DATABASE MANAGEMENT SYSTEMS

SUBJECT CODE:
3132

MODULE I

SYLLABUS

MODULE – I Database systems

1.1 To Understand Database systems
1.1.1 Define data, information, field, record, file, and database
1.1.2 Define DBMS
1.1.3 Explain the advantages of DBMS
1.1.4 Describe the applications of DBMS
1.1.5 List the Database Users.
1.1.6 Define instance, schema and subschema
1.1.7 Explain Three Schema architecture with diagram
1.1.8 Explain Data Independence – Logical Data Independence and Physical Data Independence
1.1.9 Describe Conceptual Model, representation and physical model
1.1.10 Explain hierarchical, network and relational models
1.1.11 Describe DBMS Languages – DDL, DML and DCL
1.1.12 Explain Component Modules of DBMS
1.1.13 Centralised and Client-Server Database Systems

DBMS BASIC TERMINOLOGIES

Data: Data is the known facts or figures that have implicit meaning. It can also be defined as it is the representation of facts ,concepts or instruction in a formal manner, which is suitable for understanding and processing. Data can be represented in alphabets(A-Z, a-z),in digits(0-9) and using special characters(+,-.#,$, etc).It's raw and unprocessed.
e.g: 25, "ajit" etc.

Information:
Information is the processed data on which decisions and actions are based. Information can be defined as the organized and classified data to provide meaningful values
.
Eg: "The age of Ravi is 25"

Field

In a **database** table, a **field** is a data structure for a single piece of data. The term "**fields**" refers to columns, or vertical categories of data.

Record - Related fields of data are grouped to form a *record.* Thus, a record represents a collection of *attributes* that describe an *entity*. *Fixed-length* records contain a fixed number of fixed-length data fields. *Variable-length* records contain a variable number of fields and field lengths.

Eg: Customer contact information

Row:ID no,names,steet address,city,telephoneno

File:
Collection of related record is known as a data file. It stored in secondary memory.

Database

In order to overcome the limitation of a file system, a new approach was required. Hence a database approach emerged.

- Collection of interrelated data that is organized.So that it can be easily accessed,managed and updated.
- Data is organized in to rows,columns and tables
- Database can be software based or hardware based, with one sole purpose, storing data
- A database can be of any size and varying complexity
- A small database can be handled manually but for a large database and having multiple users it is difficult to maintain it, In that case a computerized database is useful.
- It is accessible in a logical order without any difficult

for example:

consider the roll no, name, address of a student stored in a student file. It is collection of related data with an implicit meaning.

Advantages of database system over traditional, paper based methods of record keeping are:

· **compactness**:

No need for large amount of paper files

· **speed**:

The machine can retrieve and modify the data more faster way then human being

· **Accuracy:** Accurate,up-to-date information is fetched as per requirement of the user at any time.

Database System:

It is computerized system, whose overall purpose is to maintain the information and to make that the information is available on demand.

Advantages:

1.Redundency can be reduced.
2.Inconsistency can be avoided.
3.Data can be shared.

4.Standards can be enforced.
5.Security restrictions can be applied.
6.Integrity can be maintained.
7.Data gathering can be possible.
8.Requirements can be balanced

Database Management System (DBMS)

- A DBMS is a general purpose software that allows creation, definition and manipulation of database,
  allowing users to store, process and analyse data easily.
- DBMS provides us with an interface or a tool, to perform various operations like creating database, storing data in it, updating data, creating tables in the database and a lot more.
- DBMS also provides protection and security to the databases. It also maintains data consistency in case of multiple users.
- Here are some examples of popular DBMS used these days:
  -MySql
  -Oracle
  -SQL Server,IBM DB2
  -Amazon SimpleDB (cloud based) etc.

