

Problems Before Normalisation

110 → 112

(com)

the rows

All and update

IT → Inf Tech
Time

↓
100's

EID	Employee Name	Dept. ID	Dept. Name	SuperID
101	Shashank	A	CS	105
102	Suvedha	B	IT	110
103	Satyam	A	CS	105
104	Sagar	A	ECE	106

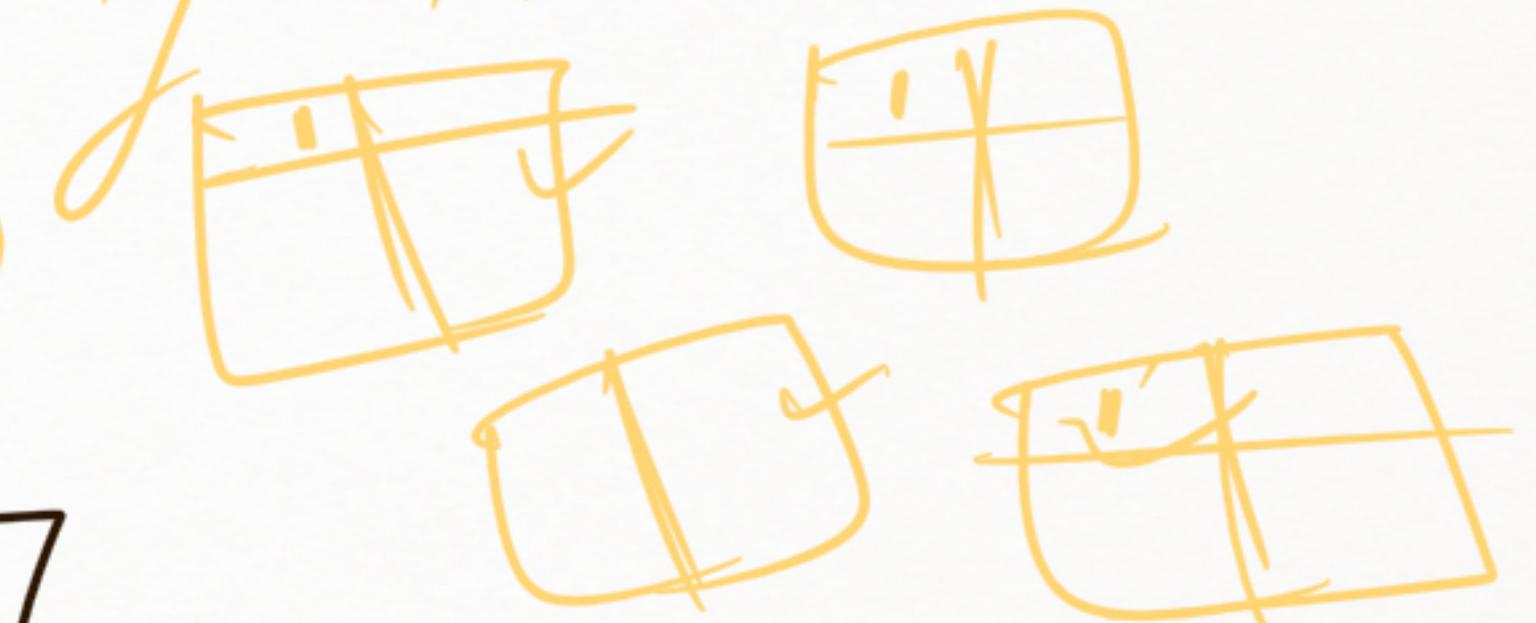
Employee

① Redundancy => Waste/Unnecessary some data is repeated

- ② Anomalies =>
- ① Insertion
 - ② Deletion => All emp of Dept A are going to leave.
 - ③ Update => IT => Information Tech.

Solⁿ => Normalisation
=> Breaking Down Big Tabk into smalles
ones by following some rules.

