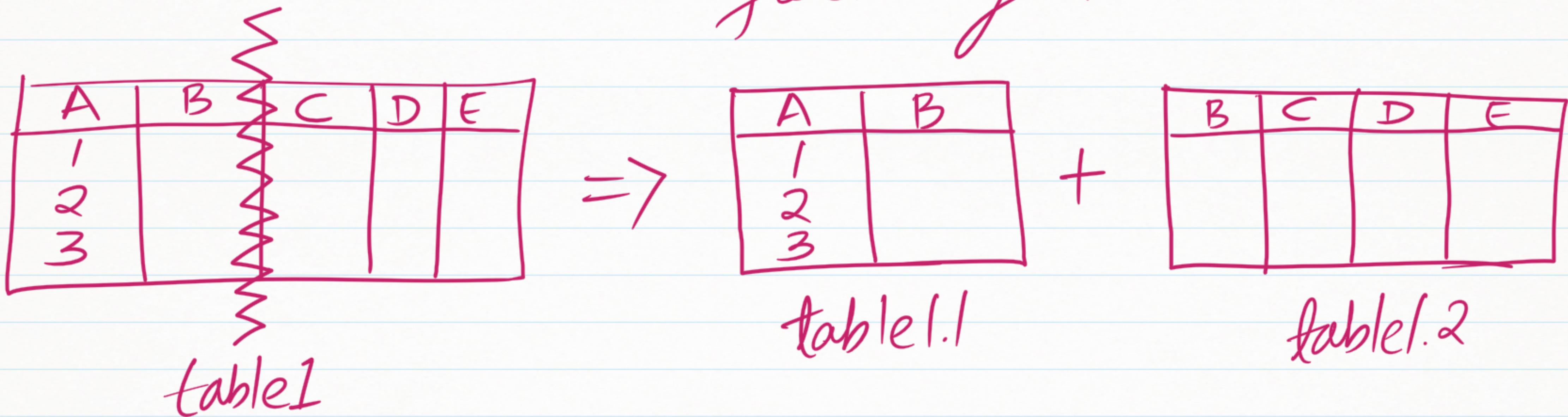


Normalization = Breaking down of table  
along the columns by  
following some rules.



## # Why we Need Normalization

~~Millions~~

EID	EName	Dep ID	Dep Name	Super ID
101	Bhavya	A	CS	101
102	Vishal	B	IT	102
103	Kanishka	A	CS	101
104	Nilesh	B	IT	102
105	Revathi	A	CS ✓	101

① Redundancy  $\Rightarrow$  Repeated Data

② Anomalies  $\Rightarrow$  ① Insertion - A new Employee  
② Deletion - When I delete all Male Emp  
③ Updation - CS  $\rightarrow$  Computer Science.

If Normalization is done properly,  
we can avoid these anomalies  
& redundancy.

EID	EName	DID
101		A ✓
102		B ✓
103		A ✓
104		B ✓
105		A

+

DD	DN	SID
A	CS	101 ✓
B ✓	IT	102 ✓

## Post Normalization

Employee

emp_id	first_name	last_name	birth_date	sex	salary	super_id	branch_id
100	David	Wallace	1967-11-17	M	250,000	NULL	1
101	Jan	Levinson	1961-05-11	F	110,000	100	1
102	Michael	Scott	1964-03-15	M	75,000	100	2
103	Angela	Martin	1971-06-25	F	63,000	102	2
104	Kelly	Kapoor	1980-02-05	F	55,000	102	2
105	Stanley	Hudson	1958-02-19	M	69,000	102	2
106	Josh	Porter	1969-09-05	M	78,000	100	3
107	Andy	Bernard	1973-07-22	M	65,000	106	3
108	Jim	Halpert	1978-10-01	M	71,000	106	3

Branch

branch_id	branch_name	mgr_id	mgr_start_date
1	Corporate	100	2006-02-09
2	Scranton	102	1992-04-06
3	Stamford	106	1998-02-13

Works\_With

emp_id	client_id	total_sales
105	400	55,000
102	401	267,000
108	402	22,500
107	403	5,000
108	403	12,000
105	404	33,000
107	405	26,000
102	406	15,000
105	406	130,000

