CHAPTER – 1

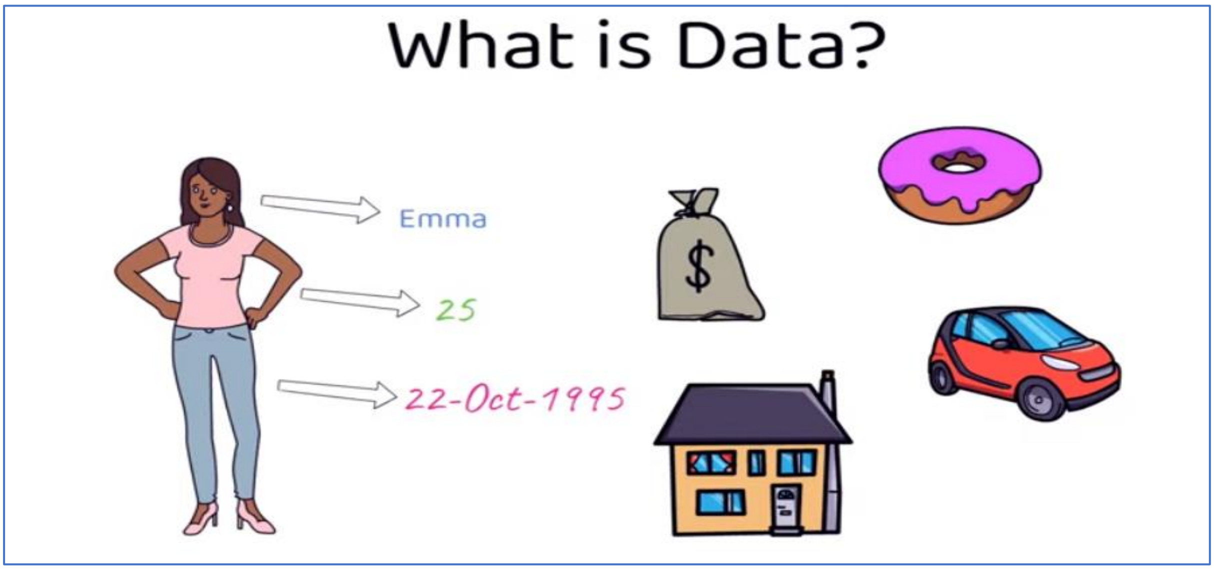
DBMS CONCEPTS

What is DATA?

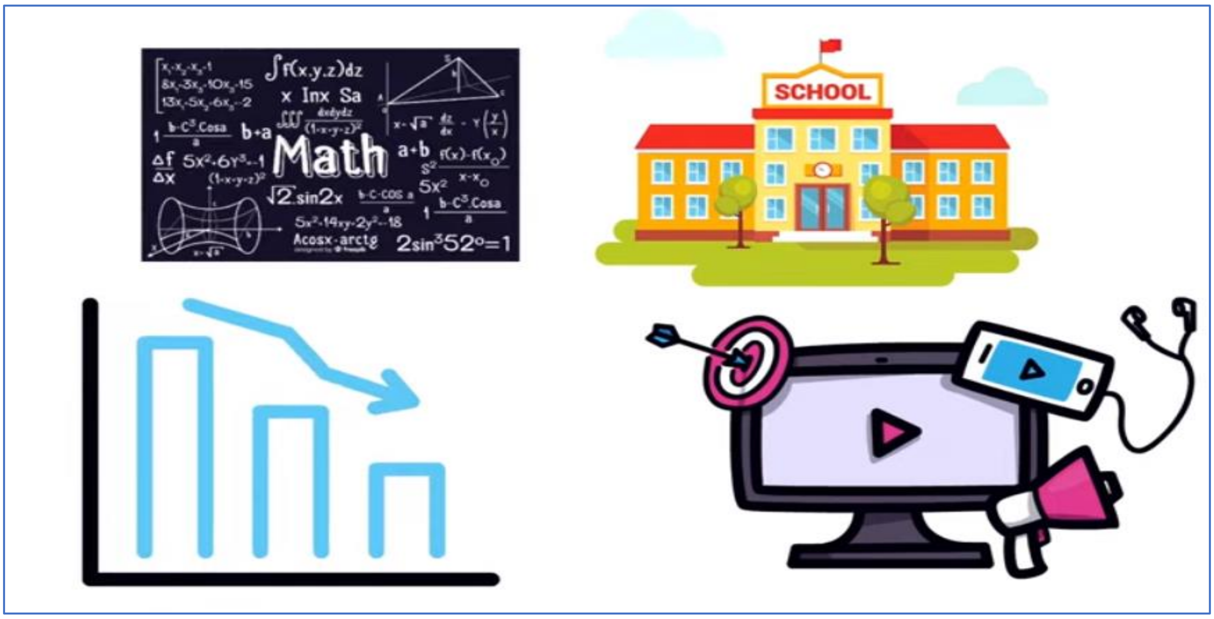
In simple words, Data can be facts related to any objects in consideration.

Ex: Your Name, DOB, Age, Height, Weight etc... are some data related to you. House Address, Bank balance, the vehicle you drive and even the food you eat can be considered as data.

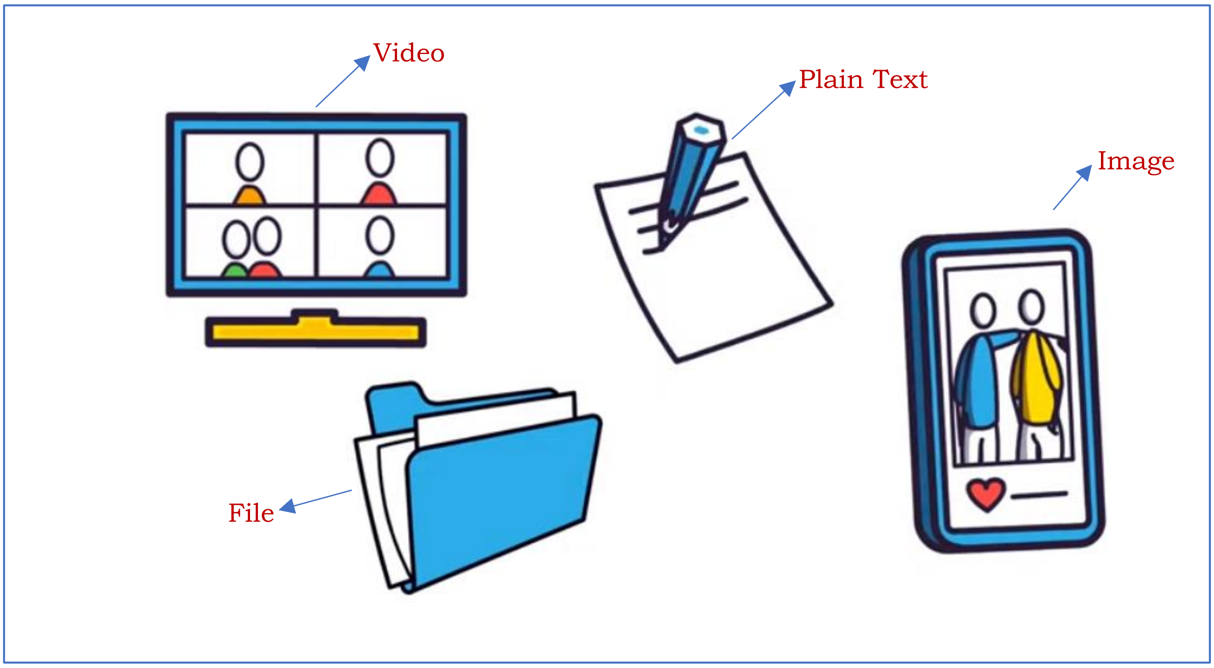
A picture, image, videos, files, etc. can also be considered as data.



