1. Good Evening

2. Lecture begins at 9:05pm

3. Topic - Normalization

Agenda

1. Exceptions for relationship creation.

2. Normalization - What?, Anomalies, 1NF, 2NF, 3NF

3. CRUD → Create, Update, Delete] Queries.

4. Data Types.

Exceptions

- 1:1 → Col on any side
  - → UK
  - → FK
- 1:m → Col on m side
  - → FK
- m:m → mapping table

Sometimes 1:1 & 1:m will also need mapping table.

1. Sparse table.

Men

id	name	sid
1	A	NULL
2	B	3
3	C	NULL
4	D	NULL
5	E	NULL
6	F	"
7	G	"
8	H	"
9	I	"
10	J	"

Women

id	name
1	Z
2	Y
3	X

~~40 bytes~~ 10.4 ≈ 40 bytes

Not preferred.

Men

id	name
1	A
2	B
3	C
4	D
5	E
6	F
7	G
8	H
9	I
10	J

Women

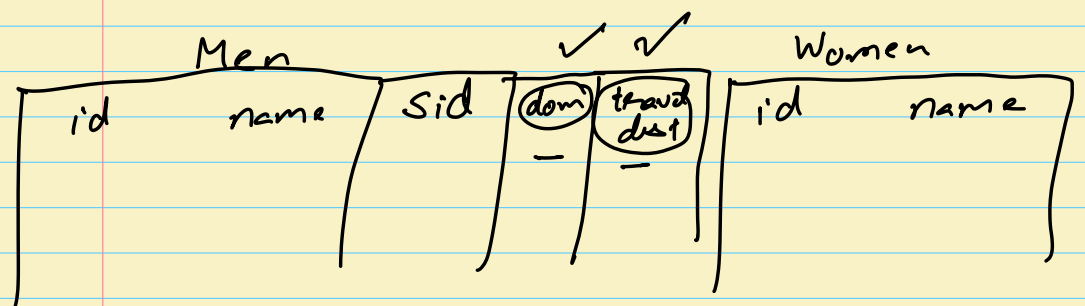
id	name
1	X
2	Y
3	Z

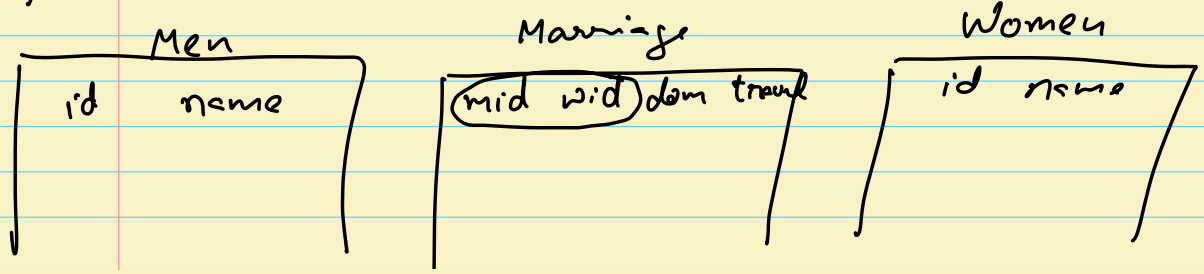
Spouse

mid	wid
2	3

4 + 4 = 8 bytes.

