INTRODUCTION TO DBMS

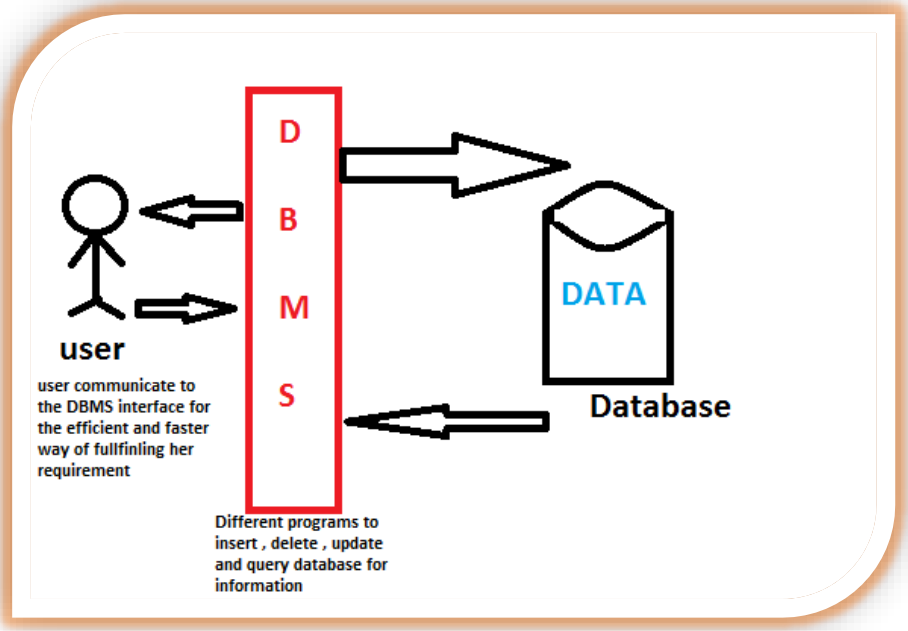
❖ Data

- **Data is a collection of a distinct small unit of information. It can be used in a variety of forms like text, numbers, media, bytes, etc. it can be stored in pieces of paper or electronic memory, etc.**
- Word 'Data' is originated from the word 'datum' that means 'single piece of information.' It is plural of the word datum.
- In computing, Data is information that can be translated into a form for efficient movement and processing. Data is interchangeable.

❖ Database

- A **database** is an organized collection of data, so that it can be easily accessed and managed.
- You can organize data into tables, rows, columns, and index it to make it easier to find relevant information.
- **Database handlers** create a database in such a way that only one set of software program provides access of data to all the users.
- The **main purpose** of the database is to operate a large amount of information by storing, retrieving, and managing data.
- There are many **dynamic websites** on the World Wide Web nowadays which are handled through databases. For example, a model that checks the availability of rooms in a hotel. It is an example of a dynamic website that uses a database.

- There are many **databases available** like MySQL, Sybase, Oracle, MongoDB, Informix, PostgreSQL, SQL Server, etc.
- Modern databases are managed by the database management system (DBMS).
- **SQL** or Structured Query Language is used to operate on the data stored in a database. SQL depends on relational algebra and tuple relational calculus.
- A cylindrical structure is used to display the image of a database.



❖ DBMS (Data Base Management System)

Database management System is software which is used to store and retrieve the database. For example, Oracle, MySQL, etc.; these are some popular DBMS tools.

- DBMS provides the interface to perform the various operations like creation, deletion, modification, etc.

- DBMS allows the user to create their databases as per their requirement.
- DBMS accepts the request from the application and provides specific data through the operating system.
- DBMS contains the group of programs which acts according to the user instruction.
- It provides security to the database.

