FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-01
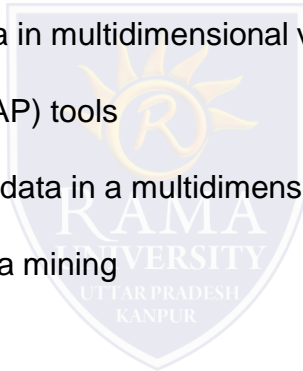
MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- ❖ **DATA WAREHOUSING - OVERVIEW**
- ❖ **DATA MINING-OVERVIEW**
- ❖ **UNDERSTANDING A DATA WAREHOUSE**
- ❖ **DATA WAREHOUSE VS OPERATIONAL DATABASES**
- ❖ **DATA WAREHOUSE FEATURES**
- ❖ **DATA WAREHOUSE APPLICATIONS**
- ❖ **DATA WAREHOUSE TYPES**
- ❖ **DIFFERENCE BETWEEN DATA WAREHOUSE AND OPERATIONAL DATABASE**
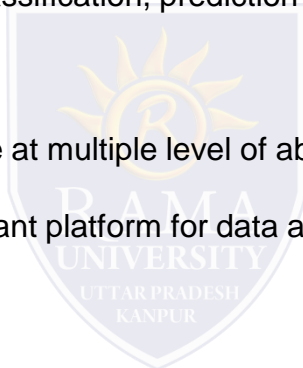- ❖ **MCQ**
- ❖ **REFERENCES**

DATA WAREHOUSING - OVERVIEW

- Term "Data Warehouse" was first coined by **Bill Inmon** in 1990Concept of Hard Computing

- a data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data

- helps analysts to take informed decisions in an organization

- provides us generalized and consolidated data in multidimensional view

- provides us Online Analytical Processing (OLAP) tools

- help us in interactive and effective analysis of data in a multidimensional space

- analysis results in data generalization and data mining

Data mining

- functions such as association, clustering, classification, prediction

- can be integrated with OLAP operations

- enhance the interactive mining of knowledge at multiple level of abstraction

- data warehouse has now become an important platform for data analysis and online analytical processing
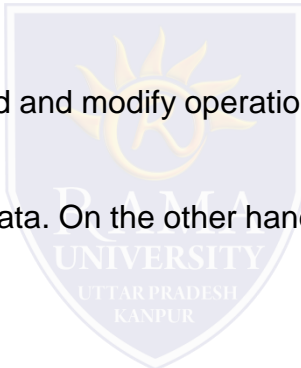
DATA WAREHOUSE

A data warehouse is a database,

- which is kept separate from the organization's operational database.

- Data warehouse systems help in the integration of diversity of application systems

- A data warehouse system helps in consolidated historical data analysis.

- helps executives to organize, understand, and use their data to take strategic decisions.

- possesses consolidated historical data, which helps the organization to analyze its business.

- no frequent updating done in a data warehouse.

DATA WAREHOUSE VS OPERATIONAL DATABASES

- An operational database is constructed for well-known tasks and workloads such as searching particular records, indexing, etc. In contract, data warehouse queries are often complex and they present a general form of data.

- Operational databases support concurrent processing of multiple transactions. Concurrency control and recovery mechanisms are required for operational databases to ensure robustness and consistency of the database.

- An operational database query allows to read and modify operations, while an OLAP query needs only read only access of stored data.

- An operational database maintains current data. On the other hand, a data warehouse maintains historical data.

DATA WAREHOUSE FEATURES

- **Subject Oriented** –
  - provides information around a subject rather than the organization's ongoing operations.
  - Example: product, customers, suppliers, sales, revenue, etc.
  - does not focus on the ongoing operations

  - focuses on modelling and analysis of data for decision making.

- **Integrated** –
  - integrating data from heterogeneous sources
  - such as relational databases, flat files, etc.
  - integration enhances the effective analysis of data.

- **Time Variant** –
  - data warehouse is identified with a particular time period.
  - provides information from the historical point of view.