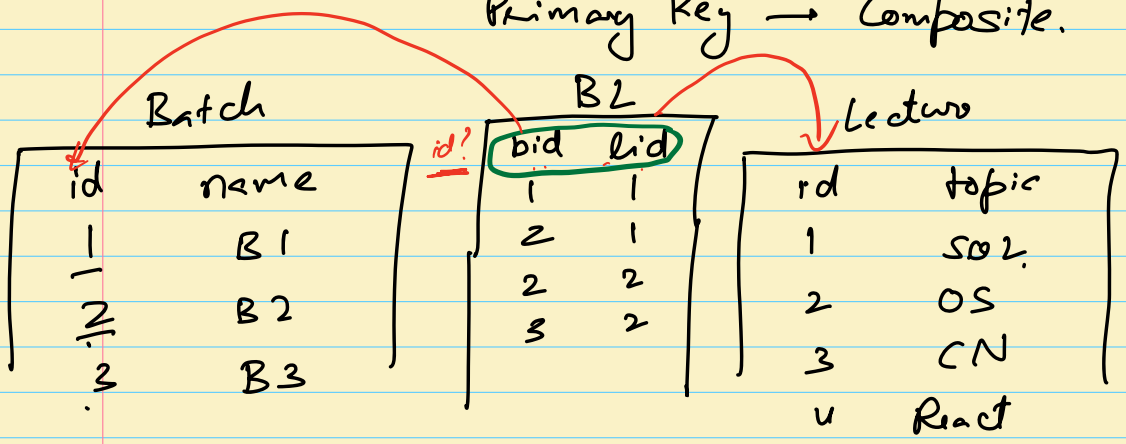
2. If there are attributes which are specific to the relationship & not the entities.





3. m : m ⟶ Mapping Table

Primary key → Composite.



