

Introduction to Database

CIS 3730

Designing and Managing Data

J.G. Zheng
Fall 2010



Overview

- ◆ What is a database?
- ◆ Why database?
- ◆ What is a *database system*?

Data

◆ Data

- Basic facts to describe something
- We need effective approaches to manage and process data, which is the basis of many human activities

◆ Basic data forms (*data type*)

- Character based
 - ◆ Number, text, special characters, etc.
- Non character based
 - ◆ Image, sound, video, other binary data, etc.

Sample Business Data

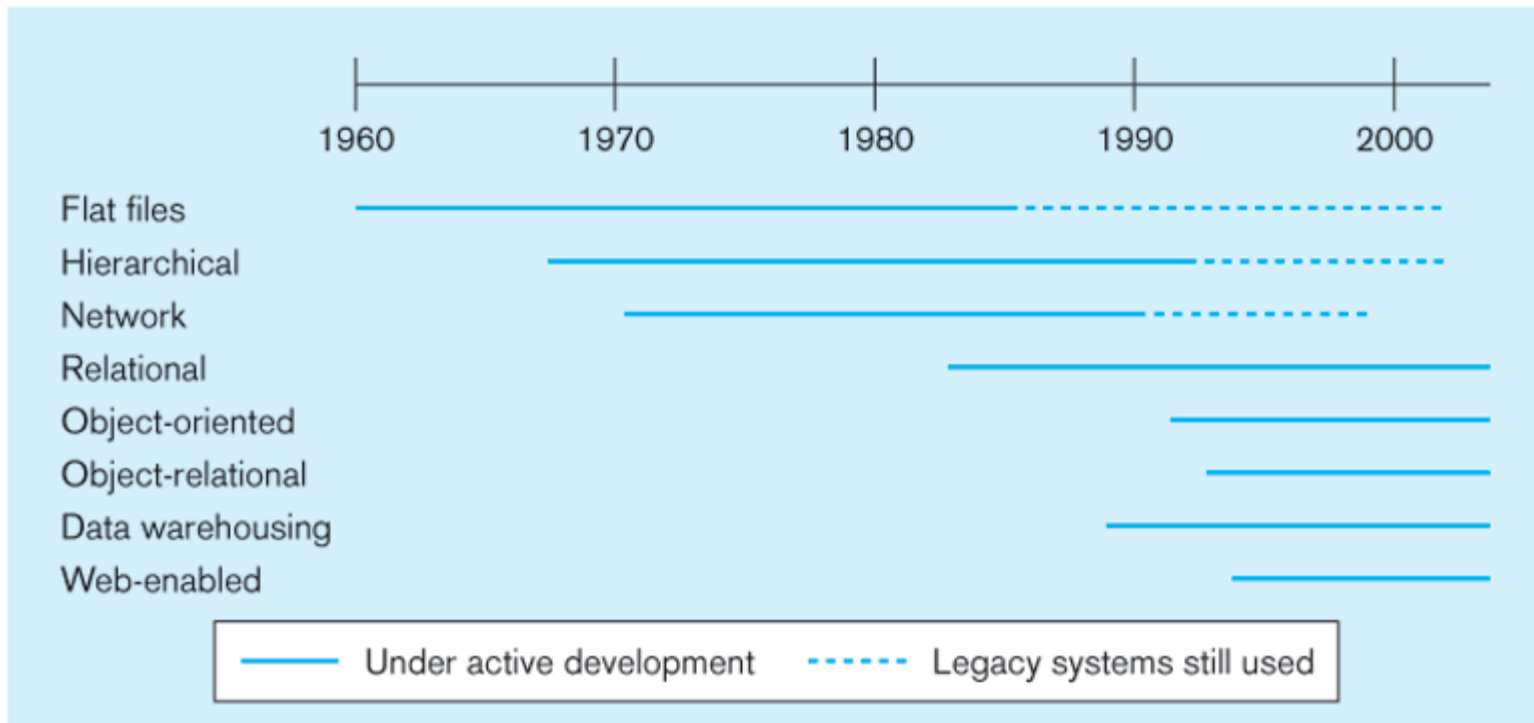
◆ How do we record and process these data?

The diagram shows a business data form for 'PREMIERE PRODUCTS'. It includes a heading section with order and customer information, a table of order lines with columns for part number, description, quantity, price, and total, and a footing section for the order total. Callouts identify these sections: 'Heading' points to the top text, 'Order lines' points to the table rows, 'Footing' points to the bottom summary row, and 'Body' points to the table area.

