Data Science & Artificial Intelligence

Warehousing

Introduction of Data Warehousing

ONE SHOT

Astha Singh Ma'am

Topics to be covered

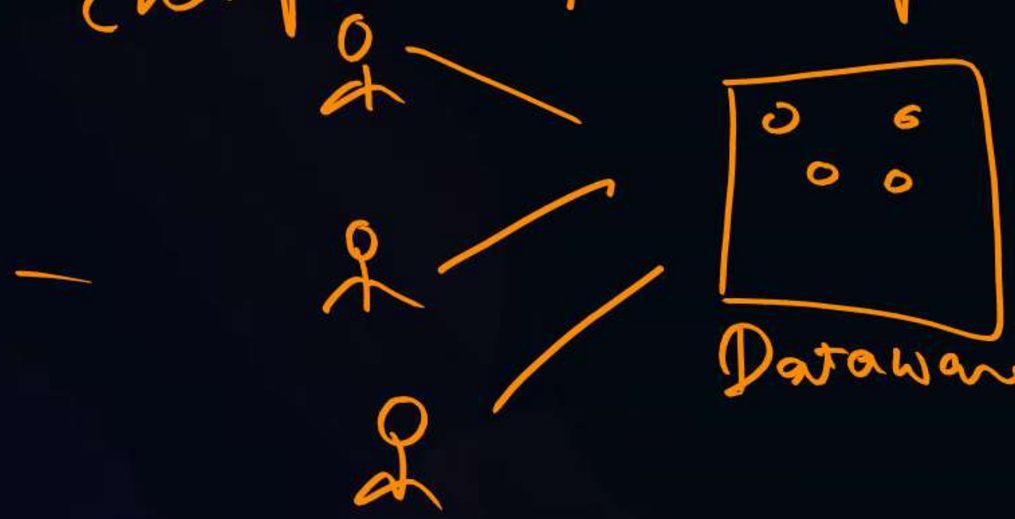
What is Data Warehouse ?
How Does Data Warehouse Work?
Need for Data Warehouse
Characteristics
Function
Data Warehouse vs DBMS
Organizational trends Motivating data Warehouses
Comparison of Operational and Informational Systems
Data Warehouse Architecture
What is Data marts?
Data Warehouse Versus Data Mart
Derived data

Topic : DEFINITIONS

What is Data Warehouse ? (DW)

