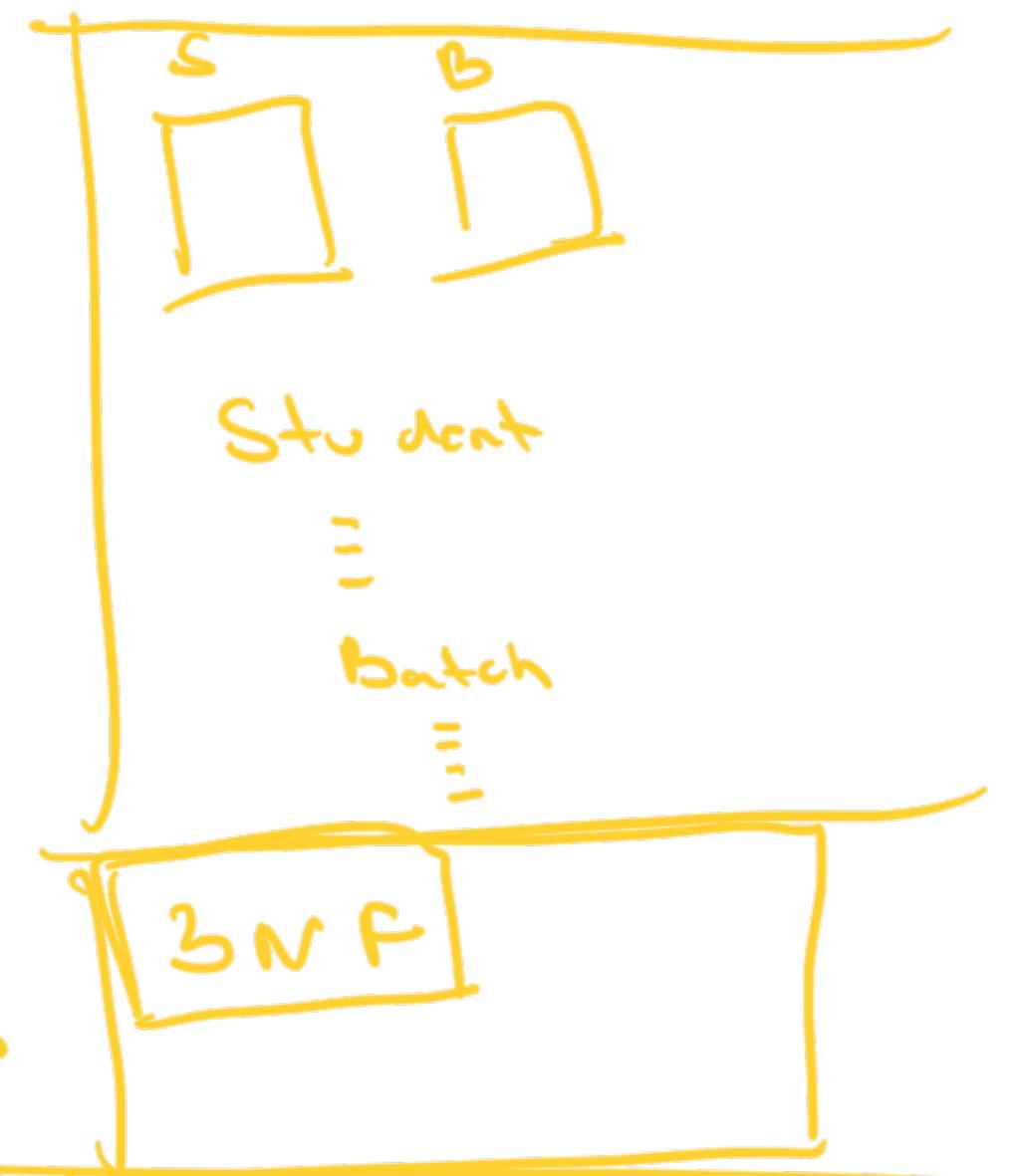


SQL - III - Data types and normalisation

Agenda

- ① Data types
 - ✗ String
 - ✗ Numeric
- ② Data Normalisation
 - ✗ Anomalies
 - ✗ Normal forms

