

We start at 7:05

# Intro to DBMS and Relational models.

Ujjwal.

SDS-2 Assignment



Blinkit



Amazon & Microsoft

## Agenda.

Theory

- ① What is a database
- ② Why are we teaching database
- ③ What, How & Why at Scales.
- ④ Type of database
  - ↳ relational
  - ↳ non relational

Practical

- ⑤ relational databases
- ⑥ keys
  - ↳ Super
  - ↳ Candidate
  - ↳ Foreign
  - ↳ Primary

MySQL.

What is a database.

daily life → data

↳ phone no, email address, photo, ~~contact~~.

Scaler

- Question
- recording
- student info → name, email, pSp
- insn or
- 

how to save this data

↳ excel

↳ screenshot

↳ one note

↳ csv

notepad → new file → vijwal-987 → vijwal.txt

Students :  $\rightarrow$  students.txt

name	batch-id	isp	attendance	coin	rank
Ujjwal	1	95	100	20	1
Amit	1	72	<del>95</del> <u>90</u> <u>89</u>	5	2
Suresh	2	63	63	1	2

Ujjwal  
Suresh

$$\frac{a_1 + a_2 + a_3 + \dots + a_n}{n}$$

⑧ Give me avg-attendance per batch name.

batch-name	avg-attendance
Aug-1, 23	98
Sep-2, 23	72
March, 6, 23	88

batches.txt

batch-name	batch-id	st-date	end-date
Aug 1, 23	1	01/08	12/12
Sep 2, 23	2	02/09	02/12

Thing that can go wrong.

- ① There is a lot of passing.
- ② mapping is not error proof.
- ③ concurrency.

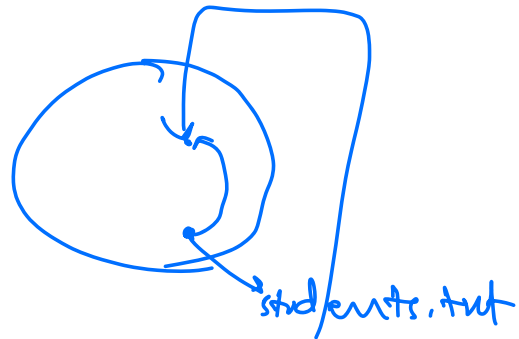
- ④ Security.
- ⑤ SLOW.

## Issues with storing data in files.

### ① Inefficient



(database use other methods like indexes)



### ② data integrity.

→ corrupted data  
→ inconsistent types

### ③ concurrency

### ④ Security

→ access control

→ read only  
→ read/actors  
→ append only  
→ admin

## What is a Database .

Airbase →  
Military base → ↓

A collection of related data .

### Scaler database

+ students  
+ instructors  
+ mentors  
+ TAs  
+ jobs  
+ payments  
+ company

+ students.txt  
+ instructors.txt  
+ ...

NOT EFFICIENT

## What is a database management system



software system that allows to easily work with database

Imp. efficiency

- 1) CRUD
- 2) Access control
- 3) concurrency

## (4) Integrity

Why are we teaching DBMS.

Early 2023, later 2022 → 15%.



Imp topics / ques for interviews  
Imp thing for day-to-day job.

Indices  
SQL queries  
→ transactions  
→ execution queries

Scalability, NOSQL, sharding

## Topics

- ① Intro to DBMS & SQL
- ② CRUD → Create / Read / Update / Delete
- ③ Joins
- ④ Agg queries
- ⑤ Sub queries & views
- ⑥ Problem Solving.

- ⑦ Indexes
- ⑧ Transaction
- ⑨ Schema Design - 1
- ⑩ Schema Design - 2
- ⑪ Window function.

Interview



DSA } SDE 1

LLD

HLD



SDE 2

→ Schema Design

→ 20%

Break till 8:10.

Types of database.