➢ Advantage of DBMS

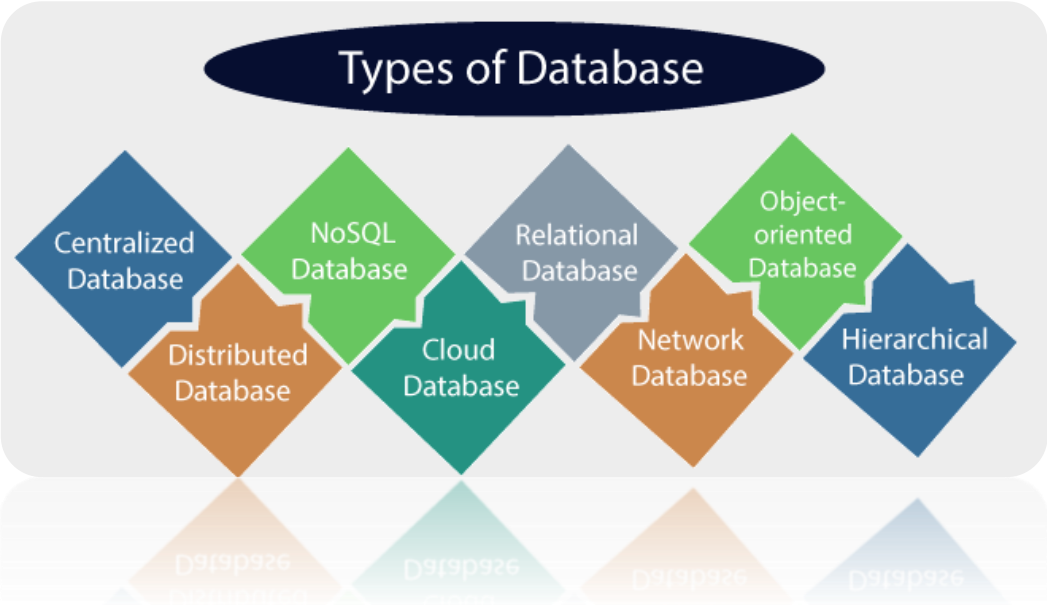
- **Controls redundancy**:It stores all the data in a single database file, so it can control data redundancy.
- **Data sharing**:An authorized user can share the data among multiple users.
- **Backup**:It providesBackup and recovery subsystem. This recovery system creates automatic data from system failure and restores data if required.
- **Multiple user interfaces**:It provides a different type of user interfaces like GUI, application interfaces.

➢ Disadvantage of DBMS

- **Size**:It occupies large disk space and large memory to run efficiently.
- **Cost**:DBMS requires a high-speed data processor and larger memory to run DBMS software, so it is costly.
- **Complexity**:DBMS creates additional complexity and requirements.

❖ Types of Databases

There are various types of databases used for storing different varieties of data:



1) Centralized Database

- It is the type of database that stores data at a centralized database system.
- It comforts the users to access the stored data from different locations through several applications.
- These applications contain the authentication process to let users access data securely.
- An example of a Centralized database can be Central Library that carries a central database of each library in a college/university.

Advantages of Centralized Database

- ✓ It has decreased the risk of data management, i.e., manipulation of data will not affect the core data.
- ✓ Data consistency is maintained as it manages data in a central repository.
- ✓ It provides better data quality, which enables organizations to establish data standards.
- ✓ It is less costly because fewer vendors are required to handle the data sets.

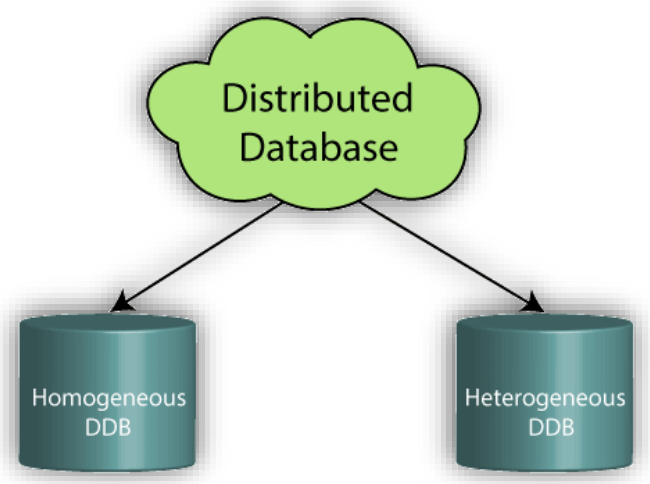
Disadvantages of Centralized Database

- The size of the centralized database is large, which increases the response time for fetching the data.
- It is not easy to update such an extensive database system.
- If any server failure occurs, entire data will be lost, which could be a huge loss.