Details related to a School or a technology or Statistics or even mathematics can be considered as data and can be stored into a Database.

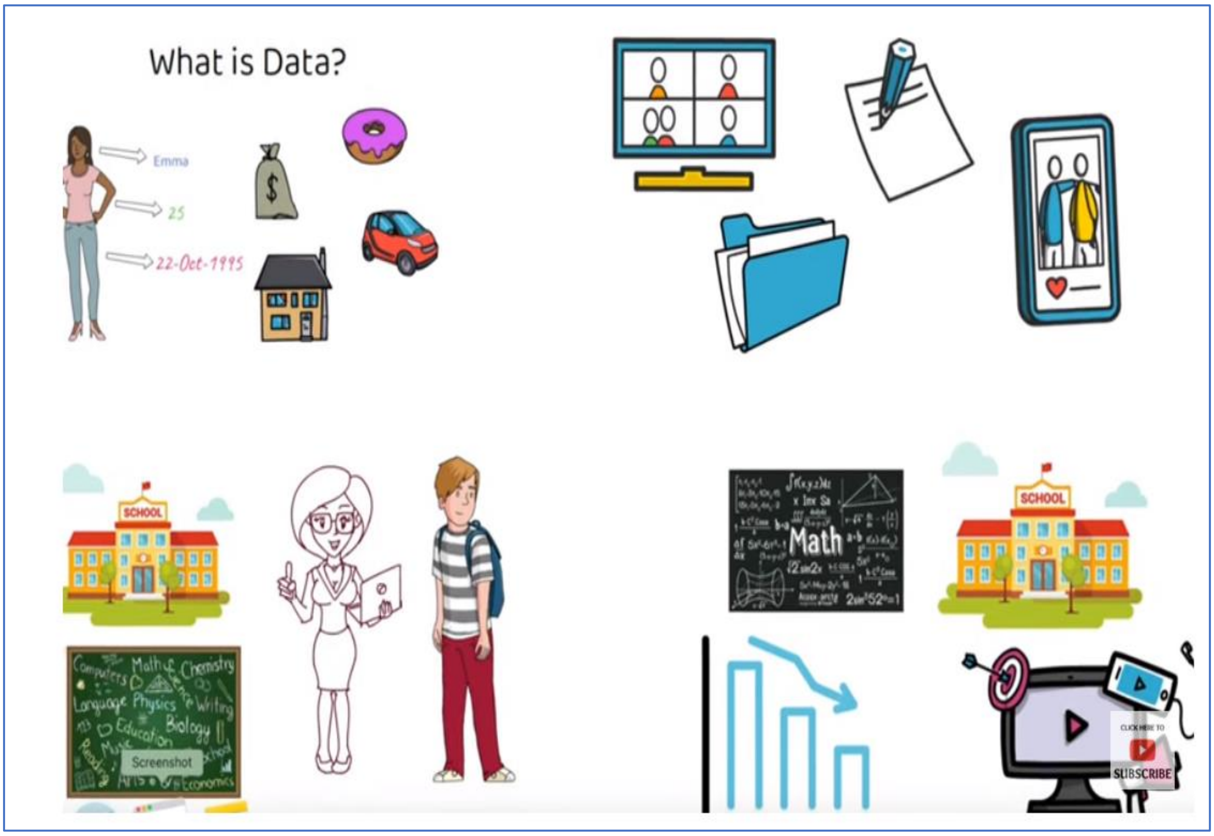


Data can be in any form such as an image or a video or a file or even a plain text. Data in any of this format can be stored into a Database.



For a School, the data can be details related to its teachers or its students or the Subjects they teach. All of these can be data and the data can be of any form.

Now you know, the data can be anything and everything and also data can be in any form.



It's an Information that is very useful and helps in making lots of business decisions when looked at different dimensions.

What is INFORMATION?

Processed or a meaningful & usable data is an Information.
Information is data that has been converted into a more useful or intelligible form.

A set of words would be data but text would be information.

DATA	INFORMATION
22, Pavan, Bengaluru	The age of Pavan is 22 who is residing in Bengaluru.
Bengaluru, 853, PRIME INTUIT, Karnataka, Near, Kothnur, J P Nagar, Main, Road, SS Tower, Floor, 560078, 3rd, etc. are some data related PI	PRIME INTUIT 853, SS Tower, 3rd Floor, Kothnur Main Rd, JP Nagar 7th Phase, Bengaluru, Karnataka 560078

In this information, the original data was interpreted, organized, and formatted according to predefined parameters. Now the meaning of the data is clear:
it is the contact information for an office named Prime Intuit.

What is Data Processing?

The data generally means raw data that needs to be collected, processed and cannot be used directly for some meaningful work such as decision making. However, once the data is processed into the information, then it can be used for some meaningful purpose which is used for decision making.

Therefore, it is said that Information is a processed Data.

For Ex: *Employee Attendance data is first recorded and then processed with the help of payroll processing software to calculate the employee salary details every month.*

Now, we need to understand where do we store the data and the purpose of storing.

DATA MANAGEMENT?

Management of Data or Records is a basic need for human society:

- Storage
- Retrieval
- Transaction
- Audit
- Archival

For:

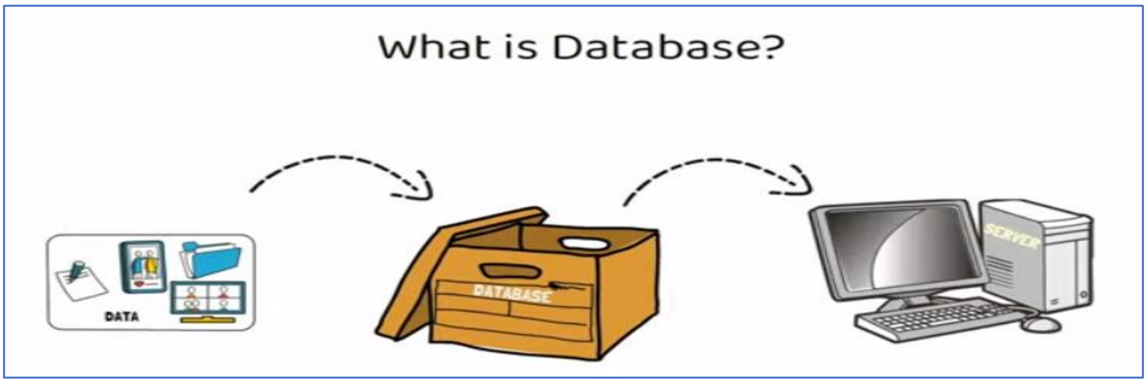
- Individual
- Small/Big Enterprise
- Global

There have been two major approaches in this practice:

- Physical – Book keeping, like physical ledgers and journals.
- Electronic - Storage devices like pen drive, Hard Drives, Servers, etc.

What is DATABASE?

