New York University Tandon School of Engineering
Computer Science and Engineering
CS-GY 6083 A Principles of Database Systems
Fall 2020
Professor Julia Stoyanovich
Time: Thursday 9:30-11am or 5-6:30pm (lecture)
Friday 9:30-10:30am or 1:30-2:30pm or 5:30-6:30pm (lab)

Location: 5 MetroTech Center, Pfizer Auditorium

To contact professor: stoyanovich@nyu.edu
Office hours: Monday 10-11am or by appointment

Course Prerequisites: Graduate standing, CS 6003 or equivalent, familiarity with basic data structures and operating system principles. No prior experience with relational databases is required or expected.

Course Description: This course broadly introduces database systems, including the relational data model, query languages, database design, index and file structures, and query processing and optimization. Additional topics will be covered to give students the flavor of current data management research. Students will acquire hands-on experience in working with database systems and in building web-accessible database applications.

Course Objectives: After successfully completing the course, students are able to:
- Explain the characteristics of Big Data applications, and articulate the main design principles of data processing platforms that respond to these characteristics.
- Use MapReduce and Apache Spark to efficiently implement large-scale data analysis of urban transportation datasets.
- Articulate the need for data profiling and data cleaning, and implement several data profiling and data cleaning methods using open-source tools and libraries.
- Implement a computer program that efficiently identifies frequently occurring patterns in large datasets.
- Articulate the difference between different interpretations of algorithmic fairness, and implement a computer program that demonstrates the trade-offs.

Course Structure: This is a blended course. The course is made up of a lecture and a lab. Your synchronous participation is encouraged but not required for full credit. Recordings of lectures and labs, along with slides and exercises, will be available online.

Lecture: Thursdays from 9:30-11am EDT over Zoom; Thursdays from 5-6:30pm EDT in person in Pfizer Auditorium. You may attend the lecture in person (provided that you did not elect remote-only instruction) or synchronously over Zoom. You may also watch the recording of the lecture, which will be posted to NYU Classes.

<u>Lab</u>: Fridays 9:30-10:30am EDT; 1:30-2:30pm EDT; 5:30-6:30pm EDT.   Labs are conducted in smaller groups over Zoom. You may attend one of the labs synchronously, or watch a recording over Zoom.   We ask that you select a lab session and do your best to stick with it over the semester.   We will make an initial assignment of students to lab sessions and will notify you. If you prefer a different assignment, please let us know and we will accommodate.

Readings: "Database Management Systems", 3rd Ed. by Ramakrishnan and Gehrke (ISBN-10: 0072465638, ISBN-13: 978-0072465631).

Course Assessment

There will be no exams. Students accumulate up to 105 points during the course. 100 points are sufficient for full credit.

Homeworks: 5 homeworks x 13 points each = 65 points. Homeworks are assigned on Monday morning, and are due at 9am on Thursday of the following week.

Homeworks must be submitted on time. If a homework is submitted late, the student will receive no credit. Each student can make use of 2 additional days for late homework submission. If a late day is used, it is used in full. That is, if a student submits a homework 2 hours late, this counts as a full day.

- **Homework 1**: ER and relational models, assigned 9/21, due 10/1 @ 9am EDT
- **Homework 2**: SQL, assigned 10/5, due 10/15 @ 9am EDT
- **Homework 3**: Normalization, assigned 10/26, due 11/5 @ 9am EDT
- **Homework 4**: Storage and indexing, assigned 11/9, due 11/19 @ 9am EST
- **Homework 5**: Transactions, datalog, assigned 11/23, due 12/3 @ 9am EST

Project: 30 points. Projects are to be conducted in teams of two. Partners should collaborate on all aspects of the project among themselves, but they should not collaborate with anyone outside of their team. Project schedule and logistics will be announced at the start of the semester.

Attendance: 5 points

- A student can miss 2 class meetings (1 lecture and 1 lab, or 2 lectures, or 2 labs) without a penalty. For every additional lecture or lab that you miss we will deduct points, up to a total of 5% of the grade.
- Each student is assigned to a particular lab and is expected to attend that lab every week. To attend a different lab section in a given week, notify your lab TA and the TA of the other lab (the one you are attending) by email.

Participation: 5 points

- Every student is expected to participate in the lecture and in the lab they attend each week.
- Participation grade will be assigned at the discretion of the instructor and the lab TA.

Grades will be determined using this scale:

Course Grade	Points Earned
A	94-100
A-	90-93
B+	87-89
B	84-86
B-	80-83
C+	76-79
C	70-75
C-	65-69
F	less than 65

Course Schedule

09/03/2020 Introduction and overview, Entity-Relationship (ER) modeling

Reading: RG 1, 2.1-2.5 (RG = Ramakrishnan & Gehrke, "Database Management Systems", 3rd Ed.)

