Chapter 1

Introduction to Data Warehousing & Dimensional modeling

Based on CO1: Identify applications which require data warehouse and select the suitable architecture required for any data ware house applications

By-Safa Hamdare

OUTLINE- Introduction to Data warehousing

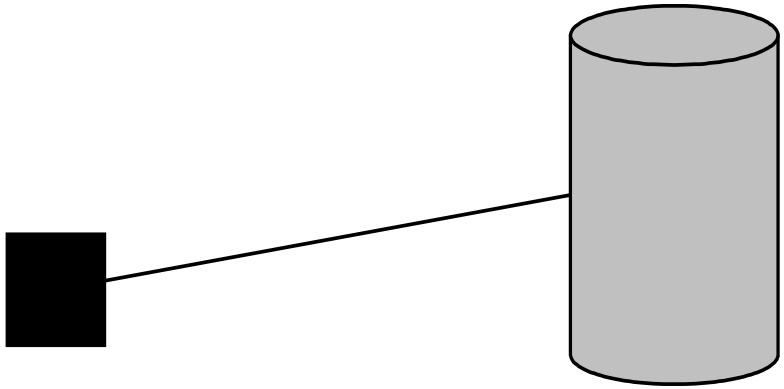
- Introduction to Strategic Information and its Need
- **Comparison between Data Warehouse i.e.** On-Line Analytical Processing (OLAP) & On-Line Transaction Processing (OLTP)
- Features of Data Warehouse
- Data warehouses versus Data Marts
- Top-down versus Bottom-up approach
- Data warehouse Architecture
- Data warehouse Infrastructure
- Metadata management

Introduction to strategic information and its need

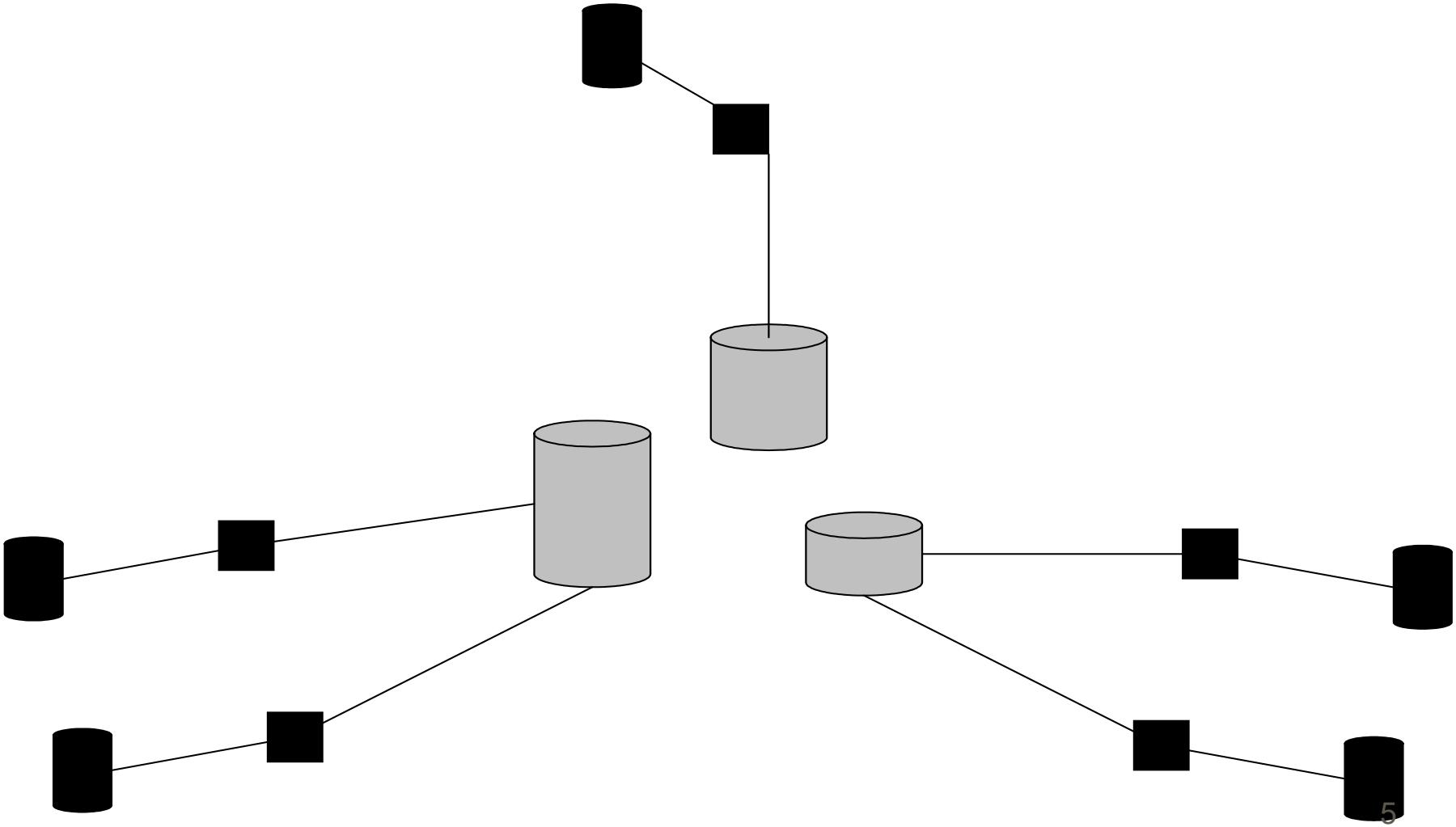
Need for Data Warehouse

– (May 2011, May 2012)

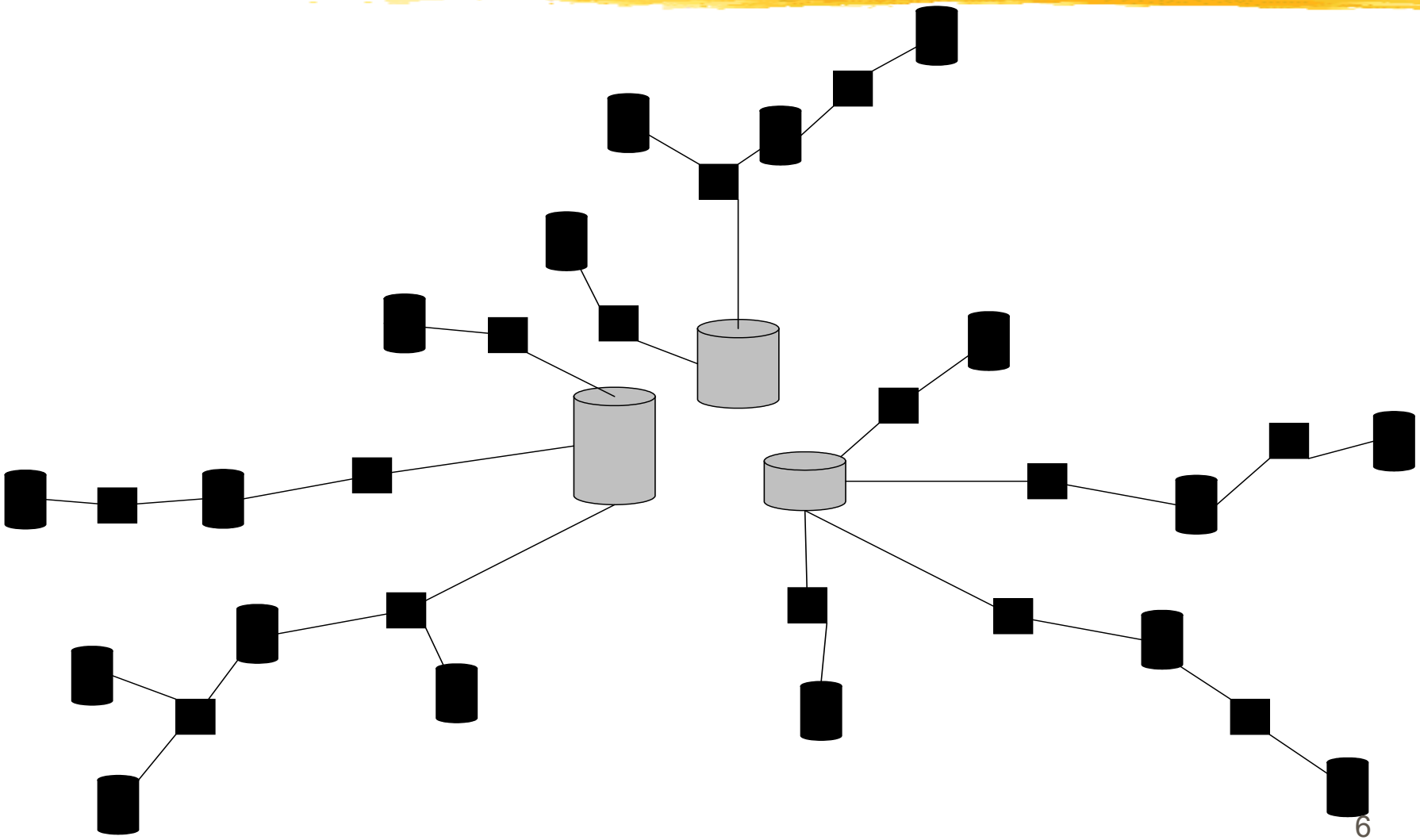
In the Beginning, life was simple…



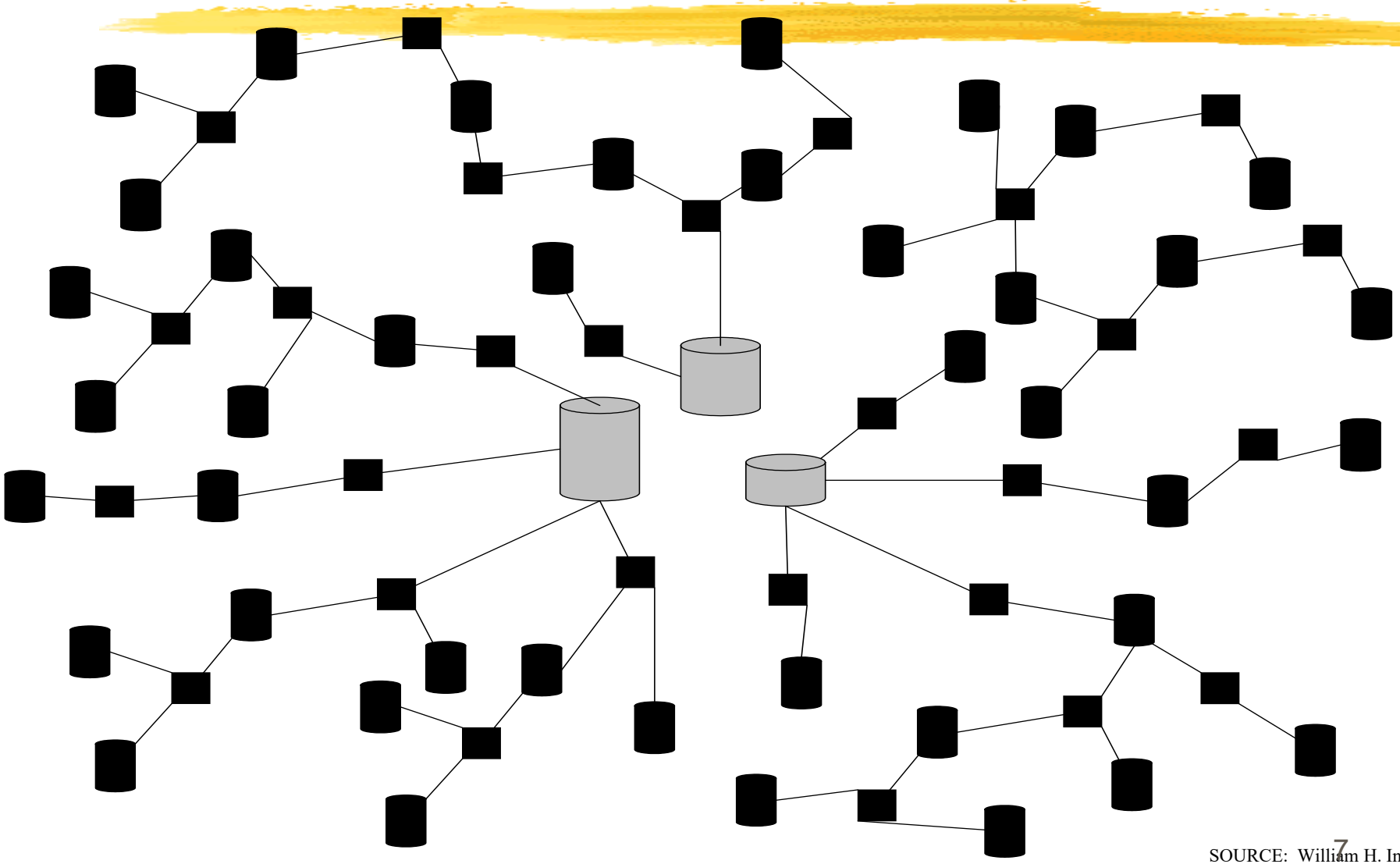
But…



Our information needs…



Kept growing. (The Spider web)



SOURCE: William H. Inmon

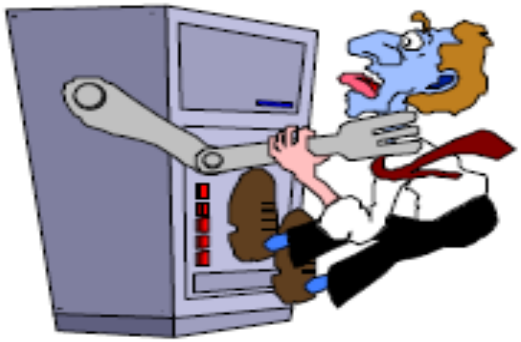
Data Warehouse Concepts

Why Do We Need A Data Warehouse ?

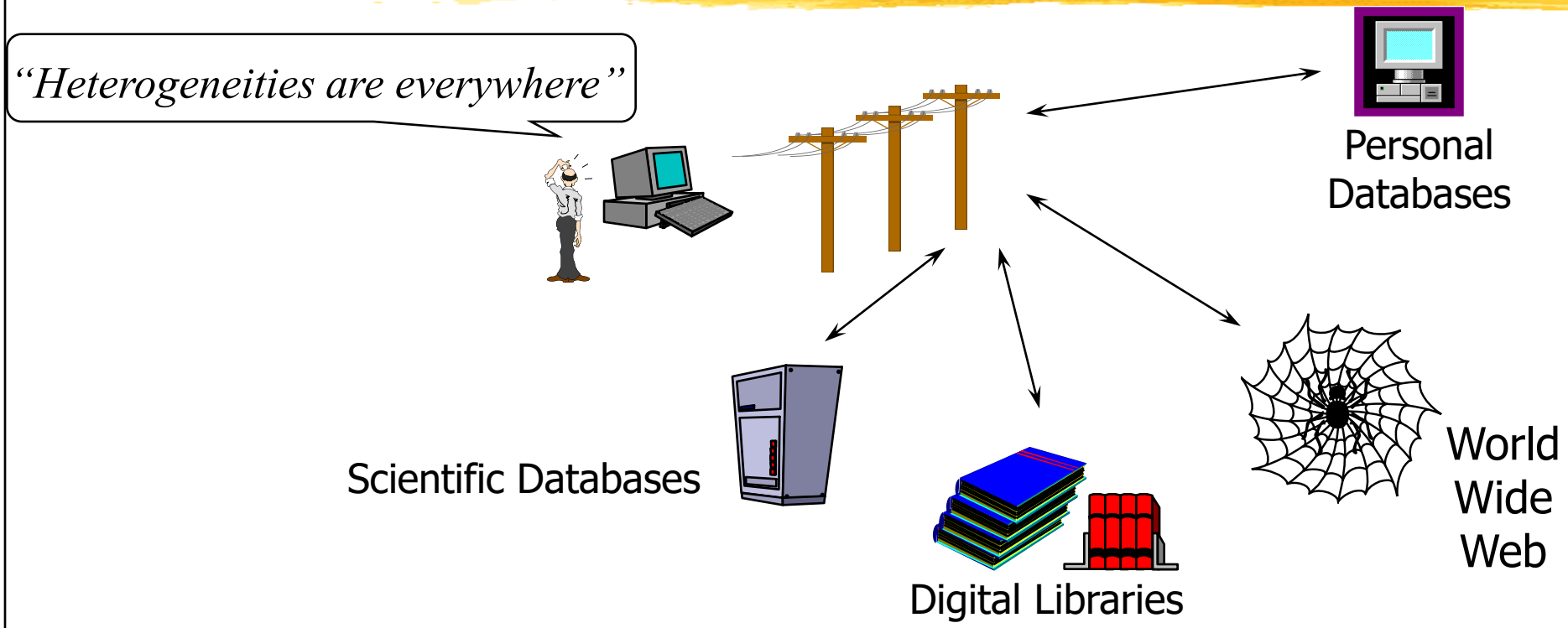
BETTER !
FASTER !
CHEAPER !

$\neq$

FUNCTIONALLY COMPLETE !

