**DBMS Topics**

* What is data and information ?
* What is databases and types and what is DBMS
* Filesystem and dbms and its advantages and disadvantages
* Logical and data independence
* 3tier architecture
* Databases langugages (like dml,tcl etc)
* How db is accessed through application (jdbc,odbc..)
* Database administrator and its role
* Tiers in dbms (1,2,3,n…)
* ER model
* Attribute , relation,entity-types
* Carinality
* Extended-ER->specialist,generalization
* How to make ER diagrams
* Relational model
* Types of keys
* Integrity constraints(domain,entity,referencial,key)
* Anamoli contrainst
* Mapping of ER to Relational Model(with egs practice)
* Functional dependencies
* Armstrong relation
* If normalization is not done (then 3 anamolies)
* Normalization (till SRP is achieved)
* ACID properties and the process by which it is achieved
* Various Indexing DB methods

What is Data?

a. Data is a collection of raw, unorganized facts and details like text, observations, figures, symbols,

and descriptions of things etc.

In other words, data does not carry any specific purpose and has no significance by itself.

Data can be recorded and doesn’t have any meaning unless processed.

What is Information?

a. Info. Is processed, organized, and structured data.

b. It provides context of the data and enables decision making.

c. Processed data that make sense to us.

DBMS

A **database**is a collection of interrelated data that helps in the efficient retrieval, insertion, and deletion of data from the database and organizes the data in the form of tables, views, schemas, reports, It allows users to create, modify, and query a database, as well as manage the security and access controls for that database. DBMS provides an environment to store and retrieve data in convenient and efficient manner

**Key Features of DBMS**

* **Data modeling:**  provides tools for creating and modifying data models, which define the structure and relationships of the data in a database.
* **Data storage and retrieval:** responsible for storing and retrieving data from the database, and can provide various methods for searching and querying the data.
* **Concurrency control:**  provides mechanisms for controlling concurrent access to the database, to ensure that multiple users can access the data without conflicting with each other.
* **Data integrity and security:**  provides tools for enforcing data integrity and security constraints, such as constraints on the values of data and access controls that restrict who can access the data.
* **Backup and recovery:**  provides mechanisms for backing up and recovering the data in the event of a system failure.
* **DBMS can be classified into two types:** Relational Database Management System (RDBMS) and Non-Relational Database Management System (NoSQL or Non-SQL)
* **RDBMS:**Data is organized in the form of tables and each table has a set of rows and columns. The data are related to each other through primary and foreign keys.
* **NoSQL:**Data is organized in the form of key-value pairs, documents, graphs, or column-based. These are designed to handle large-scale, high-performance scenarios.

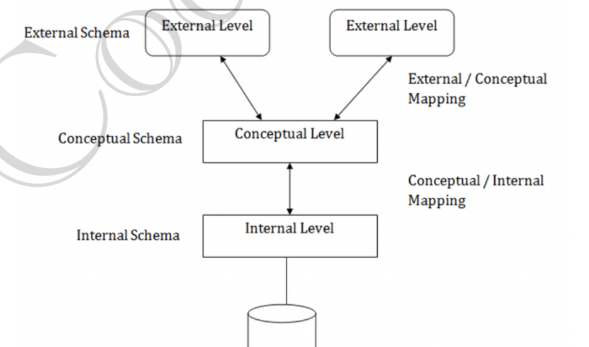
**Types of DBMS**

1. **Relational Database Management System (RDBMS):** Data is organized into tables (relations) with rows and columns, and the relationships between the data are managed through primary and foreign keys. SQL (Structured Query Language) is used to query and manipulate the data.
2. **NoSQL DBMS:** Designed for high-performance scenarios and large-scale data, NoSQL databases store data in various non-relational formats such as key-value pairs, documents, graphs, or columns

**Traditional File system : -**

* **Redundancy of data**: Data is said to be redundant if the same data is copied at many places. If a student wants to change their Phone number, he or she has to get it updated in various sections. Similarly, old records must be deleted from all sections representing that student.
* **Inconsistency of Data**: Data is said to be inconsistent if multiple copies of the same data do not match each other. If the Phone number is different in Accounts Section and Academics Section, it will be inconsistent. Inconsistency may be because of typing errors or not updating all copies of the same data.
* **Difficult Data Access**: A user should know the exact location of the file to access data, so the process is very cumbersome and tedious. If the user wants to search the student hostel allotment number of a student from 10000 unsorted students’ records, how difficult it can be.
* **Unauthorized Access**: File Systems may lead to unauthorized access to data. If a student gets access to a file having his marks, he can change it in an unauthorized way.
* **No Concurrent Access**: The access of the same data by multiple users at the same time is known as concurrency. The file system does not allow concurrency as data can be accessed by only one user at a time.
* **No Backup and Recovery**: The file system does not incorporate any backup and recovery of data if a file is lost or corrupted.

