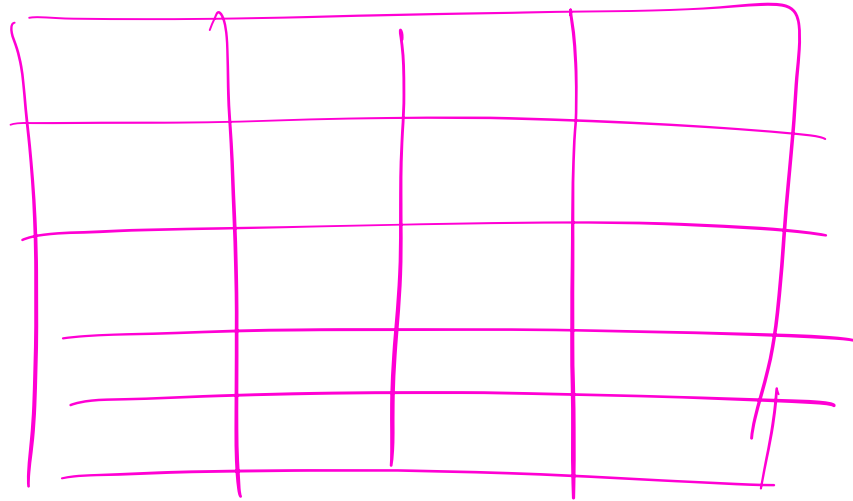# Agenda

1. what is a database?
2. what is a DBMS?
3. Types of Databases:
   - → Rel^n Database
   - → Non Rel^n Database
4. Rel^n Databases
5. Keys:
   - → Super Keys.
   - → Primary Keys
   - → Composite Keys
   - → Candidate Keys
6. Foreign Key.

(SQL)

Create a table   C
Alter a table   U

Structured data ⇒ Excel

⇒ Anything that is in a tabular form

Unstructured data → Diary

⇒ Notes

⇒ Google Doc

Reasons to store data

→ Expenses

→ Groceries

→ To do lists

⇒ Manage Finances

⇒ Planner

Similar to people, Organization need to store data in a structured manner so as to be able to retrieve them later.

Students

name, email, batch, p&p, phone number, Current Company, Resume, Linkdin profile -----

Batches

name, Students, current_topic, st_date, current_inst,

Contests

Question
Leaderboard
Score

→ Create one file per item of data that I want to store.

→ One line per item.

excel to store data.

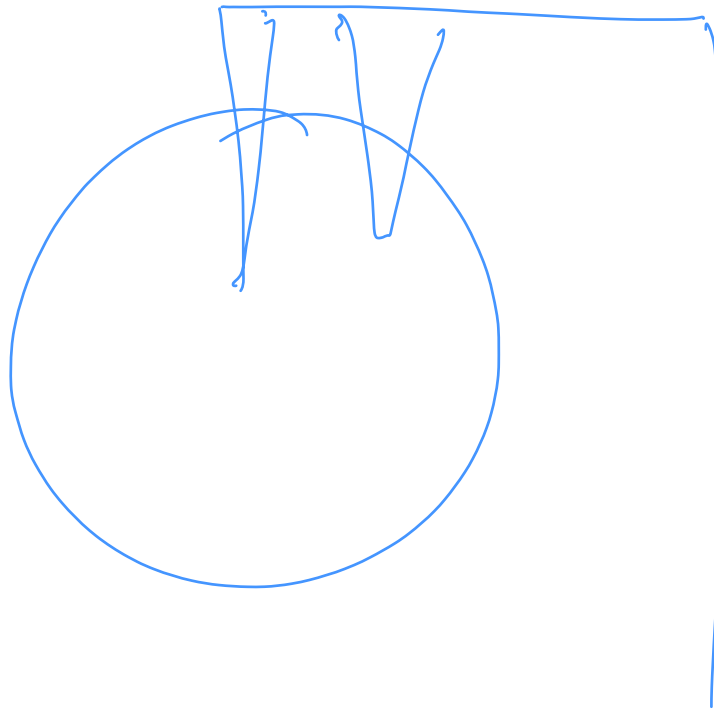Start with a simple text file to store the data.

	name	batches	psp	email	coins	rank
①						
②	Uday	1	80	. . .	1000	3
③	Ashish	1	90	. --	2000	2
④	Dil	ip	95	- - - .	2500	1
⑤	Amit	3	80 "hello"	- - - .	2000	4