also known as enterprise data warehouse

Example → framer data



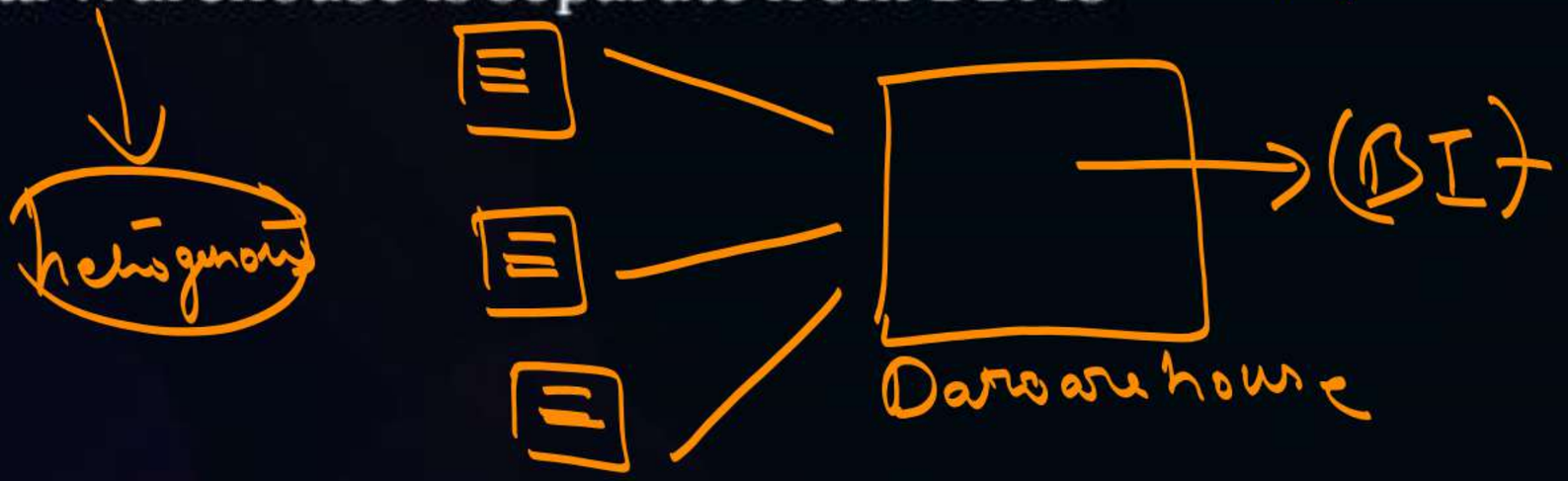
Example :-

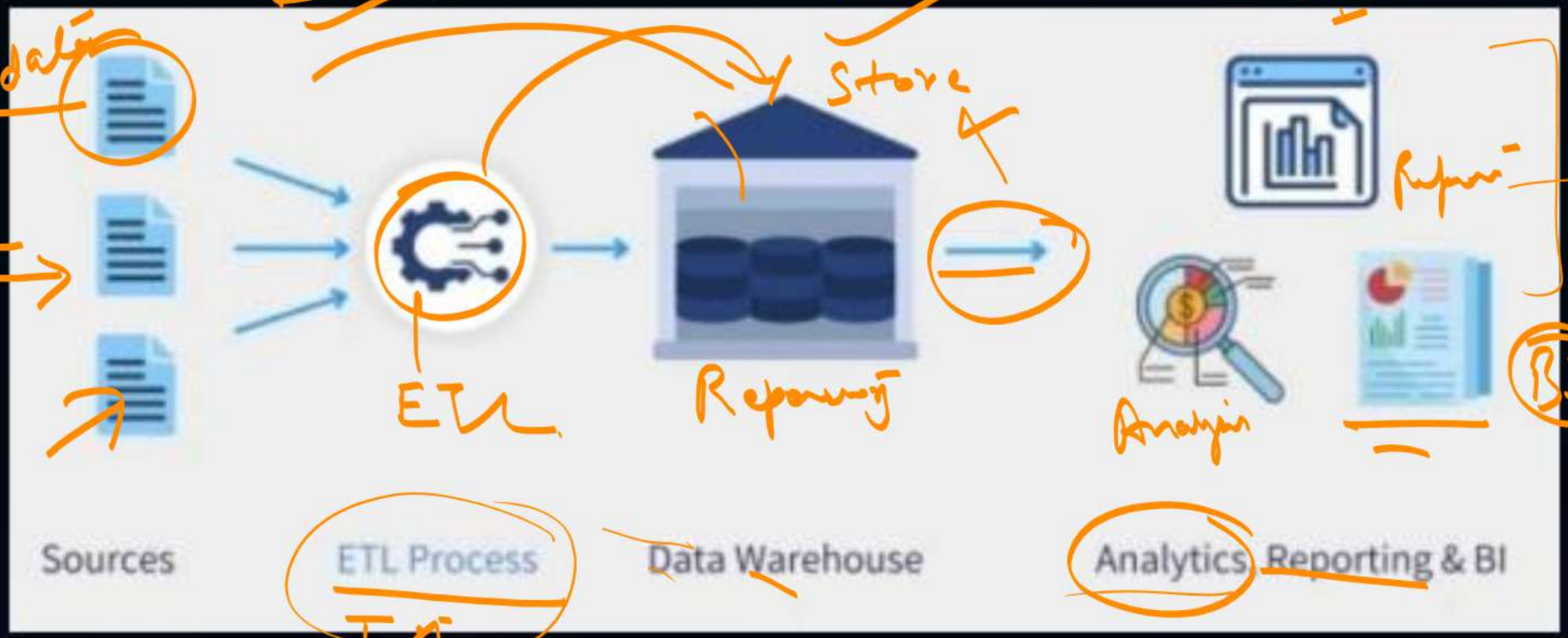
blinkit

Oracle

- A Data Warehousing (DW) is process for collecting and managing data from varied sources to provide meaningful business insights