Why Rules?
=>



EID	EN	DID
101	Shashank	A
102	Sneha	B
103	Satyam	A
104	Sagar	A

Emp

10⁰⁰⁰⁰

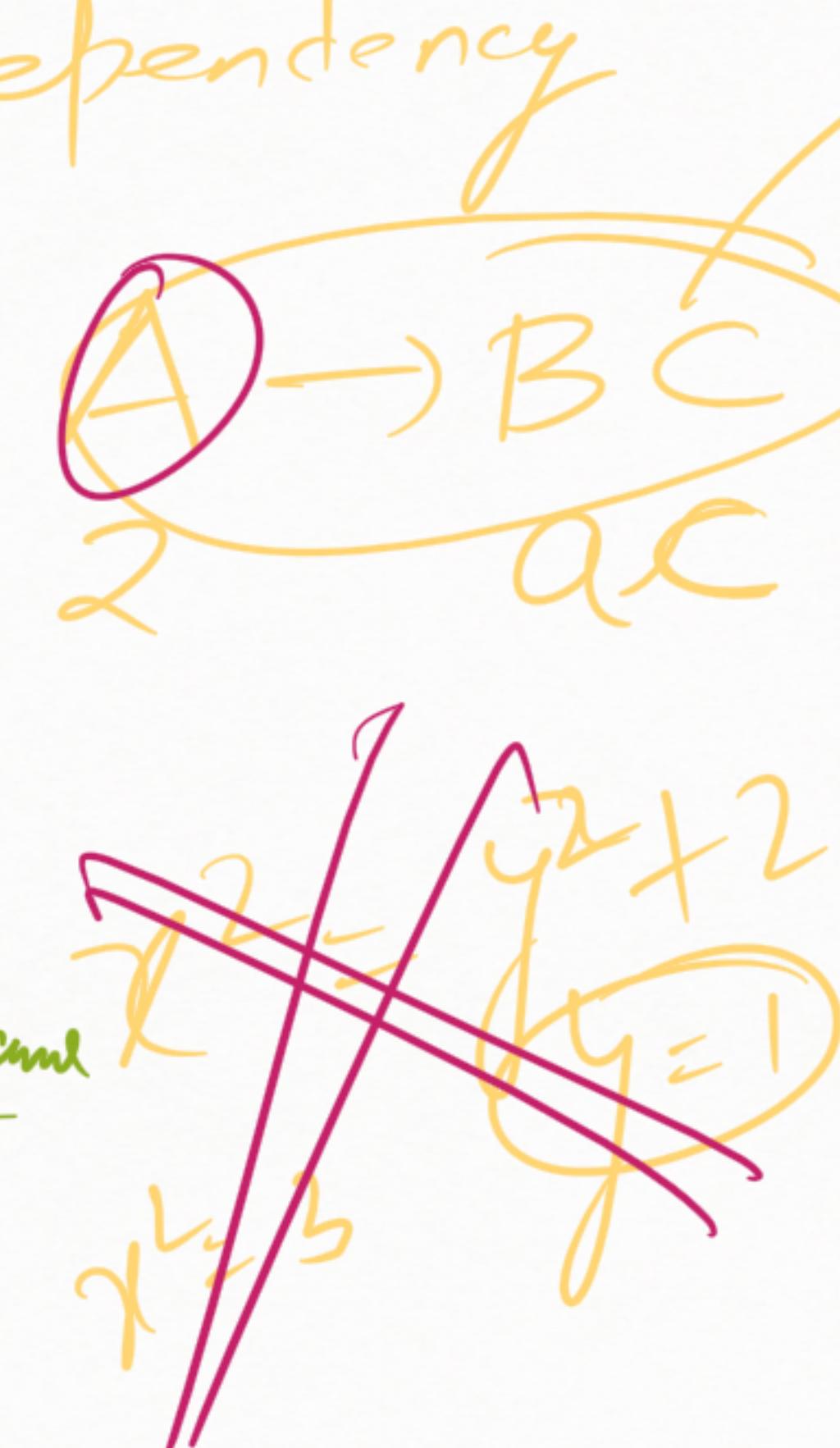
DD	DN	SID
A ✓	CS	DS
B ~	IT	110 ✓

Dep

=> So many
tables?
Query time

Functional Dependency

A	B	C
1	a	b
2	a	c
3	d	e
4	f	g



Given the value of A, you will be able to identify B & C.

