MIS

CHAPTER 3

DATABASE SYSTEMS, DATA WAREHOUSES, AND DATA MARTS

Hossein BIDGOLI

MBNA

Hong Kong Airport

A not so perfect match

ebay

proquest

learning outcomes

LO1 Define a database and a database management system.

LO2 Explain logical database design and the relational database model.

LO3 Define the components of a database management system.

LO4 Summarize recent trends in database design and use.

LO5 Explain the components and functions of a data warehouse.

learning outcomes (cont'd.)

LO6 **Describe the functions of a data mart.**

Managing data and information

- Usually too much data rather than too little in organizations
- How does an organization organize all this data and information?
  - Database – a collection of integrated and related files
    - Ebay
    - Proquest
    - **MBNA**

business is info

What is Database Technology?

- A collection of related data organized in a way that makes it valuable and useful
- Allows organizations to retrieve, store, and analyze information easily
- Is vital to an organization's success in running operations and making decisions

The Hierarchy of Data

Hierarchy of data	Example	
Database	Personnel file Department file Payroll file	(Project database)
Files	098 - 40 - 1370 Fiske, Steven 01-05-1985 549 - 77 - 1001 Buckley, Bill 02-17-1979 005 - 10 - 6321 Johns, Francine 10-07-1997	(Personnel file)
Records	098 - 40 - 1370 Fiske, Steven 01-05-1985	(Record containing SSN, last and first name, hire date)
Fields	Fiske	(Last name field)
Characters (bytes)	1000100	(Letter F in ASCII)

Databases

- Critical component of information systems
  - Any type of analysis that's done is based on data available in the database
- **Database management system (DBMS)**
  - Creating, storing, maintaining, and accessing database files
- Advantages over a flat file system

The Traditional Approach

The Traditional Approach to Data Management

U of L example

The Database Approach

all in one database

The Database Approach to Data Management

Advantages of the Database Approach

Advantages	Description
Program–data independence	Much easier to evolve and alter software to changing business needs when data and programs are independent.
Minimal data redundancy	Single copy of data assures that data storage is minimized.
Improved data consistency	Eliminating redundancy greatly reduces the opportunities for inconsistency.
Improved data sharing	Easier to deploy and control data access using a centralized system.
Increased productivity of application development	Data standards make it easier to build and modify applications.
Enforcement of standards	A centralized system makes it much easier to enforce standards and rules for data creation, modification, naming, and deletion.
Improved data quality	Centralized control, minimized redundancy, and improved data consistency help to enhance the quality of data.
Improved data accessibility	Centralized system makes it easier to provide access for new personnel within or outside organizational boundaries.
Reduced program maintenance	Information changed in the central database is replicated seamlessly throughout all applications.

Exhibit 3.2

Interaction between the user, DBMC, and Database

Types of Data in a Database

- Internal data
  - Collected within organization

transactions
online

- External data
  - Sources
    - Competitors, customers, and suppliers
    - Distribution networks
    - Economic
    - Government regulations
    - Labor and population statistics
    - Tax records
    - functional information systems

BI in Action: Law Enforcement

- Business intelligence (BI)
  - Used in law enforcement as well as in the business world
- Richmond, Virginia
  - System generates BI reports that help pinpoint crime patterns
  - Allocate manpower to days and locations where crime likely to occur

better use of resources

Calgary

Methods for Accessing Files

- **Sequential file structure**
  - Records organized and processed in numerical or sequential order
  - Organized based on a "primary key"
  - Usually used for backup and archive files
    - Because they need updating only rarely
- **Random access file structure**
  - Records can be accessed in any order
  - Fast and very effective when a small number of records need to be processed daily or weekly

Methods for Accessing Files (cont'd.)

- **Indexed sequential access method (ISAM)**
  - Records accessed sequentially or randomly
  - Depending on the number being accessed
- Indexed access
  - Uses an index structure with two parts:
    - Indexed value
    - Pointer to the disk location of the record matching the indexed value

Logical Database Design

- **Physical view**
  - How data is stored on and retrieved from storage media
- **Logical view**
  - How information appears to users
  - How it can be organized and retrieved
  - Can be more than one logical view

Logical Database Design (cont'd.)