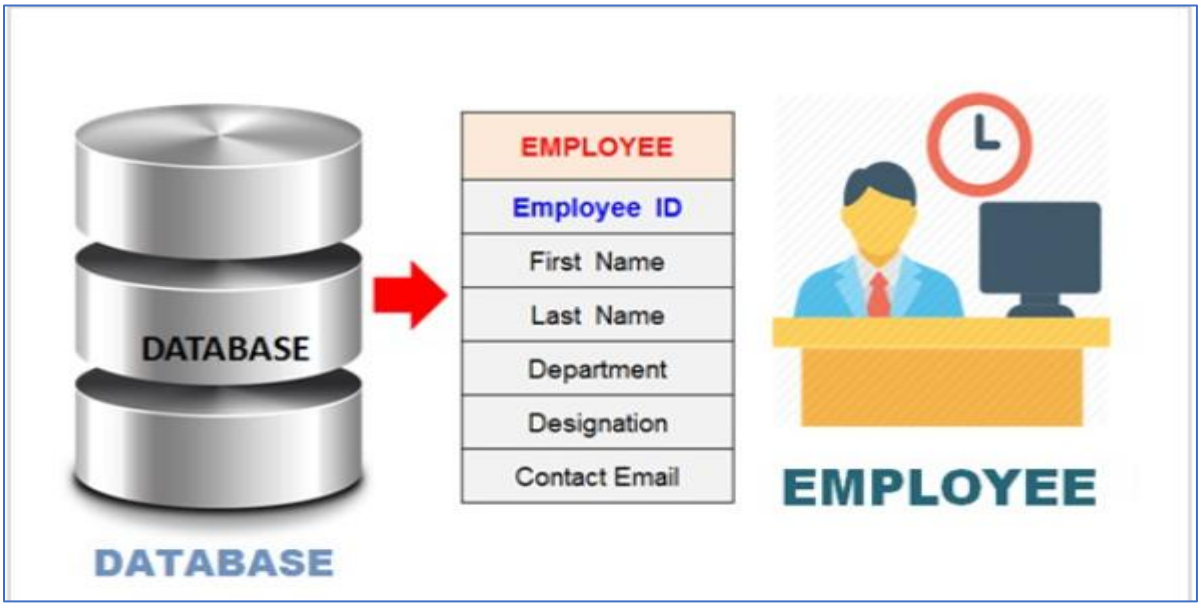
A database can be considered as a ***CONTAINER*** filled with data or information which is electronically stored in a computer system.



Database is an **organized collection of interrelated records** such as Employee Database, Student Database or Customer Database.

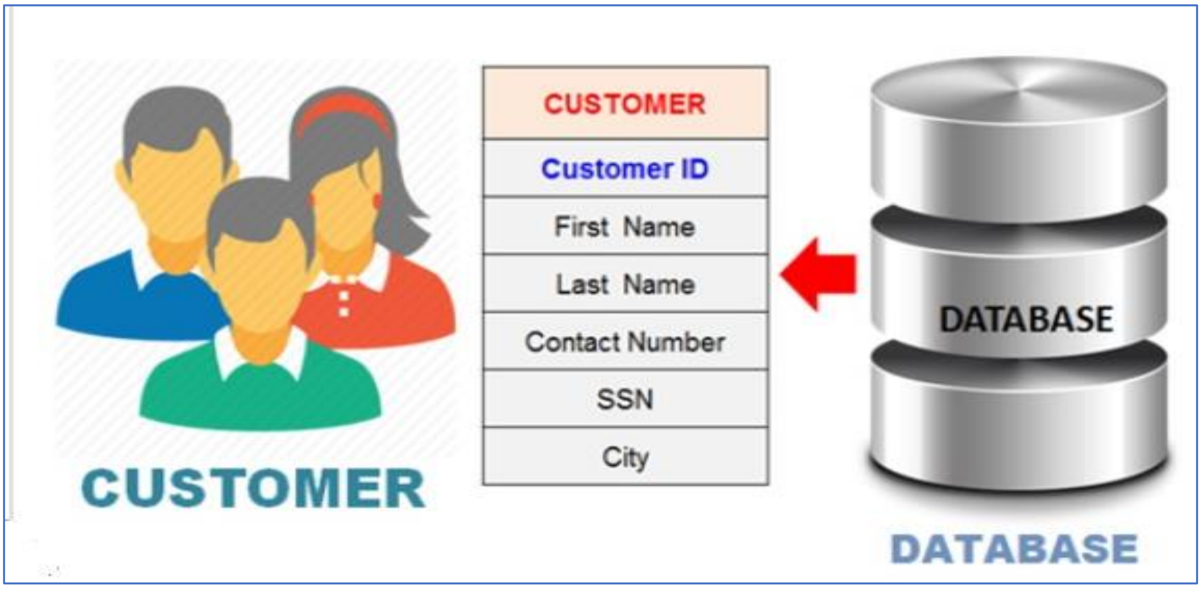
Data in a Database are Easily Accessed, Modified, Protected and Analyzed.

For Ex:
If we are designing a database to store the employee data then all the employees related information such as Name, Address, Designation, DOB, Salary can be stored as Employee Data. (Employee Database)



EMPLOYEE
Employee ID
First Name
Last Name
Department
Designation
Contact Email

The Data pertaining to the customer might include Customer Name, contact numbers, customer address, items purchased, etc…. (Customer Database)



CUSTOMER
Customer ID
First Name
Last Name
Contact Number
SSN
City

The main purpose of the database is to operate and handle large amount of information(data) by efficiently storing, retrieving and managing the data in the DB.

You may already be using Database in your day-to-day activities. Most of the websites or mobile applications like Google, Instagram, WhatsApp, Facebook already use DB to store data.

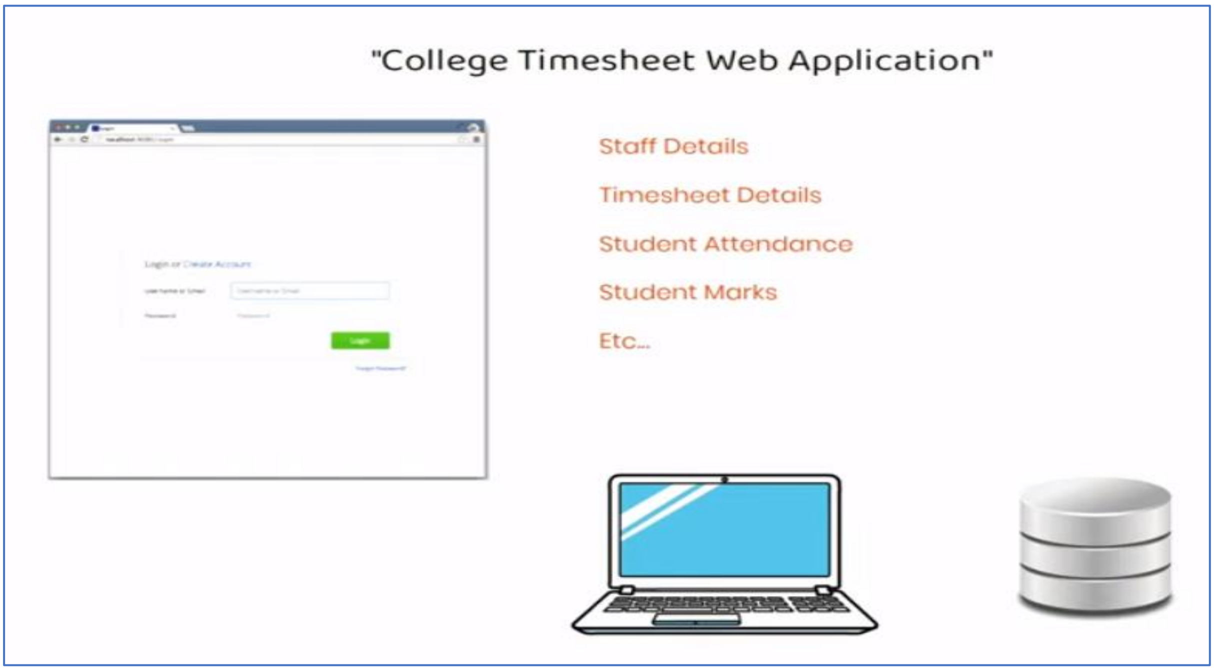
- **Facebook -** It needs to store, manipulate, and present data related to members, their friends, member activities, messages, advertisements, and a lot more.

