

Data Warehousing Concepts

Lesson 2: General Concept of
Data Warehouse

Lesson Objectives

- In this lesson, you will learn:
 - What is a Data Warehouse?
 - History of Data Warehouse
 - Need Of Data Warehouse
 - Data Warehouse Architecture
 - Data Warehouse Components
 - Features of Data warehouse
 - Data Mart
 - Application areas



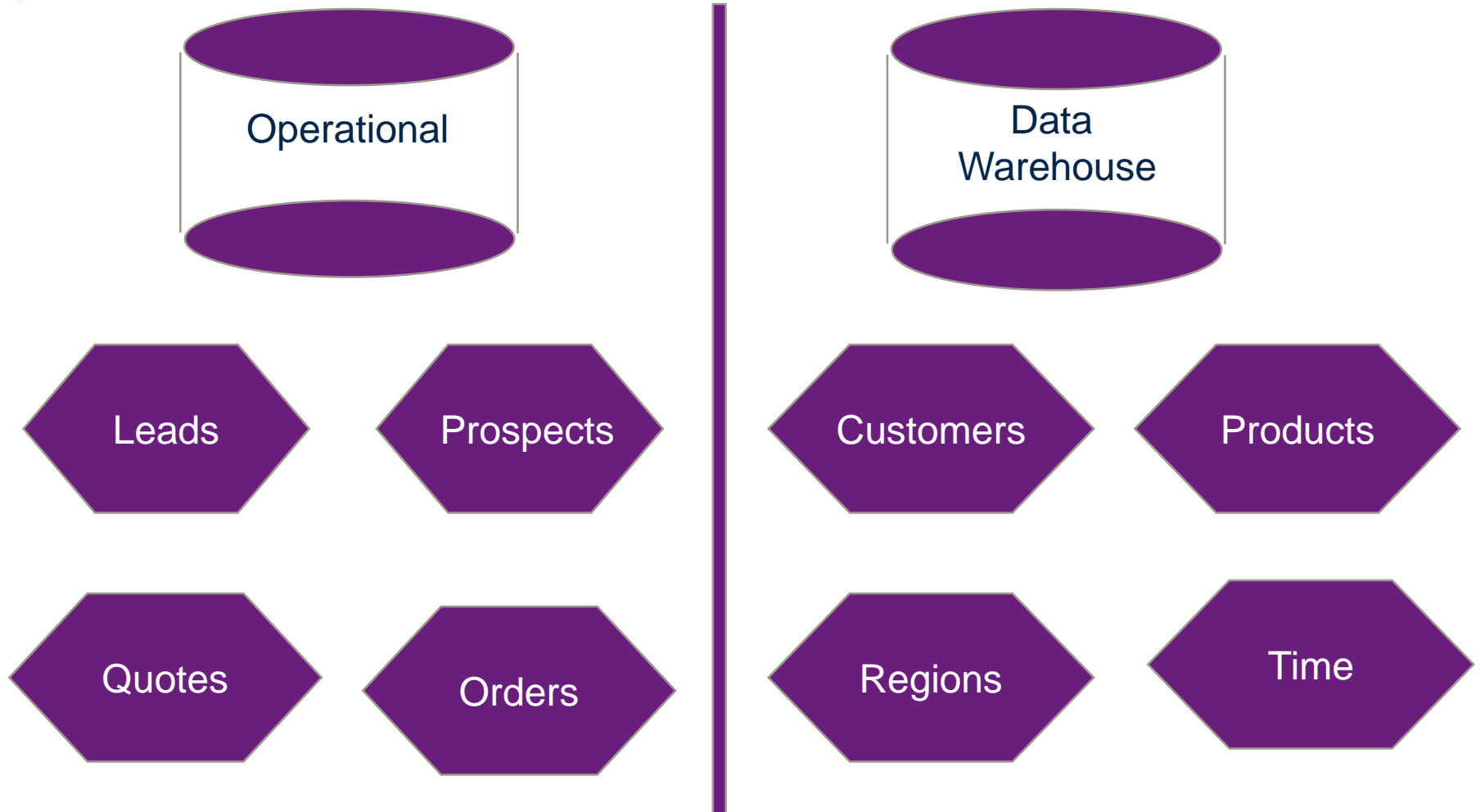
What is a Data Warehouse?

- Data Warehouse is a single, complete, and consistent store of data.
 - It is obtained from a variety of sources.
 - It is made available to users in a way they understand and use in a business context.
 - It is Central repository of information.
 - It is a collection of key information.
 - It contains read-only data.
 - It contains historical data used for analysis purpose.
 - It enables managers to make business decisions.

