MODULE – I
DATABASE SYSTEMS

## I. BASIC DEFINITIONS

### ❖ DATA

- Data is distinct piece of information.
- Data can exist in variety of forms as numbers or text on piece of paper, stored in electronic memory or stored in a person's mind.

### ❖ INFORMATION

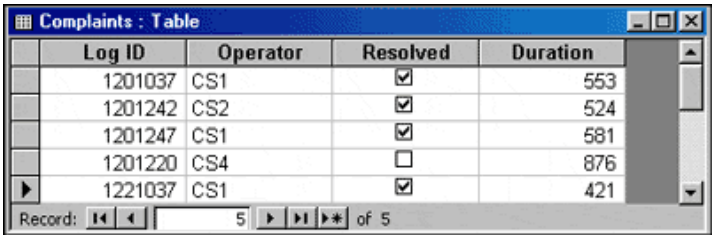
- Knowledge obtained from investigation, study, or instruction.
- In other words it is the answer to a question.

### ❖ DATA BASE

- A database is a collection of related data
- It is a collection of information that is well organized, so that which can be easily accessed, managed and updated.
- Data is organized into rows, columns and tables and it is indexed to make it easier to find relevant information.

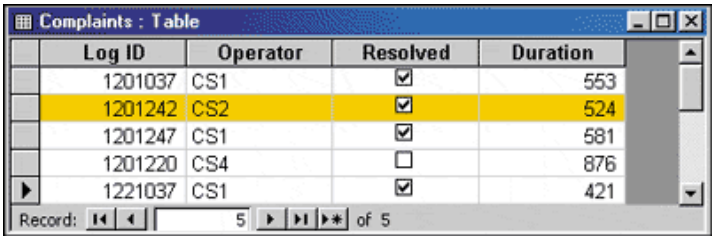
### ❖ TABLES

A database table is composed of records and fields that hold data. Tables are also called datasheets. Each table in a database holds data about a different, but related, subject.



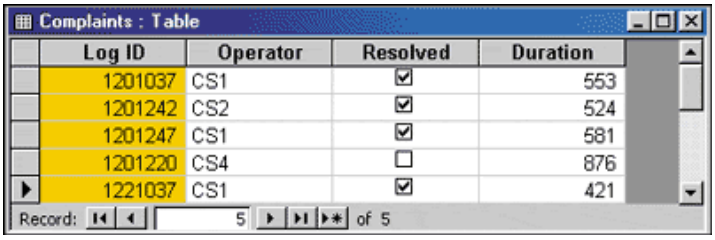

### ❖ Records

Data is stored in records. A record is composed of fields and contains all the data about one particular person, company, or item in a database. In this database, a record contains the data for one customer support incident report. Records appear as rows in the database table.




### ❖ Fields

A field is part of a record and contains a single piece of data for the subject of the record.




In the above database table , each record contains four fields:

| | |
|---|---|
| Log ID | A number assigned to this customer support incident for identification purposes |
| Operator | The code for the customer support operator who handled this incident |
| Resolved | A check box to indicate whether the incident was resolved |
| Duration | The time in seconds the operator spent on this incident |

Fields appear as columns in a database table.

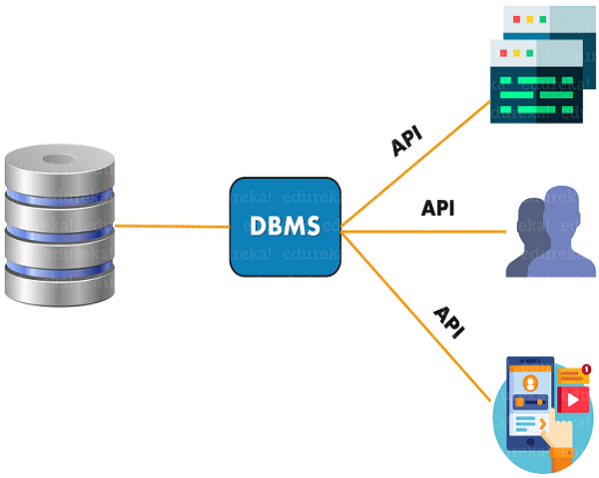
### DBMS

- **DBMS** stands for **Database Management System**.
- We can break it like this DBMS = Database + Management System.
- Database is a collection of data and Management System is a set of programs to store and retrieve those data.
- DBMS is a software to create and manage databases.
- For accessing data from a database we need DBMS.

- DBMS is a collection of inter-related data

Edit with WPS Office

and set of programs to store & access those data in an easy and effective manner.




