Date : ...../...../..... Expt. No. ..... Page No. .....

White Page

Data Mining & Data Warehousing

Teacher's Signature: _____

Data Mining ⇒ The process of extracting information to identify patterns, trends, and useful data that would allow the business to take the data-driven decision from huge sets of data is called data-mining.

⇒ Data mining refers to the analysis of data. It is the computer supported process of analyzing huge sets of data that have either been compiled by computer systems or have been downloaded into the computer.

In the data mining process, the computer analyzes the data and extract useful information from it.

It looks for hidden patterns within the data set and try to predict future behaviour.

Data mining is primarily used to discover and indicate relationships among the data sets.

⇒ Data Mining is one of the most useful techniques that help enterprenures, researchers and individuals to extract valuable information from huge sets of data.

Teacher's Signature: _____

⇒ Data mining is also called Knowledge Discovery in Database (KDD).

⇒ The Knowledge Discovery (KDD) process includes Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation, and Knowledge presentation.

⇒ Data Mining is a process used by organizations to extract specific data from huge databases to solve business problems. It primarily turns raw data into useful information.

Data Warehouse ⇒ A data warehouse refers to a place where data can be stored for useful mining.

It is like a quick computer system with huge data storage capacity.

Data from various systems are copied to the warehouse.

⇒ Data Warehouse combines data from numerous source which ensure the data quality, accuracy and consistency.

Data flows into a data~~base~~ warehouse from different databases.

A data warehouse works by sorting out data into a ~~fi~~ pattern that depicts the format and types of data.

⇒ Data Warehouses and Databases both are relative data systems, but both are made to serve different purposes.

A data warehouse is used to store a huge amount of historical data and empowers fast requests over all the data, typically using Online Analytical Processing (OLAP)

Teacher's Signature: _____

⇒ A database is made to store current transactions and allow quick access to specific transactions for ongoing business process, commonly known as Online Transaction Processing (OLTP).

⇒ A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision making process.
W. H. Inmon, 1993

⇒ Data warehouse is non-volatile. In other words, it contains only read-only data.

⇒ Database is a collection of related data that represents some elements of the real world whereas Data Warehouse is an information system that stores historical and commutative data from single or multiple sources.

Teacher's Signature: _____

White Page

Standard Definition of Data Warehouse

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

W. H. Inmon, 1993

Subject-Oriented ⇒ Cover all the individual applications concerning an organization.

The data warehouse is oriented around the major subjects of an enterprise, e.g. customer, vendor, product & productivity.

Integrated ⇒ The data collected in data warehouse are taken from different sources. These data are integrated in a singular, globally acceptable fashion.

Data warehouse become more successful when data from multiple sources are combined in such a way that data consistencies and conflicts are effectively addressed. This process is known as data scrubbing.