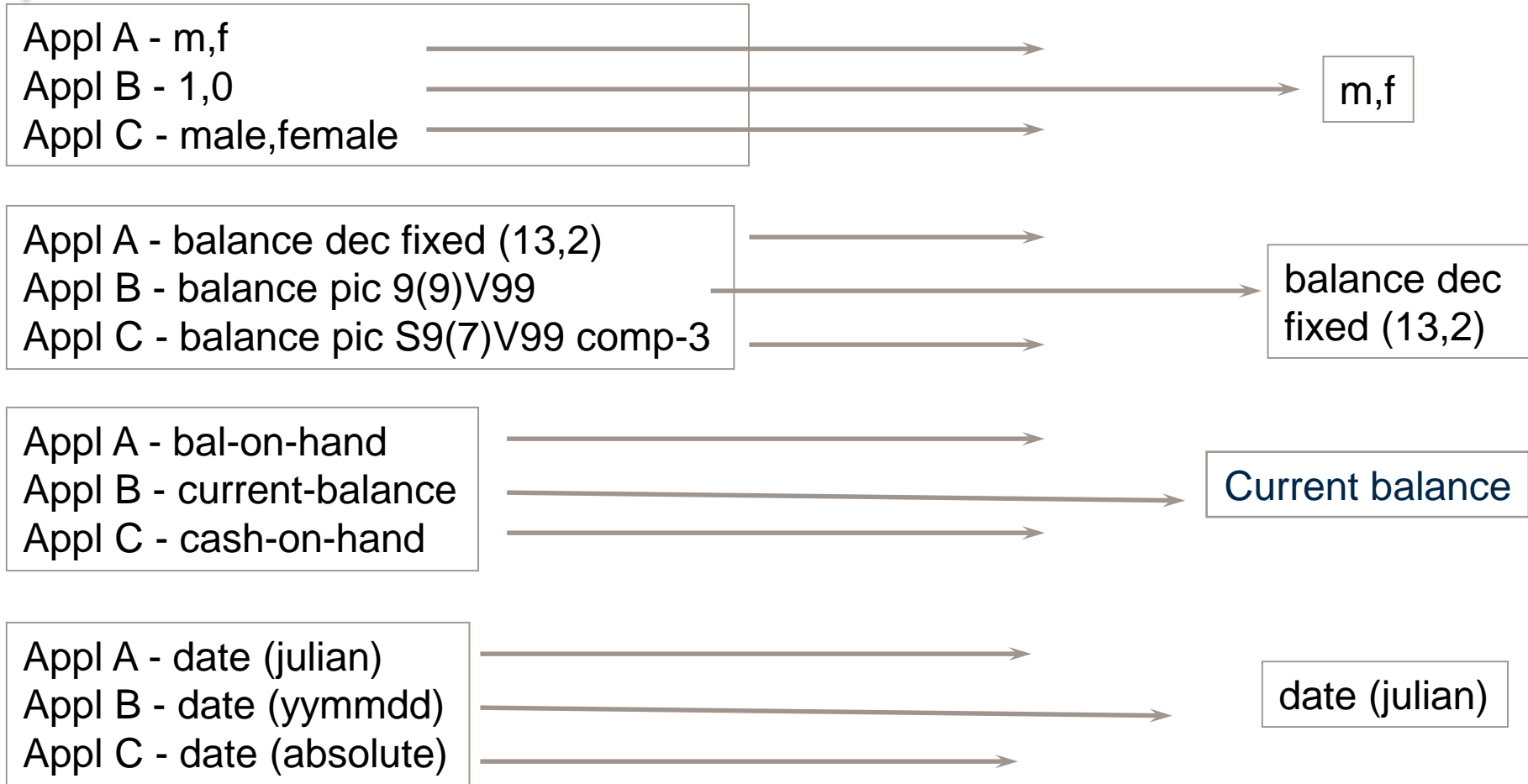
2.2 Characteristics of a Data Warehouse?

