DBMS
CHAPTER-1

Prepared By :Jagadish Shrestha

Data

Data is a collection of raw, unorganised facts and details like text, observations, figures, symbols and descriptions of things etc

Data does not carry any specific purpose and has no significance by itself.

For eg.Sensor readings, survey responses.

Information

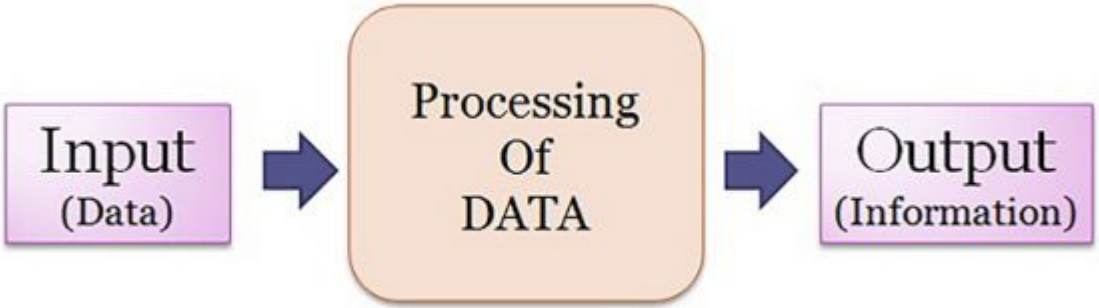
Information is processed, organised and structured data.

It provides context for data and enables decision making.

For example, when you process the data collected from survey ,research and derive some report or facts then it is called information

Data and Information

.



Database

A database is a digital repository for storing, managing and securing organized collections of data.

Database is the organized collection of related data.

Different types of databases store data in different ways. For example, relational databases store it in defined tables with rows and columns, while non relational databases can store it as a variety of data structures, including key-value pairs documents ,etc

DBMS (Database Management System)

A database management system (DBMS) is a software system for creating and managing databases.

The primary goal of a DBMS is to provide a way to store and retrieve database information that is both convenient and efficient.

A DBMS enables end users to create, protect, read, update and delete data in a database.

Eg MySql,PostgreSQL,Microsoft SQL Server,Oracle Database,etc

DBMS (Database Management System)

Database systems are designed to manage large bodies of information.

Management of data involves both defining structures for storage of information and providing mechanisms for the manipulation of information.

It is a collection of programs that allows users to specify the structure of a database, to create , query and modify the data in the database and to control access to it. (Example: limit access to the database so that only relevant staff can access details of enrolling students).

Database System

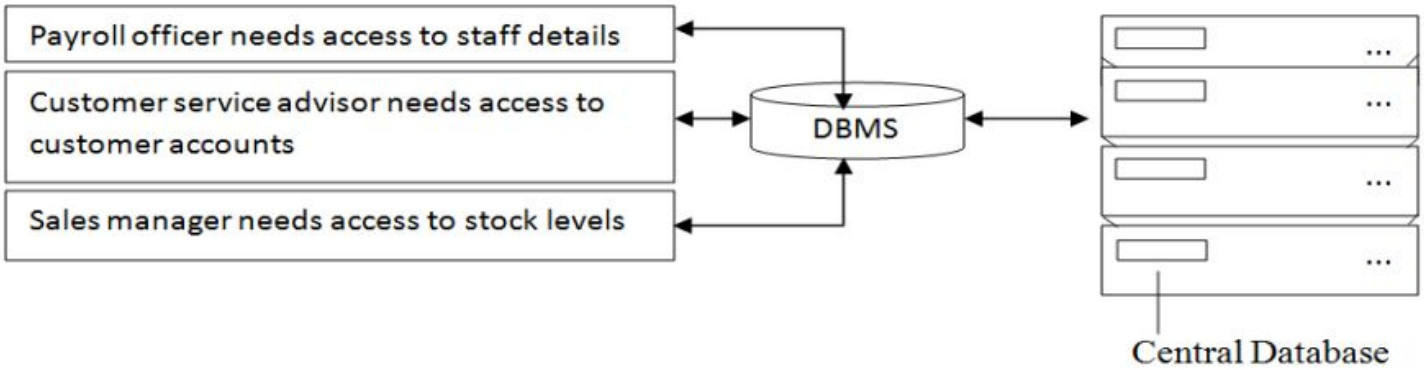
A database system refers to the combined components of a database and a database management system (DBMS).

It represents a structured collection of related data files, along with the tools and mechanisms used for their definition, int

erpretation, manipulation, and maintenance.

