Today's Agenda :-

1) Schema Design for Scale Continued..

2) Deciding Primary Keys of mapping table

3) How to represent Enums ?

4) Represent indexes

5) Netflix Schema Design Case Study

How to approach Schema Design?

1) Scaler will have multiple batches. About each batch, store their name, start month and current instructor.

Batches (m) — Instructor (1)

m : 1

2) Each batch of Scaler will have multiple Students.

Batches (1) — Students (m)

1 : m

3) Each batch has multiple classes.

Batches (1) — Classes (m)

m : m

4) For each class, store the name of the class, date & time of class, instructor of class.

Classes (m) — Instructor (1)

m : 1

5) For every Student, store their name, grad year, university name, email, phone number.

6) Every Student has a buddy who is also a Student.

Students (1) → (1) Students

1 : 1

7) A Student may move from one batch to another.

8) For each batch a student goes to we have to store the start date of that batch.

Students (m) — Batches (1)

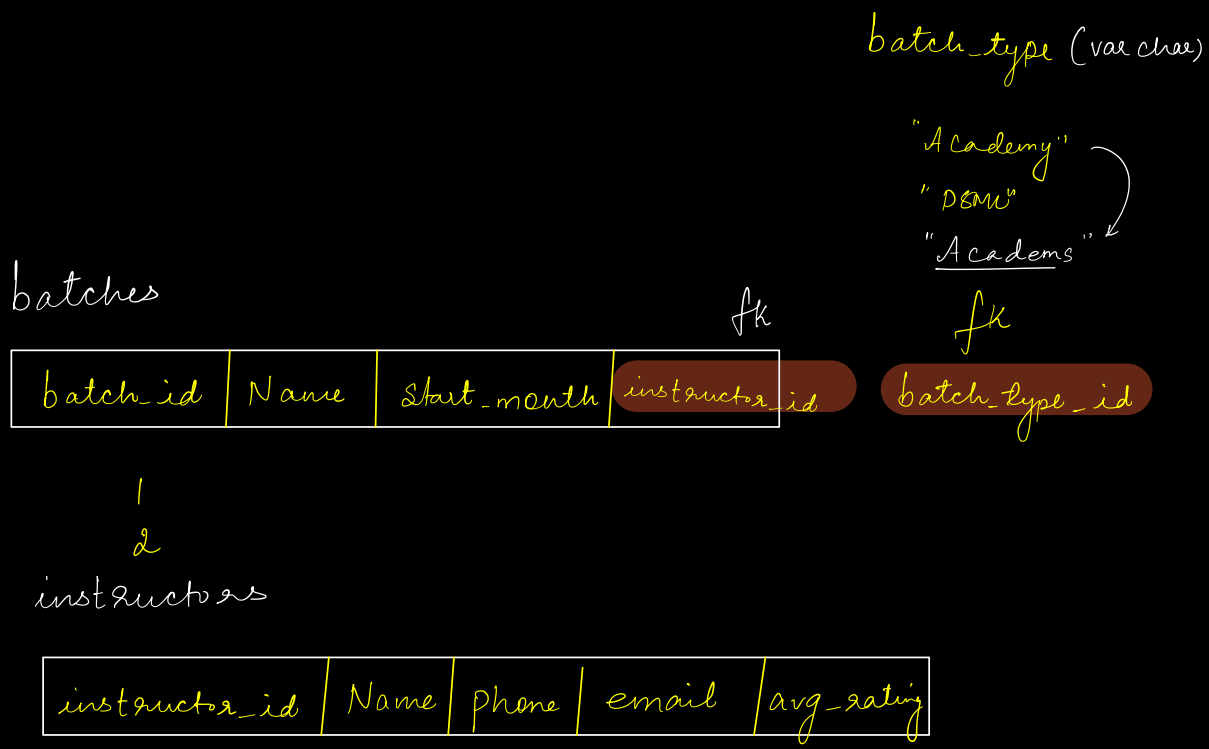
m : 1

9) Every Student also has a mentor. For every mentor, store their name and current company name.

10) We have to store info about all mentor sessions
store the time, duration, student, mentor,
stud_rating, mentor_rating.

mentor_session — student, mentor

11) For every batch we have to store info if it is Academy batch or DSML batch.



Students

Student_id	name	email	phone	gradYear	Univ_Name

batch_id mentor_id buddy_id

Self Referential → fk

buddy_id → also a St_id in the same table

classes

class_id	name	Schedule	instructor_id

mentors

mentor_id	name	current_company

mentor_Sessions

mentor_session_id	time	duration	St_rating	men_rating

St_id mentor_id

batches_classes

batch_id	class_id

Students_Batches

St_id	b_id	move_date

Batch_Types

id	type
1	"Academy"
2	"DSML"
3	"MLAI"

Enums :- used to represent well defined constants.

```
enum Gender {
    MALE,
    FEMALE
}
```

Mapping Table ✓

Batch_Types

id	type
1	"Academy"
2	"DSML"
3	"MLAI"

* Deciding Primary Keys of Mapping table :-

Students_Batches

st_id	b_id	move_time
2	3	24/05/24
2	4	2/07/24
3	3	25/06/24

⇒ (st_id, b_id) pk (1) → Create an index.

data_cv

Prakash_Batch

6	5	20th July
6	7	21th July
6	5	29th Aug

If it's allowed for a student to join the same batch again :

(st_id, b_id, move_time)

(2)

id (pk)	st_id	b_id	move_date

Advantages of separate key :-

Saves space because indexing size will be small.

Advantages of Not having sep key :-

Queries are faster on (b_id, £_id).

Representing fk & indexes :-

Sql by default creates index on pk of every table.

Always Indexing gets governed by use cases.

Please talk about indexes at the end.

1) Create the tables

2) Add the pk (id) to each table

3) Add other attributes

4) Represent Cardinalities (fks)

5) Index based on use cases

5 step approach for Schema Design

6 Min break