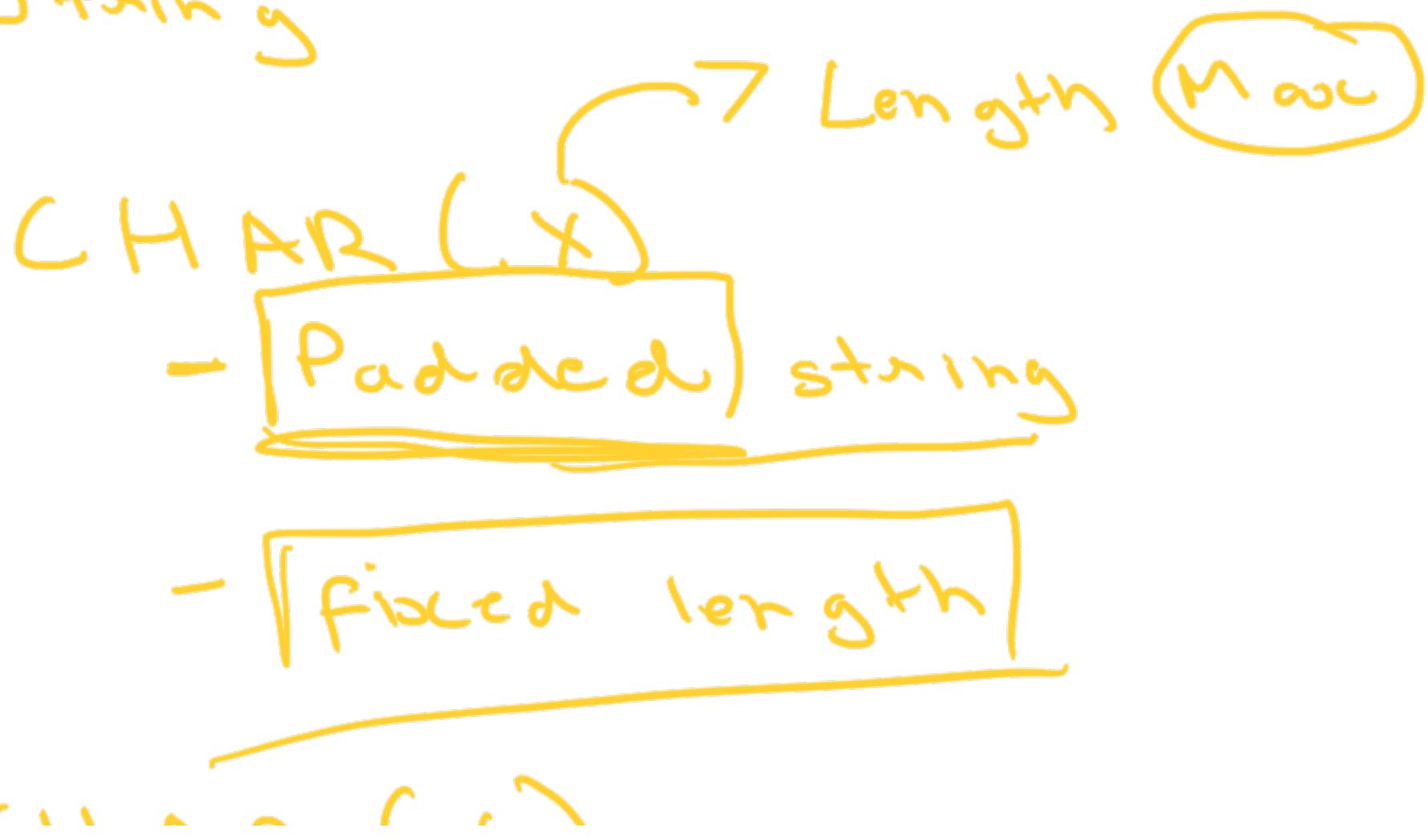


① Data types

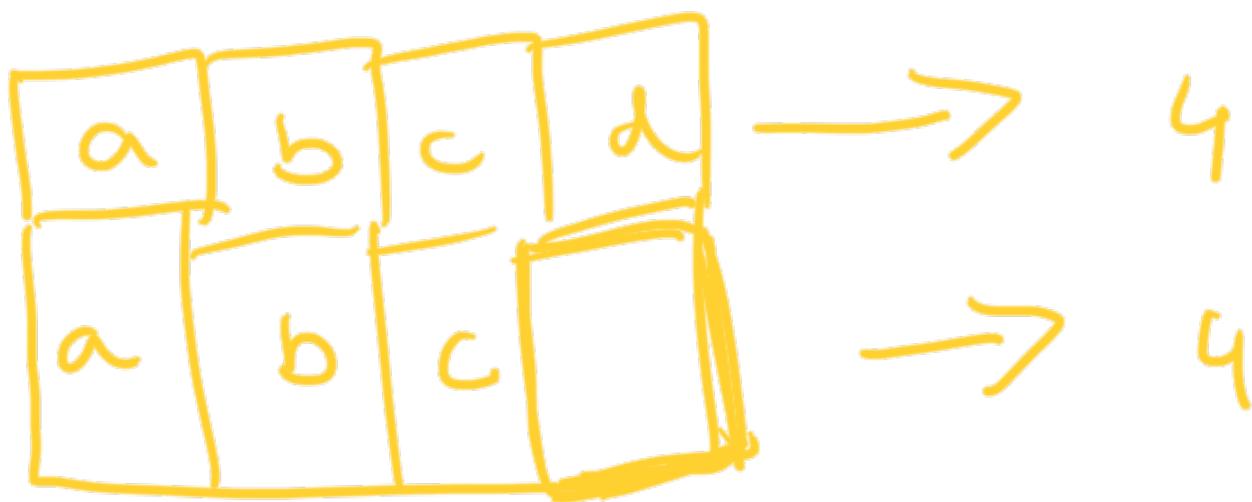
① String ↴

- 1 numeric ↗
- 2 boolean ↗
- 3 Floating ↗
- 4 Date ↗
- 5 BLOB ↗

① String



CHAR(4)
→ "abcd"
a b c



100 - 10 90

CHAR(4)
→ "a"
..

↳ "a _ _ _"
↳ Length - ④