The database stores the actual data, while the DBMS provides the software interface to access and manage that data efficiently

Database System



Importance of Data

Support Decision Making

Scientific Research & Innovation

Enables Accurate Planning and Forecasting

Foundation for Machine Learning and AI

Detect Patterns in behaviour and Trends

Evidence

Disadvantage of file processing system

Data redundancy and inconsistency

For example a student has two majors math and computer then the details of students can contain in math and computer department records resulting in redundancy and higher storage cost .

Similarly if the address is changes in math department records but not in computer department records then this result in data inconsistency.

Data inconsistency: when the same piece of data(address of student) exists in multiple locations(departments) but does not match across those locations.

Disadvantage of file processing system

Difficulty in accessing data

File processing system does not allow to retrieve the required data in flexible manner.

For example we may need to retrieve only name,address and contact No of students from a university then programmer have to write a separate program or application to extract those data.Each time we different details ,different program may be needed or we have to go through all the unnecessary data from files and extract the name,address and contact one by one which is not feasible.

Disadvantage of file processing system

Data Isolation

Data isolation mean the data are scattered across different locations and difficult to access ,combine and analyze together.

For example .if the marks,attendance,fees paid by students are stored in different flies like marks.txt,attendance.txt and fees.txt then it becomes difficult to combine the data of students and generate some kinds of reports.

Disadvantage of file processing system

Integrity problems

In file processing system the data stored are not consistent across files

Like primary key ,foreignkey ,validations etc

For eg. A student can have id 10 in math department records and 11 in computer department records. So the data are not consistent and reliable.

Disadvantage of file processing system

Atomicity problems.

Does Not have a way to roll back to the previous state in case of failure or crashes like transaction in dbms

For eg. For eg a student paid $500 and we have to make the record the transaction in student_payment.txt file and university_payment.txt file .The transaction is considered failed if we cannot make the changes on both file.

If student_payment.txt file is updated and while updating university_payment.txt file the computer shut down then there is no property way to rollback the consistent state of student_payment.txt file.

Disadvantage of file processing system

Security Problems

Does Not have a proper mechanism to provide the access to onlime some parts of the data.

For eg. A payroll officer of university need to know only the financial part of the students and staffs. A proper measures to hide the other academic details of the data are not available if academic and the financial data are in same file

Importance Data management

Data management is creating, organizing, storing and managing the data collected and created by any organization.

Reliability- Data which is managed properly is more reliable than a random piece of data that is picked up from some storage or source.Reliable data helps the organization to make proper decision making.

Security-Proper management of data helps to make the data secure by generating timely backup and provide access to only parts of the data that are authorized to the given user

Importance of Data management

Accessibility- Data management can increase the accessibility data. As the data is properly organized, users can access it easily and quickly

Scalability- Data management allows the organization to remove the unwanted and add the new data easily and scale depending upon the need of organization

How Data are Stored in Database(Data Model)

A collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints.

There are 4 categories of Data model they are

- Relational Model
- Entity-Relationship Model
- Semi-structured Data Model.
- Object-Based Data Model
- Hierarchical Model

Relational Data Model

Data are stored in tables consisting of rows and columns

Tables are linked together using relationships

Each table corresponds to an entity set or relationship set, and each row represents an instance of that entity set or relationship set.

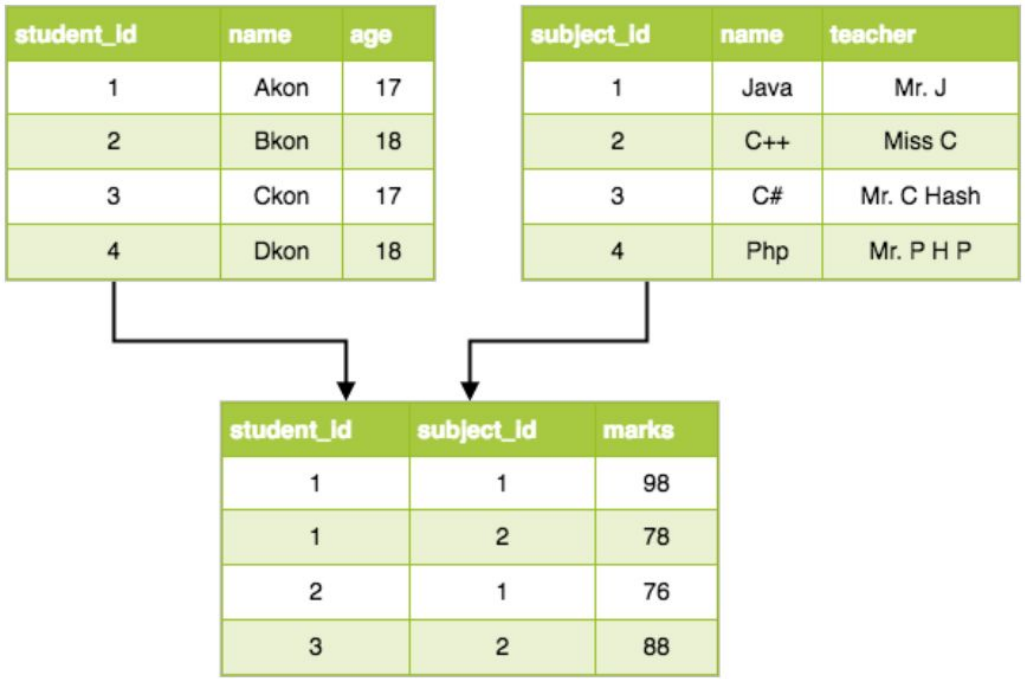
It is a lower level model that uses a collection of tables to represent both data and relationships among those data

