



Data Mining

by

Dr. P.S. RAJPUT

MEANING

- ◆ Data mining ,the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information from the data warehouse. Data mining returns to the mining or discovery of new information in terms of pattern or from vast amount of data.

What is DM

- ◆ DM refers to mining knowledge from large amount of data.
- ◆ The science of extracting useful information form large data sets or databases.
- ◆ DM is an essential step in the process of knowledge discovery in databases.

Data Mining: contd...

- ◆ Term commonly used in computer Sc.
- ◆ DM is a sub field in Computer Sc and closely associated with applied statistics
- ◆ DM is finding similarities and patterns in large sets of data using algorithms
- ◆ DM often used to apply two separate processes of Knowledge discovery and Prediction

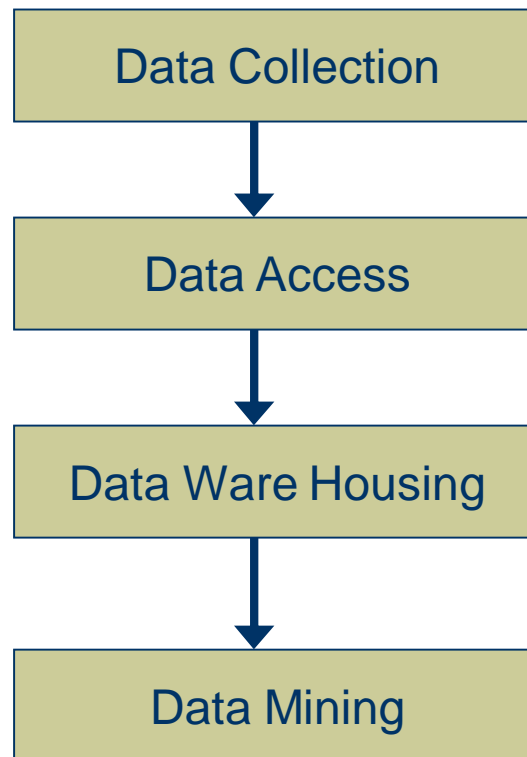


Other Names for DM



- Knowledge Discovery in Databases(KDD)
- Knowledge Extraction
- Data pattern analysis
- Data archaeology
- Data dredging
- Information harvesting

EVOLUTION



IMPORTANCE

- ◆ Availability of huge amount of data and need for using such data for useful applications.
- ◆ As a result of natural evolution of IT
- ◆ The fast growing amount of data stored in data bases has exceeded our human ability for comprehension without powerful tools

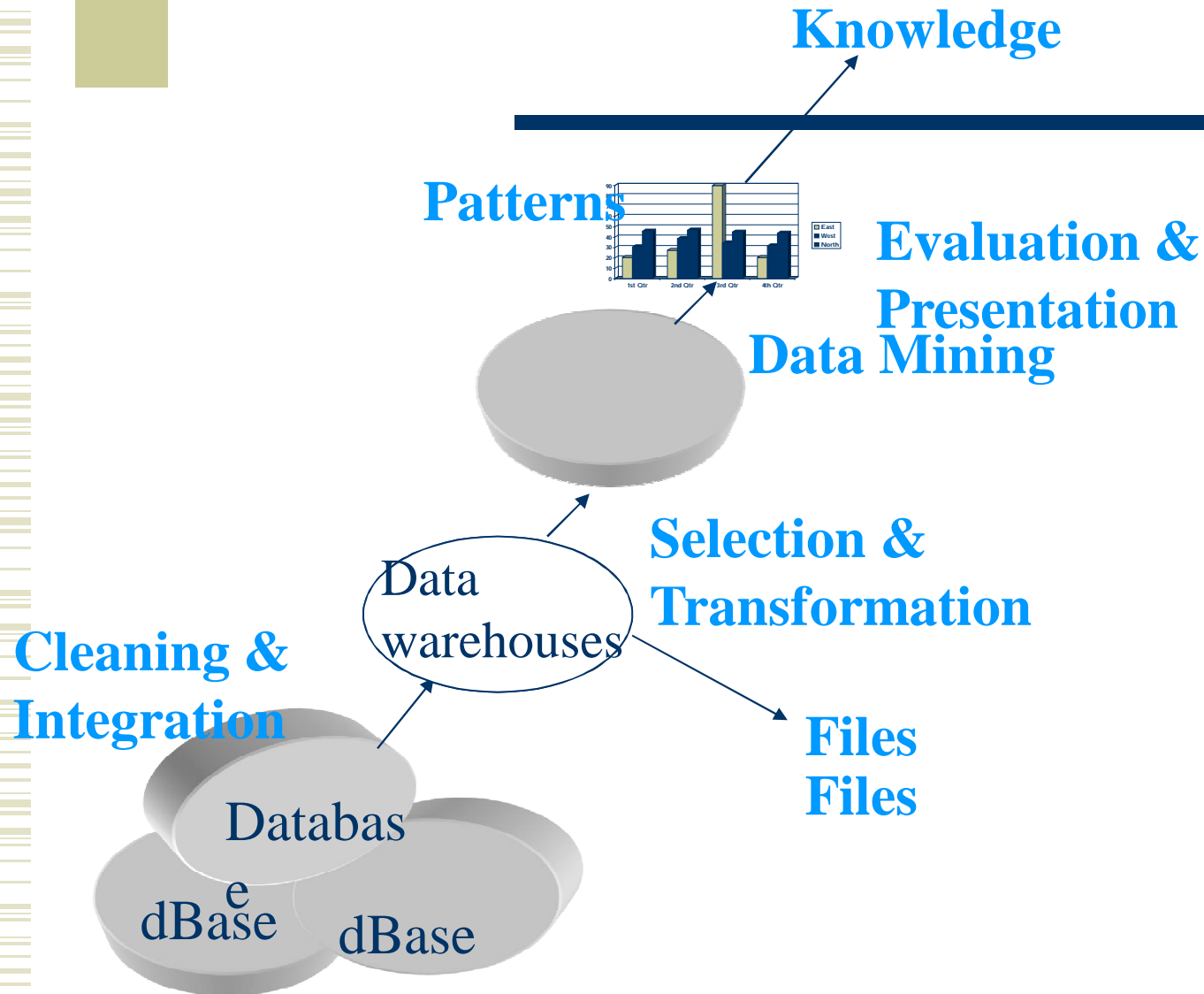
Cont...

- ◆ Information management
- ◆ Production control
- ◆ Wide availability of huge amount of data
- ◆ Automated prediction of trends and behavior
- ◆ Automated discovery of previously unknown patterns
- ◆ Large database could be used searching in-depth.

Steps of the DM

1. **Data gathering:** e.g. Data warehousing
2. **Data cleansing:** errors/ inconsistent data
3. **Data selection:** Where data relevant to the analysis task are retrieved from the dbase
4. **Data transformation**
5. **Data mining:** An essential process where intelligent methods are applied in order to extract data
6. **Pattern evaluation:** to identify the truly interesting patterns representing knowledge
7. **Knowledge presentation:** Visualization techniques used to present the mined knowledge to the user.

Steps ...cont.





PROBLEMS IN DM



- ◆ Lack of standards
- ◆ Unproven to libraries
- ◆ Technical hurdles
- ◆ Inappropriateness of mining tools
- ◆ Lack of skills



THANKS