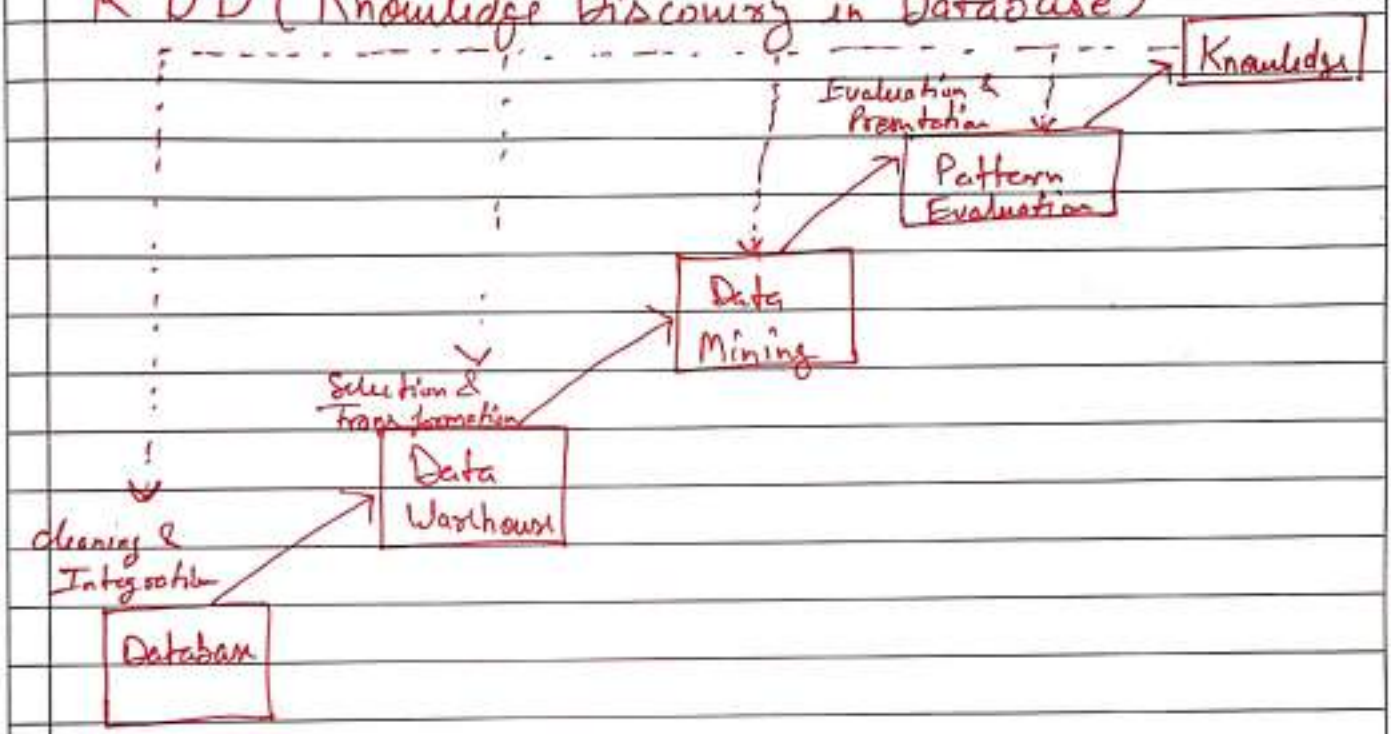
Time-Variant ⇒ The data in a data-warehouse is time-dependent. The data warehouse focus on changes over time.

Teacher's Signature: _____

Non-Volatile ⇒ The data warehouse is a non-volatile. In other words, it contains read-only data.

Once the data is loaded into the warehouse from an application-oriented environment or external sources, it does not change and is only available for accessing.

KDD (Knowledge Discovery in Database)



Data Cleaning ⇒ It is the process of removing noisy or inconsistent data.

⇒ Cleaning is performed for detection of syntax error.

⇒ Parser decides weather the given string of data is acceptable within data specification.

Data Integration ⇒ Data from multiple sources are combined.

Data Selection ⇒ Useful and relevant data are retrieved from the database for analysis.

Teacher's Signature: _____

Data Transformation ⇒ Data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance.

Data Mining ⇒ It is a process where intelligent methods are applied in order to extract data patterns.

Pattern Evaluation ⇒ To identify the truly interesting patterns representing knowledge-base on some specific criteria.

Knowledge Representation ⇒ Where visualization and knowledge representation techniques are used to present the mined knowledge

Teacher's Signature: _____

Types of Data Mining

Data mining can be performed on the following types of data -

1. Relational Database
2. Data Warehouse
3. Data Repositories
4. Object-Relational Database
5. Transactional Database

Advantages of Data Mining

Disadvantages of Data Mining

Applications of Data Mining

These are the following areas where data mining is widely used -

1. Data Mining in Healthcare
2. Data Mining in Market Based Analysis
3. Data Mining in education
4. Data Mining in Manufacturing engineering
5. Data Mining in Customer Relationship Management (CRM)
6. Data Mining in Fraud Detection
7. Data Mining in Lic Detection.
8. Data Mining in Financial Banking

Teacher's Signature: _____

Challenges in Implementation of Data Mining.

1. Incomplete and noisy data.
2. Data Distribution
3. Complex data.
4. Performance
5. Data Privacy & Security
6. Data Visualization

Data Mining Techniques

These are some important data mining techniques -

1. Classification
2. Clustering
3. Regression
4. Association Rules
5. Outer Detection
6. Sequential Patterns
7. Prediction

Teacher's Signature: _____