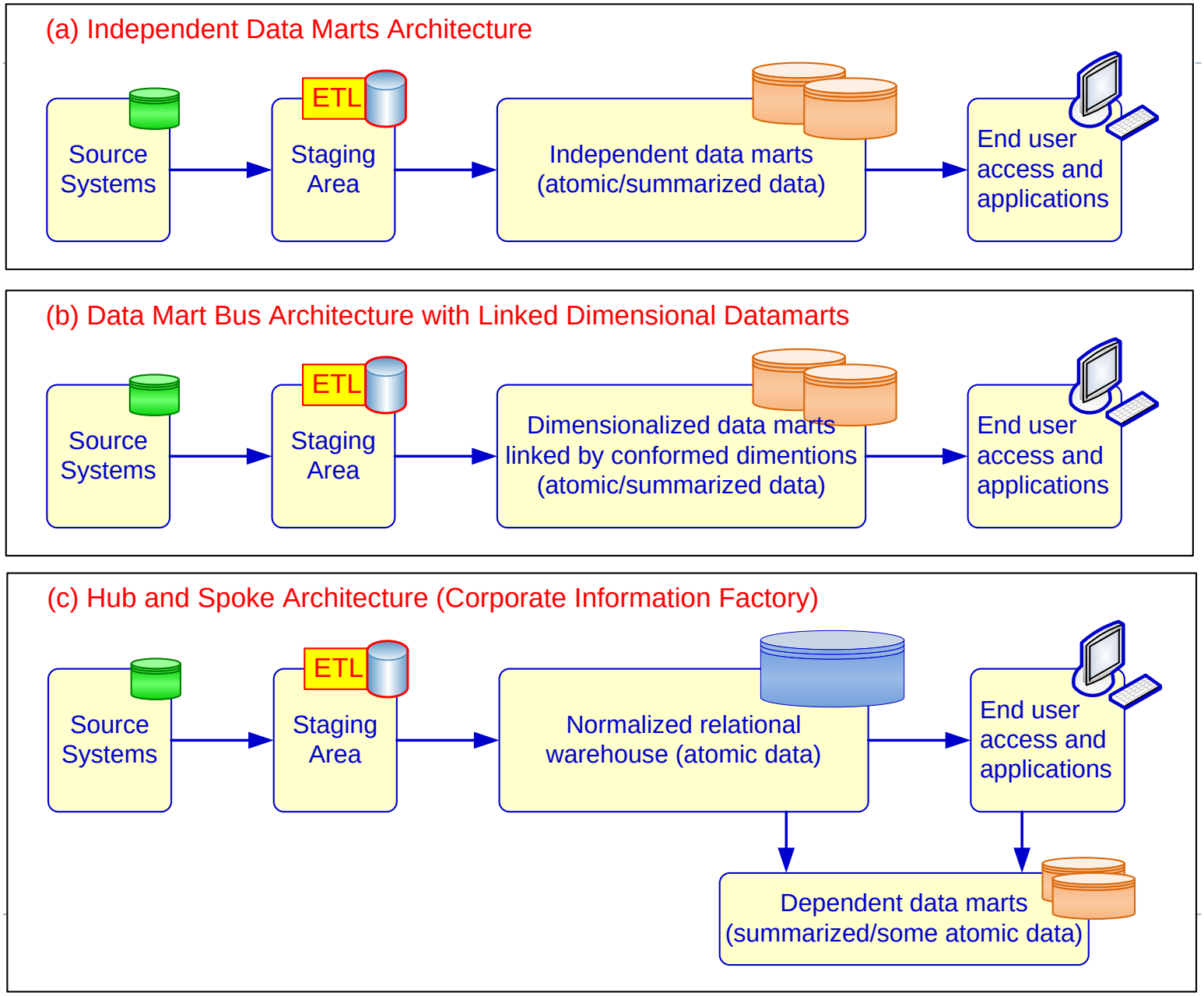
Bottom-up Architecture



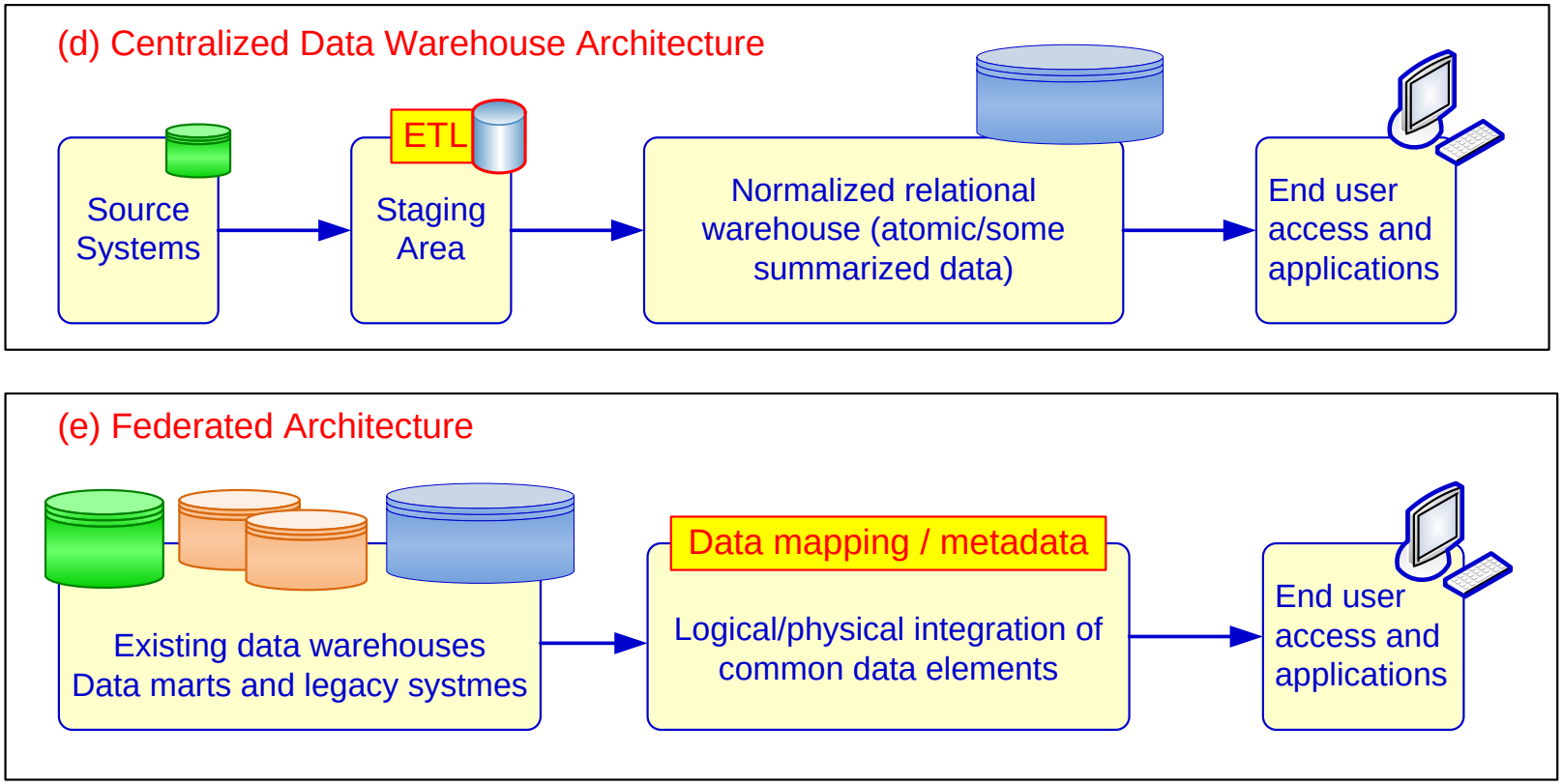
A Web-based DW Architecture



Alternative DW Architectures



DW Architectures



- Each architecture has advantages and disadvantages!
- Which architecture is the best?

DW Architectures

1. Independent Data Marts
2. Data Mart Bus Architecture
3. Hub-and-Spoke Architecture
4. Centralized Data Warehouse
5. Federated Data Warehouse

- Each has pros and cons!