## III. Characteristics / Advantages of DBMS

- **No redundant data**: Data redundancy refers to the duplication of data.Redundancy removed by data normalization. No data duplication saves storage and improves access time.
- **Data Consistency and Integrity**: If there is no data redundancy, then data consistency will be acquired automatically.
- **Data Security**: It is easier to apply access constraints in database systems so that only authorized user is able to access the data. Each user has a different set of access thus data is secured from the issues such as identity theft, data leaks and misuse of data.
- **Privacy**: Limited access means privacy of data.
- **Easy access to data** – Database systems manages data in such a way so that the data is easily accessible with fast response times.
- **Easy recovery**: Since database systems keeps the backup of data, it is easier to do a full recovery of data in case of a failure.
- **Flexible**: Database systems are more flexible than file processing systems.
- **ACID Properties** – DBMS follows the concepts of **A**tomicity, **C**onsistency, **I**solation, and **D**urability (normally shortened as ACID). These concepts are applied on transactions, which manipulate data in a database. ACID properties help the database stay healthy in multi-transactional environments and in case of failure.


## V. Disadvantages of DBMS:

- DBMS implementation cost is high .
- **Complexity**: Database systems are complex to understand
- **Performance**: Database systems may not supports features of some applications.

## V. Database Applications

- **Telecom**: There is a database to keeps track of the information regarding calls made, network usage, customer details etc. Without the database systems it is hard to maintain that huge amount of data that keeps updating every millisecond.
- **Industry**: Where it is a manufacturing unit, distribution centre, each one needs a database to keep the records of sales.

For example distribution centre should keep a track of the product units that supplied into the centre as well as the products that got delivered out from the

Edit with WPS Office

distribution centre on each day. For this a DBMS can be used.

- **Banking System**: For storing customer info, tracking day to day credit and debit transactions, generating bank statements etc. All this work has been done with the help of Database management systems.
- **Sales**: To store customer information, production information and invoice details.

- **Airlines**: To travel though airlines, we make early reservations, this reservation information along with flight schedule is stored in database.
- **Education sector**: Database systems are frequently used in schools and colleges to store and retrieve the data regarding student details, staff details, course details, exam details, payroll data, attendance details, fees details etc. There is a huge amount of inter-related data that needs to be stored and retrieved in an efficient manner.
- **Online shopping**: The online shopping websites such as Amazon, Flipkart etc... store the product information, your addresses and preferences, credit details and provide you the relevant list of products based on your query. All this involves a Database management system.

## VI. Different Types of Database Users in DBMS

1. **Application Programmers**

As its name shows, application programmers are the one who writes application programs that uses the database. These application programs are written in programming languages like **COBOL**. These programs meet the user requirement and made according to user requirements. Retrieving information, creating new information and changing existing information is done by these application programs.

2. **End Users**

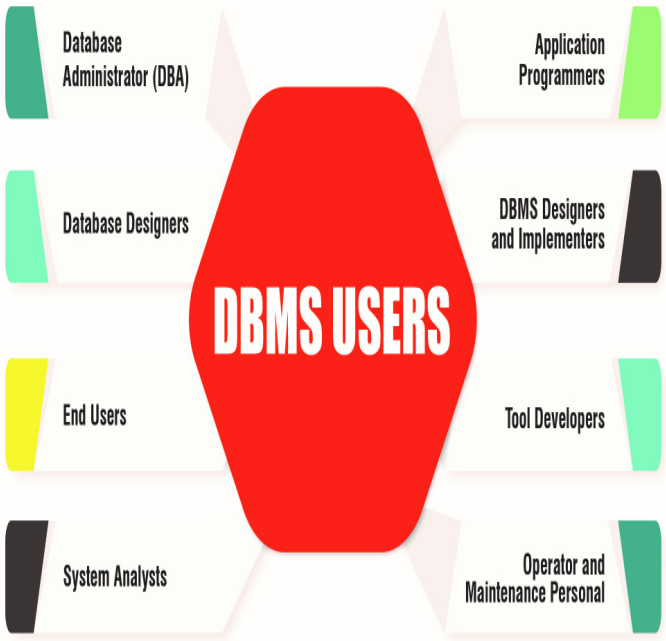
End users are those who access the database from the terminal end. They use the developed applications and they don't have any knowledge about the design and working of database. These are the second class of users and their main aim is just to get their task done.

3. **DBA (Database Administrator)**

DBA can be a single person or it can be a group of person. Database Administrator is responsible for everything that is related to database. He makes the policies, strategies and provides technical supports.

4. **System Analyst**

System analyst is responsible for the design, structure and properties of database. All the requirements of the end users are handled by system analyst. Feasibility, economic and technical aspects of DBMS is the main concern of system analyst.



