WELCOME TO 7TH SEMESTER

COURSE NAME:
DATA WAREHOUSING & DATA MINING

LECTURE 1
TOPICS TO BE COVERED

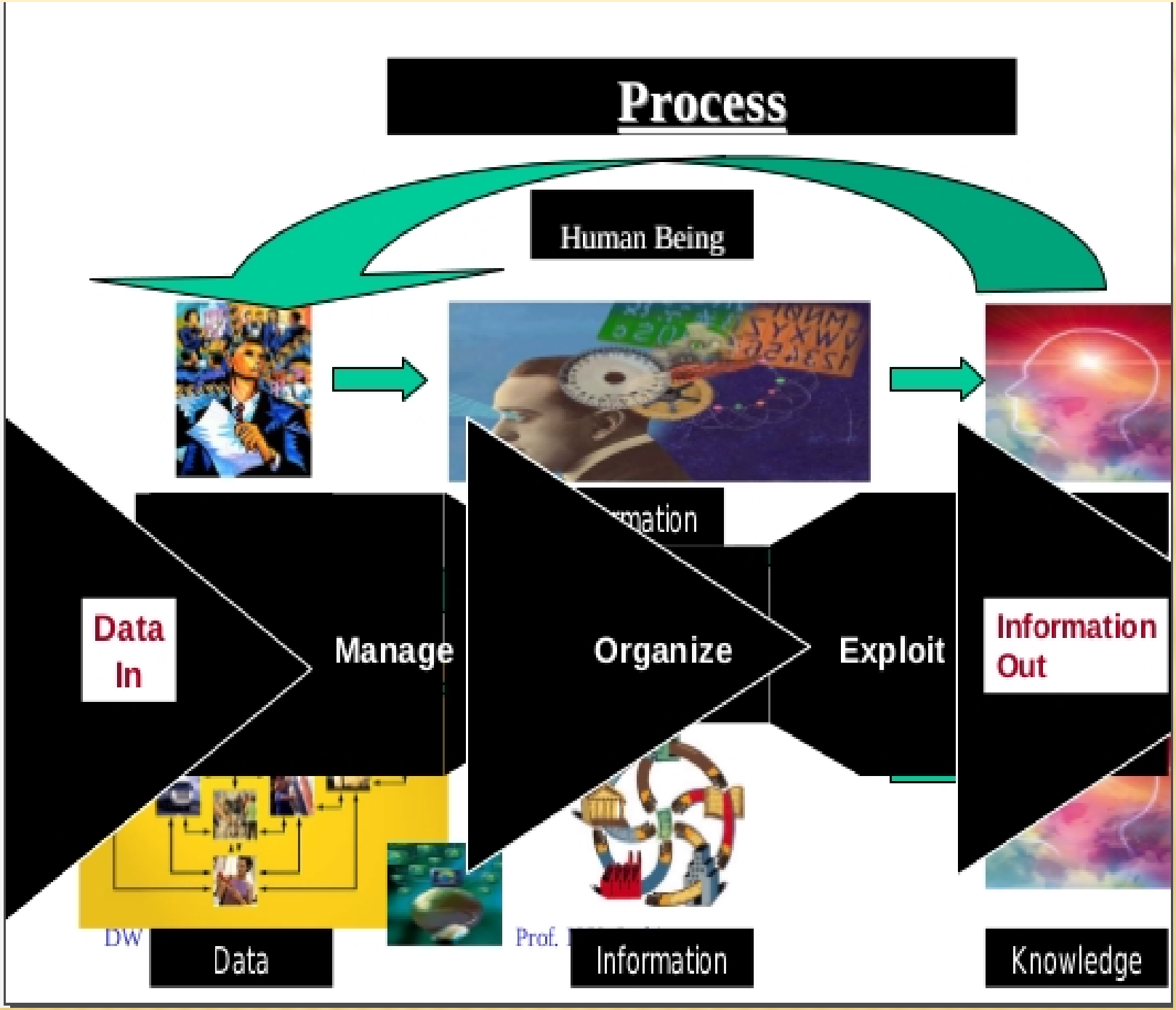
- Data, Information and Knowledge
- Operational and informational processing
- Data warehousing Definition usage and trends.
- DBMS vs data warehouse
- Data marts

DATA, INFORMATION & KNOWLEDGE

- Data is composed of observable and recordable facts that are often found in operational or transactional systems. Data are any facts, numbers or text that can be processed by a computer.

- Information is an integrated collection of facts and is used as the basis for decision making. The patterns, associations or relationships among all data can provide information.

- Information can be converted into knowledge about historical patterns and future trends.


Process
Human Being
Data In
Manage
Organize
Exploit
Information Out
Data
Information
Knowledge

OPERATIONAL & INFORMATIONAL PROCESSING

- 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.

OPERATIONAL VS INFORMATION SYSTEM

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 representation of state of the business	Historical point-in-time (snapshots) and predictions
Primary users	Clerks, 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

WHAT IS A DATA WAREHOUSE?

- The term "data warehouse" refers to a special type of database that acts as the central repository for company data. It can be thought of as a database archive that is segregated from the operational databases, and used primarily for reporting and data mining purposes.

- A fundamental concept of a data warehouse is the distinction between data and information.

