RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-17

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- TUNING THE DATA WAREHOUSE
- TESTING THE DATA WAREHOUSE
  - THREE LEVELS OF TESTING
  - DEVELOPING THE TEST PLAN
  - TESTING BACKUP RECOVERY
  - TESTING THE OPERATIONAL ENVIRONMENT
  - TESTING THE DATABASE
  - TESTING THE APPLICATION
- MCQ
- REFERENCES

OVERVIEW

- huge amount of data available in the Information Industry.
- This data is of no use until it is converted into useful information.
- necessary to analyze this huge amount of data and extract useful information from it.
- Extraction of information is not the only process we need to perform
- data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation.
- Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.

Data Mining

❑ **Data Mining** is defined as extracting information from huge sets of data. In other words

❑ procedure of mining knowledge from data.

❑ The information or knowledge extracted so can be used for any of the following applications −

- ✓ Market Analysis
- ✓ Fraud Detection
- ✓ Customer Retention
- ✓ Production Control
- ✓ Science Exploration

Applications

- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection
- also be used in the areas of production control, customer retention, science exploration, sports, astrology, and Internet Web Surf-Aid

Market Analysis and Management

Customer Profiling −

- helps determine what kind of people buy what kind of products.

Identifying Customer Requirements −

- helps in identifying the best products for different customers.
- It uses prediction to find the factors that may attract new customers.

Cross Market Analysis −

- performs Association/correlations between product sales.

Target Marketing −

- helps to find clusters of model customers who share the same characteristics such as interests, spending habits, income, etc.

Determining Customer purchasing pattern −

- helps in determining customer purchasing pattern.

Providing Summary Information −

- provides us various multidimensional summary reports.

Analysis and Risk Management

Finance Planning and Asset Evaluation −

- It involves cash flow analysis and prediction, contingent claim analysis to evaluate assets.

Resource Planning −

- It involves summarizing and comparing the resources and spending.

Competition −

- It involves monitoring competitors and market directions.

Fraud Detection

Data mining is also used in the fields of credit card services and telecommunication to detect frauds. In fraud telephone calls, it helps to find the destination of the call, duration of the call, time of the day or week, etc. It also analyzes the patterns that deviate from expected norms.

Multiple Choice Question

1. A directory to help the DSS analyst locate the contents of the data warehouse is seen in _____.
a) Current detail data.
b) Lightly summarized data.
c) Metadata
d) Older detail data.

2.. Metadata contains atleast _____.
a) the structure of the data.
b) the algorithms used for summarization.
c) the mapping from the operational environment to the data warehouse.
d) all of the above.

3. Which of the following is not a old detail storage medium?
a) Phot Optical Storage.
b) RAID
c) Microfinche
d) Pen drive.

4. The data from the operational environment enter _____ of data warehouse.
a) Current detail data.
b) Older detail data.
c) Lightly summarized data.
d) Highly summarized data.

5. The data in current detail level resides till _____ event occurs.
a) purge
b) summarization
c) achieved
d) all of the above.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture
- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education
- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier
- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education
- Mallach," Data Warehousing System", TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.