Students.txt
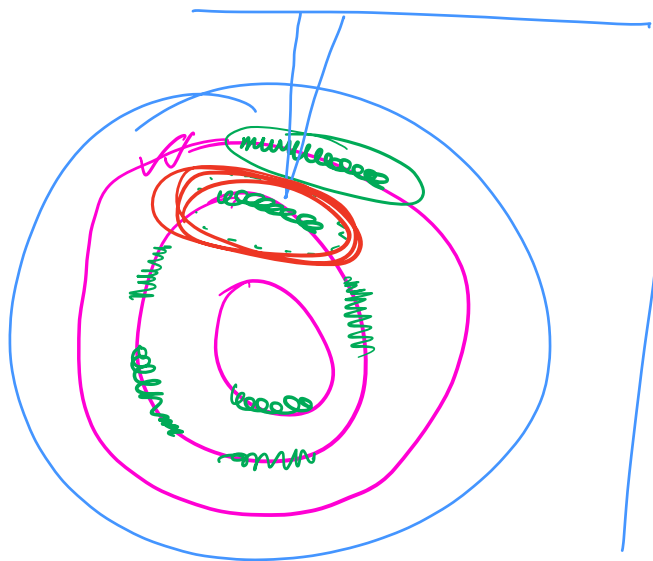
4 ⇒ " Dilip, 2, 95, email, 2500, 1 " ⇒ " Dil (ip)

But, we also need to retrieve the data as well

↓

How ??
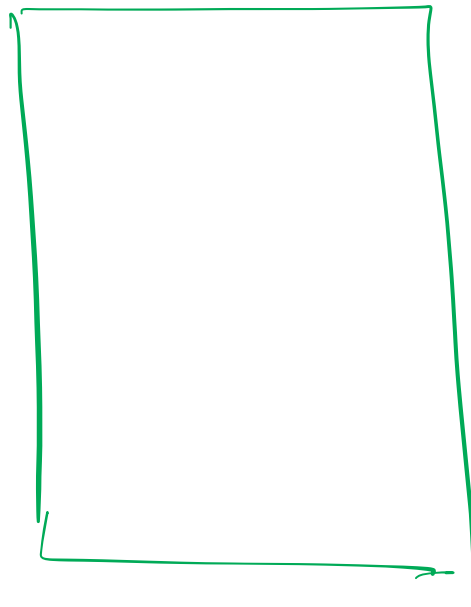
Gramophone.

Only read a file sequentially.

Q= find the average PSP of all Abhinav in Scala

1. go line by line.
2. read each line.
3. Split
4. check if name is Abhirav.

Possible issues?

1. Inefficient → Reading data is very slow. $O(n)$ for any type of operation.

2. Security
   → Access control (user level per^m)
   → Data is readable/writable by anyone

(3) Data Integrity.

→ wrong datatype.

→ wrong formatting. → any type will mess up the entire row.

(4) Concurrency

→ No two people can read any file at the same time.

login to Scaler.com

↳ email
passwords
name, coins, streak.

what is a Database?

(Airbase) → plane parking.

(Military Base) → Army stays.

(Data Base) ⇒ Data stays
→ where Data is stored.

Database : <u>Collection of related data</u>

Data about items that are related to each other based on the business use case.

Scalar database

(Students)   (batches)   (inssructors)

(Mentors)   (Contests)   (TA)
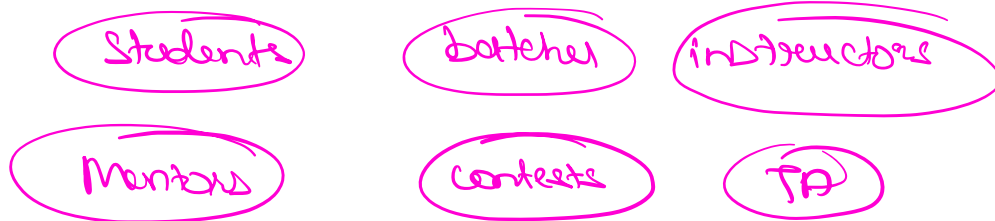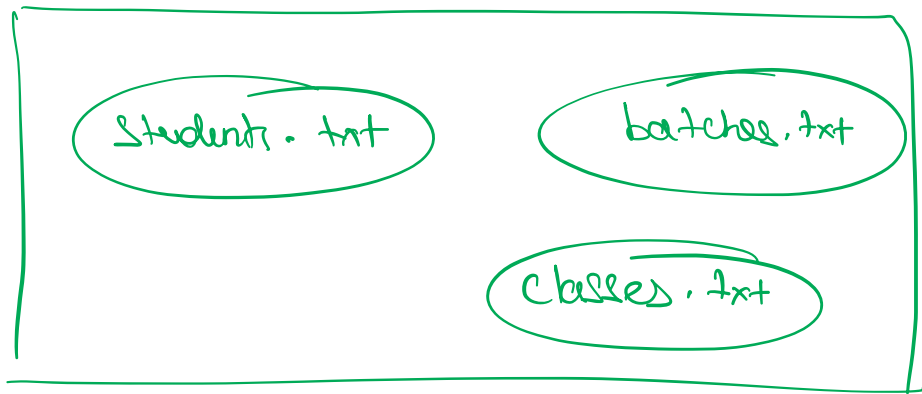
what is a DBMS (Database Mangement System)