Structured query language (SQL) is used to manipulate data stored in tables.

Most used Data model

Relational Data Model

.



student_id	name	age
1	Akon	17
2	Bkon	18
3	Ckon	17
4	Dkon	18

subject_id	name	teacher
1	Java	Mr. J
2	C++	Miss C
3	C#	Mr. C Hash
4	Php	Mr. P H P

student_id	subject_id	marks
1	1	98
1	2	78
2	1	76
3	2	88

Entity-Relationship Model

Entity-relationship (E-R) model is a high level data model based on a perception of a real world that consists of collection of basic objects, called entities, and of relationships among these entities.

An entity is a thing or object in the real world that is distinguishable from other objects.

Entities are described in a database by a set of attributes.

A relationship is an association among several entities.

Entity-Relationship Model

The entity-relationship model is widely used in database design

It define the logical structure of the database .Overall logical structure of a database can be expressed graphically by E-R diagram. The basic components of this diagram are:

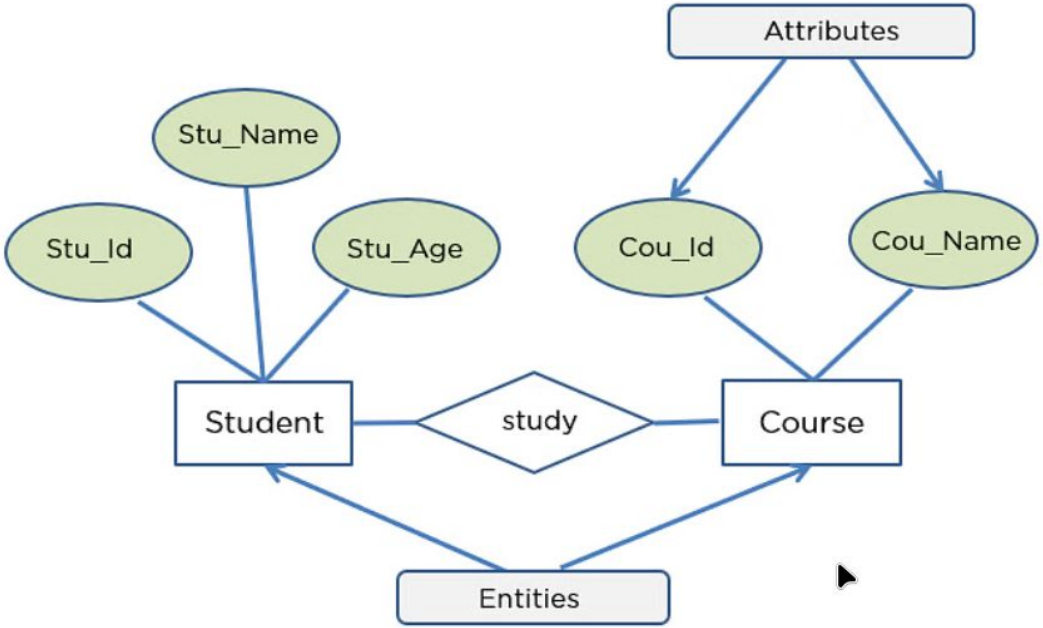
- **Rectangles** (represent entity sets)
- **Ellipses** (represent attributes)
- **Diamonds** (represent relationship sets among entity sets)
- **Lines** (link attributes to entity sets and entity sets to relationship sets)

Entity-Relationship Model

It define the logical structure of the database .Overall logical structure of a database can be expressed graphically by E-R diagram. The basic components of this diagram are:

- **Rectangles** (represent entity sets)
- **Ellipses** (represent attributes)
- **Diamonds** (represent relationship sets among entity sets)
- **Lines** (link attributes to entity sets and entity sets to relationship sets)

Entity-Relationship Model



Semi-structured Data Model.

