Smt. R. S. Kale 1 Set

SPPU New Syllabus

A Book Of

DATABASE MANAGEMENT SYSTEMS

For F.Y.B.Sc. Computer Science : Semester – I (Paper – II)

[Course Code CS102 : Credits - 2]

CBCS Pattern

As Per New Syllabus, Effective from June 2019

Dr. Ms. Manisha Bharambe
M.Sc. (Comp. Sci.), M.Phil. Ph.D. (Comp. Sci.)
Associate Professor,
Department of Computer Science,
MES's Abasaheb Garware College,
Pune 4.

Dr. A. B. Nimbalkar
M.C.S., M.Phil. (C.S.), Ph.D.(Comp. Sci.), D.C.L
Assistant Professor,
Department of Computer Science,
PDEA's A.M. College, Hadapsar,
Pune 28.

Price ₹ 180.00

NIRALI PRAKASHAN
ADVANCEMENT OF KNOWLEDGE

N5047

Syllabus ...

1. Introduction to DBMS **(3 Hrs.)**

1.1 Introduction to File Organization and DBMS
1.2 File System Vs DBMS
1.3 Levels of Abstraction and Data Independence
1.4 Structure of DBMS (Roles of DBMS Users)
1.5 Users of DBMS, Advantages of DBMS

2. Conceptual Design **(11 Hrs.)**

2.1 Overview of DB Design Process
2.2 Introduction to Data Models (E-R Model, Relational Model, Network Model, Hierarchical Model)
2.3 Conceptual Design Using ER Data Model (Entities, Attributes, Entity Sets, Relations, Relationship Sets)
2.4 Constraints (Key Constraints, Integrity Constraints, Referential Integrity, Unique Constraint, Null/Not Null Constraint, Domain, Check Constraint, Mapping Constraints)
2.5 Extended Features – Specialization, Aggregation, Generalization
2.6 Pictorial Representation of ER (Symbols)
2.7 Structure of Relational Databases (Concepts of a Table)
2.8 DBMS Versus RDBMS
2.9 Case Studies on ER Model

3. SQL **(9 Hrs.)**

3.1 Introduction to Query Languages
3.2 Basic Structure
3.3 DDL Commands
3.4 DML Commands
3.5 Forms of a Basic SQL Query (Expression and Strings in SQL)
3.6 Set Operations
3.7 Aggregate Operators and Functions
3.8 Date and String functions
3.9 Null Values
3.10 Nested Subqueries
3.11 SQL Mechanisms for Joining Relations (Inner Joins, Outer Joins and Their Types)
3.12 Views
3.13 Case Studies on SQL

4. Relational Database Design **(7 Hrs.)**

4.1 Introduction to Relational-Database Design (Undesirable Properties of a RDB Design)
4.2 Functional Dependency (Basic Concepts, F+, Closure of an Attribute Set, Armstrong's Axioms)
4.3 Concept of Decomposition
4.4 Desirable Properties of Decomposition (Lossless Join, Lossy Join, Dependency Preservation)
4.5 Concept of Normalization, Normal Forms (1NF, 2NF, 3NF, BCNF) Examples
4.6 Keys Concept with Examples: Super Keys, Candidate Keys, Composite Key
4.7 Algorithm to find Candidate Keys / Primary Key for a Relation

CHAPTER 1

Introduction to DBMS

1.0 INTRODUCTION

- With the rapid growth in computing technology and its application in all spheres of modern society, databases have become an integral component of our everyday life.
- We encounter several activities in our day-to-day life that involve interaction with a database, for example, bank database to withdraw and deposit money, movies database for booking of tickets, bus, air or railway reservation databases for booking of tickets, supermarket goods database to keep the inventory etc.
- In fact, databases and database management systems have become essential for managing for business, governments, banks, universities and every other kind of human endeavor. Thus, they are a critical element of today's software industry to support these requirements and a daunting task to solve the problems of managing huge amounts of data that are increasingly being stored.
- In today's modern (Electronic and/or Digital) environment, data (or information) and its efficient management is the most critical business objective/goal of an organization. Data is a collection of facts and figures that can be processed to produce information. Organizing data, storing data and retrieving information has been a necessity in all ages of an organization.
- Data can be defined as, "a set of isolated and unrelated raw facts with an implicit meaning". Data can be anything such as, text, number, images, sound, video etc. For example, 'Amar', '18' 'student' etc. is data.
- When the data is processed and converted into a meaningful and useful form, it is known as information. For example, 'Amar is 18 years old and he is a student', is information. Data and information are closely related and are often used interchangeably.
- A database system simplifies the tasks of managing the data and extracting useful information in a timely fashion. A database system is an integrated collection of related files, along with the details of the interpretation of the data.
- A DataBase Management System (DBMS) is a software or program that allows access/retrieval of data that contained in database. The objective of the DBMS is to provide a convenient and effective method of defining, storing and retrieving the information stored in the database.
- The database and the DBMS software are collectively known as database system.

1.1 INTRODUCTION TO FILE ORGANIZATION AND DBMS

- In Information Systems (ISs) we deal with data. This data has to be arranged in a proper way to accept process and communicate operations and results. For arranging the data, we need files. A file is a resource for storing information in a computer system.
- A DBMS (Database Management System) is a complex software system that is used to manage, store and manipulate data. It is utilized by a large variety of users, to retrieve and manipulate data under its control.
- A DBMS is an integrated set of programs used to create and maintain a database. In this section we will study basic concepts of file, file organization and DBMS.

1.1.1 Basic Concepts of Files

- A file is a named collection of related information.
- In general, a file is a sequence of bits, bytes, lines or records whose meaning is defined by the files creator and/or user.
- A manual file stores all the information relating to a particular activity. For example: Inventory activities in an inventory file, payroll activities in a payroll file and so on.
- The basic unit of information for computer and manual files is a record. The collection of data items is called a field. For example, Student_Name, Address, Phone_No etc., are known as fields of STUDENT record.
- The collection of related field is called a record. For example, Student_ Name, Address, Phone_No etc. may make a student record. A set of logically related records form or constitute a file.
- Some basic operations such as Create, Open, Locate (Find), Read (Get), Write, Close etc., on the records available in the file to modify or access the file. (April 15)
- The structure and logic rules used to manage the groups of information and their name is called as a file system. A file system is that set of system software that provides services to users and applications in the use of files.
- A file is defined as, "a collection of related sequence of records". There are two common ways of viewing files:
  1. **Logical File:** A logical file is a file viewed in terms of what data items its record contains and what processing operations may be performed on the file. The user of the file will normally adopt such view. (April 16, Oct. 16)
  2. **Physical File:** A physical file is a file viewed in terms of how the data is stored on a storage device and how the processing operations are made possible. (Oct. 17)

1.1.2 File Organization (April 15, 16)

- A file organization is a method of arranging the records in a file. The file is stored on secondary storage device called hard disk. The physical arrangement of data in a file into records on the disk is known as file organization.
- File organization is defined as, "a way of arranging the records in a file when the file is stored on secondary storage like disk, tape etc."
- File access mechanism refers to the manner in which the records of a file may be accessed. A file can be accessed or modified in different ways or methods such as Sequential Access, Direct/Random Access, Indexed sequential Access and so on.
- Some common file organization methods/techniques are explained below:

1. **Heap File Organization:** (April 15, 16)

- It is the simplest type of file organization. A heap file is an unordered set of records. A heap file is also called a pile or serial file or unsorted file.
- A heap file has no particular order. Therefore, it is equivalent to file of unordered records or sequence.Heap file organization is also called as serial file organization of pile file organization.
- In heap file organization, the records are stored in the file, in the order in which they are inserted. All the new records are stored at the end of the file.

