Unit 5. [D] Unbeaten learning

Page No.
Date: / /

↷ Net

① What is data Mining Explain ?

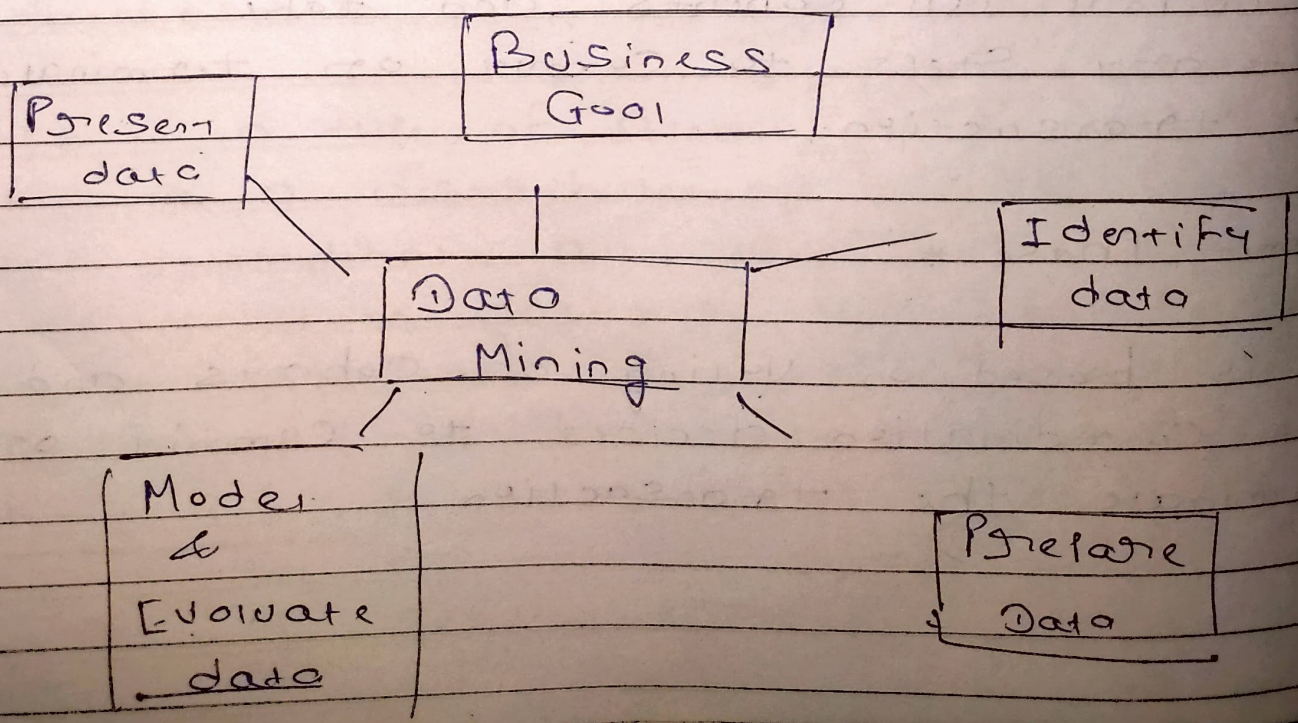
→

① Data mining is one of the most useful techniques that help entrepreneurs, researchers and individuals to extract valuable information from huge set of data.

② Data mining also called Knowledge discovery in database. (KDD)

③ The knowledge discovery process including Data Cleaning, Data integration, Data Selection, Data transformation.

④ Data mining utilize complex mathematical algorithm of data segment and evaluates the probability of future event.



Page No.
Date: / /

(5) The data can be structured, Semi-structured or unstructured and can be stored in various forms such as database, data warehouse and data lakes.

(6) The primary goal of data ~~mining~~ mining is to discover hidden patterns and relationships in the data that can be used to make informed decisions or predictions.

* Next four step from book *

BOOK

1. Data Selection
2. Data cleaning
3. Data Mining
4. Evaluation

[1] Unbalanced Learning

Page No.
Date: / /

2) What is data Warehousing explain?

→ ① ① Data warehouses are relational environments that are used for data analysis, particularly of historic data

② Organizations use data warehouses to disc Patterns and relationships in their dat that develop over time.

③ ① Data warehouse is a relational database management system. Construct to meet the requirement of transaction Processing system.

Next

3) Explain metadata in data warehousing ?

→

① Metadata is data that describe and contextualies other data.