SQL = B1, B2 ✓

OS = B2, B3 ✓

CN = B3

React = B1, B2, B3

Batch : Lecture

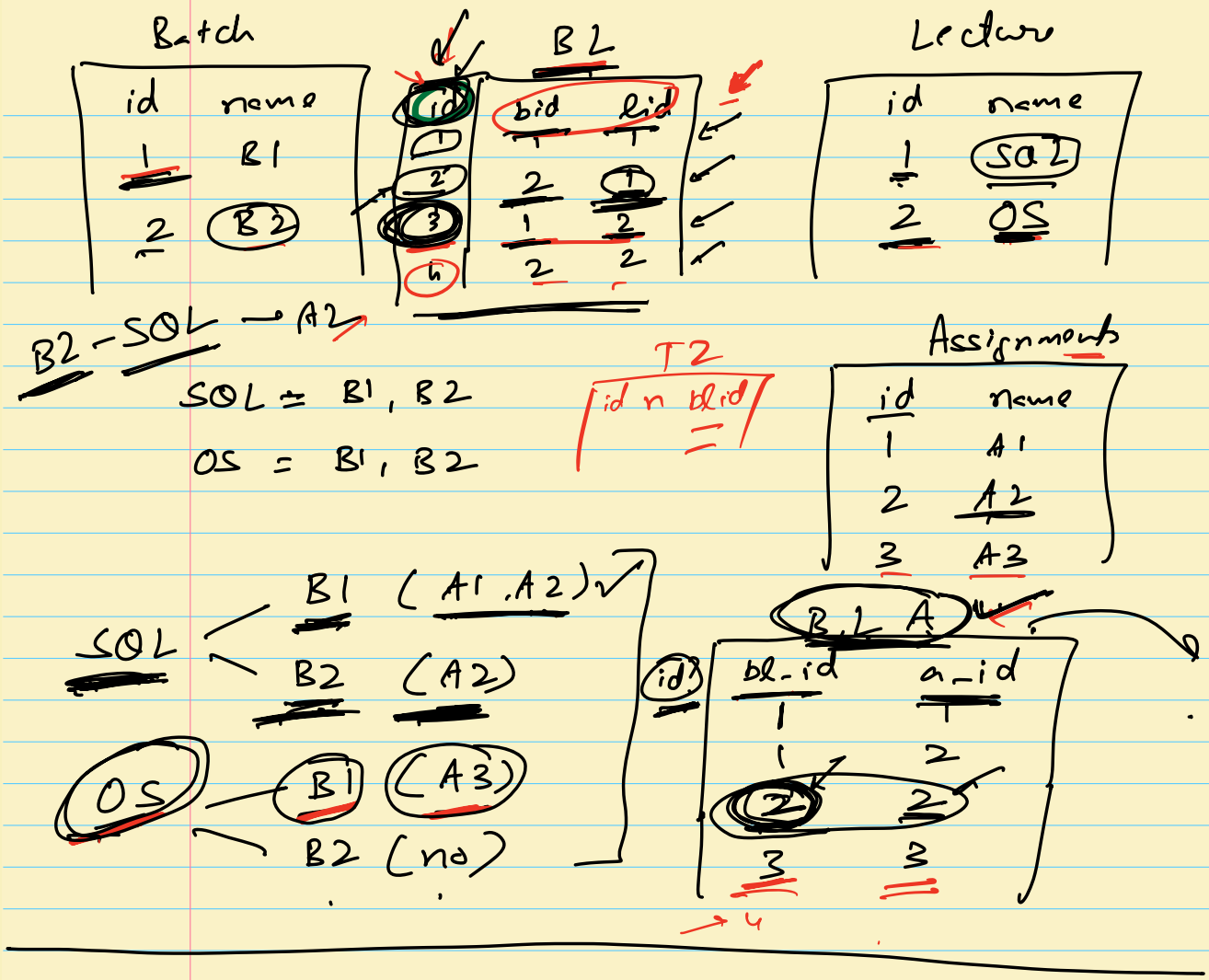
1 m

m 1

m : m

Web → a) If the mapping table has it's own attributes. [attendance for a batch > lecture combination]

Strong → b) If mapping table form it's own relationships (except the default two)



Normalization

What? ✓

Data Anomalies? ✓

[Normal Forms → 1NF, 2NF, 3NF]

What is normalization?

Technique to identify & resolve data redundancies leading to data anomalies.

Data Anomalies

→ Insert Anomaly.

Students

id	name	email	bid	bname
1	A	a@b	1	B1
2	B	b@c	2	B2
3	C	c@d	1	B1
4	D	d@e	2	B2
Null	Null	Null	3	B3
5	E	e@f	3	B3

Jan. 23

Wasting space? ✓
Low performance? ✓
Data duplicacy? ✓

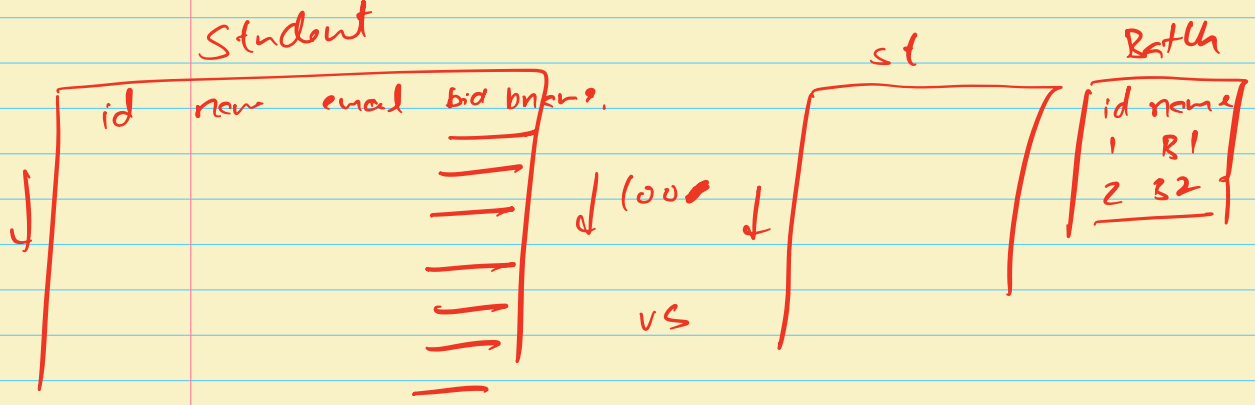
1. Can't create a batch with no students

Delete Anomaly

- While deleting B3 batch, student is also getting deleted & vice-versa.

Update Anomaly

0. Wastage of space.
1. Low Performance [B2 → B2_Jan_23]
2. Probability of data inconsistency increases.
   - All data duplicacy



Break = 9:59 to 10:06

Normal Forms

1st normal form → A db following 1NF will not have multi-valued columns.

Students

id	name	phone_numbers
1	A	xxx, yyy, zzz
2	B	aaa
3	C	bbb, ccc

Soln1.

Student

id	name	ph1	ph2	ph3
1	A	xxx	yyy	zzz
2	B	aaa	Null	Null
3	C	bbb	ccc	Null

→ Maximum phone numbers is not known.

