An Introduction to Data Warehousing

Last Updated : May 26, 2003

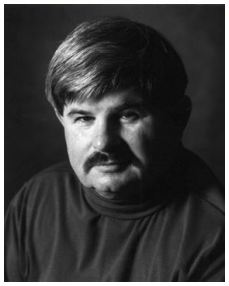
Center of Excellence
Data Warehousing

Wipro Technologies
Innovative Solutions. Quality Leadership.

Objectives

- At the end of this lesson, you will know :
  - What is Data Warehousing
  - The evolution of Data Warehousing
  - Need for Data Warehousing
  - OLTP Vs Warehouse Applications
  - Data marts Vs Data Warehouses
  - Operational Data Stores
  - Overview of Warehouse Architecture

What is a Data Warehouse ?



Can I see credit report from Accounts, Sales from marketing and open order report from order entry for this customer

Data from multiple sources is integrated for a subject

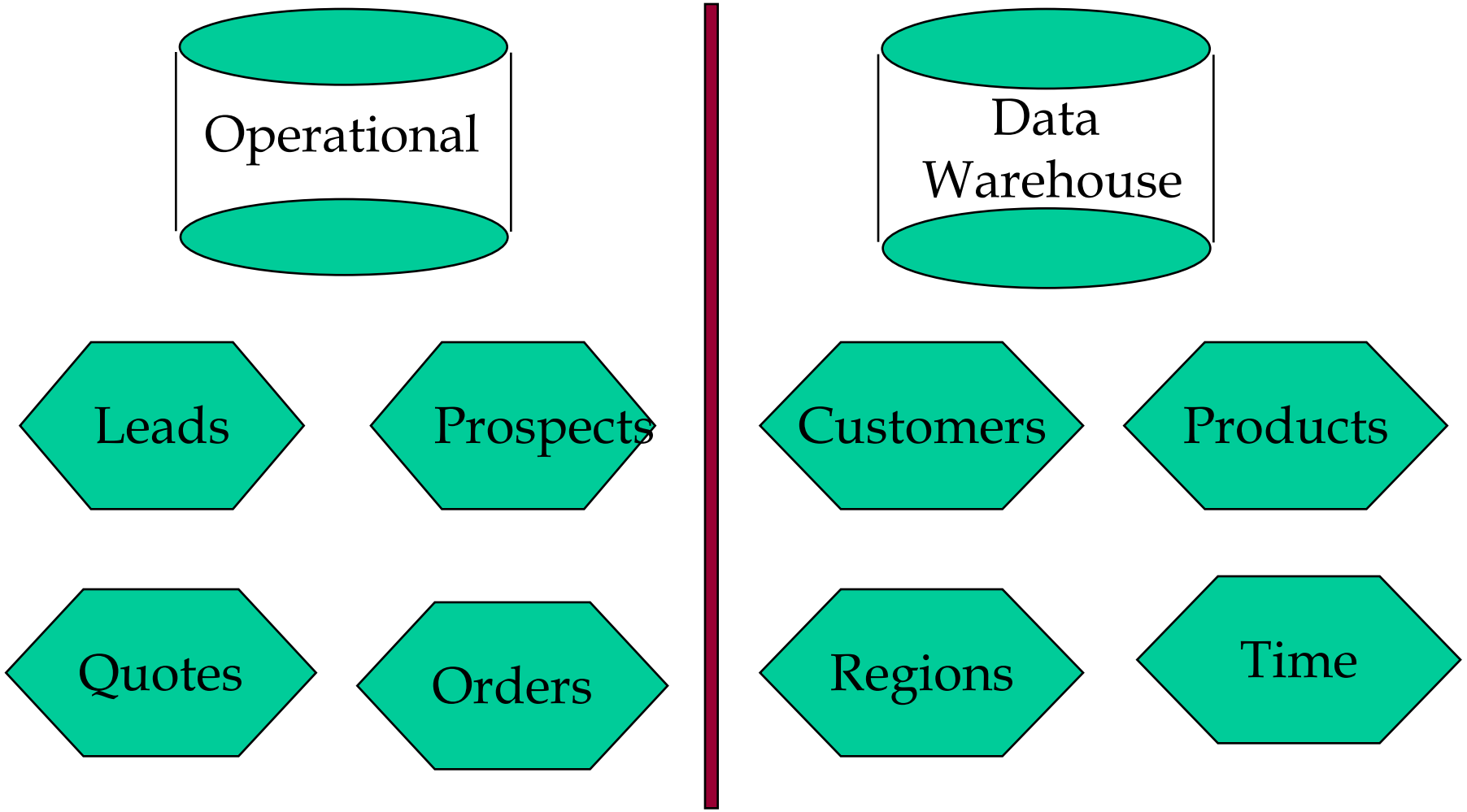
A data warehouse is a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions.

– WH Inmon

Identical queries will give same results at different times. Supports analysis requiring historical data