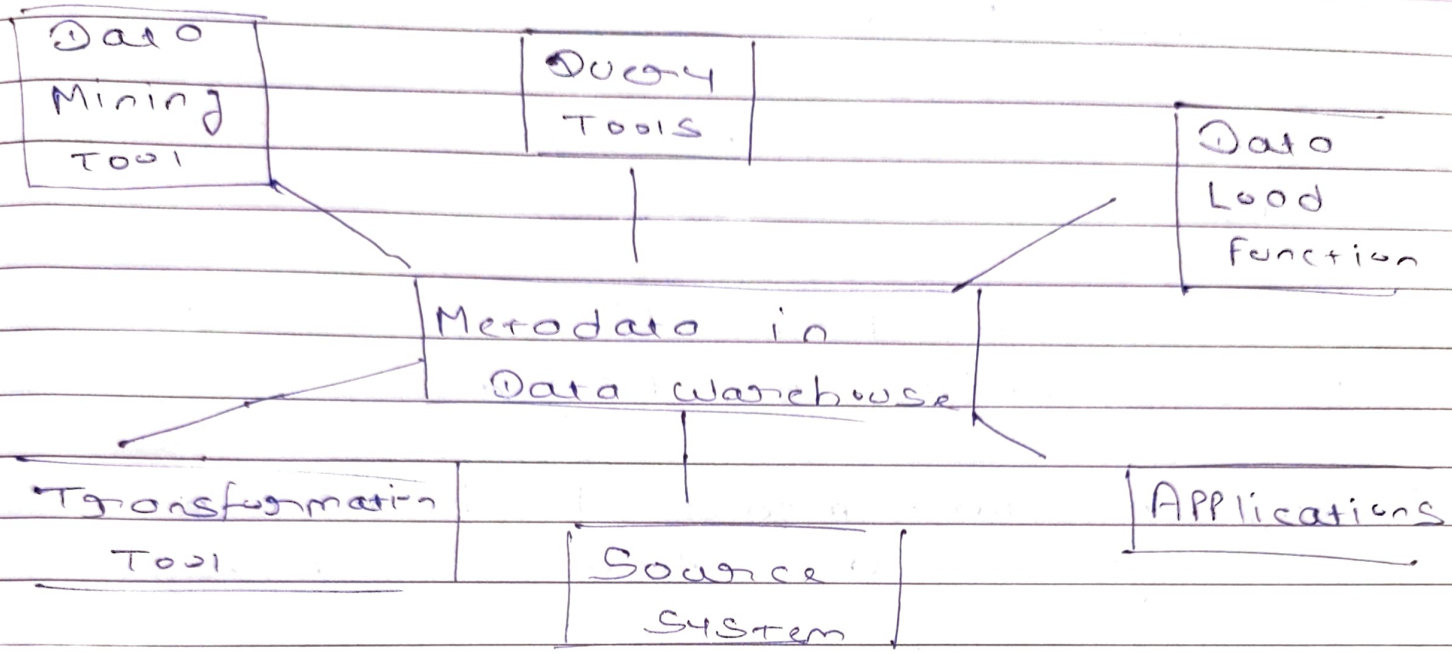
② It provides information about the content format, structure and other characteristics of data and can be used to improve the organisation, discoverability and accessibility of data.

③ Metadata can be stored in various forms such as text, XML or RDF and can be organized using metadata standards and schemes.

④ There are many metadata standards that have been developed to facilitate the creation and management of metadata

⑤ Metadata schemas define the structure and format of metadata and provide consistent framework for organizing and describing data

⑥ Metadata can be used in a variety of contexts such as libraries, museums, archives and online platform.



* Mainly three categories of Metadata *

① Business Metadata

It has the data ownership information business definition and changing Policies.

② Technical Metadata

① It include database system names, table and columns names and sizes, data types and allowed values

② Technical metadata also includes structural information such as Primary and Foreign Key attributes and indices

③ Operational Metadata.

① It includes currency of data and data lineage.

② currency of data means Wheather the data is active, archived or Purged.

③ Lineage of data means the history of data migrated and transformation applied on it.

