Paper II: CS-1102 Introduction to fundamental of DBMS

Chapter 1.

Introduction to File organization & DBMS

Chapter contents:

1.1 Introduction

1.2 Types of file organization

1.3 File system Vs DBMS

1.4 Data models

1.5 Levels of abstraction

1.6 Data independence

1.7 Structure of DBMS

1.8 Users of DBMS

1.9 Advantages of DBMS

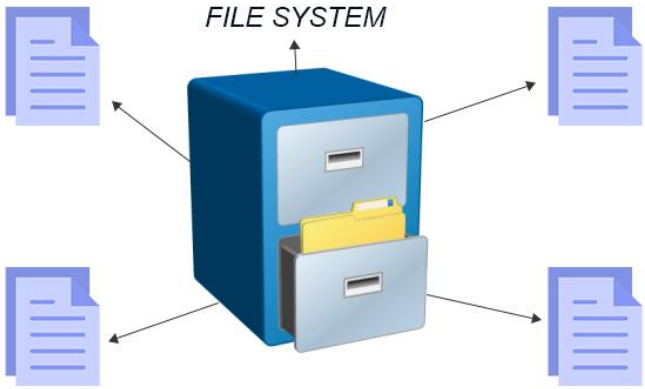
1.1 Introduction:

File system:

File: File is a collection of records which are logically related to any object. Record value can in any form like data.

For eg. : Each students records which having values of Roll no, Name, Class. For arranging data we use file.

For eg.: files of bank's customer, files of department, files of stack records etc. Files are recorded on secondary storage such as magnetic disks, magnetic tables and optical disks.



Types of files:

- **Physical file:**
  - Physical file concern with actual data that is stored.
  - It stores description about how the data is to be represented.
- **Logical file-**
  - Logical file: do not contain data.
  - They contain a description of records that are found in one or more physical files.
  - A logical file is a view or representation of one or more physical files.
- **Special character file:**
  - At the time of file creation we insert some special characters in file.
  - For eg: Control + z for end of a file which having ASCII value 26

According to records types of files:

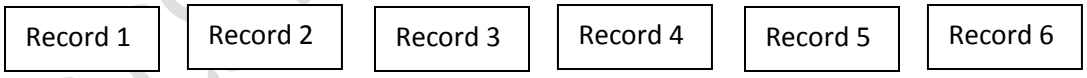
1. Fixed length record file
2. Variable length record file

1. Fixed length record file:

a. Every record in this file has same size(in bytes). Record having value set, in the fixed length record file, memory block are assign in same size. For eg., if the size for a record is assigned 30 bytes to each then records in this type are stored like as below,

Advantage: records are stored in fixed distance of memory block, so fast searching for a particular record is done.

Disadvantage: Memory blocks are unnecessarily used when record size is small as compared to assigned memory block. This useless memory block increases size of file.



2. Variable length record file:

a. Every record in this file has variable size (in bytes). Memory block are assign for a file records are in variable size. Different records in the file have different sizes. As per size of records value, memory blocks are used.

Advantage: Memory used efficiently for storing record. Whatever exact size of record that much size of memory block occupies in memory in this kind of records. Because of less memory they can move, save or transfer from one location to other in fast manner.

Disadvantage: Access for record is slower as compared to fixed length record file due to varying size of a record.

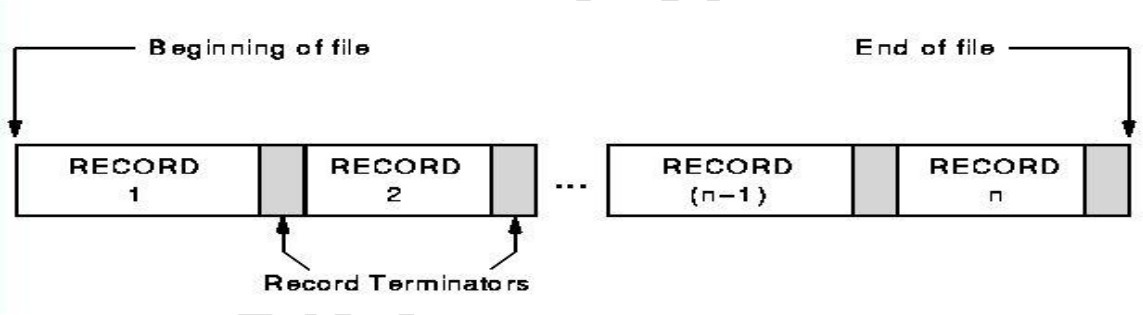
Record 1	Record 2	Record 3	Record 4	Record 5