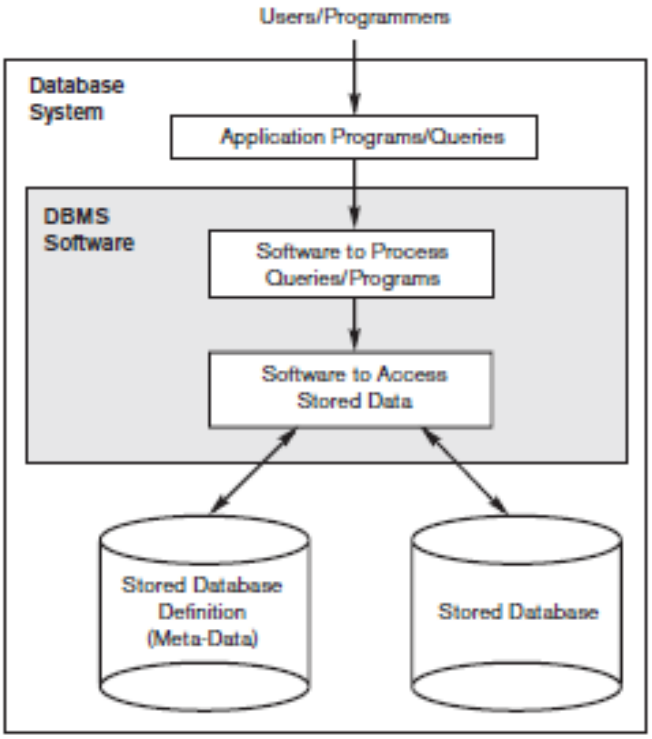


Figure 1.1
A simplified database system environment.

Characteristics of Database Management System

A database management system has following characteristics:

1. ***Data stored into Tables****:* Data is never directly stored into the database. Data is stored into tables, created inside the database. DBMS also allows to have relationships between tables which makes the data more meaningful and connected. You can easily understand what type of data is stored where by looking at all the tables created in a database.
2. ***Reduced Redundancy****:* In the modern world hard drives are very cheap, but earlier when hard drives were too expensive, unnecessary repetition of data in database was a big problem. But DBMS follows Normalisation which divides the data in such a way that repetition is minimum.
3. ***Data Consistency****:* On Live data, i.e. data that is being continuously updated and added, maintaining the consistency of data can become a challenge. But DBMS handles it all by itself.
4. ***Support Multiple user and Concurrent Access****:* DBMS allows multiple users to work on it(update, insert, delete data) at the same time and still manages to maintain the data consistency.
5. ***Query Language****:* DBMS provides users with a simple Query language, using which data can be easily fetched, inserted, deleted and updated in a database.
6. ***Security****:* The DBMS also takes care of the security of data, protecting the data from unauthorized access. In a typical DBMS, we can create user accounts with different access permissions, using which we can easily secure our data by restricting user access.
7. ***DBMS supports transactions***, which allows us to better handle and manage data integrity in real world applications where multi-threading is extensively used.

Function of DBMS:

1. **Defining database schema:** 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**.
2. **Constructing database schema:** Constructing the database is the process of storing the data on some storage medium that is controlled by the DBMS.
3. **Manipulation of the database:**Manipulating a database includes functions such as querying the database to retrieve specific data, updating the database to reflect changes in theminiworld, and generating reports from the data.
4. **Sharing of database:** Sharing a database allows multiple users and programs to access the database simultaneously.

5. **Protection of database:** protecting the database and *maintaining* 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.
6. **Database recovery:** If for any reason the system fails DBMS must facilitate database recovery.

Advantages of dbms

File Oriented approach:

The traditional file oriented approach to information processing has for each application a separate master file and its own set of personal file. In file oriented approach the program dependent on the files and files become dependent on the files and files become dependents upon the programs

Disadvantages of file oriented approach:

1) Data redundancy and inconsistency: The same information may be written in several files. This redundancy leads to higher storage and access cost. It may lead data inconsistency that is the various copies of the same data may longer agree for example a changed customer address may be reflected in single file but not else where in the system.
2) Difficulty in accessing data : The conventional file processing system do not allow data to retrieved in a convenient and efficient manner according to user choice.
3) Data isolation : Because data are scattered in various file and files may be in different formats with new application programs to retrieve the appropriate data is difficult.
4) Integrity Problems: Developers enforce data validation in the system by adding appropriate code in the various application program. How ever when new constraints are added, it is difficult to change the programs to enforce them.
5) Atomicity: It is difficult to ensure atomicity in a file processing system when transaction failure occurs due to power failure, networking problems etc.
(atomicity: either all operations of the transaction are reflected properly in the database or non are)
6) Concurrent access: In the file processing system it is not possible to access a same file for transaction at same time
7) Security problems:
There is no security provided in file processing system to secure the data from unauthorized user access.

Advantages of dbms

- **Reduction of redundancies:**Centralized control of data by the DBA avoids unnecessary duplication of data and effectively reduces the total amount of data storage required avoiding duplication in the elimination of the inconsistencies that tend to be present in redundant data files.
- **Sharing of data:**A database allows the sharing of data under its control by any number of application programs or users.