PART NUMBER	PART DESCRIPTION	NUMBER ORDERED	PRICE	TOTAL
BV06	Home Gym	2	794.95	1589.90
CD52	Microwave Oven	4	150.00	600.00
ORDERTOTAL >>				2189.90

ORDER: 21617
PREMIERE PRODUCTS
DATE: 10/23/2003
CUSTOMER: 608
Johnson's Department Store
372 Oxford
Sheldon FL 33553
SALES REP: 65
Juan Perez

Data Management Technologies



File Processing Approaches

◆ Based on file systems

- Files and folders (directories)

◆ File types

- Sequential data files
- Name-value pair files
- Spreadsheets or list files
- XML files

Sequential Data Files

- ◆ Data are sequentially recorded in a file

- For example: what does this piece of data mean?

21617-102303-02-BV06-02-079495

- Demo: another example

- ◆ Major problems

- Application dependent: the structure and meaning of the data rely on other application programs
- Difficult to maintain and share

Name-Value Pair and XML Files

◆ Name-value pair

- Uses labels (names) to store descriptions of data (metadata)

Order Number: 21617

Date: 10/20/2003

◆ XML files

- Uses tags to store metadata

```
<OrderNumber>21617</OrderNumber>
```

```
<Date>10/20/2003</Date>
```

Spreadsheet/List Files

- ◆ Data are organized in a 2D table with rows and columns
 - The first row describes the data of each column

Customer Number	Customer Name	Order Number	Order Date	Part Number	Part Description	Number Ordered	Quoted Price	Warehouse	Rep Number
148	Al's Appliance and Sport	21608	10/20/2003	AT94	Iron	11	\$21.95	3	20
148	Al's Appliance and Sport	21619	10/23/2003	DR93	Gas Range	1	\$495.00	2	20
282	Brookings Direct	21614	10/21/2003	KT03	Dishwasher	2	\$595.00	3	35
356	Ferguson's	21610	10/20/2003	DR93	Gas Range	1	\$495.00	2	65
356	Ferguson's	21610	10/20/2003	DW11	Washer	1	\$399.99	3	65
608	Johnson's Department Store	21617	10/23/2003	BV06	Home Gym	2	\$794.95	2	65
608	Johnson's Department Store	21617	10/23/2003	CD52	Microwave Oven	4	\$150.00	1	65
608	Johnson's Department Store	21623	10/23/2003	KV29	Treadmill	2	\$1,290.00	2	65

Files Processing Problems

- ◆ Loose and weak structure (a general structure may be imposed but not enforced)
 - Difficult to handle complex data
 - Low data quality: redundancy and inconsistency
- ◆ No central management
 - Difficult to maintain and share in multi-user environments
 - Limited security
- ◆ Not scalable: cannot handle large quantity of data efficiently
- ◆ Lack of specialized and standardized data management and processing capabilities

Database Approach

- ◆ Database is a structured and self-describing collection of data
 - Structured: structures and rules are consistently and rigorously organized and enforced (integrity)
 - Self-describing: the description of data (data definition, or metadata) is contained within the database
- ◆ Centralized management
 - Managed and controlled by specialized programs, database management systems (DBMS), which provides rich data management functionalities

Advantages and Disadvantages

◆ Advantages

- High data quality, integrity and consistency
- Easy access and sharing
- Scalable
- Improved security
- Specialized and productive management tool

◆ Major disadvantages

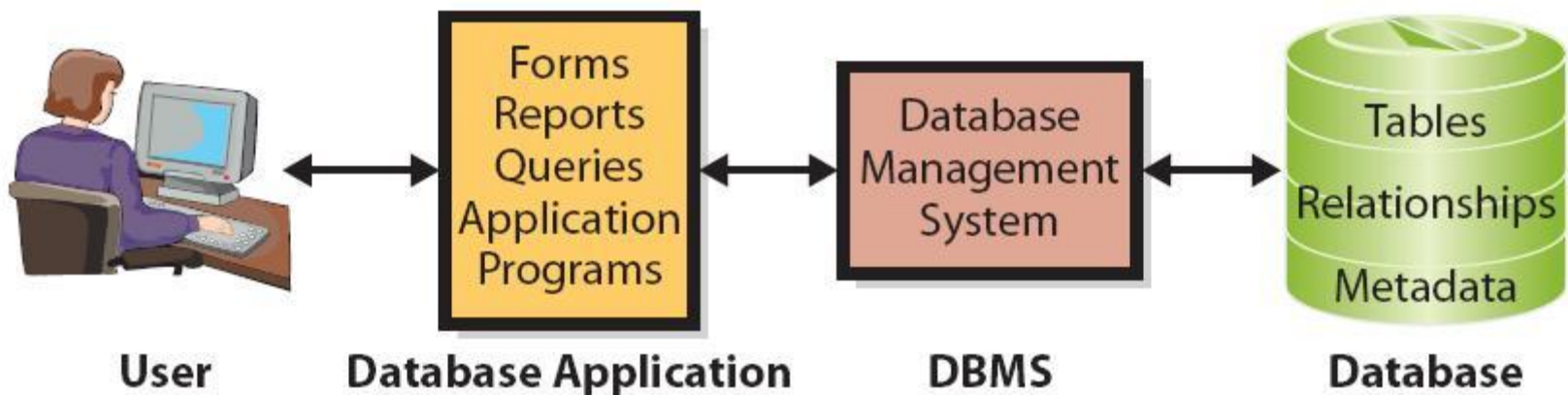
- Increased complexity
- Greater impact of failure

