Lecture 1

Types of Databases and Database Applications

- Databases and database systems are an essential component of life in modern society.
    - Eg: if we go to the bank to deposit or withdraw funds,
    - if we make a hotel or airline reservation,
    - if we access a computerized library catalog to search for a bibliographic item,
    - if we purchase something online—such as a book, toy, or computer—chances are that our activities will involve someone or some computer program accessing a database.
- These interactions are examples of **traditional database applications**, in which most of the information that is stored and accessed is either textual or numeric.

Contd…

- The proliferation of social media Web sites, such as Facebook, Twitter, and Flickr, has required the creation of huge databases that store nontraditional data, such as posts, tweets, images, and video clips.
- New types of database systems, often referred to as **big data** storage systems, or **NOSQL** systems, have been created to manage data for social media applications.
- The wide availability of photo and video technology on cellphones and other devices has made it possible to store images, audio clips, and video streams digitally.
- These types of files are becoming an important component of **multimedia databases**.

Contd…

- **Geographic information systems (GISs)** can store and analyze maps, weather data, and satellite images.

- **Data warehouses and online analytical processing (OLAP)** systems are used in many companies to extract and analyze useful business information from very large databases to support decision making.

- **Real-time and active database** technology is used to control industrial and manufacturing processes.

- And **database search** techniques are being applied to the World Wide Web to improve the search for information that is needed by users browsing the Internet.

Introduction

- A **database** is a collection of related data.
- By data, we mean known facts that can be recorded and that have implicit meaning (For example, consider the names, telephone numbers, and addresses of the people you know).
- This data can be recorded in an indexed address book or stored on a hard drive, using a personal computer and software such as Microsoft Access or Excel.
- This collection of related data with an implicit meaning is a database.

Contd…

- A database has the following implicit properties:
  - A database represents some aspect of the real world, sometimes called the **miniworld** or the **universe of discourse (UoD)**. Changes to the miniworld are reflected in the database.
  - A database is a logically coherent collection of data with some inherent meaning.
  - A database is designed, built, and populated with data for a specific purpose.
  - It has an intended group of users and some preconceived applications in which these users are interested.

Contd…

- A database has some source from which data is derived, some degree of interaction with events in the real world, and an audience that is actively interested in its contents.
- The end users of a database may perform business transactions or events may happen that cause the information in the database to change.
- In order for a database to be accurate and reliable at all times, it must be a true reflection of the miniworld that it represents; therefore, changes must be reflected in the database as soon as possible.

- A database can be of any size and complexity.
  - For example, the list of names and addresses may consist of only a few hundred records, each with a simple structure.
  - Database of greater size and complexity would be maintained by a social media company such as Facebook, which has more than a billion users.

Contd…

- A **database management system (DBMS)** is a computerized system that enables users to create and maintain a database.
- The **DBMS** is a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications.
- **Defining a database** involves specifying the data types, structures, and constraints of the data to be stored in the database.

- The database definition or descriptive information is also stored by the DBMS in the form of a database catalog or dictionary; it is called **meta-data.**
- **Constructing the database** is the process of storing the data on some storage medium that is controlled by the DBMS.
- **Manipulating a database** includes functions such as querying the database to retrieve specific data, updating the database to reflect changes in the miniworld, and generating reports from the data.
- **Sharing a database** allows multiple users and programs to access the database simultaneously.

Contd…

- An **application program** accesses the database by sending queries or requests for data to the DBMS.
- A **query** typically causes some data to be retrieved; a transaction may cause some data to be read and some data to be written into the database.
- Other important functions provided by the DBMS include **protecting the database and maintainin**g it over a long period of time.
- Protection includes system protection against hardware or software malfunction (or crashes) and security protection against unauthorized or malicious access.
- A typical large database may have a life cycle of many years, so the DBMS must be able to maintain the database system by allowing the system to evolve as requirements change over time.