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

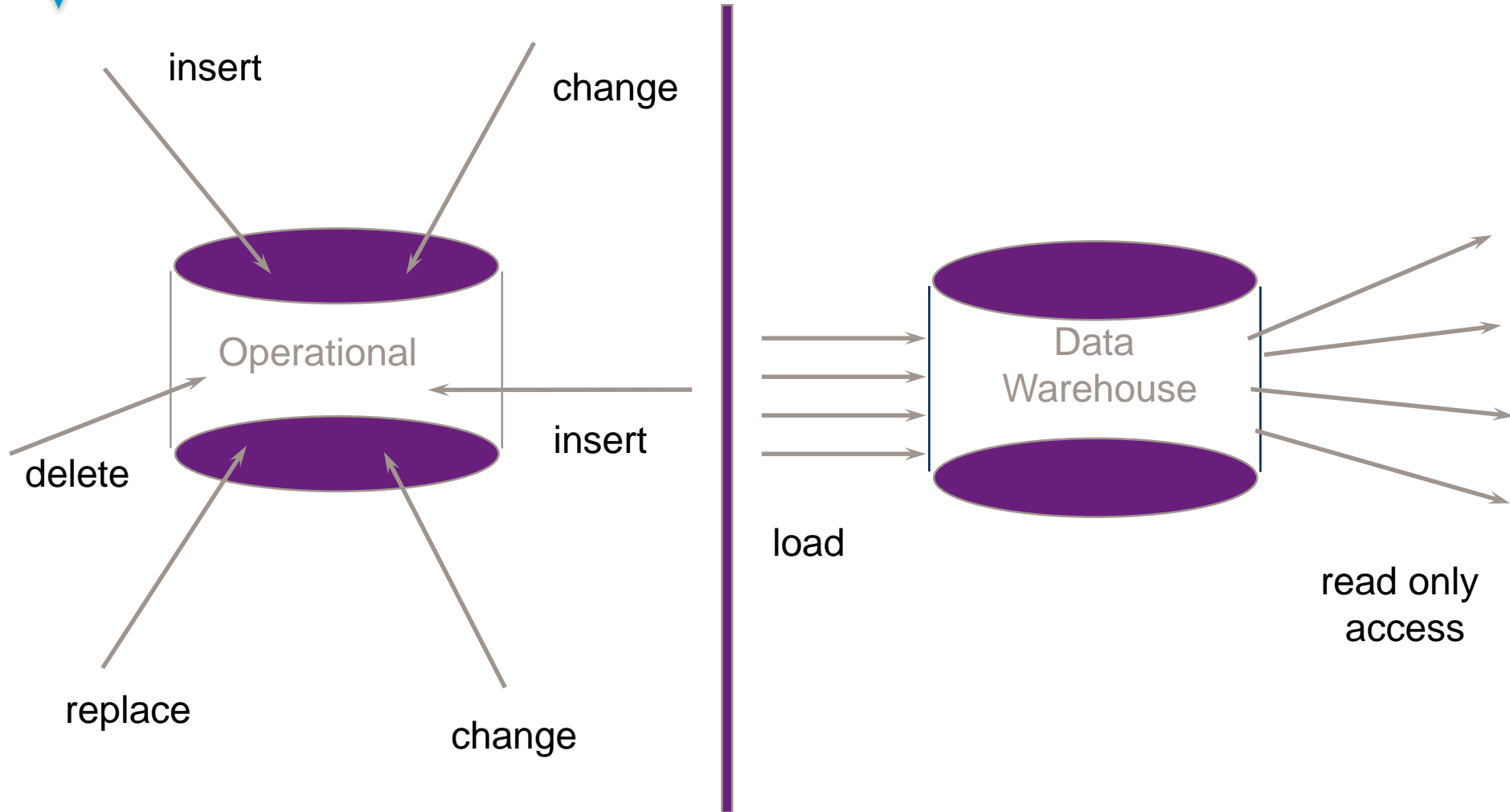
Subject-Oriented



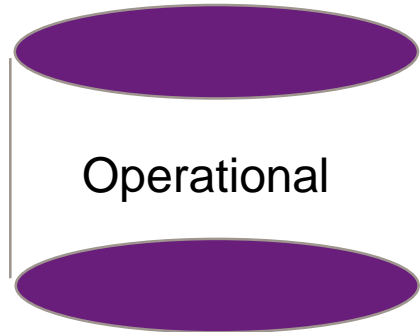
Integrated



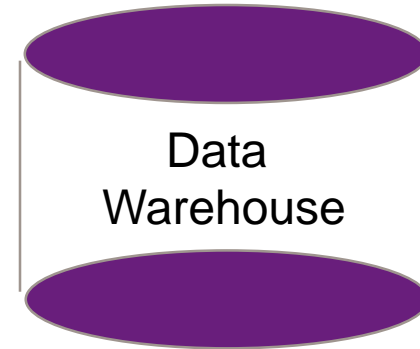
Non-volatile



Time Variant -



- 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

Evolution Of Data warehouse

- 60's: Batch reports
 - hard to find and analyze information
 - inflexible and expensive, reprogram every new request
- 70's: Terminal-based DSS and EIS (executive information systems)
 - still inflexible, not integrated with desktop tools
- 80's: Desktop data access and analysis tools
 - query tools, spreadsheets, GUIs
 - easier to use, but only access operational databases
- 90's till now: Data warehousing with integrated OLAP engines and tools, real time DW

Why Data Warehouse?