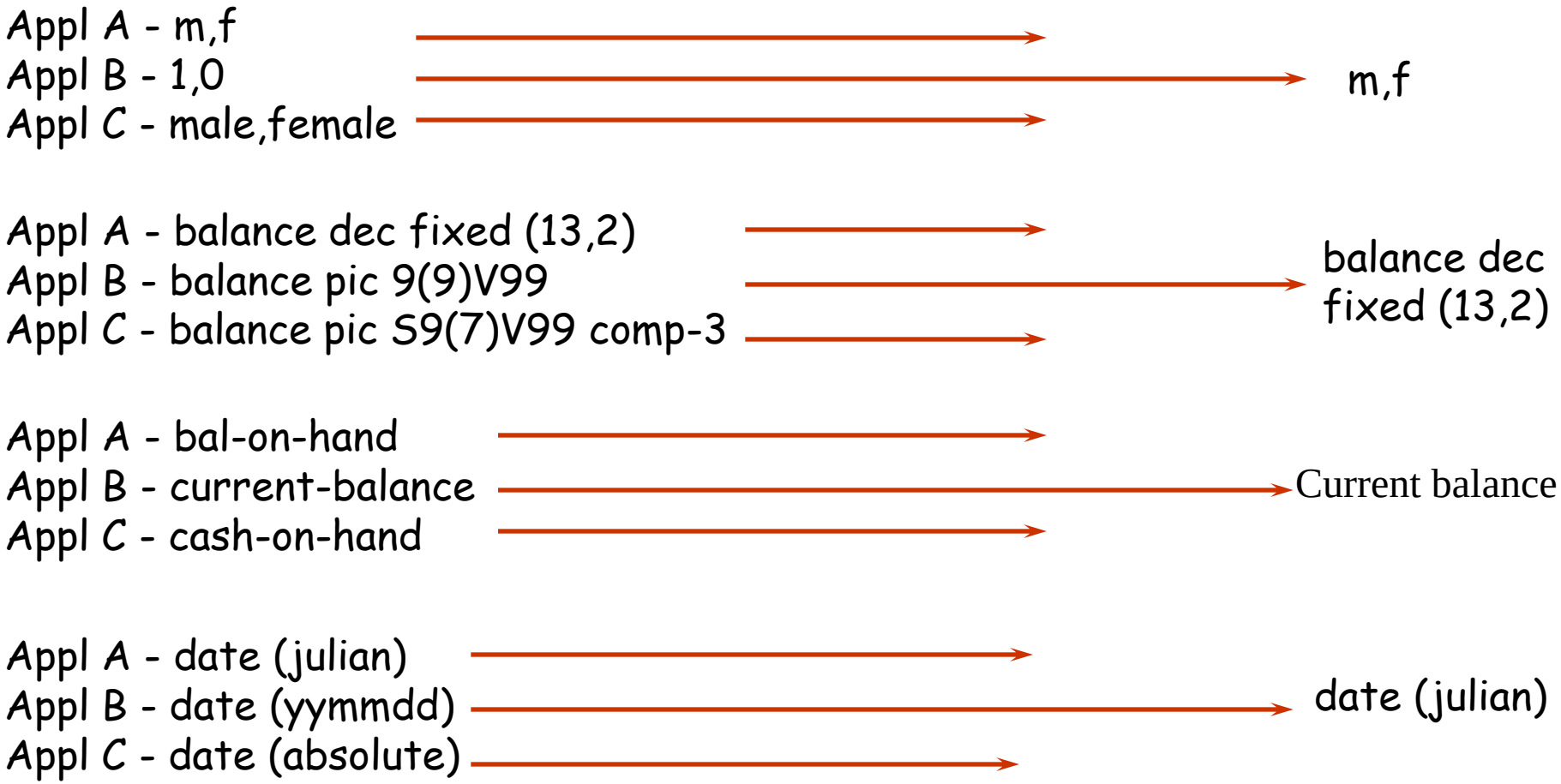
Data stored for historical period. Data is populated in the data warehouse on daily/weekly basis depending upon the requirement.