- **Data Integrity:**Data integrity means that the data contained in the database is both accurate andconsistent. Therefore data values being entered for storage could be checked to ensure that they fall with in a specified range and are of the correct format.
- **Data Security:**The DBA who has the ultimate responsibility for the data in the dbms can ensure that proper access procedures are followed including proper authentication schemas for access to the DBS and additional check before permitting access to sensitive data.
- **Conflict resolution:**DBA resolve the conflict on requirements of various user and applications. The DBA chooses the best file structure and access method to get optional performance for the application.
- **Data Independence:**The ability to modify a schema definition in one level without affecting the schema definition in higher levels.

DISADVANTAGES OF DBMS

- It's Complexity
- Except MySQL, which is open source, licensed DBMSs are generally costly.
- They are large in size.

APPLICATIONS OF DBMS

- Banking: all transactions
- Airlines: reservations, schedules
- Universities: registration, grades
- Sales: customers, products, purchases
- Online retailers: order tracking, customized recommendations
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions

Database users

Many persons are involved in the design, use and maintenance of any database. These persons can be classified into 2 types as below.

a)Actors on the Scene

The people, whose jobs involve the day-to-day use of a database are called as 'Actors on the scene', Those who actually use and control the database content, and those who design, develop and maintain database applications .

Database Administrators (DBA): Many People use the same resources database,dbms and related software. The DBA is responsible for authorizing access to the database, for Coordinating and monitoring its use and for acquiring software and hardware resources as needed. These are the people, who maintain and design the database daily.

DBA responsible for the following

- Manages overall Database system
- Installation and upgradation of dbms server
- Design and implementation of database.
- Backup and recovery.
- Security
- Documentation

Database Designers: Database Designers begin task before the Database is implemented **.**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. It communicate with prospective database users to understand the requirement and to create the design.

End Users: People who wish to store and use data in a database. End users are the people whose jobs require access to the database for querying, updating and generating reports. They need not know about the working, DB design, and Access mechanism etc.They use developed applications to yet their task done.

1. **Casual end users:** occasionally access the database, but they may need different information each time.
   For example
   Student access result from university website

2. **Naïve/Parametric end user:** Uses the database very frequently only little knowledge about the structure of the database. They Work on developed application to get the result.

For example

Clerical staff in the bank

3. **Sophisticated end user**: Include engineers, scientist,business analyst and others who thoroughly familiarize themselves with the facilities of the dbms in order to implement their application to meet their complex requirement.
4. **Stand alone 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 tax package that stores a variety of personal financial data for tax purposes.

System Analyst:

These people determine the requirements of end users and develop specifications for transactions.

Application Programmers (Software Engineers):

These people can test, debug, document and maintain the specified transactions. They are software developers who interact with database using DML quries.

For example

Writing a c program to generate the report of employees who are working in particular department will involve a query to fetch the data from database.

b) Workers behind the scene

Those who design and develop the DBMS software and related tools, and the computer systems operators are called workers behind scene

Database Designers and Implementers: design and implement the DBMS modules and interfaces as a software package. A DBMS is a very complex software system that consists of many components, or **modules**, including modules for implementing the catalog, query language processing, interface processing, accessing and buffering data, controlling concurrency, and handling data recovery and security.

Tool Developers:

Include persons who design and implement tools consisting the packages for design, performance monitoring, and prototyping and test data generation.

Tool are optional packages that are often purchase seperately.

For example

package for database design.

Operators and Maintenance control:

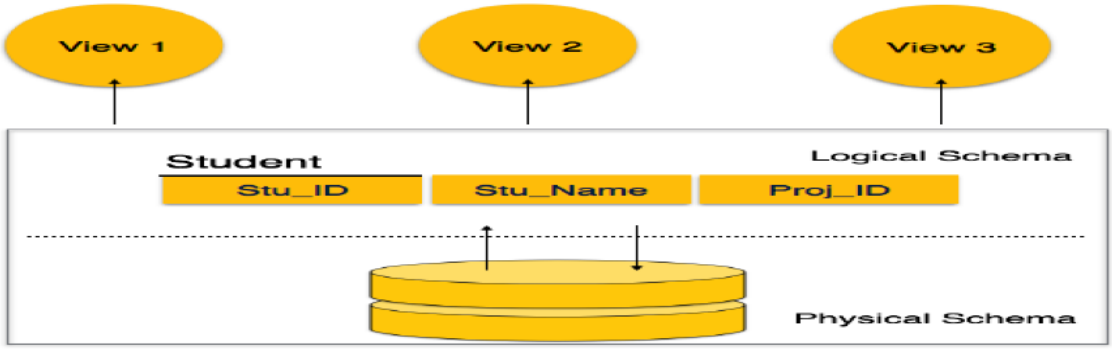
(system administration personnel)are responsible for the actual running and maintenance of the hardware and software environment for the database system.