File organization:

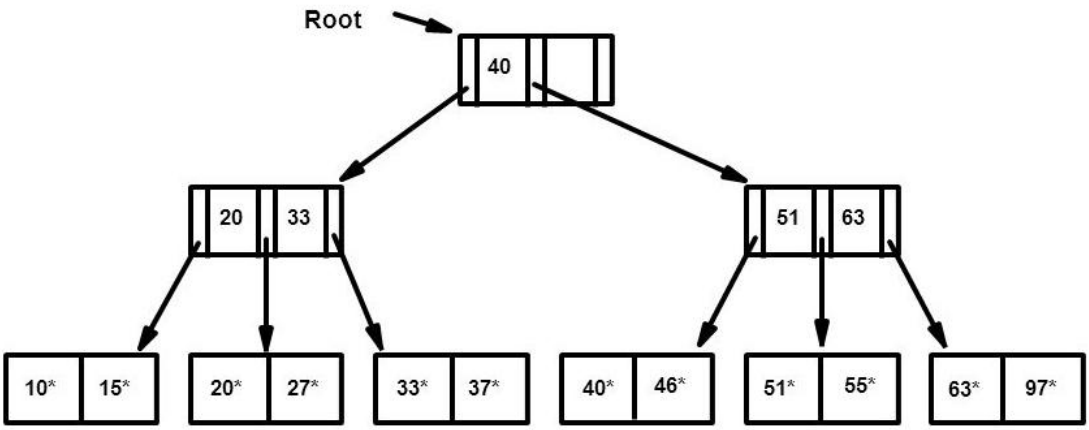
File organization refers to the logical relationships among various records that constitute the file, particularly with respect to the means of identification and access to any specific record. In short, storing the files in certain order is called file organization.

Types of file organization:

1. **Sequential file organization:** Sequential file organization is easiest method. In this method files are stored one after the other in a sequential manner.
   This method is also called as Pile or sorted file.
   This method is fast & efficient for huge amount of data.
   Sorted file is inefficient as it takes time & space for sorting records.



2. **Serial file organization:** Serial file organization is also called as heap file. In this method, records are inserted at the end of file into the data blocks. There is no requirement of sorting data.
   When huge amount of data is to be inserted at a time in a organization, that time this method is suitable.
   Accessing of data is slower as compared to sorted file method.
3. **Index Sequential Access Method (ISAM):** ISAM method is advanced sequential file organization. In this method, index value is generated and mapped with every record. Using that index, accessing of record is done.



4. **Random access / Direct access file organization:** In this file organization records are stored randomly but accessed directly. To access a file stored randomly, a record key is used to determine where a record is stored on the storage media. Magnetic and optical disks allow data to be stored and accessed randomly.

What is data?

- Data is nothing but Information or knowledge that we collect for specific purpose.
- For eg. –
    - Information about course
    - Knowledge about any skill like computer programming language, about any game etc.
    - Data about any inquiry for online reservation of bus tickets.
    - Data can be in any form like text, video, audio, images etc.

All these are considered as data.

But when we process on such type of data problems are arises related to handling such data. Any organization when want to store all the data which are related to them, then arranging such data , successfully process on that data, updating of that is little more complicated process. For solving this problem DBMS (Database Management System)are used. First step of DBMS is forming database.

What is database?

- Database(DB) is organized collection of facts(Data) that are arranged in a systematic manner.
- All data in DB are logically related.
- Data is typically stored electronically in computer system
- Data is operated through any DBMS.

- Databases make data management easy.
- For eg:
    - Electricity service provider use a database to manage billing, customer related issues etc.
    - Consider the facebook, it needs to store, manipulate and present data related to related to members, their friends, member activities, messages, advertisements etc.