Software that will (efficiently) write and retrieve data, alongwith features that maintain

data integrity.
Security
Concurrency

Q find the student with psp b/w 50 and 90
DBMS will use its advanced data streuctures to solve any Query.

Types of Databases:

name	30/6	3/7	5/7
Nohsin	✓	x	x

Register.

Tables are the most natural way to store any data.

Based on whether DBMS uses Table to store its data, it is categorised into 2 types

DBMS

Table = Rel^n DBMS

→ Stores related data in tables.

Non Rel^n DBMS.

Store related data in non-tabular form.

→ graphs
↳ Columnar → Cassandra.
→ Document
→ Key value pair

Students

name	email

Columns are attributes
rows are data entry.

Shir → Govinth
Govinth → Vish
Shir → Vish
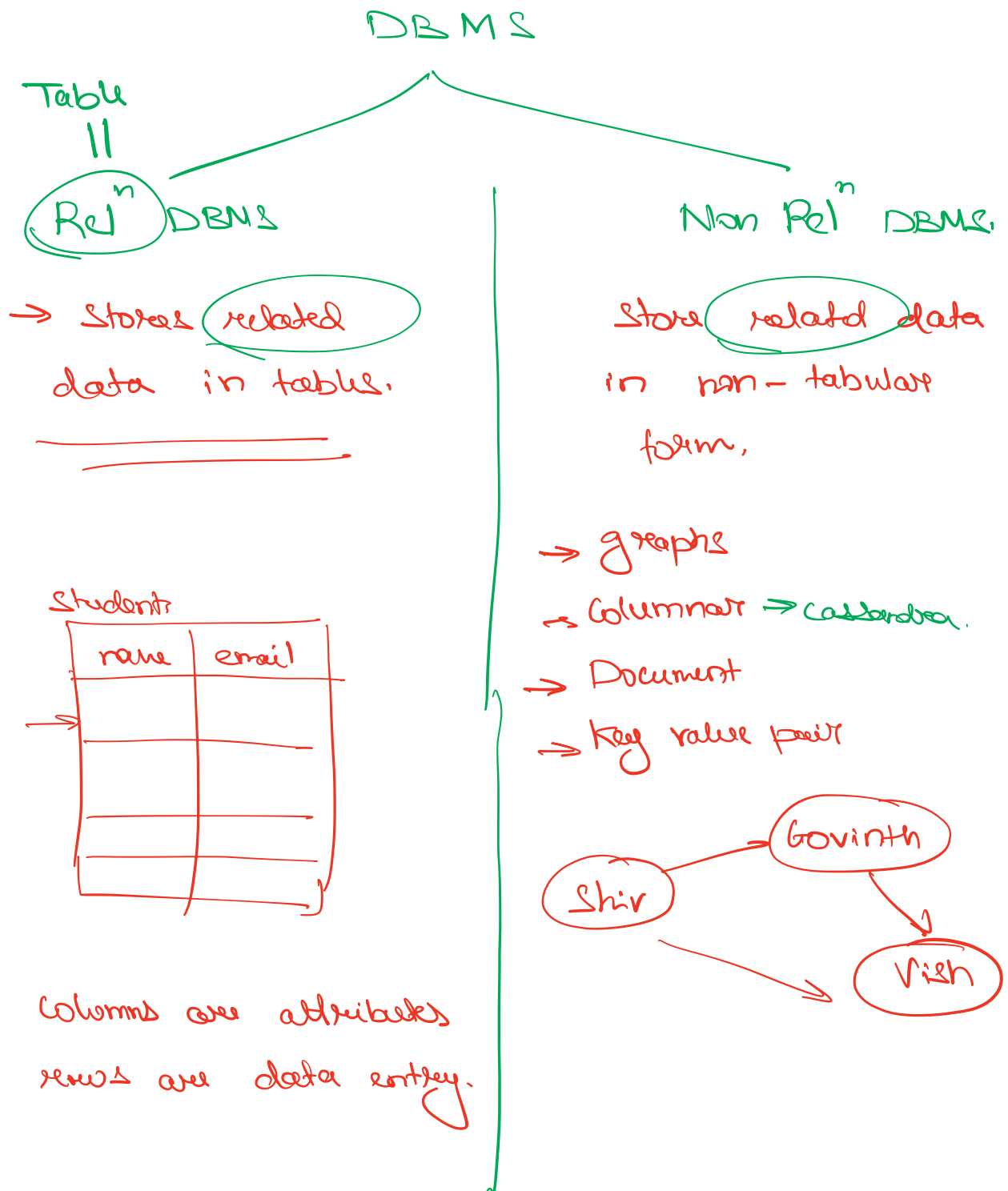
name	post_url
Somesh	pic_url
Somsh	pic2_url