Assignments: none

09/10/2020 Entity-Relationship (ER) modeling continued

Reading: RG 2.1-2.5

Assignments: none

09/17/2020 The relational model

Reading: RG 3.1-3.3

Assignments:

- **Homework 1**: ER and relational models, assigned 9/21, due 10/1 @ 9am EDT

09/24/2020 Translating ER models to relational schemas (continued)

Reading: RG 3.5

Assignments:

- **Homework 1**: ER and relational models, assigned 9/21, due 10/1 @ 9am EDT

10/01/2020 Relational algebra
Reading: RG 4.1, 4.2

Assignments:

- **Homework 1**: ER and relational models, assigned 9/21, due 10/1 @ 9am EDT

- **Homework 2**: translating ER to relational model, relational algebra, assigned 10/5, due 10/15 @ 9am EDT
- **Project (part 1)**: abstract, business rules, ER diagram, assigned 10/1, due 10/29 @ 9am EDT

10/08/2020 SQL
Reading: RG 5.1 - 5.5

Assignments:

- **Homework 2**: translating ER to relational model, relational algebra, assigned 10/5, due 10/15 @ 9am EDT
- **Project (part 1)**: abstract, business rules, ER diagram, assigned 10/1, due 10/29 @ 9am EDT

10/15/2020 SQL continued
Reading: RG 5.5, 5.6

Assignments:

- **Homework 2**: translating ER to relational model, relational algebra, assigned 10/5, due 10/15 @ 9am EDT
- **Project (part 1)**: abstract, business rules, ER diagram, assigned 10/1, due 10/29 @ 9am EDT

10/22/2020 Schema refinement and normal forms, database application development
Reading: RG 19.1-19.4

Assignments:

- **Homework 3**: SQL, assigned 10/26, due 11/5 @ 9am EDT
- **Project (part 1)**: abstract, business rules, ER diagram, assigned 10/1, due 10/29 @ 9am EDT

10/29/2020 Schema refinement and normal forms (continued), database application development
Reading: RG 19.5-19.6

Assignments:

- **Homework 3**: SQL, assigned 10/26, due 11/5 @ 9am EDT
- **Project (part 1)**: abstract, business rules, ER diagram, assigned 10/1, due 10/29 @ 9am EDT
- **Project (part 2):** ER to relational translation, queries, UI, final report, assigned 10/29, due 12/07 @ 9am EDT

11/05/2020 Storage and indexing
Reading: RG 8.1-8.4

- **Homework 3**: SQL, assigned 10/26, due 11/5 @ 9am EDT
- **Homework 4**: Normalization, storage and indexing, assigned 11/9, due 11/19 @ 9am EST

- **Project (part 2):** ER to relational translation, queries, UI, final report, assigned 10/29, due 12/07 @ 9am EST

11/12/2020 Storage and indexing (continued)
Reading: RG 13.1-13.3

Assignments:

- **Homework 4**: Normalization, storage and indexing, assigned 11/9, due 11/19 @ 9am EST
- **Project (part 2):** ER to relational translation, queries, UI, final report, assigned 10/29, due 12/07 @ 9am EST

11/19/2020 Datalog, recursive SQL
Reading: Abiteboul, Hull, Vianu, "Foundations of Databases" Ch. 12
http://webdam.inria.fr/Alice/pdfs/Chapter-12.pdf

Assignments:

- **Homework 4**:Normalization, storage and indexing, assigned 11/9, due 11/19 @ 9am EST
- **Homework 5**: Datalog, recursive SQL, assigned 11/23, due 12/3 @ 9am EST
- **Project (part 2):** ER to relational translation, queries, UI, final report, assigned 10/29, due 12/07 @ 9am EST

12/03/2020 Responsible Data Management
Reading: Data, Responsibly Comics vol. 1 "Mirror, Mirror"
https://dataresponsibly.github.io/comics/

Assignments:

- **Homework 5**: Datalog, recursive SQL, assigned 11/23, due 12/3 @ 9am EST
- **Project (part 2):** ER to relational translation, queries, UI, final report, assigned 10/29, due 12/07 @ 9am EST

12/10/2020 Course project presentations

Moses Center Statement of Disability
If you are a student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

NYU School of Engineering Policies and Procedures on Academic Misconduct – complete Student Code of Conduct here

    A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the

School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B.   Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

> 1.   Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
> 2.   Fabrication: including but not limited to, falsifying experimental data and/or citations.
> 3.   Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
> 4.   Unauthorized collaboration: working together on work meant to be done individually.
> 5.   Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
> 6.   Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

NYU School of Engineering Policies and Procedures on Excused Absences – complete policy here

> A.   Introduction: An absence can be excused if you have missed no more than **10 days of school**. If an illness or special circumstance has caused you to miss more than two weeks of school, please refer to the section labeled Medical Leave of Absence.

> B.   Students may request special accommodations for an absence to be excused in the following cases:

1. Medical reasons
2. Death in immediate family
3. Personal qualified emergencies (documentation must be provided)
4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the *Coordinator of Student Advocacy, Compliance and Student Affairs* and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

NYU School of Engineering Academic Calendar – complete list here.
The last day of the final exam period is 12/20/2019. Final exam dates for undergraduate courses will not be determined until later in the semester. Final exams for graduate courses will be held on the last day of class during the week of 12/16/2019. If you have two final exams at the same time, report the conflict to your professors as soon as possible. Do not make any travel plans until the exam schedule is finalized.

Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu