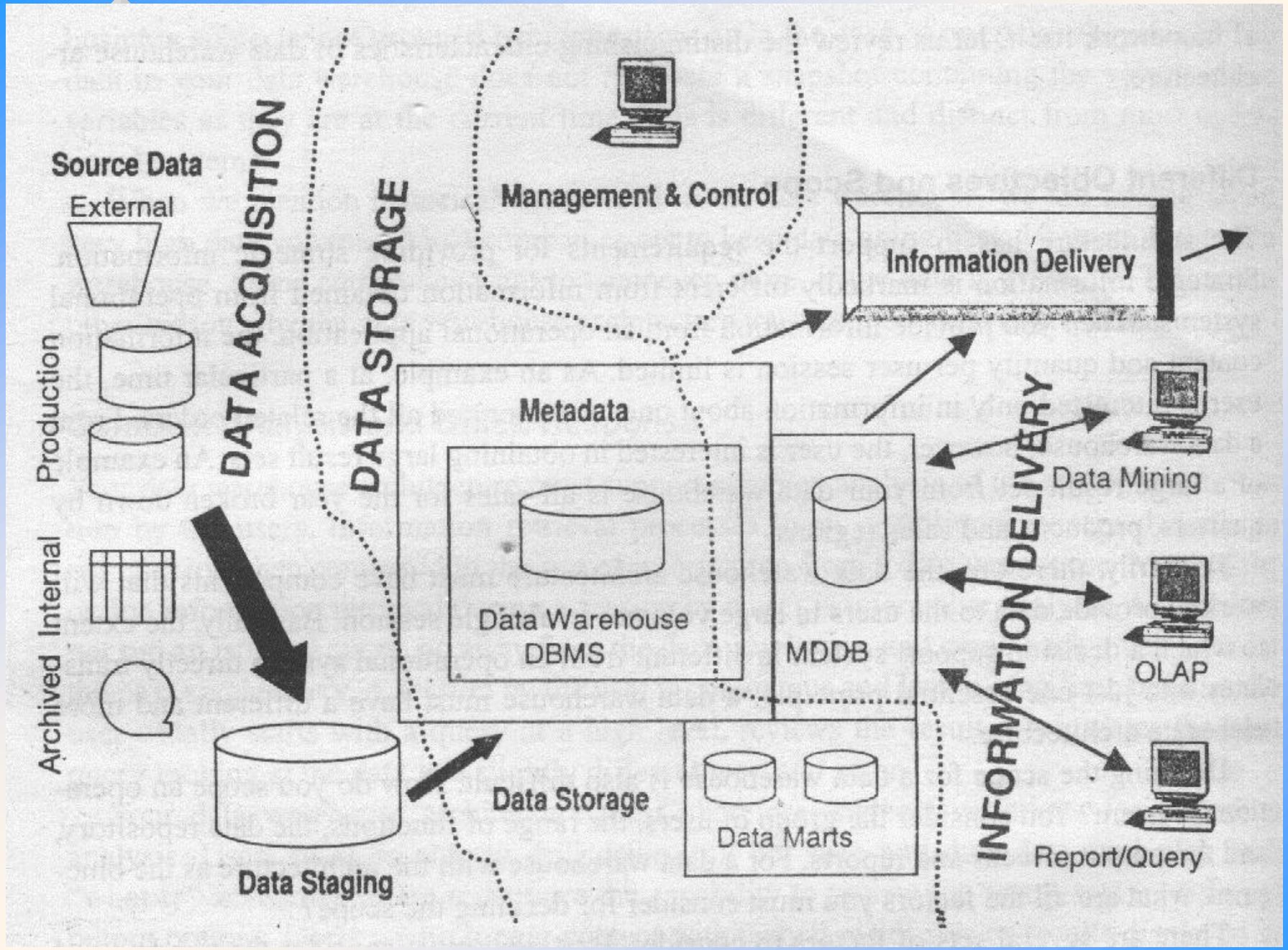
Data Warehouse Building Blocks

Objectives of Today's Lecture

- Overview of the Components
- Source Data Component
- Data Staging Component
- Data Storage Component
- Information Delivery Component
- Metadata Component
- Types of Metadata

Overview of the Data warehouse components

- Source data component
- Data staging component
- Data storage component
- Information delivery component


Source Data
External
Production
Internal
Archived
DATA ACQUISITION
DATA STORAGE
Management & Control
Information Delivery
Metadata
Data Warehouse
DBMS
MDDB
Data Mining
OLAP
Report/Query
INFORMATION DELIVERY
Data Storage
Data Marts
Data Staging

Source Data Component

- Production data
- Internal data
- Archived data
- External data

Production Data

- This category of data comes form different operational system of the enterprise.
- Different format of data
- Different platforms

Internal Data

- Data which is kept by the users “privately”
- Private spread sheets
- Documents
- Customer profiles

Archived Data

- In every system you, periodically take the old data and store it in archived files.
- Three staged archival methods are
  - Recent data is archived to a separate archival DB.
  - Older data is archived to flat files on disk storage.
  - Oldest data is archived to tape cartridges or microfilm and even kept off-site.
- Choose the range of data depending upon your data warehouse requirements

External Data

- Data form external sources
- Gives you a picture
  - What you are doing / done in the past.
  - Compare performance.
  - Derive strategies for future.

Data Staging Component

- Provides a place and an area to clean, change, combine, convert, deduplicate, and prepare data for storage and use in the data warehouse.
- Three major functions take place
  - Data extraction
  - Data transformation
  - Data loading

Commonly known as ETL

Data Extraction

- Retrieval of data from various sources like..
  - RDBMS
  - Flat files
  - Spread sheets
- After you extract data where do you keep it ?
  - Group of flat files
  - Data staging relational database
  - Or combination of both

Data Transformation

- Changing of data into clean, standardized, and summarized form.
- Conversion of data form
  - File oriented system to relational DB system.
  - Prior structure to suitable structure.
- Cleaning of data…
  - Correction of misspellings
  - Resolve conflicts like Zip codes and state codes
- Standardization
  - Standardize the data field length
  - Resolve the synonyms and homonyms

Data Transformation contd..

- When data transformation ends we have a collection of integrated data that is cleaned, standardized, and summarized. We now have the data read to load into each data set in your data warehouse.

Data Loading Component

- Sending data into data warehouse
- After completion of data transformation we perform data loading
- After completing the the design and construction of data warehouse we go for data loading.
- Moves large amount of data

Data Storage Component

- It is a separate repository (w.r.t. operational systems)
- Keep the data in the structure suitable for analysis but not for quick retrieval of individual pieces of information
- Warehouse is kept separate from the data storage for operational system
- data should represent snapshots at specified period
- In data storage component the data should not be in continual state of update

- This function is time-consuming
- Initial load moves very large volumes of data
- The business conditions determine the refresh cycles

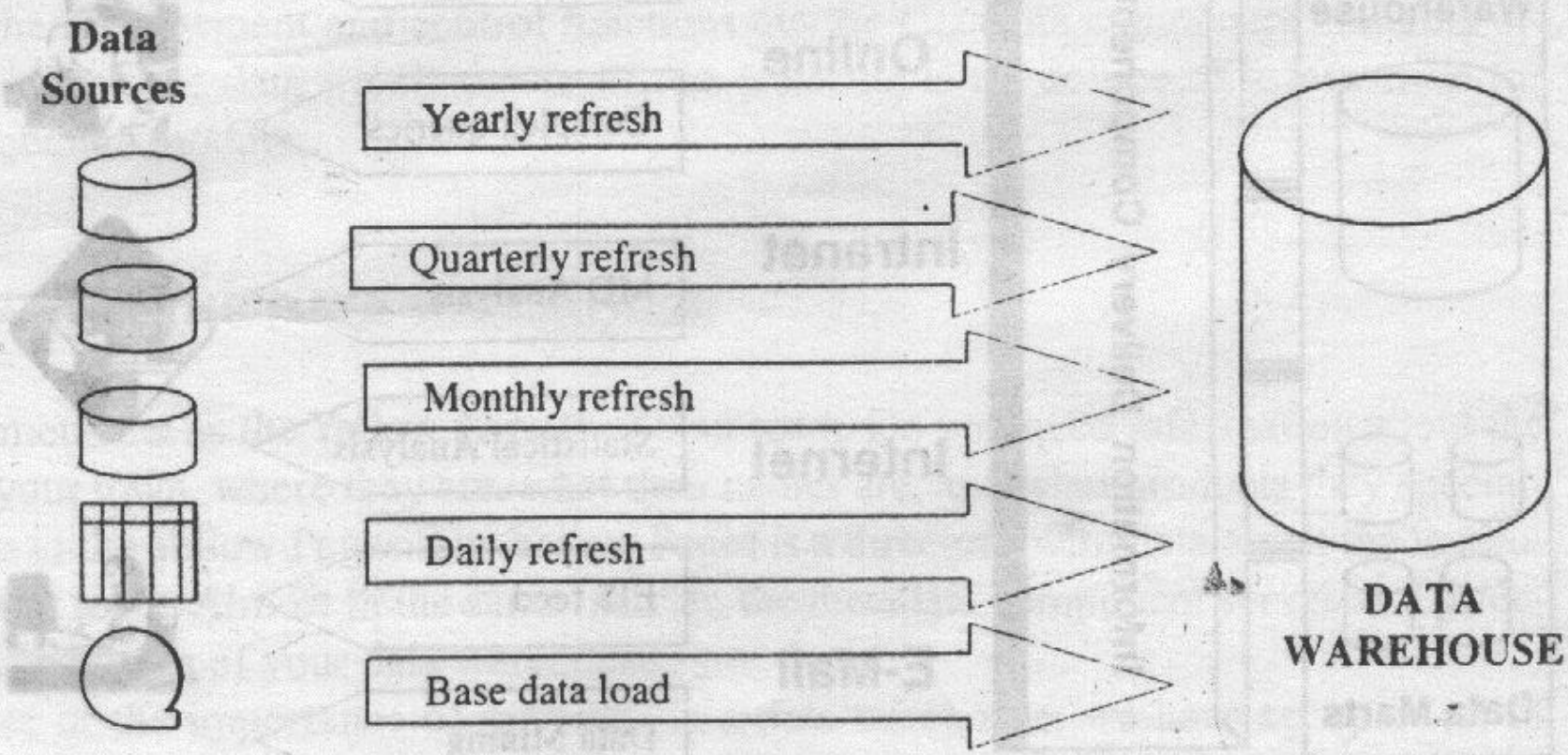


Figure 2-7 Data movements to the data warehouse.

Information Delivery Component

- Information is required for the wide community
  - Novice users
  - Casual users
  - Business analyst
  - Power users

