DECISION SUPPORT SYSTEMS AND INTELLIGENT SYSTEMS,
Seventh Edition
**---Efraim Turban**, **Jay E. Aronson, and Ting-Peng Liang**

# Chapter 5

**BUSINESS INTELLIGENCE:
WAREHOUSIMG, DATA ACQUISITION, DATAMIHING, BUSINISS
ANALYTICS, and VISUALIZATION**

**College of Computer Science and Information Technology
Department of Computer Information Systems
Prof Dr. Taleb A. S. Obaid**

# LEARNING OBJECTIVES

- Describe issues in data collection, problems, and quality.

- Describe the database management systems.

- Explain and use of a data warehouse and data mart.

- Describe how online analytical processing (OLAP), data mining, data visualization,

- Explain how the Web impacts database technologies..

- Intelligence/business analytics improve decision-making

# Business intelligence

- Organizations have amassed vast amounts of data.

- Use analytic tools to improve organizational decision-making.

- Enabling decision analysis through access to all relevant information is known as business intelligence.

- Business intelligence includes data warehousing online analytical processing, data mining, and visualization and multidimensionality.

- DW enables data mining, the ability to automatically synthesize (create) vast amounts of information in order to discover hidden truths within the data.

## OPENING VIGNETTE: INFORMATION SHARING A PRINCIPAL COMPONENT OF THE NATIONAL STRATEGY FOR HOMELAND Security

- The National Strategy for Homeland Security of the United States includes a National Vision for the sharing of information related to the detection of terrorist activities.

- Build a national environment that enables the sharing of essential homeland security information.

- Can provide the right information to the right people at all times.

- Shared Information

   "horizontally" across each level of government and

   "vertically" among federal, state and local governments, private industry and citizens.

- Proper use of people, and technology, homeland security can have complete and common awareness of threats.

- So they can anticipate (expected) threats and respond rapidly and effectively.

- The goal of the project is to create a workable model for integrating knowledge that resides across many disparate data sources

## OPENING VIGNETTE: INFORMATION SHARING A PRINCIPAL COMPONENT OF THE NATIONAL STRATEGY FOR HOMELAND Security

The five major initiatives that are identified within the strategy include:

1. To integrate information sharing across the federal government

2. To extend the integration of information sharing across state and local governments, private industry, and citizens

3. To adopt common metadata standards of electronic information relevant to homeland security

4. To improve public safety communication

5. To ensure reliable public health information.

QUESTIONS FOR THE OPENING VIGNETTE

1. Identify the challenges faced by the Office of Homeland Security in integrating disparate databases.

2. Identify the sources of information that will be required to make the information in this data portal useful.

3. What are the expected benefits?

4. Identify decisions supported by this data portal.

5. What decision support tools and techniques can be used to identify potential terrorist activities?

# 5.2 THE NATURE AND SOURCES OF DATA

- A decision-maker needs data, information, and knowledge.

- Decision-maker must be able to apply analysis tools so that the data, information, and knowledge can be utilized to full benefit.

- All decision-support systems use data, information, and/or knowledge:

1. **Data**. Items about things, events, activities, and transactions are recorded, classified, and stored **but are not organized** to convey any specific meaning. Data items can be numeric, alphanumeric, figures, sounds, or images.

# 5.2 THE NATURE AND SOURCES OF DATA

2.  **Information**. Data have organized in a manner that give them meaning for the recipient. An MSS application processes data items so that the results are meaningful for an intended action or decision.

3.  **Knowledge**. Knowledge consists of data items and/or information organized and processed to convey understanding, experience, accumulated learning, and expertise that are applicable to a current problem or activity. MSS data can include documents, pictures, maps, sound, video, and animation.

# 5.2 THE NATURE AND SOURCES OF DATA

Many MSS applications use data that come from three primary sources: internal , external and personal.

- **INTERNAL DATA:** are stored in one or more places.

  - Sales data can be stored in several places: aggregate sales data in the corporate database, and details at each region's database.

  - Internal data are available via an organization's intranet or other internal network.

# 5.2 THE NATURE AND SOURCES OF DATA

**EXTERNAL DATA**: There are many sources of external data.

- They range from commercial databases to data collected by sensors and satellites.

- Data are available on CDs and DVDs, on the Internet, as films and photographs, and as music or voices.