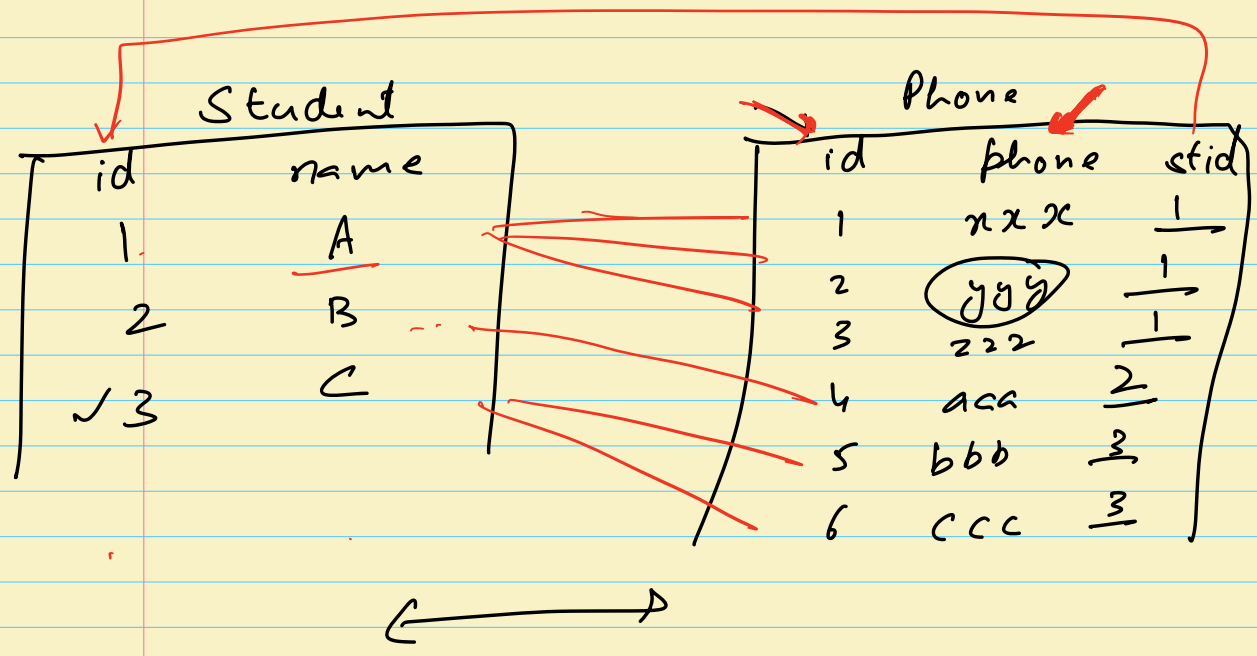
→ Sparse table.

Soln 2.

Student

id	name	phone
1	A	xxx
2	A	yyy
3	A	zzz
4	B	aca
5	C	bbb
6	C	ccc

Final Solution



2nd NF

1. It follows 1NF
2. All non PK columns should depend on entire PK.

Students

id	name
1	A
2	B
3	C

Mentor_Sessions

st_id	mid	s_date	feedback	m_name
1	1	1st Jan	Resume	ABC
1	1	10th Jan	DSA	ABC
2	1	10th Jan	Resume	ABC
3	2	1st Jan	LLD	DEF
2	2	15th Jan	HLD	DEF