- Data Warehouse is required to meet the following needs:
 - Companies want to tap on the vast potential of information to:
 - Have a separate informational system from operational systems
 - Improve quality of decision making
 - Companies seek profitability through focused action.
 - IT business requires an integrated, company-wide view of high quality information.
 - Organizations want to analyze their activities in a balanced way.
 - Organizations need to build on Customer Relationship Management.

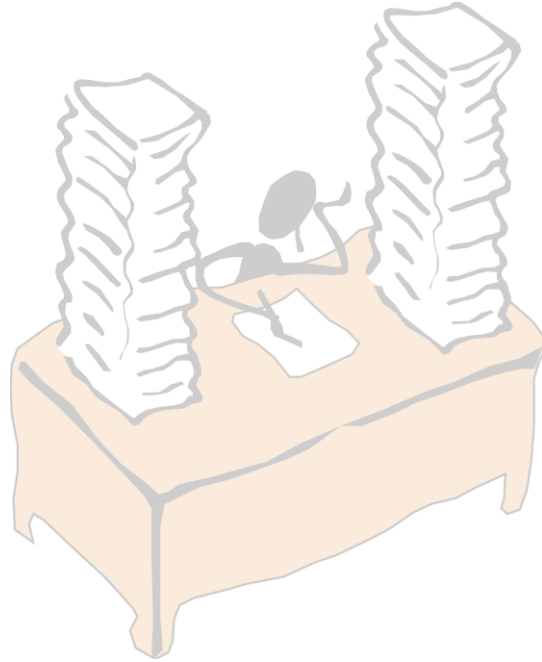
Why a separate Data Warehouse?

- A Data Warehouse helps in finding missing data.
- It provides consolidated data from multiple data sources.
- It helps in maintaining data quality coming from different sources.
- Special data organization is needed for vast volume of data.
- Complex OLAP queries degrade performance.

What is Data Warehouse Architecture?

- Data Warehouse Architecture is a description of the components and services of the Data Warehouse.
 - It provides the mechanism to achieve enterprise integration to support business.
 - It provides an organizing framework that will improve data sharing.

Data Mining works with Warehouse Data



- Data Mining provides the Enterprise with intelligence

- Data Warehousing provides the Enterprise with a memory



What makes data mining possible?

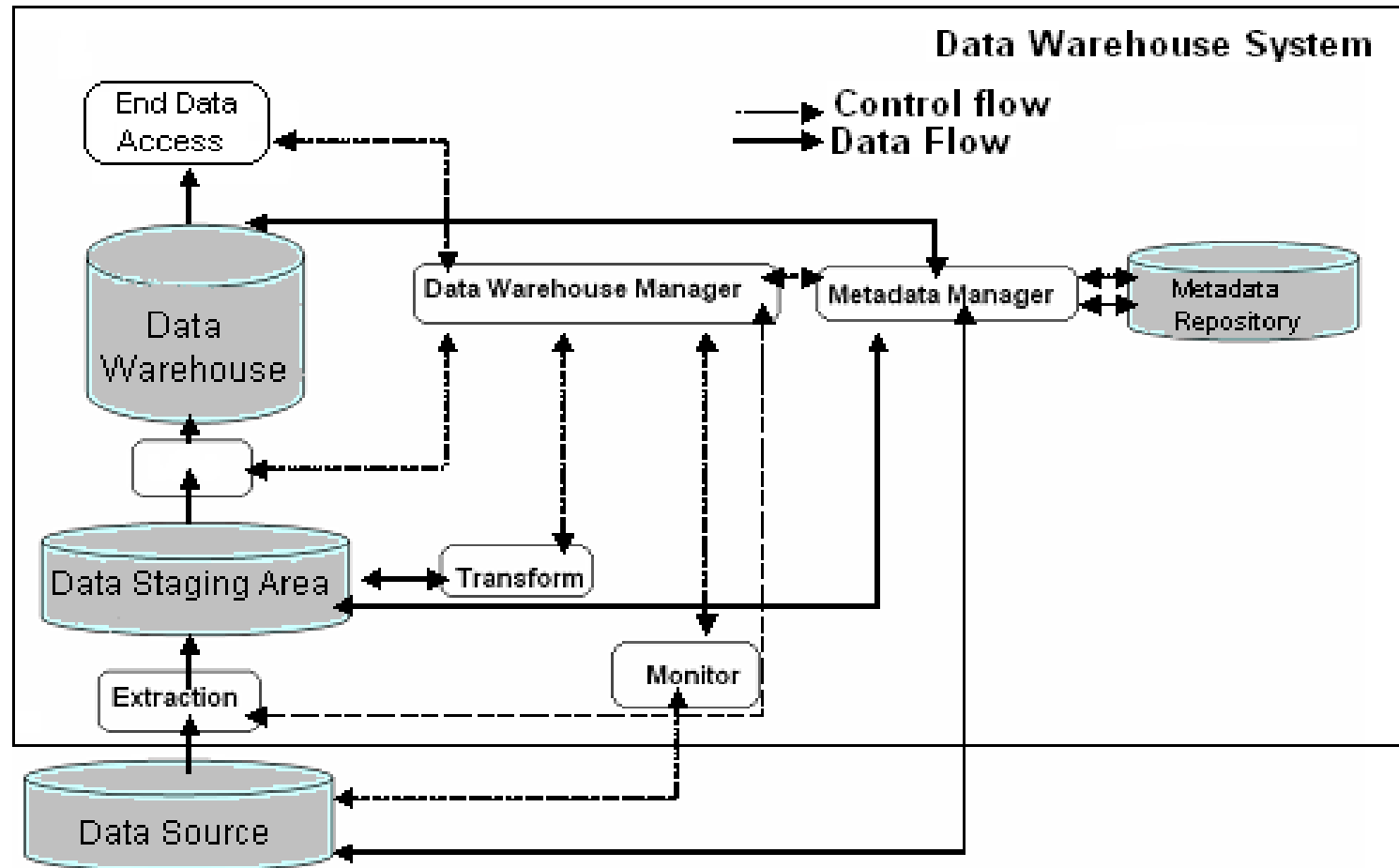
- Advances in the following areas are making data mining deployable:
 - Data warehousing
 - Better and more data (i.e., operational, behavioral, and demographic)
 - The emergence of easily deployed data mining tools and
 - The advent of new data mining techniques.

