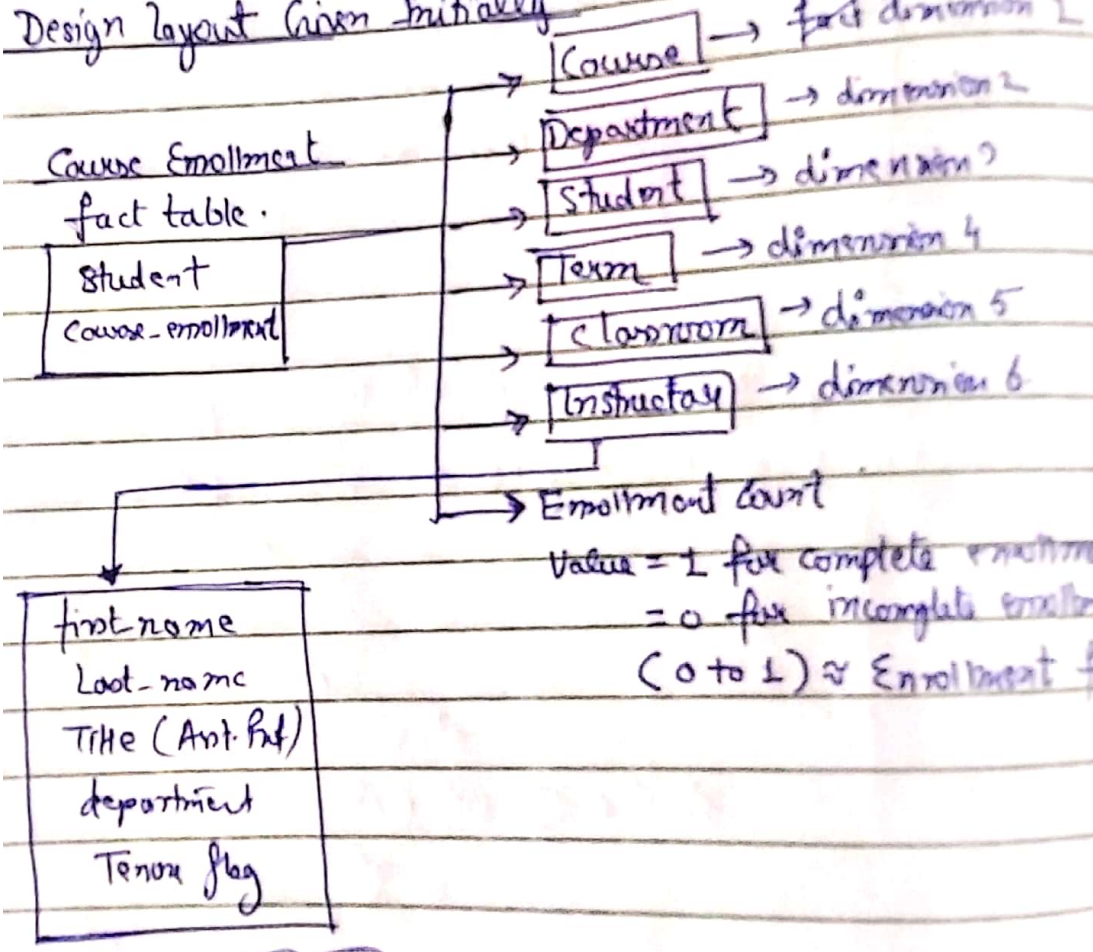
DW ASSIGNMENT 2

Scenario 1:

Agenda

1. Most Popular Courses
2. Most Popular Instructors
3. Course popular among graduate students
4. Course popular among undergrads students
5. Are there courses for which assigned classroom is too large or too small

Design Layout Given Initially -

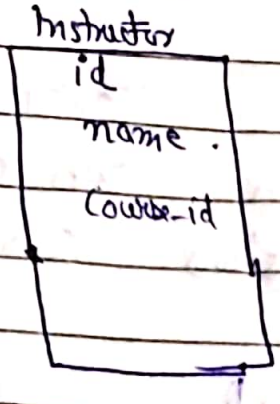


first_name
Last_name
Title (Asst. Prof)
department
Tenure flag

Challenge

What if a Course has Multiple Instructor?

Option A-



Here name can be in combined form like Manning J Raghavan as well as separate name.

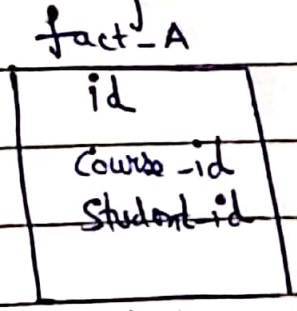
* The Problem with this option is that we will have to maintain Instructor table quite a lot as the instructors can change for a single course id frequently. So this is not suggested.

Option B

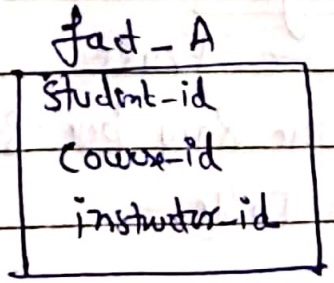
Changing grain of the fact table to be one row per student enrollment per course per instructor.
i.e has 2 fact Rows for each student.
This option is also not suggested as it will become challenging to maintain such table if Instructors keeps on changing for the courses.

Recommended → Option C

Creating 2 fact tables.



Student enrollment fact table.



Student-instructor fact-Table

Answer: Option C is recommended as its easier to Maintain.

Question 1-

	Strength	Weakness
Option A	Easier design Lesser dimensions	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 dw for an online brokage company company makes a commission when the customer buy and sells stocks					
proposed design					
fact table	trade				
grain	one row per stock trade				
dimension table	data				
	customer				
	account				
	security				
	trade type				
Two Scores are calculated		1-Each customer is placed into one of nine Customer Activity Segments based on their frequency of transactions, average transaction size, and recency of transactions.			
		2-Each customer is assigned a Customer Profitability Score based on the profits earned as a result of that customer's trades. The score can be either 1,2,3,4, or 5, with 5 being the most profitable			
Options	**Scores Criterion**		**strength**	**weakness**	

Option A	The scores are attributes of the Customer dimension.When scores change,the old score is overwritten with the new score	simple approach less memory	write operations are not recommend
Option B Recommended	The scores are attributes of the Customer dimension. When scores change, new Customer dimension rows are created using the updated scores (Type 2 Slowly Changing Dimension)	historical data entry	will require more memory Bit difficult to set up pipline for data flow
Option C	The scores are stored in a separate Customer Scores dimension which contains 45 rows, one for each combi- nation of activity and profitability scores. The Trades fact table includes a foreign key to the CustomerScores dimension.	historical data entry	will require more memory but is suitable for static data Difficult to maintain if parameters are increased from 45 to higher or lower numbers.Not a dynamic approach
Option D	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 does not). When scores change, the foreign key column in the Customer table is updated to point to the correct outrigger row	historical data	write operations are not recommend frequenctly ,chances of error and no historical data Also will occupy larger space