Advantages of Heap File Organization:

(i) This is a simple and easy file organization method.

(ii) Insertion of new record is fast and efficient.

Disadvantages of Heap File Organization:

(i) Deletion of many records results in wastage of space.

(ii) Searching and accessing of records is very slow.

(iii) Updation cost of data is comparatively high.

2. **Sorted File Organization:** (April 15, 16)

- We can physically order the records of a file on disk based on the values of one of their fields – called the ordering field. This leads to an ordered or sequential file.
- A sequential file maintains the records in the logical sequence. Every file record contains a data field (attribute) to uniquely identify that record.
- In sorted file organization, records are stored in a sequential order according to the "ordering key". A ordering key is an attribute or a set of attributes which are used to serialize the records.
- In a sorted file, first the record is inserted at the end of the file and then moved to its correct location, (ascending or descending). Records are stored in order of the values of the key field.

Advantages of Sorted File Organization:

(i) The sequential file can be created on any device with ease and simple.

(ii) The input, output media and devices used in this type of organization are inexpensive.

(iii) The average processing speed of this file organization is high.

Disadvantages of Sorted File Organization:

(i) The entire file must be processed to read only one record.

(ii) In this file system, the data redundancy is usually high.

(iii) In this file organization linking of the records of two files is not possible.

3. **Indexed File Organization:** (April 15, 16, Oct. 16)

- In the index file organization determining the storage location is a separate operation from accessing the record.
- In an indexed file organization, the rows are stored either sequentially or non-sequentially, and an index is created that allows the application software to locate individual rows.
- An index is structure that is used to determine the rows in a file that satisfy some conditions. Each entry matches a key (primary or secondary) value with one or more rows. An index can point to unique rows or to potentially more than one row.
- A sequential file that is indexed is called as index sequential file. A common indexed file organization is the indexed sequential organization often called ISAM (Indexed Sequential Access Method). (Oct. 16)

Advantages of Indexed File Organization:

(i) In this file organization there is no need to sort the transactions before updating.

(ii) In indexed organization, it is possible to link the records of two files.

(iii) Comparing with other file organization, here it is easy to update the files.

(iv) Indexed file organization is flexible in report of sequential and of random/direct processing.

Disadvantages of Indexed File Organization:

(i) Additional space is to be provide for indexes, this file Organization requires more space than sequential file.

(ii) Files in this file organization must be stored on a direct access storage device, hence, it is expensive.

4. **Hash File Organization:** (April 15, 16, 17)

- A random file (also called a hash file, direct file or relative file) has records that are stored and retrieved according to either their disk address or their relative position within the file. This means that the program which stores and retrieves the records has first to specify the address of the record in the file.

A hashed file uses a hash function applied to a particular field (hash key) to determine a record's placement on disk.

This organization also known as random access (without any specific order or sequence) files organization or relative file organization.

In this organization, the user can access any record which is situated anywhere in the file, without reading or passing through other records in the file.

In hash file Organization, mapping (also called transformation) of the search key value of a record is made directly to the address of the storage location at which that record is to reside in the file. One mechanism used for doing mapping is called hashing.

There are two types of hashing namely **Internal Hashing** (for internal files, hashing is typically implemented through the use of an array of records) and **External Hashing** (hashing for disk files and the address space of disk is divided into buckets, each of which holds multiple records).

dvantages of Hashed File Organization:

(i) The records can be accessed directly in this file organization; the file accessing and updating becomes very easy, simple and fast.

(ii) Random file organization suitable for access of on-line applications.

(iii) In random file organization there is no need to sort the input before updating.

Disadvantages of Hashed File Organization :

(i) The processing speed of random file system is low, which ranges between the sequential and indexed sequential file systems.

(ii) In random file organization, the expensive resources are used. Hence it is relatively expensive when compared withother file organization.

(iii) Random files can be opened only on direct access devices. Thus, it is not suitable for all storage media.

Comparison of Sequential (Sorted), Indexed and Hashed File Organizations:

Sr. No.	Sequential (Sorted) Organization	Indexed File Organization	Hashed/Direct Organization
1.	Random retrieval on primary key is impractical.	Random retrieval of primary key is moderately fast.	Random retrieval of primary key is very fast.
2.	Multiple key retrieval in sequential file organization is possible.	Multiple key retrieval is vary fast with multiple indexes.	Multiple key retrieval is not possible.
3.	Addition of new records requires rewriting the file.	Addition of new records is easy and requires maintenance of indexes.	Addition of new records is very easy.
4.	Deletion of records can create wasted space.	Deletion of records is easy, if space can be allocated dynamically.	Deletion of records is very easy.
5.	There is no wasted space for data.	No wasted space for data but there is extra space for index.	Extra space for addition and deletion of records.
6.	Sequential retrieval on primary is very fast.	Sequential retrieval on primary key is moderately fast.	Sequential retrieval of primary key is impractical.
7.	Updating of records generally requires rewriting the file.	Updating of records requires maintenance of indexes.	Updating of records is the easiest one.

1.1.3 Introduction to Database

- The database is used to store information useful to an organization/an enterprise. A database is a shared collection of inter-related data, which is designed to fulfill the information needs.
- In any organization, data is the basic resource needed to run the organization. This data is required by decision makers for processing and retrieving information.
- A database is a collection of information that is organized so that it can easily be accessed, managed, and updated.
- In simple word, a database is a collection of data. For example: Consider the names, telephone numbers and addresses of the people you know. This data is recorded in an indexed address book or stored on a diskette or using a personal computer. This is a collection of related data with an implicit meaning and hence is a database.

Properties of Database:

1. A database represents some aspect of the real worldsometimes called the mini-world. Changes to the real world are reflected in the database.
2. A database is a logically collection of data with some meaning.
3. A database is designed, built and populated with data for a specific purpose. It has an intended group of users.

- So, finally we can say that a database has some source from which data are derived, some degree of interaction with events in the real world, and an audience that is actively interested in the contents of the database.

1.1.3.1 Definition of Database

- A database is, "an organized collection of data from which users can efficiently retrieve the desired information". OR
- A database can be defined as, "a collection of interrelated data". OR
- A database is defined as, "a collection of logically related data stored together that is designed to meet information requirements of an organization".

1.1.3.2 Components of Database

- A database consists of four components as shown in Fig. 1.1.
- Fig. 1.1 shows following components of a database:

1. **Data Item** is a distinct piece of information and is explained in the previous section.
2. **Relationships** represents a correspondence between various data elements.
3. **Constraints** are the predicates that define correct database states.
4. **Schema** describes the organization of data and relationships within the database.