- **Data model**
  - Determines how data is created, represented, organized
  - Includes
    - Data structure
    - Operations
    - Integrity rules

The Relational Model

- **Relational model**
  - Uses a two-dimensional table of rows and columns of data
- **Data dictionary**
  - Field name
  - Field data type
  - Default value
  - Validation rule

The Relational Model – Example

Department Records

Department No	Dept Name	Location	Dean
Dept A			
Dept B			
Dept C			

Instructor Records

Instructor No	Inst Name	Title	Salary	Dept No
Inst 1				
Inst 2				
Inst 3				
Inst 4				

primary key

ID #

unique

Figure 3.12 With the relational model, we represent these two entities, department and instructor, as two separate tables and capture the relationship between them with a common column in each table.

Copyright © 2008 Pearson Education Canada

The Relational Model

- Data retrieval
  - **Select**
  - **Project**
  - **Join**
  - Intersection
  - Union
  - Difference

Table 3.4

The Customers table

Customer#	Name	Address
2000	ABC	Broadway
3000	XYZ	Jefferson
9000	TRY	Madison

Table 3.5

The Invoices table

Invoice#	Customer#	Amount	Payment
1110	2000	$2000.00	Cash
2220	3000	$4000.00	Credit
3330	3000	$1500.00	Cash
4440	9000	$6400.00	Cash
5550	9000	$7000.00	Credit

Table 3.6

Joining the Invoices and Customers tables

Invoice#	Customer#	Amount	Payment	Name	Address
1110	2000	$2000.00	Cash	ABC	Broadway
2220	3000	$4000.00	Credit	XYZ	Jefferson
3330	3000	$1500.00	Cash	XYZ	Jefferson
4440	9000	$6400.00	Cash	TRY	Madison
5550	9000	$7000.00	Credit	TRY	Madison

Components of a DBMS

- Database engine
- Data definition
- Data manipulation
- Application generation
- Data administration

Database Engine

- Heart of DBMS software
- Responsible for data storage, manipulation, and retrieval
- Converts logical requests from users into their physical equivalents

Data Definition

- Create and maintain the data dictionary
- Define the structure of files in a database
  - Adding fields
  - Deleting fields
  - Changing field size
  - Changing data type

Data Manipulation

- Add, delete, modify, and retrieve records from a database
- Query language
  - **Structured Query Language (SQL)**
    - Standard fourth-generation query language used by many DBMS packages
    - SELECT statement
  - **Query by example (QBE)** most used by users
    - Construct statement of query forms background
    - Graphical interface makes SQL request

Application Generation

- Design elements of an application using a database
  - Data entry screens
  - Interactive menus
  - Interfaces with other programming languages

Data Administration

- Used for:
  - Backup and recovery
  - Security
  - Change management
- Create, read, update, and delete (CRUD)
- **Database administrator (DBA)**
  - Individual or department
  - Responsibilities

Recent Trends in Database Design and Use

- Data-driven Web sites
- Distributed databases
- Client/server databases
- Object-oriented databases

Data Warehouses, Data Marts, and Data Mining

- **Data warehouse:** collects business information from many sources in the enterprise
- **Data mart:** a subset of a data warehouse
- **Data mining:** an information-analysis tool for automated discovery of patterns and relationships in a data warehouse or a data mart
- **Online Analytical Processing** –Graphical software tools that provide complex analysis of data stored in a database

Target store: time + sales

diapers sale ↗ on friday evening

Data Warehouses

Hong Kong Airport

- **Data warehouse**
  - Collection of data used to support decision-making applications and generate business intelligence
- Multidimensional data

- List the Different Databases that Hong Kong Airport would utilize?

speed + efficiency

OLTP

- luggage
- security
- ticketing
- food

OLAP

over time

transition

traffic control

flight info display sys

Exhibit 3.9

A Data Warehouse Configuration

Slicing and Dicing Data

- **Data warehouses** are not transaction-oriented.
- **Data warehouses** support online analytical processing (OLAP).

A not so perfect match

- With the increasing power of Data mining techniques, comes ever increasing and reaching uses of this powerful technology.
- 1. What are the benefits of DNA databases?
- 2. What problems do DNA databases pose?
- 3. Who should be included in a national DNA database? Should it be limited to convicted felons?
- 4. Who should be able to use DNA databases?

Summary

- Databases
  - Accessing files
  - Design principles
  - Components
  - Recent trends
- Data warehouses and data marts