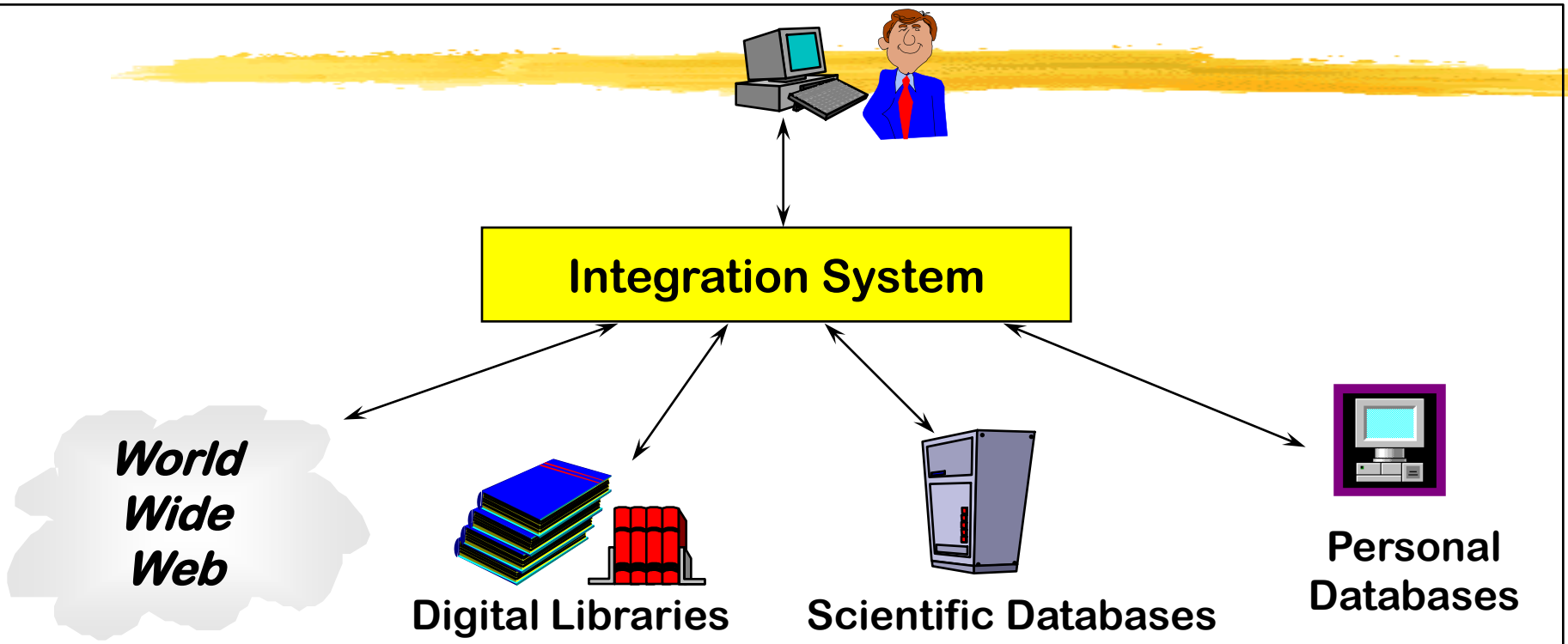


Problem: Heterogeneous Information Sources



- Different interfaces
- Different data representations
- Duplicate and inconsistent information

Goal: Unified Access to Data



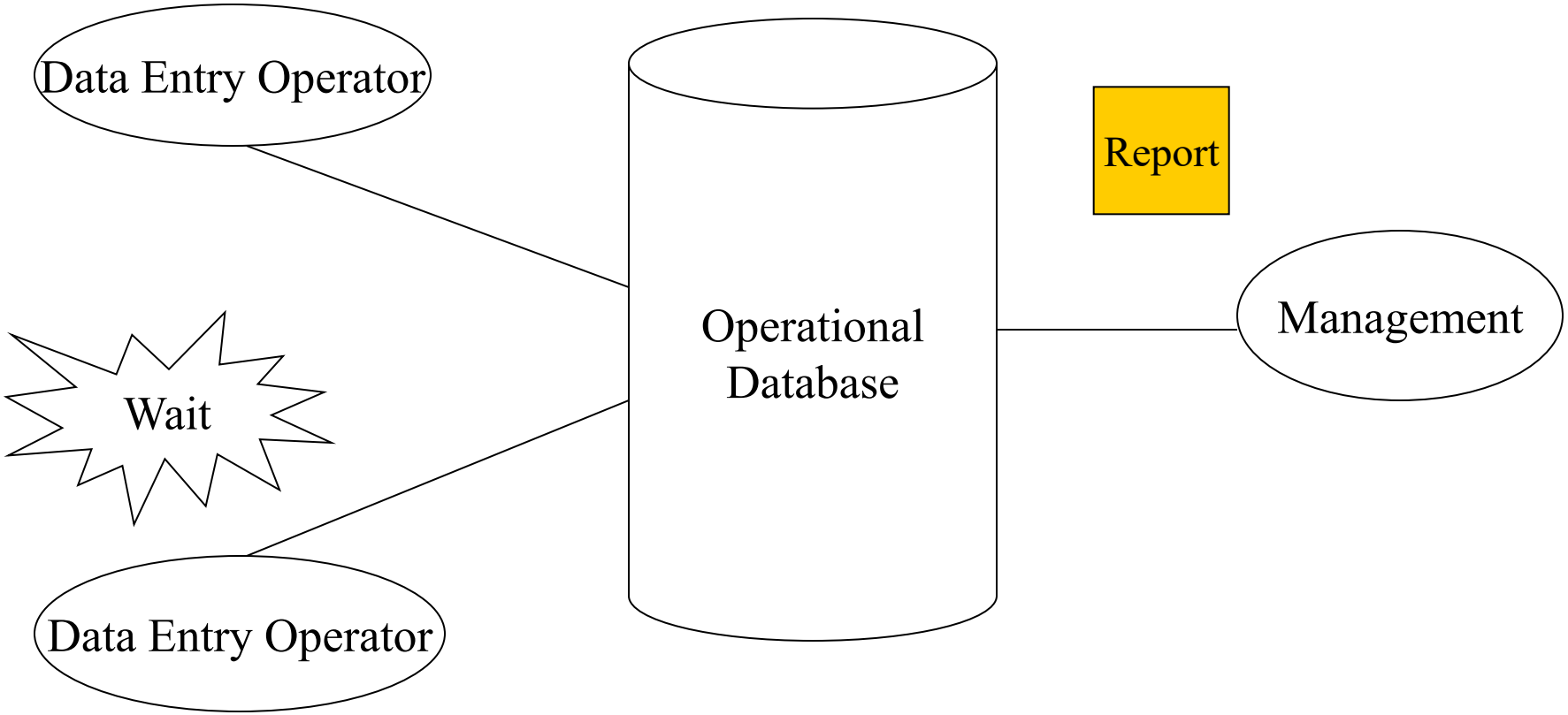
- Collects and combines information
- Provides integrated view, uniform user interface
- Supports sharing

Need for Data Warehouse – (May 2011, May 2012)

1. Consolidation of information resources from different data source
2. Improved query performance
3. Separate research and decision support functions from the operational systems
4. Foundation for data mining, data visualization, advanced reporting and OLAP tools

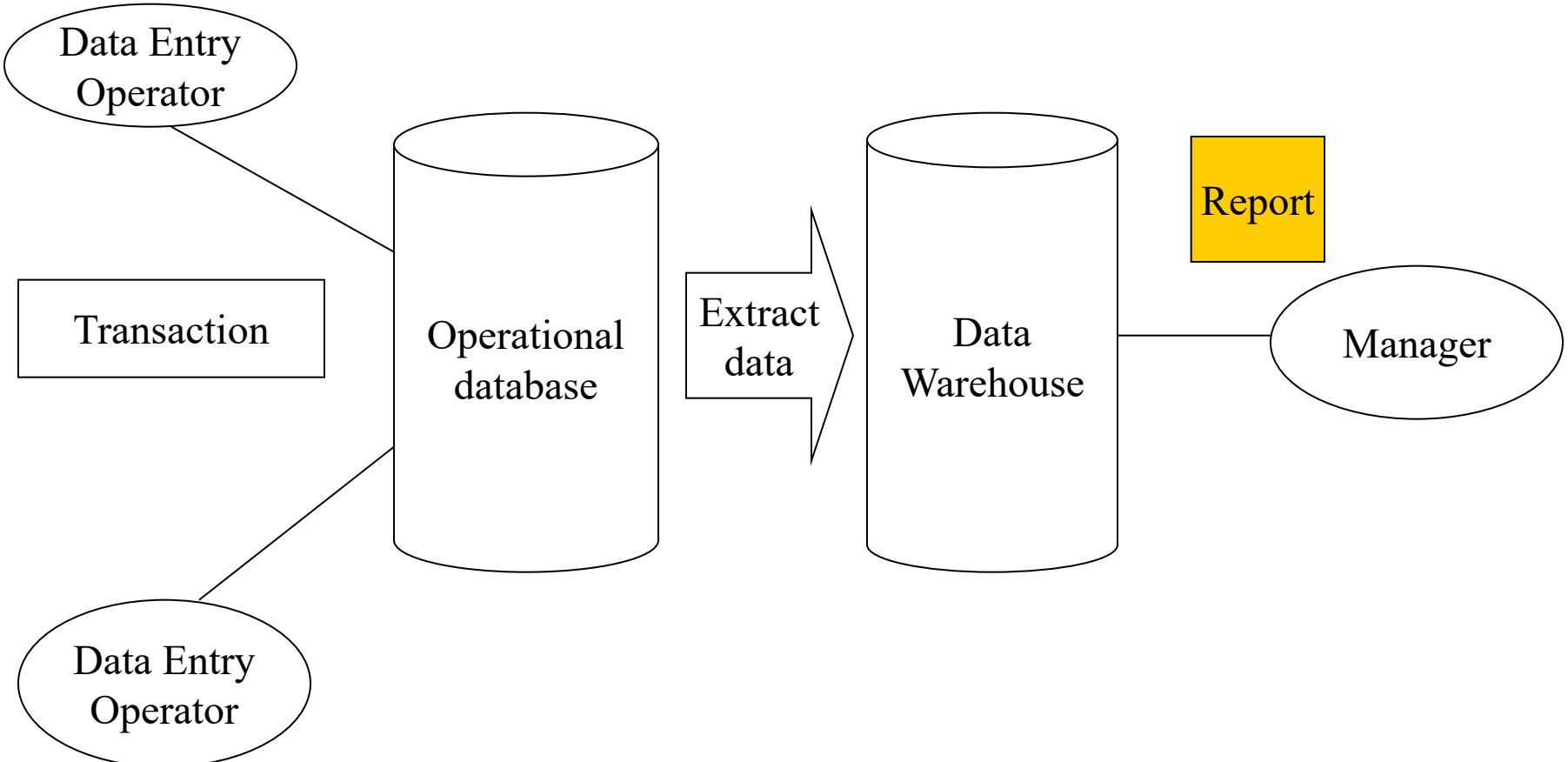
Scenario : explain need for DW

One Stop Shopping Super Market has huge operational database. Whenever Executives wants some report the OLTP system becomes slow and data entry operators have to wait for some time.



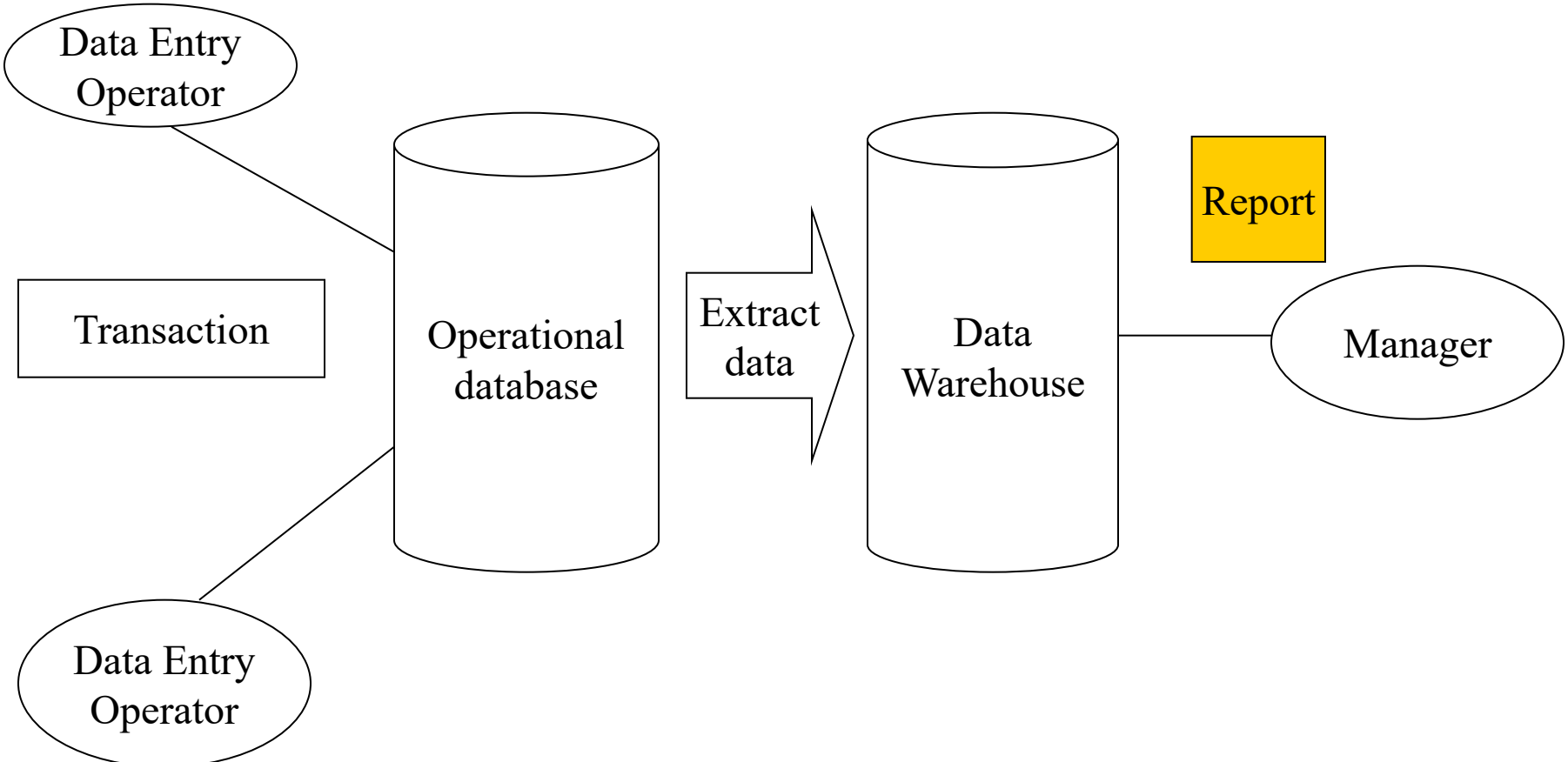
Solution to the Scenario

- Extract data needed for analysis from operational database. Store it in another system, the data warehouse.
- Refresh warehouse at regular intervals so that it contains up to date information for analysis. Warehouse will contain data with historical perspective.



Activity

- Select any operation database example
- Write 2 business rule that would be required by manager.



Summing up?

- Why do you need a warehouse?
  - Operational systems could not provide strategic information
  - Executive and managers need such strategic information for
    1. Making proper decision
    2. Formulating business strategies
    3. Establishing goals
    4. Setting objectives
    5. Monitoring results

Motivation

- "Modern organization is drowning in data but starving for information".
- ***Operational processing*** (transaction processing) captures, stores and manipulates data to support daily operations.
- ***Information processing*** is the analysis of data or other forms of information to support decision making.
- ***Data warehouse*** can consolidate and integrate information from many internal and external sources and arrange it in a meaningful format for making business decisions.

Need to separate operational (Normal Db) and information systems (Data warehouse)

- High performance for both systems
  - DBMS— tuned for OLTP: access methods, indexing, concurrency control, recovery
  - Warehouse—tuned for OLAP: complex OLAP queries, multidimensional view, consolidation
- Different functions and different data:
  1. **missing data**: Decision support requires historical data which operational DBs do not typically maintain
  2. **data consolidation**: DS requires consolidation (aggregation, summarization) of data from heterogeneous sources
  3. **data quality**: different sources typically use inconsistent data representations, codes and formats which have to be reconciled

Compare OLTP and OLAP System? 5 marks, 10 marks

- ASKED IN MU EXAM (MAY 2010,MAY 2011, MAY 2012, DEC 2016)

Data Warehouse vs. Operational DBMS

-(May 2010,May 2011, May 2012, Dec 2016) 5, 10 marks

- **OLTP (on-line transaction processing)**
  - Major task of traditional relational DBMS
  - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- **OLAP (on-line analytical processing)**
  - Major task of data warehouse system
  - Data analysis and decision making
- **Distinct features (OLTP vs. OLAP):**
  - **User and system orientation:** customer **vs.** market
  - **Data contents:** current, detailed **vs.** historical, consolidated
  - **Database design:** ER + application **vs.** star + subject
  - **View:** current, local **vs.** evolutionary, integrated
  - **Access patterns:** update **vs.** read-only but complex queries

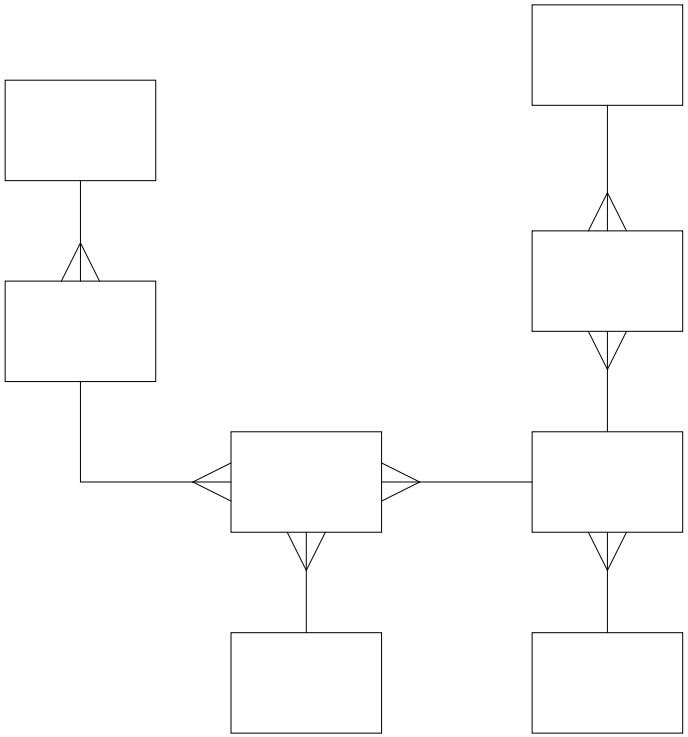
Compare OLTP (Operational System) and OLAP (Data Warehouse)

-(May 2010,May 2011, May 2012, Dec 16) 5 m,10 m

	OLTP	OLAP
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

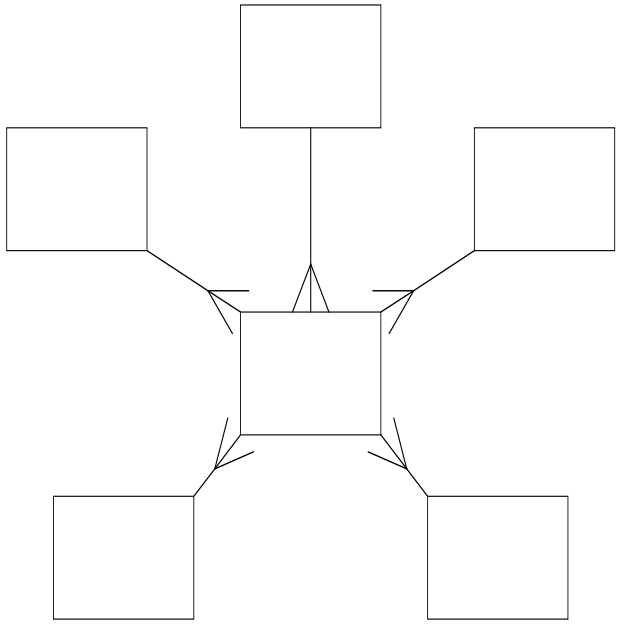
Design Differences

Operational System



ER Diagram

Data Warehouse



Star Schema

Data Warehouse: Definition, Benefits and Features of Data Ware house

What is Data Warehouse?

