Data Warehouses, Business Intelligence and Decision Support Systems

Information Technology For Management 7th Edition
Turban & Volonino
Based on lecture slides by L. Beaubien, Providence College
John Wiley & Sons, Inc.
Data Life Cycle Process

Data Sources and Databases
- Internal Data
- External Data
- Personal Data

Data Storage
- Data Warehouse
- Data Marts

Data Analysis
- OLAP, Queries, EIS, DSS
- Data Marts

Results
- Data Visualization
- Decision Support
- Knowledge and its Management

Solutions
- SCM
- CRM
- EC
- Strategy
- Others

Business Analytics
Transactional vs. Analytical Data Processing

- **Transactional processing** takes place in operational systems (TPS) that provide the organization with the capability to perform business transactions and produce transaction reports. The data are organized mainly in a hierarchical structure and are centrally processed. This is done primarily for fast and efficient processing of routine, repetitive data.

- Supplementary activity to transaction processing is called **analytical processing**, which involves the analysis of accumulated data. Analytical processing, sometimes referred to as *business intelligence*, includes data mining, decision support systems (DSS), querying, and other analysis activities. These analyses place strategic information in the hands of decision makers to enhance productivity and make better decisions, leading to greater competitive advantage.
Data Warehouse

- DW is a repository of data that are organized to be readily acceptable for analytical processing activities (DSS, querying, data mining)
- Organization by subjects (multidimensional, often employing hierarchies)
- Standardization of data
- Relational structure
- Delivery of DWH content to users on the intranet and extranet (online banking)
- Not all data are necessarily transferred to data warehouse
187 Canon printers were produced in January 2002 in Denver.

<table>
<thead>
<tr>
<th>Date</th>
<th>Intel</th>
<th>AMD</th>
<th>HP</th>
<th>Lexm</th>
<th>Canon</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>442</td>
<td>401</td>
<td>201</td>
<td>302</td>
<td>187</td>
</tr>
<tr>
<td>February</td>
<td>224</td>
<td>289</td>
<td>134</td>
<td>89</td>
<td>121</td>
</tr>
<tr>
<td>March</td>
<td>211</td>
<td>271</td>
<td>75</td>
<td>76</td>
<td>312</td>
</tr>
<tr>
<td>April</td>
<td>254</td>
<td>208</td>
<td>143</td>
<td>108</td>
<td>112</td>
</tr>
<tr>
<td>May</td>
<td>187</td>
<td>234</td>
<td>45</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>June</td>
<td>112</td>
<td>267</td>
<td>111</td>
<td>78</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>America</th>
<th>New York</th>
<th>Paris</th>
<th>Berlin</th>
<th>Los Angeles</th>
<th>Denver</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>442</td>
<td>400</td>
<td>210</td>
<td>47</td>
<td>290</td>
<td>121</td>
</tr>
<tr>
<td>February</td>
<td>224</td>
<td>150</td>
<td>43</td>
<td>176</td>
<td>222</td>
<td>121</td>
</tr>
<tr>
<td>March</td>
<td>211</td>
<td>75</td>
<td>76</td>
<td>312</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>April</td>
<td>254</td>
<td>210</td>
<td>47</td>
<td>290</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>May</td>
<td>187</td>
<td>43</td>
<td>98</td>
<td>121</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>June</td>
<td>112</td>
<td>290</td>
<td>12</td>
<td>121</td>
<td>121</td>
<td>121</td>
</tr>
</tbody>
</table>
The Data Warehouse & Data Management
Web-based Data Management Systems – content and information

Figure 3.14 Teradata Corp.’s enterprise data warehouse. (Source: Teradata Corporation [teradata.com], with permission.)
Data Warehousing

- Data warehousing is most appropriate when
  - Large amounts of data to be accessed
  - The operational data is stored in different systems
  - Large number of users (AT&T)
  - Extensive end-user computing
Datamarts

- The high cost of DWH confines their use to large companies
- A datamart is a small warehouse designed for a department
- Two types
  - Dependent
  - Standalone
Document Management

- Automated control of electronic documents (images, spreadsheets, text documents, invoices,...) through entire lifecycle
  - Production
  - Storage
  - Distribution
- Normally includes version management and access control
- Provided by Document Management Systems (DMS)
Case Studies

- Prepare Minicase Chapter 3 (Pushing Corporate Data to Employees' Facebook Pages) for next class session.
Knowledge Management (KM)

• A process that helps organizations identify, select, organize, disseminate, transfer, and apply information and expertise that are part of the organization’s memory and that typically reside within the organization in an unstructured manner.
Knowledge Management

- **Explicit Knowledge**: The more objective, rational, and technical types of knowledge
- **Tacit knowledge**: The cumulative store of subjective or experiential learning; it is highly personal and hard to formalize.
The Knowledge Management System Cycle

