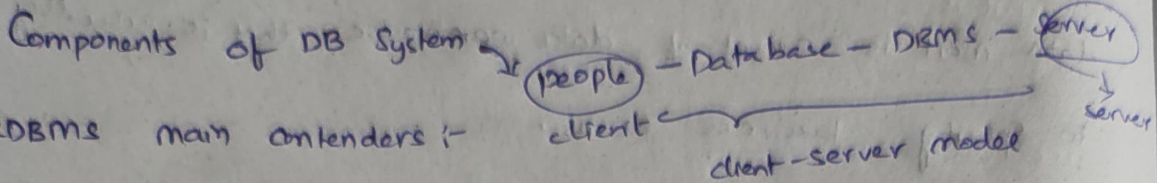


## SQL:-

→ We can connect n no. of servers to our ~~Ans~~ sql studio

→ Sql server is a database on database server



RDBMS main contenders:-

Commercial Software

→ IBM DB<sub>2</sub>

→ ms sql

→ oracle

etc.

others

→ mysql

→ postgresql

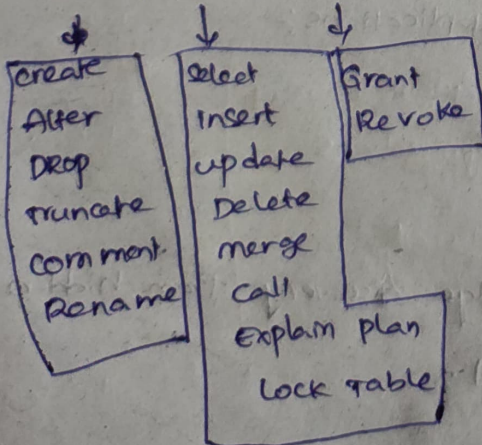
→ For any application we need servers for ~~etc~~

Ecom Appl. → app server  
→ web server  
→ DB server.

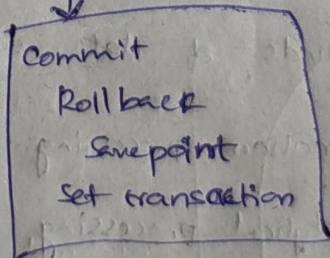
→ ISAM method used to store data in database tables.

(↓ Indexed sequential access method)

→ DDL, DML, DCL, Commands



SQL





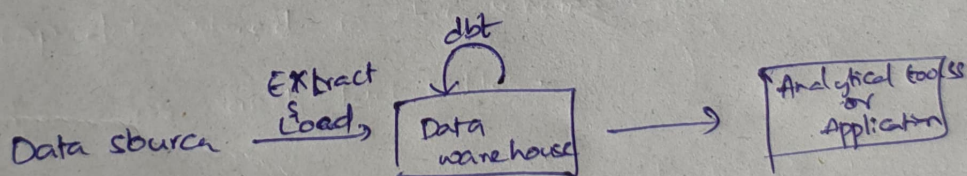
## Data cleaning:-

The process of data cleaning involves identifying inaccuracies in dataset and then fixing them. It's first step in any analysis and it includes deleting data, updating data and

→ process of fixing or removing incorrect, incomplete data ---

### 8-step data cleaning process :-

1. Remove irrelevant Data — Select \* from table-name where city = 'pune';
2. Remove duplicate data — Select distinct \* from table-name;
3. Fix structural errors
4. Do type conversion
5. Handle missing data.
6. Deal with outliers
7. Standardize / Normalize data
8. Validate data



## Data cleaning & Manipulation:-

### ETL

↑ have extracted  
If you have 10 files and that 10 files, you are transforming them to some files using SQL commands (DDL, DCL, TCL) and that transformed data is stored in data warehouse  
→ At transform you use SQL, PySpark and some tools

→ This process is data cleaning.