**3-Tier Architecture in DBMS**



**View of Data** (Three Schema Architecture)

a.The major purpose of DBMS is to provide users with an abstract view of the data. That is, the system hides certain details of how the data is stored and maintained.

b. To simplify user interaction with the system, abstraction is applied through several levels of abstraction.

c. The main objective of three level architecture is to enable multiple users to access the same data with a personalized view while storing the underlying data only once

d. **Physical level/Internal level**

i. The lowest level of abstraction describes how the data are stored.

ii. Low-level data structures used.

iii. It has Physical schema which describes physical storage structure of DB.

iv. Talks about: Storage allocation (N-ary tree etc), Data compression & encryption etc.

v. Goal: We must define algorithms that allow efficient access to data.

e. **Logical level / Conceptual level**:

i. The conceptual schema describes the design of a database at the conceptual level, describes what data are stored in DB, and what relationships exist among those data. ii. User at logical level does not need to be aware about physical-level structures. iii. DBA, who must decide what information to keep in the DB use the logical level of abstraction.

iv. Goal: ease to use.

f. **View level/External level**:

i. Highest level of abstraction aims to simplify users' interaction with the system by providing different view to different end-user.

ii. Each view schema describes the database part that a particular user group is interested and hides the remaining database from that user group.

iii. At the external level, a database contains several schemas that sometimes called as subschema. The subschema is used to describe the different view of the database.

iv. At views also provide a security mechanism to prevent users from accessing certain parts of DB.

Instances and Schemas

* 1. The collection of information stored in the DB at a particular moment is called an instance of DB. The overall design of the DB is called the DB schema.

**Data Independence**

Data independence means a change of data at one level should not affect another level. Two types of data independence are present in this architecture:

* **Physical Data Independence:** Any change in the physical location of tables and indexes should not affect the conceptual level or external view of data. This data independence is easy to achieve and implemented by most of the DBMS.
* **Conceptual Data Independence:** The data at conceptual level schema and external level schema must be independent. This means a change in conceptual schema should not affect external schema. e.g.; Adding or deleting attributes of a table should not affect the user’s view of the table. But this type of independence is difficult to achieve as compared to physical data independence because the changes in conceptual schema are reflected in the user’s view.

**Database Languages**

* **Data Definition Language**
* **Data Manipulation Language**
* **Data Control Language**
* **Transactional Control Language**

**Data Definition Language (DDL)**

**DDL** is the short name for Data Definition Language, which deals with database schemas and descriptions, of how the data should reside in the database.

* **CREATE:** to create a database and its objects like (table, index, views, store procedure, function, and triggers)
* **ALTER:**alters the structure of the existing database
* **DROP:** delete objects from the database
* **TRUNCATE:** remove all records from a table, including all spaces allocated for the records are removed
* **COMMENT:**add comments to the data dictionary
* **RENAME:** rename an object

**Data Manipulation Language (DML)**

**DML** is the short name for Data Manipulation Language which deals with data manipulation and includes most common SQL statements such SELECT, INSERT, UPDATE, DELETE, etc., and it is used to store, modify, retrieve, delete and update data in a database. **Data query language(DQL)** is the subset of “Data Manipulation Language”. The most common command of DQL is **SELECT** statement. SELECT statement help on retrieving the data from the table without changing anything in the table.

* **SELECT:**retrieve data from a database
* **INSERT:** insert data into a table
* **UPDATE:** updates existing data within a table
* **DELETE:**Delete all records from a database table
* **MERGE:** UPSERT operation (insert or update)
* **CALL:**call a PL/SQL or Java subprogram
* **EXPLAIN PLAN:** interpretation of the data access path
* **LOCK TABLE:**concurrency Control

**Data Control Language (DCL)**

**DCL**is short for Data Control Language which acts as an access specifier to the database.(basically to grant and revoke permissions to users in the database

* **GRANT:**grant permissions to the user for running DML(SELECT, INSERT, DELETE,…) commands on the table
* **REVOKE:**revoke permissions to the user for running DML(SELECT, INSERT, DELETE,…) command on the specified table

**Transactional Control Language (TCL)**

**TCL** is short for Transactional Control Language which acts as an manager for all types of transactional data and all transactions. Some of the command of TCL are

* **Roll Back:** Used to cancel  or Undo changes made in the database
* **Commit:** It is used to apply or save changes in the database
* **Save Point:** It is used to save the data on the temporary basis in the database

**Data Query Language (DQL)**

**Data query language(DQL)** is the subset of **“Data Manipulation Language”**. The most common command of DQL is the **SELECT statement**. SELECT statement helps us in retrieving the data from the table without changing anything or modifying the table. DQL is very important for retrieval of essential data from a database.

Role of DBA

Designing database structures (tables, schemas, indexes) to ensure efficient storage and access. Installing and configuring the DBMS software. Ensuring data normalization and appropriate relationships between tables. Monitoring database performance and identifying bottlenecks. Optimizing queries, indexes, and database configurations for faster performance. Implementing access controls and user roles to ensure only authorized users can access data. Creating and managing regular database backups to ensure data is not lost. Performing routine maintenance tasks, such as updating database software and applying patches. Assisting application developers and end-users in database-related issues. Ensuring that data remains accurate, consistent, and reliable across the database. Planning for database scalability to handle growing data volumes and user loads.