Database Management System(DMBS) :

Database Management System (DBMS) is a collection of programs which enables users to access database, manipulate data, representation of data, control access to the database etc.

DBMS was firstly implemented in **1960's by Charles Bachmen** called as Integrated Data Store(IDS).

Need of DBMS- File system v/s DBMS :

In file system data was stored in different files, this method having many drawbacks :

1. **Data redundancy :**
   In file system, data is stored on multiple places. This leads to produce data repetition. DBMS reduces this redundancy.
2. **Data Isolation:**
   Data in file organization is scattered at multiple places. That's why accessing appropriate data is difficult. The data isolation property is strongly maintained in DBMS.
3. **Data access & updating:**
   Due to repetition of data accessing & updating is difficult.
4. **Data Integrity:**
5. In DBMS, data must satisfy integrity constraint. Integrity means level of accuracy can be maintained from initial stage to till execution complete. Integrity is does not hold in file system. Eg- At the admission time, some criteria's are applied for this. But after sometime if the norms are changed then it is complicated to change all the condition & data in file system.
6. **Concurrent access:**
   Multiple access or parallel access is very beneficial to database system for improving performance. File system gives serial access, DBMS facilitates

concurrent access. Eg. : In bank system multiple customers uses bank database & performs different tasks at a time successfully due to concurrent access.

7. **Security problems:**

   Data in file system is in ad-hoc manner(When necessary). So security level is low. DBMS having strong security majors.

Data abstraction:

Data abstraction is simplified view of an data in which ones required data level is provided & remaining data is hided from other.

Data abstraction is also called as Abstract Data Type(ADT) of 'User Defined Type'.

Levels of data abstraction:

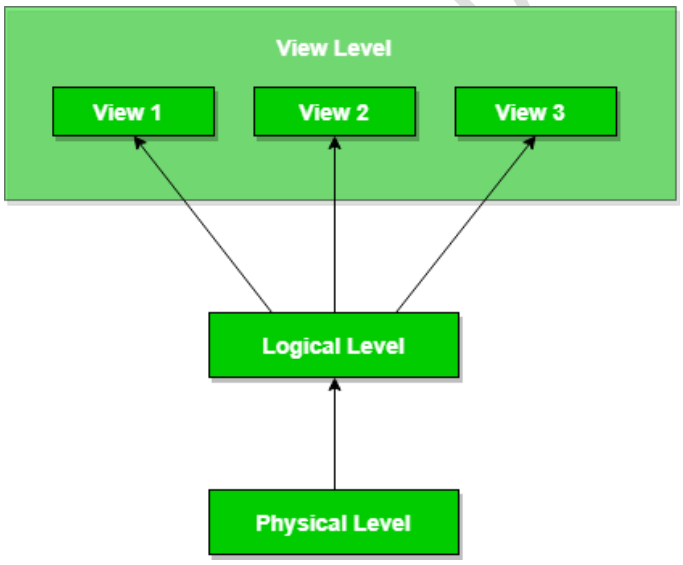
1. **Internal level(Physical level)**
   It show only how the data stored. It described physical structures of database.
2. **Conceptual level**

   What the data stored
   describe structure of whole Dbase with relations

1. **View Level : (External level)**

   Highest level in Dbase abstraction

   Show some part of Dbase & hide remaining.



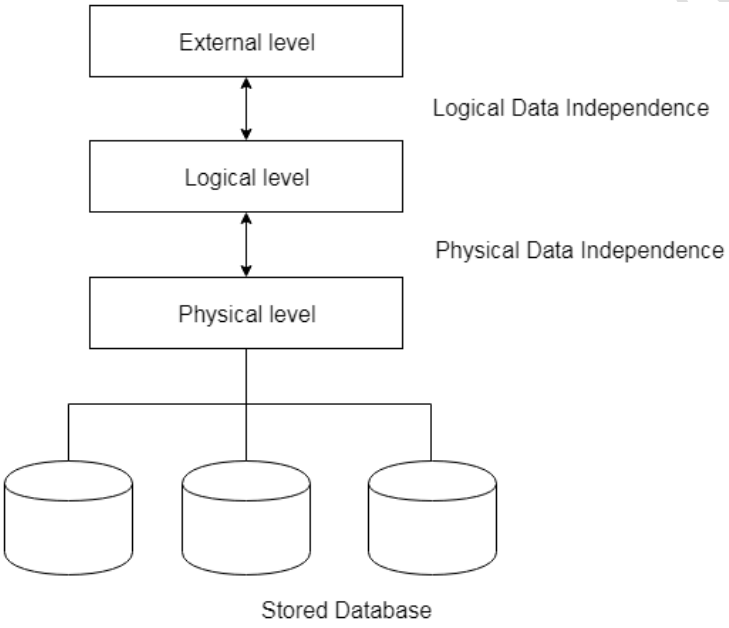
Data independency:

Data independency is the capability to change one level data without affecting its higher level data.

Two types of data independency:

1. **Conceptual data independency:**
2. **Physical data independency:**

1. **Conceptual data independency:**
   It is also called as logical data independency. In this independency, change occurred at conceptual level without affecting next external level data.
2. **Physical data independency:**
   In this independency, change occurred at physical level without affecting next logical level data.



Data model:

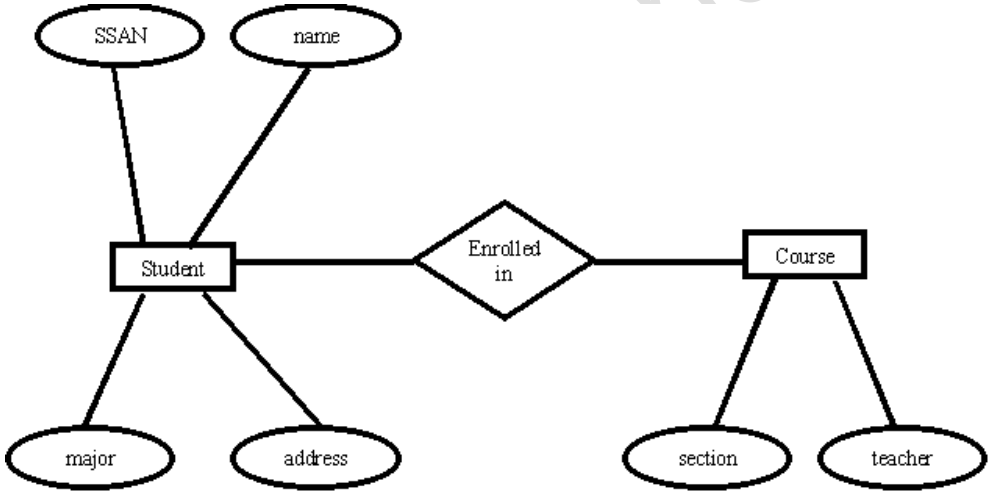
Data model are collection of concept used to describe structure of a database & maintain levels of data abstraction. Using data model retriving and storing methods are defined.

Three types of data model:

1. **Object based logical model**
   a. Entity relationship model
   b. Object oriented model
   c. Symantic data model
   d. Functional data model

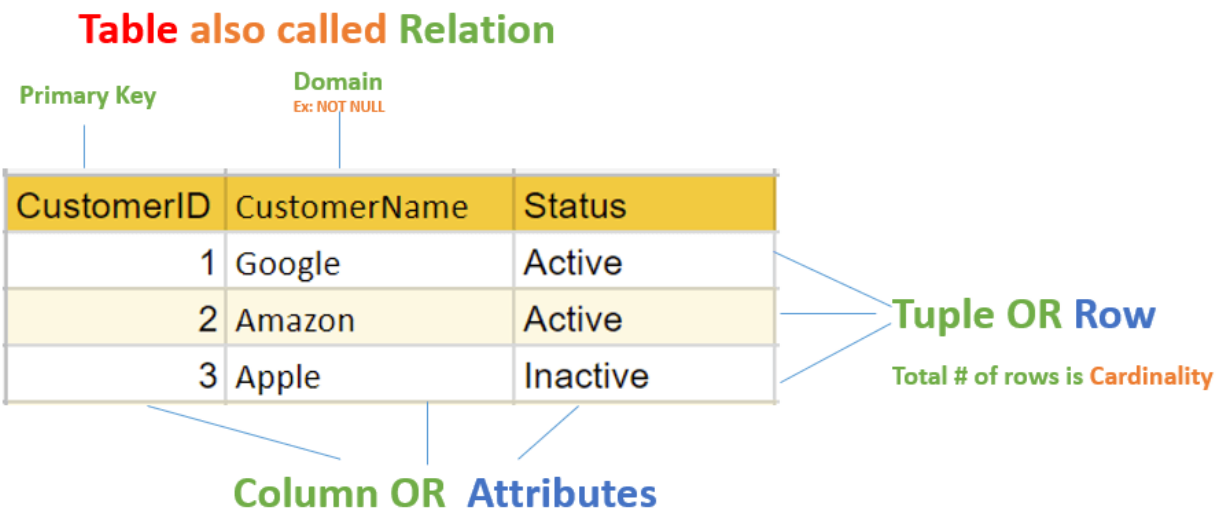
2. **Record based logical model**
    a. Relational data model
    b. Network data model
    c. Hierarchical data model
3. **Physical data model**