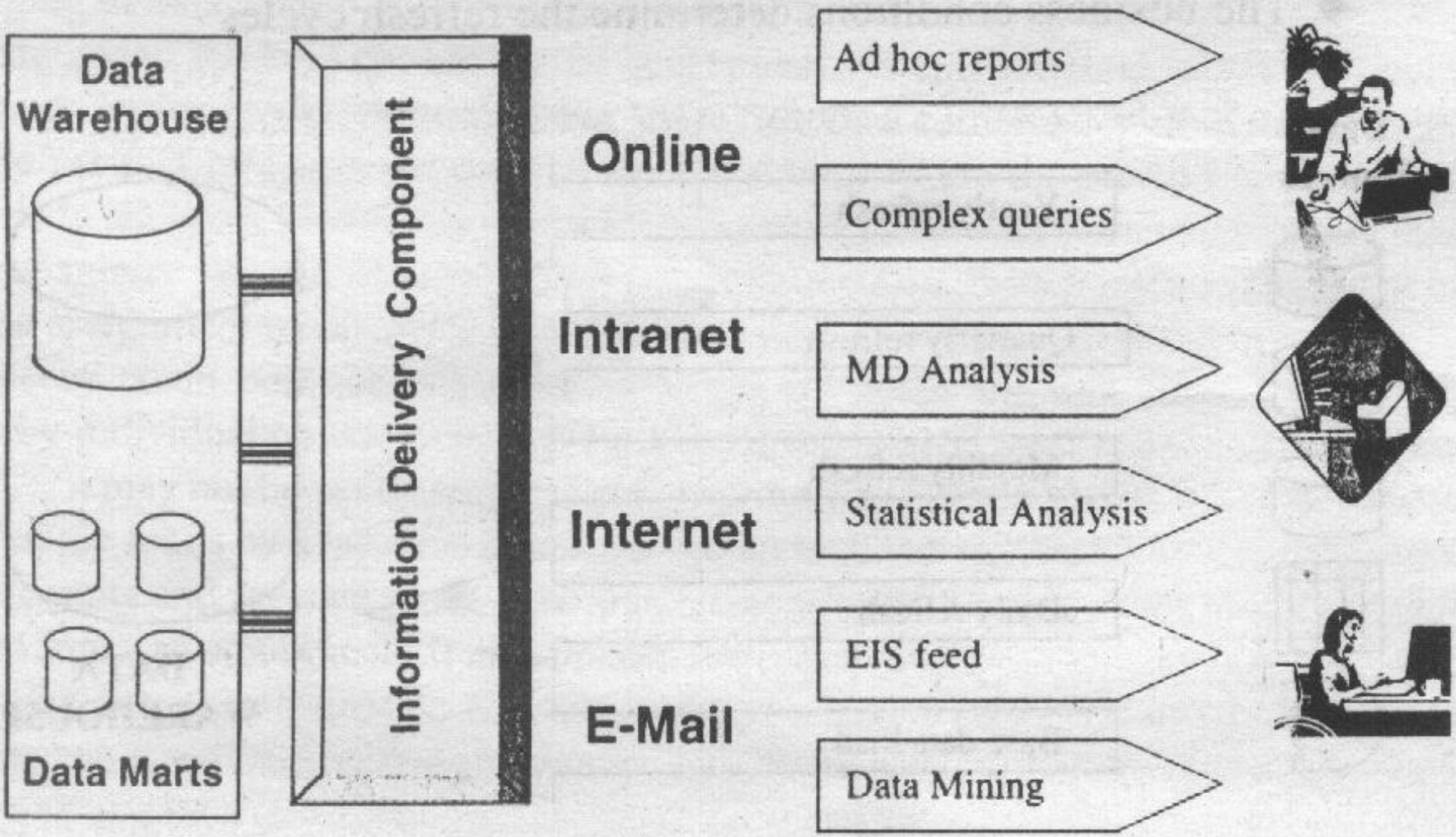
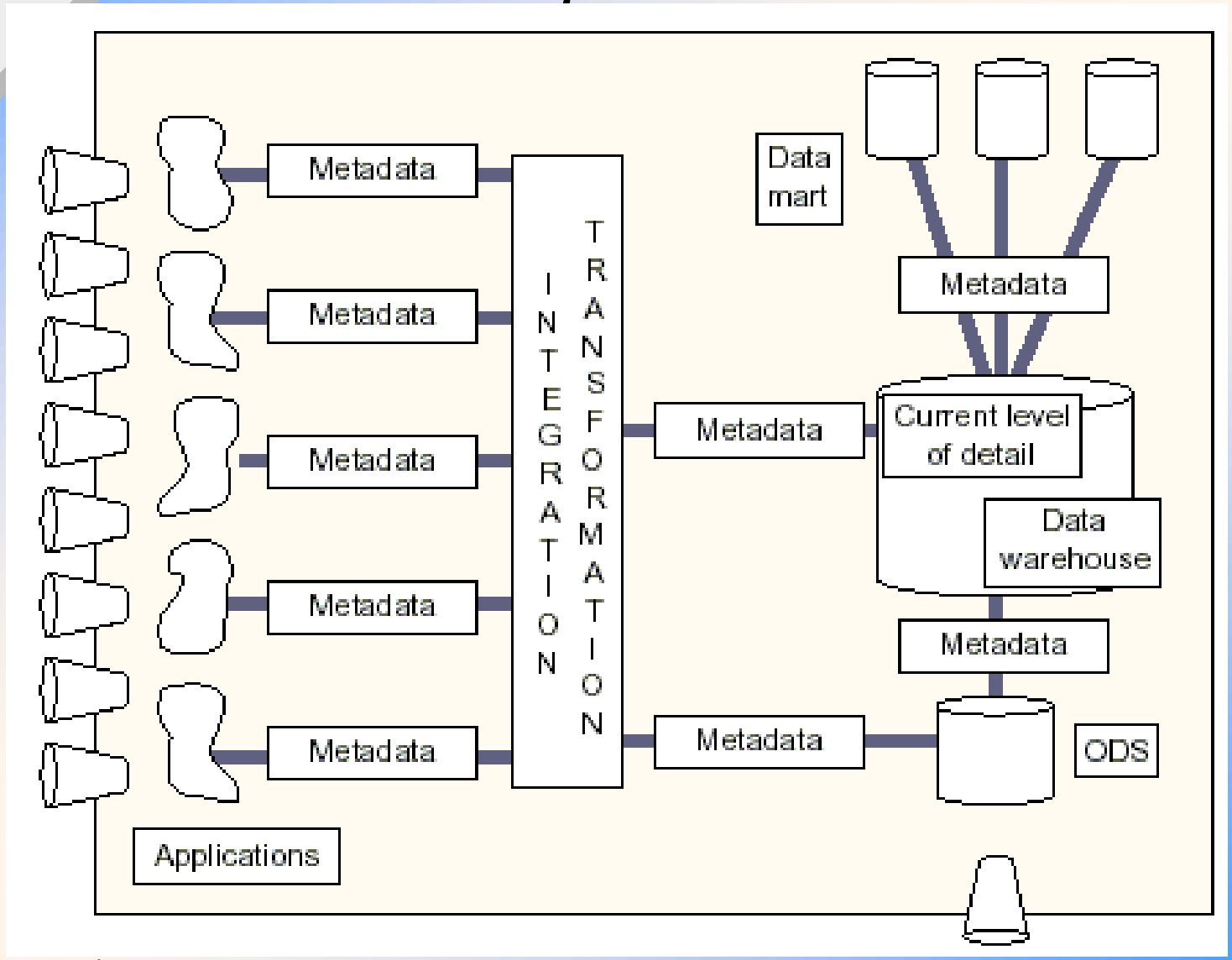


Figure 2-8 Information delivery component.

Meta Data Component

- It is similar to data dictionary
- It is at advance level than data dictionary
- Metadata is key architectural component of DWH
- There are three types of metadata
  - Operational metadata
  - Extraction an transformational metadata
  - End user meta data

Meta Data Component Contd..



Meta Data Component Contd..

- Special significance
  - First, it acts like a glue that connects all parts of DWH
  - Next, it provides information about the contents and structures to the developers
  - Finally, it opens the door to the end users and makes the contents recognizable in their own terms

Thank You Very Much