Client

client_id	client_name	branch_id
400	Dunmore Highschool	2
401	Lackawana Country	2
402	FedEx	3
403	John Daly Law, LLC	3
404	Scranton Whitepages	2
405	Times Newspaper	3
406	FedEx	2

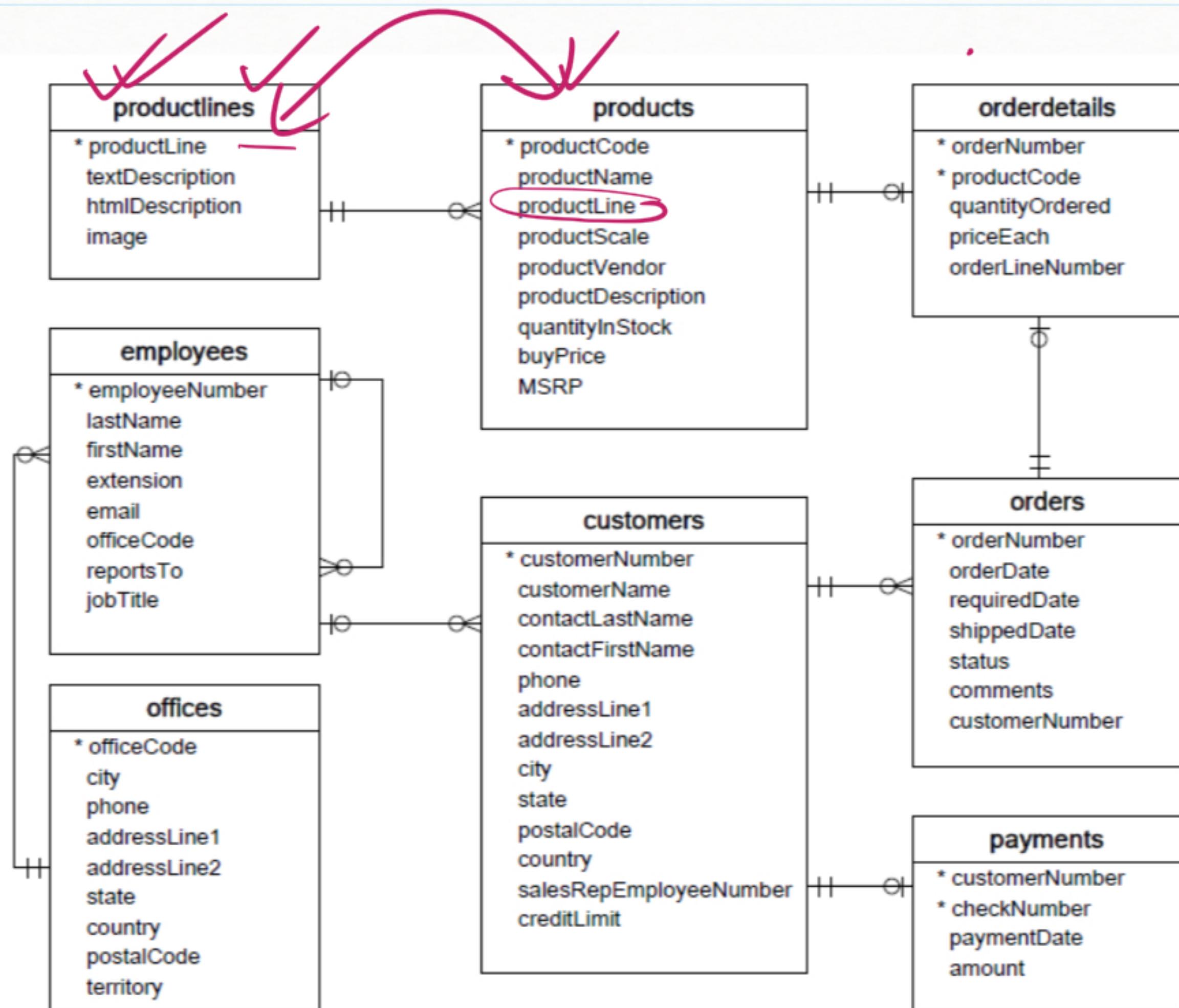
Branch Supplier

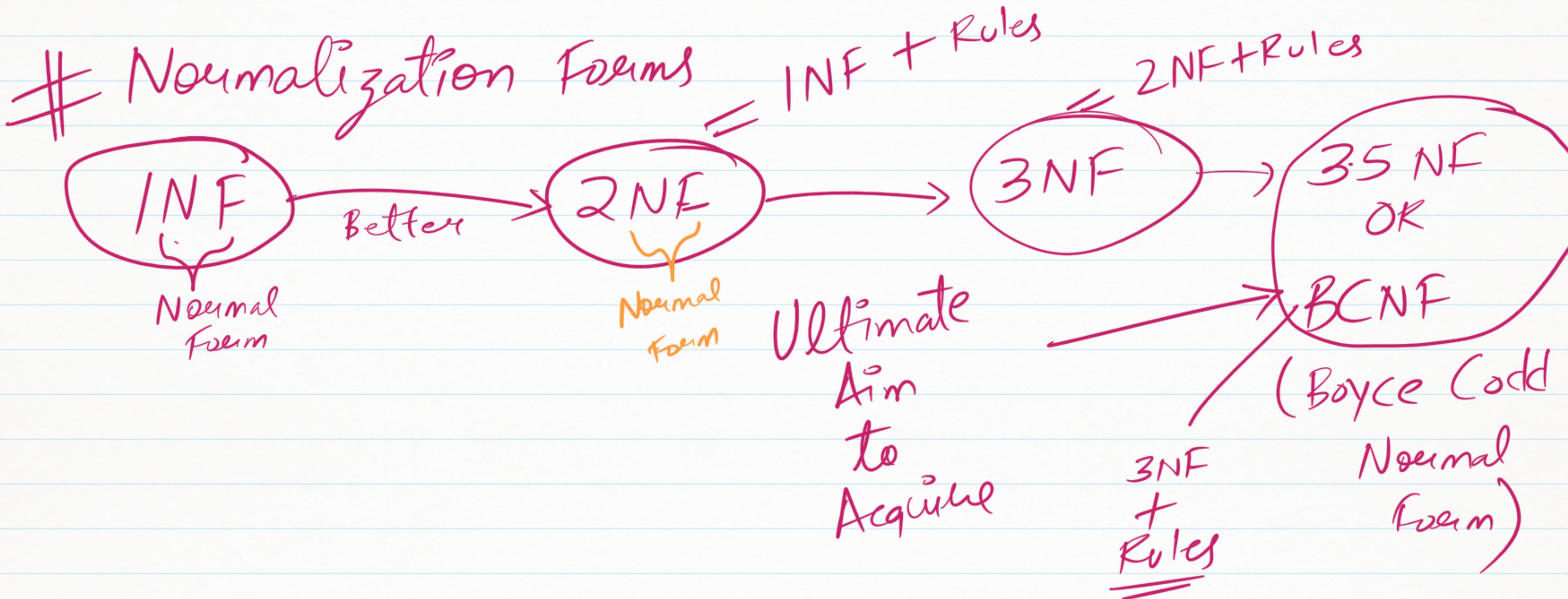
branch_id	supplier_name	supply_type
2	Hammer Mill	Paper
2	Uni-ball	Writing Utensils
3	Patriot Paper	Paper
2	J.T. Forms & Labels	Custom Forms
3	Uni-ball	Writing Utensils
3	Hammer Mill	Paper
3	Stamford Lables	Custom Forms

# What if we create just single table for this huge data?

\* Employee + Branch

Things will get complicated





## # How to do Normalization

- We will have to follow some rules
- To understand those rules, we will have to understand few concepts like