Individual data items of the same type may have different sets of attributes.

JSON and Extensible Markup Language (XML) are widely used semi-structured data representations.

Irregular structure: Items in the same collection(eg students) may have different fields.

Semi-structured Data Model.

```
{
  "name": "John Doe",
  "age": 30,
  "email": "john.doe@example.com"
}
```

Some attributes are missing in data of same type

```
{
  "name": "Jane Smith",
  "email": "jane.smith@example.com",
  "preferences": {
    "newsletter": true,
    "smsNotifications": false
  }
}
```

Object-Based Data Model

Data are stored in the form of objects and their relationships

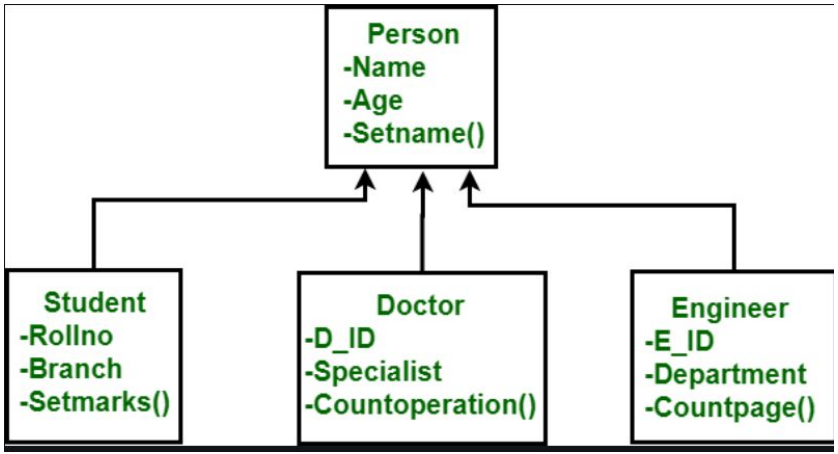
It uses the concept of object oriented programming

Easy to represent real world objects

- Components of object-Based Data Model are:
  Object:An object is an abstraction of a real world entity
- Attribute: An attribute describes the properties of object
- Method :represents the behavior of an object like eat,read .
- Class: a collection of similar objects with shared structure

Object-Based Data Model

By using concept of Inheritance(OOP),new class can inherit the attributes and methods of the old class i.e. base class. For example: as classes Student, Doctor and Engineer are inherited from the base class Person.