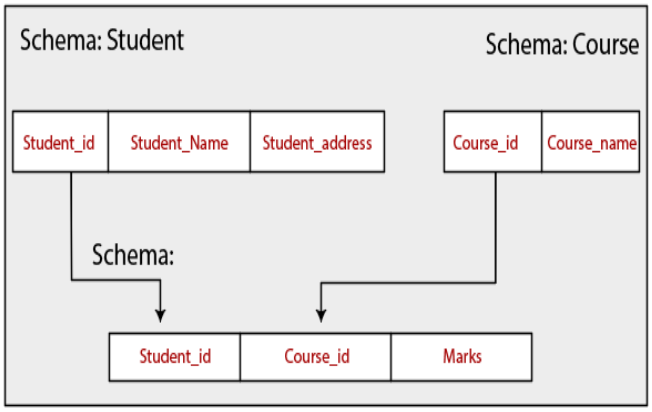

## VII. SCHEMA ,SUBSCHEMA AND INSTANCE

- **Schema in DBMS**

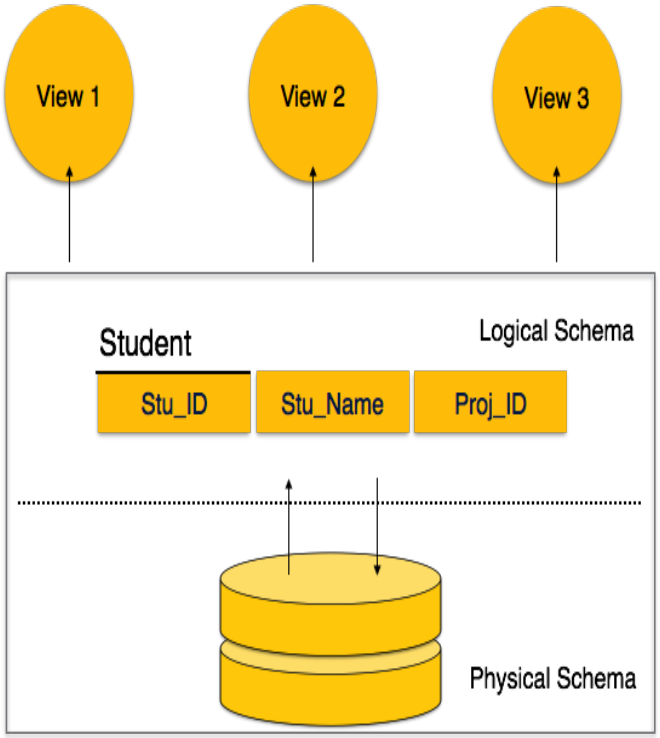
- A database schema is the skeleton structure of database.
- It can also be defined as the design of a database.
- It defines how the data is organized and how the relations among them are associated.
- Physical arrangement of data as it appears

Edit with WPS Office

in database can be defined as schema.




- It can be categorized into three parts. These are:
  a. Physical Schema
  b. Logical Schema
  c. View Schema




A database schema can be divided broadly into two categories –

a. **Physical Database Schema** – A physical schema can be defined as the design of a database at its physical level. In this level, it is expressed how data is stored in blocks of storage.
b. **Logical Database Schema** – A logical schema can be defined as the design of the database at its logical level. In this level, the programmers, as well as the database administrator (DBA), work. At this level, data can be described as certain types of data records that can be stored in the form of data structures.
c. **View Schema** :- It can be defined as the design of the database at the view level, which generally describes end-user interaction with database systems.

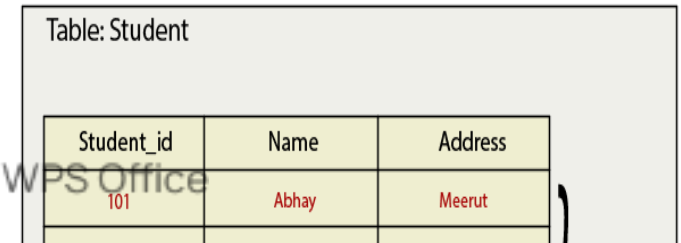
- ## Sub schema in DBMS

- It can be defined as the subset or sub-level of schema.
- That has the same properties as the schema.
- This type of schema allows a user to view only that part of the database in which user is interested.
- The logical view of data as it appears to the application can be called as sub schema.

- ## Instances in DBMS

In simple words, it is the snapshot of the database taken at a particular moment. It can also be described as the collection of the information stored in the database at that particular moment.

Every time we update the DB say we insert, delete or modify the value of the data item in the record, it changes from one state to other. At the given time, each schema has its own set of instances.