-- Gartner Group

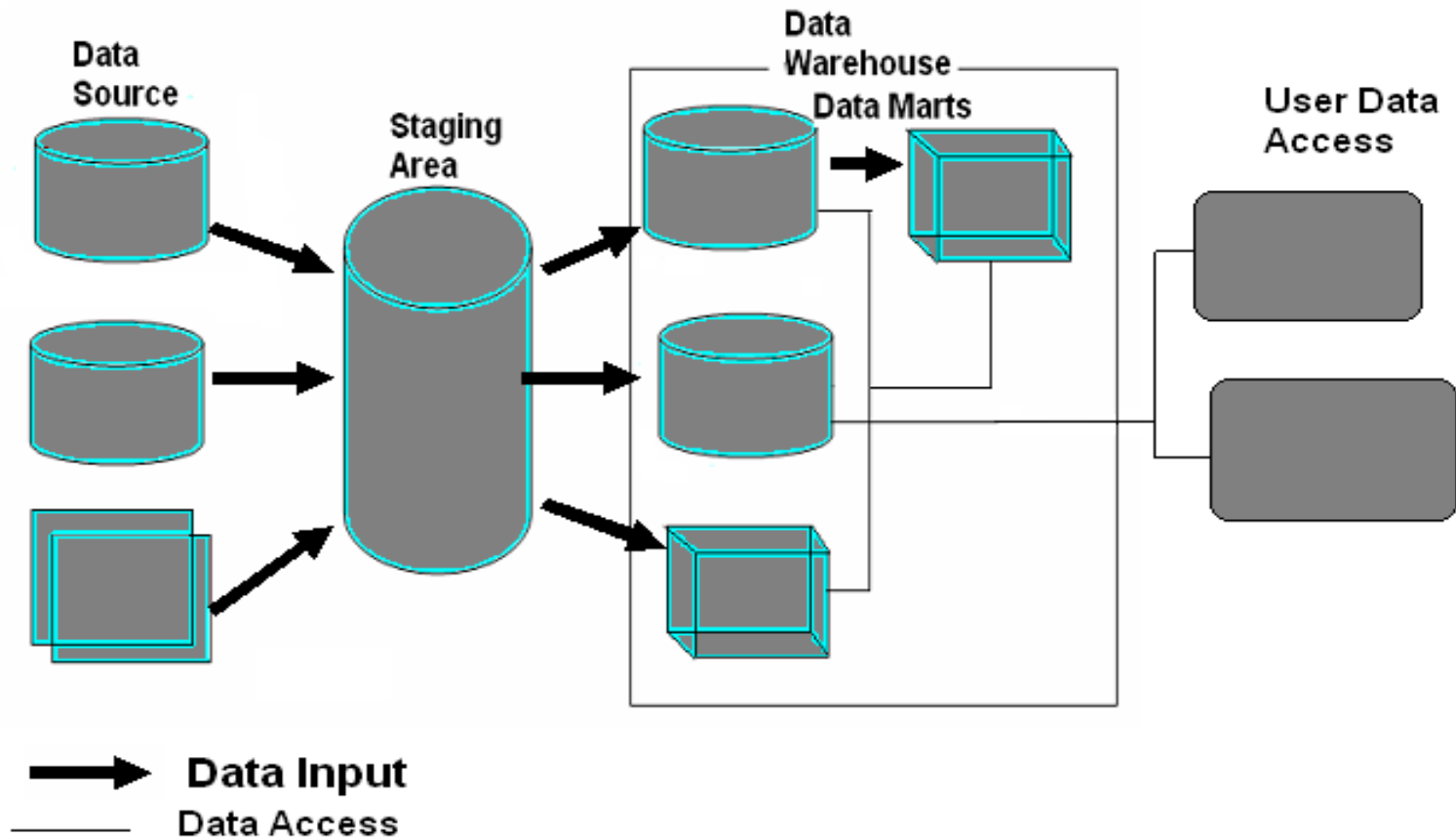
Data Warehouse Architecture Layers

- Data Warehouse Architecture consists of interrelated parts called as “layers” or “components”.
- Four layers of Data Warehouse Architecture are:
 - Operational: Functions as data storage
 - Informational: Stores business logic
 - Data access: Acts as a bridge between operational and informational layer
 - Meta data: Stores data dictionary

Block Diagram – Data Warehouse Architecture



Data Warehouse Components



Salient Features

- Here are some of the features of a Data Warehouse:
 - Time-variant data:
 - Data is meant for analysis and decision-making over the time.
 - Changes to the data are recorded against time dimension.
 - Data is stored as snapshots over past and current periods.
 - Non-volatile data:
 - Data is not needed to run the daily business.
 - Data is primarily used for query and analysis.
 - Individual transactions are not updated in a Data warehouse.
 - Data is never over-written or deleted. It is non-updatable data.

Salient Features

- Data granularity:
 - It refers to the level of detail.
 - It is inversely proportional to the amount of data stored.
 - Data is summarized at different levels.