Subject-Oriented- Characteristics of a Data Warehouse

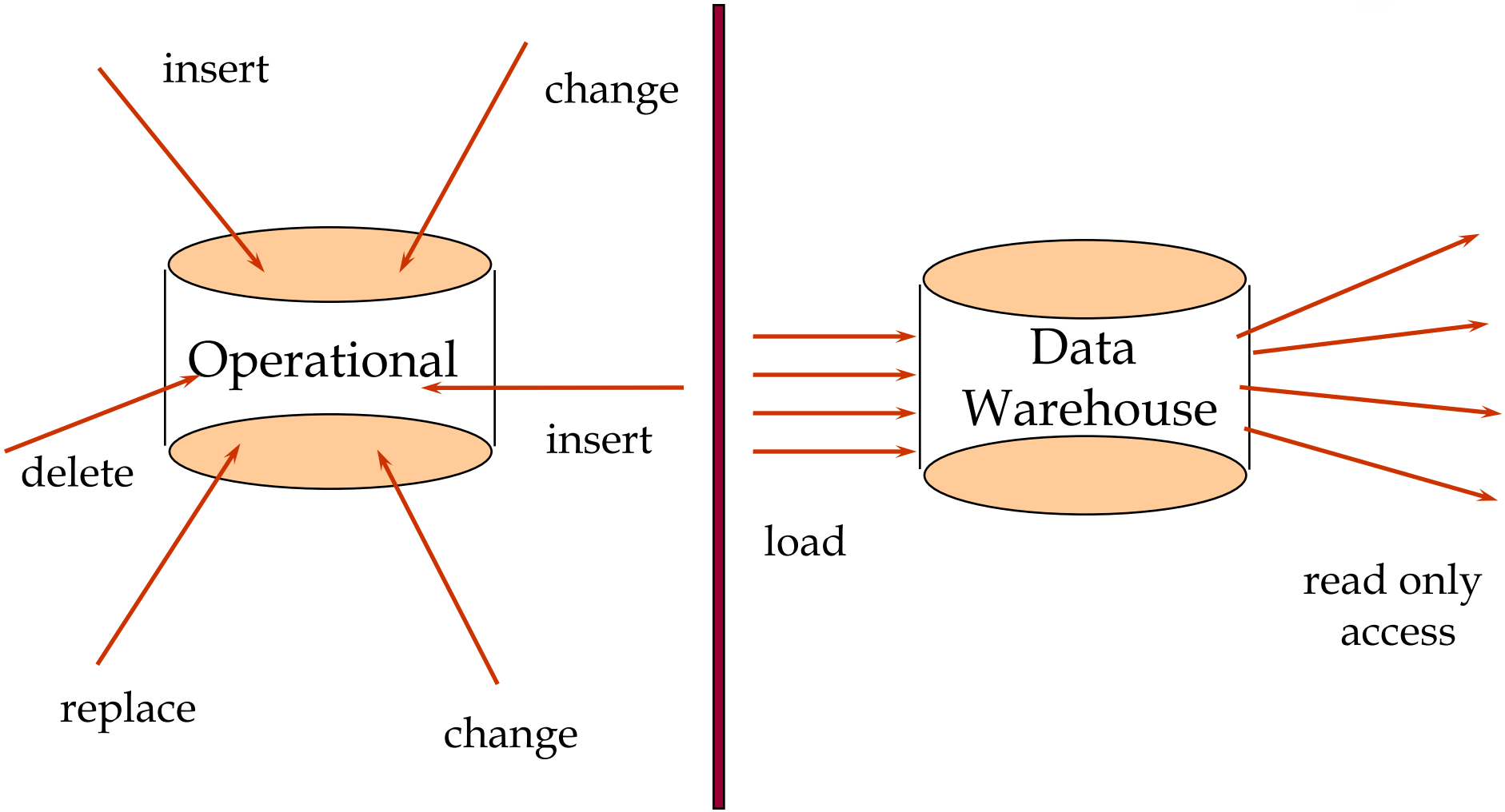


Focus is on Subject Areas rather than Applications

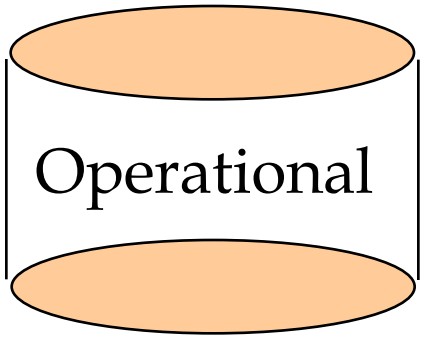
Integrated - Characteristics of a Data Warehouse



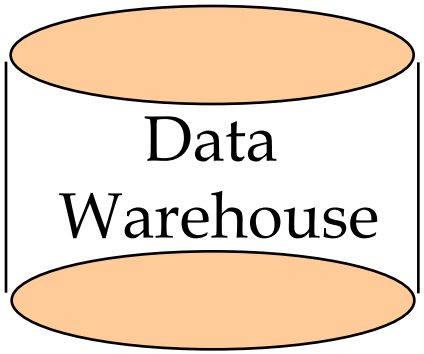
Non-volatile - Characteristics of a Data Warehouse



Time Variant - Characteristics of a Data Warehouse



Current Value data
- time horizon : 60-90 days
- key may not have element of time



Snapshot data
- time horizon : 5-10 years
- key has an element of time
- data warehouse stores historical data

Alternate Definitions

A collection of integrated, subject oriented databases designed to support the DSS function, where each unit of data is relevant to some moment of time

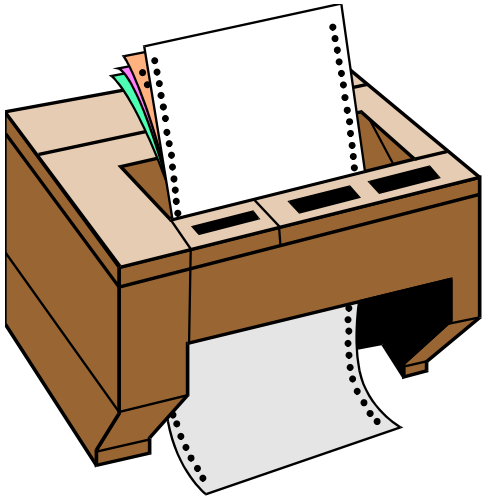
- Imhoff

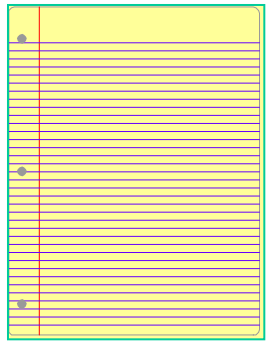
Alternate Definitions

Data Warehouse is a repository of data summarized or aggregated in simplified form from operational systems. End user orientated data access and reporting tools let user get at the data for decision support - Babcock

Evolution of Data Warehousing

1960 - 1985 : MIS Era





- Unfriendly
- Slow
- Dependent on IS programmers
- Inflexible
- Analysis limited to defined reports

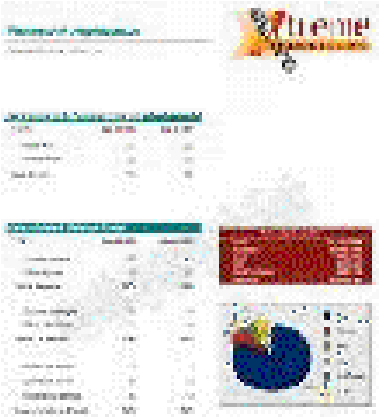
Evolution of Data Warehousing

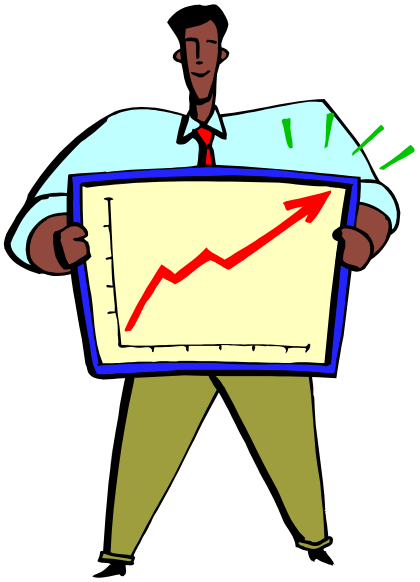
1985 - 1990 : Querying Era

- Adhoc, unstructured access to corporate data
- SQL as interface not scalable
- Cannot handle complex analysis

Evolution of Data Warehousing

1990 - 20xx : Analysis Era





- Trend Analysis
- What If ?
- Moving Averages
- Cross Dimensional Comparisons
- Statistical profiles
- Automated pattern and rule discovery