- **Create knowledge.** Knowledge is created as people determine new ways of doing things or develop know-how. Sometimes external knowledge is brought in.
- **Capture knowledge.** New knowledge must be identified as valuable and be represented in a reasonable way.
- **Refine knowledge.** New knowledge must be placed in context so that it is actionable. This is where human insight (tacit qualities) must be captured along with explicit facts.
- **Store knowledge.** Useful knowledge must then be stored in a repository so that others in the organization can access it.
- **Manage knowledge.** Like a library, the knowledge must be kept current. It must be reviewed to verify that it is relevant and accurate.
- **Disseminate knowledge.** Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and any time.
Knowledge creation or knowledge acquisition is the generation of new insights, ideas, or routines.

- **Socialization mode** refers to the conversion of tacit knowledge to new tacit knowledge through social interactions and shared experience.
- **Combination mode** refers to the creation of new explicit knowledge by merging, categorizing, reclassifying, and synthesizing existing explicit knowledge.
- **Externalization** refers to converting tacit knowledge to new explicit knowledge.
- **Internalization** refers to the creation of new tacit knowledge from explicit knowledge.

- **Knowledge sharing** is the exchange of ideas, insights, solutions, experiences to another individuals via knowledge transfer computer systems or other non-IS methods.

- **Knowledge seeking** is the search for and use of internal organizational knowledge.
Chief Knowledge Officer (CKO)

- Executive whose objectives are to maximize the firm’s knowledge assets, design and implement knowledge management strategies, and effectively exchange knowledge asset internally and externally.

- Process approach to KM: codify knowledge through controls, processes and technologies

- Practice approach to KM: assumes majority of tacit knowledge, creates necessary social environment
  - **Community of practice**: A group of people in an organization with a common professional interest.
Knowledge management systems (KMSs): Information technologies used to systematize, enhance, and expedite intra- and interfirm knowledge management.

Communication technologies: allow users to access needed knowledge, and to communicate with each other—especially with experts. E-mail, the Internet, corporate intranets, and other web-based tools provide communication capabilities.

Collaboration technologies: provide the means to perform group work. Collaborative computing capabilities such as electronic brainstorming enhance group work especially for knowledge contribution.

Storage and retrieval technologies: originally meant using a database management system to store and manage explicit knowledge. Electronic document management system and specialized storage system that are part of collaborative computing system are the tools used to capture, store, and manage tacit knowledge.
Discussion

- Which industries are especially reliant on knowledge management?
- What are applications of knowledge management in these industries?
Business Intelligence

- Encompasses architectures, tools, applications, databases and methodologies
- Objective is to enable interactive access to data, enable manipulation, and to provide managers and analysts the ability to conduct analyses
Business Analytics

- Products supporting data analysis and reports
  - Reporting and queries
  - Advanced analytics
  - Data mining
## TABLE 11.3  Stages in the Evolution of Knowledge Discovery

<table>
<thead>
<tr>
<th>Evolutionary Stage</th>
<th>Business Question</th>
<th>Enabling Technologies</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection (1960s)</td>
<td>What was my total revenue in the last five years?</td>
<td>Computers, tapes, disks</td>
<td>Retrospective, static data delivery</td>
</tr>
<tr>
<td>Data access (1980s)</td>
<td>What were unit sales in New England last March?</td>
<td>Relational databases (RDBMS), structured query language (SQL)</td>
<td>Retrospective, dynamic data delivery at record level</td>
</tr>
<tr>
<td>Data warehousing and decision support (early 1990s)</td>
<td>What were the sales in region A, by product, by salesperson?</td>
<td>OLAP, multidimensional databases, data warehouses</td>
<td>Retrospective, dynamic data delivery at multiple levels</td>
</tr>
<tr>
<td>Intelligent data mining (late 1990s)</td>
<td>What’s likely to happen to the Boston unit’s sales next month? Why?</td>
<td>Advanced algorithms, multiprocessor computers, massive databases</td>
<td>Prospective, proactive information delivery</td>
</tr>
<tr>
<td>Advanced intelligent system</td>
<td>What is the best plan to follow?</td>
<td>Neural computing, advanced AI models, complex optimization, Web Services</td>
<td>Proactive, integrative; multiple business partners</td>
</tr>
<tr>
<td>Complete integration (2000–2004)</td>
<td>How did we perform compared to metrics?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Mining Concepts

- **Data mining**: The process of searching for valuable business information in a large database, data warehouse, or data mart.