2) Distributed Database

- Unlike a centralized database system, in distributed systems, data is distributed among different database systems of an organization.
- These database systems are connected via communication links. Such links help the end-users to access the data easily.
- **Examples** of the Distributed database are Apache Cassandra, HBase, Ignite, etc.

We can further divide a distributed database system into:



1. **Homogeneous DDB:** Those database systems which execute on the same operating system and use the same application process and carry the same hardware devices.
2. **Heterogeneous DDB:** Those database systems which execute on different operating systems under different application procedures, and carries different hardware devices.

Advantages of Distributed Database

- ✓ Modular development is possible in a distributed database, i.e., the system can be expanded by including new computers and connecting them to the distributed system.
- ✓ One server failure will not affect the entire data set.

3) Relational Database

- This database is based on the relational data model, which stores data in the form of rows(tuple) and columns(attributes), and together forms a table(relation).
- A relational database uses SQL for storing, manipulating, as well as maintaining the data. E.F. Codd invented the database in 1970.
- Each table in the database carries a key that makes the data unique from others.
- **Examples** of Relational databases are MySQL, Microsoft SQL Server, Oracle, etc.

Properties of Relational Database

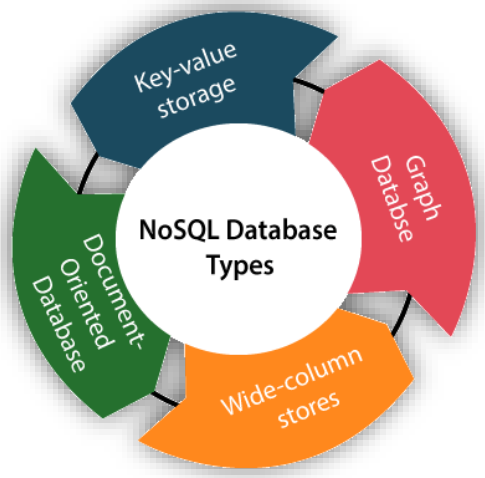
There are following four commonly known properties of a relational model known as ACID properties, where:

- **A means Atomicity:** This ensures the data operation will complete either with success or with failure. It follows the 'all or nothing' strategy. For example, a transaction will either be committed or will abort.
- **C means Consistency:** If we perform any operation over the data, its value before and after the operation should be preserved. For example, the account balance before and after the transaction should be correct, i.e., it should remain conserved.
- **I means Isolation:** There can be concurrent users for accessing data at the same time from the database. Thus, isolation between the data should remain isolated. For example, when multiple transactions occur at the same time, one transaction effects should not be visible to the other transactions in the database.

- **D means Durability:** It ensures that once it completes the operation and commits the data, data changes should remain permanent.

4) NoSQL Database

- Non-SQL/Not Only SQL is a type of database that is used for storing a wide range of data sets.
- It is not a relational database as it stores data not only in tabular form but in several different ways.
- It came into existence when the demand for building modern applications increased. Thus, NoSQL presented a wide variety of database technologies in response to the demands.
- We can further divide a NoSQL database into the following four types:



1. **Key-value storage:** It is the simplest type of database storage where it stores every single item as a key (or attribute name) holding its value, together.

2. **Document-oriented Database:** A type of database used to store data as JSON-like document. It helps developers in storing data by using the same document-model format as used in the application code.
3. **Graph Databases:** It is used for storing vast amounts of data in a graph-like structure. Most commonly, social networking websites use the graph database.
4. **Wide-column stores:** It is similar to the data represented in relational databases. Here, data is stored in large columns together, instead of storing in rows.