Types of Database

- ◆ Depending on how data are structured in the database:
 - Hierarchical
 - Network
 - Relational
 - Object-oriented
 - XML
- ◆ Relational database is the dominant type of today's database implementation.

Database System

- ◆ A database system is a complete information system
- ◆ Basic layers of a database system

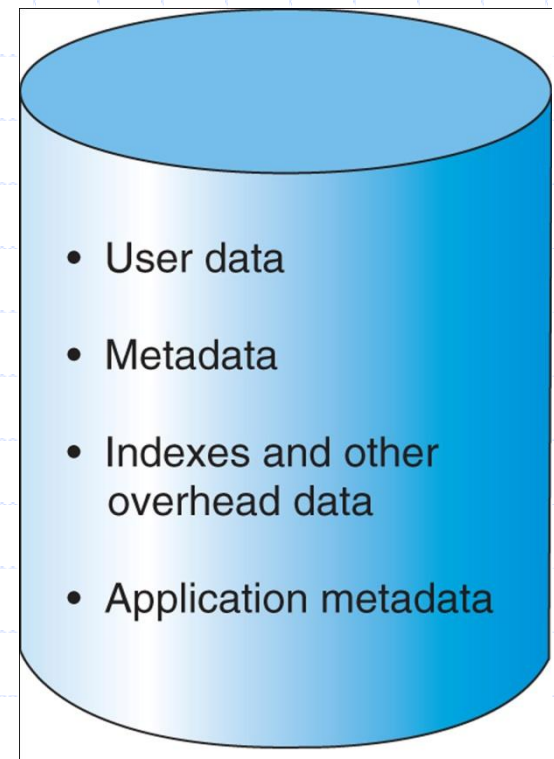


Database

◆ A database is a storage place for data

◆ What's in the database?

- Data (tables)
- Metadata
- Other data and structures



Metadata

◆ Metadata are data that describe data (data definitions)



◆ Metadata is always a part of a database.

Metadata defines tables, columns, data types, keys (relationships), constraints, etc.

USER_TABLES Table

TableName	NumberColumns	PrimaryKey
STUDENT	3	StudentNumber
CLASS	4	ClassNumber
GRADE	3	(StudentNumber, ClassNumber)

USER_COLUMNS Table

ColumnName	TableName	DataType	Length (bytes)
StudentNumber	STUDENT	Integer	4
LastName	STUDENT	Text	25
FirstName	STUDENT	Text	25
EmailAddress	STUDENT	Text	100
ClassNumber	CLASS	Integer	4
Name	CLASS	Text	25
Term	CLASS	Text	12
Section	CLASS	SmallInteger	2
StudentNumber	GRADE	Integer	4
ClassNumber	GRADE	Integer	4
Grade	GRADE	Decimal	(2, 1)