- **YouTube -** Many videos are uploaded, who uploaded, when it was uploaded, what is the length of the video, title/description of the video, status (private or public). It's all about data with different dimensions.

For Example:

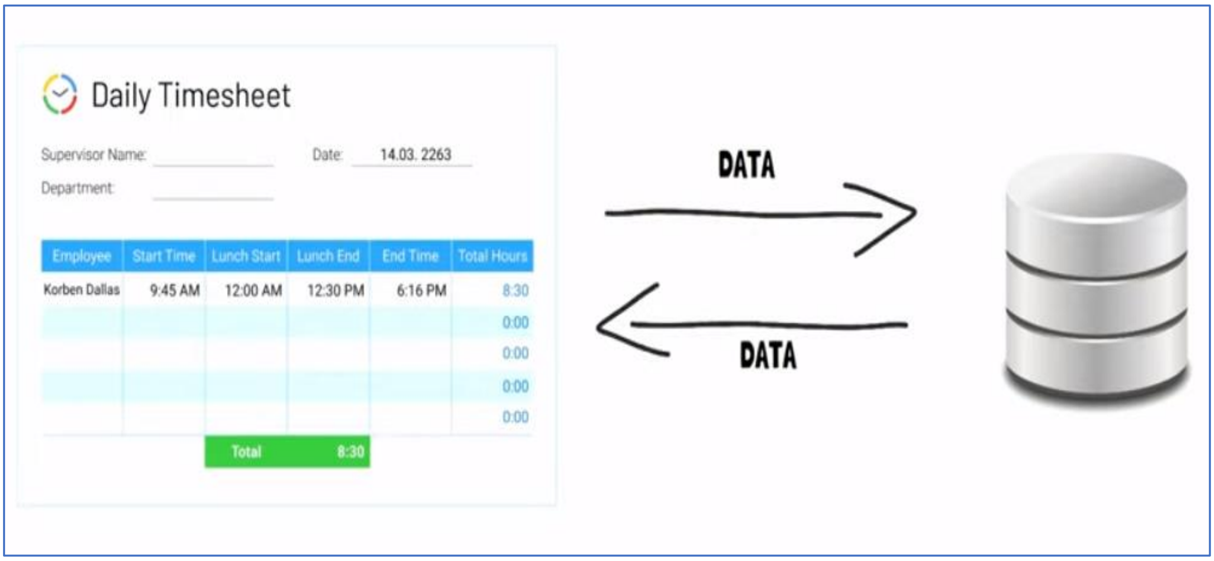
For your college demo project, you plan to develop a simple college timesheet web application. Basically, this application should allow you to login through a URL & then should have options to enter the Staff details, Timesheet details, Student Attendance, Student Marks, etc....

Since this is a college demo project, you may just use your personal laptop to develop this project. The first thing your application will need is a Database to store and retrieve the data. Hence you need to first install a Database on your laptop which will connect to this Time Sheet application.



Through your web application, whenever you try to enter, save, edit or retrieve an information, the application will connect to the database to store or modify or extract the data.

In other words, any information that is shown in your application is actually being retrieved from the database and any information that you enter in the application will be stored in the database.



Daily Timesheet

Supervisor Name: ____________ Date: 14.03. 2263

Department: ____________

Employee	Start Time	Lunch Start	Lunch End	End Time	Total Hours
Korben Dallas	9:45 AM	12:00 AM	12:30 PM	6:16 PM	8:30
					0:00
					0:00
					0:00
					0:00
		Total		8:30	

Since, the DB is installed in your laptop, the data is actually stored in your laptop memory. This is a simple example of how an application uses a small DB to store, modify and retrieve data.

Why Database?

- ***Manage large amounts of data***

A database stores and manages a large amount of data on a daily basis. This would not be possible using any other tool such as a spreadsheet as they would simply not work.

- ***Data Accuracy***

A database is pretty accurate as it has all sorts of build in constraints, checks etc. This means that the information available in a database is guaranteed to be correct in most cases.

- ***Easy to update data***

In a database, it is easy to update data using various Data Manipulation languages (DML) available. One of these languages is SQL.

- ***Data Integrity***

This is ensured in databases by using various constraints for data. Data integrity in databases makes sure, that the data is accurate and consistent in a DB.

- ***Data Security***

Databases have various methods to ensure security of data. There are user logins required before accessing a database and various access specifiers. These allow only authorised users to access the database.

<u>Evolution of Database:</u>

Started in 1960's, first type of DB was…

Flat File Database *(Stores data in simple files like csv files or fixed length files, etc.)*

↓

Hierarchical DB

↓

Network DB

(Both of these DB's Stores data through parent – child relationship. But both of these DB's were incapable of storing complex data relationships…hence were soon replaced by Relational DB's.)

In today's world there are mainly two popular Database types:

Relational DB	Non – Relational DB
Structured Data Ex: Oracle IRCTC, University DB	Unstructured Data Ex: Mongo DB Webpages

As per the usage, over 70% of the database used today are Relational DB.
But due to the immense rise in data usage over the past decade mainly due to the social media platforms, Non-Relational DB have become very popular.

However, the biggest companies today who store data related to millions of users everyday generally use combination of both Relational & Non-Relational DB.
Hence both these DB types are very popular and widely used.

File System vs DBMS: (Both are used to manage the data)

File system comes along with Operating System.

Files – Word, Excel, Notepad, PDF (Portable Document Format) …etc.

- 10 Friends Phone numbers –
  We can store it in a notepad and access.
  (File System: Folders – Sub Folders – Files….)

- Client Server Architecture:
  Data is present in the server & from all over the world the users access the data → Hence, we go with DBMS

So, reasons for DBMS over File system are:

1) Data Search:

   IRCTC application - only one train information required of 1KB data, but the whole file data of approximately - 25 GB we will get – unnecessary Memory usage in the system and time Consuming to get the file, whereas in DBMS we use a query to fetch a particular record.

2) Attributes:

   We require the file name, location, permissions, etc. i.e., we need the metadata of the data which is required to locate the file.

3) Concurrency:

   IRCTC - Multiple users can access the data – Train booking
   File system – No protocols, so there can be inconsistency for multiple users.
   No protocols exists in File System.
   DBMS – Protocols exists like RR/RW/WR/WW

4) Security:

DBMS - University data – Faculty/Students access the data
Role based access like Student role/faculty role/dean role…etc.
(Role based access control)

File System – No security.

5) Data Redundancy:(duplicacy)