- A Data Warehouse is separate from DBMS → RDBMS ⇒ Row base




Sources
ETL Process
Data Warehouse
Analytics, Reporting & BI

Topic :Characteristics

- Subject-Oriented: It organizes according to specific business subjects or areas, such as sales, marketing, or finance
- Integrated: Data from multiple sources is integrated into a consistent format that can be easily analyzed and reported.
- Time-Variant: It stores historical data over time, enabling users to analyze trends and changes in data over time.
- Non-volatile: Data in a warehouse cannot be modified or deleted once loaded.

Topic : Function

- Data Integration: One of its primary functions is to integrate data from multiple sources into a single, centralized repository
- Data Management: It also performs several functions related to data management, including data storage, organization, and retrieval.
- Data Analysis: Finally, its primary function of it is to enable advanced data analysis and reporting.

Topic : NEED FOR DATA WAREHOUSING

Topic :Data Warehouse vs DBMS

DBMS	Data Warehouse
A common Database is based on operational or transactional processing.	A data Warehouse is based on analytical processing.
Database stores current and up-to-date data which is used for daily operations.	A Data Warehouse maintains historical data over time.
Example – A database stores related data, such as the student details in a school.	**Example –** A data warehouse integrates the data from one or more databases , so that analysis can be done to get results , such as the best performing school in a city.

MCQ

Gate

1.What is the primary purpose of a data warehouse?

A) Real-time transaction processing
B) Data analysis and decision support
C) Online transaction processing
D) Data storage and retrieval

(ETL

2.Which term is used to describe the process of extracting, transforming, and loading data into a data warehouse?

A) ETL
B) OLAP
C) OLTP
D) SQL

Which SQL operation is commonly used for retrieving data from a Data Warehouse?

a) SELECT
b) UPDATE
c) DELETE
d) INSERT

Which of the following is a challenge in Data Warehousing?

a) Lack of data redundancy
b) Real-time data processing
c) Data security concerns
d) Small data volumes

Topic : ISSUES WITH COMPANY-WIDE VIEW

- Inconsistent key structures
- Synonyms
- Free-form vs. structured fields
- Inconsistent data values
- Missing data

Topic : ORGANIZATIONAL TRENDS MOTIVATING DATA WAREHOUSES

- No single system of records
- Multiple systems not synchronized
- Organizational need to analyze activities in a balanced way
- Customer relationship management
- Supplier relationship management

Topic : SEPARATING OPERATIONAL AND INFORMATIONAL SYSTEMS

- **Operational system –** a system that is used to run a business in real time, based on current data; also called a system of record

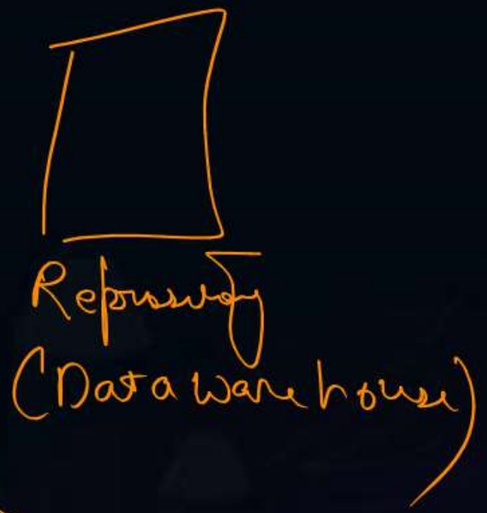
- **Informational system –** a system designed to support decision making based on historical point-in-time and prediction data for complex queries or data-mining applications

TABLE 9-1 : Comparison of Operational and Informational Systems

Characteristic	Operational Systems	Informational Systems
Primary Purpose	Run the business on a current basis	Support managerial decision making
Type of data	Current salespersons, administrators	Managers, business analysts, customers
Scope of usage	Narrow, planned, and simple updates and queries	Broad, ad hoc, complex queries and analysis
Design goal	Performance, throughput, availability	Ease of flexible access and use
Volume	Many constant updates and queries on one or a few table rows	Periodic batch updates and queries requiring many or all rows