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

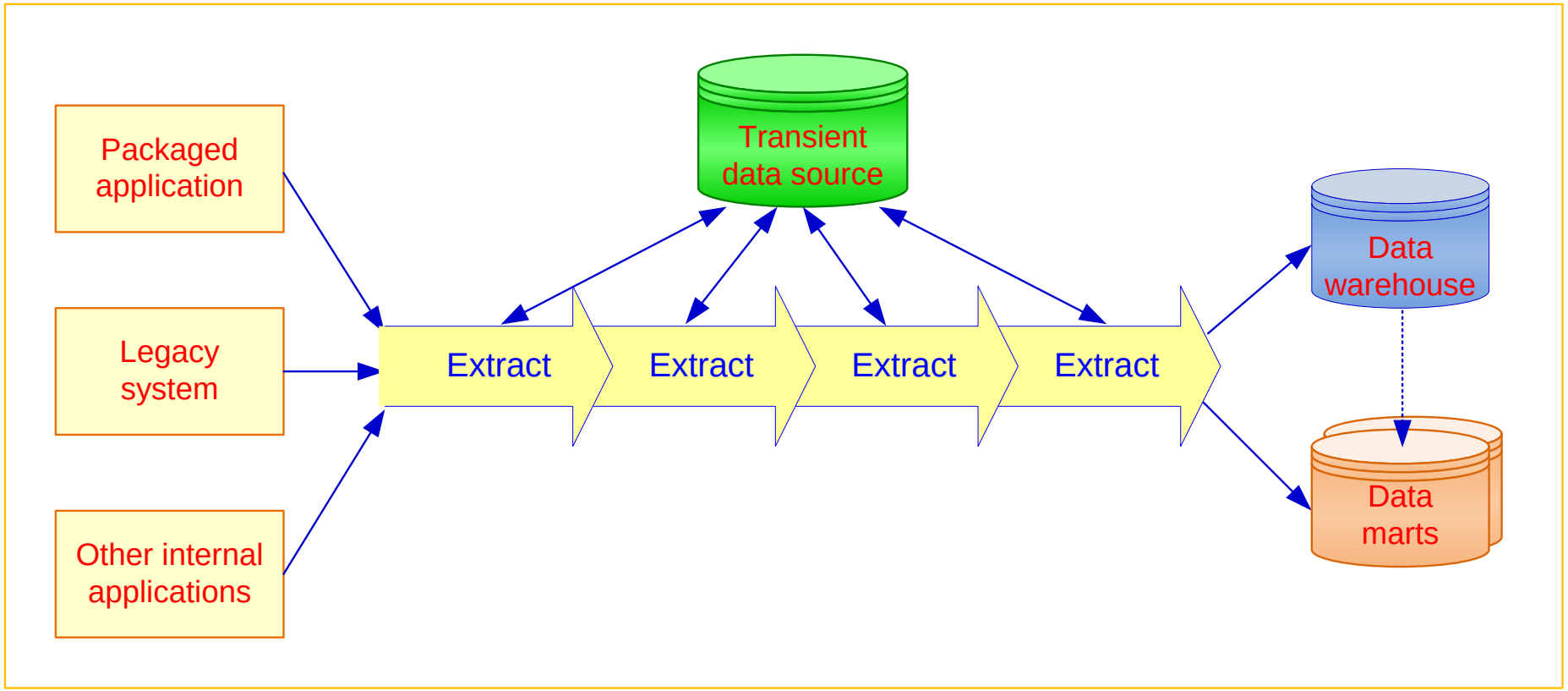
Data Feeds – Integration of Data in DW

- This process is known as Extraction Transformation Loading (ETL)
  - Help ensure that only clean data is fed into the data warehouse.
  - By tradition, its batch oriented
  - Different architecture required if real-time feeds
  - ETL is heavily driven by business rules.
  - Performance is difficult to manage (as DW expands)

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



ETL (Extract, Transform, Load)

- Issues affecting the purchase of an 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

Data Warehouses Challenges

- Significant coordination across organisational units
- Uncertain data quality in data sources
- Difficult to scale data warehouse
  - Enhancing a DW is time consuming
- They are built slowly
- DW and BI have been dominated by insights into what happened in the past.
- Data latency - Operational BI requires insights to what's happening currently