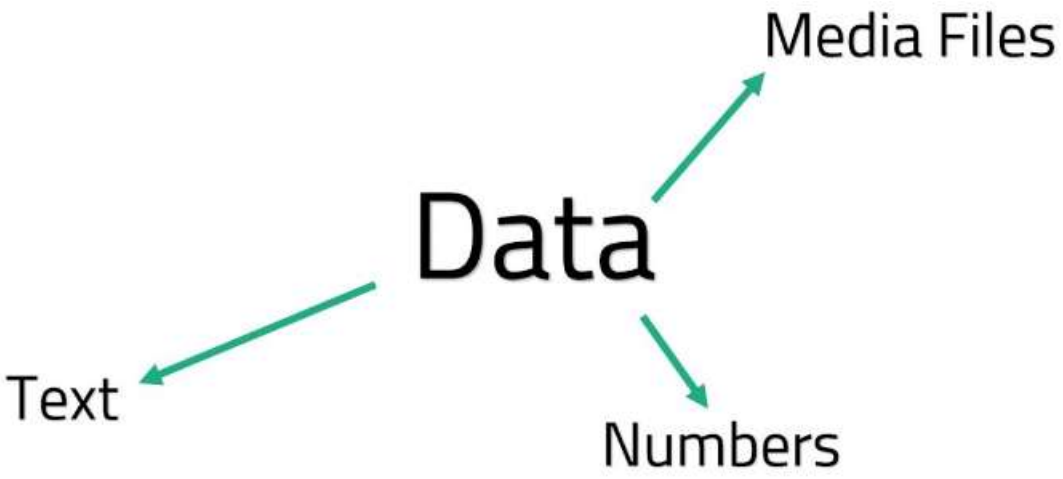
DATABASE MANAGEMENT SYSTEM

UNIT -1

DATA:

Data is defined as a collection of individual facts. Data can come in the form of text, observations, figures, images, numbers, graphs, or symbols. Data is a single unit of information or we can say information is collection of data.

The simplest definition of "data" is: **"Facts, information, or details that can be collected, stored, and used to gain knowledge or make decisions."**



INFORMATION:

Information can be understood as data that has been processed, organized, or contextualized in a way that gives it meaning and relevance. An information is collection of data and information helps in decision making. For example: name, age, address.

There are many differences between data and information, and some of them are given below:

Data	Information
Data is the raw facts	Information is a processed form of data.
Data is not significant.	Information is significant to a business.

Data	Information
Data is the atomic level piece of information.	Information is a collection of data.
Data does not help in decision making.	The information helps in decision making.
For example, 23 is a data	For example, Age 23 is an example of information

DATABASE :

A database is information that is set up for easy access, management and updating. Computer databases typically store aggregations of data records or files that contain information, such as sales transactions, customer data, financials and product information.

A **database** is an organized collection of structured data that is stored and managed in a systematic way. Databases are used to store a wide range of data, from small amounts of personal information to massive volumes of data for large organizations.

PERSISTANT DATA:

Persistent data, often referred to as "persistence," refers to data that is saved or stored beyond the current session or instance of a program or system. It remains available even when the application or system is closed, restarted, or powered off. Persistence ensures that data is retained over time and can be accessed in subsequent interactions.

persistent data refers to data that is stored and maintained within a database even after the application or system that interacted with the data has terminated or the computer has been shut down. DBMSs are designed to provide reliable and efficient ways of managing persistent data.

1. **Data Storage and Structure:**
    - **Tables:** Data in a DBMS is organized into tables, each with a defined structure and set of columns representing different attributes.

- **Rows:** Each row within a table represents a specific record or instance of data.
- **Columns:** Columns define the attributes or fields of the data, each holding a specific type of information.

2. **Data Integrity:**
    - **Constraints:** DBMSs enforce constraints such as primary keys (uniquely identifying rows), foreign keys (establishing relationships between tables), and uniqueness constraints (ensuring unique values in a column).
    - **Data Validation:** DBMSs validate data against defined data types and constraints to maintain data integrity.
3. **Transactions and ACID Properties:**
    - **Transactions:** A transaction is a sequence of database operations treated as a single unit. It ensures data consistency and integrity.
    - **ACID Properties:** Transactions adhere to ACID properties: Atomicity (all or nothing), Consistency (data remains valid), Isolation (transactions are isolated from each other), and Durability (committed data is permanent).
4. **Data Retrieval and Querying:**
    - **SQL:** Structured Query Language (SQL) is used to retrieve, manipulate, and manage data in a DBMS.
    - **Queries:** Queries allow users to retrieve specific data using conditions, sorting, and aggregations.
5. **Indexes:**
    - **Indexes:** Indexes are data structures that enhance query performance by providing efficient access paths to data. They speed up data retrieval by reducing the need for full table scans.
6. **Backup and Recovery:**
    - **Backup:** Regular database backups create copies of data that can be used for recovery in case of data loss or corruption.
    - **Recovery:** Recovery mechanisms restore the database to a consistent state after a failure.
7. **Concurrency Control:**
    - **Concurrency:** DBMSs manage multiple users accessing data simultaneously. Concurrency control ensures data consistency in a multi-user environment.