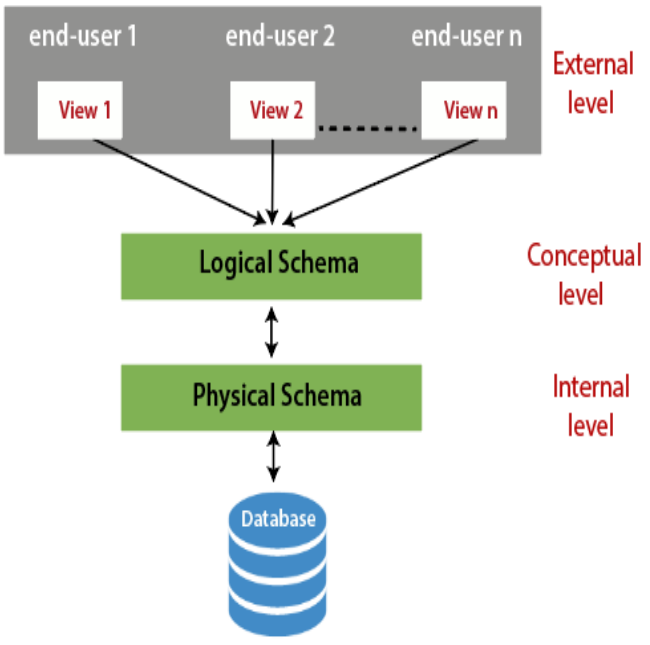



Edit with WPS Office

## VIII. THREE SCHEMA ARCHITECTURE

- The three schema architecture divides the database into three-level to create a separation between the physical database and the user application.
- In simple words, this architecture hides the details of physical storage from the user.
- The database administrator (DBA) should be able to change the structure of database storage without affecting the user's view.
- This architecture contains three layers or levels of the database management system:

1. External level
2. Conceptual level
3. Internal level




DBMS Architecture

**1. External or View level:** This is the highest level of database abstraction. External or view level describes the actual view of data that is relevant to the particular user. This level also provides different views of the same database for a specific user or a group of users. An external view provides a powerful and flexible security mechanism by hiding the parts of the database from a particular user.

**2. Conceptual or Logical level:** The conceptual level describes the structure of the whole database. This level acts as a middle layer between the physical storage and user view. It explains what data to be stored in the database, what relationship exists among those data, and what the datatypes are. There is only one conceptual schema per database.
Database administrator and the programmers work at this level. This level does not provide any access or storage details but concentrates on the relational model of the database. The conceptual schema also includes features that specify the integrity and consistency.

**3. Internal or Physical level:** This is the lowest level of database abstraction. It describes how the data is actually stored in the database and provides methods to access data from the database. It allows viewing the physical representation of the database on the computer system.

❖ **Advantages of Three-schema Architecture:**

- This architecture makes the database abstract.
- It hides the details of how data is physically stored in a computer system, which makes it easier to use for a user.
- This architecture allows each user to access the same database with a different customized view of data.
- This architecture enables a database admin to change the storage structure of the database without affecting the user currently on the system.

## X. DATA INDEPENDENCE

Edit with WPS Office

- Data Independence is defined as a property of DBMS that helps you to change the Database schema at one level of a database system without requiring to change the schema at the next higher level.
- Data independence helps you to keep data separated from all programs that make use of it.
- In many systems, data independence is an essential function for components of the system.
- There are two types of data independence,

a. Physical data independence
b. Logical data independence.

### a. Physical Data Independence

- Physical data independence helps you to separate conceptual levels from the internal/physical levels.
- It allows you to provide a logical description of the database without the need to specify physical structures.
- Compared to Logical Independence, it is easy to achieve physical data independence.
- With Physical independence, you can easily change the physical storage structures or devices with an effect on the conceptual schema.
- Any change done would be absorbed by the mapping between the conceptual and internal levels.
- Physical data independence is achieved by the presence of the internal level of the database and then the transformation from the conceptual level of the database to the internal level.

Due to Physical independence, any of the below change will not affect the conceptual layer.

- Using a new storage device like Hard Drive or Magnetic Tapes
- Modifying the file organization technique in the Database
- Switching to different data structures.
- Changing the access method.
- Modifying indexes.

### b. Logical Data Independence

- Logical Data Independence is the ability to change the conceptual scheme without changing

1. External views
2. External API or programs

- Any change made will be absorbed by the mapping between external and conceptual levels.
- When compared to Physical Data independence, it is challenging to achieve logical data independence.

Due to Logical independence, any of the below change will not affect the external layer.