SCHEMA, INSTANCE AND SUBSCHEMA

Schema :The description of a database is called the **database schema**, which is specified during database design and is not expected to change frequently. In other words, schema is an overall structure of a database.

- **Physical Schema:** The physical schema describes database design at physical level. Physical schema defines how database is stored. It defines how data is stored in secondary storage.

- **Logical schema:** The database design at the logical level is called logical schema. This schema defines logical structure of database. Usually programmer and DBA work at this level. Logical schema defines tables, constraints of a database.



A displayed schema is called a **schema diagram.** T]he diagram displays the structure of each record type but not the actual instances of records.

STUDENT

Name	Student_number	Class	Major

COURSE

Course_name	Course_number	Credit_hours	Department

Instance

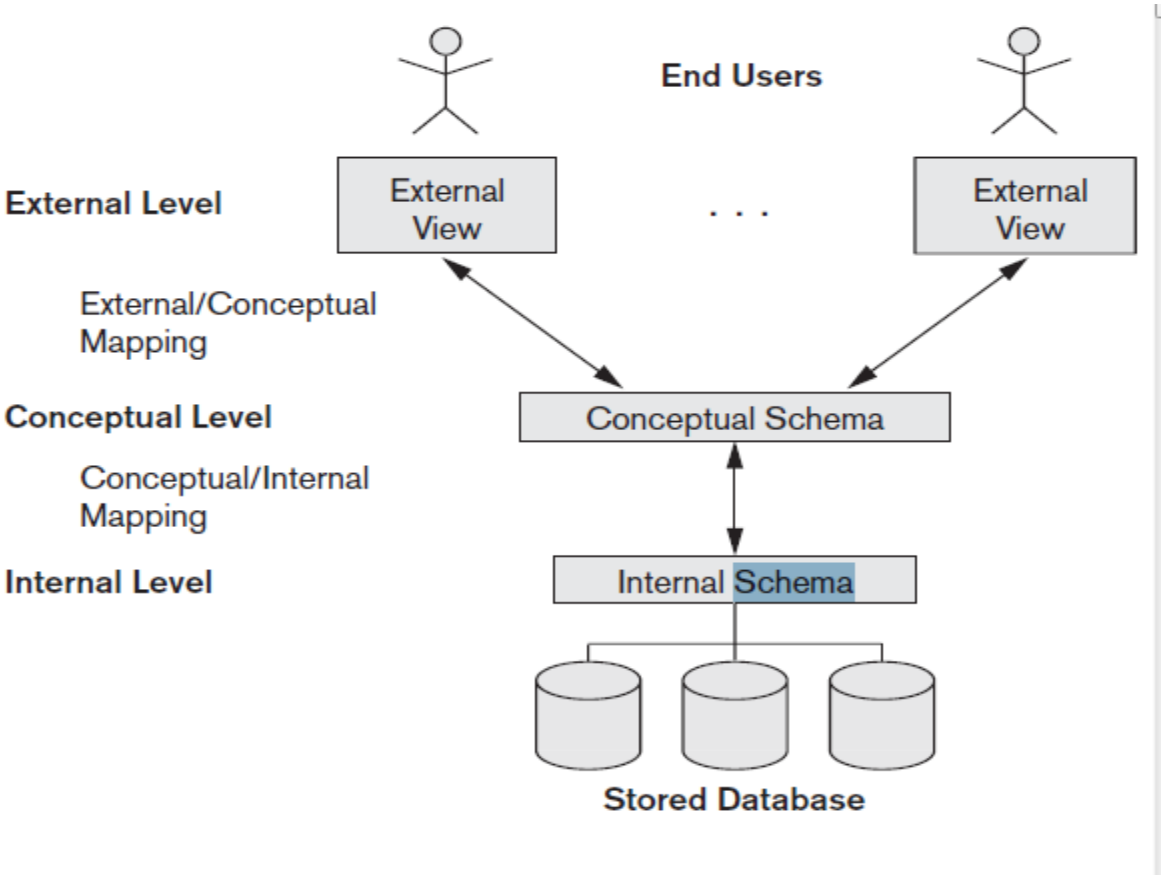
The data in the database at a particular moment in time is called a **database state** or **snapshot**. It is also called the *current* set of **occurrences** or **instances** in the database. Database changes every time when information is inserted and deleted to/from database .

