# CPS610 - Assignment 2 - Labs 4 / 5:
Distributed Databases with Relational Algebra and SQL

This assignment explores key concepts in distributed databases, including SQL queries and the challenges of managing data across multiple sites. You will work with centralized and distributed database structures to understand data distribution, query processing, and reconstruction of data from fragmented sources.
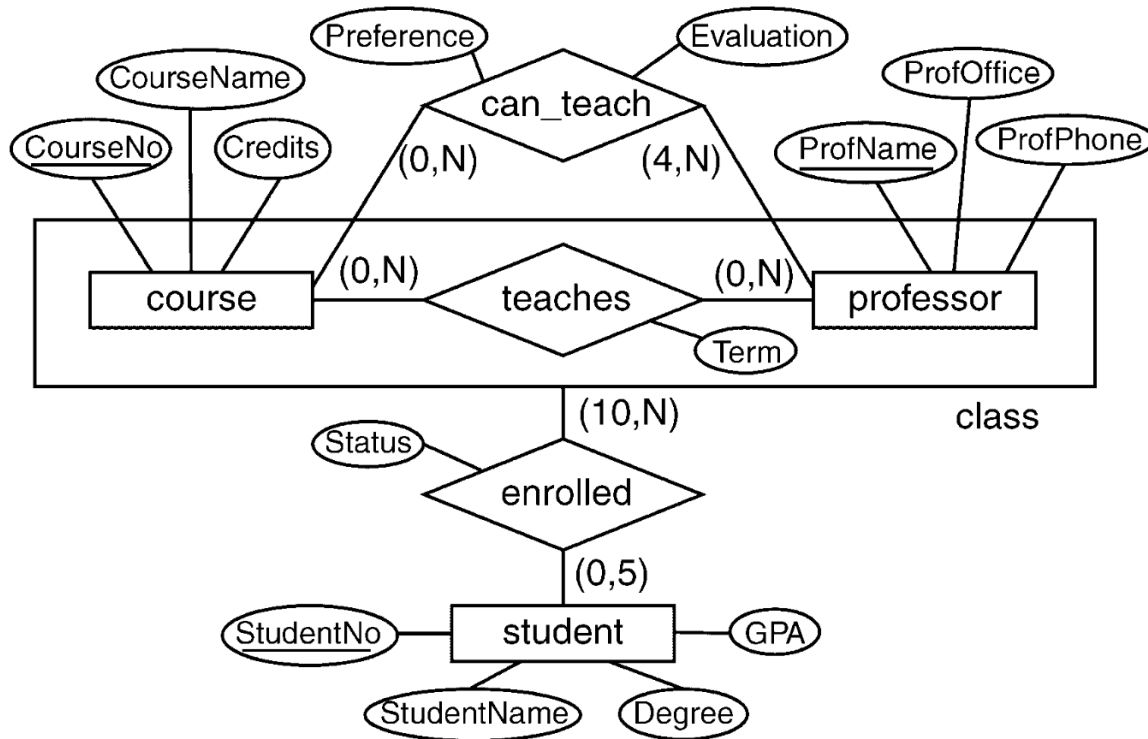
Based on the ER above of a University database, you should create a centralized database schema (Central Database) and its distributed counterparts (Engineering and Science databases). The assignment is divided into **two labs** (Lab 4 and Lab 5), each focusing on specific tasks.

Tasks:
1. **Task 1: Creating Relational Schemas:** Create the Relational Schemas for the Centralized Database and the Distributed Counterparts using our two Oracle Databases, Oracle11 and Oracle 12. You have three databases and need three instances of databases: One centralized and two distributed. How do we solve this problem? There are a few ways to do this; choose one and go with it! Your solution should involve at least one DBLink.
2. **Task 2: Relational Algebra and SQL for Populating DDBs:** Create and populate distributed database tables (Engineering and Science) with data extracted from the centralized database using **relational algebra operations** and **SQL commands**.

3. **Task3: Relational Algebra and SQL for Reconstruction of Central DB:** Ops, you lost your Central Database data!! Reconstruct data from distributed fragments into the centralized database using SQL UNION operations.
4. **Task 4: Relational Algebra and Queries:** Write SQL queries to solve the following problems:
   - List of classes taken by students in Engineering and Science faculties.
   - Retrieve students with a GPA greater than or equal to 4.
   - Calculate the average GPA for students enrolled in each course.
   - Identify professors teaching courses they have not listed as teachable.
   - Retrieve professors who can teach both Science and Engineering courses.
   - List students enrolled in Professor Harry's classes, including course names.
   - Identify courses worth 1 credit in both Science and Engineering.
   - Retrieve professors teaching during the summer term.

Marking Scheme (Total: 10 Marks):

1. **Tasks 1, 2, 3: (4 Marks):**

   - Relational Schemas (1 mark)
   - Correctly populate distributed tables from the centralized database (2 Marks).
   - Reconstruct centralized data from distributed sources (1 Mark).

2. **Task 4: (6 Marks):**

   - 0 queries = 0 marks
   - Between 1 and 3 queries = 1 mark
   - Between 4 and 6 = 2 marks
   - Between 6 and 7 = 4 marks
   - 8 queries = 6 marks

Deliverables:

Students must show all queries running to the TA and submit a report (there is a provided template you must use) containing the following:

1. Relational Schema for centralized and two distributed databases (task 1)
2. SQL scripts and relational algebra for all other tasks (tasks 2, 3, 4)

IMPORTANT: Not using the template results in a 50% mark deduction.