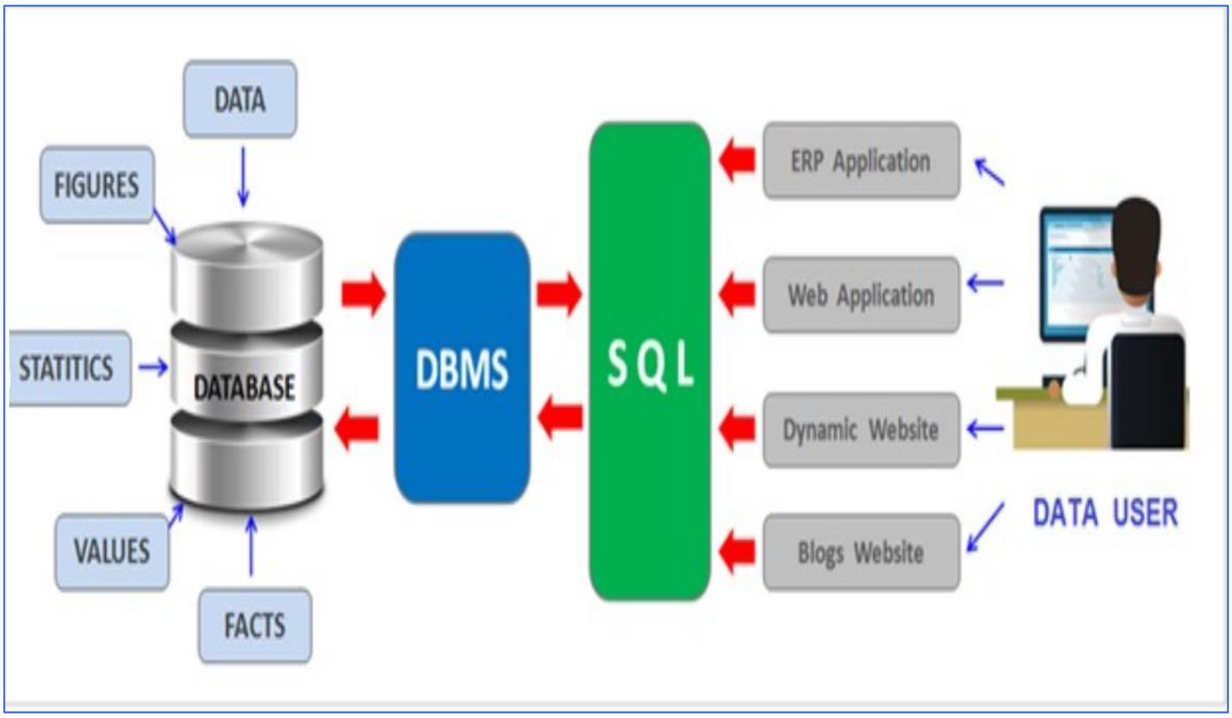
Excel file: 1) Pavan Kumar
1) Pavan Kumar
1) Pavan Kumar
But OS will check if the file names are same, it will not allow to save.

DBMS – We have constraints so that we can store unique data.

Database Management System: (DBMS)

The DBMS is a software or a group of programs designed to perform the database operations such as creating new database, administering and managing the databases by performing various user specified DB operations.

The DBMS allows the end users to perform various DB operations such as creating a new record, save data, edit data, delete data, read and manipulate data.



The DB user interacts with the application program. The application programs communicate with the DBMS using SQL (Structured Query Language) commands. The DBMS effectively functions as an interface between the application program and database. The DBMS operates on the database as per the SQL Commands.

Relational Database Management System: (RDBMS)

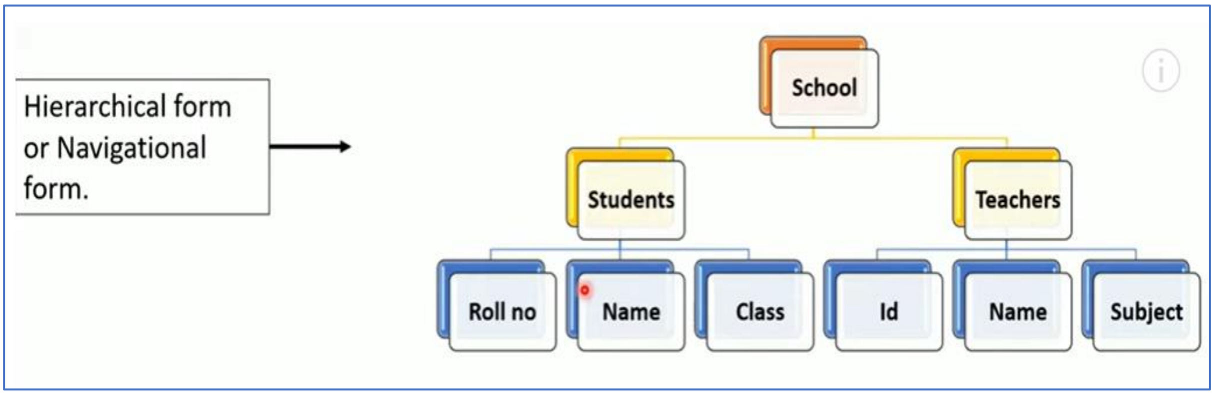
RDBMS is also a database s/w which has facility to handle more data volume, good performance,enhanced security features etc. when compared against DBMS.

Any DBMS to qualify as a RDBMS should support the Codd rules/Codd laws.

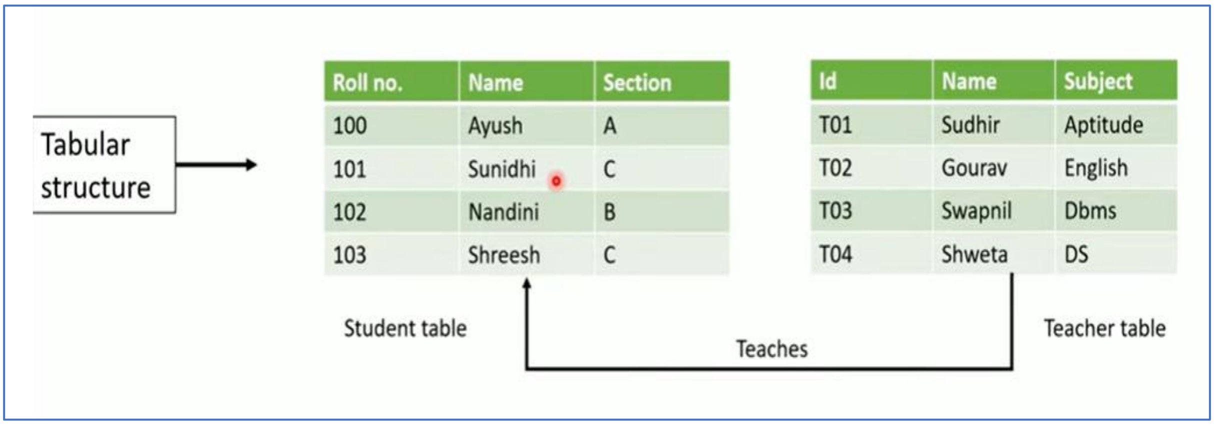
- It is a type of database that stores and provides access to data points that are related to one another.
- It stores data in tables, so it can be used in relation to other stored datasets.
- The data is often stored in many tables, also called 'relations'.
- These tables are divided into rows, also called records and columns (fields).

KEY DIFFERENCE:

DBMS stores data as a file and data is generally stored in either a hierarchical form or a navigational form. It uses level concept and also has the parent/child relations. No relation exists between the data.



In RDBMS, data is stored in the form of tables and uses a tabular structure where the headers are the column names and the rows contain corresponding values.