Use cases

- Grade - A, B, C, D
CHAR(1)

→ Gender

→ Pincode

→ Country code

→ PAN / Aadhar

0 - 253



aF - - -



Varchar

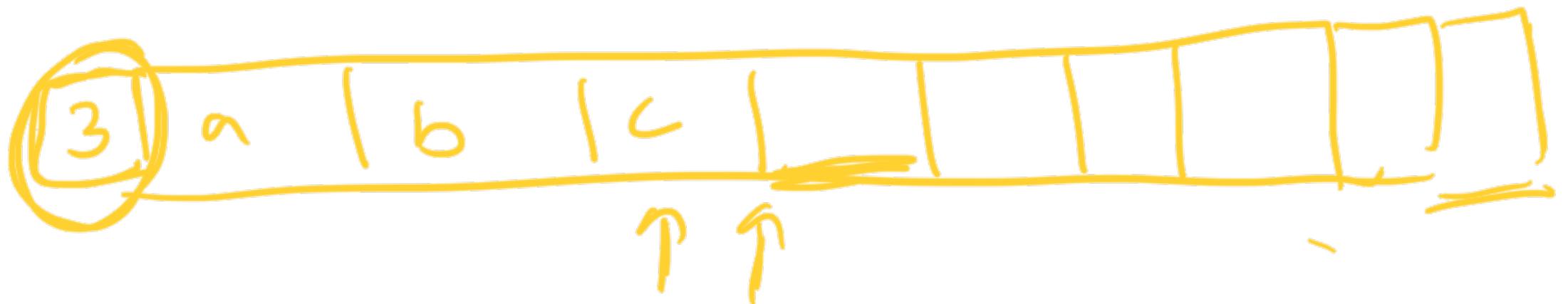
- Variable character
- VARCHAR
- = Not padded
- (4) - @

0-65,555

CHAR -
VARCHAR -



VBR CHAR (235)