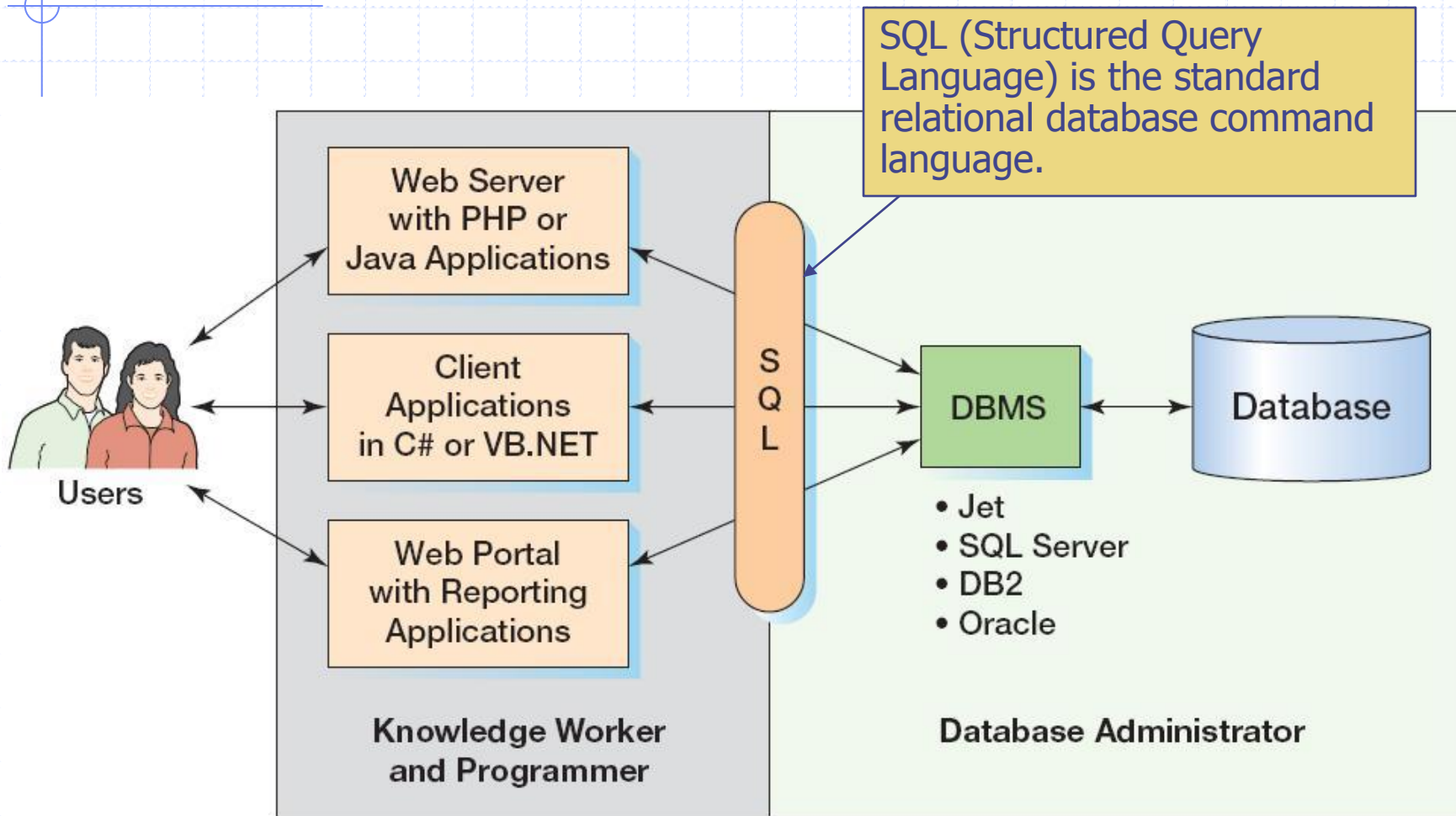
Database Management System

- ◆ DBMS serves as a controller (gatekeeper) for databases
- ◆ DBMS provides common functionalities and interfaces for managing and controlling database activities, such as
 - creating and maintaining databases and other structures
 - reading, updating and deleting data
 - data backup and recovery
 - controlling concurrency, consistency, and enforcing other rules
 - providing security