1. Add/Modify/Delete a new attribute, entity or relationship is possible without a rewrite of existing application programs
2. Merging two records into one
3. Breaking an existing record into two or more records

| Logical Data Independence | Physical Data Independence |
|---|---|
| Logical Data Independence is mainly concerned with the structure or changing the data definition. | Mainly concerned with the storage of the data. |
| It is difficult as the retrieving of data is mainly dependent on the logical structure of data. | It is easy to retrieve. |
| Compared to Logic Physical independence it is difficult to achieve logical data independence. | Compared to Logical Independence it is easy to achieve physical data independence. |

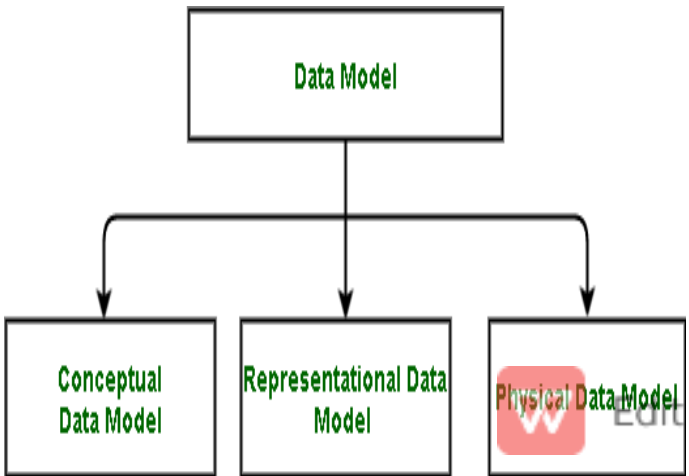
Edit with WPS Office

| | |
|---|---|
| You need to make changes in the Application program if new fields are added or deleted from the database. | A change in the physical level usually does not need change at the Application program level. |
| Modification at the logical levels is significant whenever the logical structures of the database are changed. | Modifications made at the internal levels may or may not be needed to improve the performance of the structure. |
| Concerned with conceptual schema | Concerned with internal schema |
| Example: Add/Modify/Delete a new attribute | Example: change in compression techniques, hashing algorithms, storage devices, etc |

## X. DATA MODELS

- Data models define how the logical structure of a database is modeled.
- Data models developed to summarize the description of the database.
- Data models define how data is connected to each other and how they are processed and stored inside the system.
- The very first data model could be flat data-models, where all the data used are to be kept in the same plane.
- Earlier data models were not so scientific.

### DATA MODEL CLASSIFICATION




1. **Conceptual Data Model :**
   Conceptual data model, describes the database at a very high level and is useful to understand the needs of the database. This model used in the requirement gathering process i.e., before the Database Designers start making a particular database. One such popular model is the entity/relationship model (ER model).

2. **Representational Data Model :**
   This type of data model is used to represent only the logical part of the database and does not represent the physical structure of the databases. The representational data model allows us to focus primarily, on the design part of the database. A popular representational model is Relational model.

3. **Physical Data Model :**
   All data in a database is stored physically on a secondary storage device such as discs and tapes. This is stored in the form of files, records and certain other data structures. It has all the information of the format in which the files are present and the structure of the databases and their relation to each other.

### Advantages of Data model:

- The data model is detailed enough to be used for building the physical database.
- The information in the data model can be used for defining the relationship between tables, primary and foreign keys, and stored procedures.
- Data Model helps business to communicate within and across organizations.
- Data model helps to document data mappings .
- Help to recognize correct sources of data .

### Disadvantages of Data model:

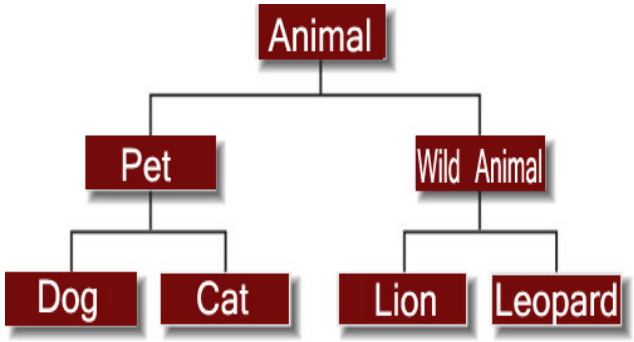
Edit with WPS Office

- To develop Data model one should know characteristics of physical data stored.
- Complex to design.
- It requires a knowledge of the biographical truth.
- Even smaller change made in structure require modification in the entire application.

