

Data Warehousing Concepts

Lesson 6: Data Mining

Lesson Objectives

- In this lesson, you will learn about:
 - Online Analytical Processing
 - Data Mining
 - The Knowledge Discovery Process
 - Why Use Data Mining Today?
 - Data Mining Usage
 - Data Mining and Business Intelligence
 - Types of Data used in Data Mining
 - Data Mining Applications



Lesson Objectives

- In this lesson, you will learn about (contd.):
 - Data Mining Products
 - Mining market



6.1: Data Mining

What is Data Mining?

- Data Mining is:
 - Subset of BI.
 - Extraction of necessary information from data in large databases.
 - Process of analyzing large databases to find valid, novel, useful, and understandable patterns.
 - Process of efficient discovery in large databases and Data warehouses.



Copyright © Capgemini 2015. All Rights Reserved 4

Data Mining:

- **Data mining** is the way of analyzing data by exploring large databases. It helps in understanding the business by extracting necessary information from the databases. It allows you to understand the pattern and helps in predicting the behavior of it.
- Data mining helps in increasing the business and forecasting the chunks related to it at early stages. It includes finding patterns that are suitable for the organization.

6.2: The Knowledge Discovery Process (KDD)

Concept of KDD

- Data Mining is also known as Knowledge Discovery in Databases (KDD).
 - It involves mining on different kinds of data.
 - It is a process of using raw data.
 - It is a collection of powerful techniques.
 - It refers to the automated extraction of hidden information from databases.
 - It helps customers to detect previously undetected facts present in their business critical data.

The Knowledge Discovery Process (KDD):

- **Data Mining** involves mining on different kinds of data such as Relational databases, Data warehouses, Transactional databases, Advanced DB systems and information repositories, Object-oriented and object-based databases, Text databases and multimedia databases, Heterogeneous and legacy databases. Data mining is the process of using raw data to infer important business information. It is a collection of powerful techniques for analyzing large amounts of data. Data mining tools can access data directly in the Data Warehouse.
- The advantage of mining is that no separate copy of data is needed for data mining. Data may not be organized in a way that is efficient for the tool. Data Mining is done by running a software that examines a database and looks for patterns in the data. Data Mining will not tell users about patterns in data that users may not have thought about. Data mining is used to try and mine key information from a Data warehouse to find **patterns in data**. Data mining allows organizations to collect information and make themselves more productive and beat their competitors.

6.3: Need of Data Mining

Why Use Data Mining Today?

- Human analytical skills are inadequate when:
 - Volume and dimensionality of the data increases.
 - Data growth rate is high.
- Data Mining is used for availability of:
 - Data
 - Data Storage
 - Computation of Data



Copyright © Capgemini 2015. All Rights Reserved 6

Need for Data Mining:

- Data mining is essential because of the following utilities:
 - Data mining helps to identify why customers buy certain products.
 - Data mining provides the ideas for very direct marketing.
 - Data mining provides the ideas for shelf placement.
 - It helps for training of employees versus employee retention.
 - It helps to identify employee benefits.

6.4: Use of Data Mining

Usage

- Here are some instances where Data Mining is essential:
 - The US Government needs to track fraudulent events.
 - A Supermarket is aspiring to become an information broker.
 - Basketball teams need it to track game strategy.
 - Cross Selling
 - Target Marketing
 - Holding on to good customers
 - Weeding out bad customers

Usage Scenarios

- Data warehouse mining is used in the following scenarios:
 - Assimilate data from operational sources
 - Mine static data
 - Mining log data
 - Continuous mining in process control
- Stages in mining:
 - Data selection -> Pre-processing-> cleaning -> Transformation -> Mining -> Result evaluation -> Visualization