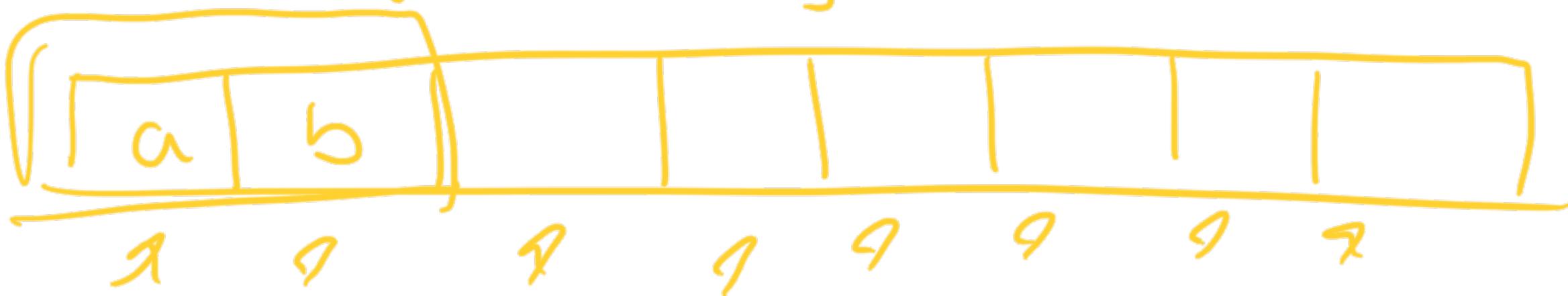
a b c \rightarrow 3 a b c

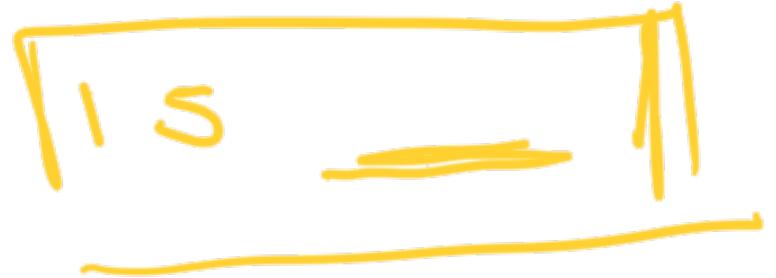
a b c d e \rightarrow 5 a b c d e

1 0 0 ... \rightarrow 1 0 0

} Run length

V CHAR C(10)





S abcde

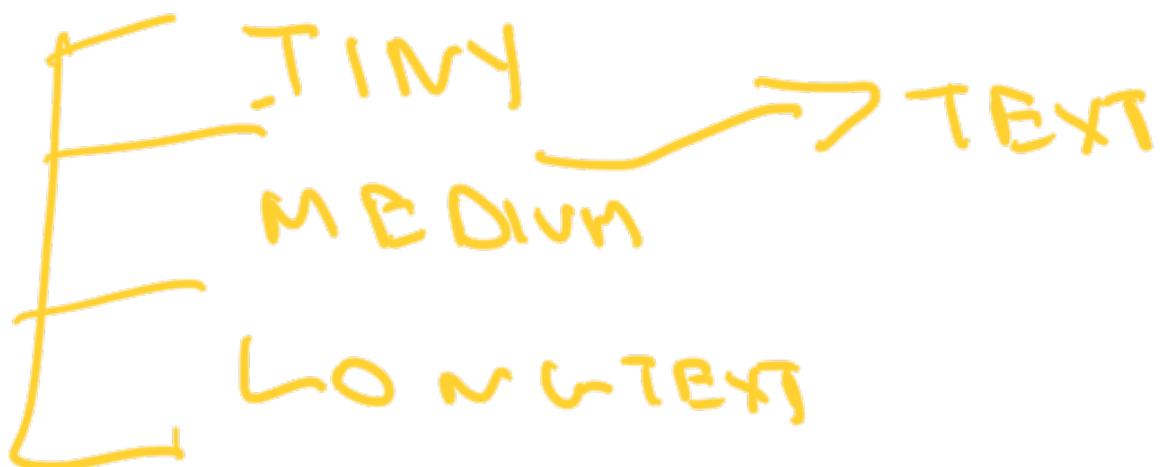


1 Byte





* Text



Tiny	=	256B
TEXT	=	64 KB
MEDIUM	=	16 MB

Long - 4 GB

Text cannot be indexed

- costly not performant

Ans

S3 - Filestore

Numeric

TINYINT

SMALL INT

MEDIUM INT

INT

Storage

1B

2B

3B

4B

Age - 100 127 3

127

1 bit | 8 bits

$$\begin{aligned} 1 \text{ bit} &\rightarrow \boxed{255} \\ &\Rightarrow -127 - 128 \\ 2 \text{ bits} &\rightarrow 32 \text{ K} \end{aligned}$$