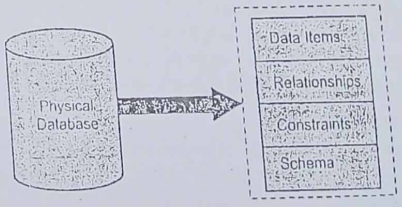


Fig. 1.1: Components of Database

1.1.4 Introduction to DBMS (April 17)

- A Database Management System (DBMS) is a software system that allows user to define, manipulate and process the data in a database, in order to produce meaningful information.
- DBMS is defined as, "the software system that allows users to define, create, maintain and control access to the database". OR

- A DBMS is, "a software system that allows user to define, manipulate and process the data in a database, in order to produce meaningful information".
- The basic functions of DBMS are:
  1. To store data in a database.
  2. To organize the data.
  3. To control access of data.
  4. To protect data i.e., provide security.
- The DBMS is hence a general purpose software system that facilitates the processes of defining, constructing and manipulating databases for various applications.
  - **Defining** a database involves specifying the data types, structures and constraints for the data to be stored in the database.
  - **Constructing** the database is the process of storing the data itself on some storage medium that is controlled by the DBMS.
  - **Manipulating** a database includes such functions as querying the database to retrieve specific data, updating the database to reflect changes in the mini-world, and generating reports from the data.
- So in general, user can write programs or queries; DBMS use the database stored on storage devices and gives meaningful information.
- The database system is illustrated in the Fig. 1.2.
- The main objective of a DBMS is to provide a convenient and effective method of defining, storing, retrieving and manipulating the data contained in the database.
- There are many DBMS like MySQL, PostgreSQL, Microsoft Access, SQL Server, FileMaker, Oracle, RDBMS, dBASE, Clipper, FoxPro and so on.
- DBMS acts as an interface between the application program and the data stored in the database, (Refer Fig. 1.3).

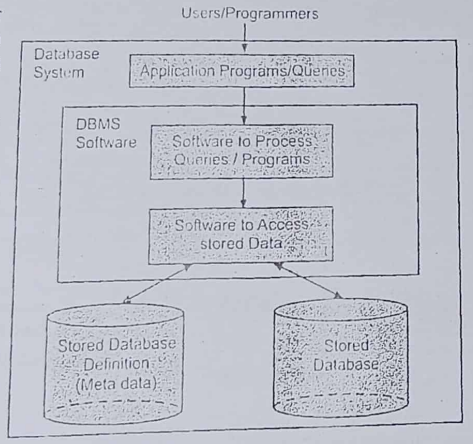


Fig. 1.2: Simple Database System

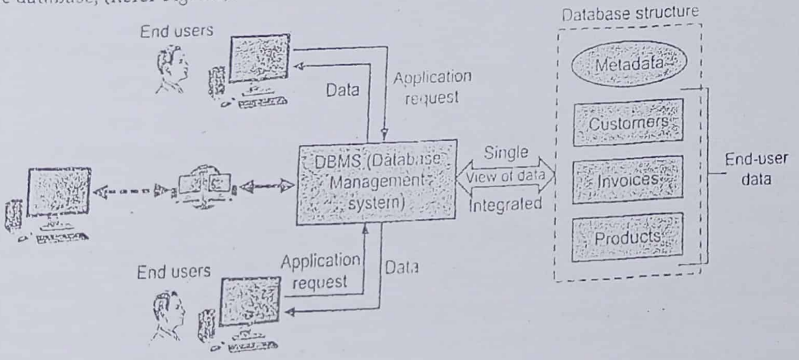


Fig. 1.3

1.2 FILE SYSTEM VS DBMS (Oct.-17)

- In this section we study concepts of file processing system and DBMS.

1.2.1 File System (April-15)

- In file processing system (Refer Fig. 1.4) the records are stored in separate files. Each file is called a flat file. To access the data from these flat files, various programs are written.
- So, that the system provides fruitful information to the end user. Actually speaking the work is tedious. Because if user wants simple change in the resultant database, lot of changes are needed in the application programs.

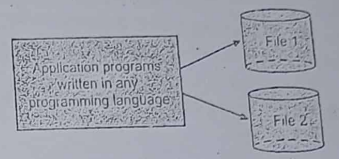


Fig. 1.4: File Processing System

- Fig. 1.5 shows an example of traditional file processing system of an organization. All functional areas in the organization creates, processes and disseminates its own files.
- The files such as Sales department and Accounting department etc. generate separate files and do not communicate with each other.

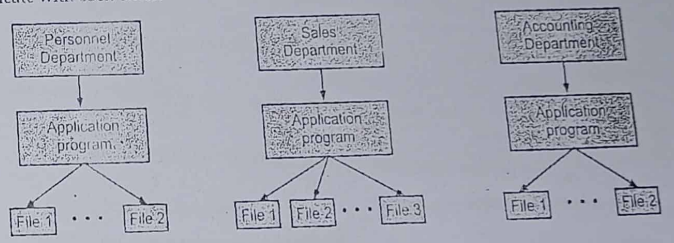


Fig. 1.5: Example of Traditional File System

- As system becomes more complex, file processing system lost its flexibility and shows many limitations or disadvantages. Some of them are listed below: (April-15)

1. Separated Data:
   - To make certain decision, a user might need data from different/separate files. Because of the involvement of system analyst and programmer, files may have different formats.
   - For example, salary of an employee stored as integer in one file and real in other file. Therefore, files were evaluated first and then relationships were determined. After that programs could be written to extract data from files.
2. Isolated Data:
   - Data are scattered in various files, and the files may be in different format, writing new application program to retrieve data is difficult in file system.
3. Data Redundancy:
   - Suppose same information was stored in more than one file. This repetition of data caused loss of data integrity, (means accurate and consistent data).

- For example, Address of an employee was stored in three different files. If change of address occurs and has been updated in one file only, then mismatch of information gives inconsistency to data.

4. **Difficulty in Data Access:**

- As it has been said, files and records were described by specific physical formats by analysts and programmers, it is difficult to access data.
- If the format of a certain record was changed, the code must be immediately updated, system provides incorrect data.

5. **Concurrent Access Anomalies:**

- In file processing system, once a file is opened by a user, it cannot be used by another user, till first closes a file. It means sharing a file by various user at the same time is not possible.

6. **Security Problem:**

- Enforcing security constraints in file processing system is very difficult as the application programs are added to the system in an ad-hoc manner.

7. **Atomicity Problem:**

- It is difficult to ensure atomicity in file processing system.

1.2.2 Database System (DBMS Environment)

- The DBMS software together with the database is called a database system. In other words, database system can be defined as, "an organization of components that define and regulate the collection, storage, management and use of data in a database".
- A database system is a system whose overall purpose is to record and maintain information. It simple terms, Database + DBMS Software = Database System.
- A database system consists of four major components as shown in Fig. 1.6 i.e., Data, Hardware, Software and Users.
- A data file is a single disk file that stores related information on a hard disk.