- Defined in many different ways, but not rigorously.
  - A decision support database that is maintained separately from the organization's operational database
  - Support information processing by providing a solid platform of consolidated, historical data for analysis.

Definition of a Data Warehouse – (May 2011, May 2012)

"A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process."

—W. H. Inmon

Data warehousing:

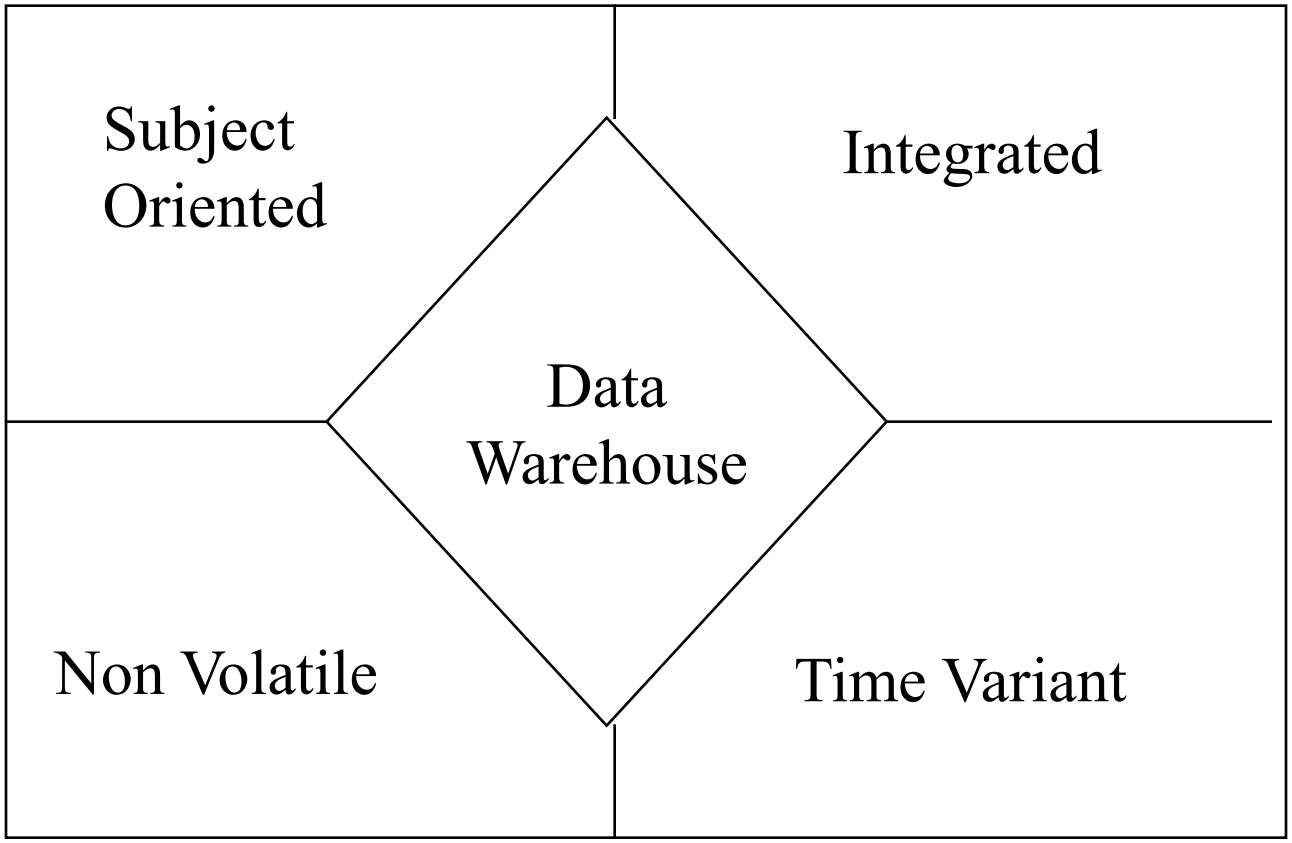
The process of constructing and using data warehouse

Building Blocks of a Data Warehouse

Why is data integration required in a data warehouse, more so than in an operational application? [5 marks, Dec 2019]

Data Warehouse Properties – (Dec 2010)

Data in the data warehouse is:



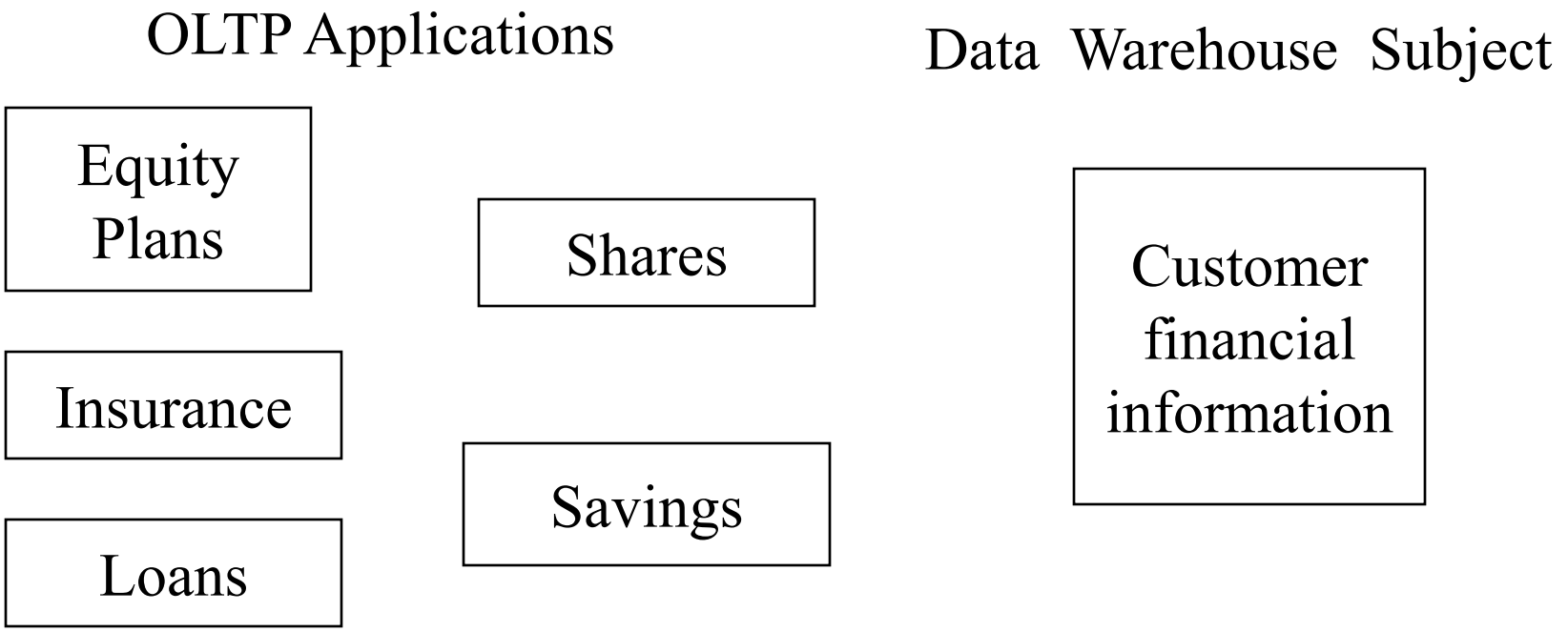
1. Data Warehouse—Subject-Oriented

- Organized around major subjects, such as customer, product, sales
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process

Subject-Oriented

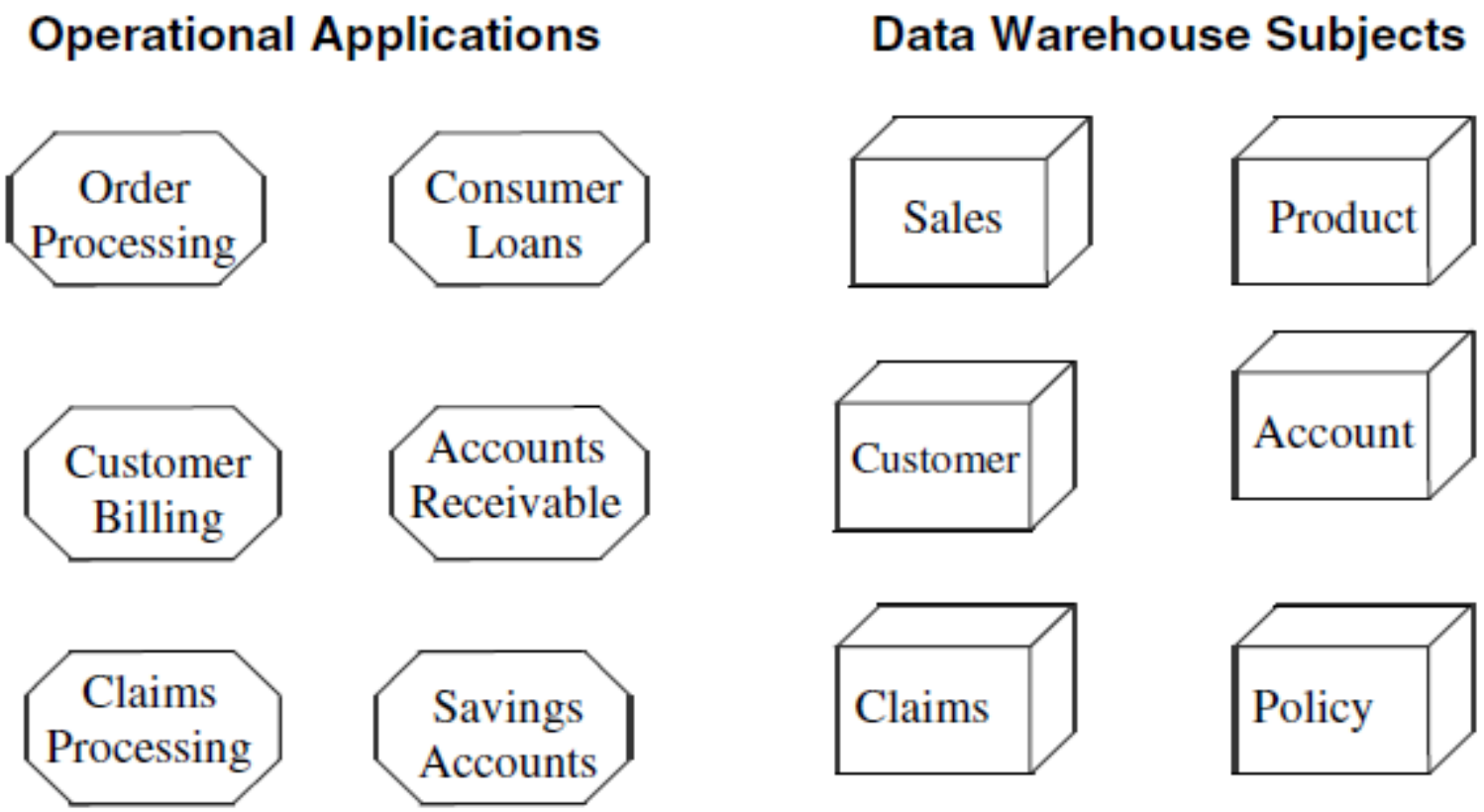
Data are organized based on how the users refer to them.

Data is categorized and stored by business subject rather than by application.



Example of Subject Oriented:

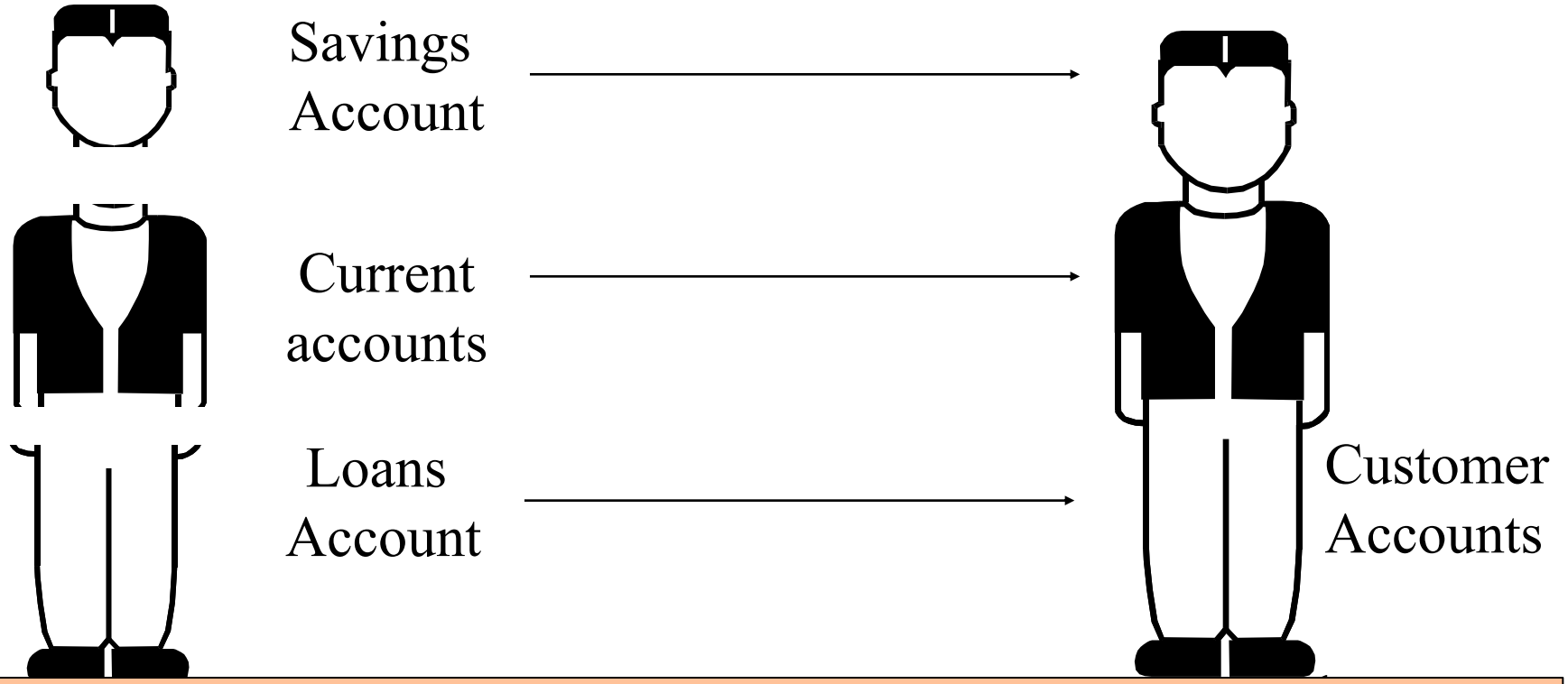
In the data warehouse, data is not stored by operational applications, but by business subjects.



2. Data Warehouse—Integrated

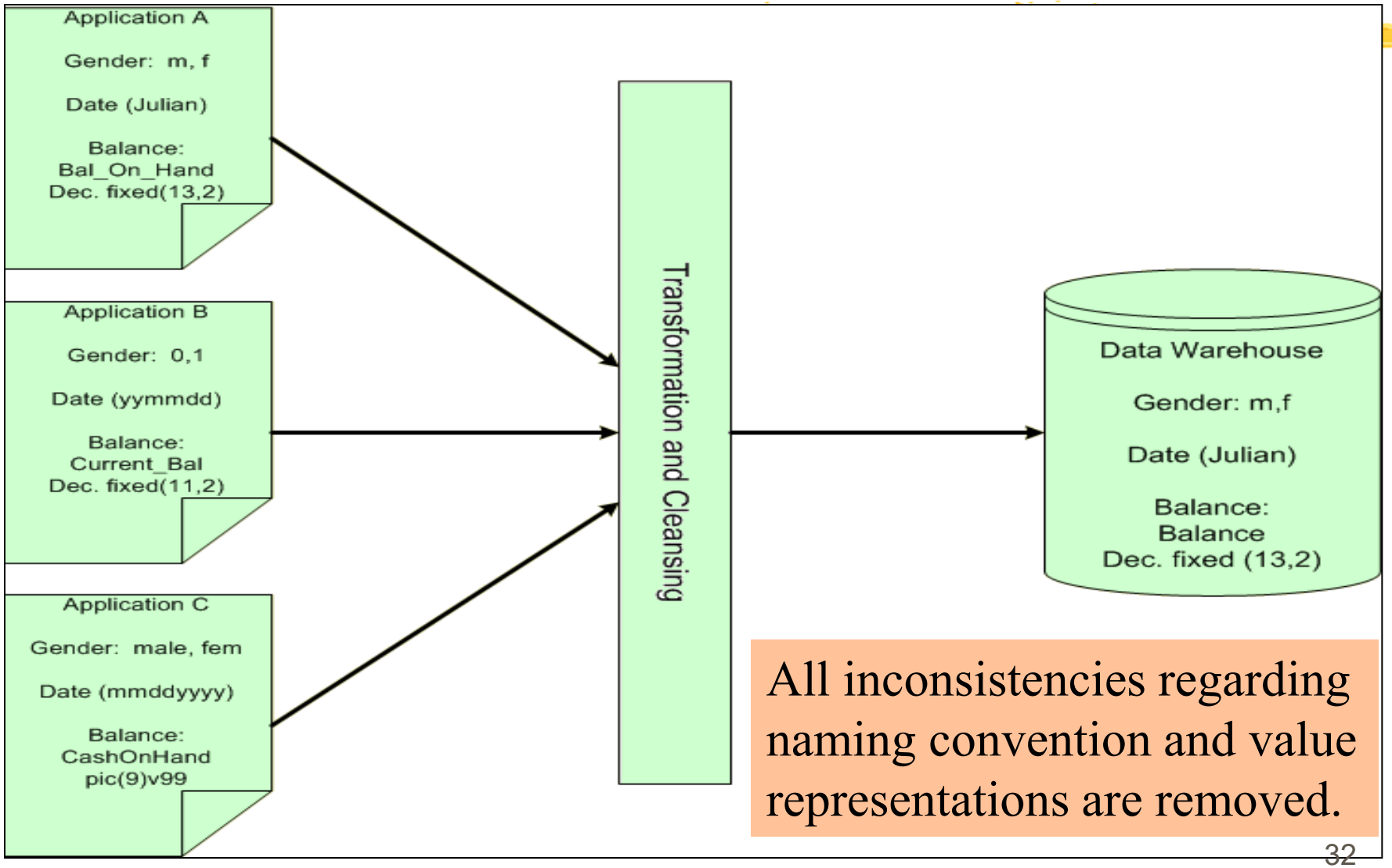
- Constructed by integrating multiple, heterogeneous data sources
  - relational databases, flat files, on-line transaction records
- Data cleaning and data integration techniques are applied.
  - Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources
    - E.g., **Hotel price: Room rent, tax, Food bill covered, etc.**
  - When data is moved to the warehouse, it is converted.