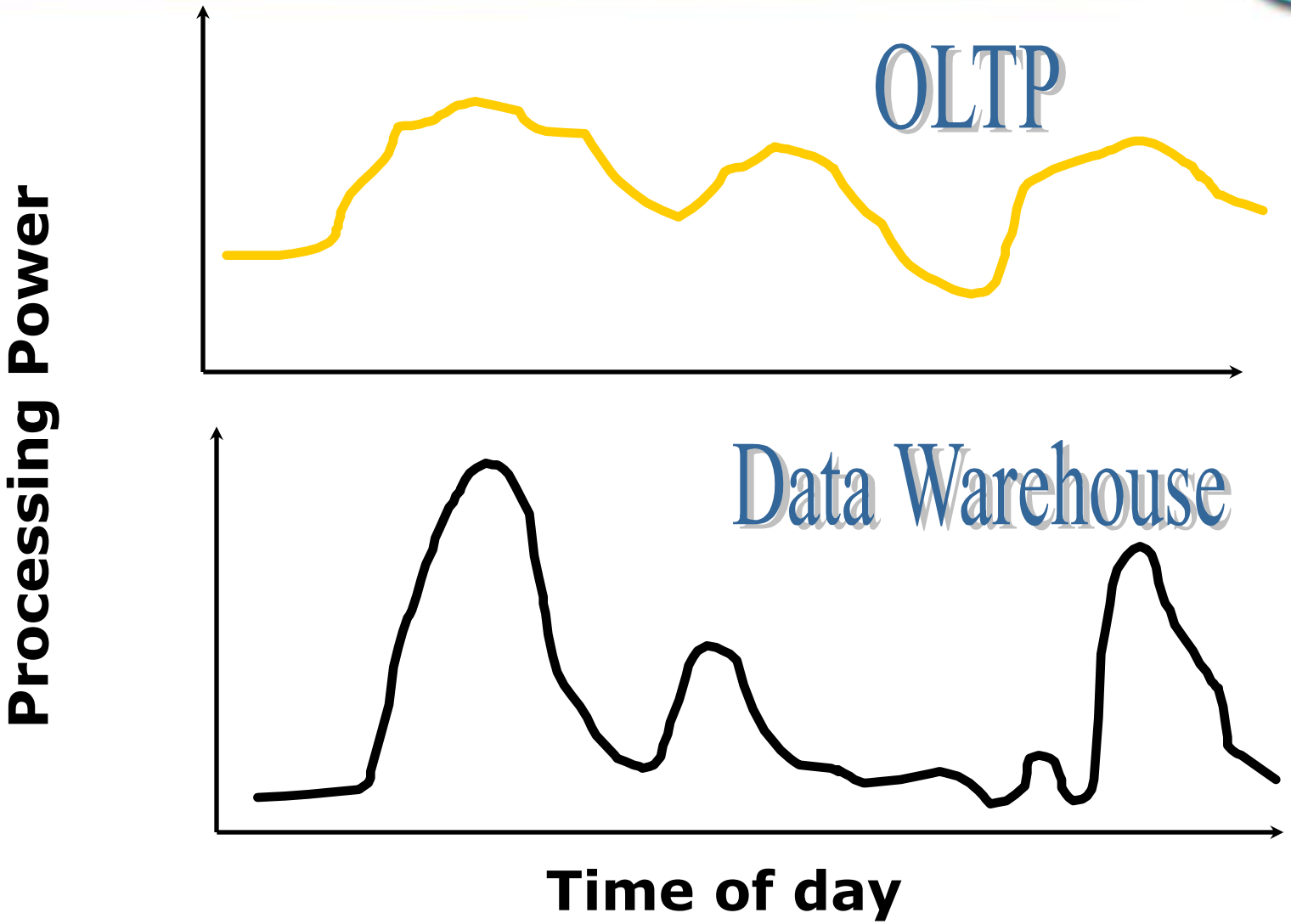
Need for Data Warehousing

- Better business intelligence for end-users
- Reduction in time to locate, access, and analyze information
- Consolidation of disparate information sources
- Strategic advantage over competitors
- Faster time-to-market for products and services
- Replacement of older, less-responsive decision support systems
- Reduction in demand on IS to generate reports

OLTP Vs Warehouse

Operational System	Data Warehouse
Transaction Processing	Query Processing
Time Sensitive	History Oriented
Operator View	Managerial View
Organized by transactions (Order, Input, Inventory)	Organized by subject (Customer, Product)
Relatively smaller database	Large database size
Many concurrent users	Relatively few concurrent users
Volatile Data	Non Volatile Data
Stores all data	Stores relevant data
Not Flexible	Flexible

Capacity Planning



Processing Load Peaks During the Beginning and End of Day

Examples Of Some Applications

Manufacturers

- Target Marketing
- Market Segmentation
- Budgeting
- Credit Rating Agencies
- Financial Reporting and Consolidation

Retailers

- Market Basket Analysis - POS Analysis
- Churn Analysis
- Profitability Management
- Event tracking

Customers

Do we need a separate database ?

- OLTP and data warehousing require two very differently configured systems
- Isolation of Production System from Business Intelligence System
- Significant and highly variable resource demands of the data warehouse
- Cost of disk space no longer a concern
- Production systems not designed for query processing

Data Marts

- Enterprise wide data warehousing projects have a very large cycle time
- Getting consensus between multiple parties may also be difficult
- Departments may not be satisfied with priority accorded to them
- Sometimes individual departmental needs may be strong enough to warrant a local implementation
- Application/database distribution is also an important factor

Data Marts

- Subject or Application Oriented Business View of Warehouse
  - Quick Solution to a specific Business Problem
  - Finance, Manufacturing, Sales etc.
  - Smaller amount of data used for Analytic Processing

Data Warehouses or Data Marts

- For companies interested in changing their corporate cultures or integrating separate departments, an enterprise wide approach makes sense.
- Companies that want a quick solution to a specific business problem are better served by a standalone data mart.
- Some companies opt to build a warehouse incrementally, data mart by data mart.

Data Warehouse and Data Mart

	Data Warehouse	Data Marts
Scope	●Application Neutral ●Centralized, Shared ●Cross LOB/enterprise	●Specific Application Requirement ●LOB, department ●Business Process Oriented
Data Perspective	●Historical Detailed data ●Some summary	●Detailed (some history) ●Summarized
Subjects	●Multiple subject areas	●Single Partial subject ●Multiple partial subjects

Data Warehouse and Data Mart

	Data Warehouse	Data Marts
Data Sources	●Many ●Operational/ External Data	●Few ●Operational, external data
Implement Time Frame	●9-18 months for first stage ●Multiple stage implementation	●4-12 months
Characteristics	●Flexible, extensible ●Durable/Strategic ●Data orientation	●Restrictive, non extensible ●Short life/tactical ●Project Orientation

Warehouse or Mart First ?