Components of database system:

1. **Data:**

- A data is collection of information or real fact which can be recorded and have implicit meaning. The whole data in the database system is stored in a single database and this data in the database are both shared, (sharing of data means individual pieces of data in the database is shared among different users and every user can access the same piece of data but may be for different purposes) and integrated, (integration of data means the database can be function of several distinct files with redundancy controlled among the files).

2. **Hardware:**

- The hardware consists of the secondary storage devices like disk, where the database resides together with other devices.

3. **Software:**

- It is a layer or interface of software exists between the physical database and the users. This layer is called the DBMS.
- All requests from the users to access the database are handled by the DBMS.

4. **Users:**

- The users are the people interacting with the database system in any way.
- The four types of users interacting with the database are Application Programmers, Online users, End users (naive users) and Database Administrator (DBA).

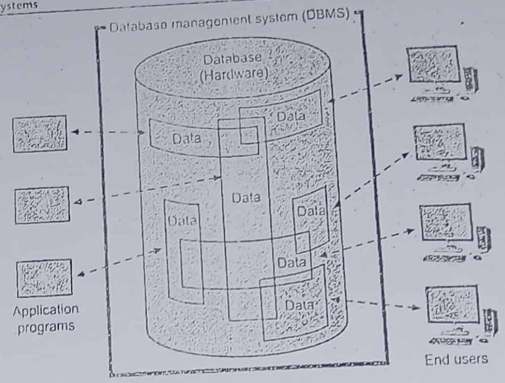


Fig. 1.6: Components of Database System

- Fig. 1.7 shows database system of an organization.

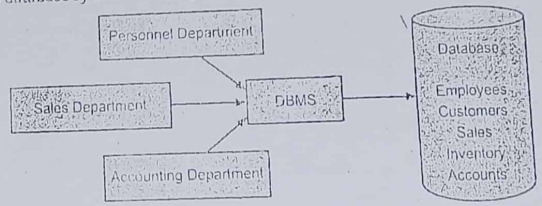


Fig. 1.7

1.2.2.1 Characteristics of DBMS

- A DBMS is a piece of software that is designed to make all above tasks easier. By storing a data in DBMS, rather than in files, we can use the DBMS features to manage the data in a robust and efficient manner.
- Characteristics of DBMS are given below:

1. **Users File Approach:**

- In file processing, each user defines and implements the files needed for a specific application, therefore more storage space is required.
- In DBMS, a single database is maintained that is defined once and then is accessed by various users. DBMS software is not written for specific application.

2. **Self Describing Nature of DBMS System:**

- A DBMS contains database as well as a complete definition of the database, which is stored in the system catalog. It contains information such as structure of each file, the type and storage format of each data item and various constraints of the data.
- The catalog is used by DBMS software by database users who need information about database. For example, a company databases, a banking database, a university database. The database definition is stored in the catalog.

- In file processing, data definition is part of the application programs. These programs works with only one specific database unlike DBMS can access diverse databases by extracting the database definitions from the catalog.

3. *Isolation between Programs and Data:*

- DBMS access programs are written independently of any specific files. The structure of data files is stored in catalog separately from the access program. (program-data independence).
- In file processing, if any change in structure of file is made then, all programs that access this file have to be changed, suppose we want to add any field in record of file (say birth-date in student file), then program has to be changed in file processing.

4. *Multiple Views of Data:*

- DBMS supports multiple viewsof the data, which a file cannot support.
- For example: one user is only interested for student marklist, other user is interested for courses attended by that student, these multi-user view are satisfied by DBMS.

1.2.3 Comparison between File System and DBMS

- Following table comparers File System and DBMS:

Sr. No.	File Processing System	Database Management System
1.	PC based small systems.	Mini mainframe based large systems.
2.	Relatively cheap.	Relatively expensive.
3.	Less number of files used.	More number of files used.
4.	Single user system.	Multiple user system.
5.	Data redundancy and inconsistency occur.	Data is independent and non-redundant.
6.	Data access is difficult.	Data access is efficient.
7.	Data is isolated.	Data is integrated.
8.	Concurrent access to a file is not possible.	Concurrent access and crash recovery possible.
9.	Security problems.	Better security.
10.	Little preliminary design.	Vast preliminary design.
11.	Examples: C++, COBOL, VB.	Examples: MS-Access, Oracle, PostgreSQL.
12.	File system data sharing is not powerful as DBMS.	DBMS offers features of data sharing efficiently.
13.	Transaction concept is not used, (transaction means set of related operations).	The concept of transaction is an important aspect of DBMS.
14.	DBMS is a collection of data and set of programs to access those data.	File system is an abstraction to store, retrieve and update a set of files.

1.3 LEVELS OF DATA ABSTRACTION AND DATA INDEPENDENCE

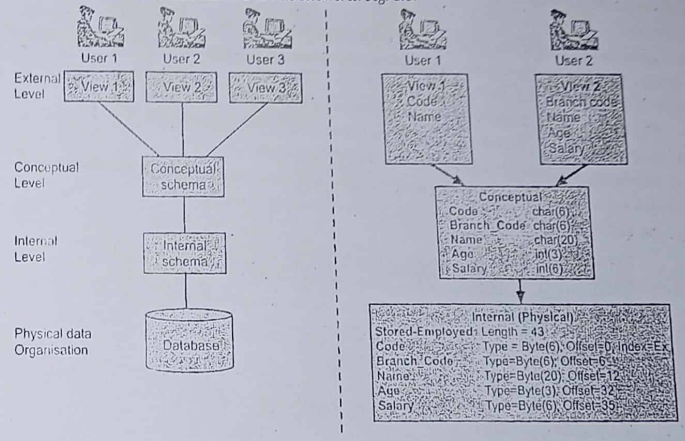
- In this section we will study levels of data abstraction and data independence.

1.3.1 Levels of Data Abstraction

(April 15, 16; Oct. 16, 17)

- A database management system is a collection of interrelated files and a set of programs that allow users to access and modify these files.

- A main purpose of a database management system is to provide users with an abstract view of the data. That is the database system hides certain details of how data are stored and maintained.
- Data abstraction means hiding the implementation details from the end users. DBMS uses the same principle means it hides the complex details of how the data is stored on the secondary storage device in DBMS.
- Data abstraction is refers to, "the hiding of certain details of how the data are stored and maintained". With several levels of abstraction, the user's view of the database is simplified and this leads to the improved understanding of data.
- There are three levels of data abstraction as shown in Fig. 1.8.



(a) Levels of Data Abstraction (b) Example of Data Abstraction

Fig. 1.8

- Fig. 1.8 shows following levels of abstraction:

1. Physical (Internal) Level:

- Internal level is the lowest level of data abstraction.
- Internal level describes in detail how the raw data is actually stored at the byte level using complex low level data structure.
- It also describes the access paths for the database.
- The database system hides many of these lowest level storage and access details from the database programmers and the end users.

2. Logical (Conceptual) Level:

- Logical level is the next higher level of data abstraction.
- Logical level is used by the database administrators to describe what data are stored in the database and the relationship between those data.
- Here, the entire database is described in terms of relatively simple, easy to understand structure like data tables, keys, indices etc.
- It describes the relationship between the different data stored in the database, the operations possible on those data and any constraints or rule to be imposed on the data.