2. Integrated- Data on a given subject is defined and stored once.



Even within 3 applications, there could be several variations. Naming conventions, attributes for data items could be different. For E.g. acc_no in savings account application could be 8 bytes long, but only 6 bytes in checking account application.

Data Integrated- Example



All inconsistencies regarding naming convention and value representations are removed.

Data Integrated- Example

Integrated

Operational Systems	
Order Processing	Order ID = 10
Accounts Receivable	Order ID = 12
Product Management	Order ID = 8

D/W

Order ID = 16

HR System	Sex = M/F
Payroll	Sex = 1/2
Product Management	Sex = 0/1

D/W

Sex = M/F

3. Data Warehouse—Time Variant

- The time horizon for the data warehouse is significantly longer than that of operational systems
  - Operational database: current value data
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Every key structure in the data warehouse
  - **Contains an element of time, explicitly or implicitly**
  - But the key of operational data may or may not contain "time element"

Time-Variant-Data is stored as a series of snapshots, each representing a period of time

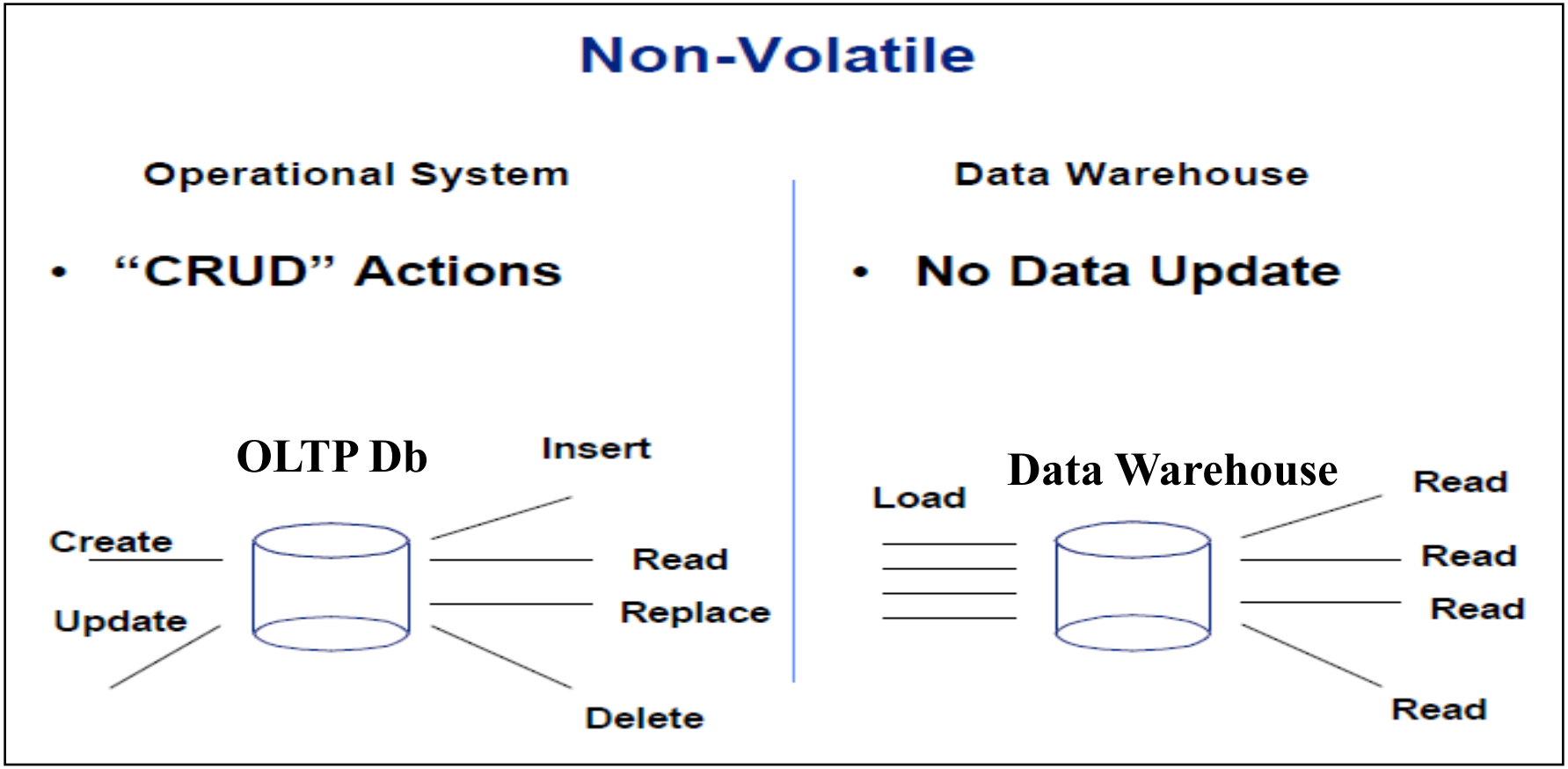
Operational System	Data Warehouse
• View of The Business Today	• Designated Time Frame (3 - 10 Years)
• Operational Time Frame	• One Snapshot Per Cycle
• Key Need Not Have Date	• Key Includes Date

4. Data Warehouse-Nonvolatile

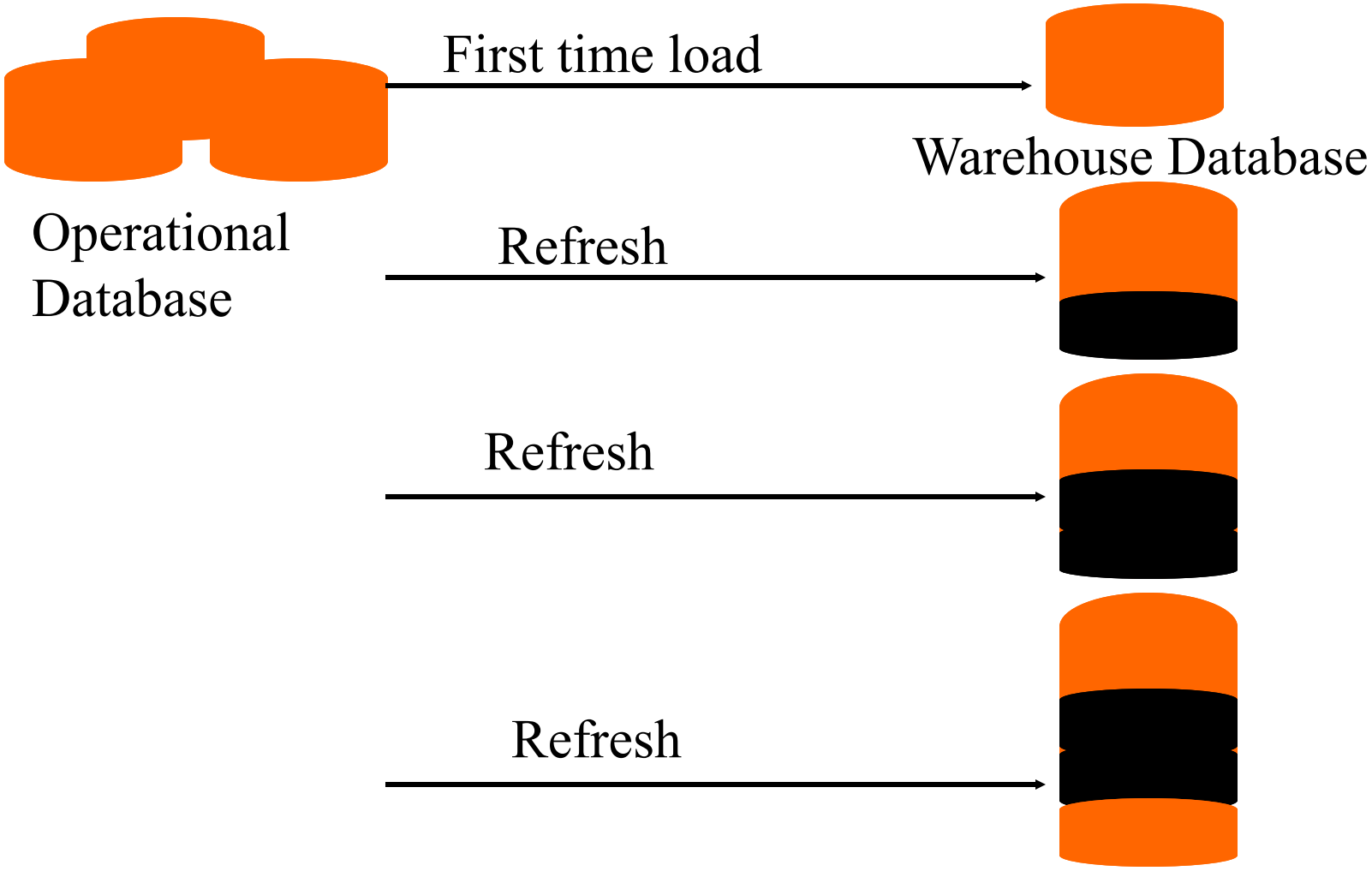
- A physically separate store of data transformed from the operational environment
- Operational update of data does not occur in the data warehouse environment
  - Does not require transaction processing, recovery, and concurrency control mechanisms
  - Requires only two operations in data accessing:
    - ***initial loading of data*** and ***access of data***

Nonvolatile: Data are stored in read-only format and do not change over time.

Typically data in the data warehouse is not updated or delelted.



Changing Data



5. Data Granularity

- In an operational system, data is usually kept at the lowest level of detail.
- In a DW, data is summarized at different levels.

E.g. Three data levels in a banking data warehouse

Daily Detail	Monthly Summary	Quarterly Summary
Account	Account	Account
Activity Date	Month	Month
Amount	No. of transactions	No. of transactions
Deposit/ Withdraw	Withdrawals	Withdrawals
	Deposits	Deposits
	Beginning Balance	Beginning Balance
	Ending Balance	Ending Balance

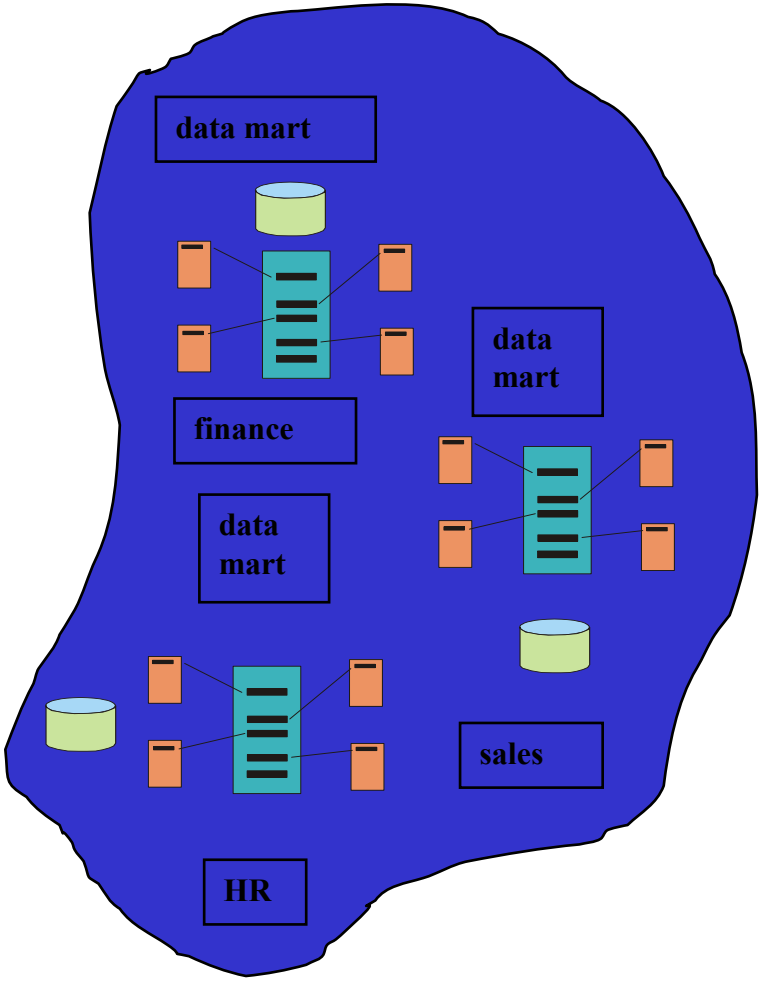
Comparison between Data Warehouse & Data Mart

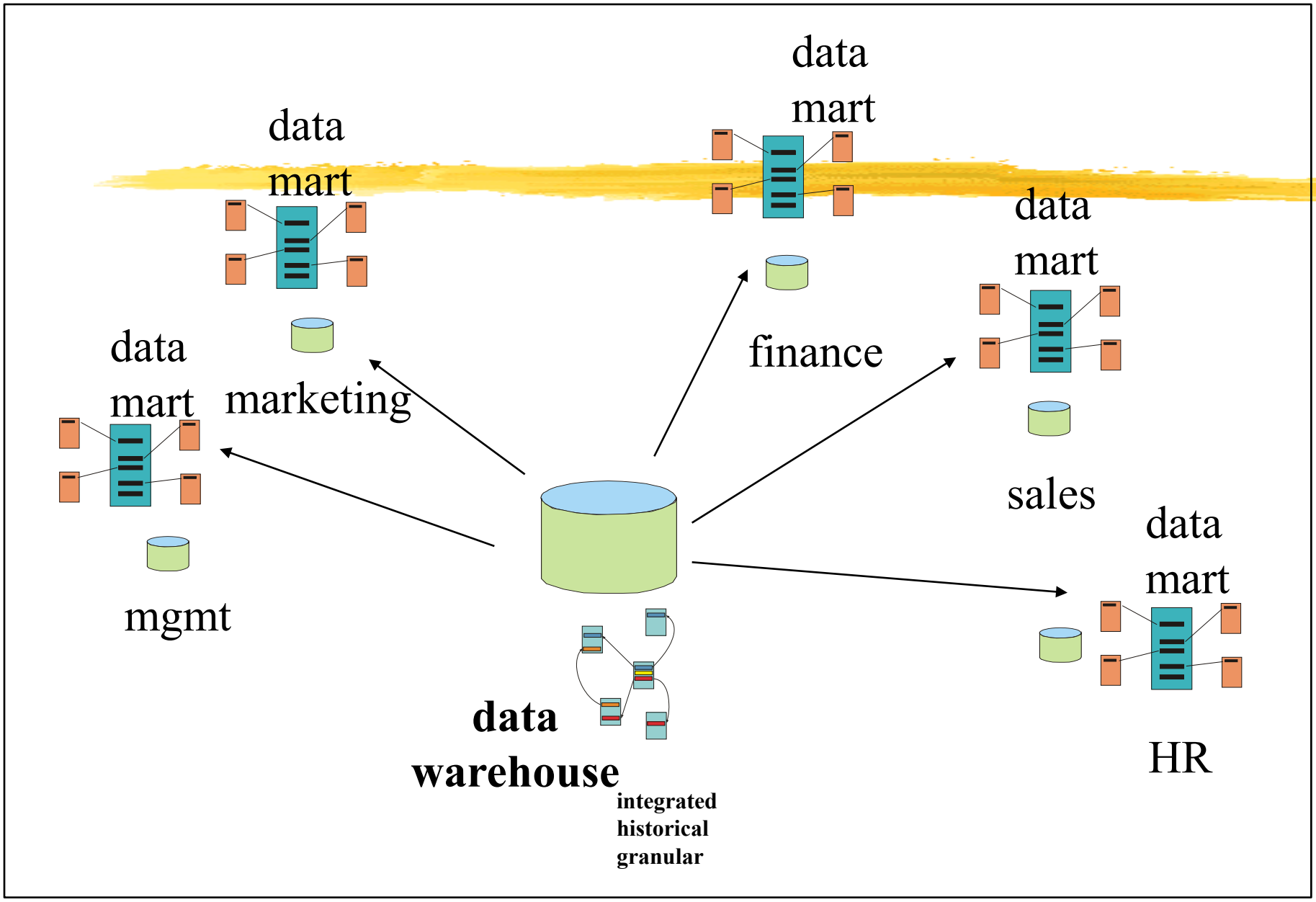
What are the differences between Data Warehouse and Data Mart? 05,10 marks

- ASKED IN MU EXAM (MAY 2013, DEC 2016, MAY 2016, MAY 2017)

Data Warehouse & Data Mart

- **Data warehouse**
  - a data warehouse is the union of all of the data marts designed specifically to support management decision making
- **Data mart**
  - A subset of a data warehouse for small and medium-size businesses or departments within larger companies
    - Do not normally contain detailed operational data unlike data warehouses.
    - May contain certain levels of aggregation




data
mart
data
mart
data
mart
data
mart
marketing
finance
data
mart
sales
mgmt
data
mart
HR
data
warehouse
integrated
historical
granular