Tabular structure
Teaches

Roll no.	Name	Section
100	Ayush	A
101	Sunidhi	C
102	Nandini	B
103	Shreesh	C

Student table

Id	Name	Subject
T01	Sudhir	Aptitude
T02	Gourav	English
T03	Swapnil	Dbms
T04	Shweta	DS

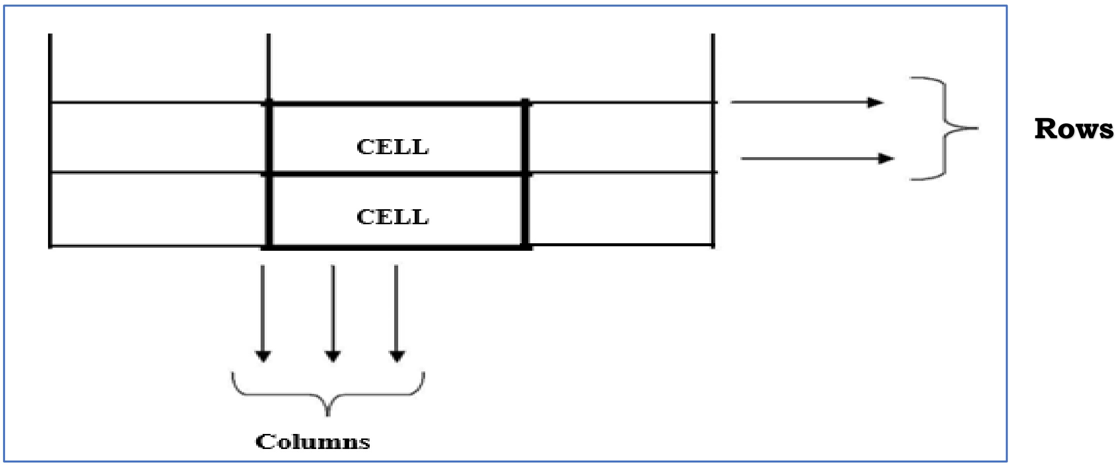
Teacher table

- ✓ DBMS supports single users, while RDBMS supports multiple users.
- ✓ DBMS does not support client-server architecture but RDBMS supports client-server architecture.
- ✓ DBMS has low software and hardware requirements whereas RDBMS has higher hardware and software requirements.
- ✓ In DBMS, data redundancy is common while in RDBMS, keys and indexes do not allow data redundancy.

The main differences between DBMS and RDBMS are given below:

DBMS	**RDBMS**
DBMS applications store data as file.	*RDBMS applications store data in a tabular form.*
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.*
Normalization is not present in DBMS.	*Normalization is present in RDBMS.*
Data redundancy is common in this model.	*Keys and indexes do not allow Data redundancy.*
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.*
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.*
Data fetching is slower for the large amount of data.	*Data fetching is fast because of relational approach.*
It supports single user.	*It supports multiple users.*
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.*
Examples of DBMS are file systems, xml, Windows Registry, dBase, FoxPro, etc.	*Example of RDBMS are MySQL, PostgreSQL, SQL Server, Oracle, etc.*

Table – A table is a collection of rows and columns.



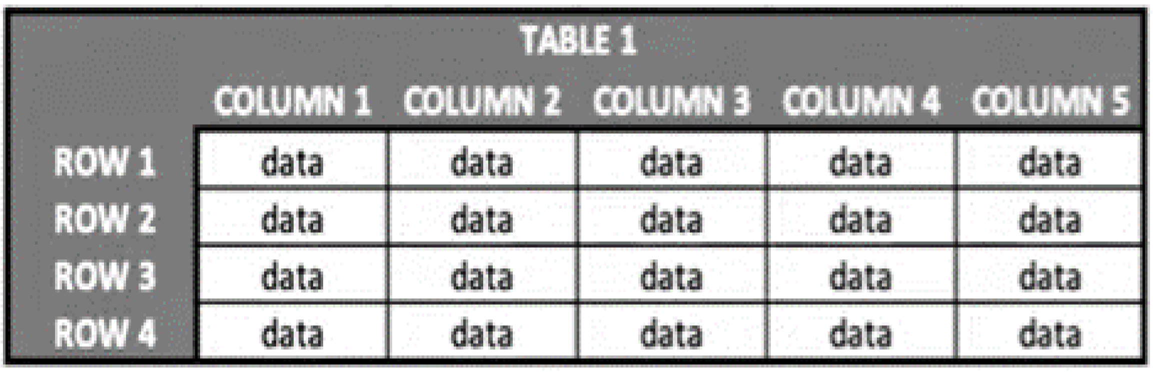
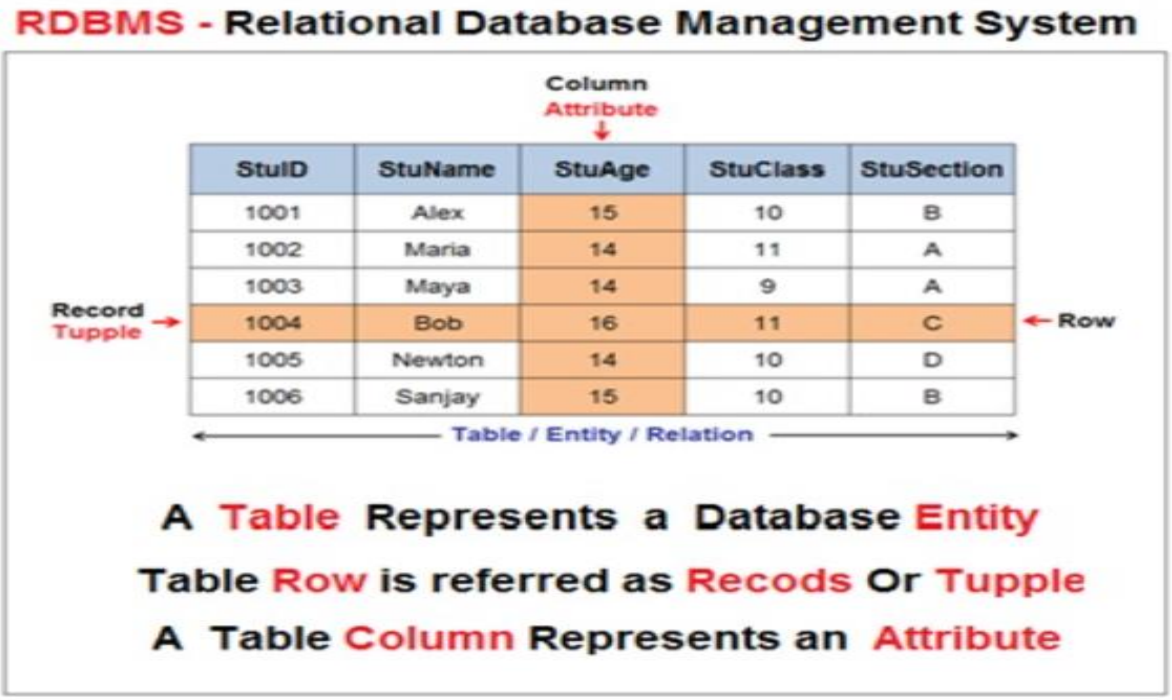


TABLE 1

	COLUMN 1	COLUMN 2	COLUMN 3	COLUMN 4	COLUMN 5
ROW 1	data	data	data	data	data
ROW 2	data	data	data	data	data
ROW 3	data	data	data	data	data
ROW 4	data	data	data	data	data

- A cell is an intersection of a row and a column
- A column is also called as a field/attribute.
- A record is also called as a row/tuple.
- A table is also called as an entity/relation.