## Types of Data Models

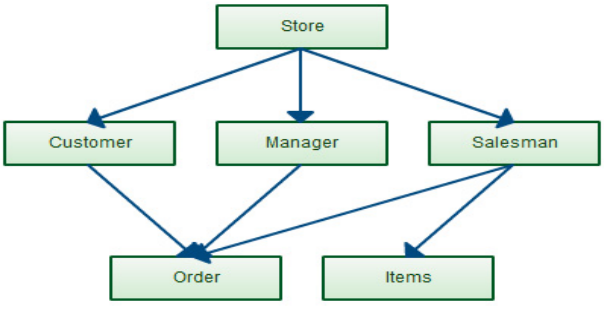
### 1. Hierarchical Data Model

Hierarchical model has one parent entity with several children entity but at the top we should have only one entity called root. For example, department is the parent entity called root and it has several children entities like students, professors and many more.



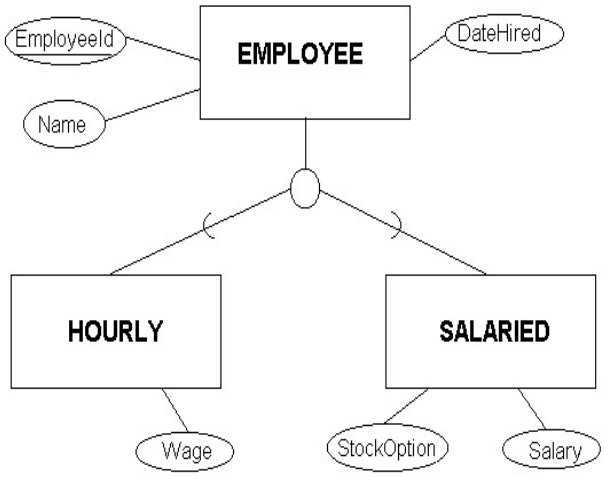

### 2. Network Data Model

Network model has the entities which are organized in a graphical representation and some entities in the graph can be accessed through several paths.




### 3. Relational Data Model

Relational model is the most extensively used model. In this model the data can be stored in the tables and this storing is called as relation, the relations can be normalized and the normalized relation values are called atomic values. Each row in a relation contains unique value and it is called as tuple, each column contains value from same domain and it is called as attribute.




## XI. DBMS Language

- A DBMS has appropriate languages and interfaces to express database queries and updates.
- Database languages can be used to read, store and update the data in the database.

## Types of Database Language

### 1. Data Definition Language

- DDL stands for Data Definition Language. It is used to define database structure or pattern.
- It is used to create schema, tables, indexes, constraints, etc. in the database.
- Using the DDL statements, you can create the skeleton of the database.
- Which deals with database schemas and descriptions, of how the data should reside in the database.

## DDL COMMANDS

- CREATE - to create a database and its

Edit with WPS Office

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

### 2. Data Manipulation Language

**DML** stands for **D**ata **M**anipulation **L**anguage. It is used for accessing and manipulating data in a database. It handles user requests.

Here are some tasks that come under DML:

- **Select:** It is used to retrieve data from a database.
- **Insert:** It is used to insert data into a table.
- **Update:** It is used to update existing data within a table.
- **Delete:** It is used to delete all records from a table.
- **Merge:** It performs UPSERT operation, i.e., insert or update operations.
- **Call:** It is used to call a structured query language or a Java subprogram.
- **Explain Plan:** It has the parameter of explaining data.
- **Lock Table:** It controls concurrency.

### 3. Data Control Language

- **DCL** stands for **D**ata **C**ontrol **L**anguage. It is used to retrieve the stored or saved data.
- The DCL execution is transactional. It also has rollback parameters.

Here are some tasks that come under DCL:

- **Grant:** It is used to give user access privileges to a database.
- **Revoke:** It is used to take back permissions from the user.

There are the following operations which have the authorization of Revoke:

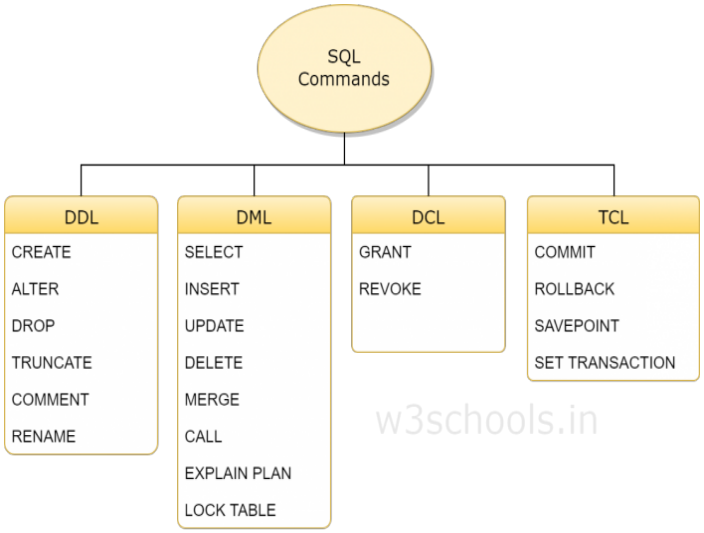
CONNECT, INSERT, USAGE, EXECUTE, DELETE, UPDATE and SELECT.