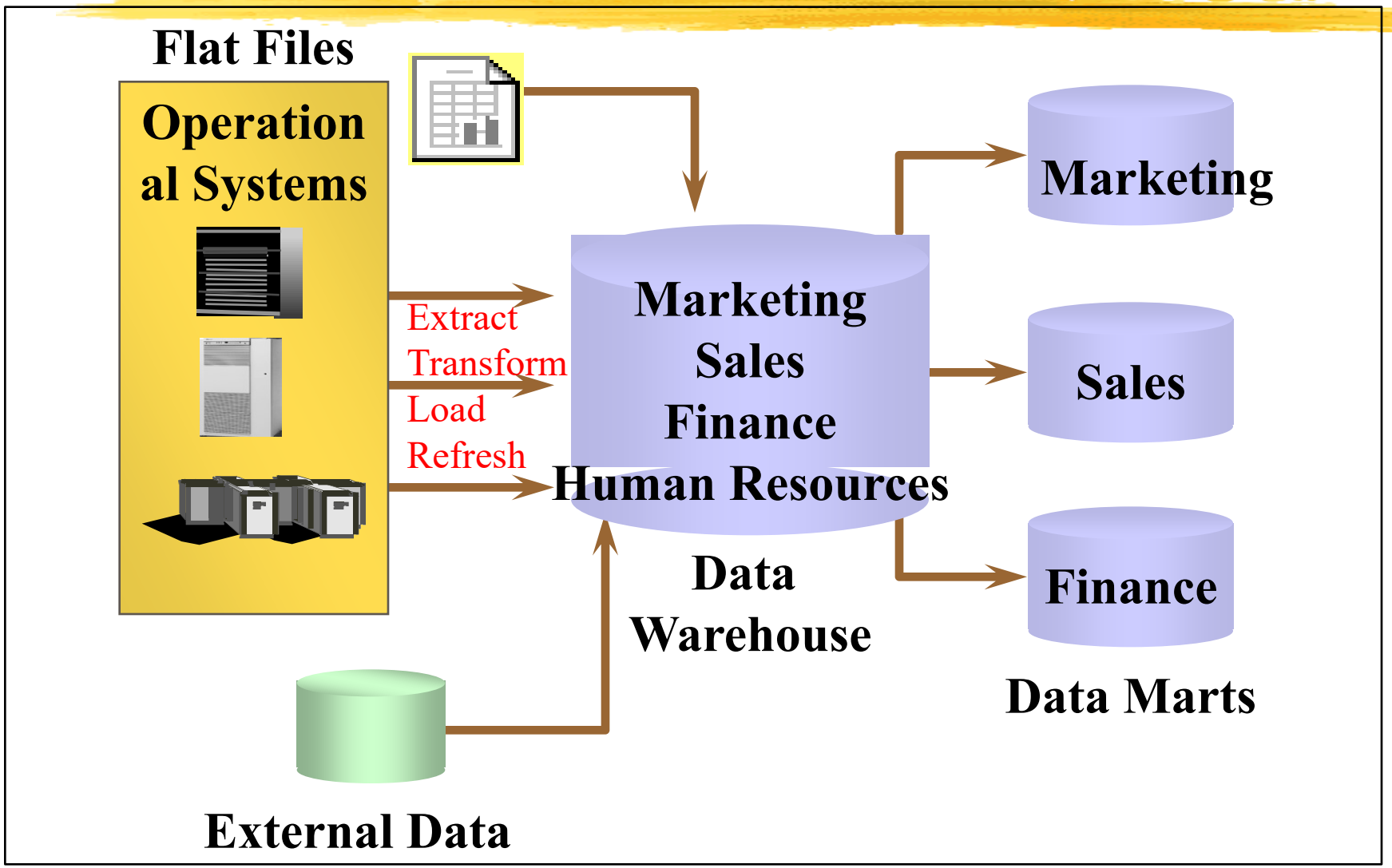
Data Warehouse verses data marts – (May 2013) 5 marks

Data Warehouse	Data Mart
A data warehouse is application independent. i.e. Corporate/ Enterprise wide	A data mart is a dependent on specific DSS application. i.e. Departmental
Union of all data marts	A single business process
It is centralized, and enterprise wide	It is decentralized by user area
Data received from staging area i.e. data source is many subject	Star- joined (facts & dimensions) i.e. data source is single subject of concern to the user.
It is well planned	It is possibly not planned
It is highly flexible	It is restrictive
Implementation takes months to year	Implementation is done usually in months
Generally size is from 100 GB to 1 TB	Generally size is less than 100 GB

Differentiate between top down & bottom up approaches for building a data warehouse. Discuss the merits and limitations of each approach? (10 marks)

- ASKED IN MU EXAM (MAY 2010,MAY 2011, MAY 2012, DEC 2017)

Top Down Approach / Dependent Data Mart- A subset that is created directly from a data warehouse



Dependent Data Mart/ Top Down Approach

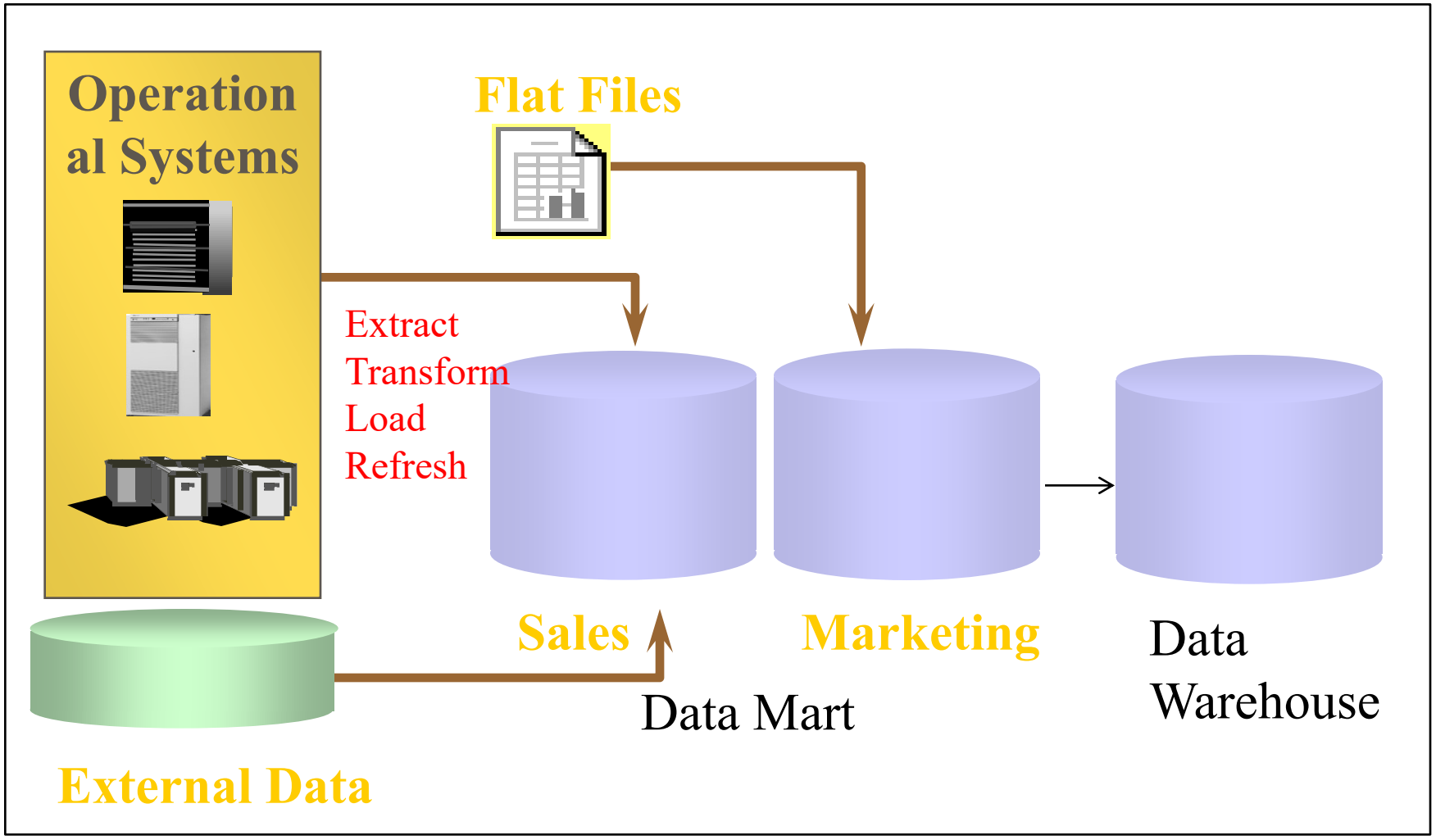
- The data flow in the top down OLAP environment begins with data extraction from the operational data sources to the loading of data into data warehouse.
- Once the Data Warehouse aggregation and summarization processes are complete, the data mart refresh cycles will extract the data from the DW and perform a new set of transformation on them.
- **Advantages:**

1. It is not just union of disparate data marts but it is inherently architected.
2. The data about the content is centrally stored and the rules and control are also centralized.
3. The results are obtained quickly if it is implemented with iterations.

- **Disadvantages:**

1. Time consuming process with an iterative method
2. The failure risk is very high.
3. As it is integrated a high level of cross functional skills are required.

Independent Data Mart / Bottom-Up Approach- A small data warehouse designed for a strategic business unit or a department



Independent Data Mart / Bottom-Up Approach-

- This architecture makes the data warehouse more of a virtual reality than a physical. All data marts could be located in one sever or could be located on different severs across the enterprise while the data warehouse would be virtual entity being a sum total of all the data marts.
- **Advantages:**
  1. This model strikes a good balance between centralized and localized flexibility.
  2. Manageable pieces are faster and are easily implemented.
  3. Risk of failure is low.
  4. Allows one to create important data mart first.
- **Disadvantages:**

1. Allows redundancy of data in every data mart.
2. Preserves inconsistent and incompatible data.
3. Grows unmanageable interfaces.

Two approaches in designing a Data Warehouse - (MAY 2010,MAY 2011, MAY 2012, DEC 2017)

Top-down approach	Bottom-up approach
Enterprise view of data	Narrow view of data
Inherently architected	Inherently incremental
Single, central storage of data	Faster implementation of manageable parts
Centralized rules and control	Each data mart is developed independently
Takes longer time to build	Comparatively less time than a DW
Higher risk to failure	Less risk of failure
Needs higher level of cross-functional skills	Unmanageable interfaces

Data Warehouse Architecture

Write short note on Data Warehouse Architecture?
(10 marks) (May 2010, Dec 2010, May 2011, May 2012)

Define Data warehouse. Explain what is the need for developing a data Warehouse and hence explain its architecture? 10 marks (MAY 2011,dec 2011, may 2012)

What is architecture?

- The structure that brings all the components of a data warehouse together is known as the architecture.
- Many factors affect the architecture of a DW
  - Integrated data
  - Data preparation and storing
  - Data delivery
  - Technology
- Comprehensive blueprint
- **Data Warehouse Architecture divided into parts:**
  1. **Data Acquisition**
  2. **Data Storage**
  3. **Information Delivery**
  4. **Management & control**

Architecture of data warehouse

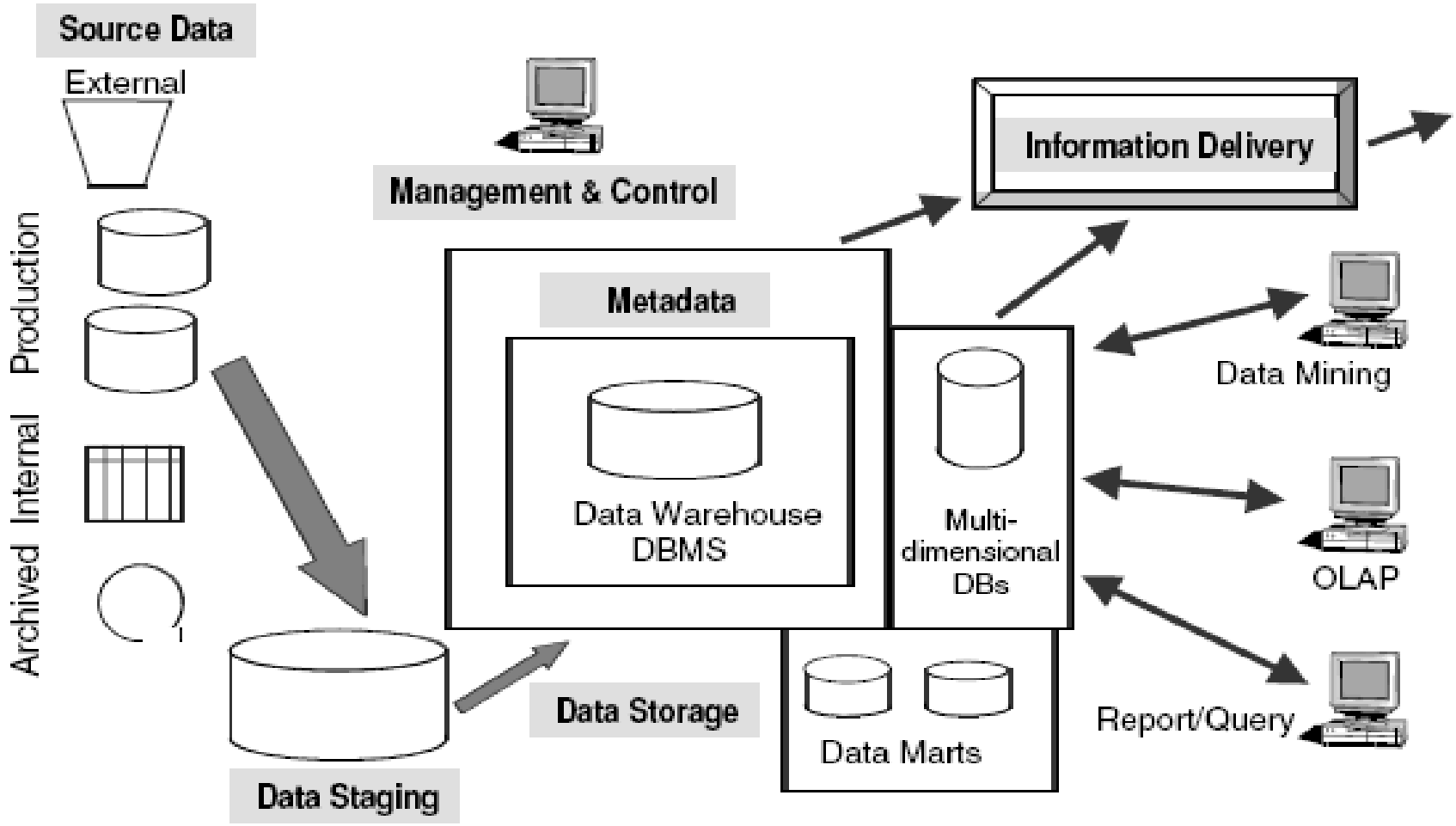
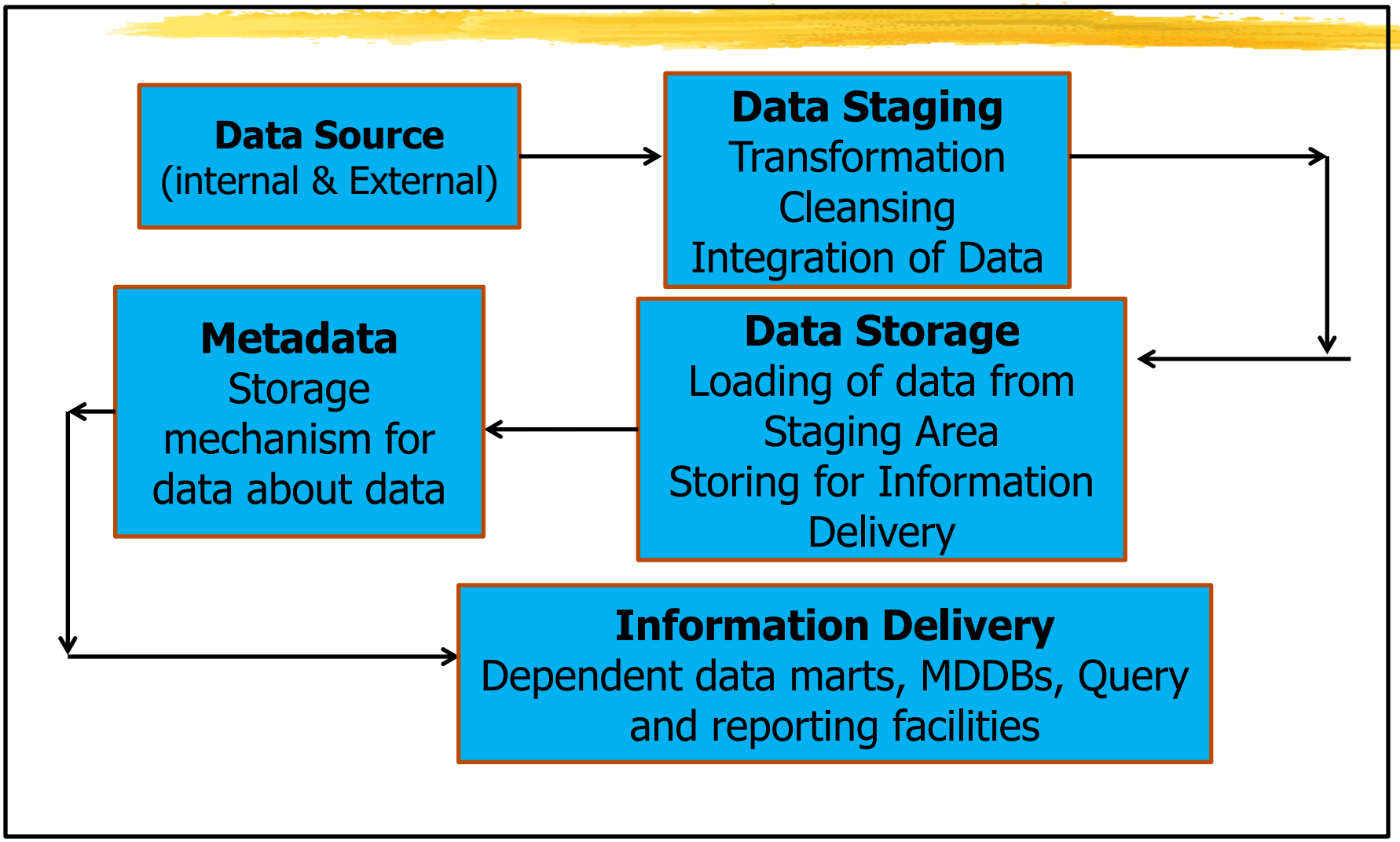
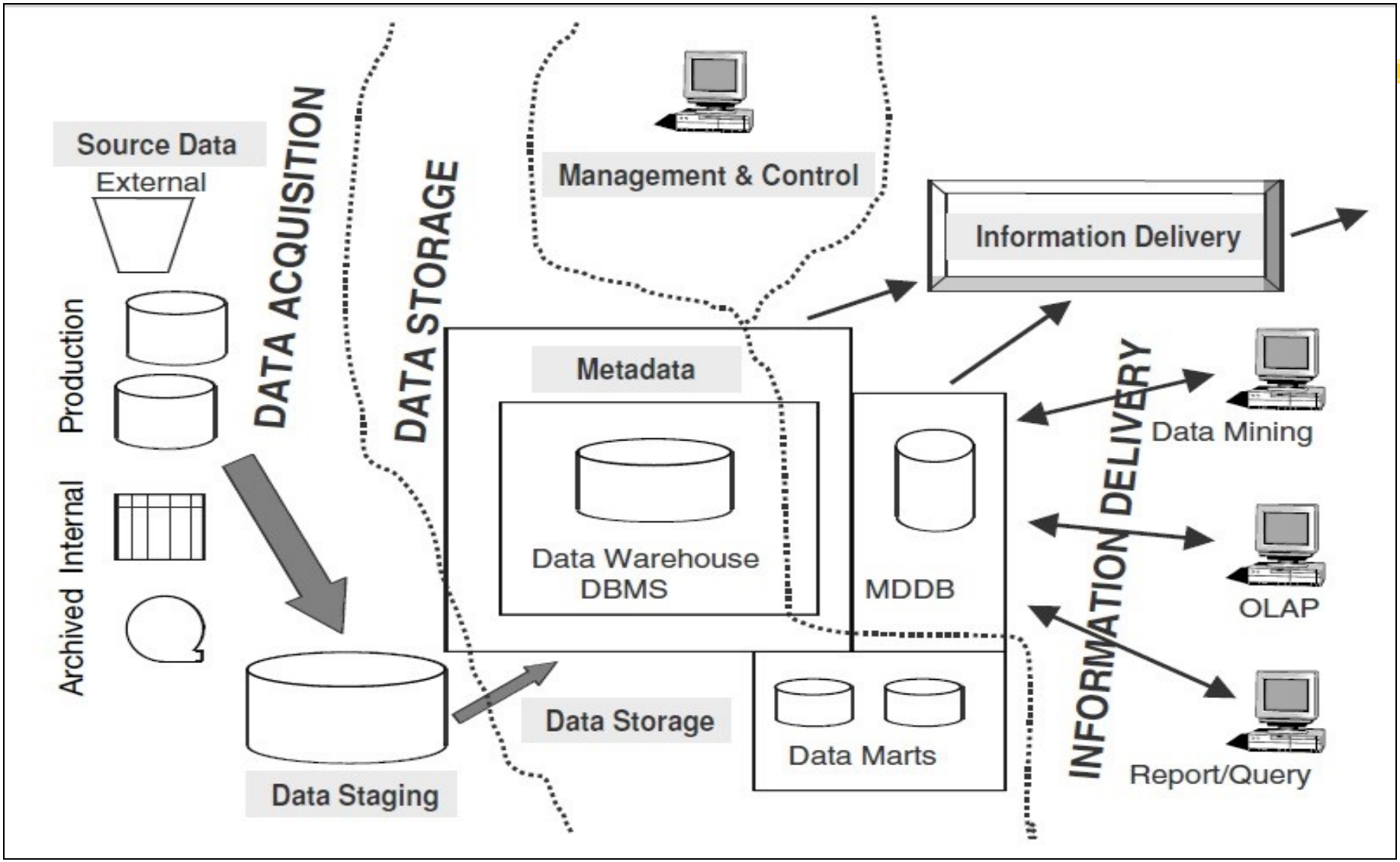