This is not exactly the key, but there is a relationship b/w Key and FD's ✓

A	B	C
a a b	1 2 2	e f g h

$A \rightarrow B$

$a \rightarrow 1$

$b \rightarrow 2$

A	B
a b	1 2

Whenever
A repeats B
also repeats

a appears 100 times -

1 appears 100 times

A	C
a a b	e f g h

Normal Forms =



1NF



Ultimate Aim
~~In Real Case~~

BCNF

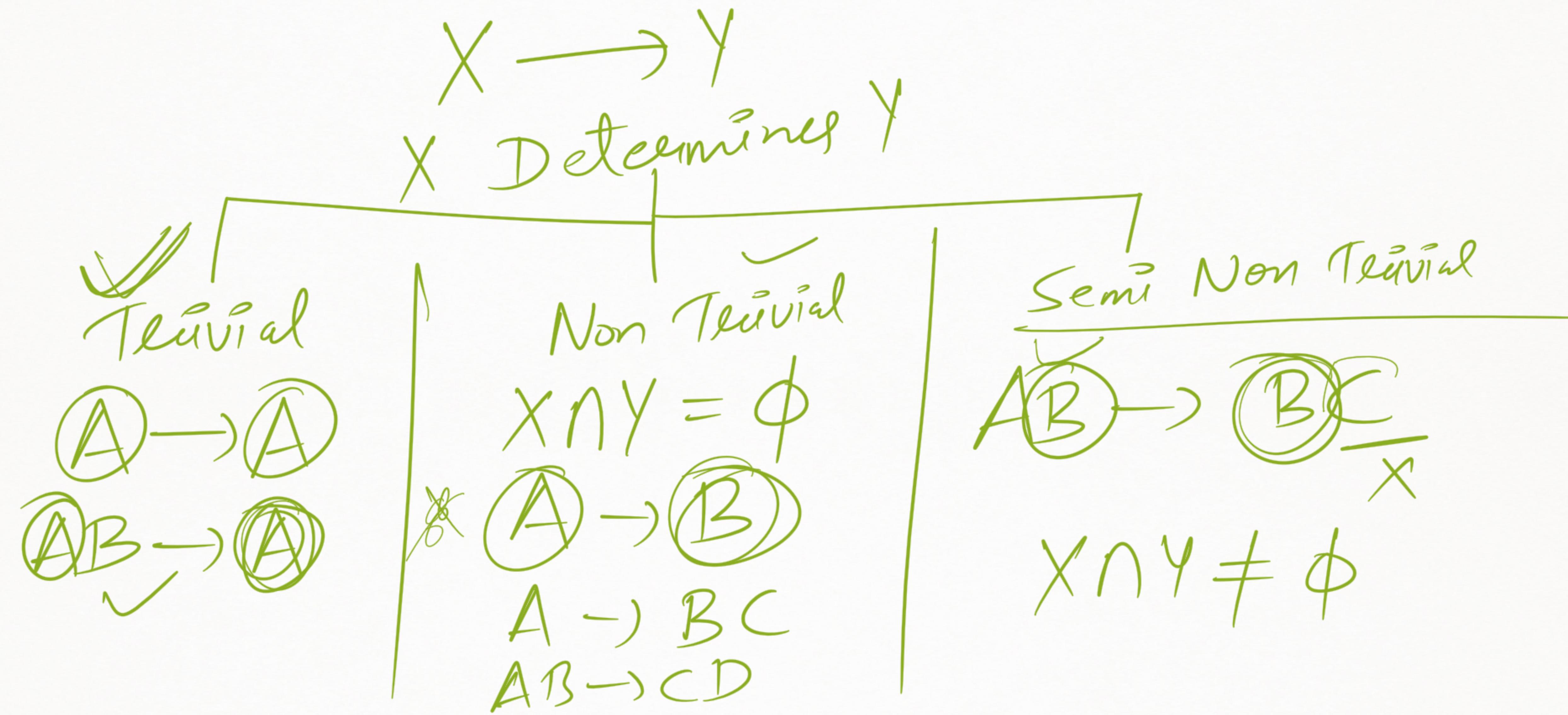
3NF

