



SNS COLLEGE OF ENGINEERING
Coimbatore-107



UNIT III

DATA MINING

Introduction



DATA MINING

Data mining refers to **extracting or mining knowledge** from large amounts of data stored either in databases, data warehouses or other information repositories.

In other words, “UNCOVERING THE HIDDEN INFORMATION”
- DATA MINING

ALTERNATIVE NAMES:

- KDD (knowledge discovery in databases)
- Knowledge extraction
- Data/pattern analysis
- Information harvesting



Applications

- **Banking: loan/credit card approval**
 - predict good customers based on old customers
- **Customer relationship management:**
 - identify those who are likely to leave for a competitor.
- **Targeted marketing:**
 - identify likely responders to promotions
- **Fraud detection: telecommunications, financial transactions**
 - from an online stream of event identify fraudulent events
- **Manufacturing and production**



Applications

- Medicine: disease outcome, effectiveness of treatments
 - analyze patient disease history: find relationship between diseases
- Molecular/Pharmacy: identify new drugs
- Scientific data analysis:
 - identify new galaxies by searching for sub clusters
- Web site/store design and promotion:
 - find affinity of visitor to pages and modify layout



KDD

- Knowledge discovery in databases
- Formalized in 1989 in the pursuit of **seeking knowledge from the data.**
- KDD tends to be highly iterative and interactive
- Data mining is a step in the KDD process



KDD

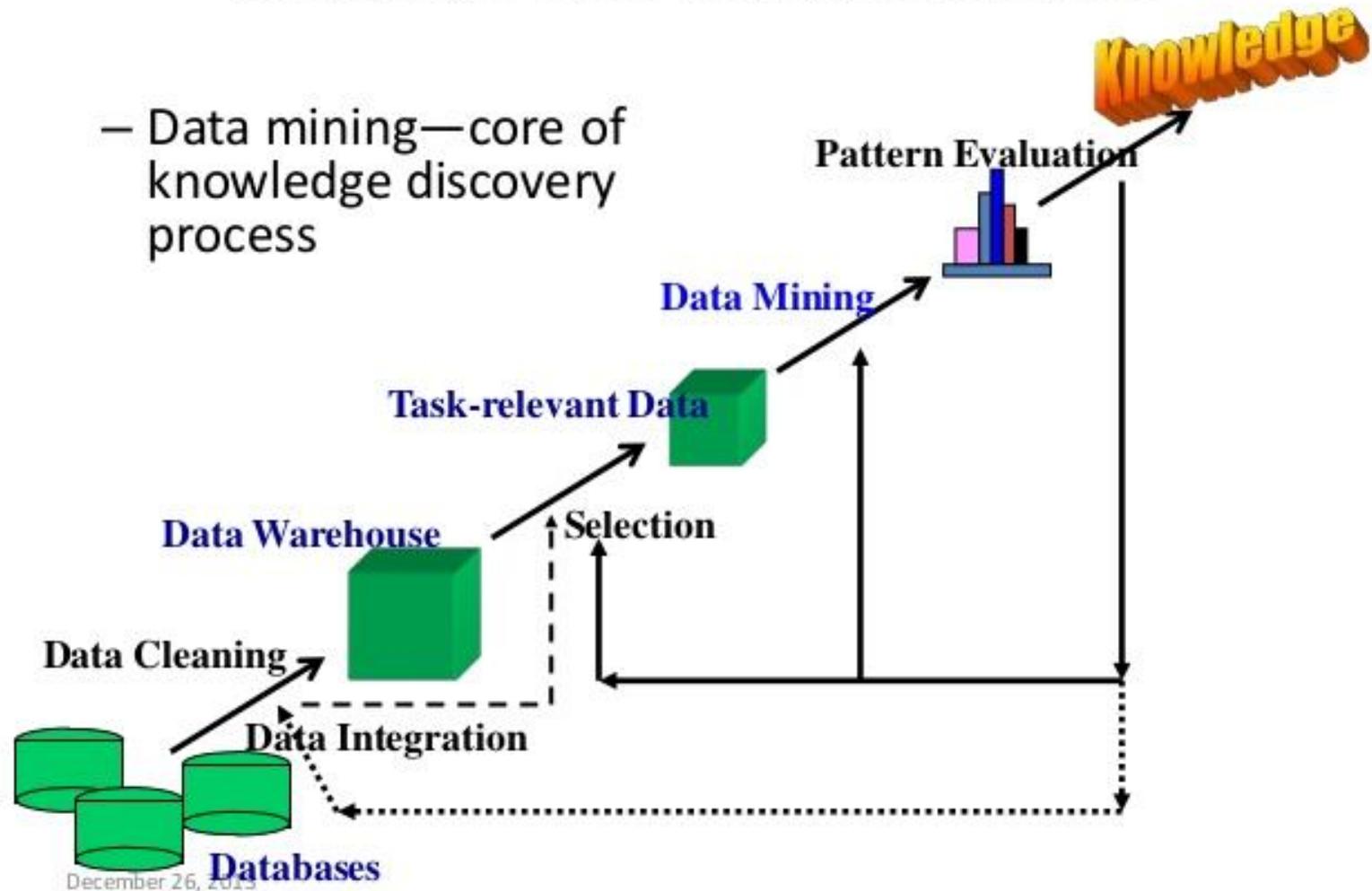
Steps in KDD process:

- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern evaluation
- Knowledge presentation



Knowledge Discovery (KDD) Process

– Data mining—core of knowledge discovery process





Steps in KDD process

- Data cleaning – process of filling in the missing values, identifying or removing outliers and resolving inconsistencies in data
- Data integration – process of combining data from multiple sources
- Data selection – process of retrieving relevant data from the databases.