Topic : DATA WAREHOUSE ARCHITECTURES

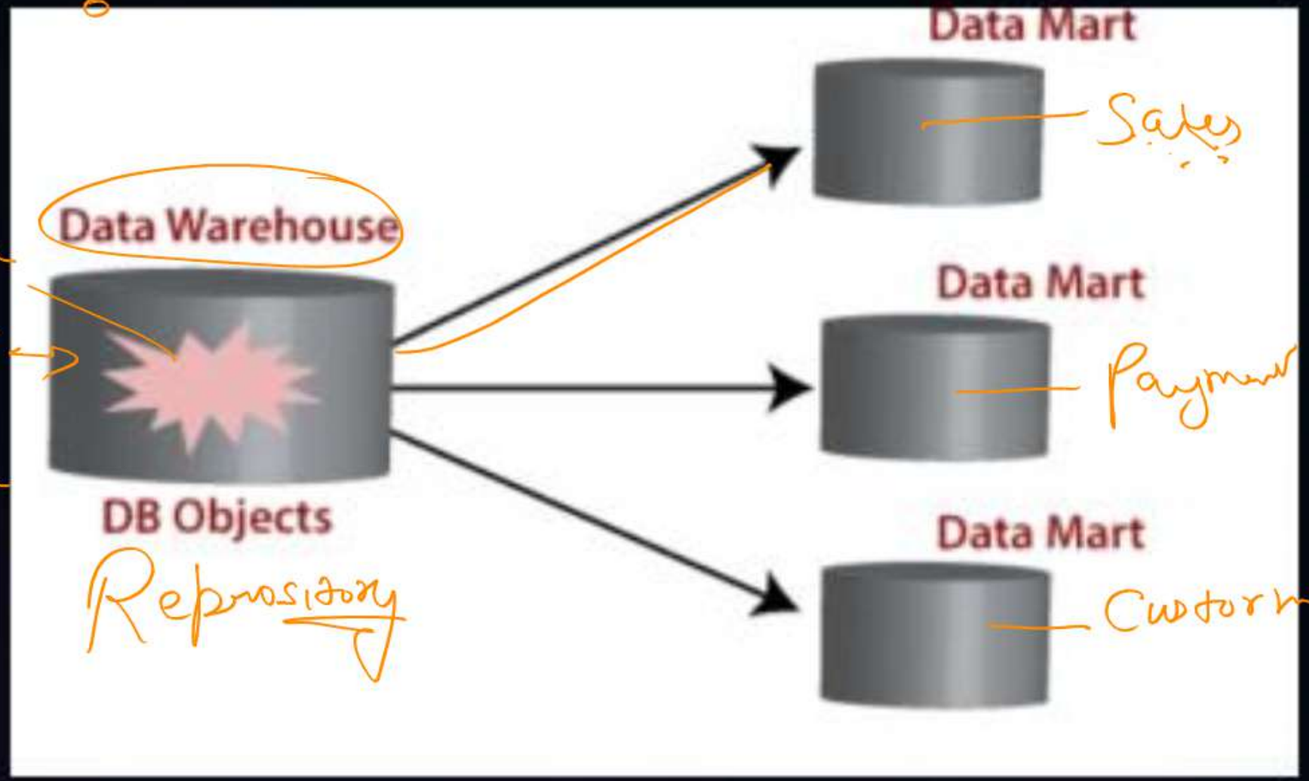
- Independent Data Mart
- Dependent Data Mart and Operational Data Store
- Logical Data Mart and Real-Time Data Warehouse
- Three-Layer architecture



All involve some form of ***extract***, ***transform*** and ***load*** (**ETL**)