3. **View (External) Level:**

- View level is the highest level of data abstraction.
- View level deals with the way a particular user application program views the data from the database.
- Each view level is used to describe a part of the database that a particular user group is interested in and hides the rest of the database from that user group.

Schemas and Instances in DBMS:

1. **Schemas:**

- The overall description of the database is called as the database schema.
- Actually, DBMS has a three level architecture and accordingly there are three different types of schemas in the database. These are given below:
  - **(i) Physical Schema:** It is at the lowest level. i.e., at Physical level.
  - **(ii) Logical Schema:** It is at the next or intermediate level i.e., at Logical level.
  - **(iii) Sub-Schema:** It is at the highest level i.e., at the View level.
- A schema is defined as, "an outline or a plan that describes the records existing at a particular level".
- The schema diagram is used to show the database schema. For example,

STUDENT_ADDRESS

Roll No.	Name	Address	Place	PIN

STUDENT_MARKS

Roll No.	Subject	Exam Date	Marks

Fig. 1.9: Schema Diagram

- A schema diagram displays only some aspects of a schema like the names of record types and data items and some type of constraints.
- The description of the database is called as the database schema which is specified during database design and is not expected to change frequently. A displayed schema is called as a schema diagram, (Refer Fig. 1.9).

2. **Instances:**

- Databases change over time when the information insert or delete.
- The collection of information stored in the database at a particular moment is called as an instance of the database.
- For example, the instances of STUDENT_ADDRESS table are as shown in following table.

STUDENT_ADDRESS

Roll No.	Name	Address	Place	PIN
01	Amar	A-2/7, JanakPuri	Pune	110012
02	Deepa	B-5, Saket	Pune	110014
03	Kiran	G-8, Rajouri Garden	Pune	110036

Difference between Instance and Schema:

Sr. No.	Instance	Schema
1.	The actual content of the database at a particular time is called database instance.	The overall design of the database is called database schema.
2.	It describes the data structure of the database.	It describes the contents of the database.
3.	Instance state is called extension.	Schema is also called intension.
4.	Instance changes every time the database is updated.	The database schema changes very infrequently.
5.	An instance is analogous to the value of a variable in a program.	An schema is analogous to the value of a variable in a program at any execution point.

1.3.2 Data Independence

(Oct. 16, 17; April 15, 18)

- Data independence means the property to change the overall logical or physical structure of the data without changing the application programs view of data i.e. the views created for users does not have any change.
- Data independence is defined as, "the characteristics of a database system to change the schema at one level without having to change the schema at the next higher level".
- Data independence means that the application is independent of the storage structure and access strategy of data.
- There are two levels of data independence namely, physical data independence and logical data independence.

1.3.2.1 Physical Data Independence

(April 15)

- In physical data independence, the physical layout, organization of data may be changed without changing either the overall logical structure of the data or the application programs.
- For example: We can add extra fields i.e. new fields to the database without disturbing the old records; say in `emp table` `char address [20]` is added.
- Actually in 'C' language or in any programming language if we do so, the program should also be changed. But in DBMS, it is not required. Because of the Data independence facility the change is automatically reflected in secondary storage device.

1.3.2.2 Logical Data Independence

(April 15)

- Inlogical data independence the overall logical structure of data may be changed without changing application program.
- This is hard to achieve because application programs are usually dependent on logical structure of the data. For example: To achieve this, attributes from different tables are considered and on those attributes query is written.
- Here, attribute name, table names are properly selected and written, otherwise result may not appear with the required format.
- Data independence is advantageous in the database environment, since it allows for changes at the level of the database without affecting other levels. These changes are absorbed by the mapping between the levels.
- Fig. 1.10 shows relationship between data independence and three-level DBMS architecture.

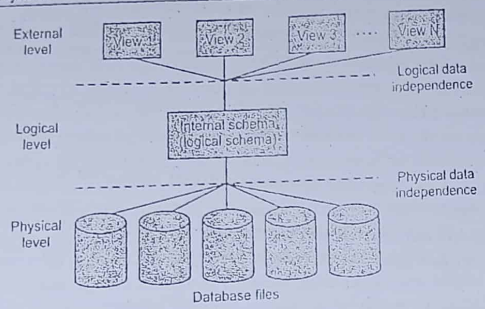


Fig. 1.10: Data Independence

Comparison between Logical and Physical Data Independence:

Sr. No.	Physical Data Independence	Logical Data Independence
1.	It is ability to modify the internal schema of a database without the need to modify the conceptual schema of the database.	It is the ability to modify the conceptual schema of a database without the need to modify the external schema.
2.	Changes in the physical level are usually done to improve the overall performance of a database.	Change in the logical level are usually done to modify the logical organisation of the data.
3.	The conceptual schema insulates the users from changes in physical storage of data.	The users are shielded from changes in the logical structure of the data or changes in the choice of relations to be stored.
4.	It leaves the conceptual (logical) schema unaffected by changes made to the physical schema, which in turn also leaves the external schema unaffected.	It leaves the external (view) schema and methods for accessing the information unaffected by changes made at the conceptual schema.
5.	Examples : Changing the internal data structures or Creating additional access paths for retrieving the same data more efficiently, etc.	Examples : Adding or deleting a field from a record type or Modifying the existing constraints, etc.
6.	Application programs do not depend much on the physical strucutre of the data. Hence it is relatively easier to achieve this.	Application programs are highly dependent on the logical structure of the data, and it is a difficult job to achieve this.

1.4 STRUCTURE OF DBMS

(April 17, Oct. 16, 17)

- The structure of DBMS defines the components are identified, the function of each component is specified and the inter-relationships and interactions among these components are defined.
- A database system is partitioned into modules that deal with each of the responsibilities of the overall system.
- In most cases, the computer's operating system provides only the most basic services and the database system must build on that base. Thus, the design of a database system must include consideration of the interface between the database system and the operation system.

- Fig. 1.11 shows overall structure of DBMS. The basic components of DBMS can be divided into following three subsystems:
  1. **Design Tool:** It allows to create database form and reports.
  2. **Runtime Facilities:** It process the application created by the design tool.
  3. **DBMS Engine:** It works as a translator between design tool and runtime facilities. It is divided into two main parts namely, Query processor components and Storage manager components (Refer Fig. 1.11).

