

- DBMS

Data : information  
raw fact figure

Nawdup

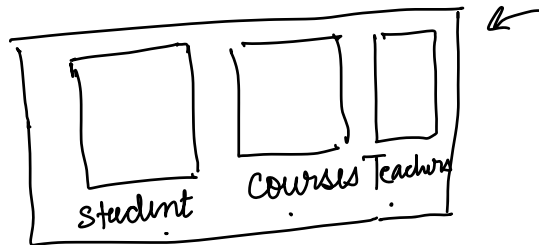
98103

CTC

- Database

collection of data in a related way

Pepecoding DB



- DBMS - software

↳ manage



efficiently

store

query

Time

• File system  
Peppcoding

- students.txt
- course.txt
- mentors.txt

Files →  
Hard disk  
 ↳ sequential

<u>student.txt</u>			
1	Sanket	2020	abhinav CTC
2			
3			
100	-	-	-

- 1) Searching - sequential [non efficient]
- 2) Redundant - multiple copies
- 3) inconsistent data

Sanket 12345  
            
           ✓

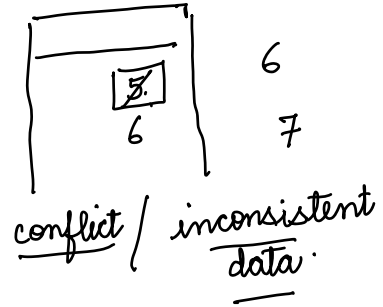
Sanket 12345  
           98123  
            
           ✓

- 4) Data back up
- 5) Security → Hard disk ✓  
           ↳ DBMS - encrypted  
                           RBAC

6) Concurrency ✓

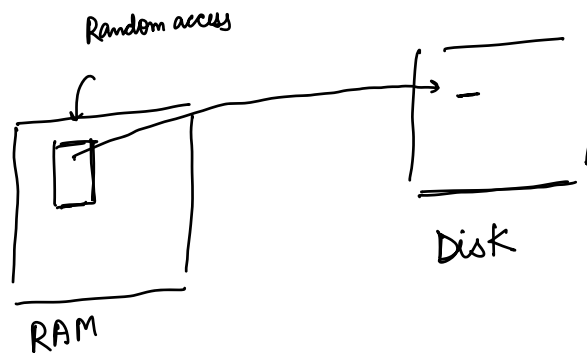
↳ File

5 →



DBMS

1)



2) No Redundant

3) Security

→ RBAC

↳ encryption

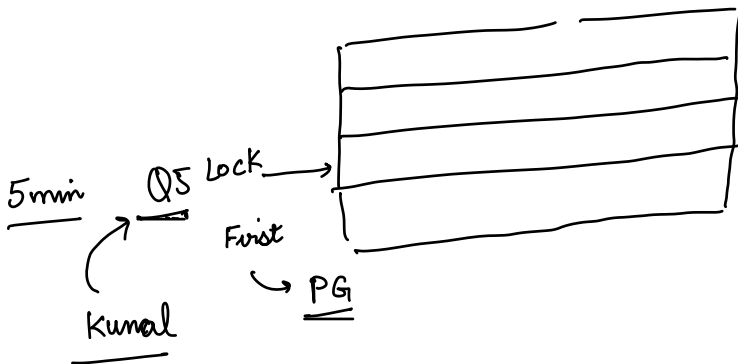
4) backup

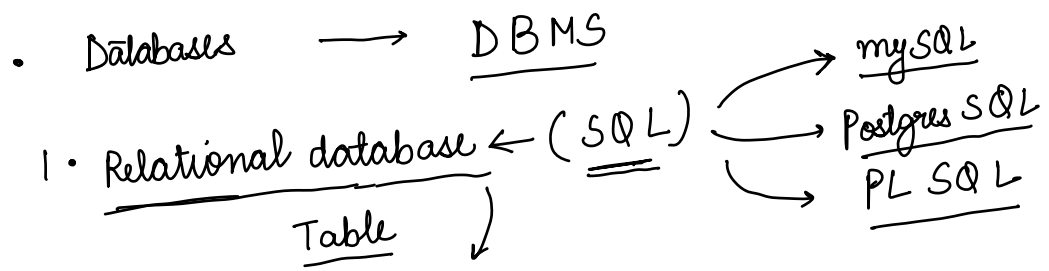
↳ 3 am

— Database

5)

Concurrency ✓





File → RDBMS  
Table — Relation

Student


mentors


Courses


batches


attributes

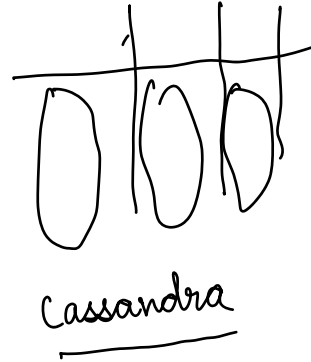
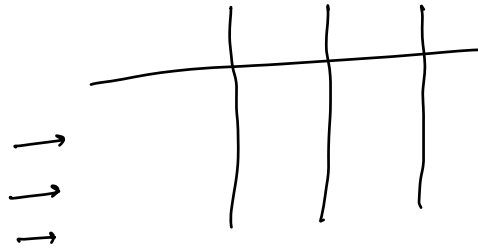
Name | email | Phone no | password | grad\_year

batches

Name | start-date | curr-ins | time

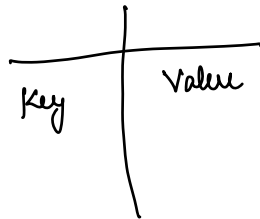
## Other Databases

1. Graph based — Neo4J
2. Columnar Databases



3. Document object json } → mongoDB  
DynamoDB

- 4) Key - value pair (Redis) → [in memory storage]



## • Relational Databases

↳ Relational Data model

↳ Theoretical model

way to store and interact with data.