Figure 2-6 Data warehouse: building blocks or components.

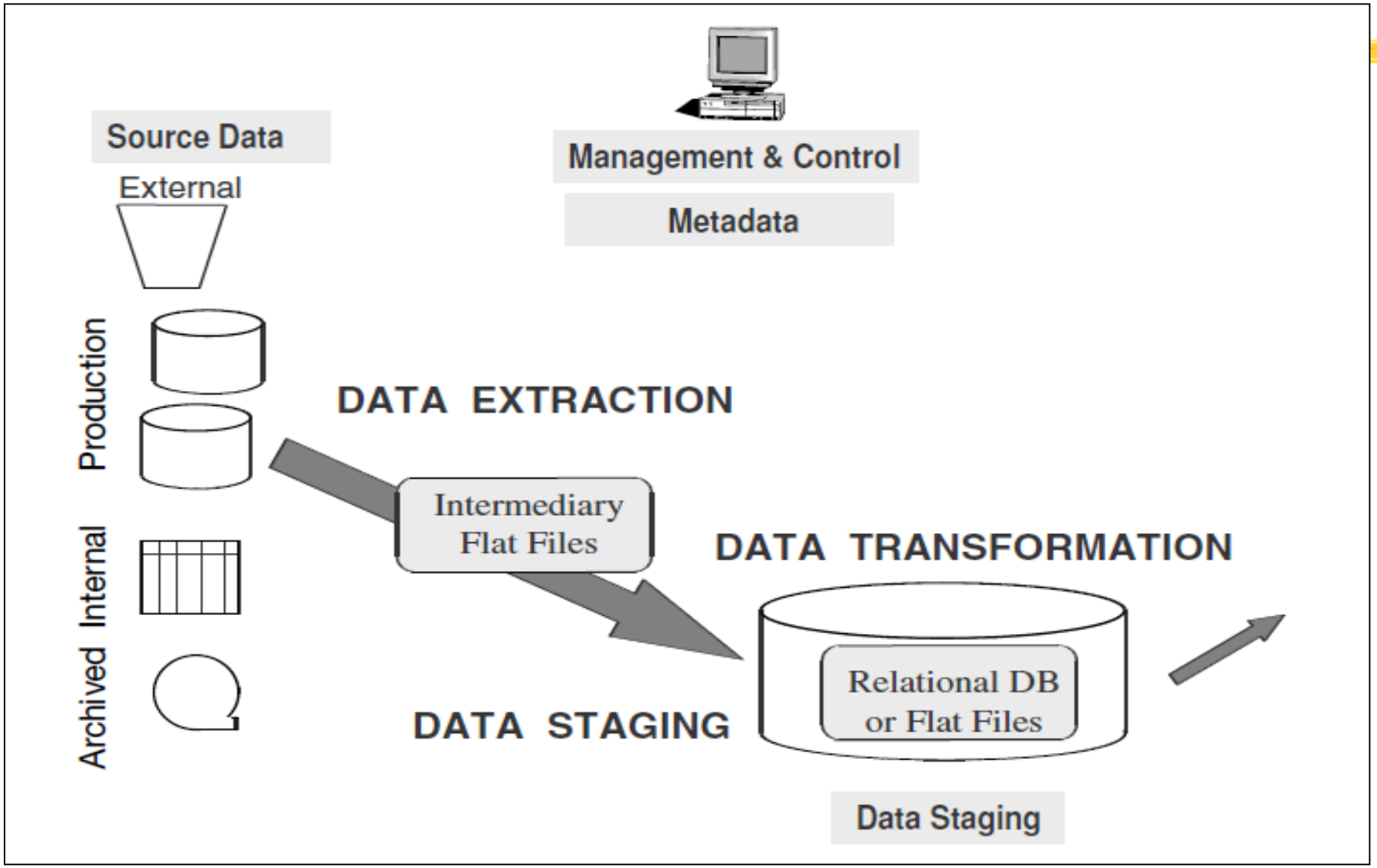
Architecture Supporting the Flow of data



Architecture of data warehouse



1. Data acquisition



Functions & Services of first stage:

Data Extraction

- Select data sources, determine filters
- Automatic replicate
- Create intermediate files

Data Transformation

- Clean, merge, de-duplicate data
- Covert data types
- Calculate derived data
- Check for referential integrity

Data Staging (Data Extraction, transformation , Loading)

- Provide backup
- Create primary & foreign keys for load tables

Data Movement to the data Warehouse

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

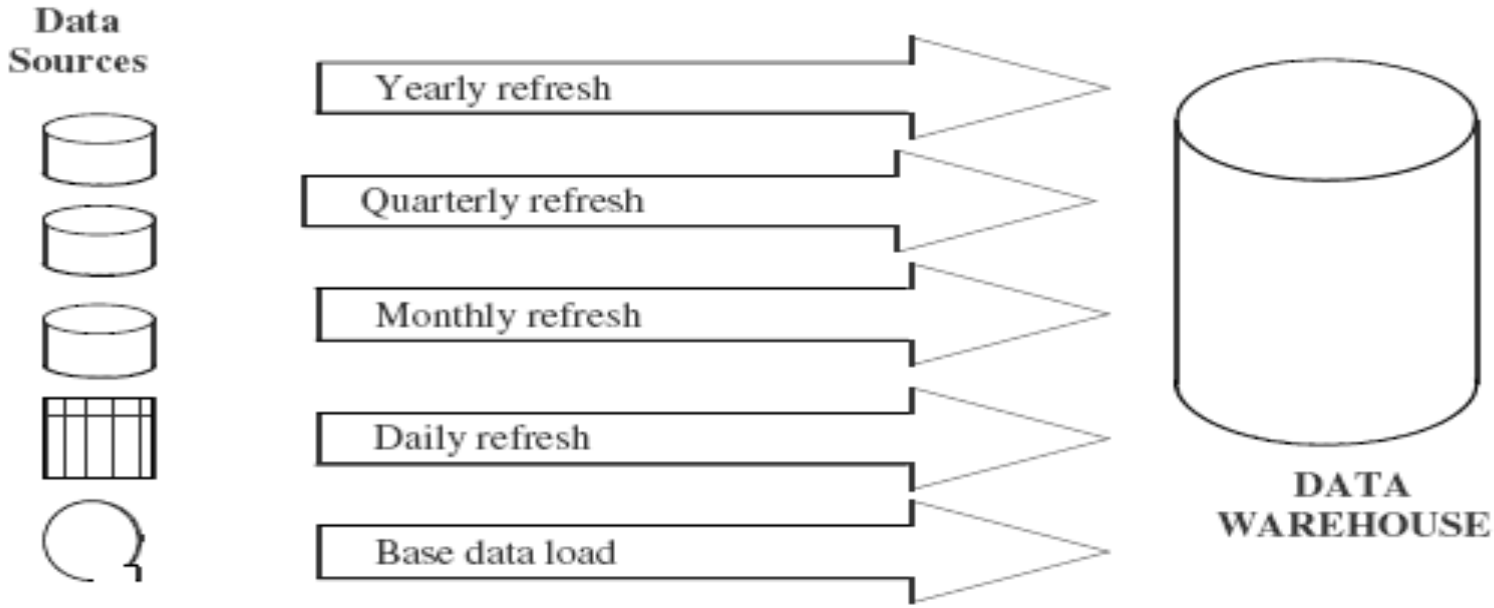
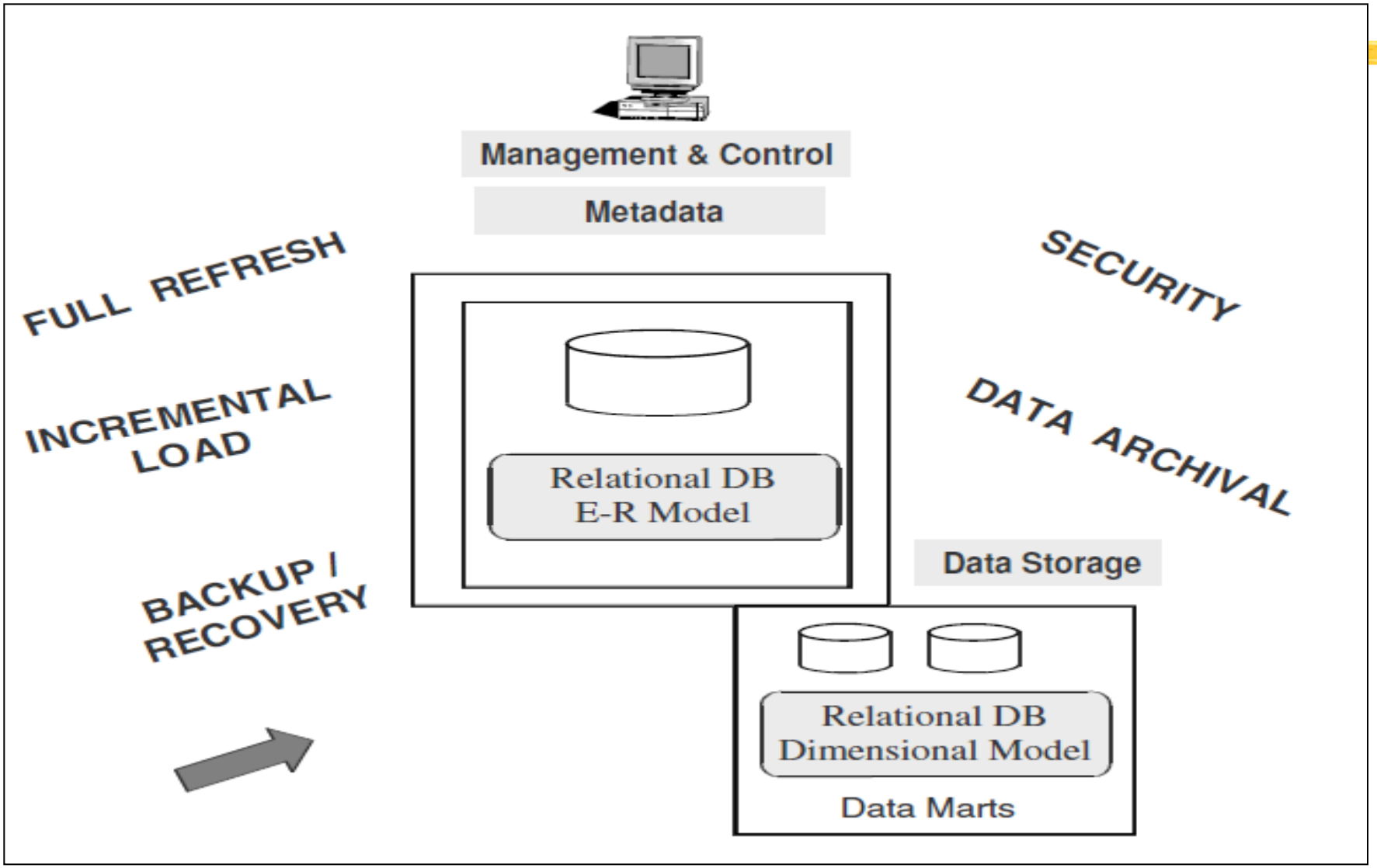


Figure 2-7 Data movements to the data warehouse.

2. Data Storage



Functions & Services

- **Initial loading of data:** loading the data from the staging area into the data warehouse repository. Before loading data into the data ware the metadata repository gets populated
- **Incremental load at regular intervals**
- **Security**
- **Backup and recovery**
- **Monitor the data warehouse**

