Database Management Systems

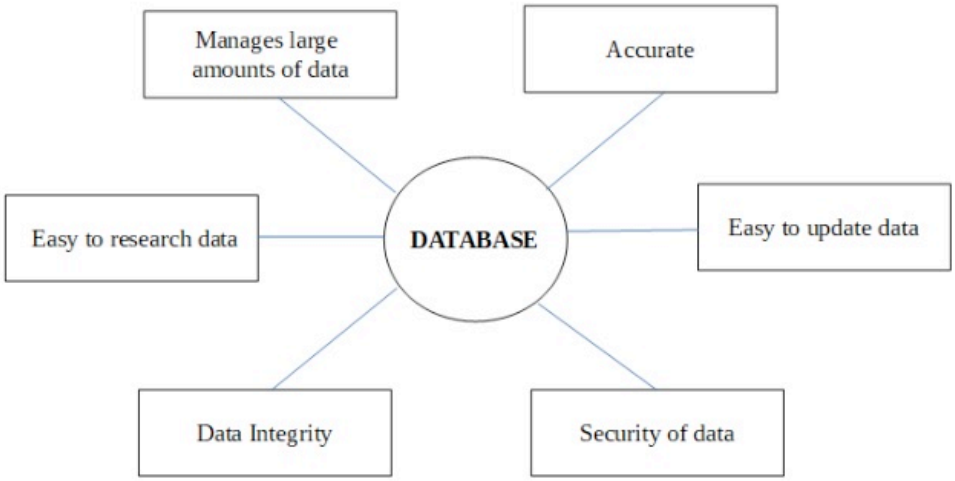
Week - 01

* **Data** → originated from word "datum"
  (means Single piece of information)

* **Database** → organised collection of data, usually stored in electronic form, so that it can be easily accessed and organised.

Why database ?

A good database is crucial to any company or organisation. This is because the database stores all the pertinent details about the company such as employee records, transactional records, salary details etc.



Earlier file systems were used to store data, but they had some drawbacks.

* **Drawbacks of file systems**

Data redundancy and inconsistency → Duplication
Difficulty in accessing data → New program for each task
Data isolation → Multiple files and formats
Integrity problems → Hard to add constraints

Atomicity of updates → failure may leave database inconsistent
Concurrent access by multiple users → multiuser can cause inconsistency
Security problems
↑ Hard to provide user access

Evolution of Data Management :→

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

1. Storage
2. Retrieval
3. Transaction
4. Audit
5. Archival

There have been two major Approach in it :

1. Physical
2. Electronic

➢ Physical Data or Records management, more formally known as Book Keeping, has been using physical ledgers and journals for centuries.

✶ Problems with such an approach of book-keeping:

1. Durability
2. Scalability
3. Security
4. Retrieval
5. Consistency

Not only small shops but large organizations also used to maintain their transaction details in book registers.

A better Solution ⟶ Spreadsheets

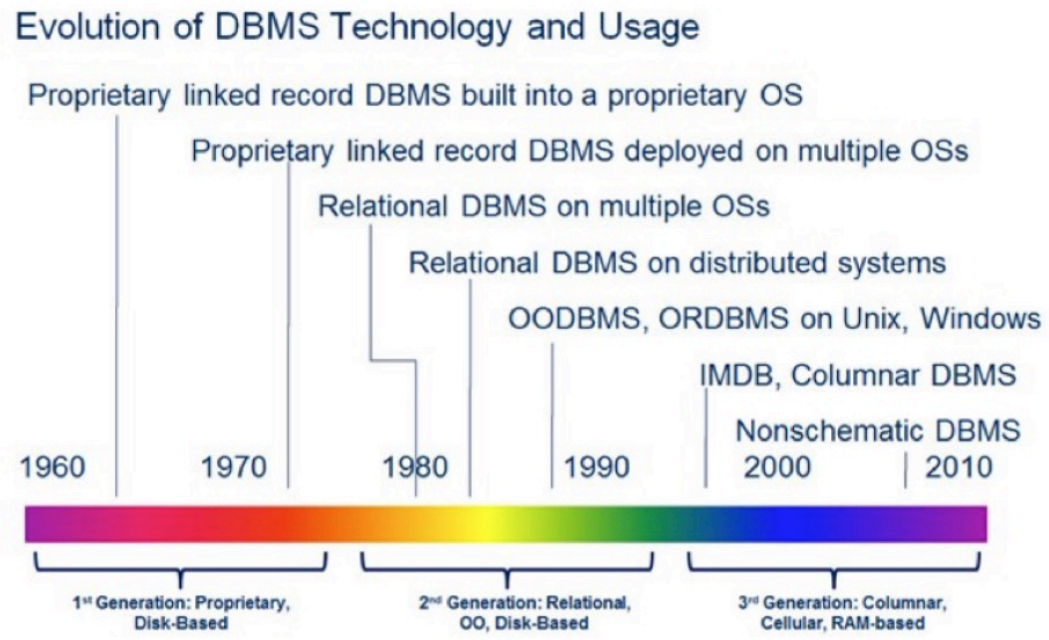
Spreadsheet Softwares like Google Sheets: Due to the disadvantages of maintaining ledger registers, organizations dealing with huge amount of data shifted to using spreadsheet softwares for maintaining their records in files.