↓  
Relational algebra

- Sets → no duplicates
- Union → unordered
- Intersection

[1, 3, 4, 9, 8]

list ↑

Table → Sets

Student

ID	Name	Email
2	Ayush	abc@xyz.com

Select \*  
where from Student  
Name = "Naman" ⇒ Subset

Relational database model

1) most widely data model

2) Data model behind SQL

3) multiple tables / relations

rep data

↓  
rows

↓  
tuples  
instances

4) values in relations

Student  
relation

ID	Name	Email
----	------	-------

- Relations
- attributes (column headings in table)
  - tuple (rows in table)
  - degree — number of attributes in a relation
  - Cardinality → number of tuples in a relation

$\cup$  — Union

$\cap$  — intersection

$*$  — Cartesian product

$R_1$  ✓

Name
A
B
C

$R_2$  ✓

City
X
Y
Z
NULL

← missing value

$R_1 * R_2$

Name	city
A	X
B	X
C	X
A	Y
B	Y
C	Y
A	Z
B	Z
C	Z

$\left[ \begin{matrix} 9 \\ - \end{matrix} \right] \rightarrow 9$

NULL ← missing values Table/R

• x →

5 columns

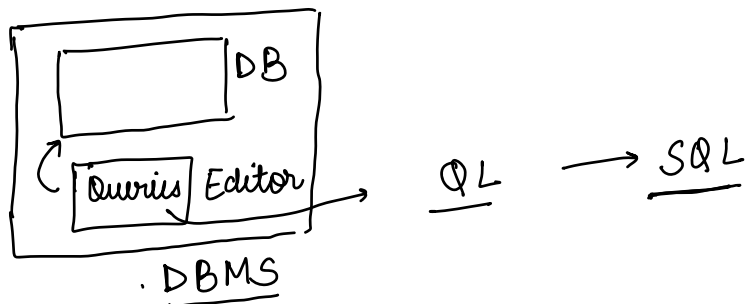

R1  
R2  
R3

R  
A  
T

degree = 5  
cardinality = 3

RDB → Tables  
RDBMS → software  
SQL — PL  
RD model ✓

10 min  
3 — 3:10





## Properties of table / relation.

1) atleast 1 column  $\rightarrow$  unique identifies each row

2)

Name	Email	Phone
⋮		

for every column  $\rightarrow$  there should be data type

3) Order of columns doesn't matter.

Select Name, batch from Students;

Name	batch
↓	↓
✓	✓

batch	Name
↓	↓
✓	✓

4) Order of rows doesn't matter.

R1	_____ ✓	R3	✓ _____
R2	_____ ✓	R2	✓ _____
R3	_____ ✓	R1	✓ _____
I		II	

⑤ every column name should be distinct.

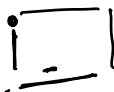
⑥ every value in column should be atomic in nature.


atomic → single values


1, true, "hello"

list → [ "hello", "world" ]

String [20] Name	int grad year	String [50] univ	double Percentages
20 B XYZ ABC	4 B 2019	50 B YZA	8 B = 182 B 92

1st row →   
add = 100

  
82 B

2nd row   
182

100 - 181  
82 182 B

Random access

List - ✓  
map - ✓

- 1. difficulty to store
- 2. difficulty to compare

• NoSQL ✓  
 ✓ too much  
writes  
 [structure]  
Relation

• SQL ✓  
 too much  
 relations  
 too much queries  
 [s/r]

## Keys

- 1) identify a record/row
- 2) identify relation between 2 tables sets relations

## Doubts system in Pepcoding

Doubts

	Student_id	Mentor_id	Time	Rating	Feedback
→	1 ✓	66	3:14pm	4	
→	2	66	4:14	5	
→	3	66	5:14	5	
→	4	66	6:14	4	
→	1 ✓	67	7:14	3	
→	1 ✓	68	8:14		
→	1	66			

## Superkey

A set of attributes which can uniquely identify a row.

Student\_id, mentor\_id

mentor\_id, time ✓ - 2

66  
time -

1 → 66  
2 → 67

→ 3:14  
→ 3:14

Rating, time, mentor id - 3

Student id, mentor id, time - 3

all the attributes - 4

Candidate Key - minimal super key

(2) mentor id, time ✓  
Student id, time ✓ → composite key

Any subset of attributes where no 2 rows will have same values and the size of subset should be minimum.

Primary Key }  
✓ one of the candidate

8 people

CK

PK

Composite Key

- Key of  $\geq 2$  attributes
- No single attribute to identify rows

CK  $\rightarrow$  (2)  $\rightarrow$  PK  
 min super key  
 (time) string  
 - time, feedback  
 int int  
 - monitor\_id, time  
 int int  
 - stud\_id, time  
 comparisons are fast.

• Foreign Key  
 relation  
 primary key in another table  
 composite

Student

Student-id	N	P	E	<u>Hostel-id</u>

T1

Foreign Key

<u>Hostel-id</u>	Name

T2