3. Information Delivery Component

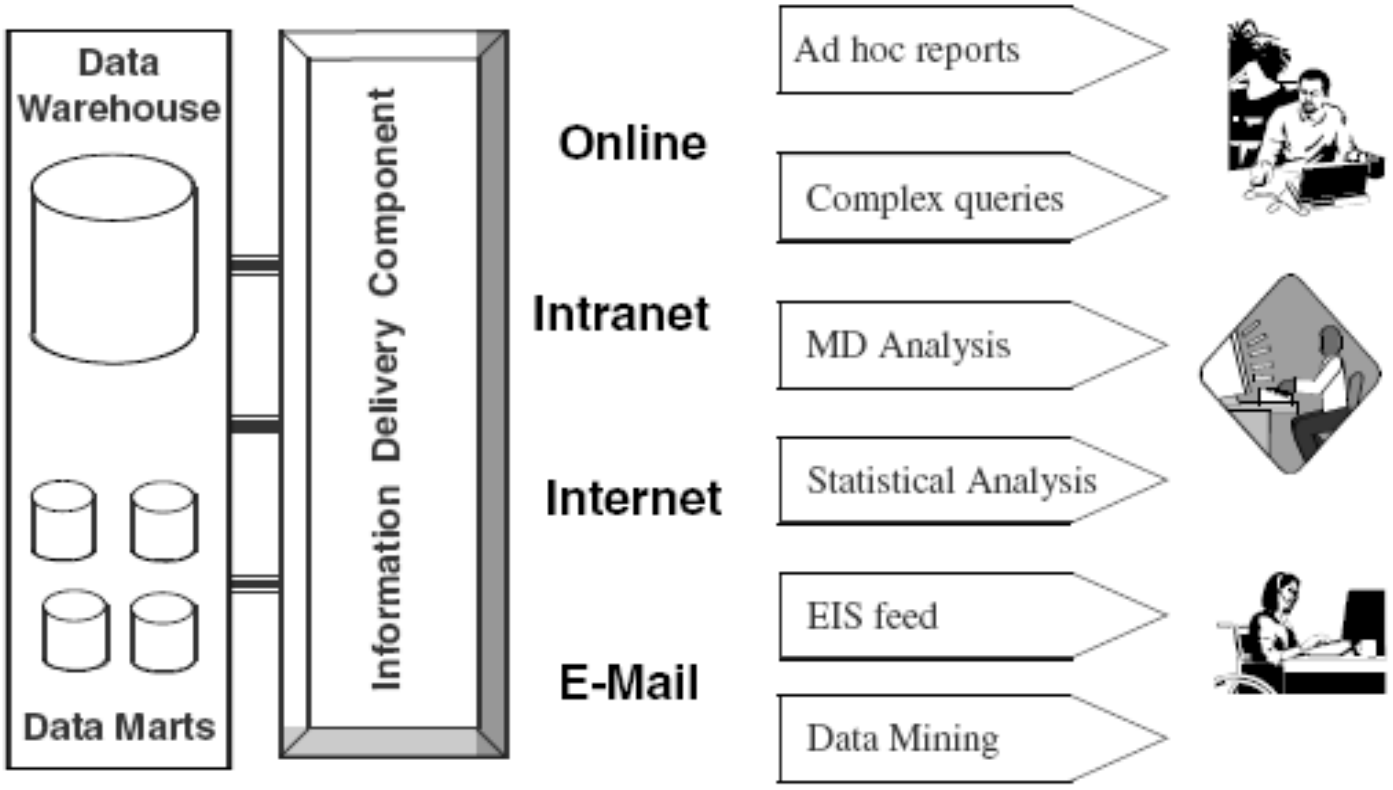
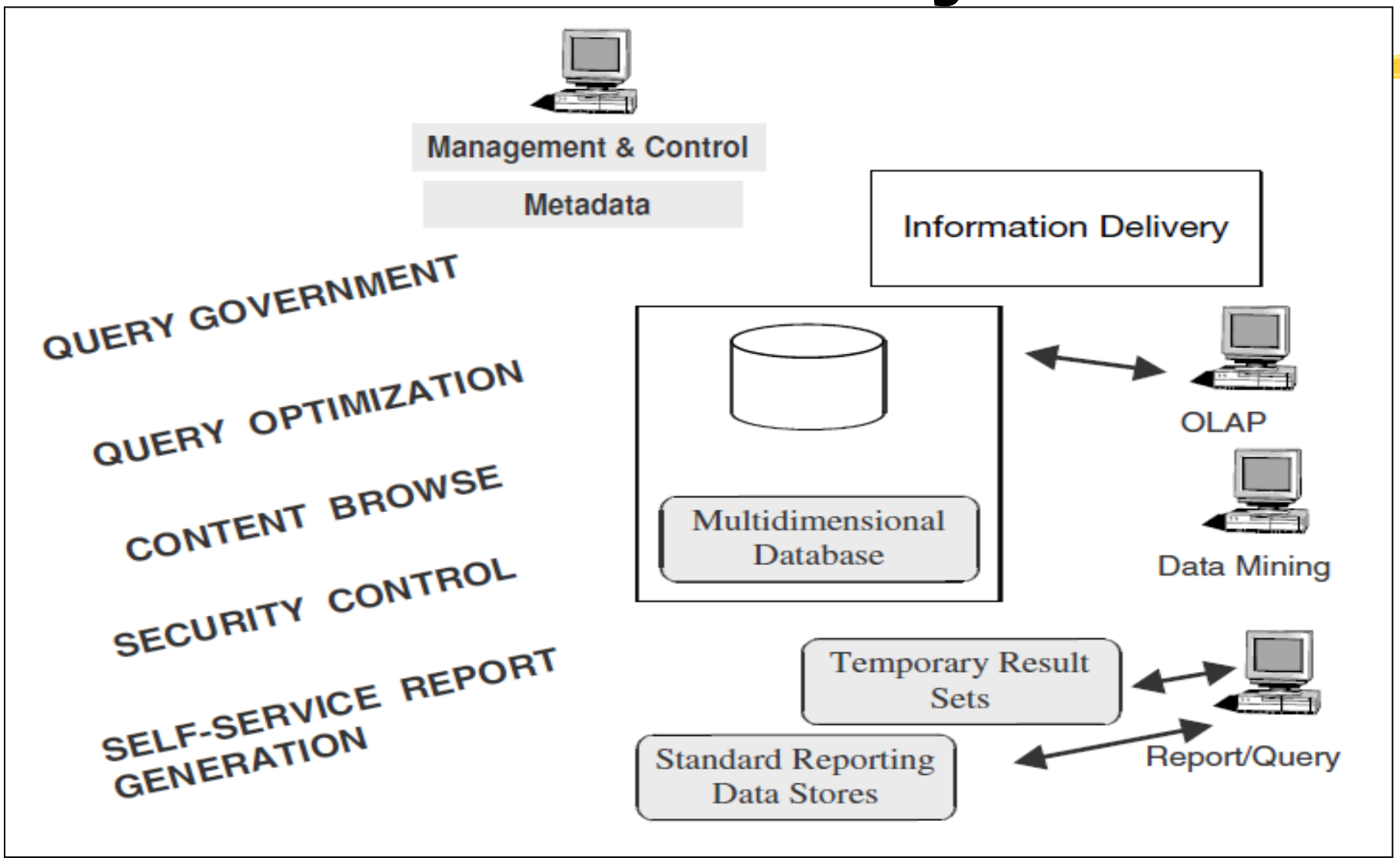


Figure 2-8 Information delivery component.

3. Information Delivery

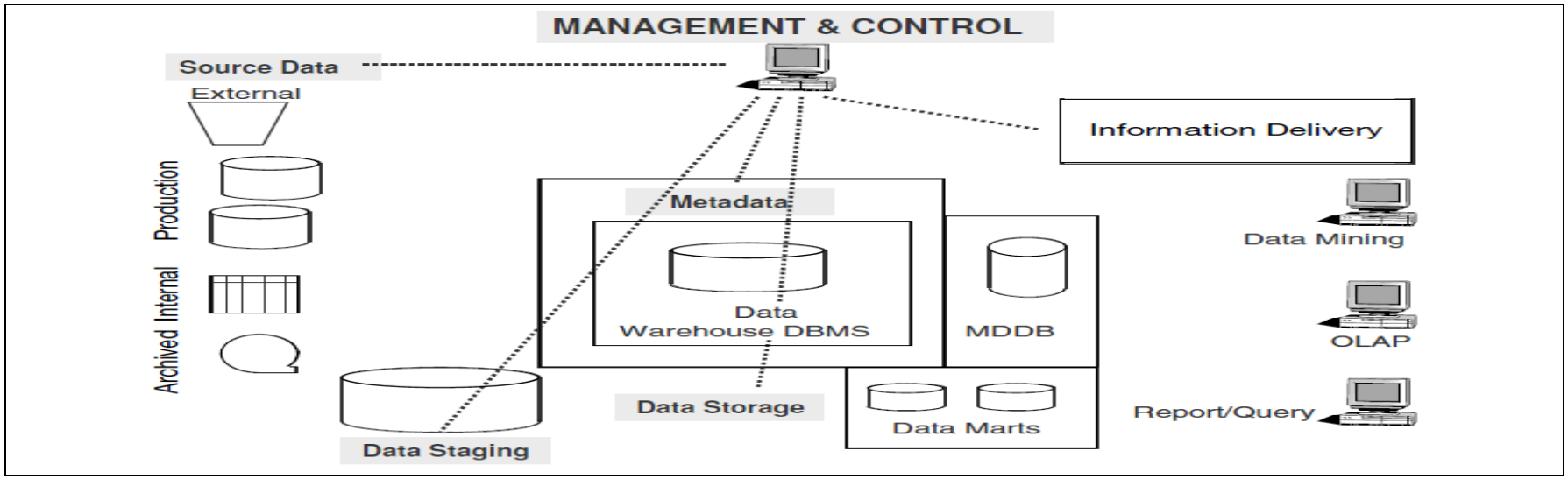


Functions & Services

- **Provide security**: to control information access and monitor user access
- **Monitor User Access**: Allow users to browse data warehouse content by hiding internal complexities
- **Automatically reformat queries:** for optimal execution, from aggregate tables as well
- **Govern Queries:** Provide self-service report generation for users, consisting of a variety of flexible options to create, schedule, and run reports
- **Store result sets of queries and reports** for future use
- **Provide multiple levels of data granularity**
- **Provide event triggers** to monitor data loading
- Make provision for the users to perform complex analysis through OLAP

4. Management and control module

- Umbrella component having two important functions
  - Monitor all ongoing operations
  - Problem recovery



Data Warehouse Infrastructure

Infrastructure

- Data Warehouse Infrastructure basically supports a data warehousing environment with the help of a combination of technologies.
- Elements that enable the architecture to be implemented.
  - Operational – help to keep the DW going
    - People
    - Procedures
    - Training
    - Management software
  - Physical
    - Hardware components
    - Operating system
    - Network, network software

Infrastructure

- **Various Factors considered for building Data Warehouse Infrastructure (contd..)**

1. Back room Infrastructure factors
2. Consideration for Hardware and Operating system Platform
3. Consideration for Database Platform
4. Front Room Infrastructure factors
5. Connectivity and Networking factors

1. Back room Infrastructure factors:

- **Data Warehouse Data Size-**
  - ✓ Grows fast in terms of size
  - ✓ Frequent additions of new dimensions, attributes and measures
  - ✓ Volume & frequency of increment of data determines the processing speed & memory of the HW platform
- **Number of Users of Data Warehouse-**
  - ✓ No. of users are essentially no. of concurrent logins which are on a data warehouse platforms.
  - ✓ There is no fixed formulae for calculating the no. of users for the purpose of estimating the infrastructure needed.

2. Features of Hardware & OS

- Hardware
  - Scalability
  - Vendor support
  - Vendor stability
- Operating System
  - Scalability
  - Security
  - Reliability
  - Availability
  - Preemptive multitasking
  - Memory protection

Hardware & Operating System Platform:

- Mainframes
  - Old hardware
  - Designed for OLTP
  - Expensive
  - Not easily scalable
- Open System Servers
  - UNIX servers are most opted
  - Robust
  - Adapted for parallel processing
- NT Servers
  - Medium-sized data warehouses
  - Limited parallel processing
  - Cost effective for small or medium DW

3. Consideration for Database Platform:

- Both Online Transaction Processing and Decision Support Systems need a computing Database platform.
- Some Data Warehouse are implemented in mainframe-based database products.
- Others are implemented using specialized multi dimensional database products called MOLAP engines.
- One of the major Considerations lies between Relational and Multi Dimensional Database.

4. Front Room Infrastructure factors:

- Front room Infrastructure factors are business and tool dependent.
- Following are the factors that affect the front room.
  - Application Server Consideration.
  - Desktop Consideration.

5. Connectivity and Networking Factors:

- This provides the link between the back room and front room.
- Connectivity Issues:
  - Bandwidth
  - Remote Access
  - Gateways
  - File Transfer
  - Database Connectivity
  - Directory Services.

Data Warehouse Meta Data

1. **Explain the role of Meta Data in Data Warehouse? Illustrate with an example.** (10 marks) (Dec 2012)
2. **What is meant by Meta Data? Explain types of Meta Data stored in Data Warehouse. Illustrate with simple customer sales Data Warehouse?** (10 marks) (May 2010,Dec 2010, Dec 2011, May 2012)
3. **What is Metadata? Why do we need metadata when search engines like Google seem so effective?** (05 marks, May 2019)

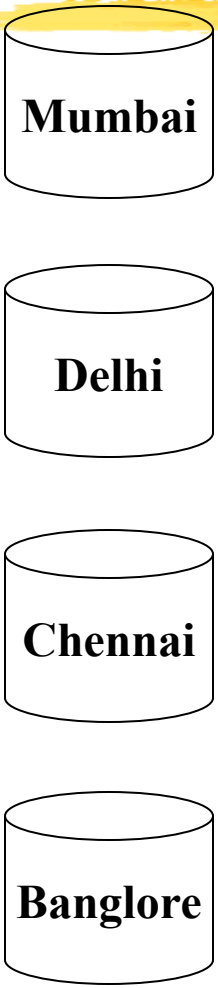
Scenario

Tanishq Jewelry is a branded Jewelry shop with branches at Mumbai, Delhi, Chennai and Bangalore.

The Executive Manager of all these branches want to find highest sale done by a Sales Person in the month of December for the year 2013.

Each branch has a separate operational system.

Scenario 1 : Tanishq Jewelry

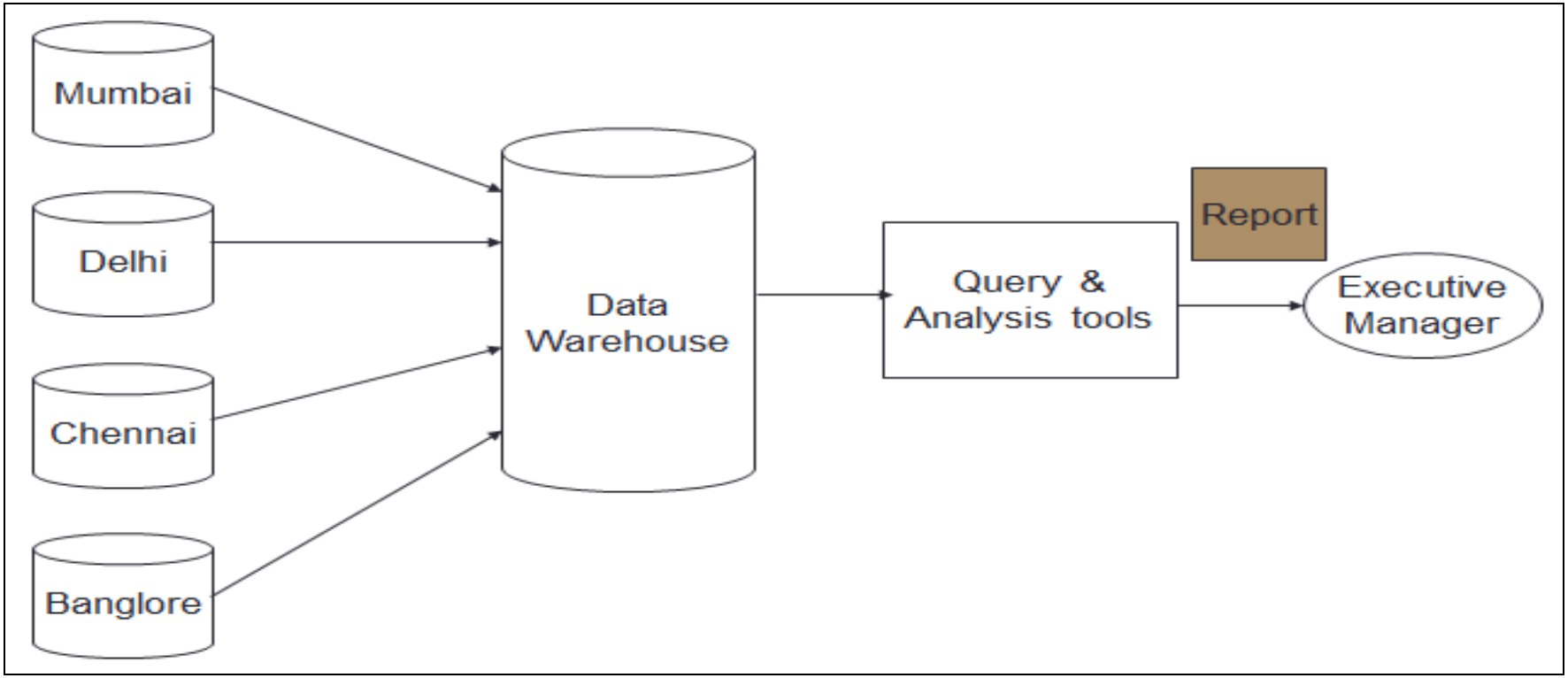


Highest Sale by a sales person per branch for December 2013

Executive Manager

Solution 1: Tanishq Jewelry

- Extract sales information from each database.
- Store the information in a common repository at a single site.



Role of METADATA in Data Warehouse

Users to compose and run the query can have several important questions:

- Are there any predefined queries I can look at?
- What are the various elements of data in the warehouse?
- Is there information about unit sales and unit costs by product?
- How can I browse and see what is available?
- From where did they get the data for the warehouse? From which source systems?
- How did they merge the data from the telephone orders system and the mail orders system?
- How old is the data in the warehouse?
- When was the last time fresh data was brought in?
- Are there any summaries by month and product?

Metadata in a data warehouse contains the answers to questions about the data in the data warehouse.

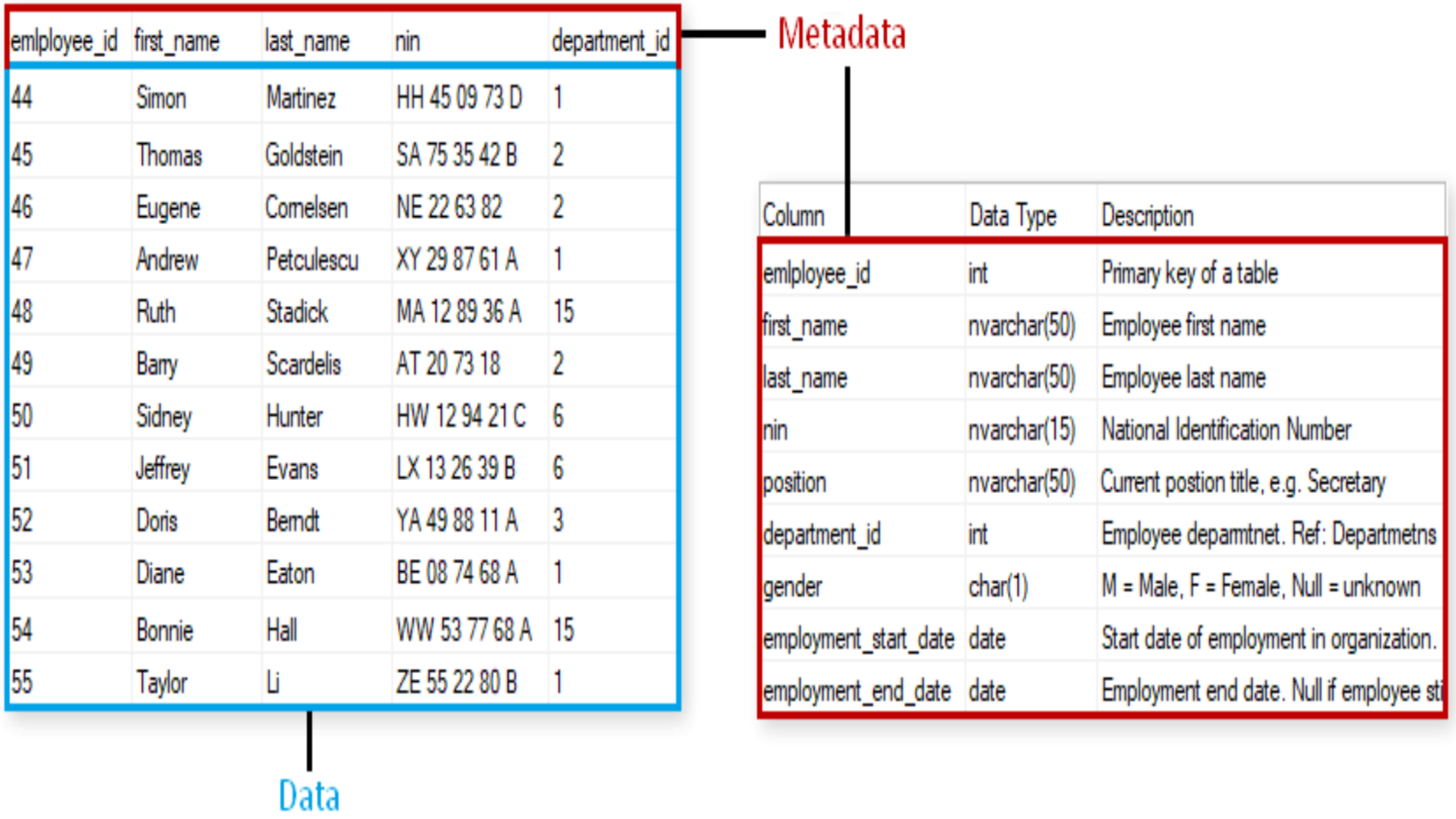
Different definitions for metadata

- Data about the data
- Table of contents for the data
- Catalogue for the data
- Data warehouse roadmap
- Data warehouse directory
- Glue that holds the data warehouse contents together

Metadata

- **Data Warehouse metadata** are pieces of information stored in one or more special-purpose metadata repositories that includes-
  - ✓ *Information on the contents* of the data warehouse, their location and their structure.
  - ✓ *Information on the process* that takes place in the data warehouse backstage, concerning the refreshment of the warehouse with clean, up-to-date, semantically and structurally reconciled data.
  - ✓ *Information on the infrastructure* and *physical characteristics* of components and the sources of the data warehouse

Meta Data Example:



emlployee_id	first_name	last_name	nin	department_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Comelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Berndt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1
54	Bonnie	Hall	WW 53 77 68 A	15
55	Taylor	Li	ZE 55 22 80 B	1

Column	Data Type	Description
emlployee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current postion title, e.g. Secretary
department_id	int	Employee depamntnet. Ref: Departmetns
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date. Null if employee sti

Metadata in OLTP

- In operational systems we do not really have any easy and flexible methods for knowing the nature of the contents of the database.
- There is no great need for user-friendly interfaces to the database contents.