Boyce Codd



2NF





Rule out FD based on table data

①

Eid	Ename
1	a ✓
2	b ✓
3	b ✓

DK

~~Eid → Ename~~

~~Ename → Eid~~

b → X 2,3

lots of
Data

②

A	B
1	1
2	2
2	2

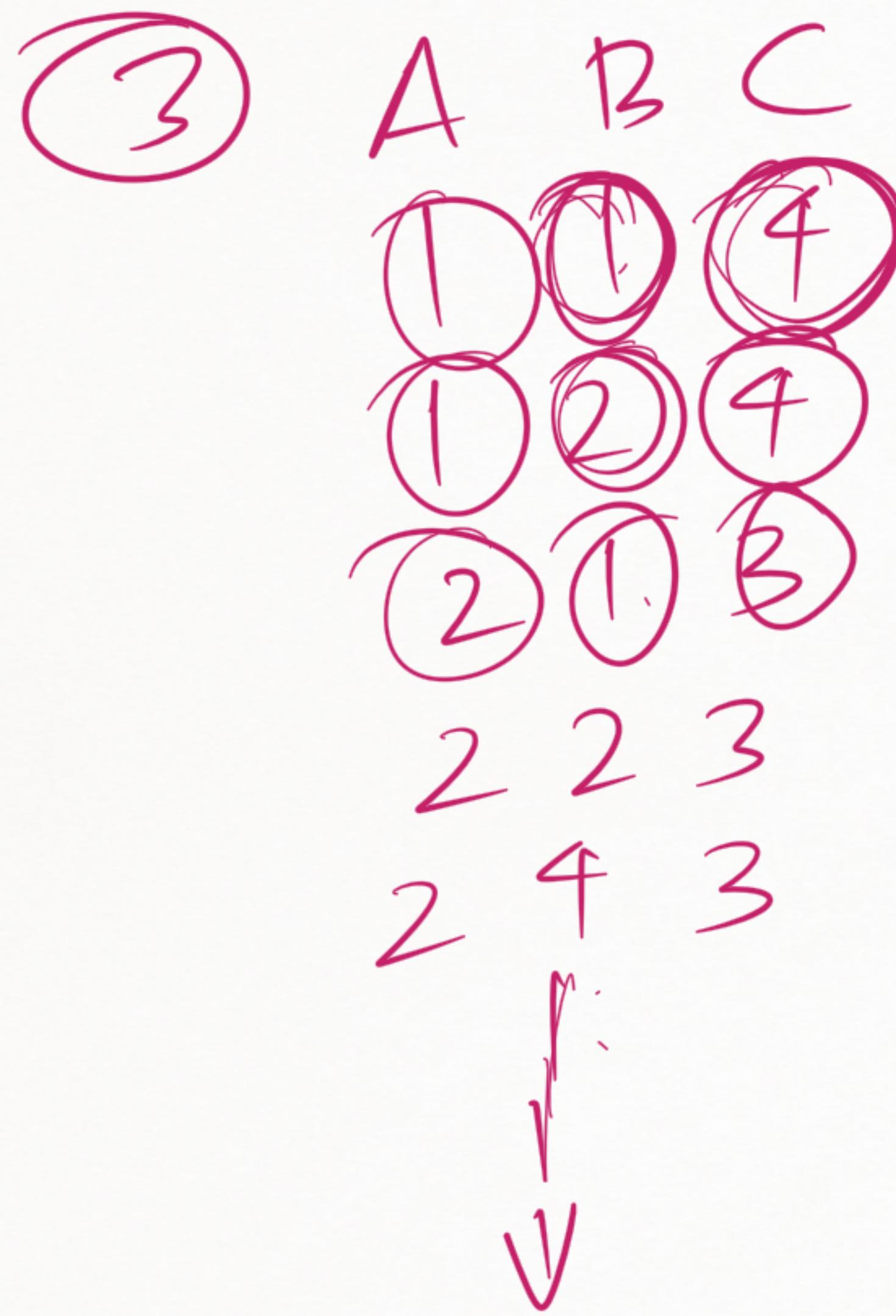
① A → B X
② B → A X

option

Wrong FD

I Don't
Know

Correct FD



FDS

① A → B X

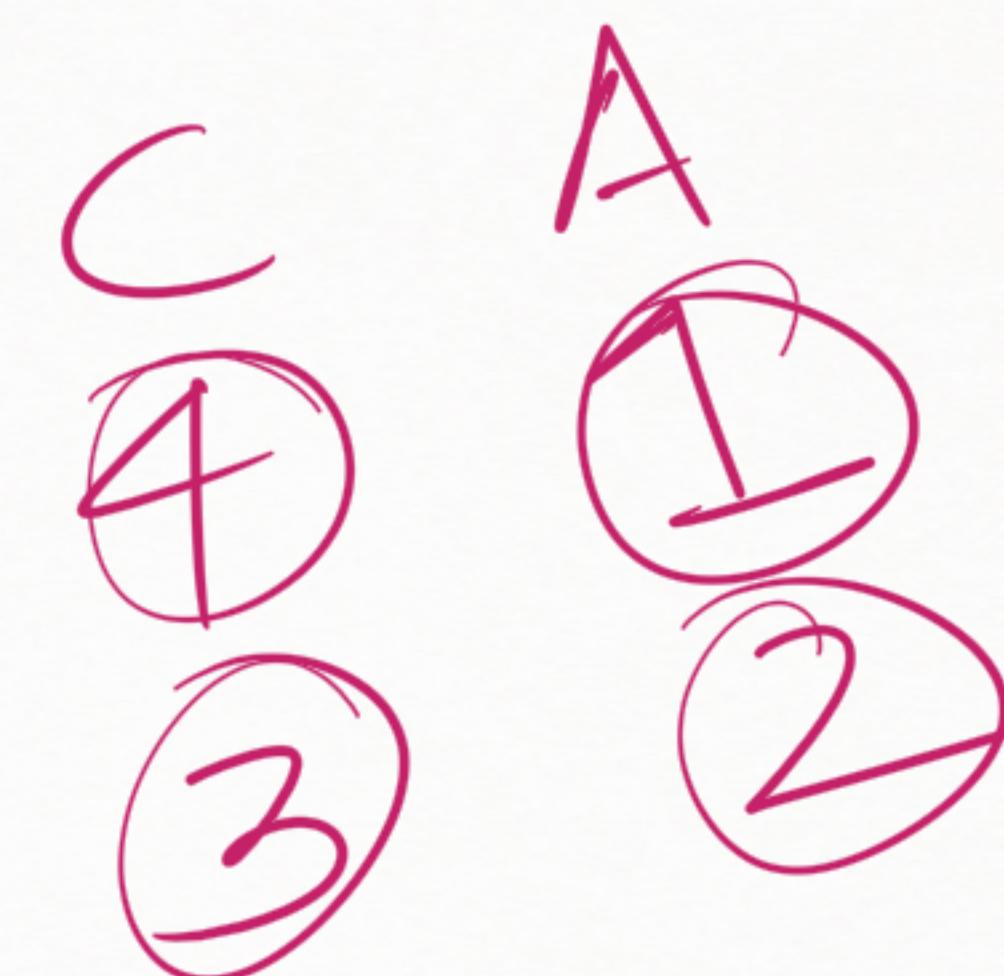
② B → C X

③ B → A X

④ C → B X

→ ⑤ C → A DK

→ ⑥ A → C DK



Formal Definition of FD
ex $\underline{AB} \rightarrow \underline{CD}$

$X \rightarrow Y$
Collection of Attributes Collection of Attributes

	A	B
t ₁	1 1 2 3 3	a a b c c
t ₂	1 1 2 3 3	a a b c c

$A \rightarrow B$ ✓

X	Y
<p>If $t_1 \neq t_2$ are same here</p> <p>If $t_1 \neq t_2$ are diff here</p>	<p>They must be same here also</p> <p>They MAY OK MAY NOT be same here</p>