- The distinction between database schema and database state is very important.When we **define** a new database, we specify its database schema only to the DBMS. At this point, the corresponding database state is the *empty state* with no data.We get the *initial state* of the database when the database is first **populated** or **loaded** with the initial data.

Subschema

Subschema is a sub part of a schema and inherits same properties of schema. Subschema describes different view of the database. Subschema is an application programmer's or user view of data item types and records type.

THREE SCHEMA ARCHITECTURE



- The **internal level /Physical level** has an **internal schema**,which describes the physical storage structure of the database. The internal schema uses a physical data model and describes the complete details of data storage and access paths for the database.It also describes how data are actually stored.
- The **conceptual level/logical level** has a **conceptual schema**, describes the structure of the whole database for a community of users. The conceptual schema hides the details of physical storage structures and concentrates on describing entities, data types, relationships, user operations, and constraints.
- The **external** or **view level** includes a number of **external schemas** or **userviews**. Each external schema describes the part of the database that a particular user group is interested in and hides the rest of the database from that user group.

The processes of transforming requests and results between levels are called **mappings**.

DATA INDEPENDENCE

The capacity to change the schema at one level of a database system without having to change the schema at the next higher level.We can define two types of data independence:

Logical data independence is the capacity to change the conceptual schema without having to change external schemas or application programs. We may change the conceptual schema to expand the database (by adding a record type or data item), to change constraints, or to reduce the database(by removing a record type or data item).
Physical data independence is the capacity to change the internal schema without having to change the conceptual schema. Hence, the external schemas need not be changed as well. Modification in the Physical level should not result in any changes in the Logical or View levels.

DATABASE LANGUAGES

Data Definition Languages (DDL)

Data Definition Language (DDL) statements are used to classify the database structure or schema. It
is a type of language that allows the DBA or user to depict and name those entities, attributes, and
relationships that are required for the application along with any associated integrity and security constraints. Here are the lists of tasks that come under DDL:

- **CREATE** - used to create objects in the database
- **ALTER** - used to alters the structure of the database
- **DROP** - used to delete objects from the database
- **TRUNCATE** - used to remove all records from a table, including all spaces allocated for the records are removed
- **COMMENT** - used to add comments to the data dictionary
- **RENAME** - used to rename an object

Data Manipulation Language (DML)

A language that offers a set of operations to support the fundamental data manipulation operations
on the data held in the database. Data Manipulation Language (DML) statements are used to manage data within schema objects. Here are the lists of tasks that come under DML:

- **SELECT** - It retrieves data from a database
- **INSERT** - It inserts data into a table
- **UPDATE** - It updates existing data within a table
- **DELETE** - It deletes all records from a table, the space for the records remain
- **MERGE** - UPSERT operation (insert or update)
- **CALL** - It calls a PL/SQL or Java subprogram
- **EXPLAIN PLAN** - It explains access path to data
- **LOCK TABLE** - It controls concurrency

Data Control Language (DCL)

The Data Control Language (DCL) is used to control privilege in Database. To perform any

operation in the database, such as for creating tables, sequences or views we need privileges.

Privileges are of two types,

- **System** - creating a session, table etc are all types of system privilege.
- **Object** - any command or query to work on tables comes under object privilege. DCL is used to define two commands. These are:
- **Grant** - It gives user access privileges to a database.
- **Revoke** - It takes back permissions from the user.

Transaction Control Language (TCL)

Transaction Control statements are used to run the changes made by DML statements. It allows statements to be grouped together into logical transactions.

- **COMMIT** - It saves the work done
- **SAVEPOINT** - It identifies a point in a transaction to which you can later roll back
- **ROLLBACK** - It restores database to original since the last COMMIT
- **SET TRANSACTION** - It changes the transaction options like isolation level and what rollback segment to use.

.