Data Warehouse Framework & Architecture

Data Warehousing Definitions

- **Operational data stores (ODS)**

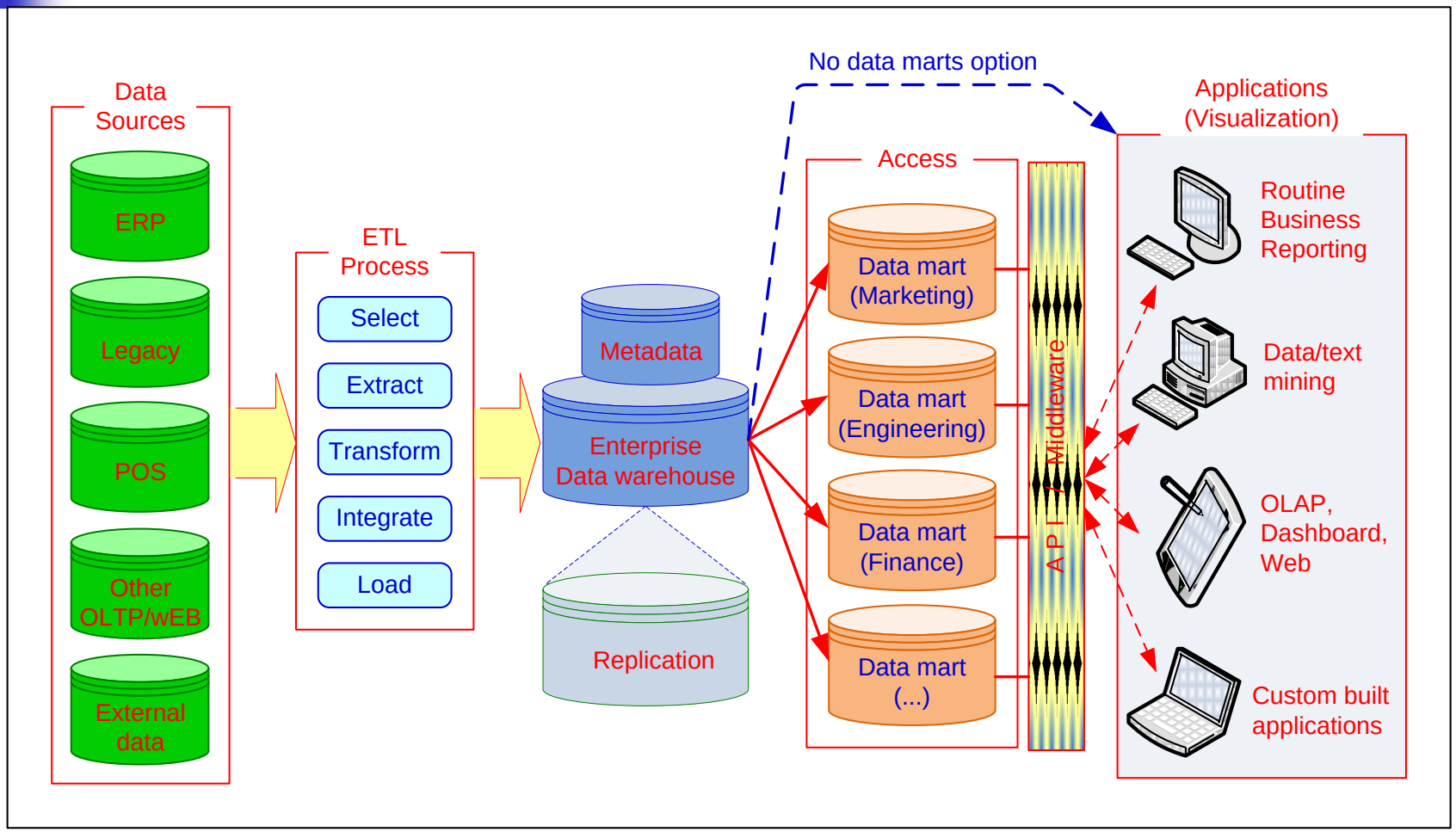
  A type of database often used as an interim area for a data warehouse
- **Oper marts**