Topic : What is Data marts

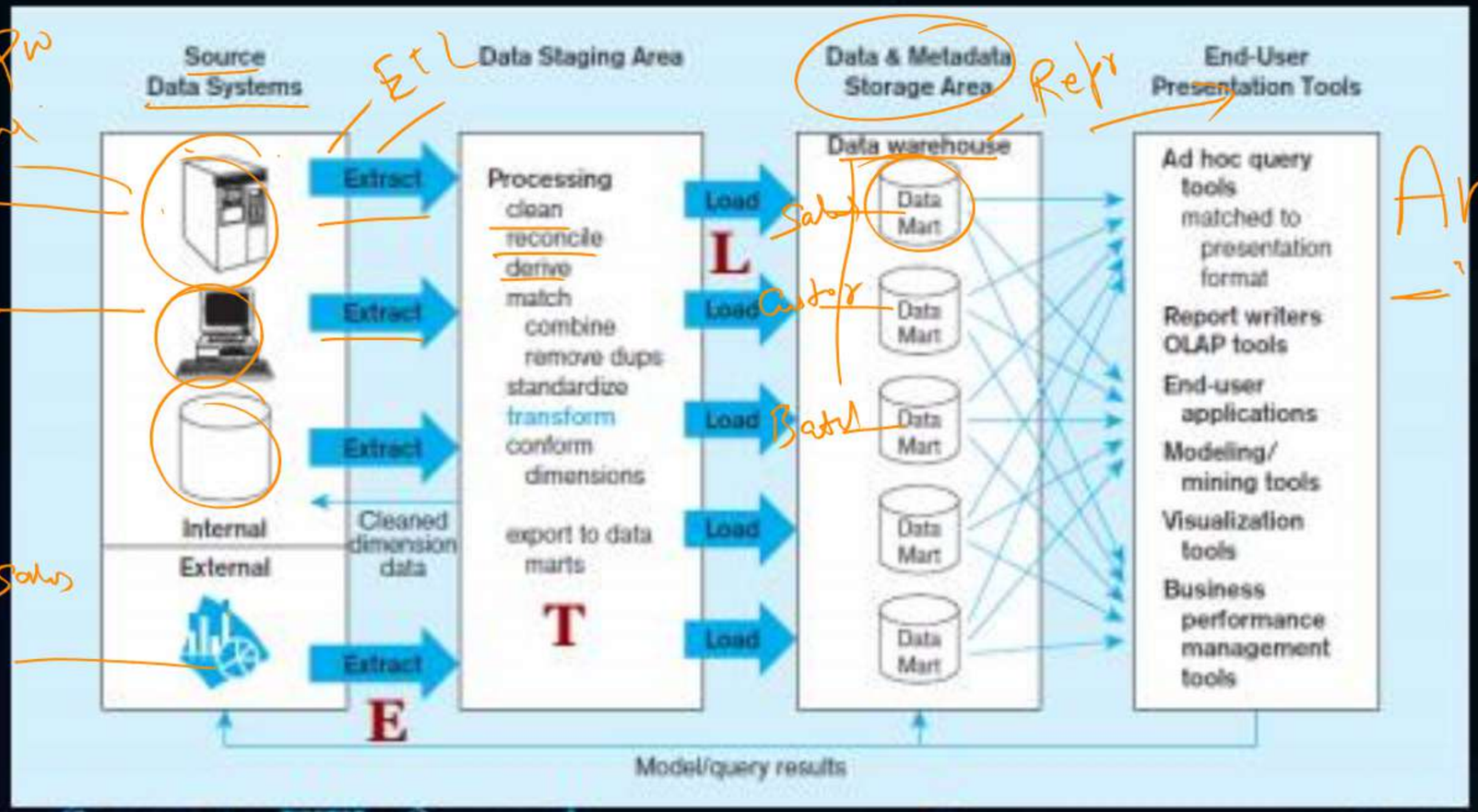
A Data Mart is a subset of a directorial information store



Topic : Data marts:

Mini-warehouses, limited in scope

Independent data mart data warehousing architecture



Separate ETL for each ***independent*** data mart

Data access complexity due to ***multiple*** data marts

What is a Data Mart?

A) A centralized repository of all organizational data.
B) A subset of a data warehouse that is focused on specific business functions or user groups.
C) An external storage for backup purposes.
D) A data modeling technique for normalization.

What is the main advantage of using Data Marts in a data warehousing architecture?