- **Non-volatile** –
  - previous data is not erased when new data is added to it.
  - kept separate from the operational database
  - frequent changes in operational database is not reflected in the data warehouse.
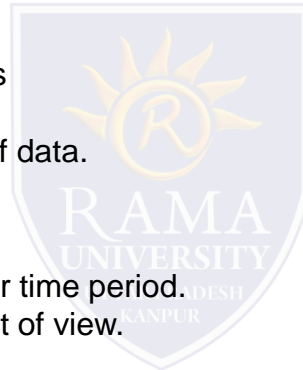
DATA WAREHOUSE APPLICATIONS

- helps business executives to organize, analyze, and use their data for decision making.

- serves as a sole part of a plan-execute-assess "closed-loop" feedback system for the enterprise management.

- widely used in the following fields −

  - Financial services

  - Banking services

  - Consumer goods

  - Retail sectors

  - Controlled manufacturing
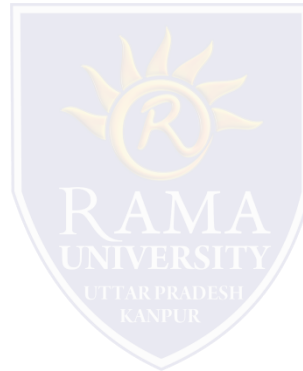
DATA WAREHOUSE TYPES
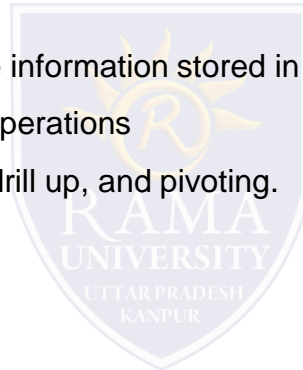
- ## Information Processing –

  - A data warehouse allows to process the data stored in it.

  - querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs.

- ## Analytical Processing –

  - supports analytical processing of the information stored in it.

  - analyzed by means of basic OLAP operations

  - including slice-and-dice, drill down, drill up, and pivoting.

- ## Data Mining –

  - supports knowledge discovery by finding hidden patterns and associations,

  - constructing analytical models, performing classification and prediction.

  - presented using the visualization tools.

DIFFERENCE BETWEEN DATA WAREHOUSE AND OPERATIONAL DATABASE

Sr.No.	DATA WAREHOUSE (OLAP)	OPERATIONAL DATABASE(OLTP)
1	It involves historical processing of information.	It involves day-to-day processing.
2	OLAP systems are used by knowledge workers such as executives, managers, and analysts.	OLTP systems are used by clerks, DBAs, or database professionals.
3	It is used to analyze the business.	It is used to run the business.
4	It focuses on Information out.	It focuses on Data in.
5	It is based on Star Schema, Snowflake Schema, and Fact Constellation Schema.	It is based on Entity Relationship Model.
6	It focuses on Information out.	It is application oriented.
7	It contains historical data.	It contains current data.

Multiple Choice Question

1. A goal of data mining includes which of the following?
a) To explain some observed event or condition
b) To confirm that data exists
c) To analyze data for expected relationships
d) To create a new data warehouse

2.. A data warehouse is which of the following?
a) Can be updated by end users.
b) Contains numerous naming conventions and formats.
c) Organized around important subject areas.
d) Contains only current data.

3. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions.
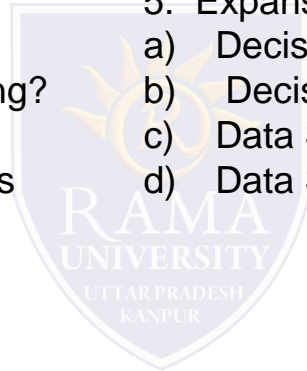a) Data Mining
b) Data Warehousing
c) Web Mining
d) Text Mining.

4.The data Warehouse is_____.
a) read only
b) write only.
c) read write only.
d) none.

5. Expansion for DSS in DW is_____.
a) Decision Support system.
b) Decision Single System.
c) Data Storable System
d) Data Support System

REFERENCES

https://www.tutorialspoint.com/dwh/dwh_overview.htm