1. **Object based logical model**
    a. **Entity relationship model**
        i. This data model consist of entity & relation present in database
        ii. Entity is nothing but real world object which having some properties to identify it.
        iii. Properties of as entity is called attributes.
        iv. Logical association between entities is called relation between them.



    b. **Object oriented model**
        i. This data model is depends upon collection of objects.
        ii. Mainly used in object oriented programming languages
    c. **Symantic data model**
        i. This data model depends upon symantic(syntax) of an objects.
    d. **Functional data model**
        i. This data model depends upon function which are defined for performing specific tasks.
        ii. This data model mainly used for saves, retrieve, and update etc. type of operations.
2. **Record based logical model**

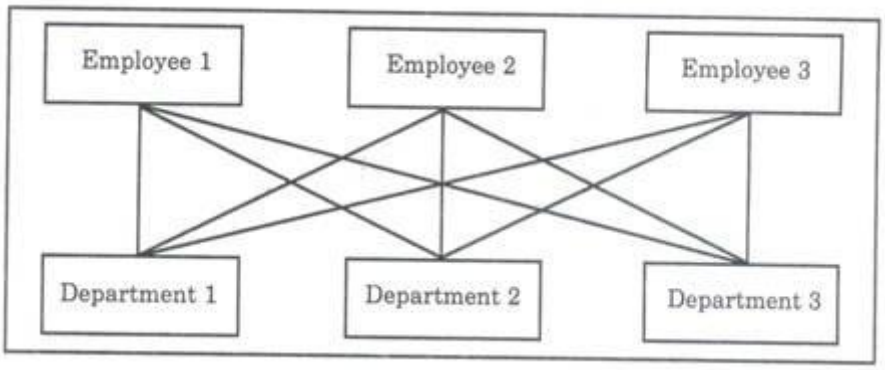
    This data model is mainly used at logical and view level of data abstraction.
    a. **Relational data model**
        i. In this data model data are stored in the form of tables.
        ii. Association between entities(relation) is specify using tuple(records) of an relation.



b. Network data model

i. This data model is known as flexible data model for representing data and relation between them.

ii. In network data model, schema is represented as a graph in which objects are defined using nodes & relation using arc.