- Government reports and files are a major source of external data, available on the Web today.

- Available by using GIS (geographic information systems).

- Chambers of commerce, local banks, research institutions, and the like, flood the environment with data and information. Data can come from around the globe.

- Most external data are irrelevant to a specific MSS.

# 5.2 THE NATURE AND SOURCES OF DATA

**PERSONAL DATA AND KNOWLEDGE**:

- MSS users have expertise and knowledge that can be stored for future use.

- These include subjective (individual) estimates of sales, opinions about what competitors are likely to do, and interpretations of news articles.

- What people really know and methodologies to capture, manage, and distribute it are the subject of knowledge management.

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

- To extract data from many internal and external sources is complicates task of MSS building.

- Sometimes need to collect raw data.

- In other cases, need to elicit data from people or the Internet.

- Data must be validated and filtered.

- A classic expression the sums up the situation is " garbage in, garbage out" (GIGO).

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

METHODS FOR COLLECTING RAW DATA

- Can be collected manually or by instruments and sensors.

- Data collection methods are time studies, surveys (using questionnaires), observations (using video cameras), and soliciting (asking) information from experts (interviews).

- In addition, sensors and scanners are increasingly being used in data acquisition.

- From point-of-purchase inventory control.

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

DATA PROBLEMS

- All computer-based systems depend on data, so, the quality and integrity of the data are critical .

-  MSS depend on data because make up information and knowledge are at the heart of any decision making system.

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

DATA QUALITY (DQ)

- DQ is an extremely important because quality determines the usefulness of data, so, the quality of the decisions based on them.

- Data in organizational databases are frequently found 'to be inaccurate, incomplete, or ambiguous.

- Poor-quality data costs billions' of dollars .

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

DATA INTEGRITY

- Major issues of DQ is data integrity.

- Lack integrity means a change made in the file in one place may not be made in the file in another place or department. This results in conflicting data.

Gray and Watson (1998) distinguish the following five issues:

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

Gray and Watson (1998) distinguish the following five issues:

1. **Uniformity** checks ensure that the data are within specified limits.

2. **Version**. Version checks are performed when the data are transformed through the use of metadata to ensure that the format of the original data has not been changed.

3. **Completeness check**. A completeness check ensures that the summaries are correct and that all values needed to create the summary are included.

4. **Conformity check**. A conformity check makes sure that the summarized data are "in the ballpark estimate." That is, during data analysis and reporting, correlations are run between the value reported and previous values for the same number.

5. **Genealogy** (descent ) **check or drill down** A genealogy check or drill down is a trace back to the data source through its various transformations.

# 5.3 DATA COLLECTION, PROBLEMS, AND QUALITY

DATA ACCESS AND INTEGRATION

- A decision-maker typically needs access to multiple sources of data that must be integrated.

- As data warehouses grow in size, the issues of integrating data exasperate.

- The complexity databases and large-scale independent MSS databases makes standard computer operating systems inadequate between the user and the database.

- A (DBMS) supplements standard operating systems by allowing for greater integration of data, complex file structure, quick retrieval and changes, and better data security.

## 5.5 DATABASE MANAGEMENT SYSTEMS IN DECISION SUPPORT SYSTEMS/BUSINESS INTELLIGENCE

- DBMS is a software program for adding information to a database and updating, deleting, manipulating, storing, and retrieving information.

- Data from the database are extracted and put in a statistical, mathematical, or financial model for further manipulation or analysis. Large, complex DSS often do this.

- The major role of DBMS is to manage data.

- By manage, we mean to create, delete, change, and display the data.

- DBMS enable users to query data as well as to generate reports.