- **Data mining capabilities include**:  
  1) Automated prediction of trends and behaviours, and  
  2) Automated discovery of previously unknown patterns.
Data Mining Applications

- Marketing
  - Retailing (recommendation services)
  - Banking
  - Insurance
  - Health care
- Manufacturing and production
- Police work
- Science (Bioinformatics)
Text Mining

- The application of data mining to non-structured or less-structured text files.
- Text mining helps organizations to do the following (1) find the “hidden” content of documents, including additional useful relationship and (2) group documents by common themes (e.g., identity all the customers of an insurance firm who have similar complaints).
Text and Data Mining

• Discussion
  ○ What could be possible applications of data and text mining in a university setting?
Web Mining

- The application of data mining techniques to discover actionable and meaningful patterns, profiles, and trends from web resources.

- Web mining is used in the following areas: information filtering, surveillance, mining of web-access logs for analyzing usage, assisted browsing, and services that fight crime on the internet.
Business Performance Management

- Framework for organizing, automating and analyzing business methodologies, metrics, processes and systems to drive the overall, performance of the enterprise
- Helps organizations translate a unified set of objectives into plans, monitor execution, and deliver critical insight to improve financial and operative performance
Business Performance Management

Diagram showing the process of Business Performance Management:

1. Strategize
   - Strategic Plans
   - Strategy Maps

2. Plan
   - Plans, Budgets, & Scenarios
   - Projects/Initiatives

Monitor
   - Scorecards
   - Reports/Analysis
   - Alerts

Act & Adjust
   - Agents, Offers, & Actions

Execution

Integrate Data
Balanced Scorecard

Financial

"To succeed financially, how should we appear to our shareholders?"

Customers

"To achieve our vision, how should we appear to our customers?"

Internal Business Processes

"To satisfy our shareholders and customers, what business processes must we excel at?"

Learning and Growth

"To achieve our vision, how will we sustain our ability to change and improve?"

Vision and Strategy

Customer satisfaction

On-time delivery

Shorter cycle time

Process quality

Lower rework

Employees' skills

Employee morale

Employee suggestions

Earnings per share

Return on assets (ROA)

Operating expenses

Objectives

Measures

Targets

Initiatives

Leading Indicators

Lagging Indicators
Decision Making

Intelligence Phase
- Organizational Objectives
- Search and Scanning Procedures
- Data Collection
- Problem Identification
- Problem Classification
- Problem Statement

Design Phase
- Formulate a Model (Assumptions)
- Set Criteria for Choice
- Search for Alternatives
- Predict and Measure Outcomes

Choice Phase
- Solution to the Model
- Sensitivity Analysis
- Selection of Best (Good) Alternative
- Plan for Implementation (Action)
- Design of a Control System

Implementation of Solution

REALITY
Problems
Opportunities

Examination

Validation of the Model

SUCCESS

Verification, Testing of Proposed Solution

FAILURE
Decision Support Systems

- DSS is a computer-based information system that combines models and data to solve semi-structured problems with intensive user involvement.
- Why do we need computers?
  - Large number of alternatives
  - Uncertainty
  - Intensive calculations
  - Decision makers and data are usually located in different places
- Decision making process can be
  - Highly structured (procedures for obtaining the best solutions are known - Management Science)
  - Highly unstructured problems (managerial judgment and intuition; ex. new product development)
  - Semi-structured problems (DSS is most suitable)
Models and modeling

- A model is a simplified representation of reality.
- With modeling, one can reach an optimum solution under certain assumptions and perform virtual experiments:
  - The cost of virtual experimentation is much lower
  - Years of operations can be simulated
  - Manipulating the model is much easier
  - What ifs
  - Evaluating large number of alternatives
Characteristics and Capabilities of DSSs

- **Sensitivity analysis** is the study of the impact that changes in one (or more) parts of a model have on other parts.
- **What-if analysis** is the study of the impact of a change in the assumptions (input data) on the proposed solution.
- **Goal-seeking analysis** is the study that attempts to find the value of the inputs necessary to achieve a desired level of output.
Structure and Components of DSSs

- **Data management subsystem** contain all the data that flow from several sources.
- **Model management subsystem** contains completed models and the building blocks necessary to develop DSS applications.
- **User interface** covers all aspects of the communications between a user and the DSS.
- **Users** are the persons faced with the problem or decision that the DSS is designed to support.
- **Knowledge-based subsystems** provide the required expertise for solving some aspects of the problem.
DSS Process
Expert systems (ESs) are attempts to mimic human experts (AI). It is decision-making software that can reach a level of performance comparable to a human expert in some specialized and usually narrow problem area. The idea is simple: expertise is transferred from an expert or other source of expertise to the computer.