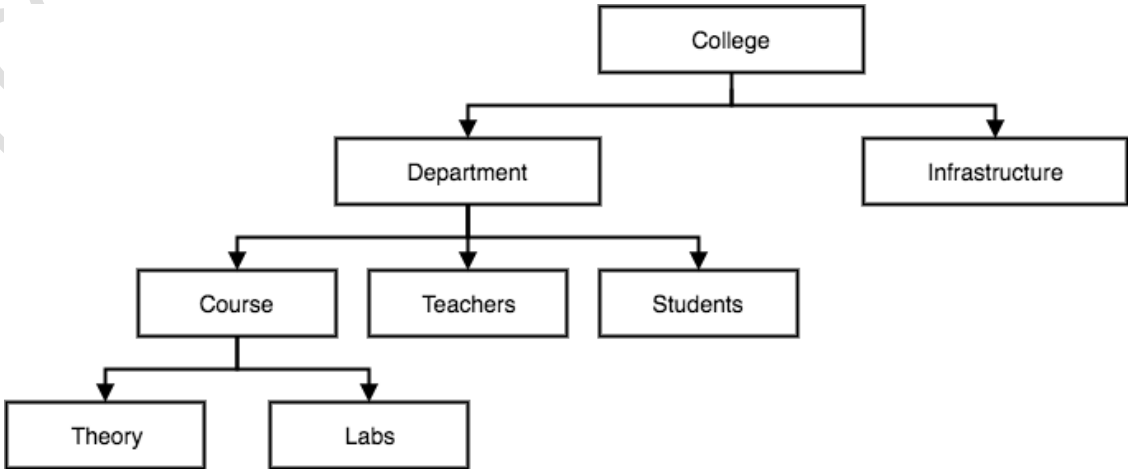


c. Hierarchical data model

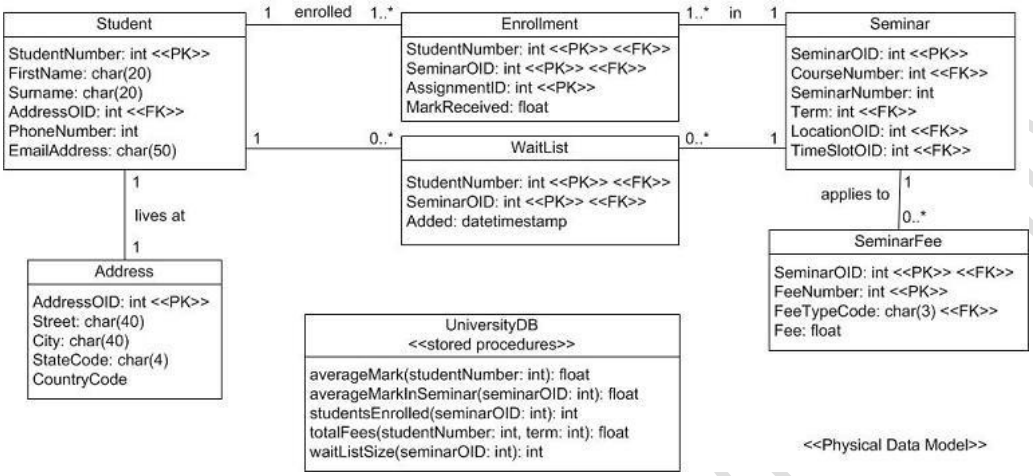
i. In this data model data is represented in tree-like structure.

ii. Data are stored as records, connected to each other using links.

iii. It is also called as tree design data model, which having root node, parent nodes and their child nodes.



3. **Physical data model**
    i. Physical data model consist of all database artifacts required to create relationship between tables.
    ii. It is used to achieve performance goal like indexes, constraint definitions, linking tables etc.



Components of DBMS:

1. **DBMS Languages:**
    a. Data Defination language(DDL)
    b. Data Manipulation Language(DML)
    c. Data Control Language(DCL)
    d. Transaction Control Language(TCL)
2. **DBMS interfaces**
    a. Manu based interface
    b. Form based interface
    c. Graphical user interface
    d. Natural language interface
3. **DBMS users**
    a. Naïve users
    b. Casual users
    c. Application programmer
    d. End user
    e. Database Adminstrator(DBA)

1. **DBMS Languages:**
    a. **Data Defination language(DDL)**
        i. DDL is used define conceptual schema & also define how to implement this.
        ii. DDL Compiler is used to process DDL commands.
        iii. For this processing DDL compiler uses Data dictionary or system catalog.
        iv. DDL is also called as Storage Definition Language(SDL).

b. Data Manipulation Language(DML)

i. DML is used to manipulate data which are defined using DDL.
ii. Two types of DML commands:

Procedural DML	Non-Procedural DML
It is also called as low level DML	It is also called as high level DML
This is embedded in general purpose programming language.	This DML uses on its own to specify processing of operations.
In this user must specify what kind of data needed for processing and how to access it.	In this user must specify what kind of data needed for processing.
For eg: Relational algebra	For eg: postgersql

c. Data Control Language(DCL)

i. DCL used to conrol access to data in database.
ii. DCL is subset of SQL.
iii. DCL gives access control to tables, indexes, views and other elements of database.
iv. For eg: CALL, RETURN, SET etc..

d. Transaction Control Language(TCL)

i. TCL deals with the transaction(any operation in database) control.
ii. Examples of TCL are:
    1. Commit: successful completion of transaction.
    2. Rollback : rollbacks transaction in case of any error occurs.
    3. Savepoint : sets a temporary save point till error free execution.

