# **SOEN 363 - Data Systems for Software Engineers**

Concordia University Winter 2022

## • Course Description

This course is designed to introduce software engineering students to the concepts and practical aspects of modern data systems. Data systems have widely diversified in the past decade, with the emergence of NoSQL, the prevalence of big data, the convergence between data storage and analytics materialized in particular in distributed file systems and the recurrent promises of linked data. Through a specific emphasis on examples and case studies, this course will teach students the design and architectural concepts required to choose the right data system for the proper application. Introduction to the current data ecosystem, relational databases and NoSQL systems, such as key-value databases, document databases, column databases, and graph databases. Lectures: three hours per week. Tutorial: one hour per week.

### • Lectures from campus

MoWe 4:15PM - 5:30PM Room: H 920 SGW Time: 10:15 a.m. - 11:30 a.m. Note: Lectures will *NOT* be recorded and *NOT* live.

## • Course Team

Instructor: Dr. Essam Mansour (<u>essam.mansour@concordia.ca</u>) Instructor's Website: <u>emansour.com</u>

### TAs:

Philippe Carrier (philippe.c@hotmail.com) Ahmed Amer (ahmed.aly.20211@mail.concordia.ca) Yasaman Sabbagh (yasaman.sabbagh@yahoo.com) Hussein Abdallah (hussein.abdallah@mail.concordia.ca) Mossad Helali (mossad.helali@mail.concordia.ca) Reham Omar (reham.omar@mail.concordia.ca)

## Lab Sessions

There will be a weekly lab session, *starting Jan 17, 2022*. Note that the lab sessions are a mandatory part of the course and the exams include them. All lab sessions will be in person.

## • Course Textbook

Raghu Ramakrishnan and Johannes Gehrke, "Database Management Systems", Third Edition, McGraw-Hill, 2002.

## • Course Organization

The participation of students in the course will involve four forms of activities:

- Attending lectures and labs online.
- Solving assignments (involving writing and/or coding).
- A programming project of two phases.
- Participating in class discussions.

### Course Outline

- Data Models
- Query Languages
- Data Storage (disk and files)
- Indexing
- Query Optimization
- NoSQL-Databases
- Graph databases (Neo4j)
- Document stores (Elasticsearch)

### • Assessment

Each student will receive a numeric score with a corresponding letter grade, based on a weighted average of the following:

**Project:** The project will count for a total of **35% of your final score**. There will be **two phases** throughout the course in teams. Phase one covers relational data systems and phase two includes NoSQL systems. Each team consists of at least 3 up to 4 students. The team also should find a real dataset of at least one GB and demo the capabilities of the chosen system. Note that attendance at the project demo is mandatory; if you do not participate in both group and individual project demonstrations, you will not receive marks for the project. The deliverables of each phase are due as the following:

Phase 1 on (Release Date): Mon, Jan 31, 2021 (Due Date): Sunday, Feb 27, 2022, Phase 2 on (Release Date): Mon, Feb 28, 2022 (Due Date): Wed, April 06, 2022,

If you submit **one day late**, we will **deduct 25%** of the project score as a penalty. If you submit **two days late**, **50% will be deducted**. The project will not be graded (and you will receive a zero score on it) if you are more than two days late.

**Exams:** There will be <u>*two midterm exams*</u>, which combined will account for **35% of your final score**, as the following:

Midterm onefor 1 hour on Feb 16, 2022Midterm twofor 1 hour on March 28, 2022

**Assignments:** There will be 5 assignments that will test you on some problem analysis and solving skills. The 5 assignments will altogether contribute **30% towards your final score**. These assignments are:

ID	Торіс	Weight	Release Date	Due Date
1	E.R. diagram	7%	Mon Jan 17	Sun Jan 30
2	SQL	8%	Mon Jan 31	Sun Feb 13
3	Indexing	8%	Mon Feb 28	Sun Mar 13
4	Graph Queries	7%	Mon Mar 14	Sun Apr 3

**Class/Recitation Participation:** Your attendance of both classes and labs as well as your participation in discussions during presentations, will help you to do well in your exams and assignments.

Table 1 below shows the breakdown of the main forms of activities that the course involves, alongside the quantity and the overall weight of each activity. Take into account that small differences in scores can make the difference between two letter grades. Letter grades will be determined by absolute standards. The total score will be plotted as a histogram. Cutoff points are determined by examining the quality of students' work on the borderlines. Individual cases, especially those near the cutoff points <u>may be adjusted upward or downward based</u> <u>on factors such as attendance, class participation</u>, the improvement observed throughout the course, exam performance, and special circumstances.

Туре		Weight
Project	2	35%
Exams		35%
Assignments	4	30%

**Table 1:** Breakdown of the main activities involved in the course.

### Policies

#### **Working Alone on Assignments**

Assignments that are assigned to students should be *performed individually*.

#### Handing in Assignments/Projects

All assignments/projects are due at 11:59 PM (one minute before midnight) on the specified due date.

#### **Appealing Grades**

After each exam, assignment, and/or project is graded, you have **7** calendar days to appeal your grade. All your appeals should be provided in writing. If after appealing you are still not satisfied, please visit the instructors. If you have questions about an exam, an assignment or a project grade, please visit the instructors directly.

## • Cheating

Each project or assignment must be the sole work of the student(s) turning it in. Projects and assignments will be closely monitored, and <u>students may be asked to explain suspicious</u> <u>similarities with any write-up or piece of code available</u>. The following are guidelines on what cheating is and is not:

### What is cheating?

- 1. Sharing code or other electronic files: either by copying, retyping, looking at, or supplying a copy of a file.
- 2. Sharing written assignments: either by re-writing, looking at, or supplying a copy of an assignment.

#### What is NOT cheating?

- 1. Clarifying ambiguities or vague points in-class handouts.
- 2. Helping others use the computer systems, networks, compilers, debuggers, profilers, or other system facilities.
- 3. Helping others with high-level design issues.
- 4. Helping others debug their codes.

Consequently, be aware of what constitutes cheating (and what does not) when interacting with your colleague students. The same rules of cheating as above apply when collaborating with other students. In short, you cannot share written assignments, code, and/or other electronic files with other students. If you are unsure, ask the teaching staff.

Finally, be sure to store your work in protected directories. The penalty for cheating is severe, and might jeopardize your whole career as a student – cheating is not worth the trouble. By cheating in the course, you are cheating yourself; the worst outcome of cheating is missing an opportunity to learn. Besides, you will be removed from the course and assigned a failing grade. We also place a record of the incident in your permanent university profile.