Unit I: Introduction to Database

Chapter 01: Databases and Database Users

Basic Definitions:

Database: A collection of related data.

Data: Known facts that can be recorded and have an implicit meaning.

Mini-world: Some part of the real world about which data is stored in a database. For example, student grades and transcripts at a university.

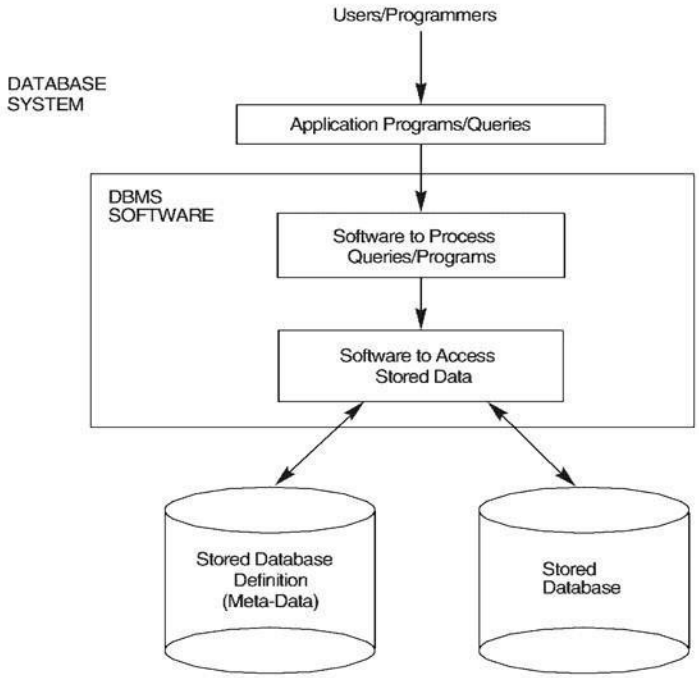
Database Management System (DBMS): A software package/ system to facilitate the creation and maintenance of a computerized database.

Database System: The DBMS software together with the data itself. Sometimes, the applications are also included.

Components of a DBMS

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

Figure 1.1 A simplified database system environment, illustrating the concepts and terminology discussed in Section 1.1.



Example of a Database(with a conceptual Data Model)

Mini-world for the example: Part of a UNIVERSITY environment.

Some mini-world *entities*:
- STUDENTs
- COURSEs
- SECTIONs (of COURSEs)
- (academic) DEPARTMENTs
- INSTRUCTORs

Some mini-world *relationships*:
- SECTIONs *are of* specific COURSEs
- STUDENTs *take* SECTIONs
- COURSEs *have* prerequisite COURSEs
- INSTRUCTORs *teach* SECTIONs
- COURSEs *are offered by* DEPARTMENTs
- STUDENTs *major in* DEPARTMENTs

NOTE: The above could be expressed in the *ENTITYRELATIONSHIP* data model.

Figure 1.2 An example of a database that stores student records and their grades.

STUDENT	Name	StudentNumber	Class	Major
	Smith	17	1	CS
	Brown	8	2	CS

COURSE	CourseName	CourseNumber	CreditHours	Department
	Intro to Computer Science	CS1310	4	CS
	Data Structures	CS3320	4	CS
	Discrete Mathematics	MATH2410	3	MATH
	Database	CS3380	3	CS

SECTION	SectionIdentifier	CourseNumber	Semester	Year	Instructor
	85	MATH2410	Fall	98	King
	92	CS1310	Fall	98	Anderson
	102	CS3320	Spring	99	Knuth
	112	MATH2410	Fall	99	Chang
	119	CS1310	Fall	99	Anderson
	135	CS3380	Fall	99	Stone

GRADE_REPORT	StudentNumber	SectionIdentifier	Grade
	17	112	B
	17	119	C
	8	85	A
	8	92	A
	8	102	B
	8	135	A

PREREQUISITE	CourseNumber	PrerequisiteNumber
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

Characteristics of database approach

1)Self-describing nature of a database system:

A DBMS **catalog** stores the *description(*structure/data types/storage format/constraints *)* of the database. The description is called **meta-data**). This allows the DBMS software to work with different databases.

In File Processing System(FPS), the programs can work with only one specific database(i.e file) as the structure of data files is embedded in the application programs.