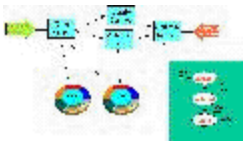
Data Warehouse First	Data Mart first
Expensive	Relatively cheap
Large development cycle	Delivered in < 6 months
Change management is difficult	Easy to manage change
Difficult to obtain continuous corporate support	Can lead to independent and incompatible marts
Technical challenges in building large databases	Cleansing, transformation, modeling techniques may be incompatible

OLTP Systems Vs Data Warehouse

Remember
Between OLTP and Data Warehouse systems

users are different

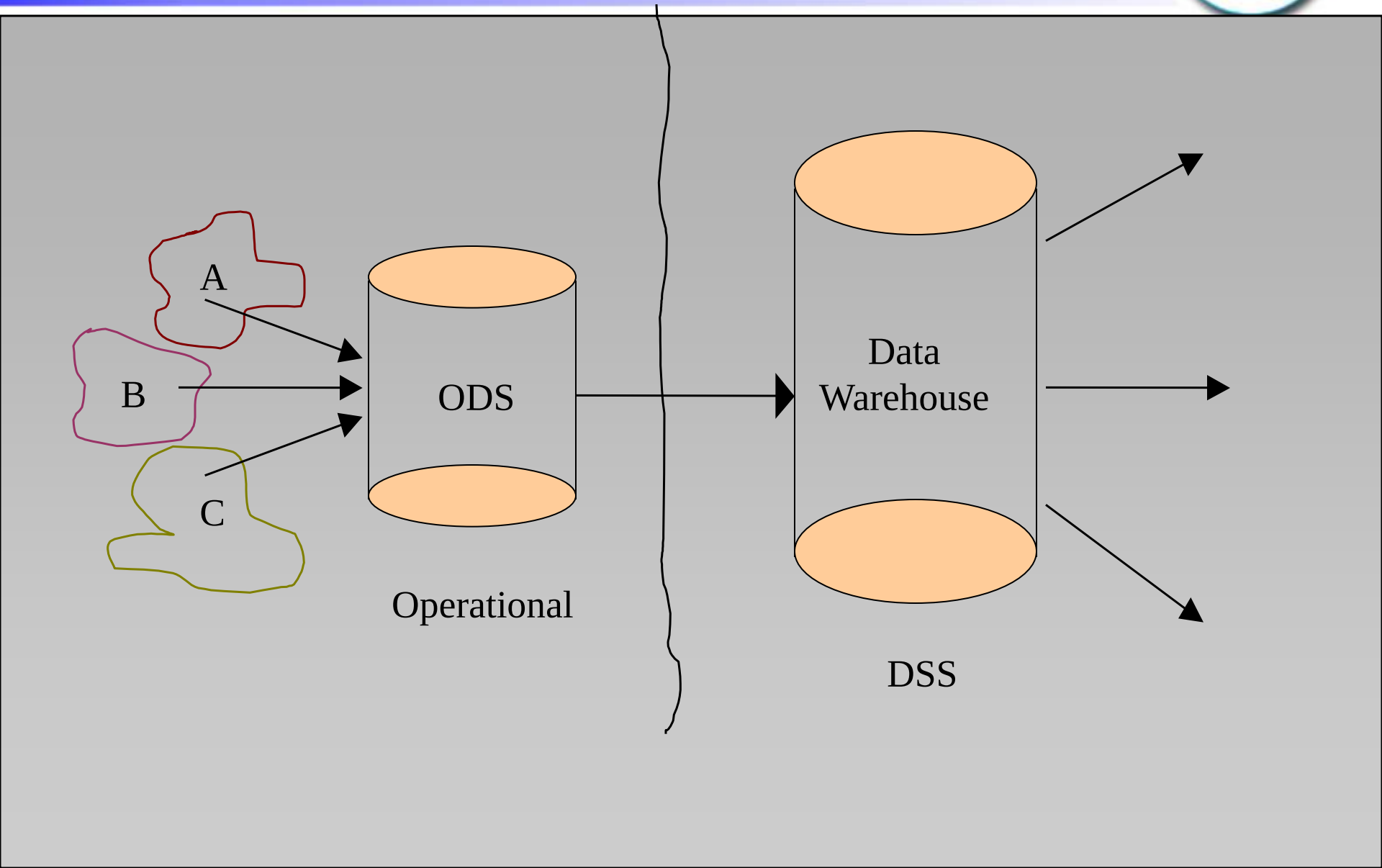
data content is different,



data structures are different

hardware is different

Operational Data Store - Definition



Operational Data Store - Definition

Can I see credit report from Accounts, Sales from marketing and open order report from order entry for this customer

Data from multiple sources is integrated for a subject

A subject oriented, integrated, volatile, current valued data store containing only corporate detailed data

Identical queries may give different results at different times. Supports analysis requiring current data

Data stored only for current period. Old Data is either archived or moved to Data Warehouse

Operational Data Store

- The ODS applies only to the world of operational systems.
- It is Point of Integration for operational systems.
- The ODS contains current valued and near current valued data.
- The ODS contains almost exclusively all detail data
- The ODS requires a full function, update, record oriented environment.
- ODS is used for low-level decision support.

Operational Data Store

- Functions of an ODS
- Converts Data,
- Decides Which Data of Multiple Sources Is the Best,
- Summarizes Data,
- Decodes/encodes Data,
- Alters the Key Structures,
- Alters the Physical Structures,
- Reformats Data,
- Internally Represents Data,
- Recalculates Data.