Hierarchical Model

Store data in a tree-like structure

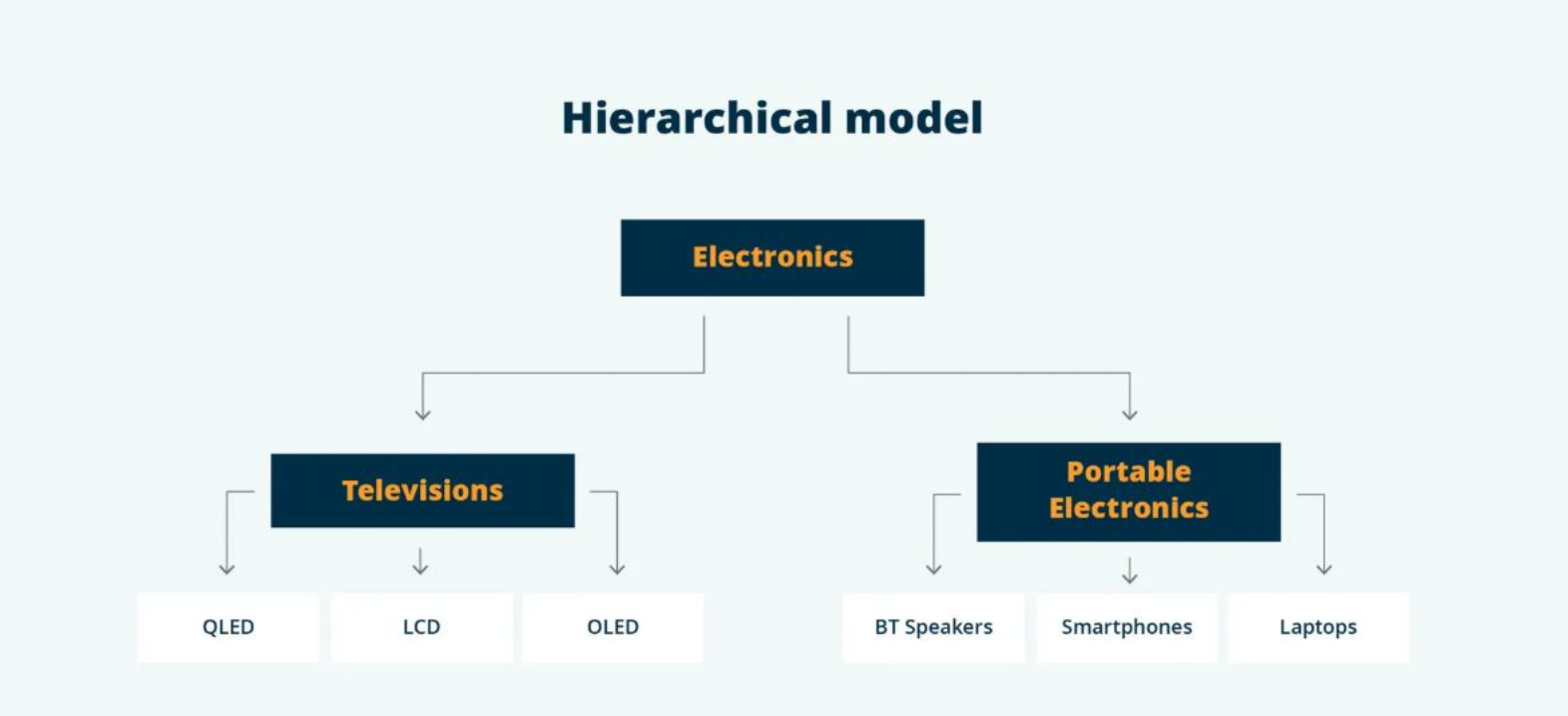
data is organized into a hierarchy, with each level representing a different category of information

where each record (called a **node**) has a single **parent**, but can have **multiple children** — similar to a file system

One to many relation.

Child cannot have many parent

Hierarchical database management system



Network Model

Designed to store, retrieve, and manage data in a networked environment.

The basic structure involves records and sets:

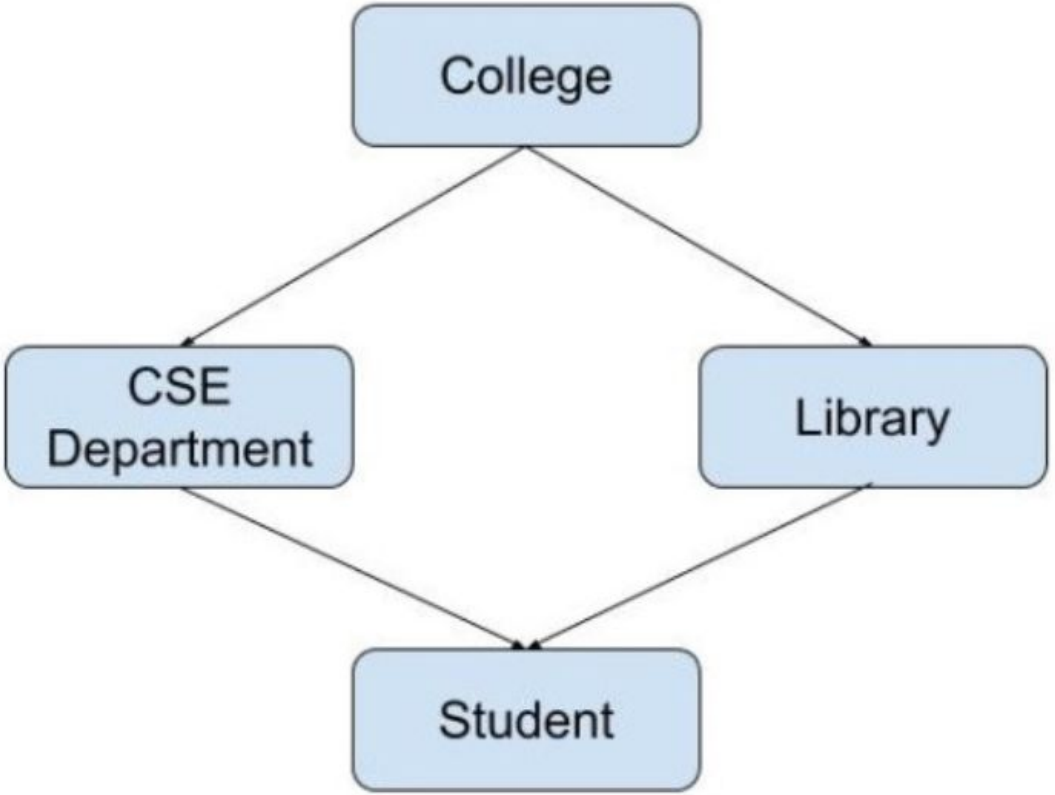
- Record: A collection of fields (similar to a row in a table) eg.student1,student2.
- Set: A one-to-many relationship between a parent record and child records (enroll,reads).