Break till 10:43 PM

when to use which DBMS

HLD

Properties of Rel^n Databases. = DBMS

1) Rel^n database stores tabular data that are (related to) each other.

Scale: Students, batches, instructors. ----

2) Every row is unique.
No two rows have the same values for all the columns.

name	age
Tushar	24
Prasun	25
Prasun	25

Problem with duplicates
↓
identity cannot be known
(who is who)

3) All the values in a column have to be of the same type else data integrity is broken.

4) values in a cell are (atomic)

Smallest indivisible
unit

→ Cannot have an array, map, json....

name	phone numbers
Saharsh	{ ——, ——]

Problem : Even Other SQL that do allow
storing lists, they are generally slow.

4B	10B
id	names
1	[Amit, Shah] → 14B
2	[Rishi, Singh,-] → 14B
3	Dilip → 14B

100 ↓

| 4B | 10B | 4B | 10B |
| 1 | 2 |

find 5th row of this table.

$$100 + (i-1)^* 14$$

5) Column sequence in queries is (not) guaranteed.

Students

name	batch	psp	email

name, batch, psp, email. ⎫
email, psp, name, batch   ⎬ Same

Select (name, psp) from Students .

ASSUME

→ | A ss | U | ME |
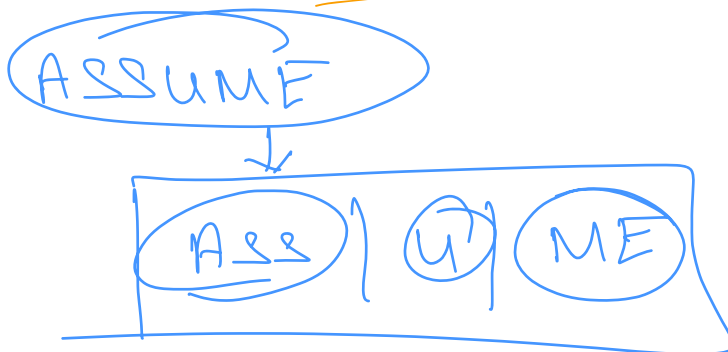
6) Order of rows is not guaranteed.
   Select * from students

Amit	Rishi
Rishi	Amit

to fix an order, we have to use
   ORDER BY (sorting on name)

Amit
Rishi   every time.

why is the order not guaranteed
→ Inorder to have efficiency.

7) No 2 columns can have the same name.

Students table

Roll No	name	name

Super Keys.

Every row should be unique

Key: a col^n or a set of col^n in a table whose values are guaranteed to give you a unique row.

Primary Key $\cong$ Key $\cong$ Super Key.

Students

name	PSP	batches,	phones.	email

Super Keys.

email   ✓
phoneno.   ✓
name   ✗
psp   ✗
batch   ✗
email, name   ✓
email, phono.   ✓
email, psp   ✓
psp, batch   ✗
psp, name   ✗
email, ph, psp, name   ✓

Super Keys

Super Keys

name	ph1	ph2	email 1	email 2 .....