Steps in KDD process

- Data transformation – data are **transformed** or **consolidated** into forms appropriate for mining by performing summary or aggregation operations
- Data mining – It is an essential process where intelligent methods are applied to extract data patterns

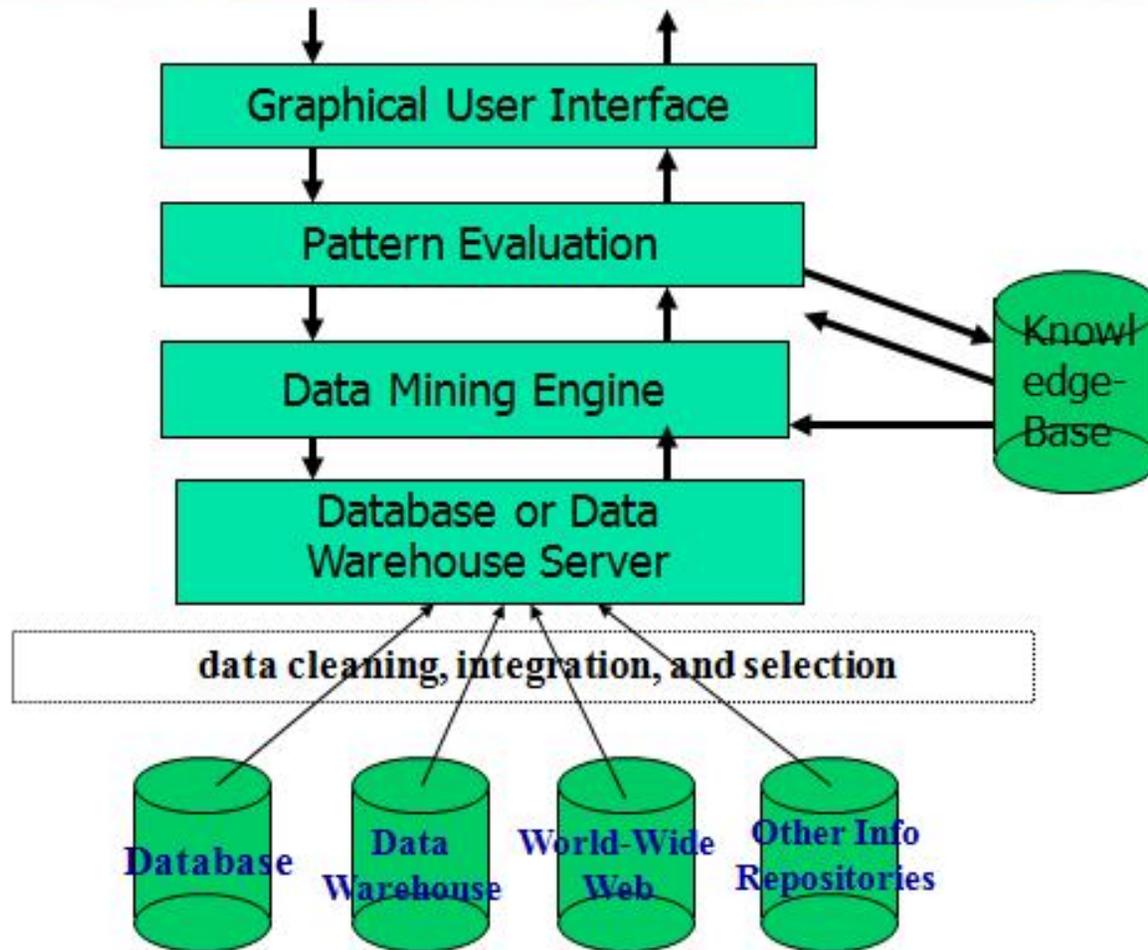


Steps in KDD process

- Pattern evaluation – **patterns** obtained in DM stage are converted into **knowledge** based on some interesting measures
- Knowledge presentation - **visualization** techniques are used to present the mined knowledge to the user.



Architecture: Typical Data Mining System





DM SYSTEM

- **Database, data warehouse or other information repository:** This is a single or a collection of multiple databases, data warehouse, flat files, spread sheets or other kinds of information repositories
- **DB or DW server:** The server fetches the relevant data, based on the users data mining request



DM SYSTEM

- **Knowledge base** – used to guide the search or evaluate the patterns
- **DM engine** – consists of a set of functional modules for tasks such as association, classification, cluster analysis, etc
- **Pattern Evaluation module** – this module interacts with DM module to focus the search towards increasing patterns



DM SYSTEM

- **GUI** – this module allows the users to interact with the system by specifying a task or data mining query, evaluate mined patterns and visualize the pattern in different forms such as maps, charts etc.



Thank You