Multiple parent and child records (Many to many relation)— more flexible than hierarchical DBMS.

Each child can have multiple parents, unlike the hierarchical model, which restricts children to a single parent.

Network model

.



DBMS Architecture(Three schema Architecture)

The DBMS architecture proposed by ANSI/SPARC (American National Standards Institute, Standards Planning And Requirements Committee) (ANSI/SPARC architecture) is defined at three levels. This architecture is also called three-schema architecture.

This architecture provides t**hree levels of abstraction** to simplify users interaction with the system.

It provides users with an abstract view of data. The system hides certain details of how data are stored and maintained.

The goal of this architecture is to separate the user applications from physical database.

Data Abstraction-DBMS Architecture

Data abstraction in dbms means hiding the implementation details from the user and providing a simplified interface to interact with the data

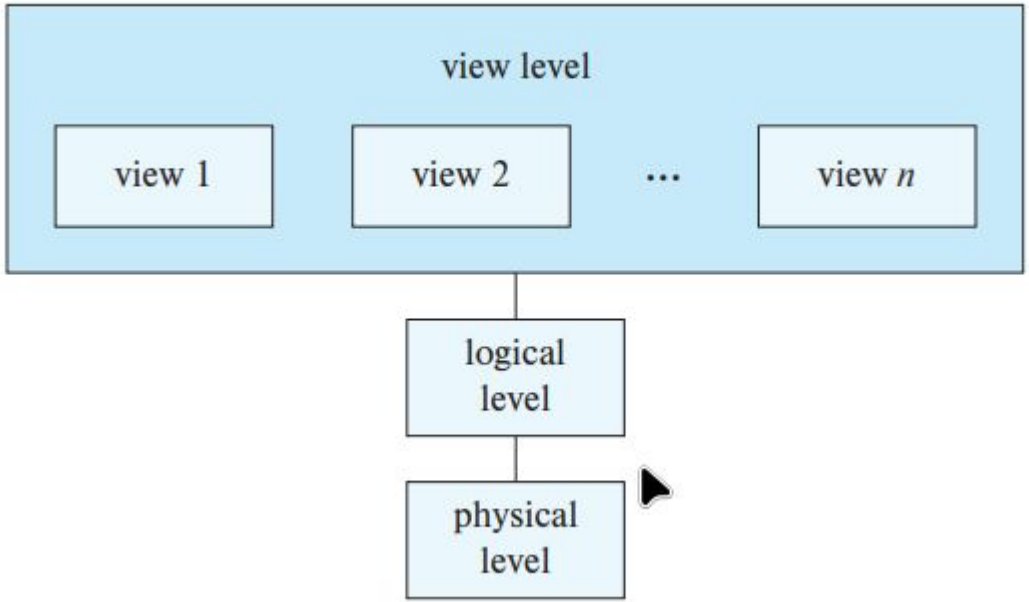
It simplifies the user interaction and provide security

There are three level of data abstraction

- Physical Level
- Logical Level
- View Level

Data Abstraction

.



Physical Level(Internal Level)

Lowest level of data abstraction

Describes how the data are actually stored using data structures like (B-tree,indexes,hashing,etc.)

User never interact with this level

Eg.Storing the binary data in a sequential memory locations

Logical Level(Conceptual Level)

Describes what data are stored in the database, and what relationships exist among those data.

Example : Defining tables, attributes, types, constraints, relationships in the schema

It is used by database administrators and developers

View Level(External Level)

Top level data abstraction

Describes only part of the entire database

Different views of the database are defined for different users so that they can access only part of the database not the entire database

Provide the security and authorization measures for the database.

Data Independence

Data Independence is the ability to change the schema at one level without having to change the schema at next higher level

There are two types of data independence:

- Logical Data independence
- Physical Data independence

Logical Data independence

The capacity to change the conceptual schema without having to change the external schemas and their associated application programs.