  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

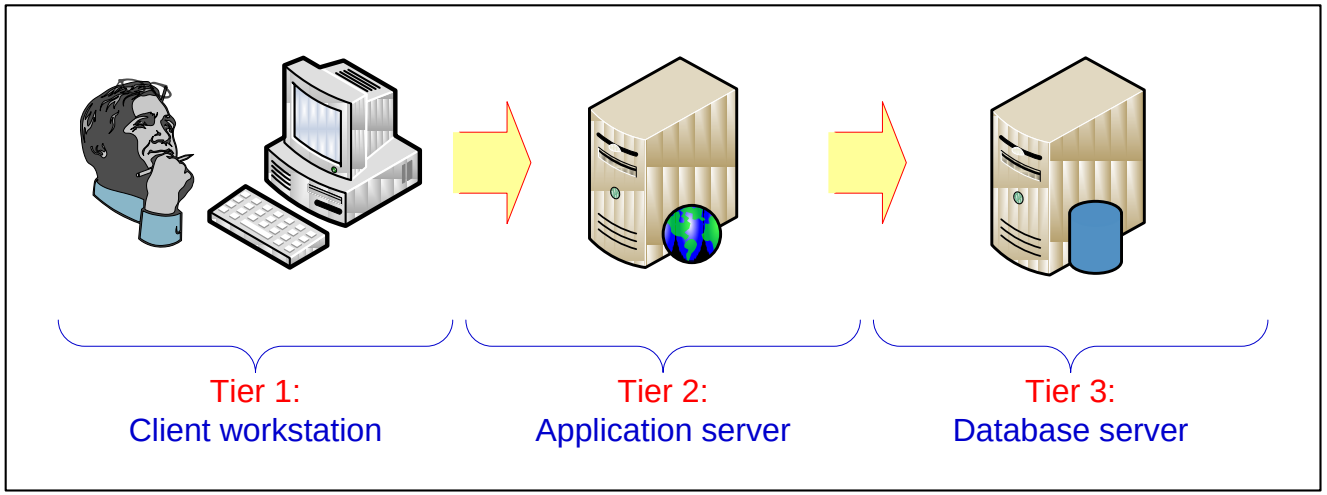
- **Three-tier architecture**
  1. Data acquisition software (back-end)
  2. The data warehouse that contains the data & software
  3. Client (front-end) software that allows users to access and analyze data from the warehouse
- **Two-tier architecture**

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

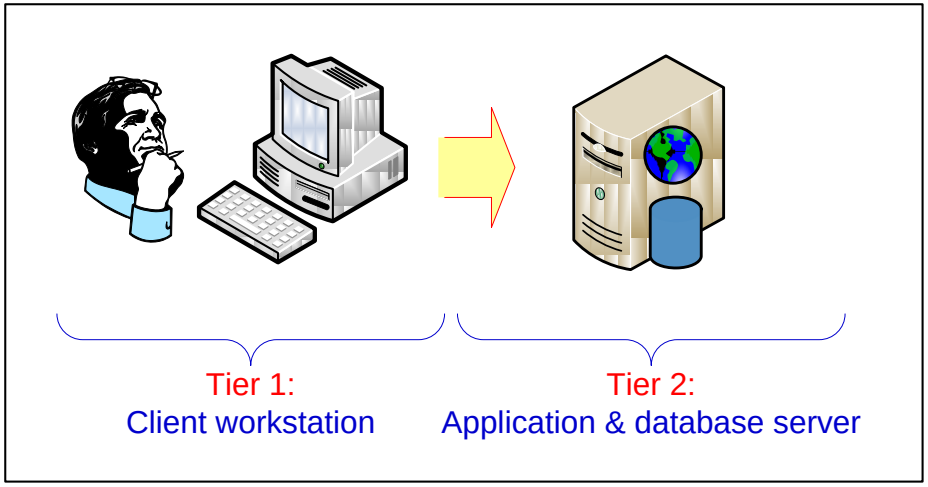
  … sometime there is only one tier?