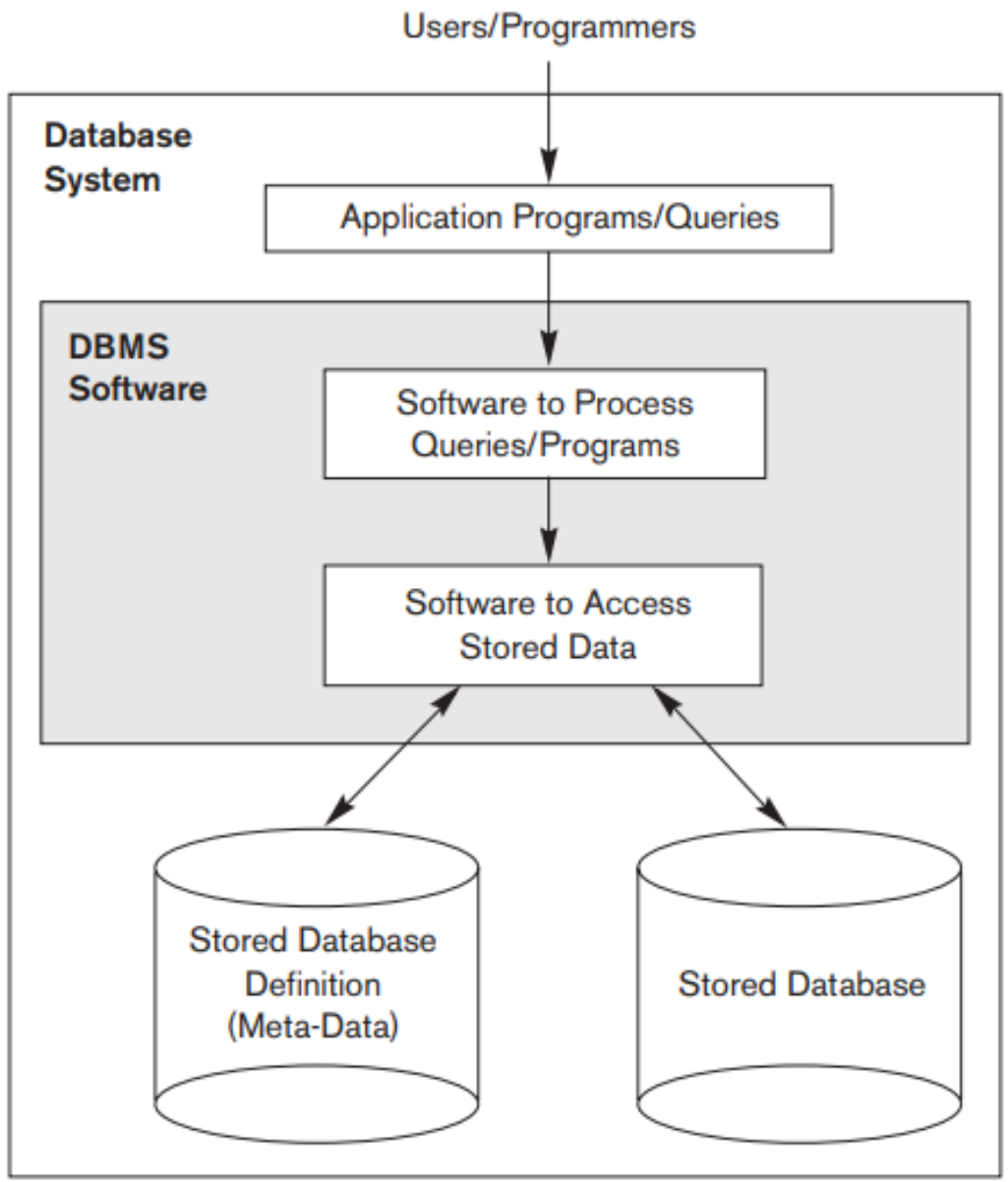

Figure 1.1
A simplified database system environment.

Examples

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Characteristics of the Database Approach

- Self-describing nature of a database system
- Insulation between programs and data, and data abstraction
- Support of multiple views of the data
- Sharing of data and multiuser transaction processing

Self-describing nature of a database system

- A fundamental characteristic of the database approach is that the database system contains not only the database itself but also a complete definition or description of the database structure and constraints.

- This definition is stored in the DBMS catalog, which contains information such as the structure of each file, the type and storageformat of each data item, and various constraints on the data.

- The information stored in the catalog is called meta-data, and it describes the structure of the primary database.

Contd…

- The catalog is used by the DBMS software and also by database users who need information about the database structure.
- A general-purpose DBMS software package is not written for a specific database application.
- Therefore, it must refer to the catalog to know the structure of the files in a specific database, such as the type and format of data it will access.

Insulation between Programs and Data, and Data Abstraction

- The structure of data files is stored in the DBMS catalog separately from the access programs (**program-data independence**).
- In some types of database systems, such as object-oriented and object-relational systems, users can define operations on data as part of the database definitions.
- An operation (also called a function or method) is specified in two parts.
- The interface (or signature) of an operation includes the operation name and the data types of its arguments (or parameters).
- The implementation (or method) of the operation is specified separately and can be changed without affecting the interface.
- User application programs can operate on the data by invoking these operations through their names and arguments, regardless of how the operations are implemented (**program-operation independence)**.

Contd…

- The characteristic that allows program-data independence and program-operation independence is called **data abstraction**.
- A DBMS provides users with a conceptual representation of data that does not include many of the details of how the data is stored or how the operations are implemented.
- Informally, a **data model** is a type of data abstraction that is used to provide this conceptual representation.
- The data model uses logical concepts, such as objects, their properties, and their interrelationships, that may be easier for most users to understand than computer storage concepts.
- Hence, the data model hides storage and implementation details that are not of interest to most database users.

Support of multiple views of the data

- A database typically has many types of users, each of whom may require a different perspective or view of the database.
- A view may be a subset of the database or it may contain virtual data that is derived from the database files but is not explicitly stored.
- A multiuser DBMS whose users have a variety of distinct applications must provide facilities for defining multiple views.