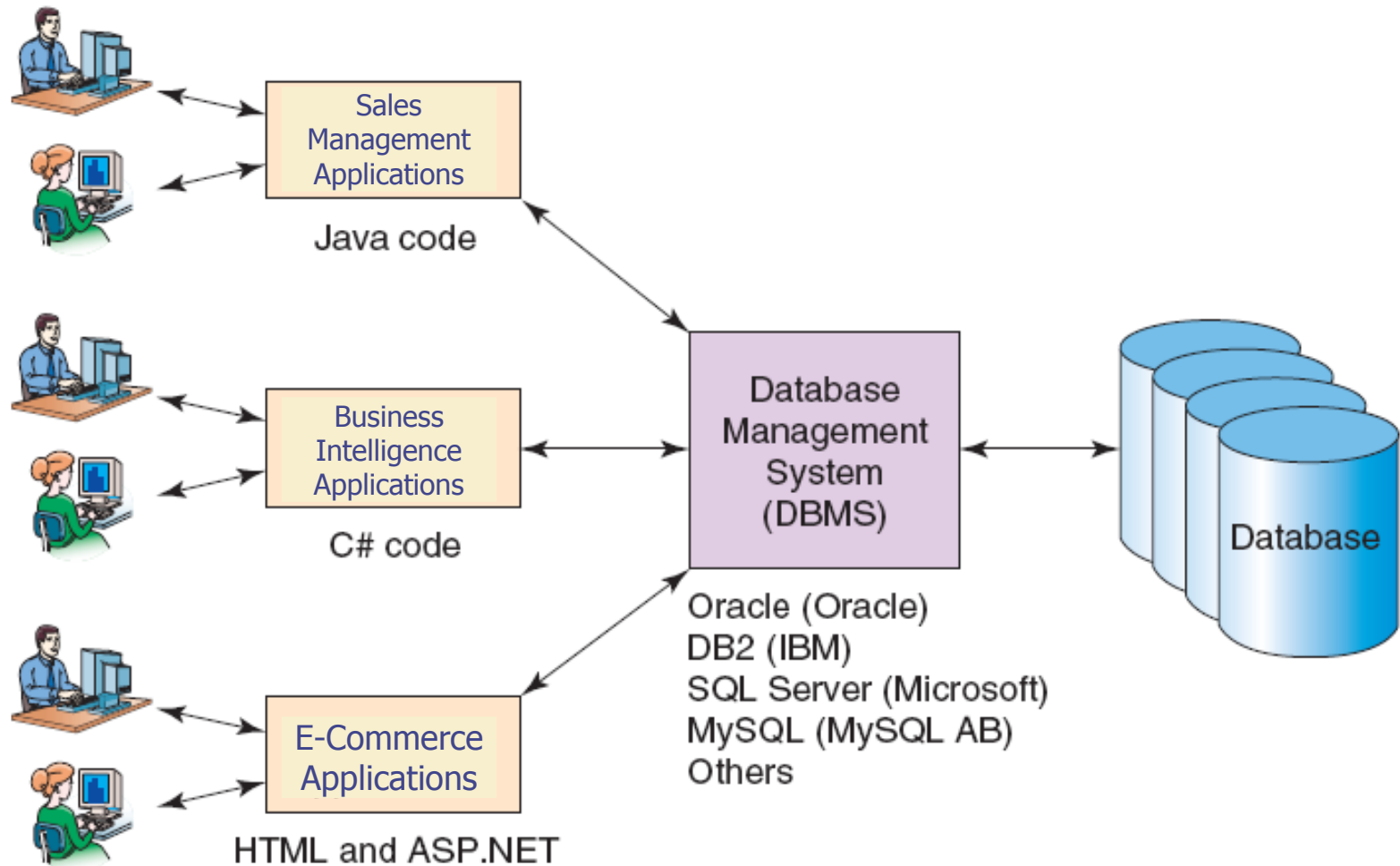
Applications/Programs

- ◆ Applications (programs) use the data in databases to perform all kinds of information processing tasks
- ◆ Example functions
 - Executing business logic/process (calculation, analysis, etc.)
 - Providing a user interface (forms, reports, etc.) to support interactions (data entry, query, navigation, etc.)
 - Presenting data in reports
 - ...

Database Users



Enterprise Database System



Enterprise vs. Personal Database System

Enterprise

- Large organizational and workgroup databases
- Thousands of users
- Many different database applications
- 24/7 operations

Personal

- Smaller, simpler applications
- Personal or small workgroup applications
- Often includes applications in the database
- More similar to file processing

Major Commercial Products

- ◆ Desktop database system / personal database
 - Microsoft Access: a low-end product intended for individual users and small workgroups.
- ◆ Enterprise level database systems
 - Oracle (Oracle)
 - DB2 (IBM)
 - SQL Server (Microsoft)
 - MySQL (Oracle)

Summary

◆ Key concepts

- Compare file processing vs. database processing
- Database features, advantages and disadvantages
- Components (layers) of database systems
 - ◆ User, Application, DBMS, Database
- Data vs. metadata
- DBMS
- SQL
- Enterprise vs. personal database system
- Know major commercial DBMS products

Good Readings and Resources

◆ A short database history

- <http://math.hws.edu/vaughn/cpsc/343/2003/history.html>

◆ The database report, July 2010

- <http://www.tdan.com/view-featured-columns/14063>

◆ Largest databases

- http://wintercorp.com/VLDB/2005_TopTen_Survey/TopTenWinners_2005.asp