Closure set of Attributes

FD's

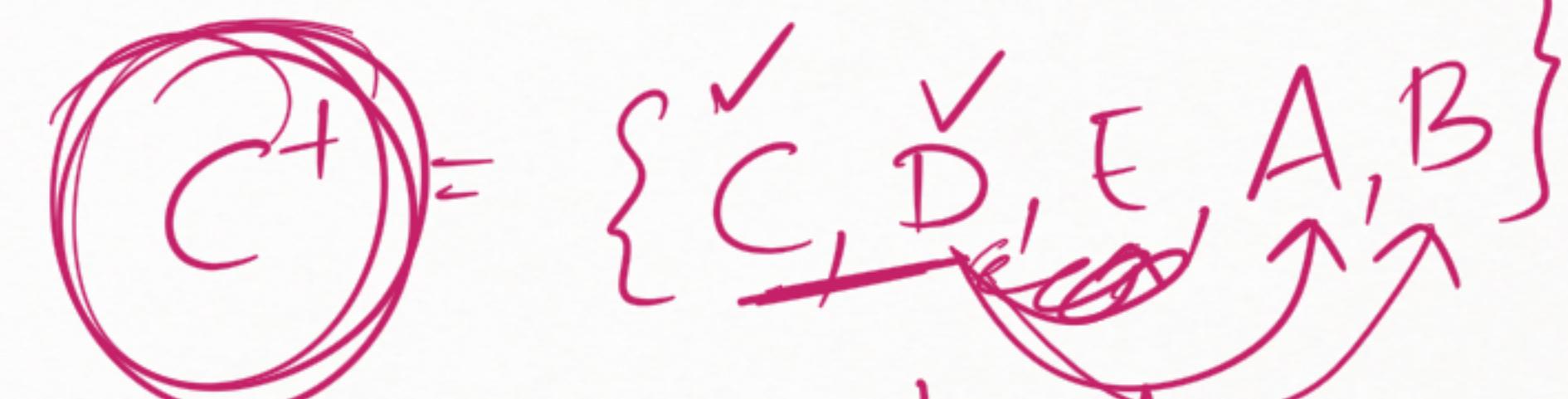
$R(A, B, C, D, E)$																														
<table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>2</td> <td>2</td> <td>2</td> <td>2</td> <td>2</td> </tr> <tr> <td>3</td> <td>3</td> <td>3</td> <td>3</td> <td>3</td> </tr> <tr> <td>4</td> <td>4</td> <td>4</td> <td>4</td> <td>4</td> </tr> <tr> <td>5</td> <td>5</td> <td>5</td> <td>5</td> <td>5</td> </tr> </tbody> </table>	A	B	C	D	E	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
A	B	C	D	E																										
1	1	1	1	1																										
2	2	2	2	2																										
3	3	3	3	3																										
4	4	4	4	4																										
5	5	5	5	5																										

- $A \rightarrow B$
- $B \rightarrow D$
- $C \rightarrow DE$
- $D \rightarrow AB$

$$A^+ = \{A, B, D\}$$

Closure Set of A

$$B^+ = \{B, D\}$$



$$D^+ = \{D\}$$

$$AC^+ = \{A\}$$

$$E^+ = \{E\}$$

$$AD^+ = \{A, B, D\}$$

$$CD^+ = \{C, D, E, A, B\}$$

Candidate Key and Super Key

- ① Candidate Key =) Given an Attribute or a collection of attributes, we will be able to uniquely identify a row.
⇒ If collection of Attribute, than those attribute has to be Minimal.

* Primary Key is just one of the Candidate key that we wish to choose.

ID1	ID2	Name	Class
1 ✓	A ✓		
2	B		
3	C		

Any 1 can
become a Primary
Key. [Candidate Key]
PRI

Cand. keys = $(\{ID_1\},$
 $\{ID_2\})$

PRI ID_1

~~$ID_1 + Name + Class$~~
 Super Key

~~Break~~

Super Key = Candidate Key +
Waste

example = ID + Name
 ↑ ↑
 Candidate
 Key Waste

Sub	Teacher Name	Dep	Y
A	Palash	IT	e
B	Ankit	CS	e
C	Mudit	CS	e
	Palash	Maths	f

Uniquely identify each Row