2. DBMS interfaces

DBMS interface is a user interface which allows user interface to the database to input quiry without using query language.

a. Manu based interface:

i. In this interface, list of options called menus are present.
ii. Using menus user feel free because of there is no need to remember any syntax of query.

b. Form based interface:

i. This interface displays a form to each other.
ii. User insert new record in database by filling and submitting such forms.

c. Graphical user interface:

i. This interface uses diagrammatical representation of option.
ii. GUI utilizes both menu based and form based interface.

d. Natural language interface:

i. This interface accept request written in English or any other language and attempt to understand them.

3. **DBMS users**

Differentiation between users of database is done according to the interaction to database.

a. Naïve users

i. Naïve users ,need not to be aware of presence of database.
ii. Also called as parametric end userd/
iii. For eg: Clerk of bank office.

b. Casual users:

i. These users having great knowledge about query processing.
ii. They do not write program but they fire query on database.

c. Application programmer

i. These users writes program that uses the database.
ii. Accessing data, creating new data, modifying data is done by using these application programs.

d. End user

i. End user access data from the terminal end.
ii. They use developed application and they don't have any knowledge about the design and working of database.

e. Database Adminstrator(DBA)

i. It can be single person or group, responsible for everything that is related to database.
ii. DBA makes policies, strategies and provides technical supports.

DBMS Structure :

A database system is partitioned into modules that deal with each of the responsibilities of the overall system. The functional components of a database system can be broadly divided into the **storage manager** and the **query processor** components.

Query Processor:

The query processor components include

· **DDL interpreter,** which interprets DDL statements and records the definitions in the data dictionary.

· **DML compiler,** which translates DML statements in a query language into an evaluation plan consisting of low-level instructions that the query evaluation engine understands.

A query can usually be translated into any of a number of alternative evaluation plans that all give the same result. The DML compiler also performs **query optimization**, that is, it picks the lowest cost evaluation plan from among the alternatives.

