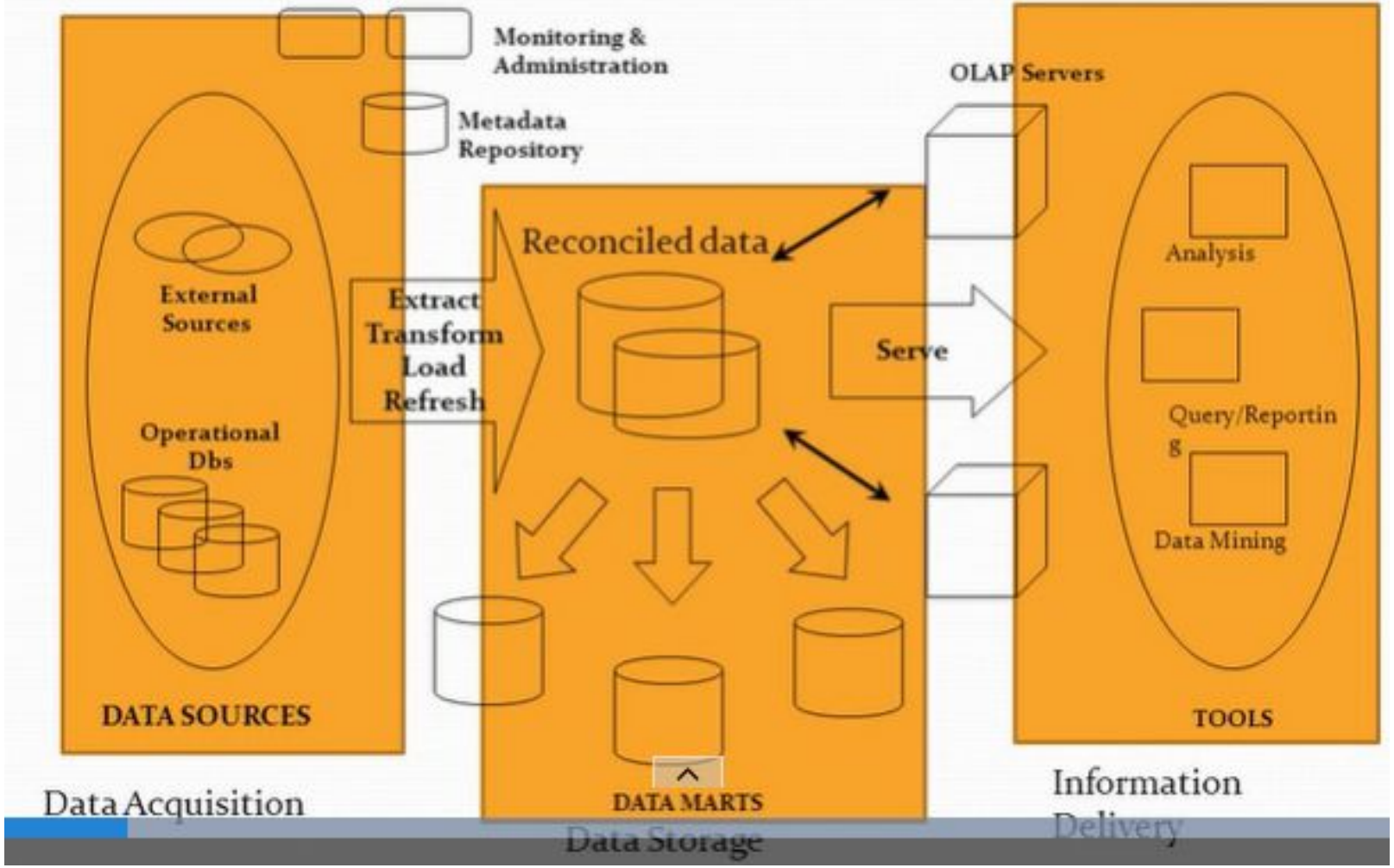
Data Warehouse Components

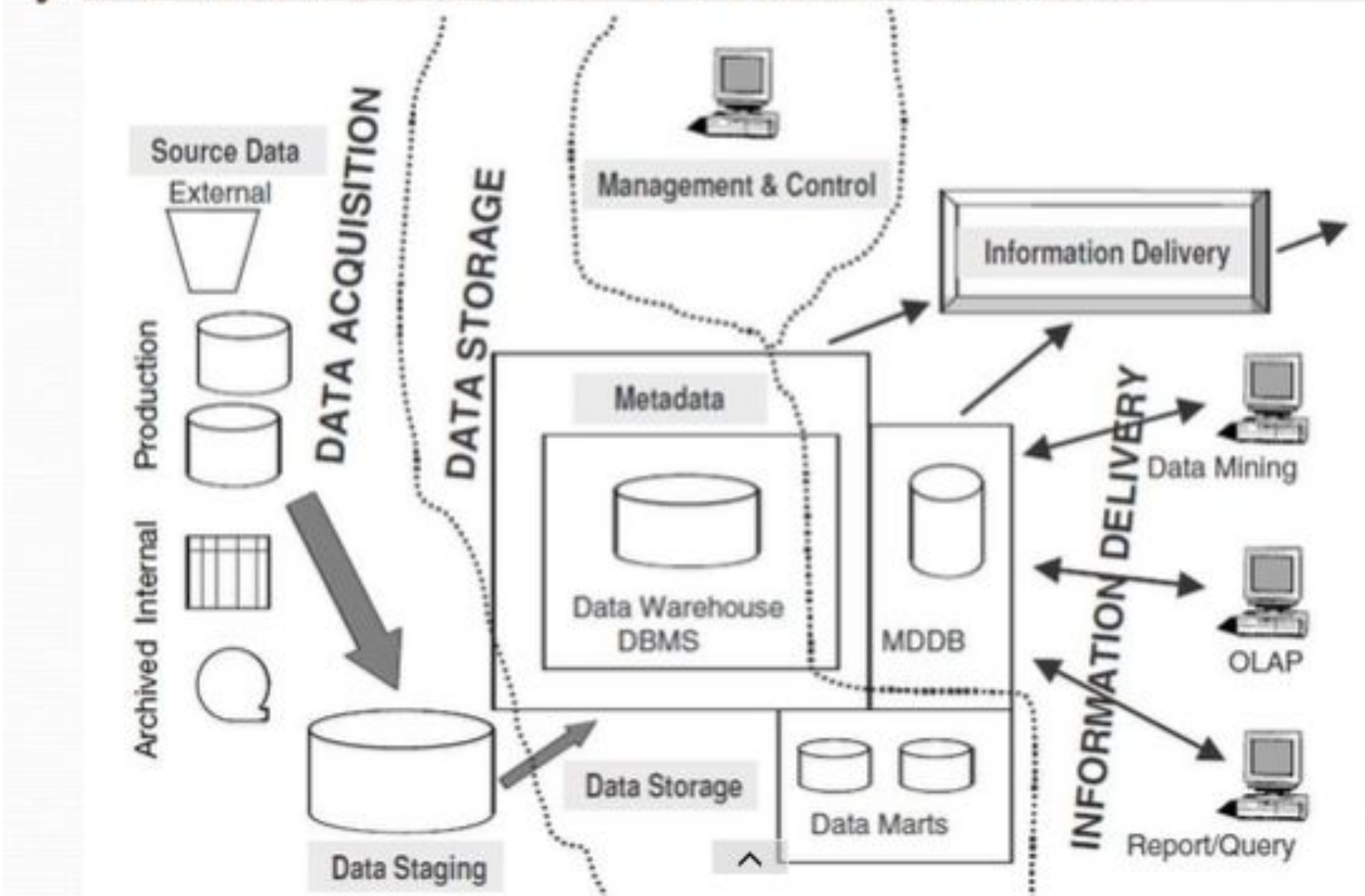
Overview of the Components

- Source Data Component
  - Production data
  - Internal data
  - Archive data
  - External data
- Data staging component
  - Extraction
  - Transformation
    - Cleaning
    - standardization
  - Loading
- Data storage component
- Information delivery component
- Metadata component
- Management and control component

Data Warehouse Architecture



Architectural Framework



Data Acquisition

You are the data analyst on the project team building a DW for an insurance company. List the possible data sources from which you will bring data into DW

- Production data: data from various operational systems
- External data: for finding trends and comparisons against other organizations.
- Internal data: private confidential data important to an organization
- Archived data: for getting some historical

Data Staging

- Performs ETL
  - Extraction
    - Select data sources, determine filters
    - Automatic replicate
    - Create intermediary files
  - Transformation
    - Clean, merge, de-duplicate data
    - Covert data types
    - Calculate derived data
    - Resolve synonyms and homonyms
  - Loading
    - Initial loading
    - Incremental loading

Why is a separate data staging area required?

- Data is across various operational databases
- It should be subject-oriented data
- Data staging is mandatory

Characteristics of data storage area

- Separate repository
- Data content
  - Read only
  - Integrated
  - High volumes
  - Grouped by business subjects
- Metadata driven
- Data from DW is aggregated in MDDBs

Information delivery component

- Depends on the user
  - Novice user: prefabricated reports, preset queries
  - Casual user: once in a while information
  - business analyst: complex analysis
  - Power users: picks up interesting data

Information delivery component

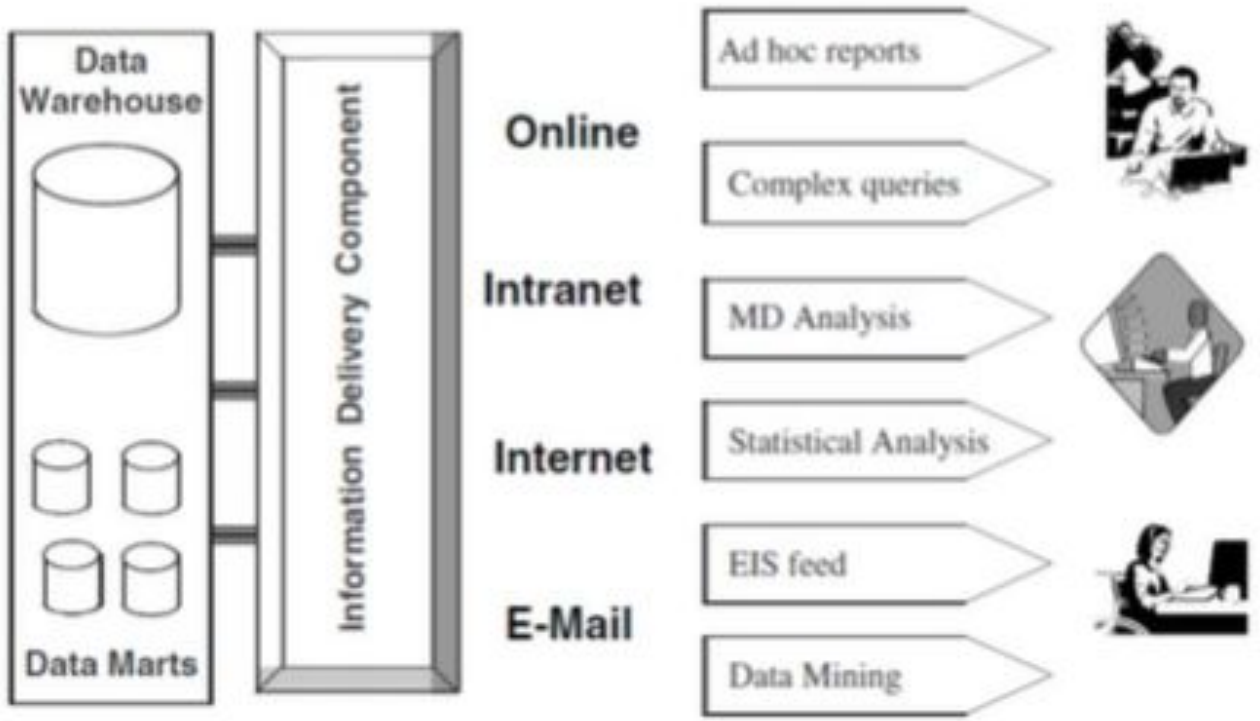


Figure 2-8 Information delivery component.

Metadata component

- Data about data in the datawarehouse
- Metadata can be of 3 types
  - Operational metadata: contains information about operational data sources
  - Extraction and transformation metadata: Details pertaining to extraction frequencies, extraction methods, business rules for data extraction
  - End-user metadata: navigational map of DW

Why is metadata especially important in a data warehouse?

- It acts as the glue that connects all parts of the data warehouse.
- It provides information about the contents and structures to the developers.
- It opens the door to the end-users and makes the contents recognizable in their own terms.

Management and Control

- Sits on top of all components
  - Coordinates the services and activities within the DW
  - Controls the data transformation and transfer in DW storage

Summing up

- Data warehouse building blocks or components are: source data, data staging, data storage, information delivery, metadata, and management and control.
- In a data warehouse, metadata is especially significant because it acts as the glue holding all the components together and serves as a roadmap for the end-users.