8. **Redundancy and Replication:**
    - **Redundancy:** Storing redundant copies of data provides fault tolerance and minimizes the risk of data loss due to hardware failures.
    - **Replication:** Replicating data across multiple servers improves availability and load distribution.
9. **Archiving and Purging:**
    - **Archiving:** Older data can be archived to improve performance, as archived data is less frequently accessed.
    - **Purging:** Removing unnecessary or obsolete data from the database helps maintain efficient data storage.
10. **Schema Evolution:**
    - **Schema Changes:** DBMSs allow modifications to the database schema (structure) while preserving existing data, often through techniques like migration scripts.
11. **Distributed Databases:**
    - **Distribution:** Data can be distributed across multiple servers or nodes to improve performance and scalability.
    - **Consistency:** Distributed DBMSs maintain data consistency across distributed nodes using techniques like two-phase commit.
12. **Security and Access Control:**
    - **Access Control:** DBMSs enforce user authentication and authorization to ensure that only authorized users can access and modify data.
    - **Encryption:** Some DBMSs support data encryption to protect sensitive information.

These concepts collectively ensure that data within a DBMS is stored, accessed, and managed effectively, while maintaining data integrity, security, and availability over time.

FILE:

A file refers to a sequence of records stored in binary format. It is a way of organizing and storing data in a structured manner. In a Database Management System (DBMS), a "file" typically refers to a collection of related data stored together. However, the

term "file" in the context of DBMS is not the same as a traditional file on an operating system. Instead, it's used to describe a higher-level organization of data within a database.

A higher-level organization of data within a database is referred to as a data hierarchy. It is a systematic arrangement of data in a hierarchical form, starting from the smallest unit of data, such as a bit, to the largest unit, such as a database. The data hierarchy is organized into several levels, including bits, fields, records, files, and databases. Each level is organized from the components below it, and the relationships among the individual records in databases are based on one of several logical data structures or models. A data model is a method for organizing databases on the logical level, and it represents the relationships among database records. The data hierarchy is essential in databases with referential integrity, third normal form, or perfect key, and it helps to show the relationships between smaller and larger components in a database or data file.

FILE MANAGEMENT SYSTEM:

A file management system is a type of software used for file maintenance operations.
It is designed to manage individual or group files, such as special office documents and records. A file management system has limited capabilities and is used to manage data files in a computer system. It displays report details, such as owner, creation date, state of completion, and similar features useful in an office environment. A file management system is also known as a file manager.

Here are some key differences between a file system and a database management system (DBMS)

- A file system is a software that manages and organizes files in a storage medium, while a DBMS is a software for storing and retrieving user's data while considering appropriate security measure.
- A file system enables you to handle the way of reading and writing data to the storage medium, while a DBMS gives an abstract view of data that hides the details.
- A file system provides the details of data representation and storage of data, while a DBMS is efficient to use as there are a wide variety of methods to store and retrieve data.
- A file system is a DBMS that allows access to single files or tables at a time, while a DBMS controls redundancy by maintaining a single repository of data that is defined once and is accessed by many users.

In summary, a file management system is a type of software used for file maintenance operations.

Basic File Operation:

Basic file operations refer to the fundamental operations that can be performed on a file in a database management system (DBMS).

Operations on database files can be classified into two categories:

- Update Operations
- Retrieval Operations

Update operations change the data values by insertion, deletion or update.

Retrieval operations on the other hand do not alter the data but retrieve them after optional conditional filtering. In both types of operations, selection plays significant role. Other than creation and deletion of a file, there could be several operations, which can be done on files.

Here are some of the basic file operations:

- **Open**: A file can be opened in one of two modes, read mode or write mode. Opening a file allows the user to access the contents of the file.
- **Locate**: Every file has a file pointer, which tells the current position where the next read or write operation will occur. The locate operation is used to move the file pointer to a specific position in the file.
- **Read**: The read operation is used to retrieve data from a file. It reads the data from the file and stores it in a buffer.
- **Write**: The write operation is used to add data to a file. It writes the data from the buffer to the file.
- **Update**: The update operation is used to modify the existing data in a file. It changes the data values by insertion, deletion, or update.
- **Delete**: The delete operation is used to remove data from a file. It deletes the record from the file.
- **Close**: The close operation is used to close the file after it has been opened. It releases all the buffers and file handlers.

Disadvantages of File system

- Data redundancy and inconsistency.
- Difficulty in accessing data
- Data isolation
- Integrity problems.
- Atomicity problems.
- Concurrent-access anomalies
- Security Problem
    - Advanced recovery techniques in a database
    - Indexing Techniques in a Database

File Structure and organization:

File structure:

Relative knowledge and data are collectively stored in file formats. A file is a record sequence that is stored in binary format. A disk drive that can store records is formatted into several blocks. On such disk blocks, file records are mapped.

File organization is a way of organizing the data in such way so that it is easier to insert, delete, modify and retrieve data from the files.