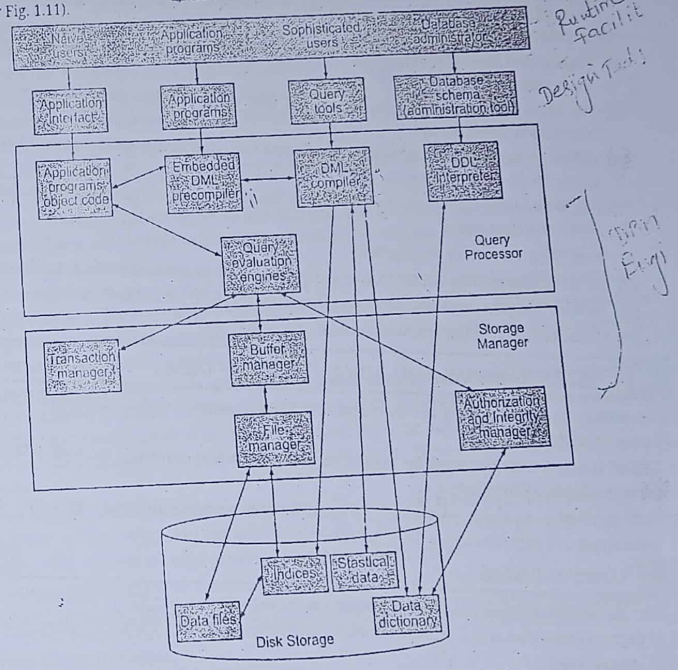


Fig. 1.11: Overall System Structure of DBMS or DBMS Engine

- Lets us see these main parts/components of DBMS overall structure in detail:

1. **Query Processor Components:** It has following sub-components:
   (i) **DML Compiler:** When query is given i.e. DML statement, it translates that DML statement to low level instruction which a query evaluation engine understands. It also translates the users request into more efficient form so that query engine finds a good strategy to execute that query.
   (ii) **Embedded DML Pre-compiler:** It converts the DML statements embedded in application program to normal procedure call in host language. So that it interacts with DML compiler.
   (iii) **DML Interpreter:** It translates data definition language statement into the set of tables in the form of metadata. The metadata is data about data and called data dictionary.
   (iv) **Query Processor:** It executes the low level instruction passed by DML compiler to fetch the necessary data.

2. **Storage Management Components:** These components provide an interface between low level data stored in database and application program and the queries submitted to the system. Following are the sub-components:
   (i) **Authorization and Integrity Manager:** When a query is submitted to a system, it is first checked for authorized data access i.e. person who executes query has a right or not. Then query is checked for integrity constraint.
   (ii) **Transaction Manager:** Despite of system failure, this ensures about the consistency of database. It also maintains the atomicity.
   (iii) **File Manager:** It manages the allocation of disk space and data structure used to represent information stored on disk.
   (iv) **Buffer Manager:** It is responsible to fetch data from disk into main memory and also decides which data to cache in memory.

- Other than these two main parts, there are some data structures available which resides in DBMS. These data structures are:
  (i) **Data Files:** This stores the database.
  (ii) **Data Dictionary:** This stores the information about data in database called as metadata.
  (iii) **Indices:** This provides fast access to the data items that hold particular files.
  (iv) **Statistical Data:** This stores statistical data about the data.

1.5 USERS IN DBMS AND ADVANTAGES OF DBMS

- Database users are nothing but the people or group of peoples who access/interact the database directly or indirectly.
- DBMS is used by various users for various purposes. Some may involve in retrieving data and some may involve in backing it up.
- The users of a database system can be classified into various categories depending upon their interaction and degree of expertise of the DBMS.

1.5.1 Users in DBMS

(April 15, 18)

- The users of database systems can be classified in the following groups. These groups are on the basis of degree of expertise or the mode of their interactions with DBMS.
  1. DBMS users.
  2. Database managers.
  3. Database administrators.

1.5.1.1 DBMS Users

(April 15, 18)

- DBMS provide an environment to store and retrieve information. On the basis of interaction with the system, users are differentiated in four types as explained below:

1. **Application Programmer:**
- These are the computer professionals. They are responsible for developing application programs.
- For development of application programs, these people used a general purpose programming language like C, COBOL, Pascal.
- Through these program application programmers handle or manipulate the databases.
2. **Sophisticated Users (On-Line Users):**
- These people are also computer professional but they do not write programs. To interact with the system, they use query languages like SQL.

- These are users who may communicate with the database directly via an on-line terminal or indirectly via a user interface and application program.
- These users are aware of the presence of the database system and have limited interaction with database through application programs.
- The more sophisticated of these users may also use a data manipulation language to manipulate the database directly. They are also called as on-line users.

3. **Specialized Users:**
- They are sophisticated users. They write specialized database application which do not fit into traditional dataframe work.
- **For example:** Computer-Aided Design (CAD) system, knowledge base and expert systems etc., that store data with complex data types (For example: Graphics data and audio data) and environment-modeling systems.

(April 17)

4. **Naive Users (Un-experienced Users):**
- There are the users who are not aware of the presence of database system or any other system.
- These users are called as unsophisticated users who interact with the system through already written permanent programs.
- In simple words, naive users are the end users.

1.5.1.2 Database Manager

- Databases require a large amount of storage space, so storing the database requires secondary storage.
- It is essential that the database systems, structure the data to minimize the need to move data between disk and main memory.
- A database manager (DB Manager) provides basic database management functionalities like creation and maintenance of databases.
- A database manager is a program module, which provides the interface between the low-level data stored in the database and the application programs and queries submitted to the system.
- The database manager is responsible for the following tasks/functions:

1. **Interaction with File Manager:**
- The raw data is stored on the disk using the file system, which is usually provided by a conventional operating system.
- The database manager translates the various DML statements into low-level file system commands. Thus, the database manager is responsible for the actual storing, retrieving and updating of data in the database.
2. **Integrity Enforcement:**
- The data values stored in the database must satisfy certain types of consistency constraints. **For example:** the number of hours an employee may work in one week may not exceed than some specific limit (say 80 hrs) otherwise appropriate action is taken by database manager.
3. **Security:**
- Security to users is provided by database manager i.e., to access only required data.
4. **Backup and Recovery:**
- In case of hard disk crash or power failure or software errors, database manager have responsibility to defect such failures and restore the database.
5. **Concurrency Control:**
- Controlling the interaction among the concurrent users is another responsibility and database manager.

1.5.1.3 Database Administrator (DBA)

(April 16, 17, 18, Oct. 17)

- The DBA is a person who must have a good knowledge in data processing. He/She must have sufficient technical knowledge to make wise decisions in the organization.
- The DBA must manage the staff to ensure orderly development of the database project, to satisfy database users and to plan for future database requirements.
- DBA is responsible for overall performance of database. Some of the main responsibilities (functions) of DBA are as follows:

1. **Monitoring the Performance:** DBA is responsible for overall performance of database.
2. **Deciding the Internal Schema of Structure of Physical Storage:** DBA decides how the data is actually stored at physical storage, how data is represented at physical storage.
3. **Deciding the Conceptual Schema or Contents of Database:** DBA decides the data fields, tables, queries, data types, attributes, relations, entities or we can say that he/she is responsible for overall logical design of database.
4. **Deciding User View:** DBA decides different views for different users.
5. **Deciding Constraints:** DBA decides various constraints over database for maintaining consistency and validity in database.
6. **Deciding Users:** DBA gives permission to user to use database. Without having proper permission, no one can access data from database.
7. **Granting of Authorities:** DBA gives authorities or rights to data access. User can use only that data on which access rights is granted to him/her.
8. **Liaisioning with Users:** Another task of the DBA is to liaisioning with users and ensure the availability of the data they require and write the necessary external schemas.
9. **Removal of Dump and Maintain Free Space:** DBA is responsible for removing unnecessary data from storage and maintain enough free space for daily operations.
10. **Checks:** DBA also decides various security and validation checks over database to ensure consistency.
11. **Backup:** DBA takes regular backup of database, so that it can be used during system failure.
12. **Security:** DBA takes various steps to make data more secure against various disasters and unauthorized access of data.