Copyright © Capgemini 2015. All Rights Reserved 8

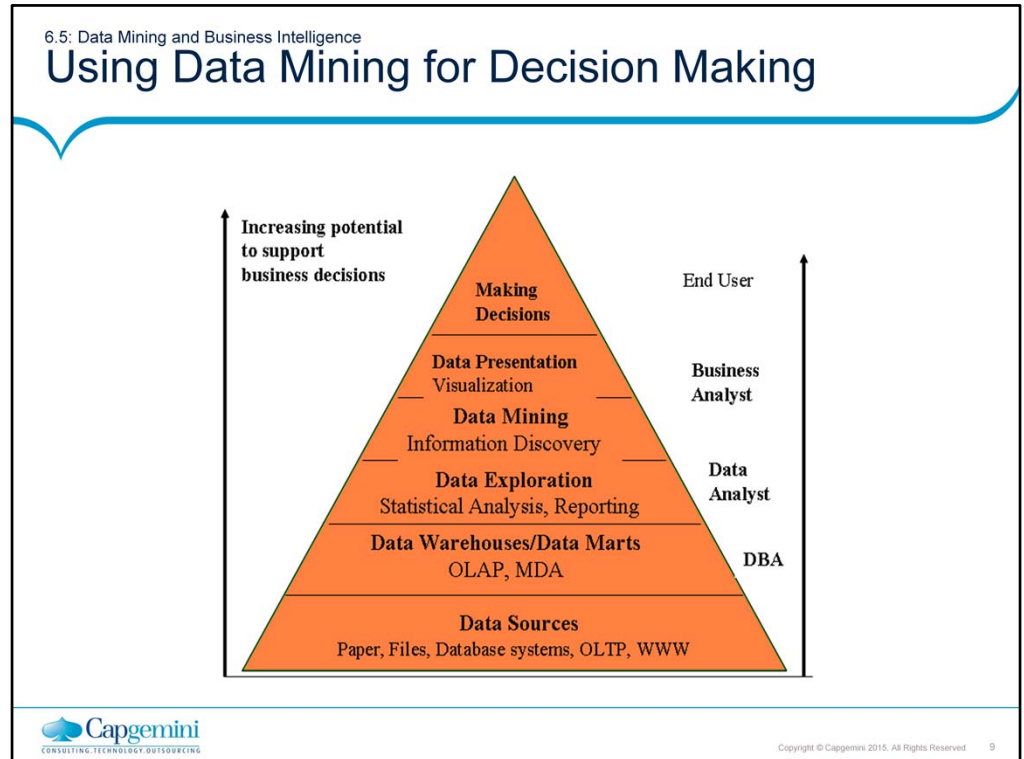
Use of Data Mining:

Usage scenarios:

- Data warehouse mining assimilates data from operational sources.
- Data warehouse mining mines static data.
- Mining log data.
- Continuous mining in process control.

Stages in mining:

1. Data selection
2. Pre-processing: cleaning
3. Transformation
4. Mining
5. Result evaluation
6. Visualization



Data Mining and Business Intelligence:

- Data Mining has grown drastically in many businesses. Data Mining has become very popular since it helps in increasing organization's profit and achieving the target.
- When Data mining gets involved in Business Intelligence, it actually helps in understanding the functionality of the organization. It helps in increasing the potential for supporting the business decisions. It makes the data visible in a visual form to the business analysts. It helps in exploring data in terms of reporting and statistical analysis.
- Data Mining along with Business Intelligence takes the following steps in logical progression:
 - **Data Source:** Typically data is sourced from transaction processing systems (Manufacturing, ERP, Sales).
 - **Data Marts (OLAP & MDA)/DBA:** You may want to customize your warehouse's architecture for different groups within your organization. You can do this by adding data marts, which are systems designed for a particular line of business.
 - **End User/Making Decision:** The principle purpose of Data warehousing is to provide information to the business user for strategic decision making.

6.6: Types of Data

Types of Data used in Data Mining

- The following types of data is drilled in Data Mining:
 - Relational data and transactional data
 - Spatial and temporal data, spatio-temporal observations
 - Time-series data
 - Text
 - Images, video
 - Mixtures of data
 - Sequence data
 - Features from processing other data sources

6.7: Data Mining Applications

Examples of Data Mining Applications

- Here are some examples of Data Mining Applications:
 - Banking: Loan / Credit card approval
 - Customer Relationship Management
 - Targeted marketing
 - Fraud detection: Telecommunications, Financial transactions
 - Manufacturing and Production
 - Web site/store design and promotion



Copyright © Capgemini 2015. All Rights Reserved 11

Data Mining Applications:

Let us discuss some examples of Data Mining 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 events, identify fraudulent events.
- Manufacturing and production:
 - Automatically adjust knobs when process parameter changes.
- Web site/store design and promotion:
 - Find affinity of visitor to pages and modify layout.

6.8: Data Mining Products

Examples of Data Mining Products

- Here are some examples of Data Mining Products:
 - DataMind: neurOagent
 - Information Discovery: IDIS
 - SAS Institute: SAS/Neuronets

6.9: Data Mining Market

Mining Market and Vendors

- There are around 20 to 30 mining tool vendors.
- Major tool players:
 - Clementine
 - IBM's Intelligent Miner
 - SGI's MineSet
 - SAS's Enterprise Miner
- Many embedded products:
 - Fraud detection
 - Electronic commerce applications
 - Health care
 - Customer Relationship Management: Epiphany

Summary

- In this lesson, you have learnt:
 - Data Mining is the way of analyzing data by exploring the large databases.
 - Data Mining is used to mine key information from a data warehouse.
 - It helps in exploring data in terms of reporting and statistical analysis.



Review Question

- Question 1: Data exploration for statistical analysis is done by:
 - Option 1: DBA
 - Option 2: Business analyst
 - Option 3: Data analyst
- Question 2: Data Mining is a subset of DW.
 - True/ False
- Question 3: Mining is also known as ____.



Review - Match the Following

1. End user

2. Business analyst

3. Data analyst

A. Data mining

B. Data warehouse

C. Data presentation

D. Making decisions

E. Data exploration