Metadata in DWH

- Users need sophisticated methods for browsing and examining the contents of the data warehouse.
- Users need to know the meanings of the data items.
- **Users have to prevent them from drawing wrong conclusions from their analysis through their ignorance about the exact meanings.**
- *Without adequate metadata support, users of the larger data warehouses are totally handicapped.*

Ways to classify Metadata

- Administrative/End-user/Optimization
- Development/Usage
- In the data mart/At the workstation
- **Building/Maintaining/Managing/Usage Technical/ Business**
- Back room/Front room
- Internal/External

Types of Metadata

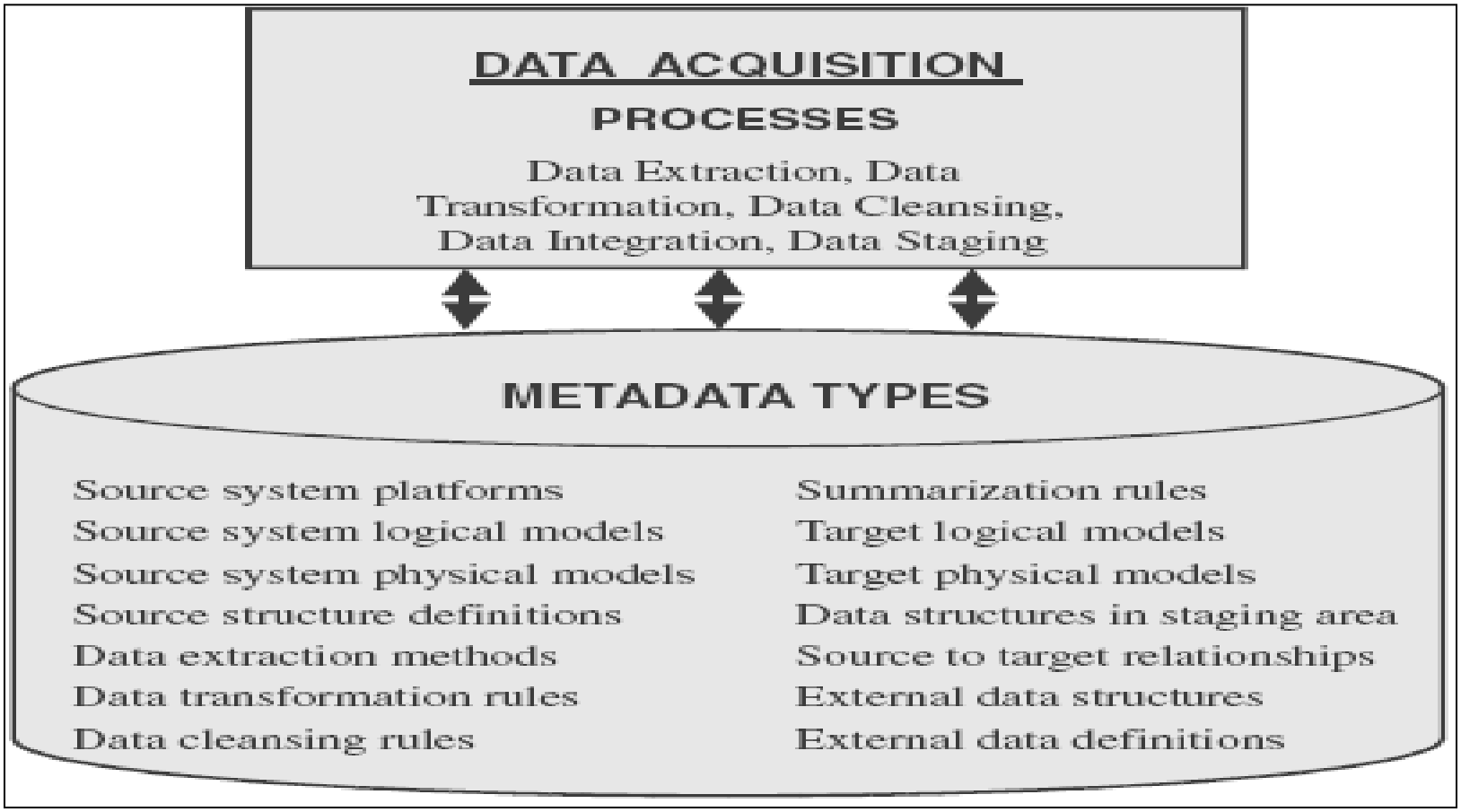
Metadata in a data warehouse fall into three major categories based on *Building/ Maintaining/ Managing/ Usage* :

1. Operational Metadata
2. Extraction and Transformation Metadata
3. End-User Metadata

1. Operational Metadata

- Data for the data warehouse comes from several operational systems of the enterprise.
- These source systems contain different data structures.
- The data elements selected for the data warehouse have various field lengths and data types.
- In selecting data from the source systems for the data warehouse, we
  - split records,
  - combine parts of records from different source files, and
  - deal with multiple coding schemes and field lengths.
- When you deliver information to the end-users, you must be able to tie that back to the original source data sets.
- *Operational metadata contain all of this information about the operational data sources.*

Data Acquisition



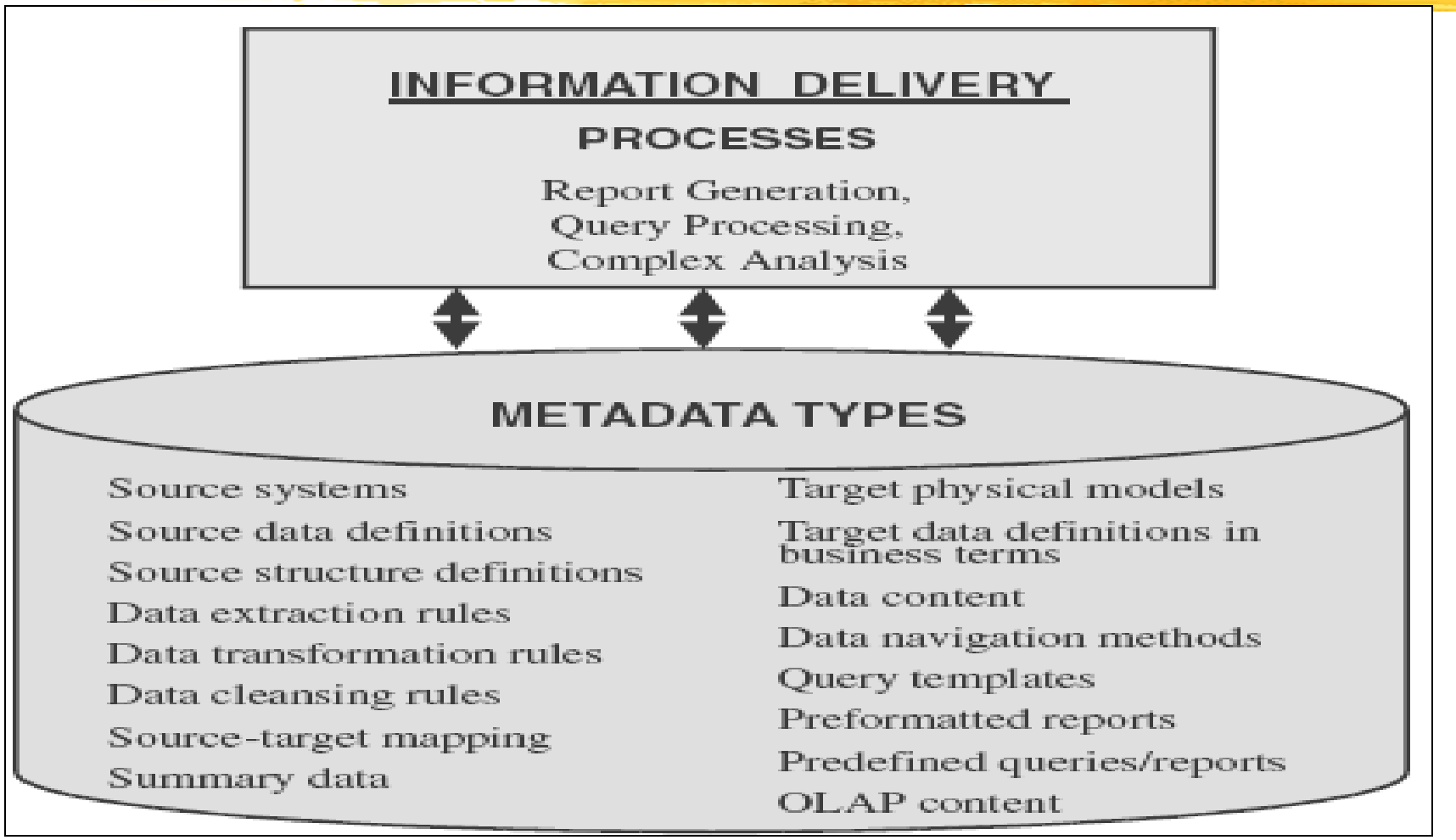
2. Extraction and Transformation Metadata

- Contains data about the extraction of data from the source systems
  - **extraction frequencies**
  - **extraction methods**
  - **business rules for the data extraction**
- *Meta data keeps here- Information about all the data transformations and extraction that take place in the data staging area.*

3. End-User Metadata

- Navigational map of the data warehouse
- Enables the end-users to find information from the data warehouse.
- Allows the end-users to use their own business terminology and look for information in those ways in which they normally think of the business.

Information Delivery



Types of Metadata

1. **Business metadata**
   - Portrays DW from the end user perspective
   - Shows business names, not actual file names
   - Less structured as compared to technical metadata
   - Used by business analysts and other end users.
2. **Technical metadata**
   - Shows the actual structure and content of the DW
   - Acts as a guide to build, maintain and administer the DW
   - Used the data warehouse administrator, and other IT staff working on the DW.

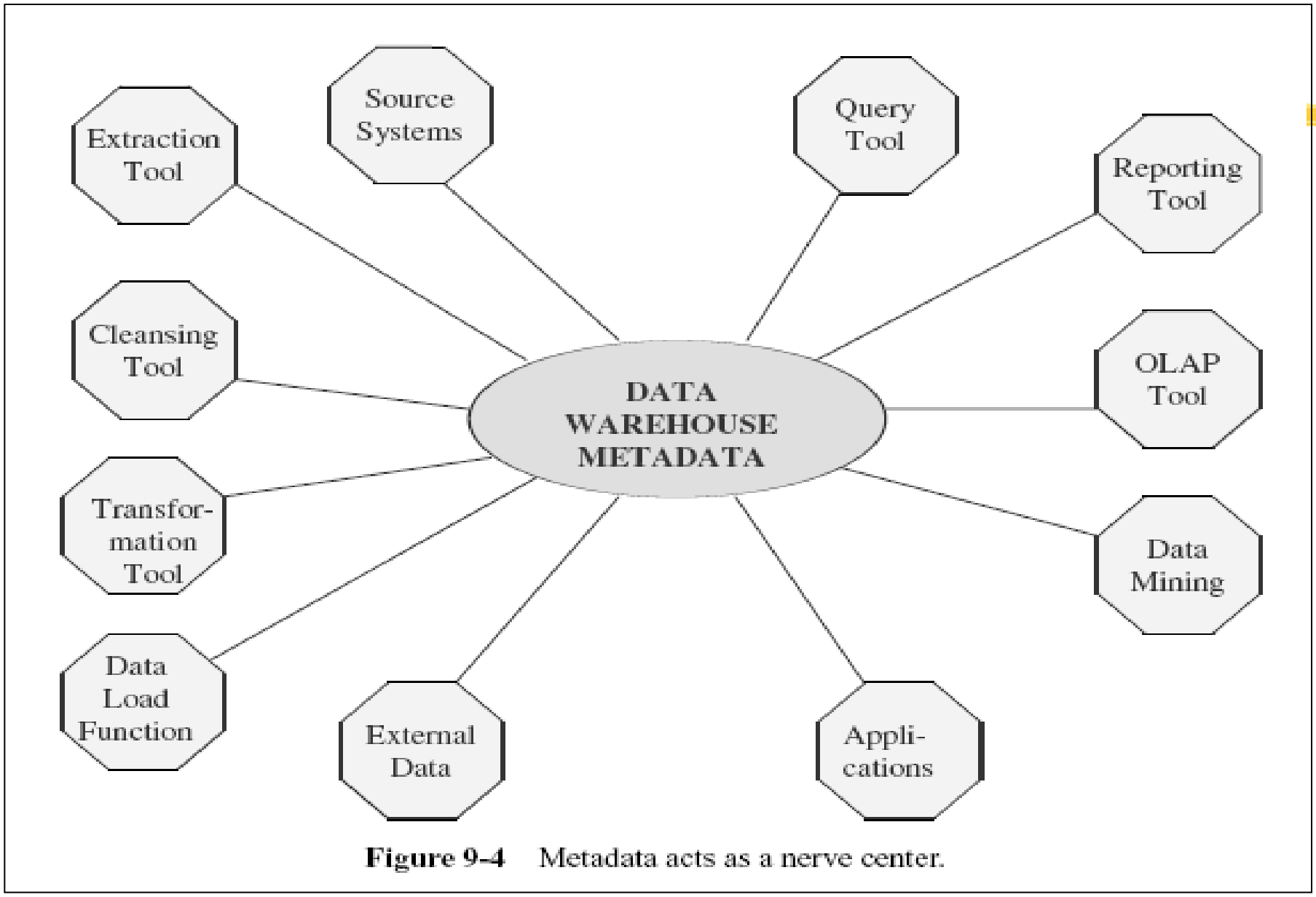


Figure 9-4 Metadata acts as a nerve center.

Illustrate Metadata with simple customer sales Data Warehouse: (May 2010,Dec 2010, Dec 2011, May 2012)

- Entity Name:
- Alias Names:
- Definition:
- Remarks:
- Source Systems:
- Create Date:
- Last Update Date:
- Update Cycle:
- Last Full Refresh Date :
- Full Refresh Cycle:
- Data Quality Reviewed Date:
- Last DE duplication Date:
- Planned Archival:
- Responsible User:

Illustrate Metadata with simple customer sales Data Warehouse:

(May 2010,Dec 2010, Dec 2011, May 2012)

- **Entity Name:** Customer
- **Alias Names:** Account, Client
- **Definition:** A person or an organization that purchase goods or services
- **Remarks:** Customer entity includes regular, current and past customers
- **Source Systems:** Finished Goods orders, Maintenance Contracts, Online Sales
- **Create Date:** June 15, 2008
- **Last Update Date:** December 31, 2013
- **Update Cycle:** Weekly
- **Last Full Refresh Date :** January 15,2014
- **Full Refresh Cycle:** Every Six Months
- **Data Quality Reviewed:** July 15, 2013
- **Last DE duplication:** June 15, 2013
- **Planned Archival:** Every Six Months
- **Responsible User:** Jenny Jose

Why do we need metadata when search engines like Google seem so effective : (May 2019)

- **Search engines can** crawl a website and guess its general purpose based on these elements;
- **Metadata** enables webmasters to tell **search engines** what a page's title **is**, which says a lot about what **search** queries it may be relevant for.
- **Metadata** is a way to tell **search engine** crawlers what to expect from a page, blog, image, or paragraph, providing the **search engine** advice on how to best index the site.

University Exam Questions of Chapter 1

1. Define Data warehouse. Explain, what is the need for developing a data Warehouse is, hence explain its architecture? 10 marks– (May 2011, May 2012)
2. Why is data integration required in a data warehouse, more so than in an operational application? 5 marks (Dec 2019)
3. Define Metadata. Discuss the types of metadata stored in a data warehouse. Illustrate with an example. (10 marks Dec 17, Dec 16)
4. Write short note on Role of Meta data (05 marks Dec 2012,May 18, May 17)
5. What is meant by Meta Data? Explain types of Meta Data stored in Data Warehouse. Illustrate with simple customer sales Data Warehouse? (10 marks) (May 2010,Dec 2010, Dec 2011, May 2012)
6. What is Metadata? Why do we need metadata when search engines like Google seem so effective? (05 marks, May 2019)
7. Illustrate the architecture of Data Warehouse system. Differentiate Data Warehouse and Data Mart. (10 marks May 17, May 16)
8. Explain Data warehouse Architecture in detail. (10 marks May 2010,Dec 2010,May 2011, May 2012, May 18)
9. Differentiate between top down & bottom up approaches for building a data warehouse. Discuss the merits and limitations of each approach? (10 marks Dec 17)
10. What are the basic Building Blocks of Data Warehouse? 10 marks
11. Differentiate. (10 marks, Dec 16)
    1. OLTP Vs. OLAP
    2. Data Warehouse Vs. Data Mart

THANK YOU