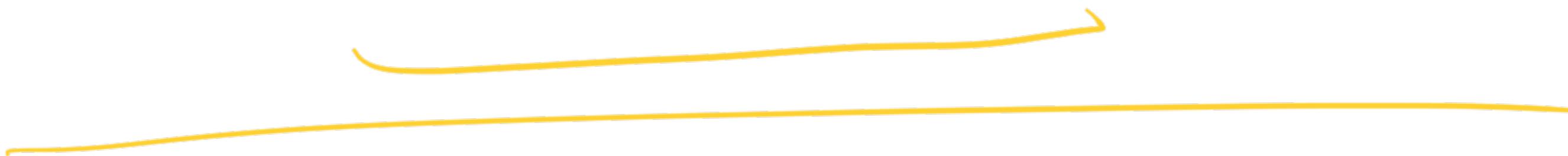
1 bit = 8 bits

‘ ’ 8 bits

$$2^8 = \underline{\underline{256}}$$
$$\frac{0 - 255}{12}$$

= 1

-2 - 12 / - 128



Floating point

① DECIMAL (P,S)

precision



length

scale



decimal points



$$\begin{array}{r} 1 \\ p = 5 \\ s = 2 \end{array}$$

DECIMAL (5, 2)

$$100.98 \rightarrow 101.92$$

$$\begin{array}{r} 302.505 \\ \hline p = 6 \\ s = 3 \end{array}$$

$$302.505 \overline{)6}$$

Truncate

S

302.505

302.505 DC(6,5)
P → length → 6
S → decimal → 3

DC(6,5)
302.500

② Float

- approximate

JEE 754



Epsilon \rightarrow buffer

$$1.9 - 0.9 \pm \delta = 1.0 \pm \delta$$

$$0.9001 - 0.901$$

$$1.9 - 0.9 \neq 1.0$$

$$1.89 - 0.91 \pm \delta = 1.0 \pm \delta T$$

BREAK

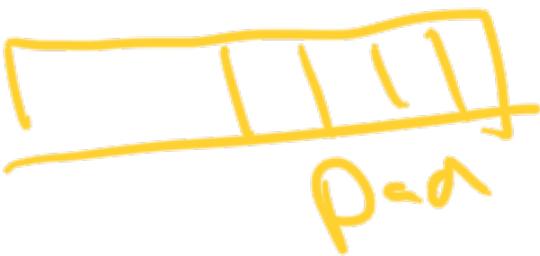
$$6:00 - 6:05$$

10:35

VAR(15)

maximum length

CHAR



VBR CHAR



Data Normalisation

Anomalies - Error

- Outlier

Student - Batch

Student + Batch

ID | NAME | EMAIL | BATCH_ID | BATCH_NAME

Insertion Anomaly

ID	NAME	EMAIL	BID	BN
1	Muruk	nuc	1	April 23

→ 1 | | June 25

Deletion An anomaly

ID	Name	BID	BNBnE
1	Sherlock	1	S
2	Mary	2	S
3	Kit with	0	O-



Update Anomaly

SID

1

2

3

Name

A

B

C

BID

1

1

2

Batch A

BNAME

Batch

AB

CB

Anomalies

- * ① Insertion - We cannot create a batch w/o students
- * ② Deletion - When we delete a student then we also delete batch
- * ③ ...

⑤ Update operation - Use to human errors
→ Miss updating all rows
→ In consistency

Normalisation → Avoid anomalies



BCNF Boyce - Codd

5NF

6NF

INF

- Only atomic values
- single valued
- No lists, collections

Students

ID	Name	Phone number
1	A	[+91 ... , +44 ...]
2	B	+91 ...
3	C	+44...

①

Add a new row ✗

②

Add a new column ✗

③

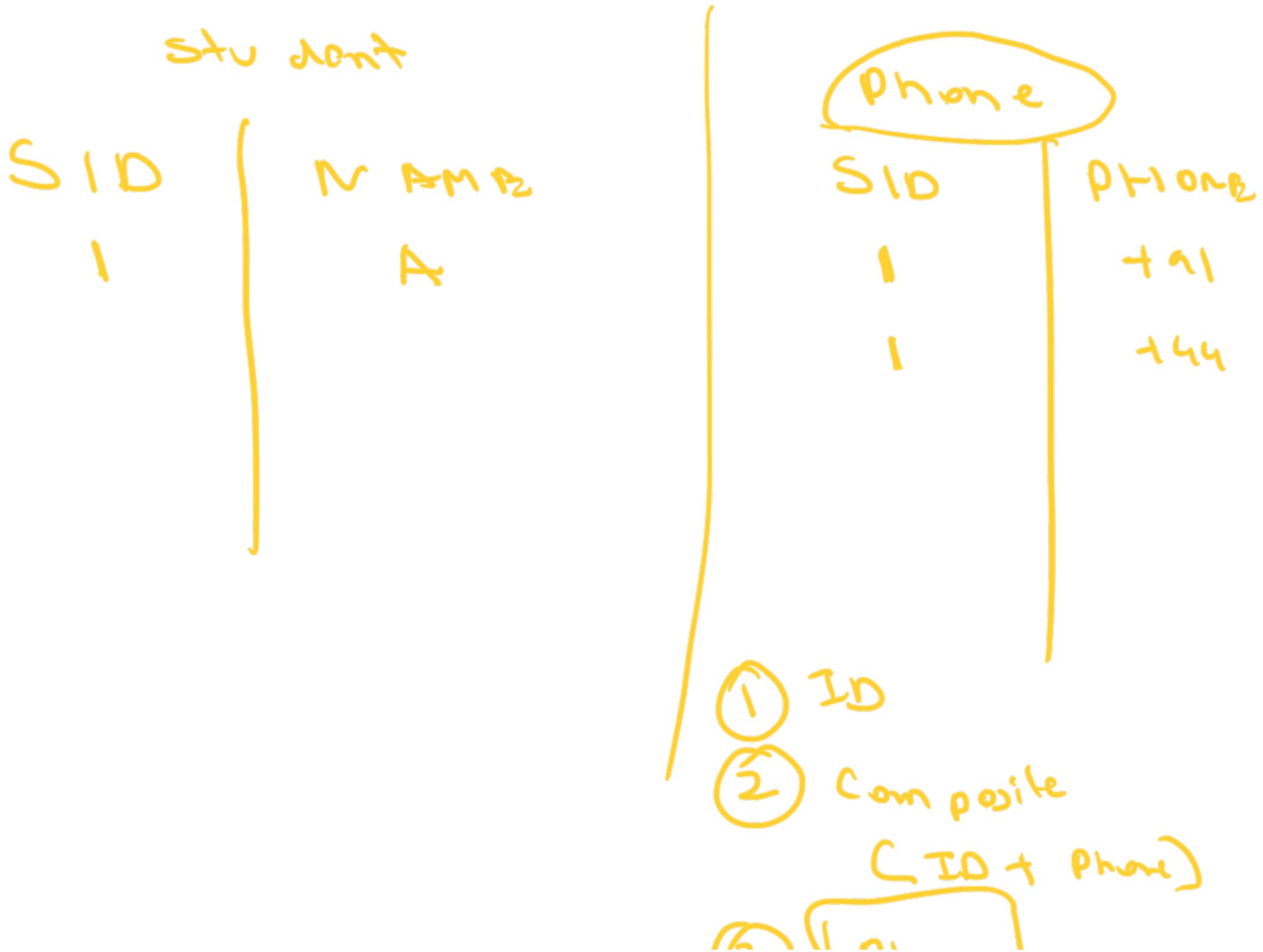
mapping table



④

Sparse column

- (2) How to decide # of commas