- Durability: These are computer applications and hence data is less prone to physical damage.
- Scalability: Easier to search, insert and modify records as compared to book ledgers
- Security: Can be password-protected
- Easy of Use: Computer applications are used to search and manipulate records in the spreadsheets leading to reduction in manpower needed to perform routine computations
- Consistency: Not guaranteed but spreadsheets are less prone to mistakes than registers.

Why leave file systems :→

Lack of efficiency in meeting growing needs

- With rapid scale up of data, there has been considerable increase in the time required to perform most operations.
- A typical spreadsheet file may have an upper limit on the number of rows.
- Ensuring consistency of data is a big challenge.
- No means to check violations of constraints in the face of concurrent processing.
- Unable to give different permissions to different people in a centralized manner.
- A system crash could be catastrophic.

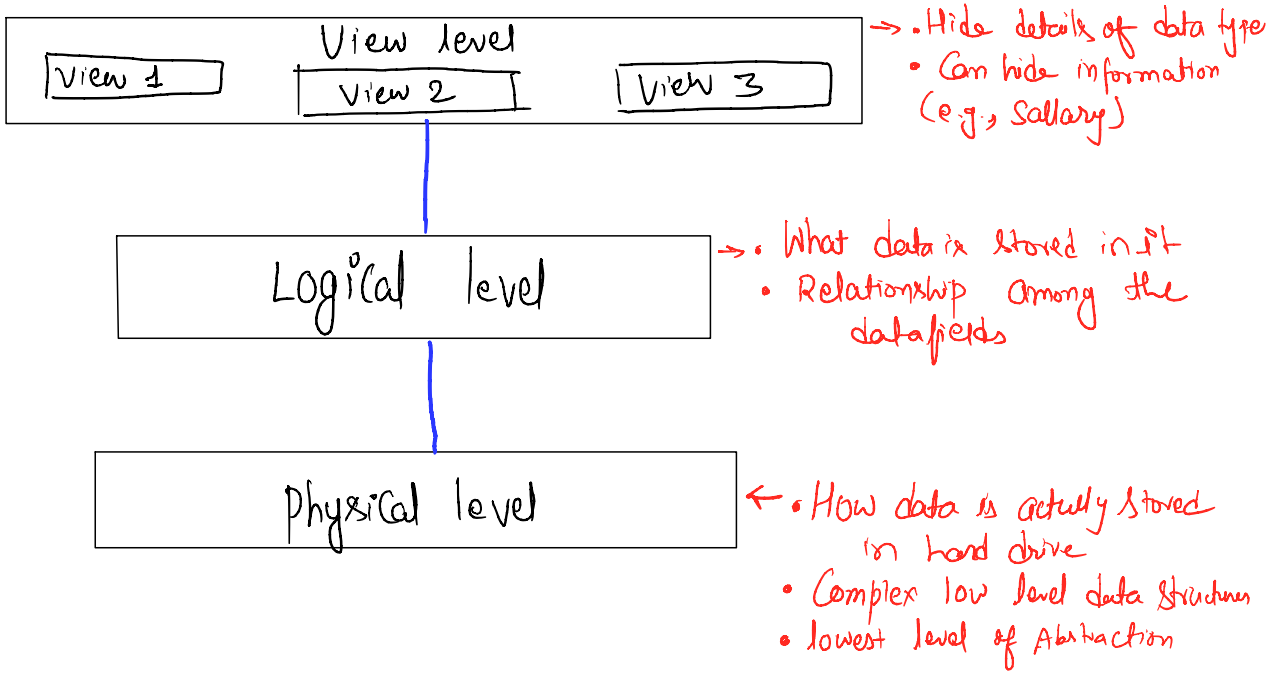


Parameter	File Handling via Python	DBMS
Scalability with respect to amount of data	Very difficult to handle insert, update and querying of records	In-built features to provide high scalability for a large number of records
Scalability with respect to changes in structure	Extremely difficult to change the structure of records as in the case of adding or removing attributes	Adding or removing attributes can be done seamlessly using simple SQL queries
Time of execution	In seconds	In milliseconds
Persistence	Data processed using temporary data structures have to be manually updated to the file	Data persistence is ensured via automatic, system induced mechanisms
Robustness	Ensuring robustness of data has to be done manually	Backup, recovery and restore need minimum manual intervention
Security	Difficult to implement in Python (Security at OS level)	User-specific access at database level
Programmer's productivity	Most file access operations involve extensive coding to ensure persistence, robustness and security of data	Standard and simple built-in queries reduce the effort involved in coding thereby increasing a programmer's throughput