### 4. Transaction Control Language

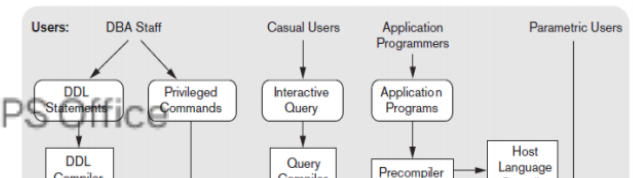
TCL is used to run the changes made by the DML statement. TCL can be grouped into a logical transaction.

Here are some tasks that come under TCL:

- **Commit:** It is used to save the transaction on the database.
- **Rollback:** It is used to restore the database to original since the last Commit.
- **SAVEPOINT** - to rollback the transaction making points within groups
- **SET TRANSACTION** - specify characteristics of the transaction




## III. Component Modules of DBMS




Edit with WPS Office

The figure is divided into two halve.

- The top half of the diagram refers to the various users of the database environment and their interfaces.

- The lower half demonstrates the internals of the DBMS responsible for storage of data and processing of transaction.

The database and the DBMS catalogue are usually stored on disk. Access to the disk is principally controlled by operating system (OS). This includes disk input/Output. A higher point stored data manager module of DBMS controls access to DBMS information that is stored on the disk.

If we consider the top half of the figure it shows interface to casual users, DBA staff, application programmers and parametric users.

The DDL compiler specified in the DDL, processes schema definitions as well as stores the description of the schema in the DBMS Catalogue. The catalogue includes information such as names and sizes of the files and data types of data items.

Storage particulars of every file mapping information among schemas as well as constraints.

Casual users as well as persons with occasional need of information from database interact using some of interface which is interactive query interface. The queries are parsed analyse for correctness of the operations for the model. The names of the data elements as well as therefore on by a query compiler that compiles them into internal form.

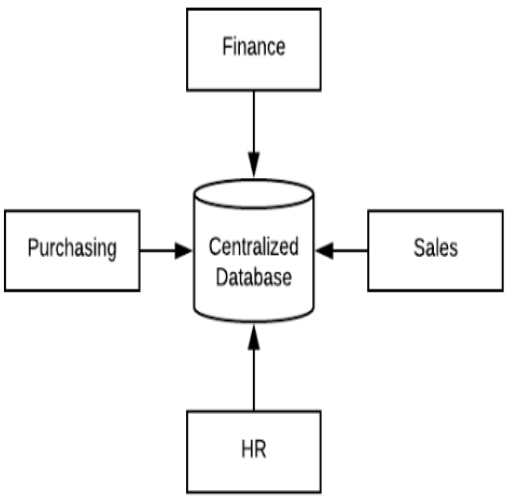
The interior query is subjected to query optimization. The query optimizer is worried with rearrangement and possible recording of operations and eliminations of redundancies.

Application programmer inscribes programs in host languages. The precompiled take out DML commands from an application program.

## III. Centralized Database Systems

A centralized database is stored at a single location such as a mainframe computer. It is maintained and modified from that location only and usually accessed using an internet connection such as a LAN or WAN. The centralized database is used by organisations such as colleges, companies, banks etc.

All the information for the organisation is stored in a single database. This database is known as the centralized database.




Edit with WPS Office

### Advantages

- The data integrity is maximised as the whole database is stored at a single physical location.
- It is easier to coordinate the data and it is as accurate and consistent as possible.
- The data redundancy is minimal in the centralised database.
- All the data is stored together and not scattered across different locations.
- Stronger security measures.
- Data is easily portable because it is stored at the same place.
- The centralized database is cheaper than other types of databases.
- It requires less power and maintenance.
- All the information in the centralized database can be easily accessed.

### Disadvantages

- Since all the data is at one location, it takes more time to search and access it.
- If the network is slow search takes even more time.
- There is a lot of data access traffic for the centralized database.
- Since all the data is at the same location, if multiple users try to access it simultaneously it creates a problem. This may reduce the efficiency of the system.
- If there are no database recovery measures in place and a system failure occurs, then all the data in the database will be destroyed.

## XIV. Client - Server Database Systems

Client server database consists of two logical components.

- Client
- Server

Clients are those who send the request to perform a specific task to the server. Servers are normally receive the command sent by the clients, perform the task and send the appropriate result back to the client.

**Example** of client is PC where as the server is a large work station.

The Client machine runs own copy of an operating system. It runs one or more applications through client's CPU and memory.