Ability to change the table constraints,attributes,relationships,etc without have to change the views of the database.

Physical Data Independence

The ability to change the physical storage structure or data access methods without altering the logical schema or application programs(views).

Example:Changing from HDD to SSD ,changing sequential storing mechanism to b-tree,etc

Instances and Schema

Databases change over time as information is inserted and deleted.

The collection of information stored in the database at a particular moment is called an **instance** of the database.

The overall design or structure of the database is called the **database schema**. It includes data types,constraints,relationships,etc. It is a blueprint of the databse.

Distributed Database

A distributed database is a collection of multiple interconnected databases, which are spread physically across various locations that communicate via a computer network.

Even Though they are physically separated,they are logically connected.

Types of data distribution:

- Horizontal partitioning (Sharding):
  - Rows of a table are distributed (mostly used in large datasets)
  - Joins across nodes are expensive
- Vertical partitioning (Sharding):
  - Columns are distributed (rarely used performance issues)
  - Harder to manage the query
- Table-wise Partitioning:
  - Tables are distributed (useful in microservices)
  - No facility for joins and relations between nodes

Distributed Database Management System

A distributed database management system (DDBMS) is a centralized software system that manages a distributed database in a manner as if it were all stored in a single location.

It is used to create, retrieve, update and delete distributed databases.

Example of Distributed database are Apache Cassandra, HBase, Ignite, etc.

Types of Distributed Database

Homogeneous DDB:

All the nodes of the database in distributed system runs on same database software,use same application logic ,have same operating system and similar hardware

Eg.All nodes use PostgreSQL,run on Linux and have same application logic

Heterogeneous DDB:

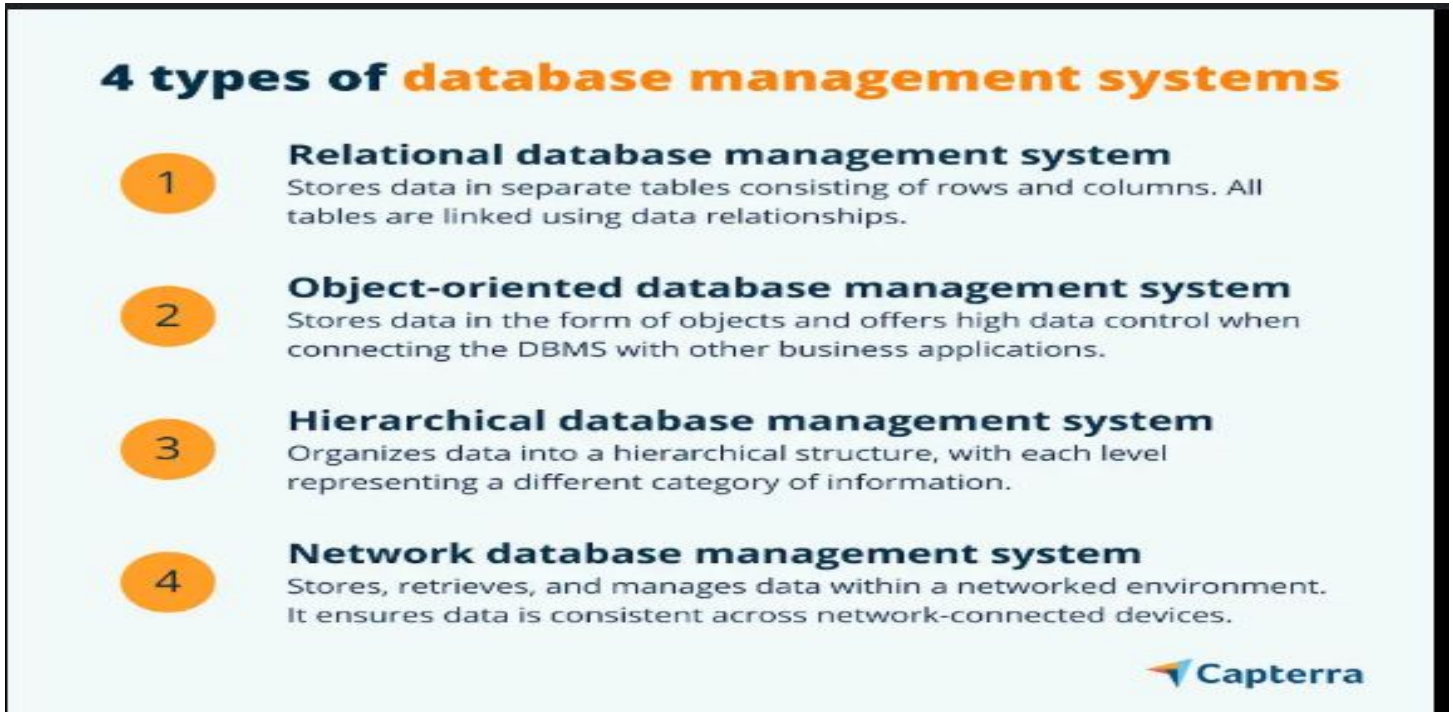
Each nodes of the database run different database software,runs on different operating system,have different application logic and different hardware.