Students

Info

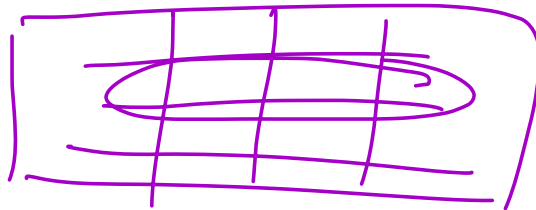
Tables

Relational database

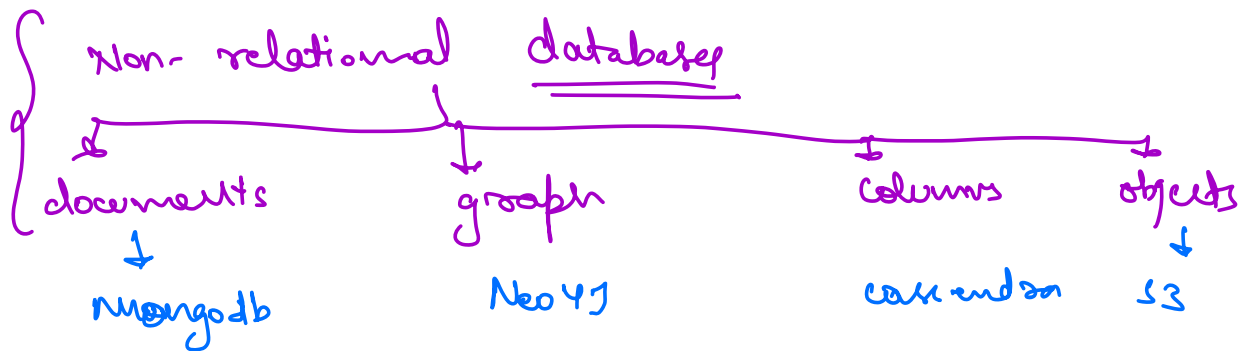
→ student  
→ teacher  
→ instructor

→ A way to store data as a collection of multiple related tables.

students



teachers



Properties of relational databases

① every table may be an entity or relation b/w entities.

Flipkart inventory					model / <u>sum</u>	
price	<u>length</u>	<u>breadth</u>	size			



Marriage		
husband	wife	marriage/date

## Attendance

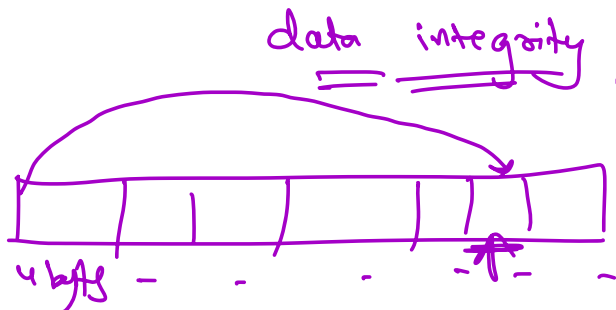
student-id    batch-id    attendance

- ② Every row of table should be unique

## Students

id	Name	age	batch	attendance	
1	Amit	26	2	100%	=
2	Amit	26	2	100%	=

- ③ All values of coln should have the same data type.



students . int

Name	attendance
—	—
—	—
—	—
—	—
—	—

optimizing / fetching values.

## Faster Queries

④ Each value in a coln is **atomic**

a coln cant have

list | map | struct

Ugwal = ( 98714, 98716, 9213 )

SOL 8 → USAs | sons

even though SOL has added these  
USAs / son options they are  
not optimal with these 4  
following queries.

⑤ Coln sequence is not guaranteed.

Students

		<u>phone</u>		<u>addresses</u>	
(id	name	<del>age</del>	<del>batch</del>	attendance	)
1	—	—	(— — —)	51	✓
2	—	—	—	52	
3	—	—	—	90	✓
4	—	—	—	63	✓

<sup>5</sup> result = <sup>1</sup> select \* from students <sup>28</sup> where id = 1,  
<sup>28</sup> result[0] → 10 X

⑥ order of records is also not guaranteed.

select \* from students where attendance > 50.

⑦ Name of every column is unique.

students

<u>id</u>	name	attendance
-----------	------	------------

batches

<u>id</u>	name	st
-----------	------	----

## KEYS

students

id	name	bsp	batch
1	vijwal	100	1
2	Ajoay	19	1
3	Saloni	22	1
4	abhishek	22	1
5	Ajoay	19	1

Each row should be unique.

Keys → help you identify a row

UNIQUELY

- super key
- candidate key
- primary key
- composite key
- foreign key

}

SUPER KEY

Students

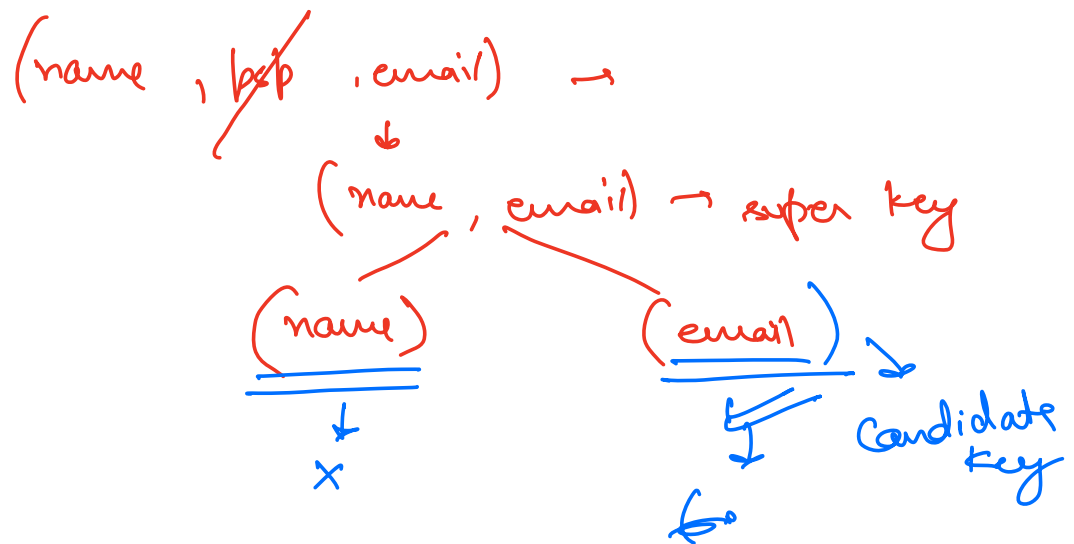
name	psp	email	batch	phone no.
Nauman	80	abc@xyz.com	1	123
Ujjwal	100	xy2@jkt	2	456
Rohit	2	def@xy2	3	789