Generic DW Architectures

3-tier architecture



2-tier architecture

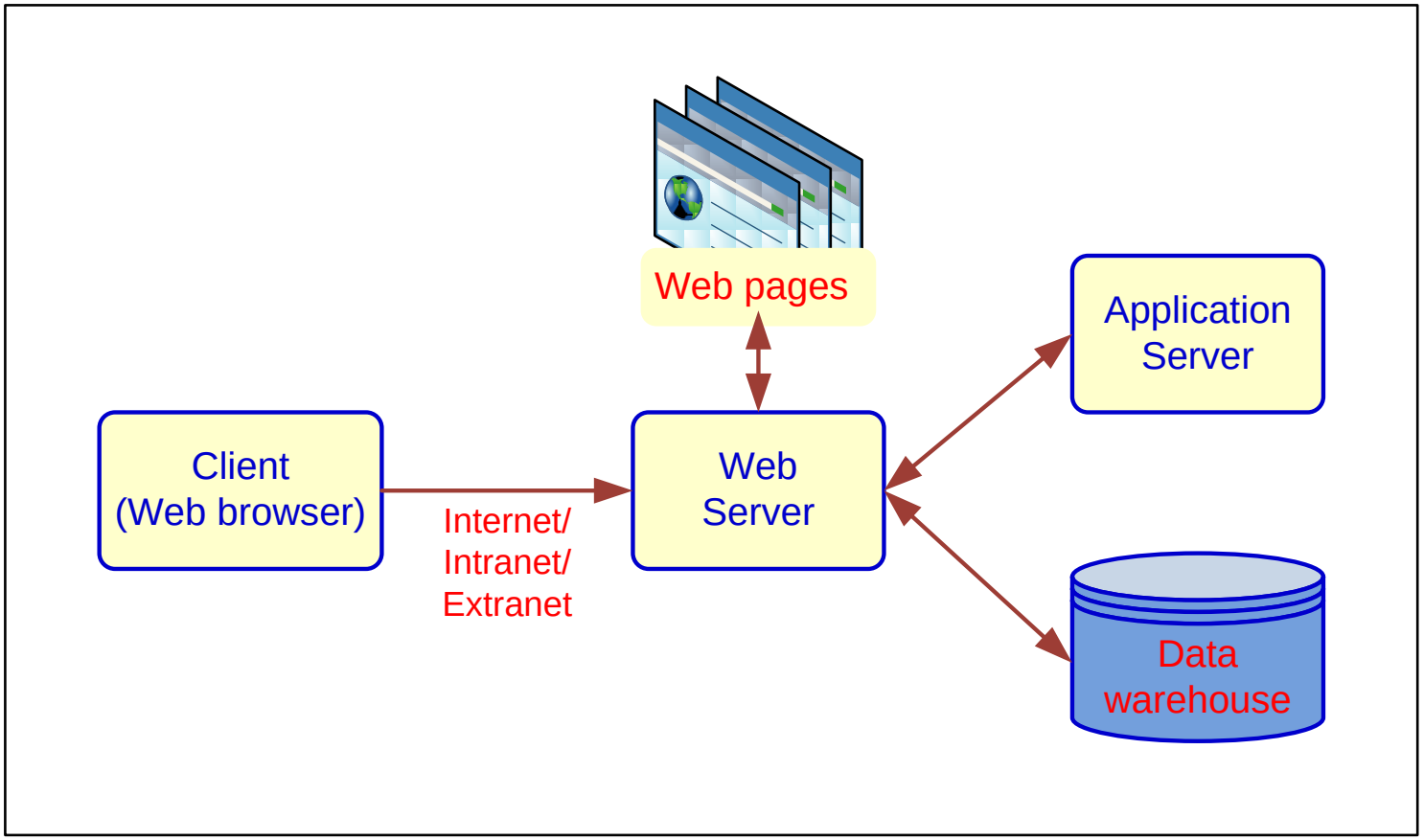


1-tier Architecture ?