Purpose of File Organization

1. File organization makes it **easier & faster** to perform operations (such as read, write, update, delete etc.) on data stored in files.
2. **Removes data redundancy.** File organization make sure that the redundant and duplicate data gets removed. This alone saves the database from insert, update, delete operation errors which usually happen when duplicate data is present in database.
3. It **saves storage cost**. By organizing the data, the redundant data gets removed, which lowers the storage space required to store the data.
4. **Improves accuracy**. When redundant data gets removed and the data is stored in efficient manner, the chances of data get wrong and corrupted goes down.

Types of File Organization

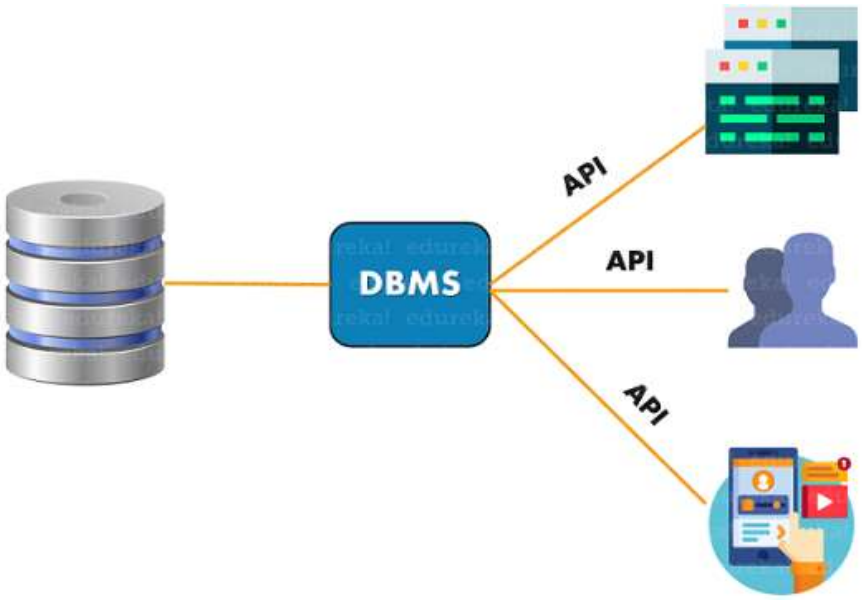
There are various ways to organize the data. Every file organization method is different from each other; therefore, each file organization method has its own advantages and disadvantages. It is up to the developer which method they choose in order to organize the data. Usually, this decision is made based on what kind of data is present in database.

Types of file organization:

1. **Sequential File Organization**
2. **Heap File Organization**
3. **Hash File Organization**
4. **B+ Tree File Organization**
5. **Clustered File Organization**
6. **Indexed sequential access method (ISAM)**

DBMS:

Database Management Systems (DBMS) are software systems used to store, retrieve, and run queries on data. It provides an interface for users and applications to interact with the data stored in the database. With the help of DBMS, you can easily create, retrieve and update data in databases. A DBMS consists of a group of commands to manipulate the database and acts as an interface between the end-users and the database.



A Database Management System (DBMS) is a software system that allows users to create, define, manipulate and manage databases. It provides a way for organizations to store, organize and retrieve large amounts of data quickly and efficiently in an organized manner.

Components of a DBMS

A DBMS is made up of several components that work together to ensure the efficient use and management of data. At its core, we can summarize six components:

1. Hardware

This refers to the physical devices, including the computer itself, that are used to store the data. Some examples of hardware used in a DBMS include:

- Computer
- Hard disks
- I/O channels

2. Software

The software component of a DBMS refers to the computer programs used to manage the database.This could include the DBMS itself or supporting software like drivers or utilities.
Simply put, the software is a layer that is supported by the hardware to create a visual interface for the user to interact with to control the database.

3. Data

The data generated by an organization is the lifeblood of a DBMS. Data is the resource that is stored within a database.
Metadata, which is information about the data, is stored within a DBMS as well. This information provides some context about the data and helps to ensure that it is maintained, secured and accessed correctly. For example, metadata could include information like the type of data, its size, and the time it was created.

4. Procedures

DBMS procedures are the guidelines that dictate how the database is set up and should be used. These procedures govern how data is accessed, modified and archived. In simple words, these are the instructions on how a DBMS should be used.

5. Database access query language

When accessing a database, users must use a specific query language that the DBMS understands. These languages can be generally split into four categories:

1. Data definition language (DDL)
2. Data manipulation language (DML)
3. Data control language (DCL)
4. Transaction control language (TCL)

These languages play a part in searching, fetching, updating and adding data to the database. Without the use of languages, the database cannot be accessed easily through interfaces.
Common query languages include:

- Structured Query Language (SQL)
- Object Query Language (OQL)
- XQuery

6. Users

Users are the people who access the database. Depending on their level within the organization, users might access different parts of the database and use different software tools to interact with the data.
Some examples of users include:

- **Database Administrators (DBAs)** create the database and set up access permissions.
- **Data Analysts** analyze data in the database to gain insights and make decisions.
- **Software developers** write programs that interact with the database.