CSCI-GA.2433-001, CSCI-GA.2433-002, DS-GA.2433-001
Database Systems
Fall 2021

Zvi M. Kedem

Updated 2021-08-27

Contents

1 Log of Major Changes After the First Day of the Semester

2 Required Background

It is assumed that as the students are taking a class in a graduate Computer Science program they have a suitable background, including sufficient exposure to data structures, algorithms, and programming experience. They will also need to know how to use a standard computing server environment, such as one running Linux, so as to be able to execute simple commands on it; and use SSH (secure shell) and SFT (secure file transfer protocol). Make sure before registering for the class that you have the needed background. At least install clients for those two applications and make sure that they run on your machine. We will not be able to tutor you in these, so do not register for the course until you are able to do what's described above.

3 Please Read Carefully

We need to follow some procedures so as to make the best use of time by the students, the assistants, and the instructor.

Please read this document carefully as it is also effectively the syllabus for the course. **I will assume that you are**

familiar with this document and that you will follow the instructions in it. It is particularly important that you fully understand Section 5.9. If anything in this document is unclear, please send me email in a way specified in Section 5.4.

Please refer to this document, as needed, during the semester. It may be updated with logistical information, as needed. Nontrivial changes will be listed in Section 1. Some such changes may also be marked by teal color in the text.

4 Course Description

4.1 Goals and Material Covered

We follow the official GSAS syllabus for the course.

After taking the course, you will learn how database management systems (especially relational ones, such as Oracle and MySQL) work, how to write SQL queries, and what the key issues are that need to be addressed in designing database systems, implementing them, and using them. Material covered will include all or most of:

- Modeling an enterprise's data requirements using Extended Entity Relationship Diagrams.
- Relational implementation of Extended Entity Relational Diagrams, mainly using SQL Power Architect and possibly other tools, such as MySQL Workbench.
- SQL as a Data Definition Language, a Data Control Language, Data Query Language, a Data Manipulation Language, and a Transaction Control Language. Focus on Oracle's implementation as the reference.
- Logical Design and Normalization of relational databases to optimize the logical design.
- Physical Design and concepts of query planning and optimization.
- Transaction Management: Recovery and Concurrency.
- Brief introduction to On-Line Analytical Processing (OLAP) and Data Warehouses.
- Brief introduction to NoSQL ("Not SQL" or "Not Only SQL") database management systems (document-oriented, graph-oriented, column-oriented), including new issues arising in distributed databases.

The main resources for covering the material will be:

- Detailed presentations, which will serve in lieu of a textbook.

  They are posted on the class web site, but may be slightly updated before the class using them takes placem will be posted

  It is highly recommended that you go over the slides on paper. They can be printed so several fit on one page. For an argument on why to use paper, see e.g., https://theconversation.com/why-we-remember-more-by-reading-especially-print-than-from-audio-or-video-159522
- Possible additional writeups produced by me.
- Material on the Web, as useful, to complement the posted material. The quality of the material on the web is very variable, so be careful with it.

As an integral part of the course you will practice using non-commercial and commercial software, which will be made available to you for free:

- Software for producing Entity-Relationship (ER) diagrams for specifying the information maintained by an enterprise.

  We will use draw.io https://www.draw.io/, which is free and is web-based with no installation, though local installation may work too
- Software for specifying relational databases using the standard Crow's feet notation

  We will use SQL Power Architect Community edition http://www.bestofbi.com/page/architect. It requires minimal installation.
- Oracle database system for formulating relational queries in SQ. Note that if I assign homework in which Oracle is to be used, you must use Oracle on the CIMS server with your own credentials and not any other software system. Information about obtaining access to Oracle will be provided later in the course.

Once the class roll is finalized, a file `oracle_pass.txt` with a password for you to access the Oracle Database Management System on a CIMS server will be put in your home directory on CIMS. *Do not delete it.*

Please make sure *as early as you can* that you have a CIMS email account, which you will need in order to run Oracle.

You will be asked to download and run some software. It is up to you to check that what you download is virus- and malware-free, so it is recommended that you run protective software that you find satisfactory.

5 Administrative Details

5.1 Class Organization

The departmental current plan is that the class will be conducted online.

5.2 Instructor

Zvi M. Kedem, https://www.cs.nyu.edu/kedem/.

5.3 Office Hours

They will be conducted online using Zoom.

I encourage students who have any questions or difficulties to see me as soon as possible in person. Do not fall behind!

5.4 Email

To make email communications as efficient as possible, please follow the following procedures *for all course-related emails*:

1. Please use the address mailto:zk1@nyu.edu and send your email from your official NYU email account, using your Net ID.
2. Please put in the Subject line the course number, that is **2433**, followed by an *informative phrase summarizing the contents of your email*. It is even better if your email can be fully stated in a reasonable-length subject without a body, so it can be quickly seen and responded to. *And the best subject line is phrased as a request for me to do something specific, if applicable.*

   Email with a subject not including the course number may be automatically classified as low priority by my mail reader and may even be filtered out and not seen.

   If your email requires a response from me, I will respond generally no later than the next business day during the semester.
3. Do not include several unrelated topics in a single message.
4. If you write on a new topic, do not do that by replying to an irrelevant old email.
5. Do not include irrelevant material from old email messages.

If I think that the topic of your email is of more general interest, I may post your email after redacting your identifying information. Of course, you may not want that. If you object to me possibly doing that, you need to let me know that in the body of your email.

My emails' subjects may include timestamps, such as 20300419154051Z, but you may ignore them.

5.5 Graders

Please see Section 6.2.