Advantages of NoSQL Database

- ✓ It enables good productivity in the application development as it is not required to store data in a structured format.
- ✓ It is a better option for managing and handling large data sets.
- ✓ It provides high scalability.
- ✓ Users can quickly access data from the database through key-value.

5) Cloud Database

- A type of database where data is stored in a virtual environment and executes over the cloud computing platform.
- It provides users with various cloud computing services (SaaS, PaaS, IaaS, etc.) for accessing the database.
- There are numerous cloud platforms, but the best options are:

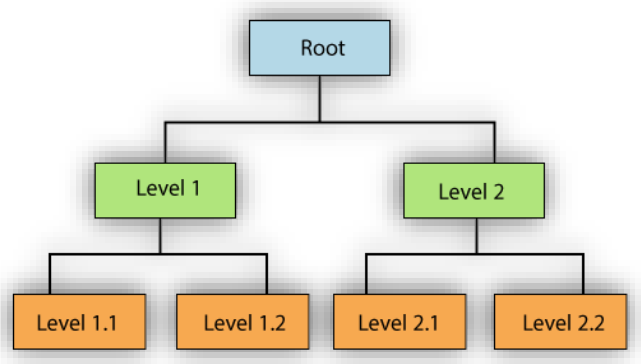
- ✓ Amazon Web Services(AWS)
- ✓ Microsoft Azure
- ✓ Kamatera
- ✓ PhonixNAP
- ✓ ScienceSoft
- ✓ Google Cloud SQL, etc.

6) Object-oriented Databases

- The type of database that uses the object-based data model approach for storing data in the database system.
- The data is represented and stored as objects which are similar to the objects used in the object-oriented programming language.

7) Hierarchical Databases

- It is the type of database that stores data in the form of parent-children relationship nodes.
- Here, it organizes data in a tree-like structure.



Hierarchical Database

- Data get stored in the form of records that are connected via links. Each child record in the tree will contain only one parent. On the other hand, each parent record can have multiple child records.

8) Network Databases

- It is the database that typically follows the network data model.
- Here, the representation of data is in the form of nodes connected via links between them.
- Unlike the hierarchical database, it allows each record to have multiple children and parent nodes to form a generalized graph structure.

9) Personal Database

- Collecting and storing data on the user's system defines a Personal Database.
- This database is basically designed for a single user.

Advantage of Personal Database

- ✓ It is simple and easy to handle.
- ✓ It occupies less storage space as it is small in size.

10) Operational Database

- The type of database which creates and updates the database in real-time.
- It is basically designed for executing and handling the daily data operations in several businesses.
- For example, An organization uses operational databases for managing per day transactions.

11) Enterprise Database

- Large organizations or enterprises use this database for managing a massive amount of data.
- It helps organizations to increase and improve their efficiency. Such a database allows simultaneous access to users.

Advantages of Enterprise Database

- ✓ Multi processes are supportable over the Enterprise database.
- ✓ It allows executing parallel queries on the system.

❖ RDBMS (Relational Database Management System)

- **RDBMS** stands for Relational Database Management System.
- All modern database management systems like SQL, MS SQL Server, IBM DB2, ORACLE, My-SQL, and Microsoft Access are based on RDBMS.
- It is called Relational Database Management System (RDBMS) because it is based on the relational model introduced by E.F. Codd.

➢ How it works

- Data is represented in terms of tuples (rows) in RDBMS.
- A relational database is the most commonly used database. It contains several tables, and each table has its primary key.
- Due to a collection of an organized set of tables, data can be accessed easily in RDBMS.

➢ Table/Relation

- Everything in a relational database is stored in the form of relations.
- The RDBMS database uses tables to store data.
- A table is a collection of related data entries and contains rows and columns to store data.
- Each table represents some real-world objects such as person, place, or event about which information is collected.
- The organized collection of data into a relational table is known as the logical view of the database.

Properties of a Relation:

- Each relation has a unique name by which it is identified in the database.
- Relation does not contain duplicate tuples.
- The tuples of a relation have no specific order.
- All attributes in a relation are atomic, i.e., each cell of a relation contains exactly one value.