A) Simplifies the ETL process.

B) Provides a centralized view of all data.

C) Allows for more focused and quicker access to data for specific user groups.

D) Reduces the need for indexing.

What is the purpose of indexing in a data warehouse?

A) To store metadata

B) To optimize query performance

C) To enforce data integrity

D) To facilitate real-time processing

Which term is used to describe a subset of a data warehouse focused on a specific business area?

A) Data Lake

B) Data Cube

C) Data Mart

D) Data Repository

Which type of schema is commonly used in data warehousing to organize tables into a star-like structure?

A) Snowflake schema
B) Star schema
C) Hierarchical schema
D) Relational schema

In data warehousing, what is the purpose of a data dictionary?

A) Stores historical data
B) Manages user authentication
C) Stores metadata and data definitions
D) Performs complex data calculations

Topic : Data Warehouse Versus Data Mart

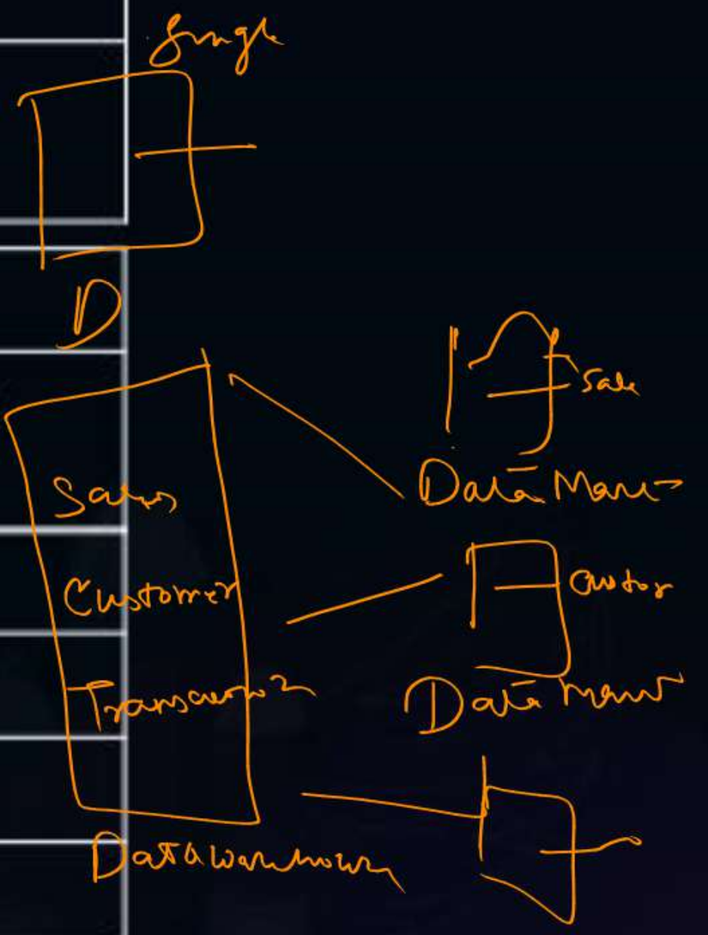
Data Warehouse	Data Mart
Scope	Scope
Application independent	Specific DSS application
Centralized, possibly enterprise-wide	Decentralized by user area
planned	Organic, possibly not planned

Data	Data
Historical, detailed and summarized	Some history, detailed, and summarized
Application independent	Highly denormalized

Topic : Data Warehouse Versus Data Mart

Subjects	Subjects
Multiple subjects	One central subject of concern to users

Sources	Sources
Many internal and external sources	Few internal and external sources
Other Characteristics	Other Characteristics
Flexible	Restrictive
Data oriented	Project oriented
Long life	Short life
Large	start small, becomes large
Single complex structure	Multi, semi-complex structures, together complex



Topic : DERIVED DATA

Objectives:

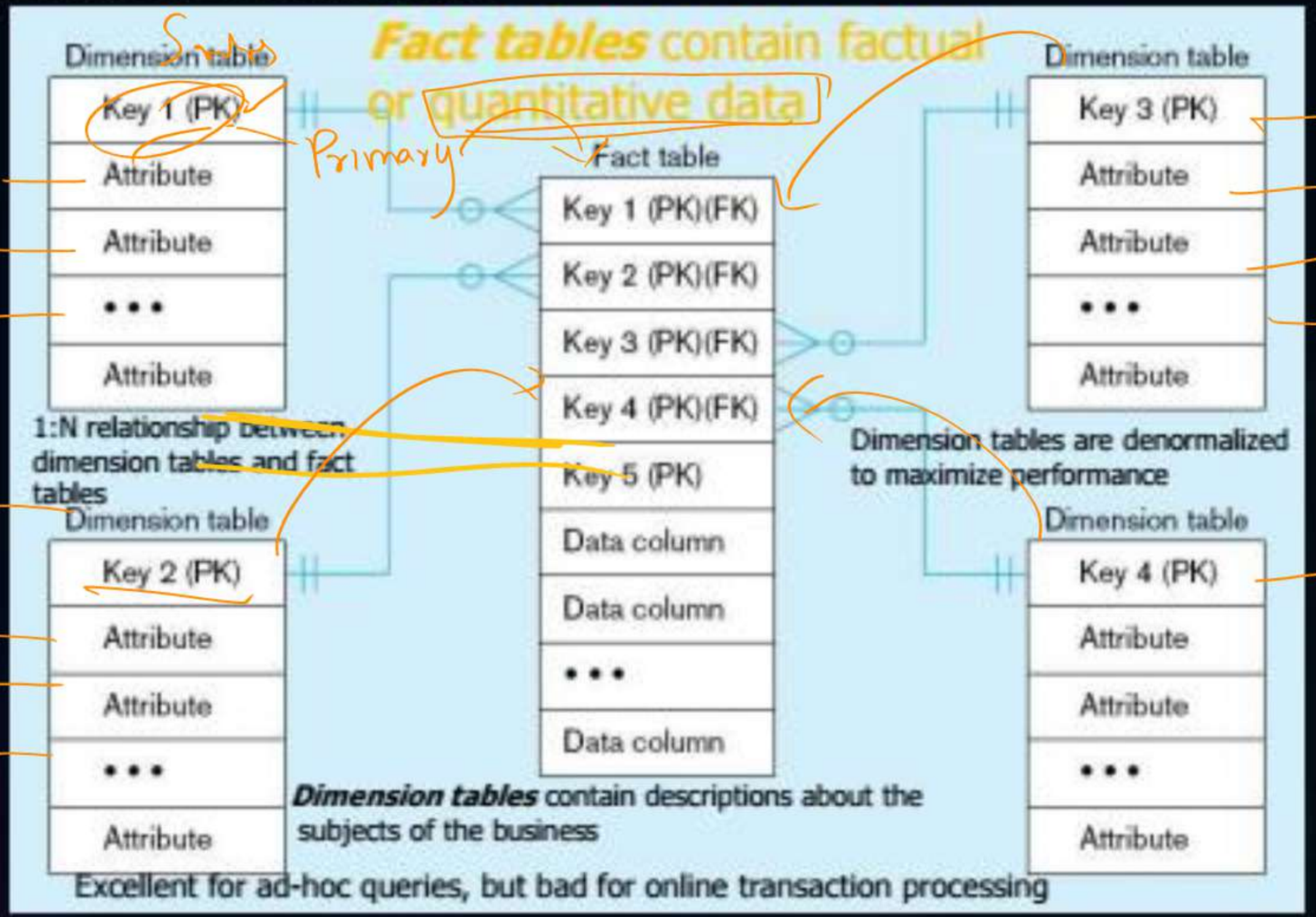
- Ease of use for decision support applications
- Fast response to predefined user queries
- Customized data for particular target audiences
- Ad-hoc query support
- Data mining capabilities

Characteristics

- Detailed (mostly periodic) data
- Aggregate (for summary)
- Distributed (to departmental servers)
- Most common data model = **dimensional model**
- (usually implemented as a **star schema**)