An assignment may have an assistant or a grader assigned to it. For each assignment, if there is an assistant/grader, it will identified in the assignment. Unless stated otherwise, you should only communicate with an assistant concerning the assignment for which it is responsible and for *no* other course-related issues.

You should only contact the graders when you need clarification of the assignments or the solutions once they are posted. But make sure that you have read *all* the files in the assignment before formulating your questions.

Do not ask a grader for any hints or help in solving the assignment; that is my job, so you should ask me. But you should review the class material before asking questions as maybe you could find an answer to your questions there. You will generally receive a reply within one business day if your email follows the format described in Section 5.4.

Do not wait with your questions until the day the assignment is due.

In communicating with a grader, to save time and effort, please follow the format I specified above for communicating with me in Section 5.4 (using, of course, the grader's email address), such as listing the course number in the subject, etc. Do not copy me on your email to a grader, unless there is a specific reason to do so. If you do not get a response within two business days, write to me.

On all "non-assignment" issues do not communicate with the graders unless I specify otherwise.

5.6 NYU-Provided Site

We will be working with an NYU-provided site.

5.7 Announcements

Announcements will be posted as needed. If you join the class late please make sure that you read the announcements already posted.

5.8 The Status Files

The tiles, one for the Monday section and one for the Friday section have been placed in the Brightspace site for the class. The file for each section lists all the dates on which the class meets and it is updated with what we have covered in each class and what we are likely to cover in the following class. There may also be some requests to go over the material and to execute some minor tasks.

In addition, the file may contain corrections and some instructions on what to do next.

The file will be updated generally within two days following a class. You need to read this file regularly, at least once between two class meetings after it was updated.

5.9 Academic Integrity

You have to read and follow all the university's, the graduate school's, and the departmental academic integrity rules as listed in https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-for-students-at-nyu.html, https://gsas.nyu.edu/about-gsas/policies-and-procedures/gsas-statement-on-academic-integrity.html , and https://www.cs.nyu.edu/web/Academic/Graduate/academic¡ntegrity.html, unless I explicitly relax them as part of the written description of an assignment.

To make some of the rules even more explicit, I add that whereas it is fine to discuss the lectures, the readings, and similar, with anybody you like, everything that you submit must be your own work only. You may neither look at anybody else's work nor show your work to anybody else. You are only permitted to discuss the assignments (other than reading assignments, if any) with me and the designated course grader, if any, as listed in the assignment. If anything is unclear, then ask. We will be available for that to the extent that you need that. Of course, whatever you state about what you did as part of your submission must not be false or misleading.

However, once a solution has been posted, you may discuss anything concerning the assignment with anybody.

Actions that violate the academic integrity rules may result in a decrease in the course grade, failing the course, failing the course without an option to repeat it, and in a dismissal from the graduate school. All of these have happened for a small number of students who took this class before. My most likely action in the case of the violation of the academic integrity rules is as follows. For a violation that I will consider as minor, I may decrease the course grade by up to one letter grade, e.g., from A to B and from A- to B-. For a violation that I will consider as major, I may decrease the grade down to F.

6 Lectures, Assignments, Exams, and Grading

6.1 Lectures

They will be conducted online using Zoom.

Although I provide very extensive written material, reading it is not a substitute for attending the classes. Please do not register for the class if you are not sure to attend the class regularly. Don't ask questions about the material if you did not attend a class fully—from its beginning to its end—unless that was unavoidable; if you did not attend the class fully, please explain why while asking your question.

I recommend not just listening passively while in the class but also taking notes to make sure that you do not miss anything that was covered. Have paper and pen near your computer.

It is important to understand the material covered in a class before the following class. Therefore, if you have any questions about the material covered in the class or in assigned reading as listed in `Status.pdf`, please review it and ask about it within a week or so.

6.2 Assignments

All the sections of the class will be given the same assignments. As the class on Monday is later than the class on Friday (sic), for example, the first Monday class takes place on September 13 and the first Friday class takes place on September 3, the assignments will be given after a Monday's class.

The assignments may include reading assignments, written assignments, and relatively short programming assignments. The number of points for each assignment will reflect its estimated difficulty and the time required to solve it. Grades and solutions will be posted. You need to make sure that the correct grade for your assignment is properly recorded once the posting of the grade is announced. Your total grade for the homework will be the sum of the grades you got on the individual assignments.

You need to follow the homework instructions precisely. Homework solutions may be rejected, required to be resubmitted with points deducted, or accepted with points deducted if they do not conform to the instructions.

When accessing CIMS machines to work on the homework, you must use only your own account.

It is strongly advised that you print out the homework assignment and you check off everything that you have finished and not do what's not requested. It is difficult, especially on a laptop with a relatively small screen, to alternate between viewing the assignment and the solution you are working on, so it is better to work with a printed copy.

If you have questions about the posted solution or your own grade for a specific homework, please communicate by email with the grader for the homework within one week from the time the grade was posted. But before that, review the relevant class material and the posted solution and any comments enclosed with it and compare it with your own. If possible, write to the grader explaining why you believe that your solution is correct. That is likely to be the most efficient way of handling the matter. If the matter is not resolved to your satisfaction, please email me.

I may be reviewing and regrading some or all of your homework. I may also review your interactions with me and the graders. Your course grade may be impacted if you do not do what's asked for or your questions show that you formulate them carelessly or without reviewing the relevant material first.

6.3 Exams

There will be no exams.

6.4 Course Grade

The numeric course grade will be the sum of the grades on the individual assignments.

The letter course grades submitted to the registrar will be "curved" to determine appropriate boundaries between the various letter grades.