Student

id	name
1	A
2	B
3	C

Mentor_Session

stid	mid	date	feedback
1	1	1st Jan	Resume

Mentor

id	name
1	ABC
2	DEF

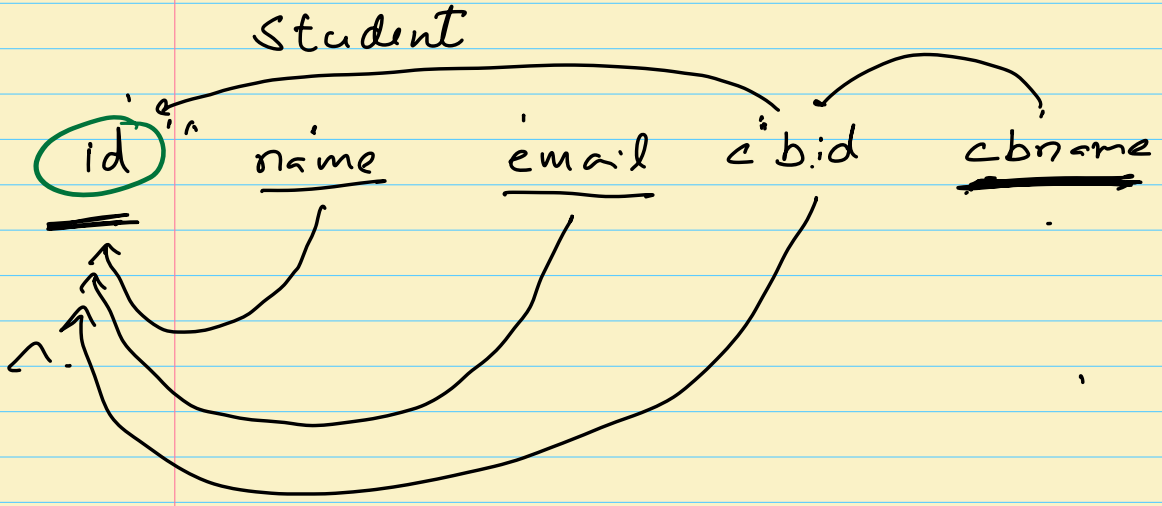
1NF → No multi-valued columns

2NF → Partial dependency of non pk columns on pk.

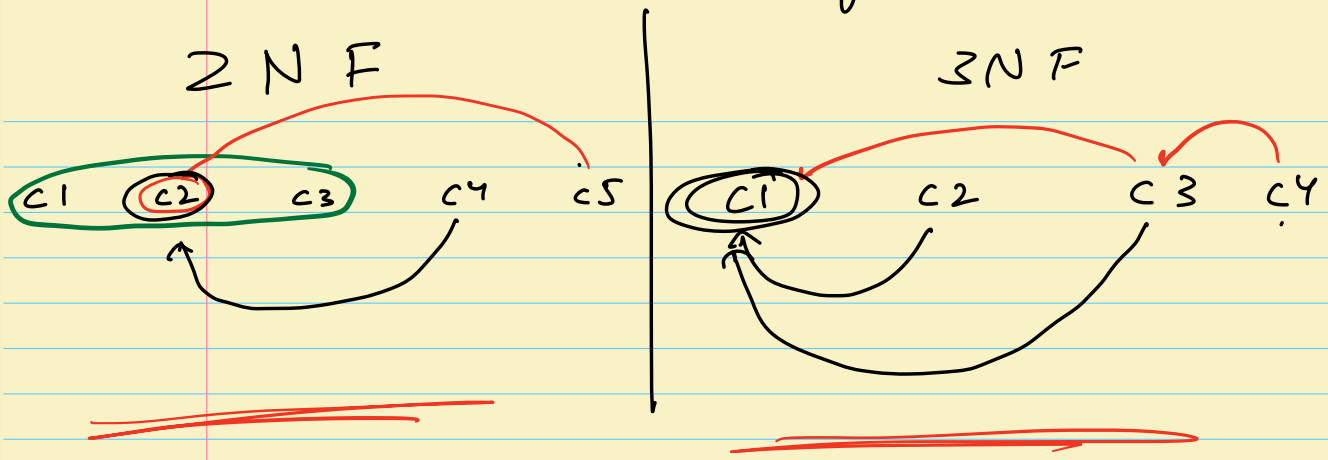
3NF → 1. 1NF

2. 2NF

3. All non pk columns should directly depend on pk. There should not be a transitive relationship.



Violations of



1 N F → Non multi-valued columns.

1 N F, 2 N F, 3 N F [BCNF, 4 N F]

Schema Design [Previous Lec]

→ 3rd NF compliant ←

Queries. [C R U D]