Different kinds of Information Needs

- Current
- Recent
- Historical

Is this medicine available in stock

OLTP

What are the tests this patient has completed so far

ODS

Has the incidence of Tuberculosis increased in last 5 years in Southern region

Data Warehouse

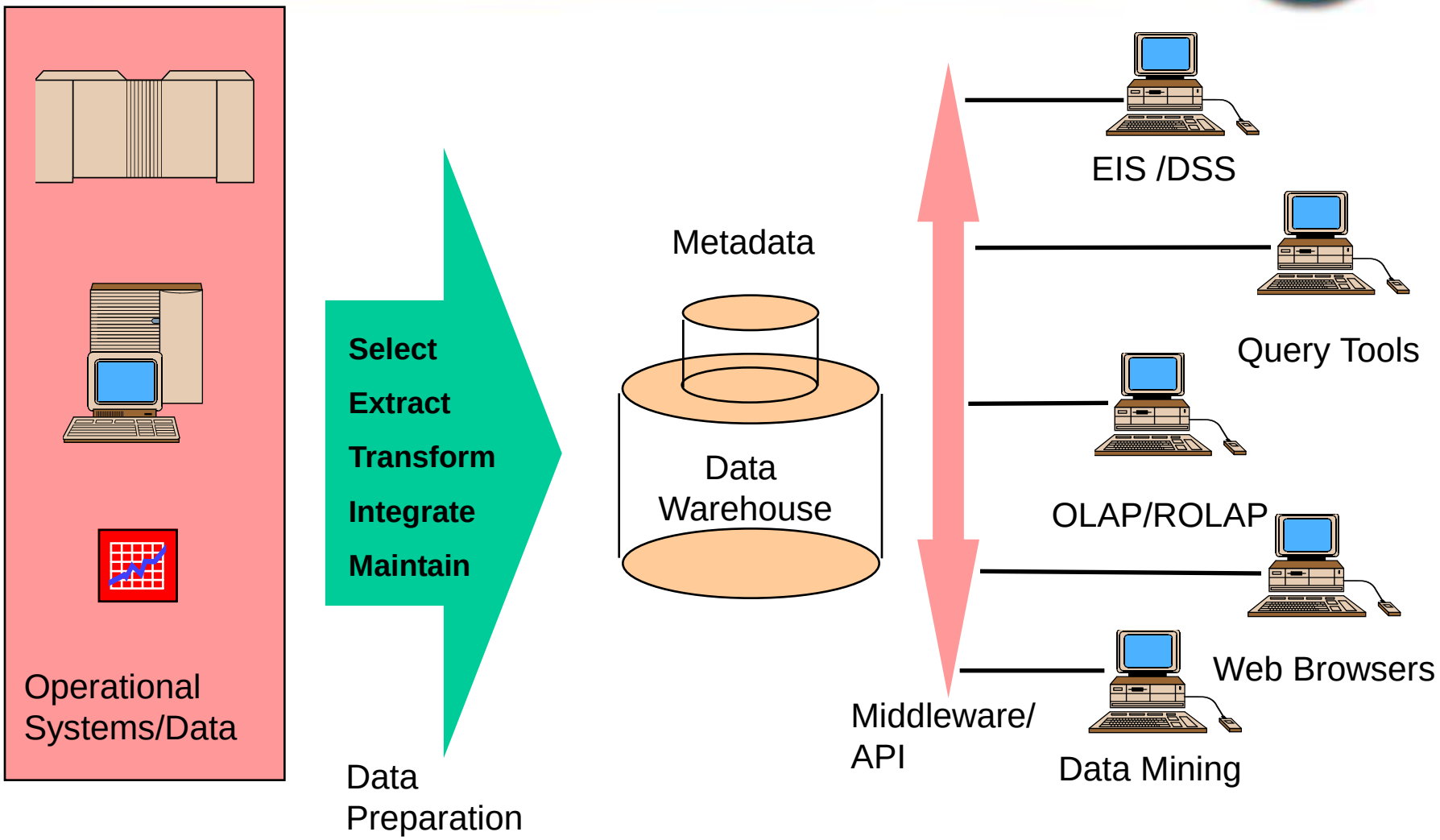
OLTP Vs ODS Vs DWH

Characteristic	OLTP	ODS	Data Warehouse
Audience	Operating Personnel	Analysts	Managers and analysts
Data access	Individual records, transaction driven	Individual records, transaction or analysis driven	Set of records, analysis driven
Data content	Current, real-time	Current and near-current	Historical
Data Structure	Detailed	Detailed and lightly summarized	Detailed and Summarized
Data organization	Functional	Subject-oriented	Subject-oriented
Type of Data	Homogeneous	Homogeneous	Vast Supply of very heterogeneous data