 - Many Data warehouses have at least two levels of granularity.
 - Summarized data is stored.
 - It reduces storage costs.
 - It reduces CPU usage.
 - It increases performance since smaller number of records have to be processed.
 - Design is around traditional high level reporting needs.
 - Tradeoff with volume of data to be stored and detailed usage of data.

Salient Features

- Subject oriented:
 - Data is stored by subjects, not applications.
 - Data is organized in the Data Warehouse such that it will infer the real world.
 - Data is organized around major subjects, such as customer, product, sales.
 - Focus is on the modeling and analysis of data for decision makers.
 - DW provides a simple and concise view around a particular subject.
 - DW is organized around the key subject of the enterprise.
 - Major subjects may include customers, patients, students, products, and time.

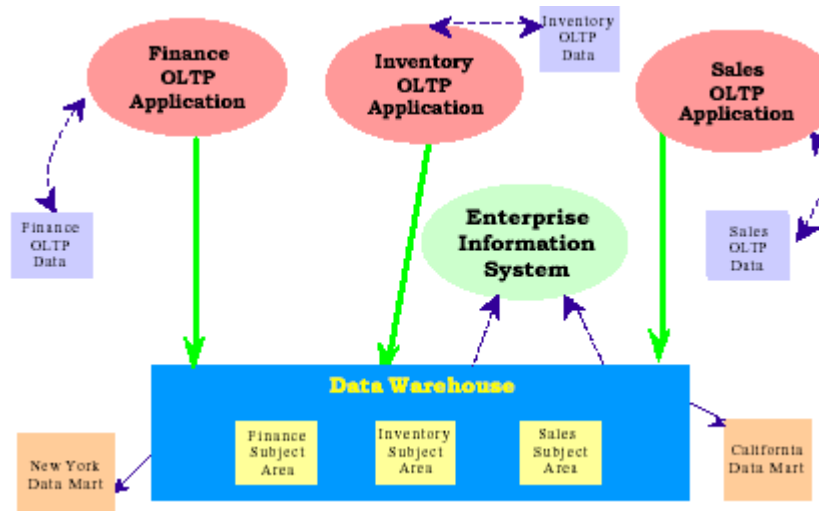
Salient Features

- Integrated data:
 - Data is pulled from various databases from all applications.
 - Operational platforms and operating systems for the data could be different.
 - Data has to undergo a process of transformation, consolidation, and integration.
 - Data inconsistencies are removed, standardization is achieved.

What is a Data Mart?