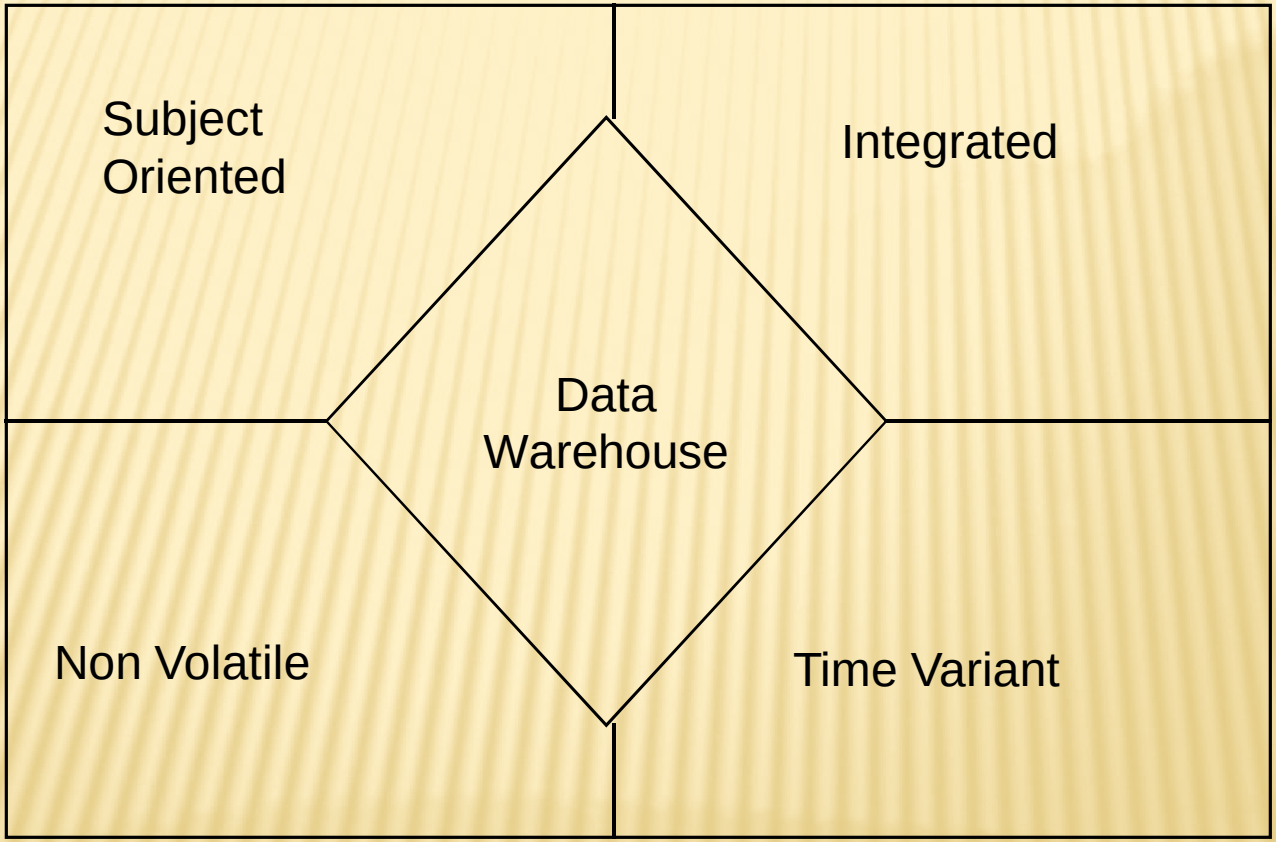
HISTORY

- Data warehouses were first developed in the 1980s in response to the growing demand for management information analysis, which operational databases could not perform without drastically affecting response time.

DATA WAREHOUSE

- Data warehouse is a subject oriented, integrated, time-variant and non-volatile collection of data in support of management decision making process.

DATA WAREHOUSE PROPERTIES



DATA WAREHOUSE PROPERTIES

- **Subject Oriented:-** It means that the data warehouse focuses on he modelling and analysis of data for decision makers.
- **Integrated:-** Datawarehouse is constructed by integrating multiple heterogeneous sources like relational database , flat files and OLT records.
- **Time-Variant:-** In datawarehouse we store the data to provide information from historical perspective.
- **Non-Volatile:-** It means that everything that we put in datawarehouse remains in one way or another.

DATA WAREHOUSING

- It is a process of constructing and using datawarehouse
- Construction of datawarhouse requires data integration and cleaning

USAGES AND TRENDS

- How the organizations use the information from data to obtain information?
- Many organizations use the information from data to obtain information.
- Many organizations use the information from data warehouse to support decision making activities.
- Increase customer focus, it means to analyze the customer buying patterns.
- Repositioning of products and managing product by comparing the performance quarterly or yearly.
- In heterogeneous database integration organizations usually collects diverse kind of data.
- Maintain large databases for this kind of work data warehouse sounds very effective .

DIFFERENCE BETWEEN DATA WAREHOUSE AND DBMS

Features	DBMS	DW
Characteristics	Operational Processing	Information Processing
Orientation	Transaction	Analysis
Users	Client ,clerks, IT professionals	Knowledge workers
Functions	Day to day	Long term information
DB Design	ER based, application oriented	Star ,Snowflake,subject – oriented.
Data	Current-guaranteed up to data	Historical accuracy maintained over time.
Summarization	Highly detailed	Summarized
Access	Read/write	Mostly read
DB size	100 mb to 1 gb	Above 1gb
View	Detailed	Multidimensional