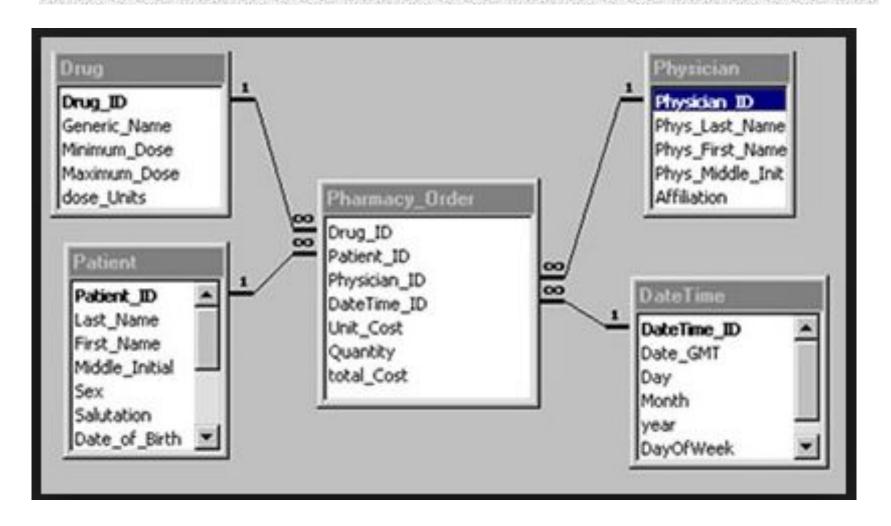# 5.6 DATABASE ORGANIZATION AND STRUCTURES

- The relationships between the many individual records stored in a database can be expressed by several logical structures.

- The three conventional structures--relational, hierarchical, and network.
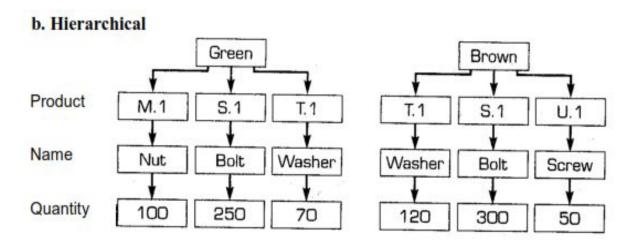
-

# RELATIONAL DATABASES

- The relational form of DSS database organization allows the user to reports as two-dimensional tables.

- Relational DBMS allow multiple access queries.

- Data file consists of a number of columns are considered a separate field.

- The rows represent individual records made up of several fields.

- The names of common fields must be spelled exactly alike, and the fields must be the same size (the same number of bytes) and type.
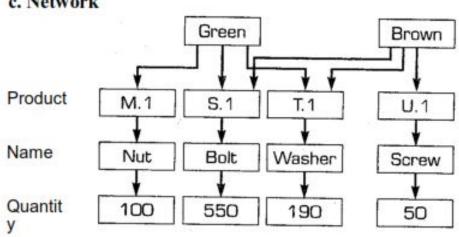
# RELATIONAL DATABASES

# HIERARCHICAL DATABASES

A hierarchical model orders data items in a top-down fashion, creating logical links between related data items. It looks like a tree or an organization chart. It is used mainly in transaction processing, where processing efficiency is a critical element.

**b. Hierarchical**

| | Green | | | | Brown | | |
|---|---|---|---|---|---|---|---|
| Product | M.1 | S.1 | T.1 | | T.1 | S.1 | U.1 |
| Name | Nut | Bolt | Washer | | Washer | Bolt | Screw |
| Quantity | 100 | 250 | 70 | | 120 | 300 | 50 |

# NETWORK DATABASES

The network database structure permits more complex links, including lateral connections between related items. This structure is also called the CODASYL model. It can save storage space through the sharing of some items. For example, in Figure 5.1, Green and Brown share S.1 and T.1.

c. Network

| | Green | | | Brown |
|---|---|---|---|---|
| Product | M.1 | S.1 | T.1 | U.1 |
| Name | Nut | Bolt | Washer | Screw |
| Quantity | 100 | 550 | 190 | 50 |

# OBJECT-ORIENTED DATABASES

- Comprehensive MSS applications require accessibility to complex data, which may include pictures and elaborate relationships.

- Such situations cannot be handled efficiently by hierarchical, network, or even relational database architectures. Even SQL may not be effective.

- For such applications, a graphical representation, such as the one used in objected-oriented systems, may be useful.

- Object oriented database systems combine the characteristics of an object-oriented programming language, such as UML,

# MULTIMEDIA-BASED DATABASES

- (MMDBMS) manage data in a variety of formats, in addition to the standard text or numeric field. These formats include images, such as digitized photographs, and forms of bit-mapped graphics, such as maps or .pic files, hypertext images, video clips, sound, and virtual reality (multidimensional images).

- It is critical to develop effective ways to manage such data for GIS and for many other Web applications.

- Managing multimedia data continues to become more important for business intelligence (see D' Agostino, 2003).