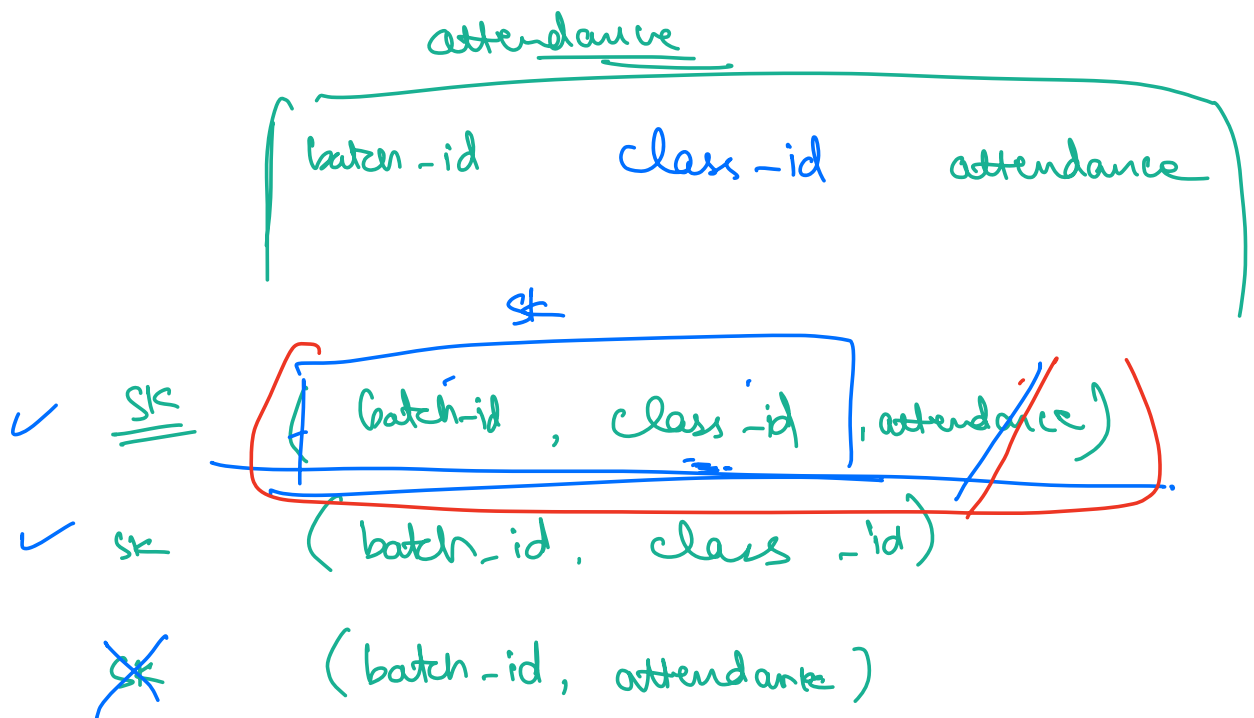
which column values can help us identify row uniquely.

- email
- phone - number
- email & phone - number
- name & phone - numbers ✓
- ⇒ (name, ~~psp~~ & email) ✓
- (name & batch, phone)

super key is a coln or a group of coln values of which is guaranteed to allow you to uniquely identify a row.

Candidate key  $\rightarrow$  A coln or set of colns from which if you remove a coln the remaining part will not let you identify record uniquely  $\rightarrow$  will not remain a super key.



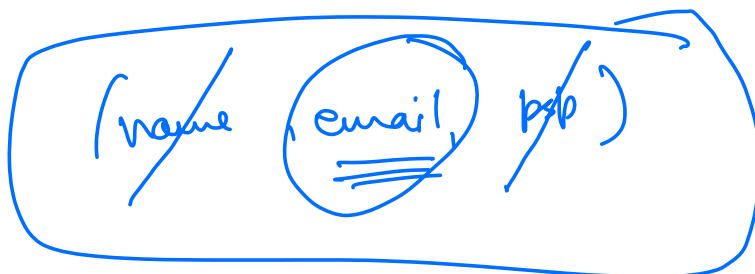


( - - + + + - )

( - - - - )

every CK is a SK ✓✓

every SK is a CK ✗



$(\text{name}, \text{email}, \text{psk}) \rightarrow \text{SK}$

CK  $\rightarrow$   $\text{SK}$