- (3) Querying is not easy



(b) iPhone

2NF ✓

①

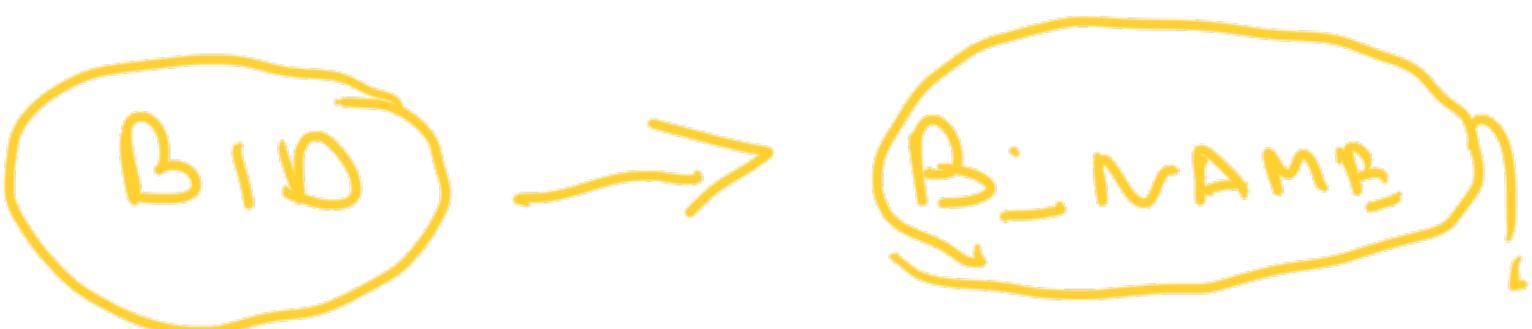
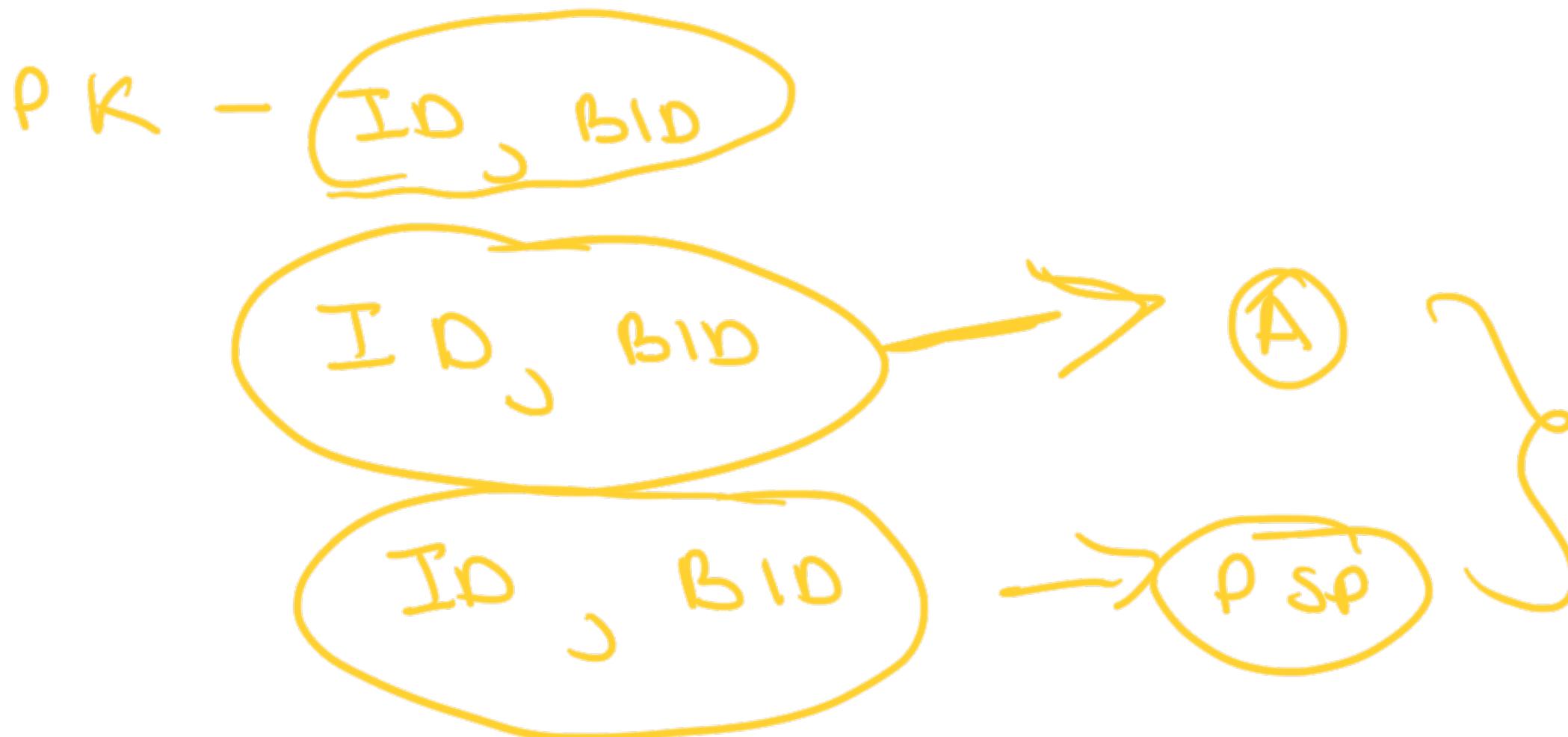
1NF ✓