A table is the simplest example of data stored in RDBMS.

Let's see the example of the student table.

ID	Name	AGE	COURSE
1	Ajeet	24	B.Tech
2	aryan	20	C.A
3	Mahesh	21	BCA
4	Ratan	22	MCA
5	Vimal	26	BSC

➢ Row or Record

A row of a table is also called a record or tuple. It contains the specific information of each entry in the table. It is a horizontal entity in the table. For example, The above table contains 5 records.

Properties of a row

- No two tuples are identical to each other in all their entries.
- All tuples of the relation have the same format and the same number of entries.
- The order of the tuple is irrelevant. They are identified by their content, not by their position.

Let's see one record/row in the table.

ID	Name	AGE	COURSE
1	Ajeet	24	B.Tech

➢ Column/attribute

A column is a vertical entity in the table which contains all information associated with a specific field in a table. For example, "name" is a column in the above table which contains all information about a student's name.

Properties of an Attribute

- Every attribute of a relation must have a name.
- Null values are permitted for the attributes.
- Default values can be specified for an attribute automatically inserted if no other value is specified for an attribute.
- Attributes that uniquely identify each tuple of a relation are the primary key.

➢ Data item/Cells

The smallest unit of data in the table is the individual data item. It is stored at the intersection of tuples and attributes.

Properties of data items

- Data items are atomic.
- The data items for an attribute should be drawn from the same domain.

In the below example, the data item in the student table consists of Ajeet, 24 and Btech, etc.

ID	Name	AGE	COURSE
1	Ajeet	24	B.Tech

➢ Degree

The total number of attributes that comprise a relation is known as the degree of the table.

➢ Cardinality

The total number of tuples at any one time in a relation is known as the table's cardinality. The relation whose cardinality is 0 is called an empty table.

➢ Domain

The domain refers to the possible values each attribute can contain. It can be specified using standard data types such as integers, floating numbers, etc.

For example, An attribute entitled Marital_Status may be limited to married or unmarried values.

➢ NULL Values

The NULL value of the table specifies that the field has been left blank during record creation.

It is different from the value filled with zero or a field that contains space.

❖ Data Integrity

There are the following categories of data integrity exist with each RDBMS:

1) **Entity integrity**: It specifies that there should be no duplicate rows in a table.
2) **Domain integrity**: It enforces valid entries for a given column by restricting the type, the format, or the range of values.
3) **Referential integrity** specifies that rows cannot be deleted, which are used by other records.
4) **User-defined integrity**: It enforces some specific business rules defined by users. These rules are different from the entity, domain, or referential integrity.

❖ Difference between DBMS and RDBMS

Although DBMS and RDBMS both are used to store information in physical database but there are some remarkable differences between them.

The main differences between DBMS and RDBMS are given below:

No.	DBMS	RDBMS
1)	DBMS applications store **data as file**.	RDBMS applications store **data in a tabular form**.
2)	In DBMS, data is generally stored in either a hierarchical form or a navigational form.	In RDBMS, the tables have an identifier called primary key and the data values are stored in the form of tables.
3)	**Normalization is not** present in DBMS.	**Normalization is** present in RDBMS.
4)	DBMS does **not apply any security** with regards to data manipulation.	RDBMS **defines the integrity constraint** for the purpose of ACID (Atomocity, Consistency, Isolation and Durability) property.
5)	DBMS uses file system to store data, so there will be **no relation between the tables**.	in RDBMS, data values are stored in the form of tables, so a **relationship** between these data values will be stored in the form of a table as well.

6)	DBMS has to provide some uniform methods to access the stored information.	RDBMS system supports a tabular structure of the data and a relationship between them to access the stored information.
7)	DBMS **does not support distributed database**.	RDBMS **supports distributed database**.
8)	DBMS is meant to be for small organization and **deal with small data**. it supports **single user**.	RDBMS is designed to **handle large amount of data**. it supports **multiple users**.
9)	Examples of DBMS are file systems, **xml** etc.	Example of RDBMS are **mysql**, **postgre**, **sql server**, **oracle** etc.