**UVA School of Data Science**

**DATA ENGINEERING**

**Database Project: UVA SDS Online MSDS Program Database**

Last updated: November 29, 2021

**EXECUTIVE SUMMARY**

There have been several questions from students, faculty, and staff on the topic of courses and curriculum in the UVA SDS Online MSDS Program. Given your knowledge of database design, you decide it will be helpful to construct a database containing some information about the courses. In particular, these things must be in the database:  
  
1) Learning Outcomes (LOs) for each course

2) The instructors assigned to teach the different courses for each term in 2021 (Spring, Summer, Fall)

**NOTE ON LEARNING OUTCOMES**

LOs capture what students should be able to do after taking a course (e.g., design and build a MySQL database). You will see that the LOs vary in terms of detail (they were written by different instructors). The good news is that someone has collected the LOs and other high-level information about the courses; that data will be provided to you.

**NOTE ON INSTRUCTOR ASSIGNMENTS**

This data was collected for you, although it will require a little cleanup and organization before database insertion.

**INSTRUCTIONS**

The project is divided into two parts, and you will submit your work separately for each part.  
The first project part involves understanding the problem and uses cases, and designing a relational database to support the needs. The second project part involves database buildout and query construction and testing. You will useMySQL Workbench to accomplish most of the tasks.  
  
**MySQL Workbench** [download site](https://dev.mysql.com/downloads/workbench/)**Data**: included in this project package

**PART 1**  
Part 1 of the assignment is to understand the problem and design the database.   
You will then answer the following Design Questions and submit your responses.  
  
**Part 1 Total Points: 20**  
  
**Design Questions**

1) (3 PTS) What tables should you build?

2) (2 PTS) For each table, what field(s) will you use for primary key?

3) (2 PTS) For each table, what foreign keys will you use?  
  
4) (2 PTS) Learning outcomes, courses, and instructors need a flag to indicate if they are currently active or not. How will your database support this feature? In particular:  
  
If a course will be taught again, it will be flagged as active. If the course won’t be taught again, it will be flagged as inactive.   
  
It is important to track if an instructor is a current employee or not.

Learning outcomes for a course can change. You’ll want to track if a learning outcome is currently active or not.   
  
5) (1 PT) Is there anything to normalize in the database, and if so, how will you normalize it? Recall the desire to eliminate redundancy.

6) (1 PT) Are there indexes that you should build? Explain your reasoning.

7) (2 PTS) Are there constraints to enforce? Explain your answer and strategy.  
For example, these actions should not be allowed:  
- Entering learning objectives for a course not offered by the School of Data Science  
- Assigning an invalid instructor to a course

8) (5 PTS) Draw and submit a Relational Model for your project. For an example, see *Beginning Database Design Solutions* page 115 Figure 5-28.

9) (2 PTS) Suppose you were asked if your database could also support the UVA SDS **Residential** MSDS Program. Explain any issues that might arise, changes to the database structure (schema), and new data that might be needed. Note you won’t actually need to support this use case for the project.

Aside: Would it have been helpful to know this information before designing this database? You bet! Does this sort of thing happen in real life? Certainly. A good engineer/developer would have addressed this need during the planning stage with the right question: *Any other use cases we’d want this database to support?*

**PART 2**  
Part 2 of the assignment is to build the database, insert all data, and write SQL queries to answer use case questions. Results of SQL queries can be exported using a button in MySQL Workbench.

Your submitted file must include:  
1) (6 PTS) Table creation file. This is the SQL code to construct database tables. INSERT statements should NOT be included. You run this code to create tables.  
  
2) (14 PTS; broken out below) SQL queries to answer User Case Questions, with resulting output

Note: it is good practice to save ALL code in files so it can be replicated and edited as needed.

**Part 2 Total Points: 20**

**Use Case Questions**

After building and populating your database, you will write SQL queries to answer the questions below.  
Include your SQL code and results in your submitted file.

1) (1 PT) Which courses are currently included (active) in the program? Include the course mnemonic and course name for each.

2) (1 PT) Which courses were included in the program, but are no longer active? Include the course mnemonic and course name for each.

3) (1 PT) Which instructors are not current employees?

4) (1 PT) For each course (active and inactive), how many learning outcomes are there?

5) (2 PTS) Are there any courses with no learning outcomes? If so, provide their mnemonics and names.

6) (2 PTS) Which courses include SQL as a learning outcome? Include the learning outcome descriptions, course mnemonics, and course names in your solution.

7) (1 PT) Who taught course ds5100in Summer 2021?

8) (1 PT) Which instructors taught in Fall 2021? Order their names alphabetically, making sure the names are unique.

9) (1 PT) How many courses did each instructor teach in each term? Order your results by term and then instructor.  
  
10a) (2 PTS) Which courses had more than one instructor for the same term? Provide the mnemonic and term for each. Note this occurs in courses with multiple sections.

10b) (1 PT) For courses with multiple sections, provide the term, course mnemonic, and instructor name for each. Hint: You can use your result from 10a in a subquery or WITH clause.

Note: Question 10 is good preparation for SQL interview questions.