Partial Dependency & Functional Dependency

## # Functional Dependency

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

$$A \rightarrow BC$$

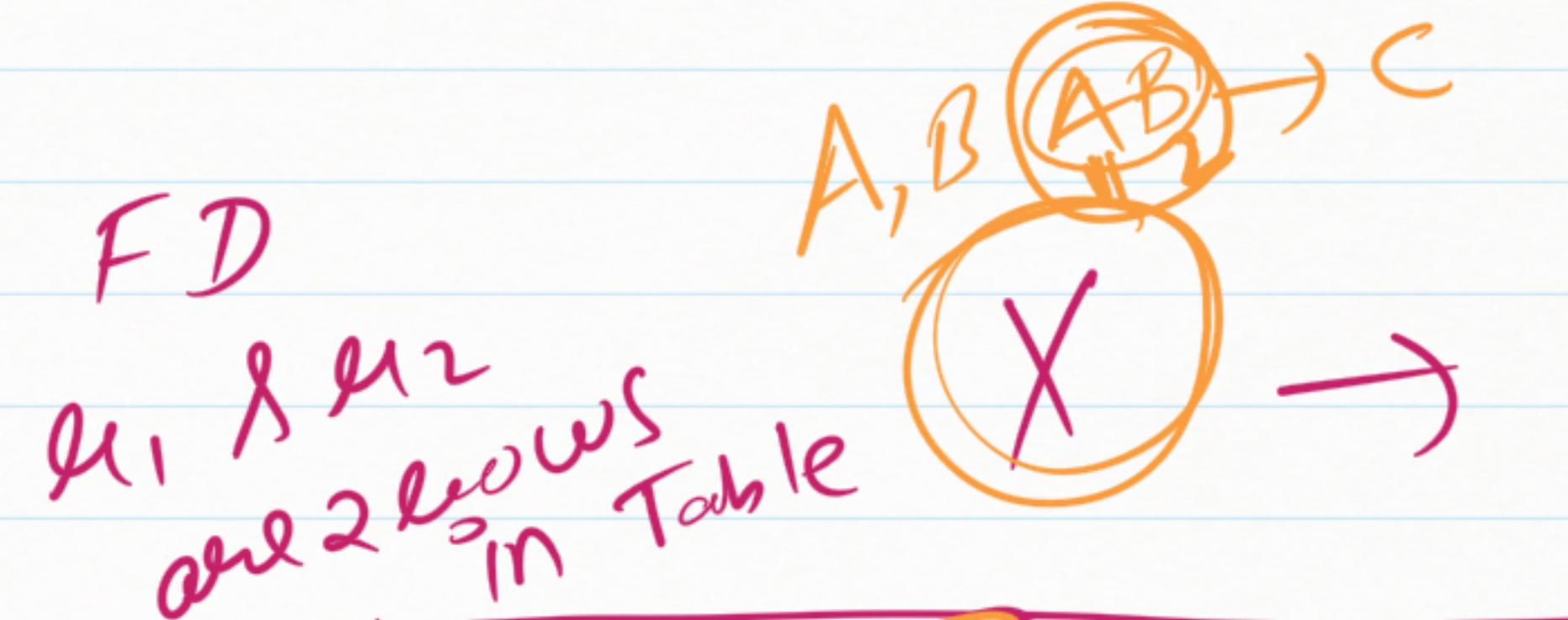
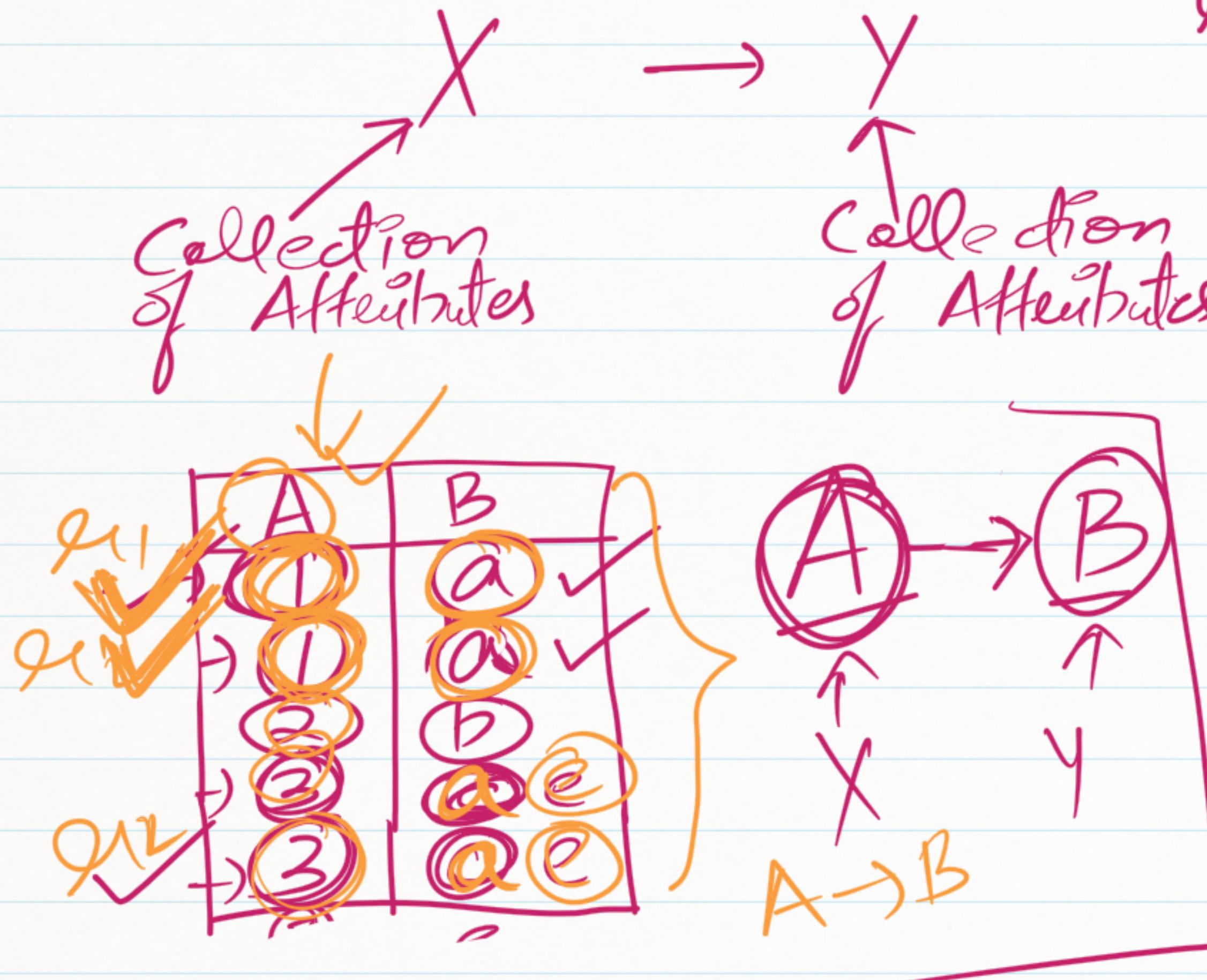
A determines BC