1.5.2 Advantages of DBMS

(April 16, 17, 18, Oct. 17)

- Various advantages of DBMS over file system are listed below:

1. **Sharing of Data:** In DBMS the data is centrally controlled and can be shared by all authorized users.
2. **Improved Data Integrity:** In database systems, data integrity means that the data contained in the database is both accurate and consistent. The centralized control of DBMS allow adequate checks can be incorporated to provide data integrity.
3. **Data Consistency:** In DBMS the problem of inconsistent data is automatically solved by controlling the redundancy.
4. **Controlled Redundancy:** In DBMS, the duplication of data can be carefully controlled, that means the database system is aware of the redundancy and it assumes the responsibility for propagating updates.
5. **Efficient Data Access:** The DBMS utilizes different sophisticated techniques to access the stored data very efficiently.
6. **Program Data Independence:** The DBMS provide an independence between the file system and application program, that allows for changes at one level of the data without affecting others.

7. **Enforcement of Standards:** In DBMS, data being stored at one central place, standards can easily be enforced by the DBA. This ensures standardised data formats to facilitate data transfers between systems.
8. **Improved Security:** Database security means protecting the data contained in the database from unauthorized users. In DBMS the DBA ensures that proper access procedures are followed, including proper authentical schemes for access to the database and additional checks before permitting access to sensitive data.
9. **Improved Backup and Recovery Facility:** Through backup and recovery subsystem of DBMS provides the facilities for recovering from hardware or software failures.
10. **Concurrency Control:** The DBMS are designed to manage simultaneous (concurrent) access of the database by many users. They also prevents any loss of information (loss of integrity) due to these concurrent accesses.
11. **High Data Quality:** The quality of data in database systems are very high as compared to traditional file systems.
12. **Minimal Program Maintenance:** In a traditional file system, high maintenance effort are required. These are reduced to minimal in database systems due to independence of data and application programs.
13. **Economical to Scale or Low Cost:** In DBMS, the operational data of an organization is stored in a central database. The application programs that work on this data can be built with very less cost as compared to traditional file system. This reduces overall costs of operation and management of the database that leads to an economical scaling.
14. **Increased Programmer Productivity and Reduced Development Time:** The DBMS provides many standard functions and these functions allow the programmers to concentrate on the specific functionality required by the users without worrying about the implementation details. This increases the overall productivity of the programmer and also reduced the development time and cost.
15. **Good Data Accessibility and Responsiveness:** The database systems provide query languages (or report writers) that allow the users to ask ad-hoc queries to obtain the needed information immediately, without the requirement to write application programs.
14. **Language Interface:** DBMS provide a language support for definition and manipulation of data.

PRACTICE QUESTIONS

Q. I: Multiple Choice Questions:

1. A collection of related record is called _____.
   (a) file (b) record (c) byte (d) field
2. Which of the following file operation modifies the file?
   (a) insert (b) delete (c) modify (d) all
3. Data is better organized and is easily accessible by using _____.
   (a) DBMS (b) FMS
   (c) EMS (d) all
4. A _____ is a set of operations that must be performed completely or not at all.
   (a) query (b) command
   (c) data sharing (d) transaction
5. Which of the following is an advantages of database?
   (a) reduction in redundancy (b) avoidance of inconsistency
   (c) security enforcement (d) all above

6. _____ is defined as, "the software system that allows users to define, create, maintain and control access to the database".
   (a) DBMS (b) File system
   (c) Both (a) and (b) (d) None
7. Which is a software or program that allows access/ retrieval to (of) data (that) contained in database.
   (a) File system (b) DBMS
   (c) Both (a) and (b) (d) None
8. Different methods of accessing files include _____.
   (a) Sequential Access (b) Direct/Random Access,
   (c) Indexed sequential Access (d) All of these
9. _____ is a collection of data.
   (a) DBMS (b) File system
   (c) Database (d) None
10. Levels of abstraction includes _____.
    (a) Physical (Internal) Level (b) View (External) Level
    (c) Logical (Conceptual) Level (d) All of these
11. Monitoring jobs running on database, should be supervised by _____.
    (a) Database system (b) Database Manager
    (c) Database users (d) Database administrator

Answers

1. (a) 2. (c) 3. (a) 4. (d) 5. (d) 6. (a) 7. (b) 8. (d) 9. (c) 10. (d) 11. (d)

Q. II: Fill in the Blanks:

1. A_____ is a named collection of related information.
2. _____ can be defined as, "a set of isolated and unrelated raw facts with an implicit meaning".
3. A_____ is a complex software system that is used to manage, store and manipulate data.
4. The physical arrangement of data in a file into records on the disk is known as_____.
5. A _____ is a collection of information that is organized so that it can easily be accessed, managed, and updated.
6. In file processing system the records are stored in separate _____
7. _____ means the property to change the overall logical or physical structure of the data without changing the application programs view of data.
8. Data _____ is refers to, "the hiding of certain details of how the data are stored and maintained".
9. A _____ provides basic database management functionalities like creation and maintenance of databases.

Answers

1. file 2. Data 3. DBMS 4. file organization 5. database
6. files 7. Data independence 8. Abstraction 9. database manager

Q. III: State True or False:

1. Using a set of files is better than using a database.
2. RDBMS is the most popular of the possible DBMS.

3. Data independence means the property to change the overall logical or physical structure of the data without changing the application programs view of data.
4. External view of a database present the view that the end users have.
5. Sophisticated Users are also computer professional, they can write programs.
6. A file is a collection of related sequence of records.
7. A DBMS is a collection of logically related data stored together that is designed to meet information requirements of an organization.
8. Schema is a distinct piece of information and is explained in the previous section.
9. A database is an organized collection of data from which users can efficiently retrieve the desired information.
10. DBMS is the software system that allows users to define, create, maintain and control access to the database.
11. File processing system supports data redundancy and data isolation.
12. In logical data independence, the physical layout, organization of data may be changed without changing either the overall logical structure of the data or the application programs.
13. The overall description of the database is called as the database instance.
14. View level describes in detail how the raw data is actually stored at the byte level using complex low level data structure.

Answers

1. (F) 2. (T) 3. (T) 4. (T) 5. (F) 6. (T) 7. (F) 8. (F) 9. (T) 10. (T) 11. (F) 12. (F) 13. (F) 14. (F)

Q. IV: Answer the following Questions:

(A) Short Answer Questions:

1. What is file? Define it.
2. What is file organization?
3. Enlist file organization techniques.
4. What is record? Define it.
5. What is DBMS?
6. List two functions of DBMS.
7. Define database?
8. Enlist elements of a database.
9. What is DBA?
10. What is schema?
11. Define database instance.
12. What is meant by file processing system?
13. What are the levels of data abstraction?
14. Give two types of data independence.
15. Give two differences between file processing system and DBMS.
16. Enlist users of DBMS.