2) Data Abstraction:

A DBMS supports both **program-data independence** and **program-operation independence** which allow changing data storage structures and operations without having to change the DBMS access programs. For this, DBMS provides users with conceptual representation of data (Data Model) that include only logical concepts (*objects & their properties/inter-relationships*) and hides the details of how the data is stored and how operations are implemented.

In FPS, any changes to the structure of a file may require changing all programs that access to this file as the structure of data files is embedded in the application programs.

3)Support of multiple views of the data:

A DBMS provides facilities for defining a multiple views for multiple users. Each user may see a different perspective or view of the database, which describes *only* the data of interest to that user.

A view may be subset of the database or it may contain virtual data that is derived from the database but is not explicitly stored.

4)Sharing of data and multiuser transaction processing :

A DBMS allows multiple users to have concurrent accesses to the database. Concurrency control within the DBMS guarantees that each **transaction** is correctly executed or completely aborted. OLTP (Online Transaction Processing) is a major part of database applications.

The concept of transaction is the process that includes one or more database accesses such as reading or updating of database records.

Database Users

Users may be divided into

- those who actually use and control the content (called "Actors on the Scene") and
- those who enable the database to be developed and the DBMS software to be designed and implemented (called "Workers Behind the Scene").

Actors on the scene:

- **Database administrators:** responsible for authorizing access to the database, for co-ordinating and monitoring its use, acquiring software, and hardware resources, controlling its use and monitoring efficiency of operations.
- **Database Designers:** responsible to define the content, the structure, the constraints, and functions or transactions against the database. They must communicate with the end-users and understand their needs.
- **End-users:** they use the data for queries, reports and some of them actually update the database content.

- ✓ **Casual End-users**: access database occasionally when needed
- ✓ **Naïve or Parametric End-users** : they make up a large section of the end-user population. They use previously well-defined functions in the form of "canned transactions" against the database. Examples are bank-tellers or reservation clerks who do this activity for an entire shift of operations.
- ✓ **Sophisticated End-users** : these include business analysts, scientists, engineers, others thoroughly familiar with the system capabilities. Many use tools in the form of software packages that work closely with the stored database.

Stand-alone End-users: mostly maintain personal databases using ready-to-use packaged applications. An example is a tax program user that creates his or her own internal database

Workers behind the scene:

1. DBMS system designers and implementers
2. Tool developer
3. Operators and maintaince personal

Advantages of Using the Database Approach

1. Controlling redundancy:

By integrating the views of different user groups during database design, all logically related data are placed together in the database. This ensures the data consistency and saves the storage spaces as well as maintenance efforts.

In FPS, every user group maintains its own files for handling its data processing applications and exists duplication of some or all of the data in these various files. This data redundancy will lead to several problems such as data inconsistent, wastage of storage space and duplication of efforts in maintaining the several data files

2. Providing Data Integrity and Security

A DBMS provides capabilities for defining and enforcing certain integrity constraints (e.g specifying data type/key/NOT NULL value)that must be hold for the data. These constraints are derived from the meaning of the data, specified to the DBMS and automatically enforced to ensure integrity of data on the database.

A DBMS provides a security and authorization subsystem to ensure the security of data. DBA creates the accounts for each group of users and specify the account restrictions to be enforced automatically by the DBMS when the user has un-authorize access.

3. Providing Storage Structure for Efficient Data Access

The database is typically store on the external storage device. A DBMS provides specialized data structure(Indexes) to speed up disk search for desire records in the database. Therefore, queries and updates get processed efficiently and optimized.

4. Providing Backup and Recovery Services

A DBMS provides facilities for recovering from hardware/software failures. If the computer system fails in the middle of update transaction, DBMS ensures either of:

--the database is restored to the state it was in before the transaction started executing.

--the transaction is resumed from the point at which was interrupted so that its full effect is recorded in the database

5. Providing Data Abstraction.

A DBMS provides an abstract view of the data to users so that users do not need to know the details of data representation and storage for writing the application programs that access the data on the database.

Types of Databases and Database Applications

- Numeric and Textual Databases
- Multimedia Databases
- Geographic Information Systems (GIS)
- Data Warehouses
- Real-time and Active Databases