Given the value of A

you will be able to identify  
B & C.

$$x^2 = y + 2$$

## # Formal Definition of FD

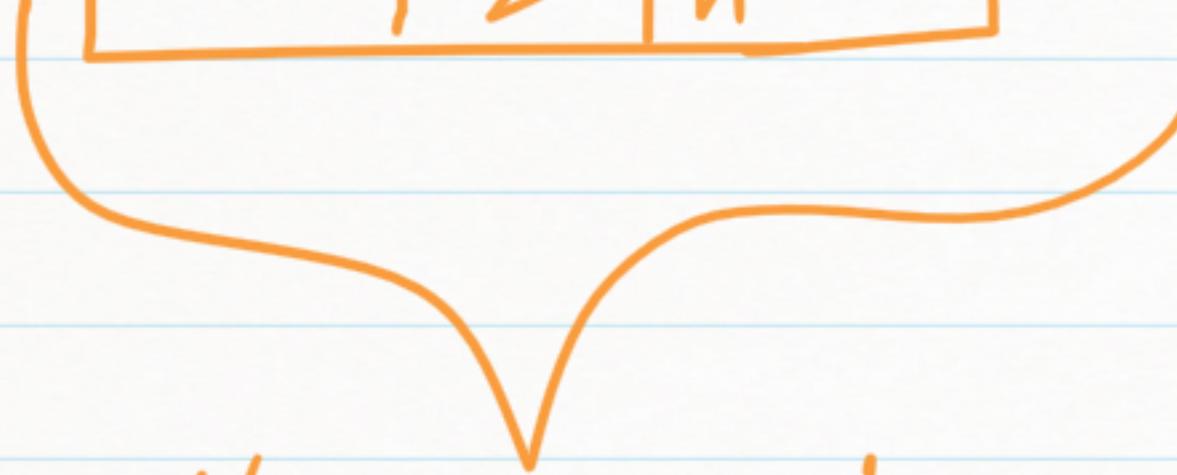


CASE	X	Y
1	If $a_1, a_{12}$ are same here	They must be same here also
2	If $a_1, a_{12}$ are diff here	They may or may not be diff here

We don't have worry about it

# How to rule out the wrong FD?

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



There can be so  
many rows in  
this table, but  
only few are  
shown to you.