But server runs a database management system which manages the whole database.

Edit with WPS Office

### Client system vs Server system

A client system is that who sends the request to the server system and the server system has to response the request as result.

A client system is managed by users where as the server system is managed by a computer expert.

No need of a computer expert in client system. The client system is a system on which the results are displayed whereas the server system is a system in which the information is prepared for the client.

### Types of client server architecture

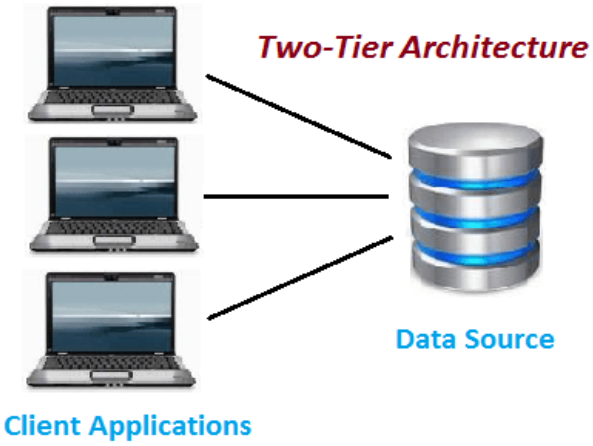
There are three types of client server architecture available in database management system.

(a) Single tier client server computing model
(b) Two tier client server computing model
(c) Three tier client server computing model

### a) Single Tier Achitecture

- Single tier architecture is the first type of client server computing model.
- In single tier client server computing model, the client server database system used on a personal computer.
- In single tier system, the database is centralized, which means the DBMS software and the data in one location and the dumb terminals were used to access the database management system.

b) Two Tier Architecture



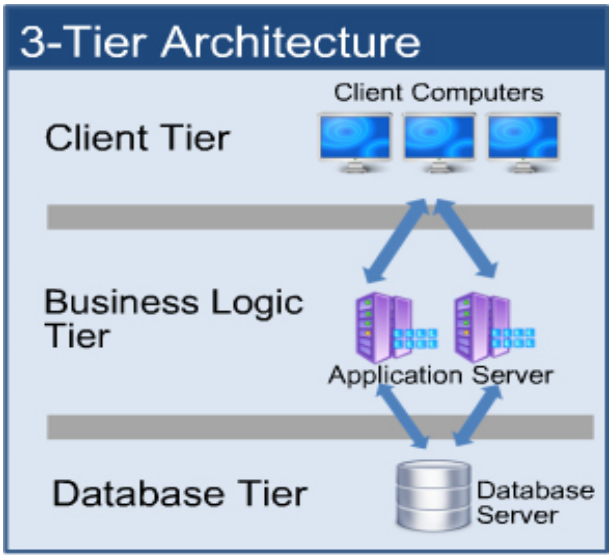

- Two-tier client server architecture is the second type of client server computing model.
- It uses two computers and one client/server system.
- At the early stages, client server computing it was called two-tier computing model in which client is considered as data capture and validation tier and server was considered as data storage tier.

Problems in two tier architecture

In 1990s, when applications were more complex and used with hundreds or thousands of end users, the two tier server model could not meet the need of the enterprise.

- An efficient client requiring considerable resources on client's computer to run effectively. This includes disk space, RAM, CPU.
  - Client machines require administration which results overhead at the server end.

c) Three Tier Architecture




- Three tier architecture is the improvement over two tier architecture.
- Three tier architecture has three layers.
- The first layer is the user interface which runs on client system.
- The second layer is called the application server . It is used for business logic and data

Edit with WPS Office

processing.

- The third layer is known as database server. It is a database management system which stores the data as needed by the middle layer.

♦ **Advantages of client server database system**

(a) Client server system has cost effective platforms to support the applications.

(b) Client offers graphical menu driven interface, which is better than traditional command line.

(c) Client server environment facilitates in more productive work by the users and efficient use of existing data.

(d) Client server database system is more flexible.

(e) A single database on server can be shared across several distinct client systems.

♦ **Disadvantages of client server database system**

(a) In initial phases, the programming cost is high.

(b) There is a lack of management tools for performance monitoring and timing.

## XV. Classification Of Database Management System

**Based on the data model**

- Relational database
- Hierarchical database
- Object oriented database
- Object relational database

**Based on the number of users**

- Single user
- Multiple users

**Based on the sites over which network is distributed**

- Centralized database system
- Distributed database system
- Client-server database system

**Based on the cost**

- Low cost DBMS
- Medium cost DBMS
- High cost DBMS

**Based on the access**

- Sequential access
- Direct access
- Inverted file structures

Edit with WPS Office