- The transfer of expertise from an expert to a computer and then to the user involves four activities:
  - Knowledge acquisition (from experts or other sources)
  - Knowledge representation (organized as rules or frames in the computer)
  - Knowledge inferencing is performed in a component called the inference engine of the ES and results in the recommendation.
  - Knowledge transfer to the user (the expert’s knowledge has been transferred to users).
Artificial Neural Networks (ANN)

- Pattern recognition and learning are the key characteristics of ANN.
- ANN can analyze large quantities of data to establish patterns where the logic rules are not known.
- Neural networks are particularly good at identifying subtle, hidden, and newly emerging patterns within complex data as well as interpreting incomplete inputs.
- Example 1: Loan application. By reviving many historical cases of applicants’ questionnaires, ANN can create patterns or profiles of applicants that should be approved or denied.
Artificial Neural Network
**Inputs.** Each input corresponds to a single attribute. For example, if the problem is to decide on approval or disapproval of a loan, some attributes could be the applicant’s income level, age, and home ownership. Several types of data, such as text, pictures, and voice, can be used as inputs. Preprocessing may be needed to convert the raw data to meaningful inputs from symbolic data or to scale the data. The inputs are multiplied by weights—when they enter the processing elements (PEs).

**Weights.** Key elements in an ANN are the weights. Weights express the relative strength (or mathematical value) of the input data or the many connections that transfer data from layer to layer. In other words, in the case of a loan application, weights express the relative importance of each input applicant’s attribute, based on past experiences. Weights are crucial in that they store learned patterns of information. It is through repeated adjustments of weights that the network learns.

**Summation function.** The summation function (represented by the symbol) calculates the weighted sum of all the input elements entering each processing element. A summation function multiplies each input value by its weight and totals the values for a weighted sum.

**Transformation function.** A transformation function (represented by) integrates the information produced by all PEs, and transforms it to meaningful outputs.

**Outputs.** The outputs of the network contain the solution to a problem. For example, in the case of a loan application it can be “yes” or “no.” The ANN assigns numeric values, like 1 for “yes” and 0 for “no.” The purpose of the network is to compute the values of the output.
Applications of Neural Networks

- Data mining: Finding data in large and complex databases
- Credit card fraud detection: Analyzing purchasing patterns for fast detection of fraud
- Tax fraud: Identifying, enhancing, and finding irregularities
- Evaluation of personnel: Matching personnel data to job requirements and performance criteria job candidates
- Loan application evaluation: Judging worthiness of loan applications based on patterns in previous application information (customer credit scoring)
- New product analysis: Sales forecasting and targeted marketing
- Insurance fraud detection: Finding fraud patterns
Fuzzy Logic

- **Fuzzy logic** deals with the uncertainties by simulating the process of human reasoning, allowing the computer to behave less precisely and logically than conventional computers do.
  - Involves decision in gray areas.
  - Uses creative decision-making processes.
Simulation generally refers to a technique for conducting experiments (such as "what-if") with a computer on a model of a management system. Because DSS deals with semi structured or unstructured situations, it involves complex reality, which may not be easily represented by optimization or other standard models but can often be handled by simulation. Therefore, simulation is one of the most frequently used tools of DSSs.

- Advantages of Simulation.
  - Allows for inclusion of the real-life complexities of problems.
  - Is descriptive.
  - Can handle an extremely wide variation in problem types.
  - Can show the effect of compressing time.
  - Can be conducted from anywhere.
Automated Decision Support (ADS)

- ADS systems are rule-based systems that are particularly useful for repetitive managerial problems.
- ADS is achieved by capturing a business user expertise in a set of business rules.
- These rules determine what actions needs to be taken in particular situations
- Examples
  - Loan approval
  - Dell
  - Yield optimization
  - Traffic lights
  - Dynamic Forecasting
BI Issues

- **Cost justification, intangible benefits.** While some of the benefits of management support systems are tangible, it is difficult to put a dollar value on the intangible benefits of many such systems.

- **Documenting personal DSS.** Many employees develop their own DSSs to increase their productivity and the quality of their work. It is advisable to have an inventory of these DSSs and make certain that appropriate documentation and security measures exist.

- **Ready-made commercial DSSs.** With the increased use of Web-based systems and ASPs, it is possible to find more DSS applications sold off the shelf, frequently online. The benefits of a purchased or leased DSS application sometimes make it advisable to change business processes to fit a commercially available DSS.

- **Embedded technologies.** Intelligent systems are expected to be embedded in at least 20 percent of all IT applications in about 10 years. It is critical for any prudent management to closely examine the technologies and their business applicability.
Activity

- Develop a recommender system for Bogazici University Department of Management students.
  - What is to be recommended (and why)?
  - Which data is used as input (registration system data, data from graduates)?
Case Studies

- Prepare the Mini Case of Chapter 12 (Lexmark) from the textbook for next class session.