- Data Mart is a subset of the Data warehouse.
 - It is typically fed from the Data warehouse.
 - It is a Data warehouse that has limited scope.
 - It is a repository of data gathered from operational data and other sources.
 - It is used for decision making by a particular end-user group.
 - Emphasis is on meeting the specific demands of a particular group of knowledge users.
 - Maintain the ability to access the underlying base data.

What is a Data Mart?



Types of Data Marts

- Dependent Data Mart

- A Data Mart whose source is the Data Warehouse
- All dependent Data Marts are loaded from the same source – the Data Warehouse

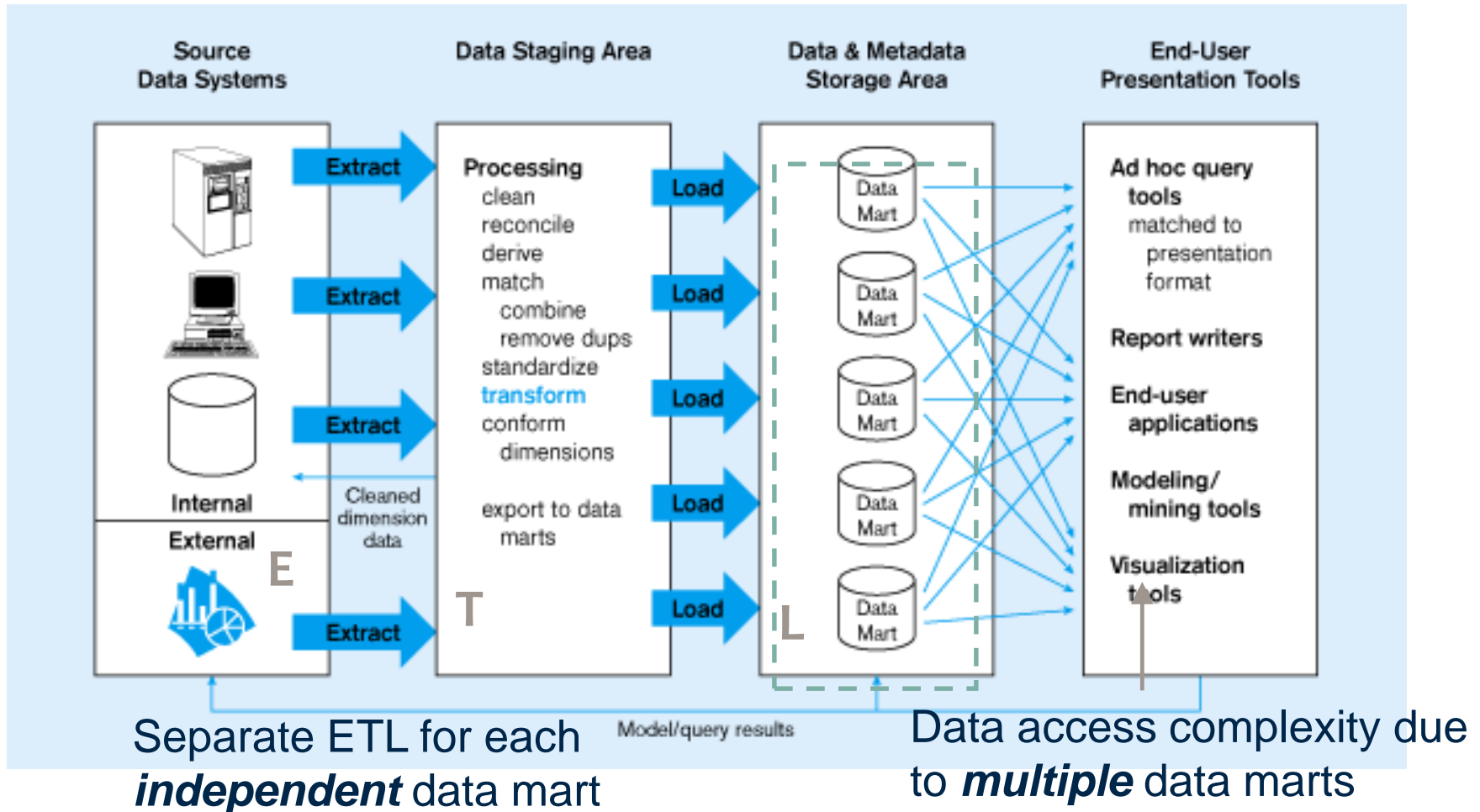
- Independent Data Mart

- A Data Mart whose source is the legacy application environment
- Each independent Data Mart is fed uniquely and separately by the individual source systems

