

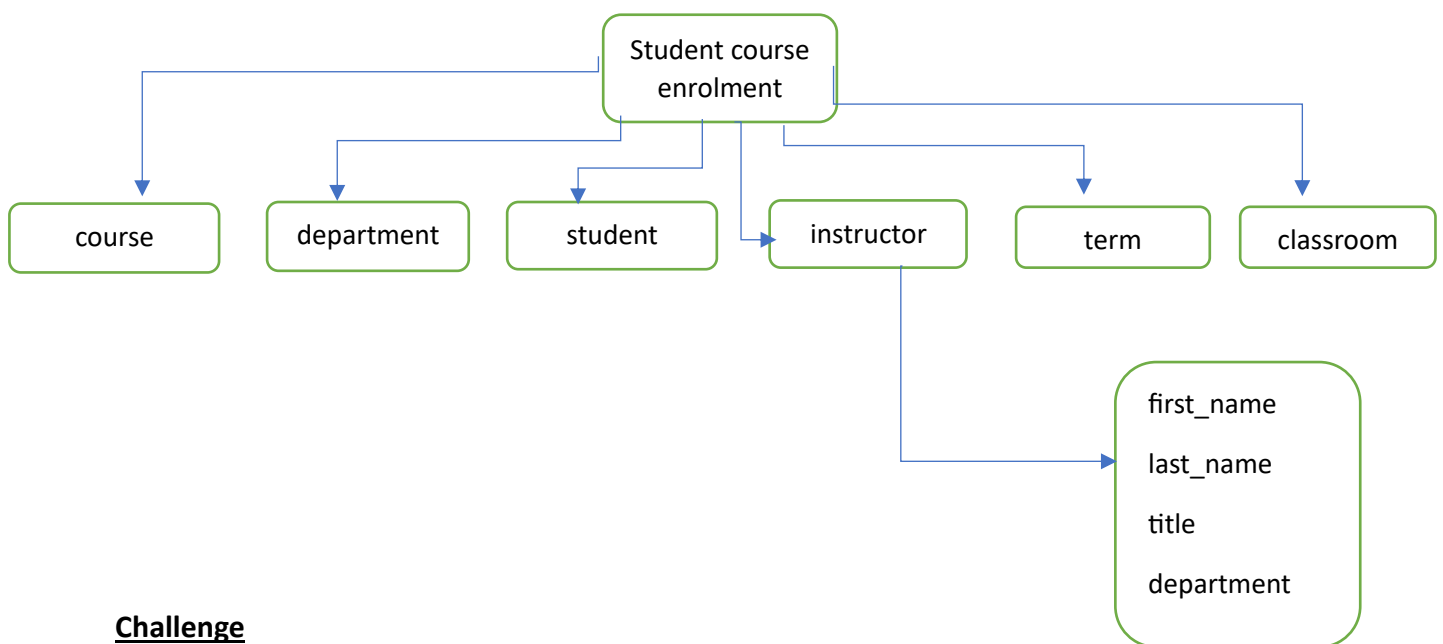
## Data Warehouse Assignment – 02

### Scenario 1:

#### Agenda:

1. Most popular courses.
2. Most popular instructors.
3. Courses popular among graduate students.
4. Courses popular among undergraduate students.
5. If there are courses for which assigned classroom is too large and too small.

Design layout given initially:-



### Challenge

**Q.** What if a course has multiple instructor?

**Ans.** Option A

Id
Name
Course_id

Instructor, here name can be in combine form like Manning/Raghavan as well as separate name.

- The problem with this option is that we will have to maintain instructor table quite a lot as the instructors can change for a single course id frequently. So this not suggested.

Option B:

Changing grain of the fact table to be one row per student enrolment per course per instructor. i.e., has 2 fact rows for each student.

This option is also not suggested as it will become challenging to maintain such if instructors keep on changing for the course.

Recommended -> Option C

Creating 2 fact tables.

fact\_A: Student enrolment fact table.

Id
course_id
student_id

fact\_B: student instructor fact table.

student_id
course_id
instructor_id

Answer: **Option C** is recommended as its easier to maintain.

### Question 1

	Strength	Weakness
Option A	Easier design lesser dimension	Difficult to maintain, suitable for static data.
Option B	Easier design	Difficult to maintain congested fact table instructors needs to be static.
Option C	Easy to maintain dynamic data can be handled.	More dimensions.

### Scenario 2:

Building a data warehouse for an online brokage company. Company makes a commission when the customer buy and sell stocks.

#### **Proposed design**

Fact table	Trade
Grain	One row per stock trade
Dimension table	Data
	Customer
	Account
	Security
	Trade type

#### **Two scores are calculated:**

- Each customer is placed into one of nine customer activity segments based on their frequency of transactions, average transaction size, and recency of transactions.
- Each customer is assigned a customer profitability score based on the profits earned because of that customers' trades. The score can be either 1,2,3,4 or 5, with 5 being the most profitable.

Option	Score criterion	strength	weakness
<b>Option A</b>	The scores are attributes of the Customer dimension. When scores change, the old score is overwritten with the new score.	Simple approach	Write operations are not recommended.
<b>Option B Recommended</b>	The scores are attributes of the Customer dimension. When scores change, new customer dimension rows are created using the updated score (type 2 slowly changing dimensions).	Historic data entry	Will require more memory, but difficult to set pipeline for data flow.
<b>Option C</b>	The scores are stored in a separate Customer Scores dimension which contains 45 rows, one for each combination of activity and profitability scores dimensions.	Historic data entry	Will require more memory but is suitable for static data. Difficult to maintain if parameters are increased to maintain from 45 to higher or lower numbers. Not a dynamic approach.
<b>Option D</b>	The scores are stored in a customer scores outrigger table which contains 45 rows. The customer dimension includes a foreign key to the outrigger table (but the fact table doesn't). When scores change, the foreign key column in the Customer table is updated to point to the correct outrigger row.	Historic data entry	Write operations are not recommended frequently, chances of error and no historical data also will occupy larger space.