TRANSCRIPT

Student_name	Student_transcript				
	Course_number	Grade	Semester	Year	Section_id
Smith	CS1310	C	Fall	08	119
	MATH2410	B	Fall	08	112
Brown	MATH2410	A	Fall	07	85
	CS1310	A	Fall	07	92
	CS3320	B	Spring	08	102
	CS3380	A	Fall	08	135

(a)

COURSE_PREREQUISITES

Course_name	Course_number	Prerequisites
Database	CS3380	CS3320
		MATH2410
Data Structures	CS3320	CS1310

(b)

Figure 1.5
Two views derived from the database in Figure 1.2. (a) The TRANSCRIPT view. (b) The COURSE_PREREQUISITES view.

Sharing of data and multiuser transaction processing

- A multiuser DBMS, as its name implies, must allow multiple users to access the database at the same time.
- This is essential if data for multiple applications is to be integrated and maintained in a single database.
- The DBMS must include concurrency control software to ensure that several users trying to update the same data do so in a controlled manner so that the result of the updates is correct.
- For example, when several reservation clerks try to assign a seat on an airline flight, the DBMS should ensure that each seat can be accessed by only one clerk at a time for assignment to a passenger. These types of applications are generally called **online transaction processing (OLTP)** applications.
- A fundamental role of multiuser DBMS software is to ensure that concurrent transactions operate correctly and efficiently

Database Users

- **Database Adminstrators**
  - Administering database, DBMS and related software is the responsibility of the **database administrator (DBA)**.
  - The DBA is responsible for authorizing access to the database, coordinating and monitoring its use, and acquiring software and hardware resources as needed.
  - The DBA is accountable for problems such as security breaches and poor system response time.

- **Database Designers**

  - Database designers are responsible for identifying the data to be stored in the database and for choosing appropriate structures to represent and store this data.
  - These tasks are mostly undertaken before the database is actually implemented and populated with data.
  - It is the responsibility of database designers to communicate with all prospective database users in order to understand their requirements and to create a design that meets these requirements.
  - Database designers typically interact with each potential group of users and develop views of the database that meet the data and processing requirements of these groups.
  - Each view is then analyzed and integrated with the views of other user groups.
  - The final database design must be capable of supporting the requirements of all user groups.

End Users

- End users are the people whose jobs require access to the database for querying, updating, and generating reports; the database primarily exists for their use.
- There are several categories of end users:

1. **Casual end users**
   - occasionally access the database, but they may need different information each time.
   - They use a sophisticated database query interface to specify their requests and are typically middle- or high-level managers or other occasional browsers.

2. Naive or parametric end users

- Their main job function revolves around constantly querying and updating the database, using standard types of queries and updates— called canned transactions—that have been carefully programmed and tested.
- The tasks that such users perform are varied.
  - Bank customers and tellers check account balances and post withdrawals and deposits.
  - Reservation agents or customers for airlines, hotels, and car rental companies check availability for a given request and make reservations.
  - Employees at receiving stations for shipping companies enter package identifications via bar codes and descriptive information through buttons to update a central database of received and in-transit packages.
  - Social media users post and read items on social media Web sites.

Sophisticated end users

Sophisticated end users include engineers, scientists, business analysts, and others who thoroughly familiarize themselves with the facilities of the DBMS in order to implement their own applications to meet their complex requirements.

Standalone users

Standalone users maintain personal databases by using ready-made program packages that provide easy-to-use menu-based or graphics-based interfaces.

An example is the user of a financial software package that stores a variety of personal financial data.

Contd…

- A typical DBMS provides multiple facilities to access a database.
- Naive end users need to learn very little about the facilities provided by the DBMS; they simply have to understand the user interfaces of the mobile apps or standard transactions designed and implemented for their use.
- Casual users learn only a few facilities that they may use repeatedly.
- Sophisticated users try to learn most of the DBMS facilities in order to achieve their complex requirements.
- Standalone users typically become very proficient in using a specific software package.

System Analysts and Application Programmers (Software Engineers)

- **System analysts** determine the requirements of end users, especially naive and parametric end users, and develop specifications for standard canned transactions that meet these requirements.
- Application programmers implement these specifications as programs; then they test, debug, document, and maintain these canned transactions.
- Such analysts and programmers—commonly referred to as s**oftware developers or software engineers**—should be familiar with the full range of capabilities provided by the DBMS to accomplish their tasks.

Advanatages of using the DBMS approach

- Controlling redundancy in data storage and in development and maintenance efforts.
  - Sharing of data among multiple users.
- Restricting unauthorized access to data.
- Providing Storage Structures (e.g. indexes) for efficient Query Processing
- Providing backup and recovery services.

Contd…

- Providing multiple interfaces to different classes of users.
- Representing complex relationships among data.
- Enforcing integrity constraints on the database.
- Drawing inferences and actions from the stored data using deductive and active rules.
- Reduced application development time:
  - Incremental time to add each new application is reduced.

- Flexibility to change data structures:
  - Database structure may evolve as new requirements are defined.

- Availability of current information:
  - Extremely important for on-line transaction systems such as airline, hotel, car reservations.