②

No partial dependencies

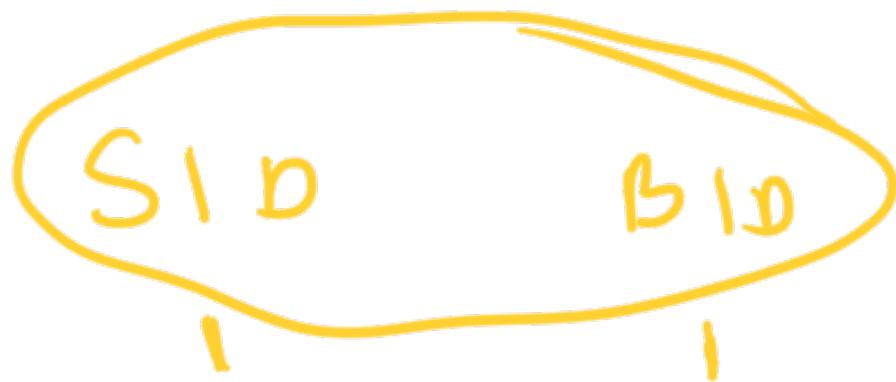
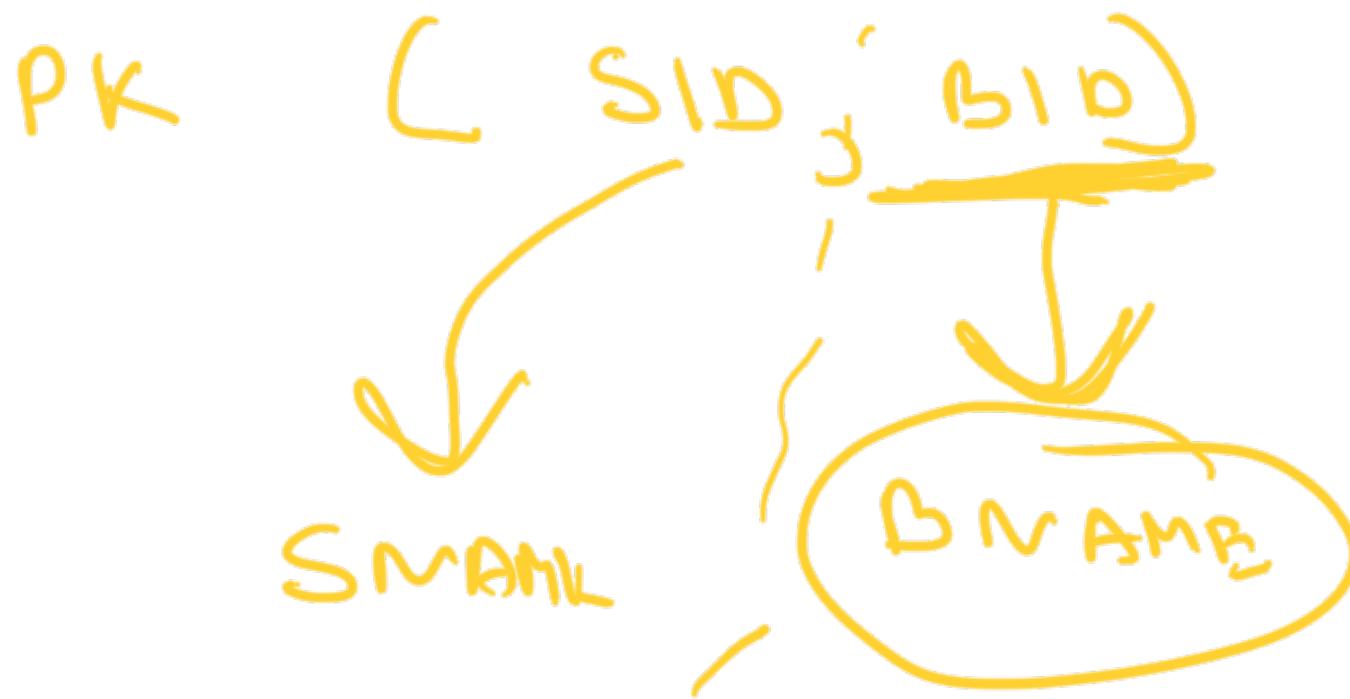
an attribute should not
depend on part of the PK

ID	BID	NAME	PSN	BONANZA
A	B	C	D	E



Partial Dep.

Solve - Mapping table



1 2
2 1

Student

ID	NAME	PSO
----	------	-----

BATCH

ID	NAME
----	------

Student batches

SID	BID
-----	-----

MySQL

Client

- ①
- ②

Workbench

Tdbe Plus



API
↳ CRUD

Frontend - Backend

Java - Spring Boot

Python - FastAPI