Eg. One node running mongodb on linux,another node run postgres in macos

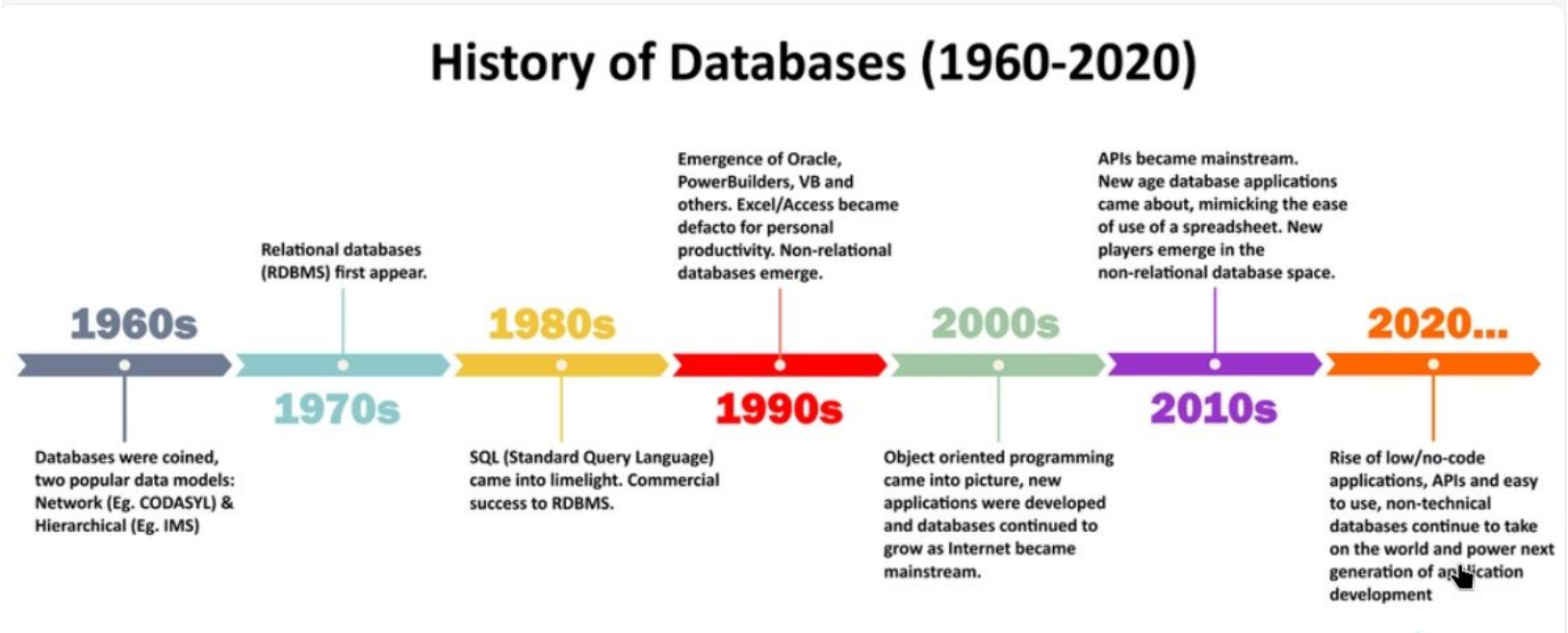
Difference Between distributed and relational(centralized) database

Sr. No.	Distributed DBMS	Centralized DBMS
1.	The database files are stored at geographically different locations across the network.	The database is stored at centralized location.
2.	As data is distributed over the network , it requires time to synchronize data and thus difficult to maintain.	A centralized database is easier to maintain and keep updated since all the data are stored in a single location.
3.	If one database fails, user can have access to other database files..	If the centralized database fails, then there is no access to a database.
4.	It can have data replication as database is distributed. Hence there can be some data inconsistency.	It have single database system, hence there is no data replication. Therefore there is no data inconsistency.

Types of Database Management System



History of Database(Self Study)



Thank You