---

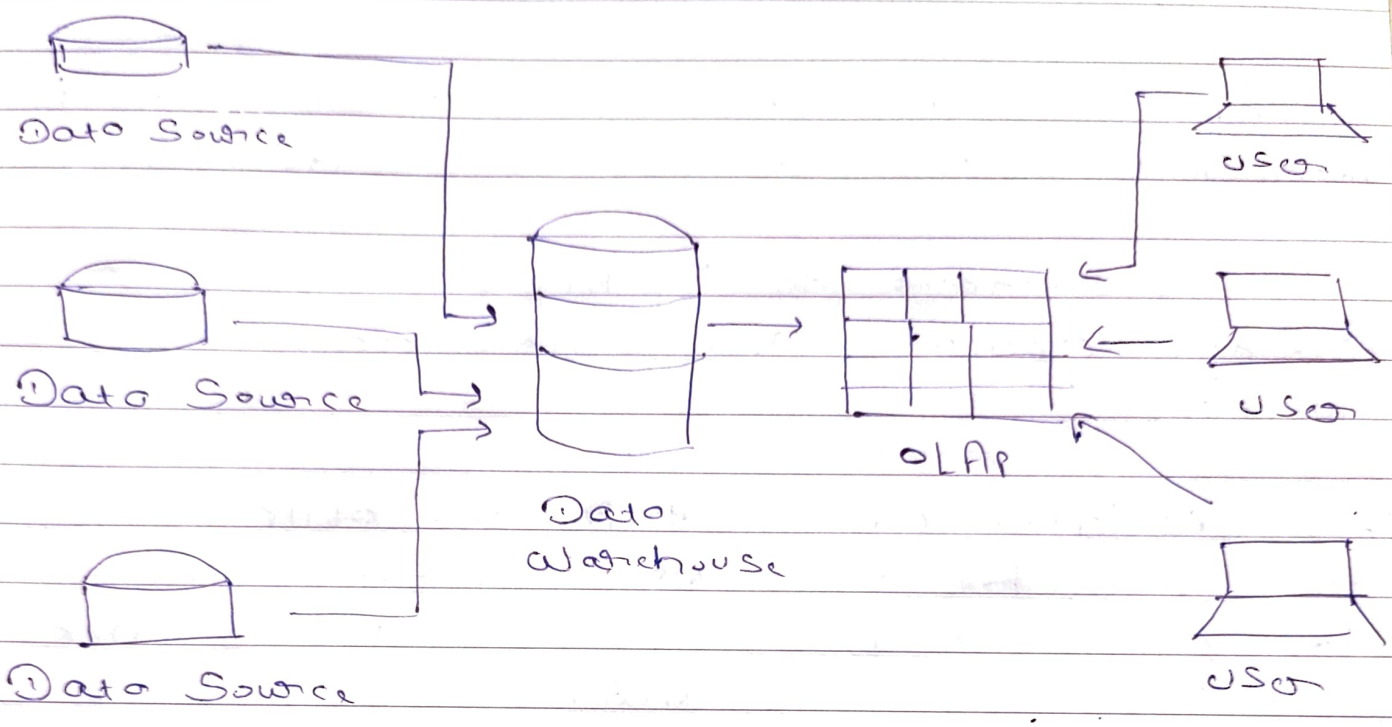
④ OLAP Explain

① OLAP Stands For online Analytical Processing Server.

② It is a Software Technology that allows user to analyze information From multi database System at the Some time.

③ It is based on multidimensional data model and allow the user to query on multi-dimensional data.

④ OLAP is a database analysis technology that involves querying, extracting and Studying Summarized data



⑤ It allows managers and analysts to get an insight of the information through fast, consistent and interactive access to information.

⑥ OLAP can be used for data mining or the discovery of previously undiscovered relationships between data items.

⑦ There are mainly three type of OLAP System.

i) Multidimensional OLAP (MOLAP)
ii) Relational OLAP (ROLAP)
iii) Hybrid OLAP (HOLAP)

① Multidimensional OLAP (MOLAP)

① MOLAP is OLAP that indexes directly into multidimensional database

② The advantage of MOLAP are that it can manage analyze and store data very quickly

③ It is also easist type of OLAP to use and can be very useful helpful when an analytic and business leader need to sort data in variety ways

② Relational OLAP (ROLAP)

① ROLAP works directly with the data in relational database without first reorganizing into cube.

② ROLAP is that better suited to analyzing large quantities of data to answer very specific question.

③ Hybrid OLAP

① HHbrid OLAP preferred to organizes data into OLAP cubes but also works directly with data in a relational database

Page No.
Date : / /

② H-LAP it combines the speed of MOLAP with the scalability of ROLAP So it can quickly analyze large volumes of detailed data in variety of ways.