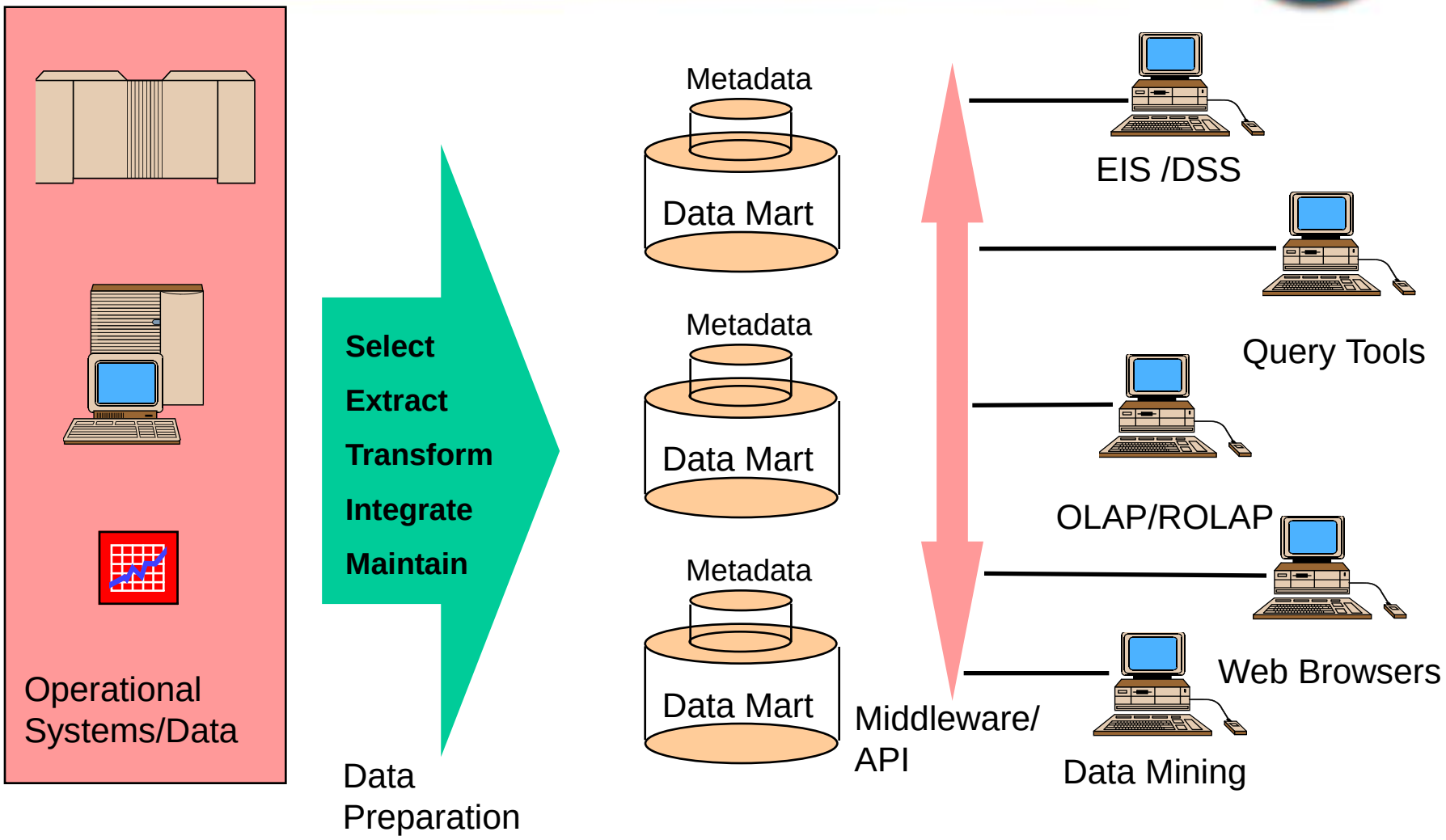
OLTP Vs ODS Vs DWH

Characteristic	OLTP	ODS	Data Warehouse
Data redundancy	Non-redundant within system; Unmanaged redundancy among systems	Somewhat redundant with operational databases	Managed redundancy
Data update	Field by field	Field by field	Controlled batch
Database size	Moderate	Moderate	Large to very large
Development Methodology	Requirements driven, structured	Data driven, somewhat evolutionary	Data driven, evolutionary
Philosophy	Support day-to-day operation	Support day-to-day decisions & operational activities	Support managing the enterprise