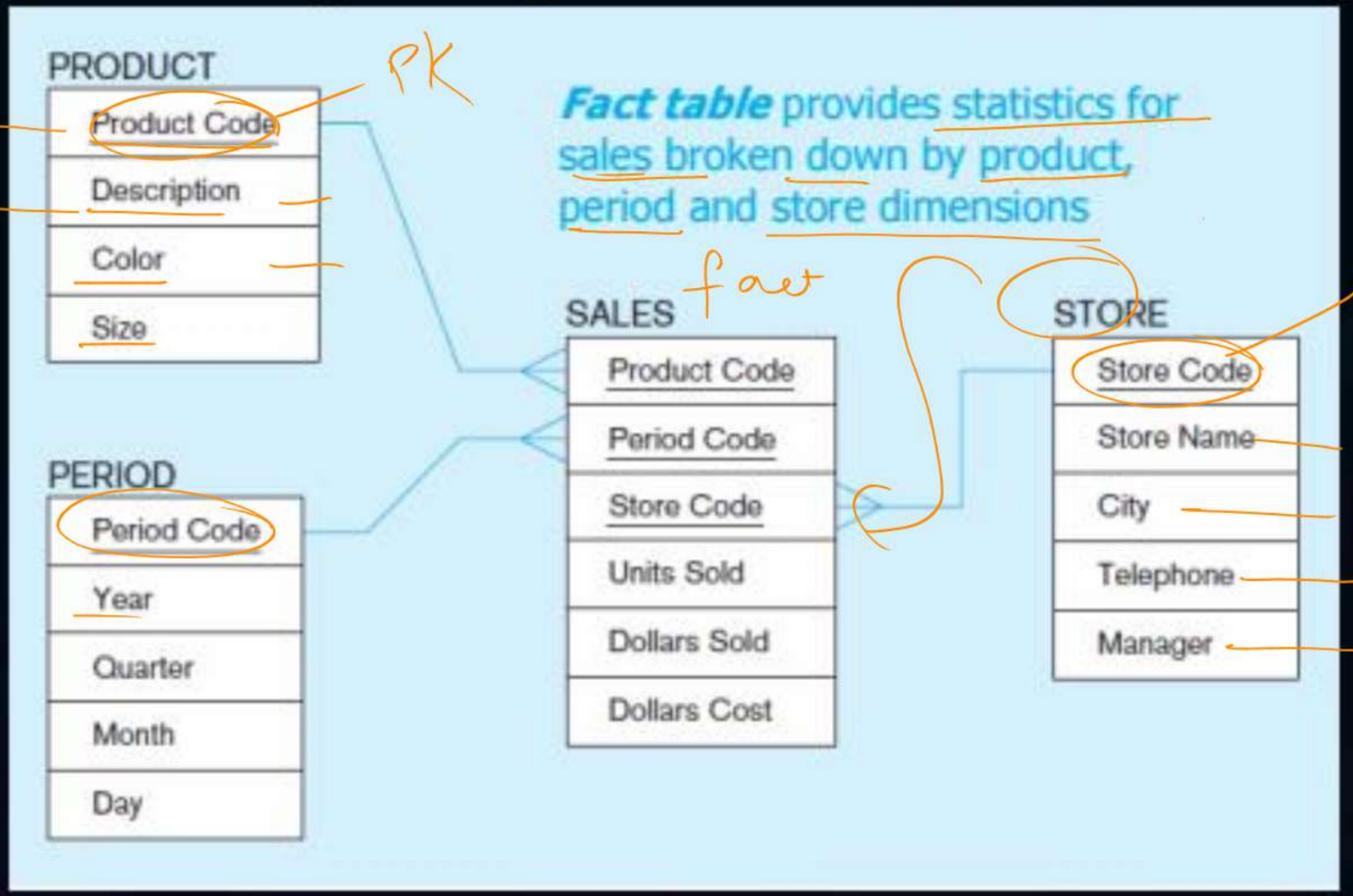
Topic : DERIVED DATA

Components of a star schema



Topic : DERIVED DATA

10 Star schema example



Topic :Next Session

- OLAP Technology
- Dimensional Modelling
- Schemas

2 mins Summary

THANK - YOU