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



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

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)**

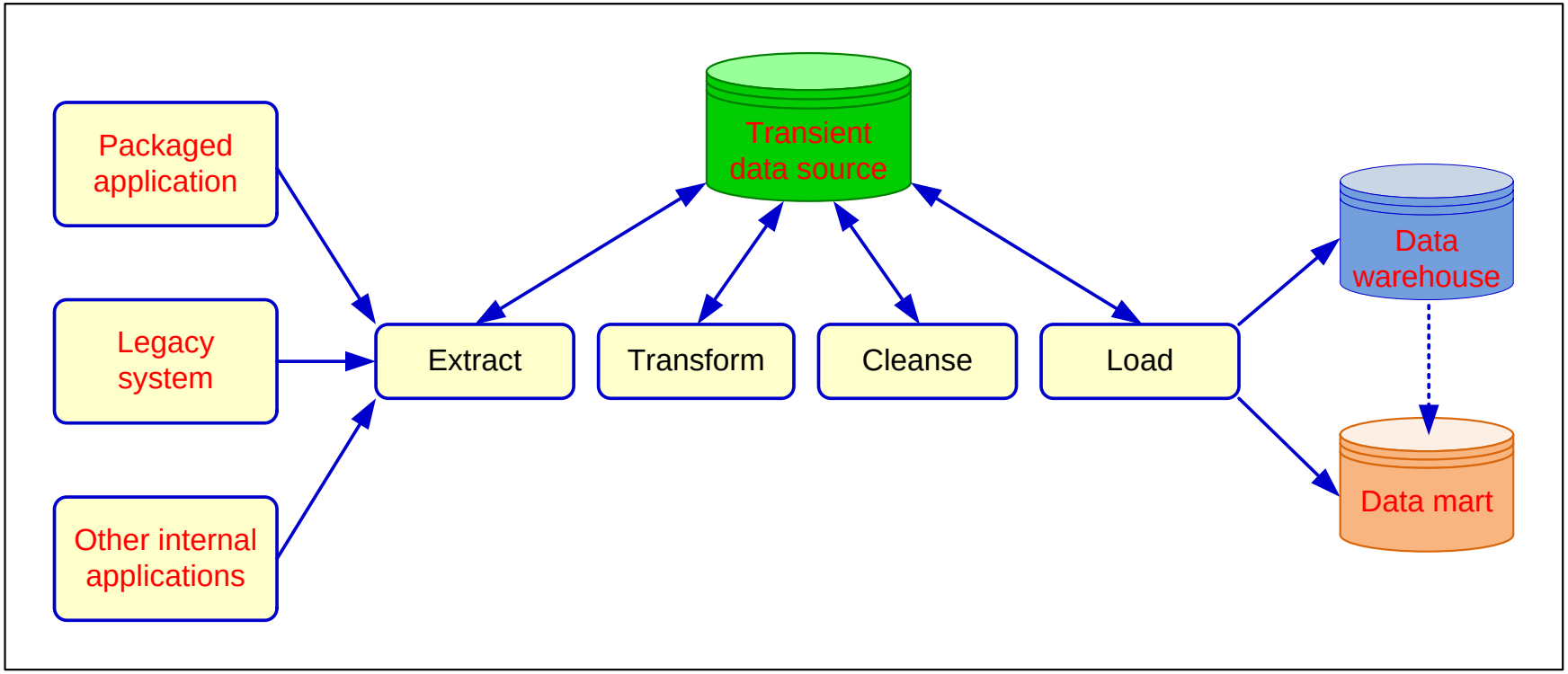
  A technology that provides a vehicle for pushing data from source systems into a data warehouse

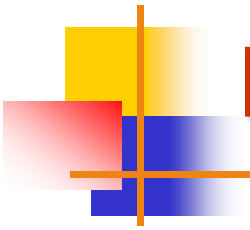
- **Enterprise information integration (EII)**

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

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

Extraction, transformation, and load (ETL) process





ETL

- 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