Arithmetic operations	Easy to do arithmetic computations	Limited set of arithmetic operations are available
Costs	Low costs for hardware, software and human resources	High costs for hardware, software and human resources

Levels of Abstraction :→

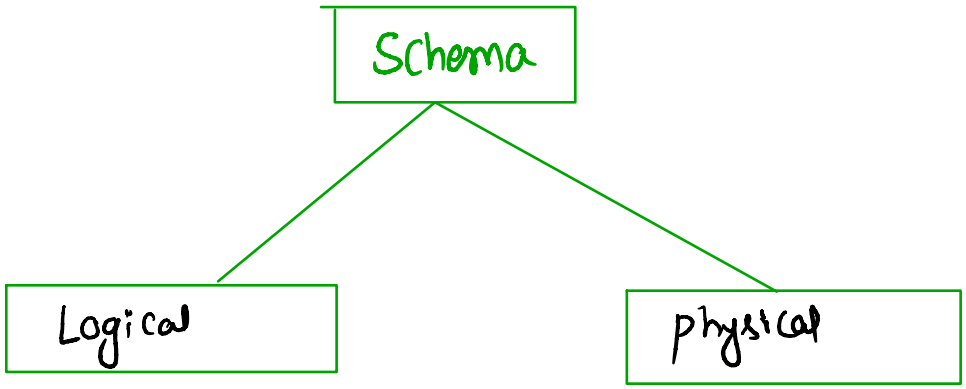
process of adding certain details and showing only the essential information to the user.

We have three levels of abstraction.



Schema :->

A Schema is the logical representation of the database.



- the overall logical structure of the database
- Analogus to type information of a variable in any program
- eg., Customer Schema, Account Schema

- the overall physical structure of the database.

Instance :->

The actual content of the database at a particular point of time.

-> Analogus to the value of the variable.

Data Independence :->

physical Data Independence :->

The ability to modify the physical Schema without changing the logical Schema.

- Modifying one level without influencing the upper levels

Logical Data Independence :->

The ability to modify the logical level without influencing the View level.

Data Models :→

It describes how a logical database structure should look like after it has been fully implemented

It describes

- Data
- Data relationships
- Data Semantics
- Data Constraints

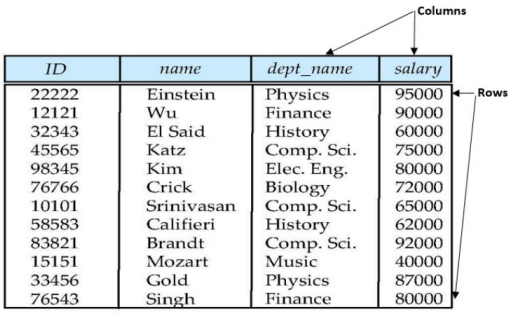
There are many types of data models in DBMS, some are;

- Hierarchal model
- Network model
- Entity - relationship model
- Relational model
- Objected Oriented data model
- flat data model

Relational Model :-

- All data is stored in various tables

- Example of tabular data in the relational model



ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

(a) The *instructor* table

dept_name	building	budget
Comp. Sci.	Taylor	100000
Biology	Watson	90000
Elec. Eng.	Taylor	85000
Music	Packard	80000
Finance	Painter	120000
History	Painter	50000
Physics	Watson	70000

(b) The *department* table

Query Processing :-

a) Parsing and translation
b) Optimization
c) Evaluation