RDBMS - Relational Database Management System

StuID	StuName	StuAge	StuClass	StuSection
1001	Alex	15	10	B
1002	Maria	14	11	A
1003	Maya	14	9	A
1004	Bob	16	11	C
1005	Newton	14	10	D
1006	Sanjay	15	10	B

A Table Represents a Database Entity

Table Row is referred as Recods Or Tupple

A Table Column Represents an Attribute

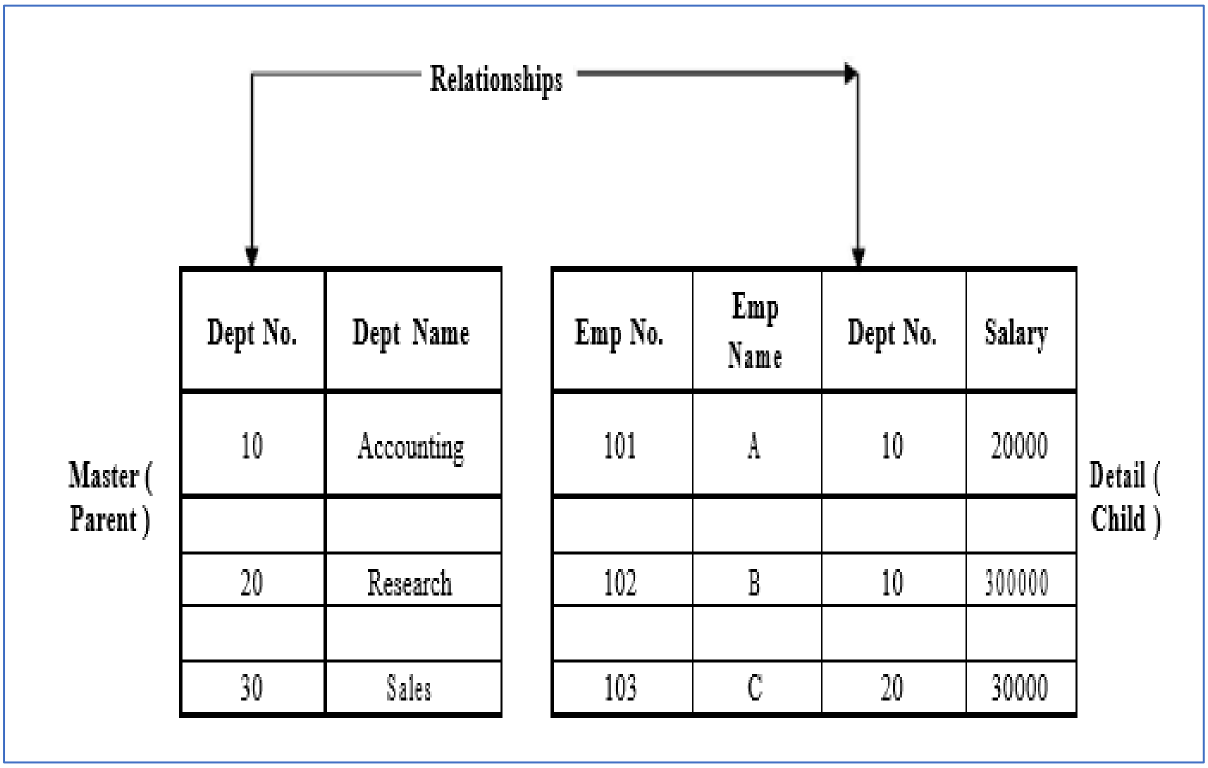
<u>Note</u>: -

- If we install any of the database related software(s) – we can create our own database, we can create our own tables and we can store the data inside it.
- When we install any database s/w(s) – a part of hard disk will be designated/reserved to perform database related activities.
- A database can also contain other database objects like views, indexes, stored procedures, functions, triggers etc...., apart from tables.

Some of the relational database software(s) we have are: Oracle, MySQL, Microsoft SQL Server, PostgreSQL...widely used is ORACLE.

RELATIONSHIPS

A relationship is the association between any two tables which preserves data integrity. (Used to maintain Accuracy and Consistency of data in the Table.)



Dept No.	Dept Name
10	Accounting
20	Research
30	Sales

Emp No.	Emp Name	Dept No.	Salary
101	A	10	20000
102	B	10	300000
103	C	20	30000

Relationship helps to prevent the incorrect data in the child tables.
Once the relationship is created, one table becomes Master (or Parent) and the other one becomes the Child (or Detail).

Whatever we insert into the child table should be present in the master table, else the record will be rejected from the child table.

The Master table contains the master data which will not change frequently.
The Child table contains the transactional data which will change quite often.

CONSTRAINTS:

A Constraint is a condition/set of rules which restricts the invalid data into the table. A constraint can be provided for a column level or table level –
Column level constraints apply to a column and table level constraints apply to the whole table.

Types of constraints:

1) NOT NULL
2) UNIQUE
3) Primary Key
4) Foreign Key
5) Check
6) Default

First, let us understand what is – NULL?

- ✓ NULL is nothing, it is neither zero nor blank space.
- ✓ It will not occupy any space in the memory.
- ✓ Two NULLS are never same in Oracle.
- ✓ NULL represents unknown value.

Some common reasons why a value may be NULL →

The value may not be provided during the data entry or the value is not yet known. Age of a particular student is not available in the age column of student table then it is represented as NULL but not as ZERO.

Any arithmetic operation we perform on NULL will result in NULL itself.
For ex, 100 + NULL = NULL; 100 * NULL = NULL; 100/NULL = NULL

NOT NULL

- ✓ NOT NULL will ensure at least some value should be present in a column.

UNIQUE

- ✓ It will not allow any duplicates in a column.
- ✓ UNIQUE column can take multiple NULL(s).

PRIMARY KEY

- ✓ It is the combination of NOT NULL and UNIQUE.
- ✓ Only one PK is allowed in a table.
- ✓ PK identifies a record uniquely in a table.
- ✓ Creation of PK is not mandatory, but it is highly recommended to create.

FOREIGN KEY

- ✓ FK creates relationship between any two tables.
- ✓ FK is also called as REFERENTIAL INTEGRITY constraints.
- ✓ FK is created on the child table.
- ✓ FK can take both NULL and duplicate values.
- ✓ To create FK, the master table should have PK defined on the common column of the master table.
- ✓ We can have more than 1 FK in a given table.

CHECK

- ✓ It is used to provide additional validations as per the customer requirements.

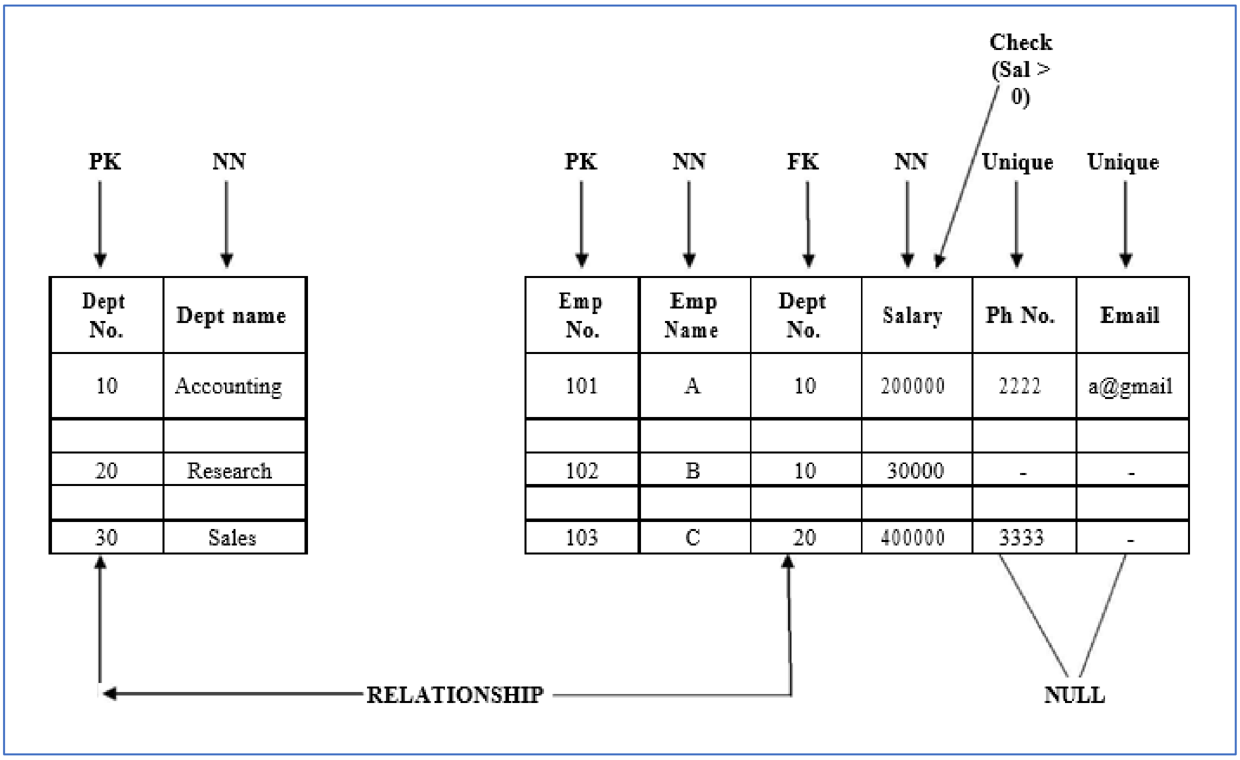
  Ex:

  CHECK SALARY > 0.
  EMPLOYEE_ID should start with 101
  COMMISSION should be between 1000 & 5000

DEFAULT

- ✓ The DEFAULT constraint is used to set a default value for a column.
- ✓ The default value will be added to all new records, if no other value is specified.