Independent Data Mart

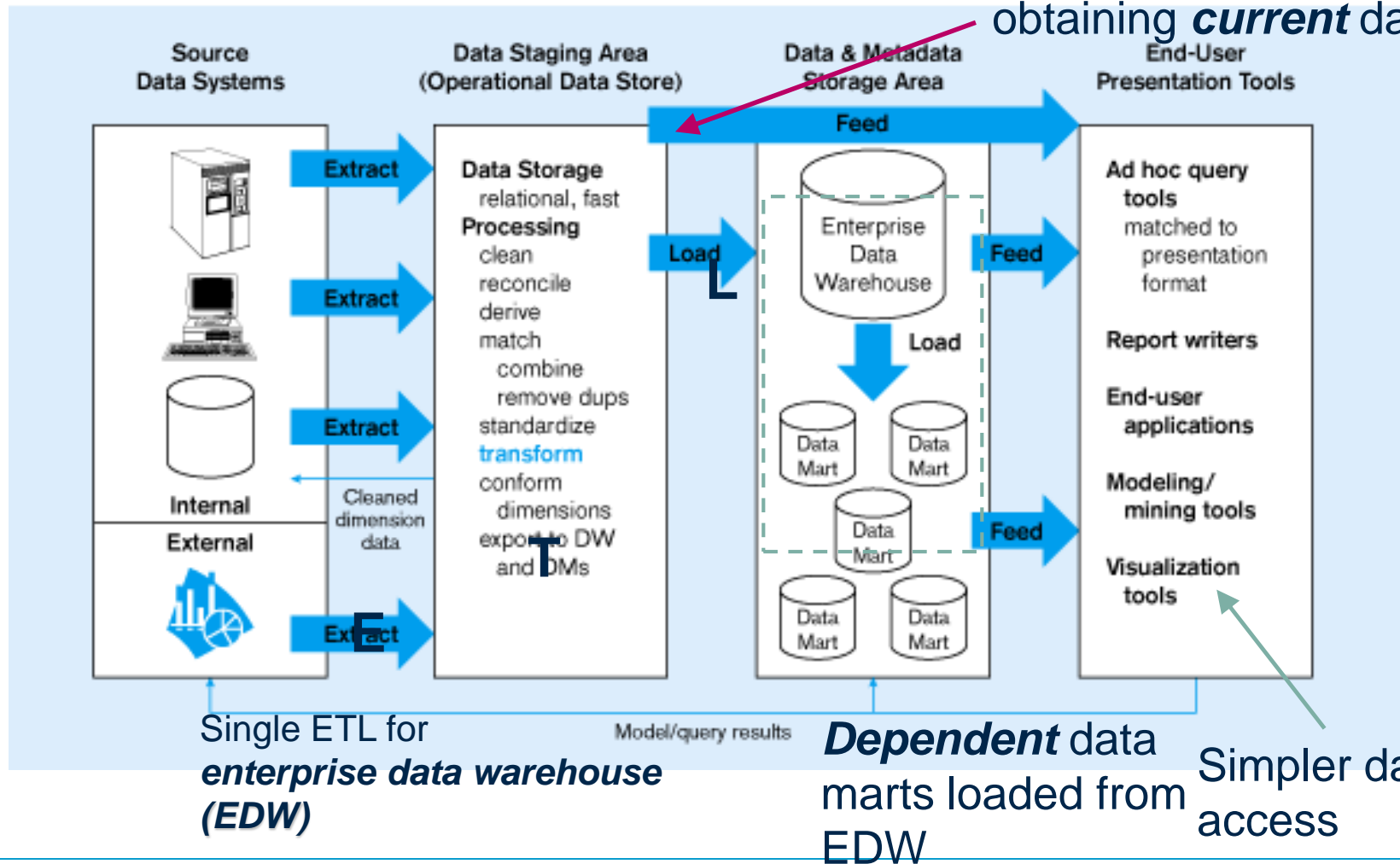
Data marts:

Mini-warehouses, limited in scope



Dependent data mart with operational data store

ODS provides option for obtaining **current** data



Industry-wise Application

Industry	Application
Finance	Credit Card Analysis
Insurance	Claims, Fraud Analysis
Telecommunication	Call record analysis
Transport	Logistics management
Consumer goods	Promotion analysis
Data Service providers	Value added data
Utilities	Power usage analysis

Summary

- In this lesson, you have learnt:
 - Data Warehouse stores historical data.
 - Data Mart emphasizes on meeting the specific demands of a particular group of knowledge users.
 - Features of Data Warehouse are:
 - Time variant data
 - Non volatile data
 - Data granularity
 - Subject oriented
 - Integrated data



Review Question

- Question 1: _____ is a subset of data warehouse.
- Question 2: Data Mart is a structure for corporate view of data.
 - True/ False
- Question 3: ____ is used for decision making by a particular end-user group.