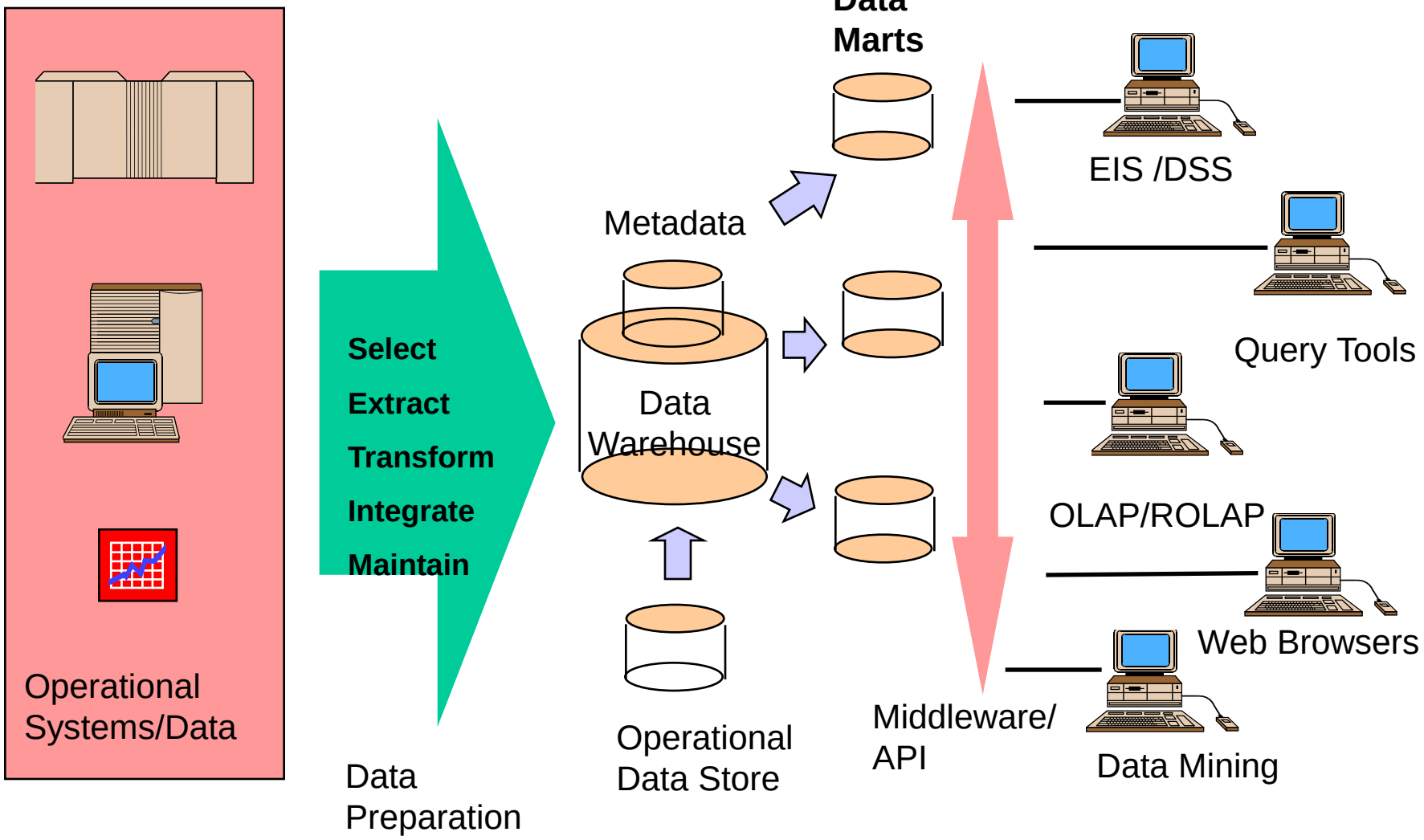
Warehouse Architecture - 1



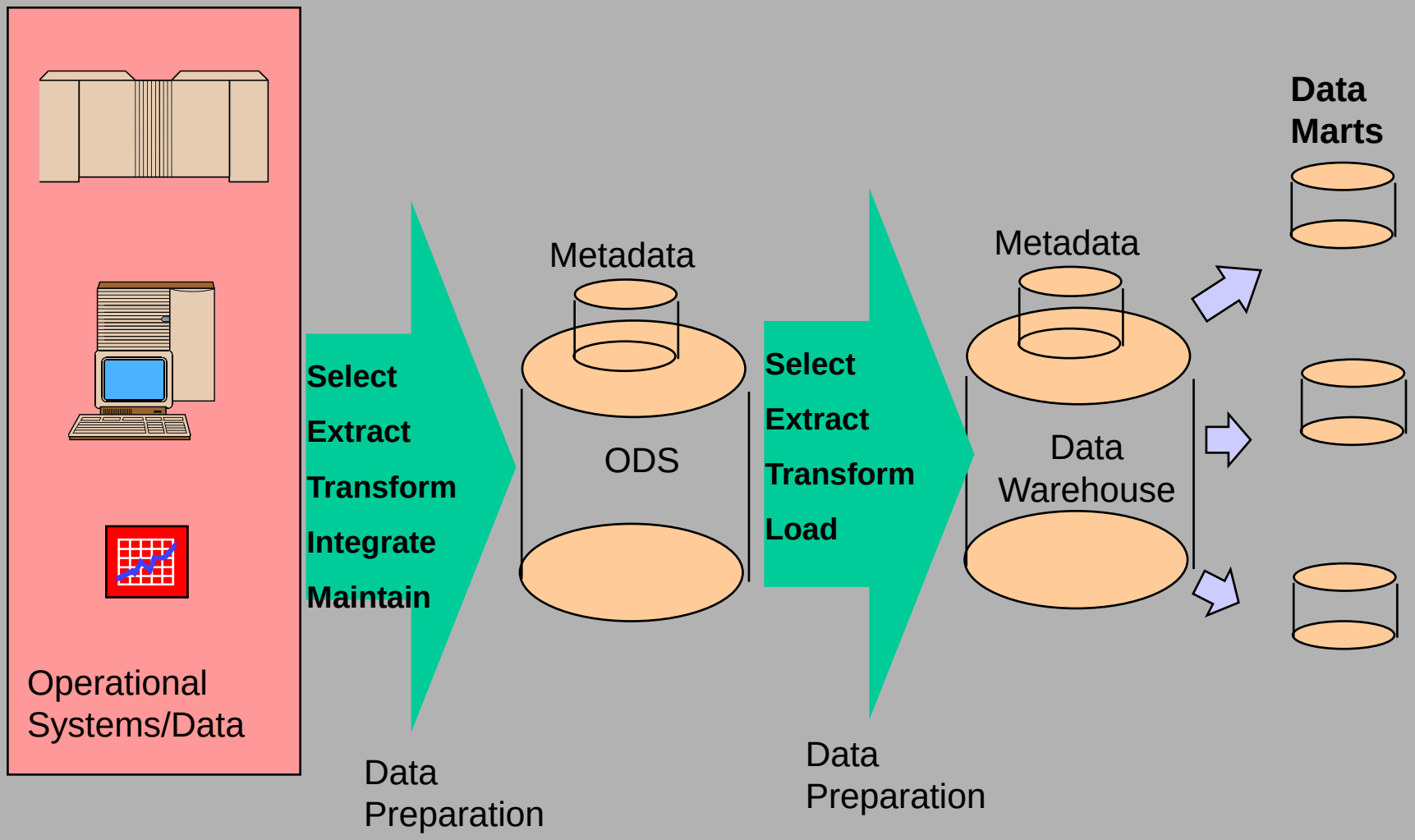
Warehouse Architecture - 2



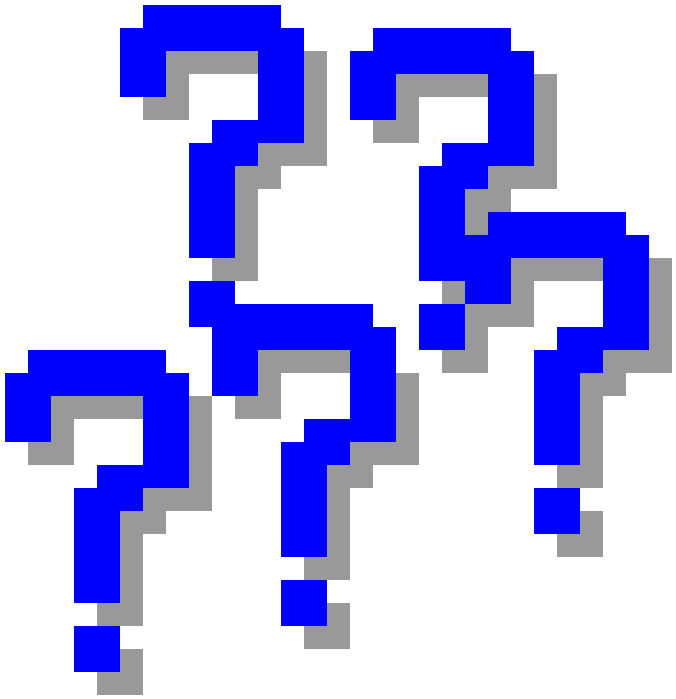
Warehouse Architecture - 3



Warehouse Architecture with ODS



Multi-tiered Data Warehouse with ODS



Questions