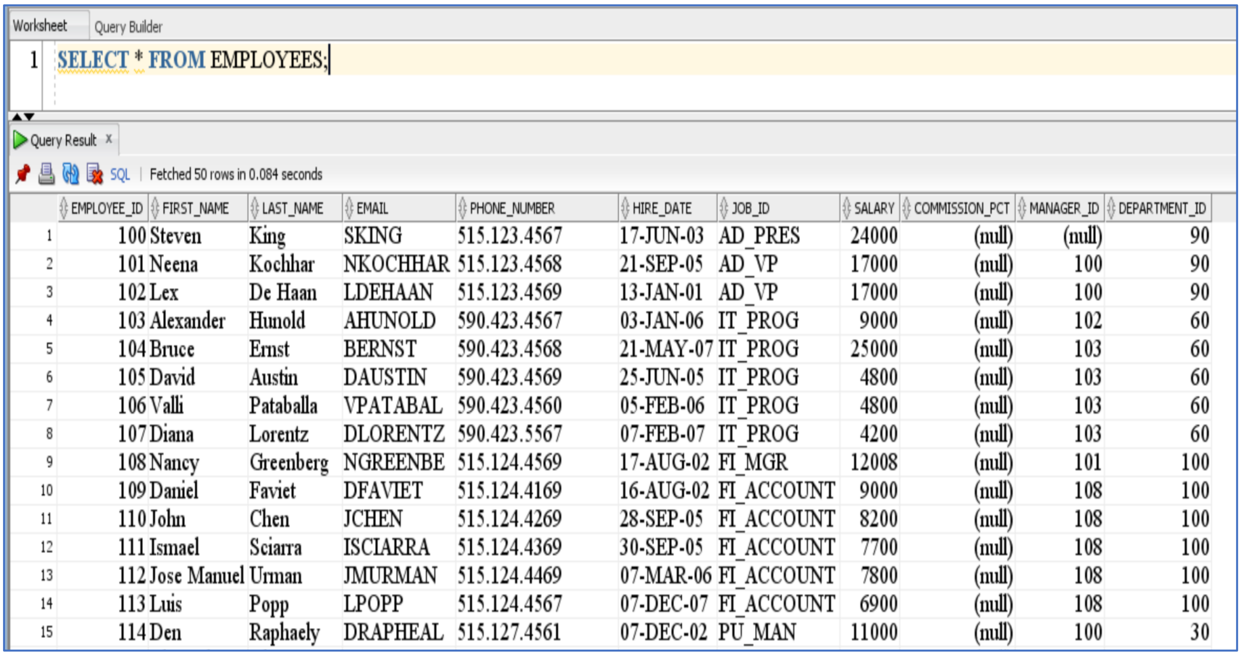
Ex: SALARY NUMBER (8,2)   DEFAULT 5000;
      CITY   VARCHAR2 (50) DEFAULT ‘BENGALURU’;



Dept No.	Dept name
10	Accounting
20	Research
30	Sales

Emp No.	Emp Name	Dept No.	Salary	Ph No.	Email
101	A	10	200000	2222	a@gmail
102	B	10	30000	-	-
103	C	20	400000	3333	-

HR Schema - Employee Table Data:



```
SELECT * FROM EMPLOYEES;
```

Fetched 50 rows in 0.084 seconds

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
1	100	Steven	King	SKING	515.123.4567	17-JUN-03	AD_PRES	24000	(null)	(null)	90
2	101	Neena	Kochhar	NKOCHHAR	515.123.4568	21-SEP-05	AD_VP	17000	(null)	100	90
3	102	Lex	De Haan	LDEHAAN	515.123.4569	13-JAN-01	AD_VP	17000	(null)	100	90
4	103	Alexander	Hunold	AHUNOLD	590.423.4567	03-JAN-06	IT_PROG	9000	(null)	102	60
5	104	Bruce	Ernst	BERNST	590.423.4568	21-MAY-07	IT_PROG	25000	(null)	103	60
6	105	David	Austin	DAUSTIN	590.423.4569	25-JUN-05	IT_PROG	4800	(null)	103	60
7	106	Valli	Pataballa	VPATABAL	590.423.4560	05-FEB-06	IT_PROG	4800	(null)	103	60
8	107	Diana	Lorentz	DLORENTZ	590.423.5567	07-FEB-07	IT_PROG	4200	(null)	103	60
9	108	Nancy	Greenberg	NGREENBE	515.124.4569	17-AUG-02	FI_MGR	12008	(null)	101	100
10	109	Daniel	Faviet	DFAVIET	515.124.4169	16-AUG-02	FI_ACCOUNT	9000	(null)	108	100
11	110	John	Chen	JCHEN	515.124.4269	28-SEP-05	FI_ACCOUNT	8200	(null)	108	100
12	111	Ismael	Sciarra	ISCIARRA	515.124.4369	30-SEP-05	FI_ACCOUNT	7700	(null)	108	100
13	112	Jose Manuel	Urman	JMURMAN	515.124.4469	07-MAR-06	FI_ACCOUNT	7800	(null)	108	100
14	113	Luis	Popp	LPOPP	515.124.4567	07-DEC-07	FI_ACCOUNT	6900	(null)	108	100
15	114	Den	Raphaely	DRAPHEAL	515.127.4561	07-DEC-02	PU_MAN	11000	(null)	100	30

What is a Schema in DB?

Logical representation of a database,

STUDENT: → ***This is an ENTITY***

ROLL NO	STUD NAME	STUD ADDRESS	MOBILE NO

→ ***SCHEMA***

The Schema is designed/implemented using SQL with DDL commands.

- A database schema is the logical representation of a database, which shows how the data is stored logically in the entire database. It contains list of attributes and instruction that informs the database engine that how the data is organized and how the elements are related to each other.
- A database schema contains schema objects that may include tables, **fields, packages, views, relationships, primary key, foreign key**
- In actual, the data is physically stored in files that may be in unstructured form, but to retrieve it and use it, we need to put it in a structured form. To do this, a database schema is used. It provides knowledge about how the data is organized in a database and how it is associated with other data.
- The schema does not physically contain the data itself; instead, it gives information about the shape of data and how it can be related to other tables or models.
- A database schema object includes the following:

  - ✓ Consistent formatting for all data entries.
  - ✓ Database objects and unique keys for all data entries.
  - ✓ Tables with multiple columns, and each column contains its name and datatype

What is a Instances in DB?

Instances are the collection of information stored at a particular moment. The instances can be changed by certain CRUD operations as like addition, deletion of data. It may be noted that any search query will not make any kind of changes in the instances.

****