(B) Long Answer Questions:

1. Define file organization? Enlist file organization techniques with their advantages and disadvantages..
2. What is physical and logical files ? Compare them.
3. With the help of diagram describe elements of database.

4. Differentiation between sequential, indexed and hashed file organization.
5. What is database system? Explain its role in detail.
6. What are the functions of DBMS?
7. What are drawbacks of file system?
8. Differentiate between the file system and DBMS.
9. State the advantages of DBMS.
10. What is data abstraction? What are different levels of data abstractions?
11. What is data independence? What are its different types?
12. What is DBMS user?
13. What are tasks or functions of database manager?
14. What are functions of DBA?
15. What are different types of database users?
16. Explain overall DBMS structure with a neat diagram.
17. Give the difference between logical and physical data independence.
18. Explain following terms:
    (i) Data independence, (ii) Data redundancy.
19. "DBA is responsible for organizing the system such that it should give best performance", explain.
20. How is DBMS better than file system?
21. What is schema and instance? Define and compute them.

UNIVERSITY QUESTIONS AND ANSWERS

April 2015

1. What are the levels of abstraction of a database schema? (1 M)

Ans. Refer to Section 1.3.1.

2. Give different operations performed on a file. (1 M)

Ans. Refer to Page 1.2.

3. Give any two disadvantages of traditional file system. (1 M)

Ans. Refer to Page 1.7.

4. Explain various types of users in DBMS. (5 M)

Ans. Refer to Section 1.5.1.

5. What are the different types of file organizations? Explain any two in detail. (5 M)

Ans. Refer to Section 1.1.2.

6. What is data independence? Explain the two types of data independence in brief. (5 M)

Ans. Refer to Section 1.3.2.

April 2016

1. What is a logical file? (1 M)

Ans. Refer to Page 1.2 Point (1).

2. State and explain the different levels of abstraction. (5 M)

Ans. Refer to Section 1.3.1.

3. Explain any two types of file organizations. (5 M)

Ans. Refer to Section 1.1.2.

4. Explain the advantages of DBMS over traditional file system. (5 M)

Ans. Refer to Section 1.5.2.

5. What are the different functions performed by DBA? (5 M)

Ans. Refer to Section 1.5.1.3.

October 2016

1. List out any two disadvantages of DBMS. (1 M)

Ans.: Disadvantages of DBMS :

(a) **Requirement of More Memory Space:** The wide functionality and more complexity increase the size of DBMS. Thus, it requires much more memory space to store and run than the traditional file system.

(b) **Additional Cost of Hardware:** The cost of DBMS installation is much more and it depends on environment and functionality, size of the hardware and maintenance costs of hardware.

(c) **Complexity Increases:** The data structure in DBMS, may become more complex and critical because of the centralized database supporting many applications in an organization.

(d) **Organizational Conflict:** A centralised and sheared database system requires a consensus on data definitions as well as responsibilities for accurate data maintenance.

(e) **Need for Backup and Recovery:** For a DBMS to be accurate and available all times, a procedure is required to be developed and used for providing backup copies to all its users when damage occurs.

(f) **More Installation and Management Cost:** The big and complete database systems are more costly. They require trained manpower to operate the system and has additional annual maintenance and support costs.

(g) **High Cost of Conversion:** The cost of conversion from old file-system to new database system is very high.

2. State any two functions of DBA. (1 M)

Ans. Refer to Section 1.1.5.3.

3. Define logical file. (1 M)

Ans. Refer to Page 1.2 Point (1).

4. What is ISAM? (1 M)

Ans. Refer to Page 1.3 Point (3).

5. What is file organization? Explain any two types with example. (5 M)

Ans. Refer to Section 1.1.2.

6. Explain different functional components of DBMS. (5 M)

Ans. Refer to Section 1.4.

7. What is data independence? (3 M)

Ans. Refer to Section 1.3.2.

8. What are different levels of data abstraction? (3 M)

Ans. Refer to Section 1.3.1.

April 2017

1. What is Hashing? (1 M)

Ans. Refer to Page 1.3 Point (4).

2. Who are Naive users? (1 M)

Ans. Refer to Section 1.5.1.1 Point (4).

3. What is DBMS? What are the advantages and disadvantages of DBMS? (5 M)

Ans. Refer to Sections 1.1.4 and 1.5.2.

4. Write a note on functional components of DBMS. (5 M)

Ans. Refer to Section 1.4.

October 2017

1. What is a physical file? (1 M)

Ans. Refer to Page 1.2 Point (2).

2. State any two functions of DBA. (1 M)

Ans. Refer to Section 1.5.1.3.

3. State any two advantages of DBMS over file system. (1 M)

Ans. Refer to Section 1.5.2.

4. What are the types of Data Independence? (1 M)

Ans. Refer to Section 1.3.2.

5. Write short note on Data Abstraction. (5 M)

Ans. Refer to Section 1.3.1.

6. Explain overall DBMS structure with neat diagram. (5 M)

Ans. Refer to Section 1.4.

April 2018

1. List any two advantages of DBMS. (1 M)

Ans. Refer to Section 1.5.2.

2. State different types of users of DBMS. (1 M)

Ans. Refer to Sections 1.5.1 and 1.5.1.1.

3. Write a short note on data independence. (5 M)

Ans. Refer to Section 1.3.2.

❖❖❖

CHAPTER 2

Conceptual Design

2.0 INTRODUCTION

- Database design is the process of producing/developing a detailed data model of a database. Database design is the process of constructing a stable database structure from user requirements.
- Database design enables us to represent the real-world entities in a form that can be processed by the computer.
- Database design requires understanding both the operational and business requirements of an organization as well as the ability to model and realize those requirements using a database.
- Conceptual modeling is an important in designing a successful database. Conceptual modeling is also called as semantic modeling.
- Conceptual modeling represents various pieces of data and their relationships at a very high-level of abstraction. Conceptual design mainly focuses on what data is required and how it should be organized rather than what operations are to be performed on the data.
- Database desgin involves a set of activities, which include requirements analysis, conceptual database design, selection of a database management system, mapping from conceptual model to physical model and the physical database design.

2.1 OVERVIEW OF DB DESIGN PROCESS

- The process of designing a database begins with an analysis of what information the database must hold and what are the relationships among components of that information.
- Design the logical and physical structure of one or more databases to accommodate the information needs of the users in an organization for a defined set of applications.
- The database design is a process of designing the logical and physical structure of one or more databases. The reason behind it is to accommodate the information needs or queries of the users for a defined set of applications and support the overall operation and objectives of an enterprise/organization.
- The goals of database (DB) design are explained below:
  1. Provide a natural and easy-to-understand structuring of the information.
  2. Satisfy the information content requirements of the specified users and applications.
  3. Support processing requirements and any performance objectives, such as response time, processing time and storage space.
- Various phases in DB design in Fig. 2.1 are explained below:

Phase 1: Requirements Collection and Analysis:

- Before we can effectively design a database, we must know and analyze the expectations/ requirements of the users and the intended uses of the database in as much detail as possible. This process is called requirements collection and analysis.
- To specify the requirements, we must first identify the other parts of the information system that will interact with the database system.
- These include new and existing users and applications, whose requirements are then collected and analyzed.