Storage Manager

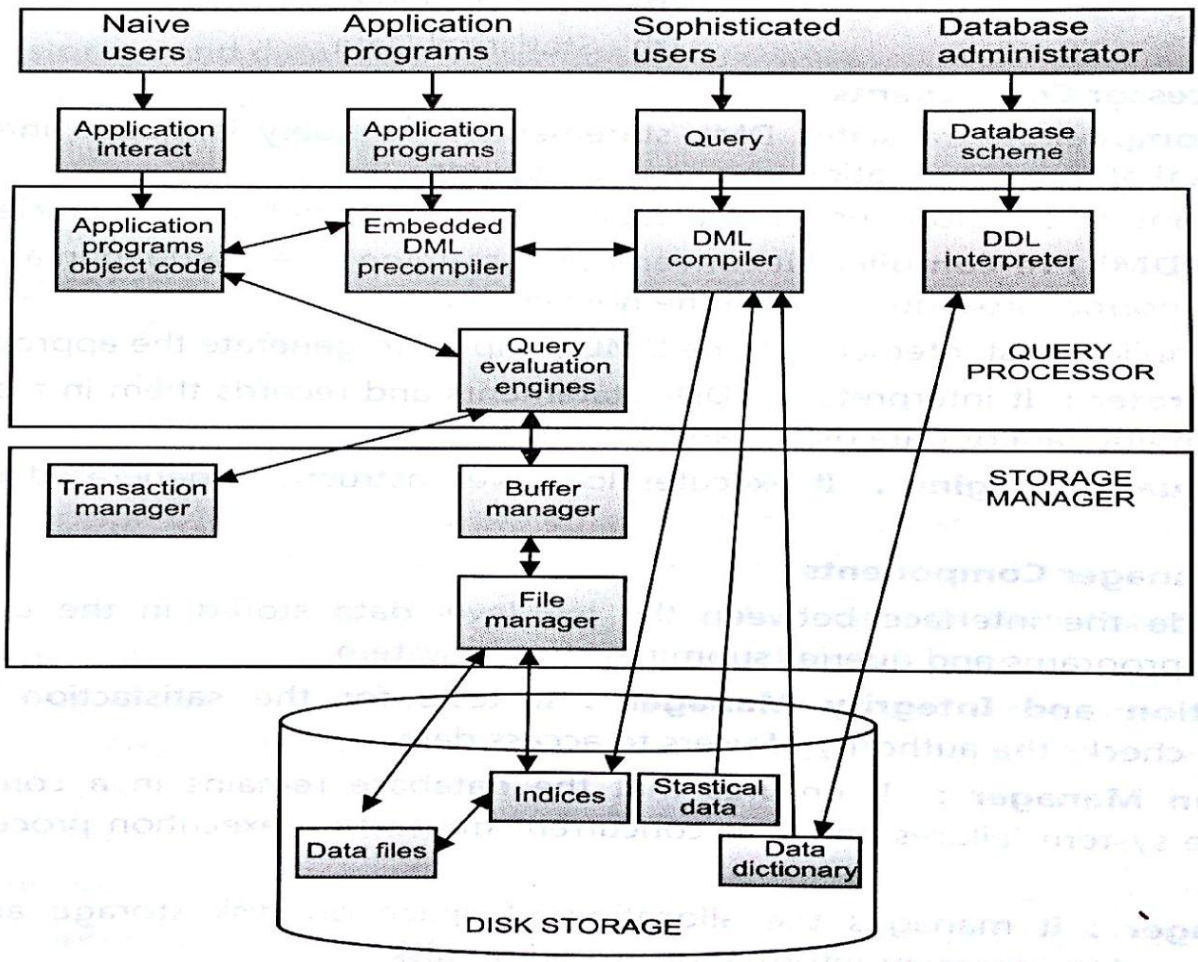
A *storage manager* is a program module that provides the interface between the low level data stored in the database and the application programs and queries submitted to the system. The storage manager is responsible for the interaction with the file manager. The raw data are stored on the disk using the file system, which is usually provided by a conventional operating system. The storage manager translates the various DML statements into low-level file-system commands. Thus, the storage manager is responsible for storing, retrieving, and updating data in the database.

The storage manager components include:

- **Authorization and integrity manager**, which tests for the satisfaction of integrity constraints and checks the authority of users to access data.
- **Transaction manager**, which ensures that the database remains in a consistent (correct) state despite system failures, and that concurrent transaction executions proceed without conflicting.
- **File manager**, which manages the allocation of space on disk storage and the data structures used to represent information stored on disk.
- **Buffer manager**, is a critical part of the database system, since it enables the database to handle data sizes that are much larger than the size of main memory.

Transaction Manager

A **transaction** is a collection of operations that performs a single logical function in a database application. **Transaction - manager** ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures.


Naive users
Application programs
Sophisticated users
Database administrator
Application interact
Application programs
Query
Database scheme
Application programs object code
Embedded DML precompiler
DML compiler
DDL interpreter
Query evaluation engines
QUERY PROCESSOR
Transaction manager
Buffer manager
STORAGE MANAGER
File manager
Indices
Stastical data
Data files
Data dictionary
DISK STORAGE

Assignment –1

Issued date :- 26/07/19 ***Complition date :- 02/08/19***

Q. 1 Answer the following. [5M]

1. Explain types of files?
2. What are the types of records?
3. “No one file organization is uniformaly superior in all situations”, Comment.
4. What are the levels of data abstraction?
5. Explain in brief data independency.
6. Explain the object based logical data model.
7. Explain record based logical data model & physical data model.
8. Explain DBMS language with type?
9. Explain types of DBMS interfaces?
10. Which are the types of DBMS users?
11. Explain DBMS structure in brief.
12. Sorted file is best if range selection is desired-Comment.

Q. 2 Write differentiate point between:

1. Physical file & logical file
2. Procedural DML & non-procedural DML.
3. File organization & DBMS.

Q.3 Answer in short. [2M]

1. What is data?
2. What is database?
3. List the record based Logical Models
4. Which are two types of DMLs? Give examples.