COMP-421: Database Systems

Joseph D'silva
joseph.dsilva@cs.mcgill.ca
McConnel 205N

Names and Numbers

- Class:
  - Wednesdays, Fridays 04:05pm - 05:25pm

Lectures will be recorded

- Lecturer:
  - Joseph D'silva
  - joseph.dsilva@cs.mcgill.ca
  - Office Hours:
    - Fridays 10am – 12noon
    - Time permitting - after class / appointments

TAs

- TAs:
  - Yash Patel <yash.patel@mail.mcgill.ca> ➔ [In charge of projects]
  - Jianhao Cao <jianhao.cao@mail.mcgill.ca>
  - Ting Gu <yu.t.gu@mail.mcgill.ca>
  - Laetitia Fesselier <laetitia.fesselier@mail.mcgill.ca>
  - Christopher Zheng <christopher.zheng@mail.mcgill.ca>
  - Yunwen Ji <yunwen.ji@mail.mcgill.ca>

- There are no regular TA office hours. (Use my office hours if you have questions).
- TA Office hours will be scheduled based on assignment deadlines and will be announced in mycourses accordingly.
- TAs are only responsible for Projects / Assignments.
- The TAs responsible for each of the deliverables will be announced when the assignment is given out.

Marking Scheme

- Marking Scheme:
  - 3 Individual Assignments total of 12%
    - Each worth 4%
  - 1 MapReduce Exercise total of 5% (TBD)
  - 3 Project Deliverables total of 15%
    - Each worth 5%
  - Midterm 10%,
  - Final 58%

Optional – Make your midterm count towards 20% of grade and make the final 48%.

- Midterm: March – 11th, 6:00 – 9:00 pm
- Final: 3hr final covers all

Course Prerequisites

- COMP 206 Introduction to Software Systems
- COMP 251 Algorithms and Data Structures
- COMP 302 Programming Languages and Paradigms

If you do not satisfy the prerequisites, please email me a list of COMP courses, other courses with a programming element that you have taken and any other relevant experience you have which you believe will help you in this course.

Marking Scheme

- Written Assignments:
  - Solving concrete problems, from easy to difficult
  - Around one assignment every two – three weeks, three in total
  - Usually to be done within one or two weeks.
  - Late turn-in: for each day late 10% of the maximal achievable points will be subtracted from the achieved points. Maximum of 4 late days allowed unless otherwise mentioned.
  - Possible that there is one assignment around the midterm (A2) with no possibility of late turn in.
  - SQL Assignment (Part of A2) will be graded by an automated script (details will be provided later).
- There will be a 4th assignment. This will not be graded if the Map-Reduce assignment is given out. Otherwise this will instead constitute the 5% of the grade currently attributed to the Map-Reduce assignment.

Marking Scheme

- Map-Reduce Assignment: (TBD)
  - Second half of the semester
  - Using the map-reduce framework to solve large-scale queries
  - In groups
    - Same group as the project group
  - If the Map-Reduce assignment is not given out, the 4th assignment will be graded. The 4th assignment will have to be done individually like the prior the assignments.

Marking Scheme

- Project Deliverables:
  - Designing and building a database, working with the database, and writing applications using the database
  - Practical work with DB2 / PostgreSQL
  - 3 project deliverables
    - (stepwise project development)
  - To be done in a given time period
  - Four people will build a team

Project Groups

- To be done by next week!
- Find partners to build a group of 4 (exactly 4)
- Once you find sufficient people to build a group,
  - Go to mycourses
  - Go to groups menu
  - Choose an empty group and sign up to it (all 4 of you must sign up for the same group number).
- If you are not enough to build a group
  - Go to mycourses / post under "Find Partners for Building your Project Group" to find teams/members.
- If you accidentally enrolled yourself in the wrong group, email Yash to move you out of it.

Project Special

- Do you work for a company, a small start up, and you have the feeling you could need a database for some of your data management tasks?
  - Write me an email

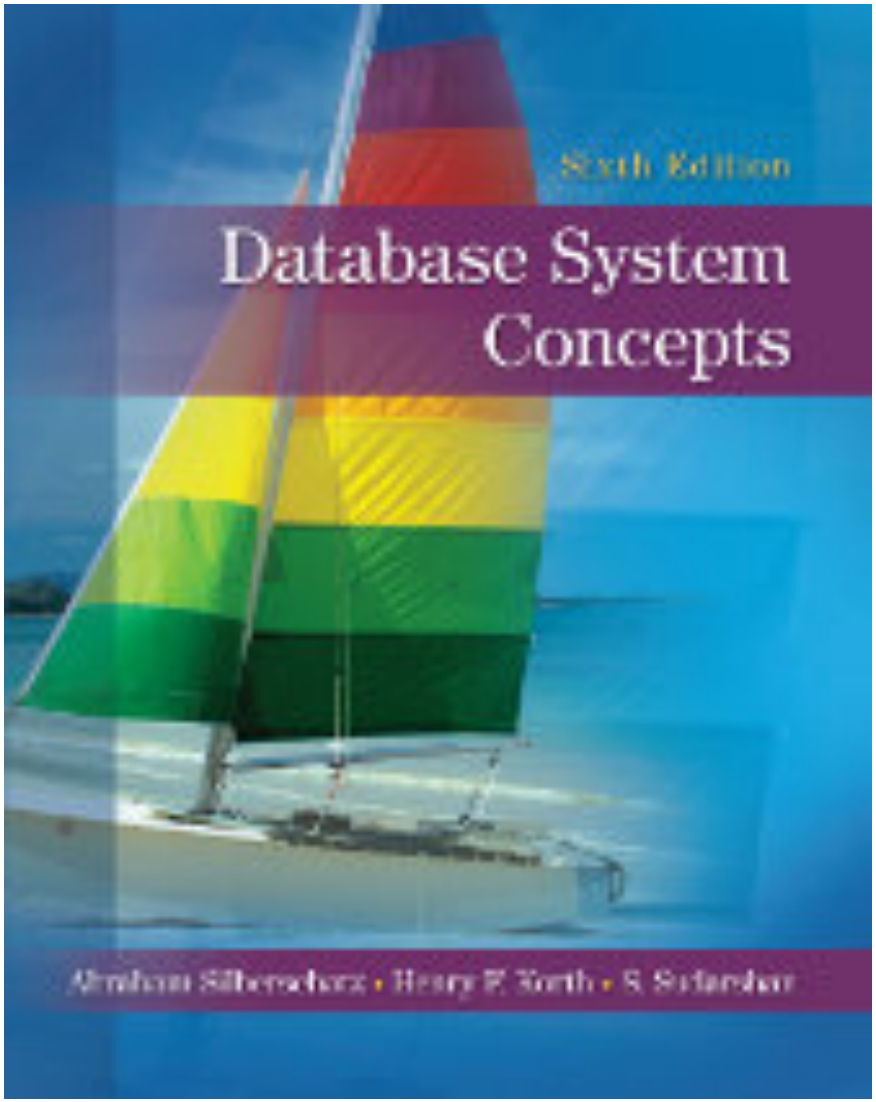
Literature

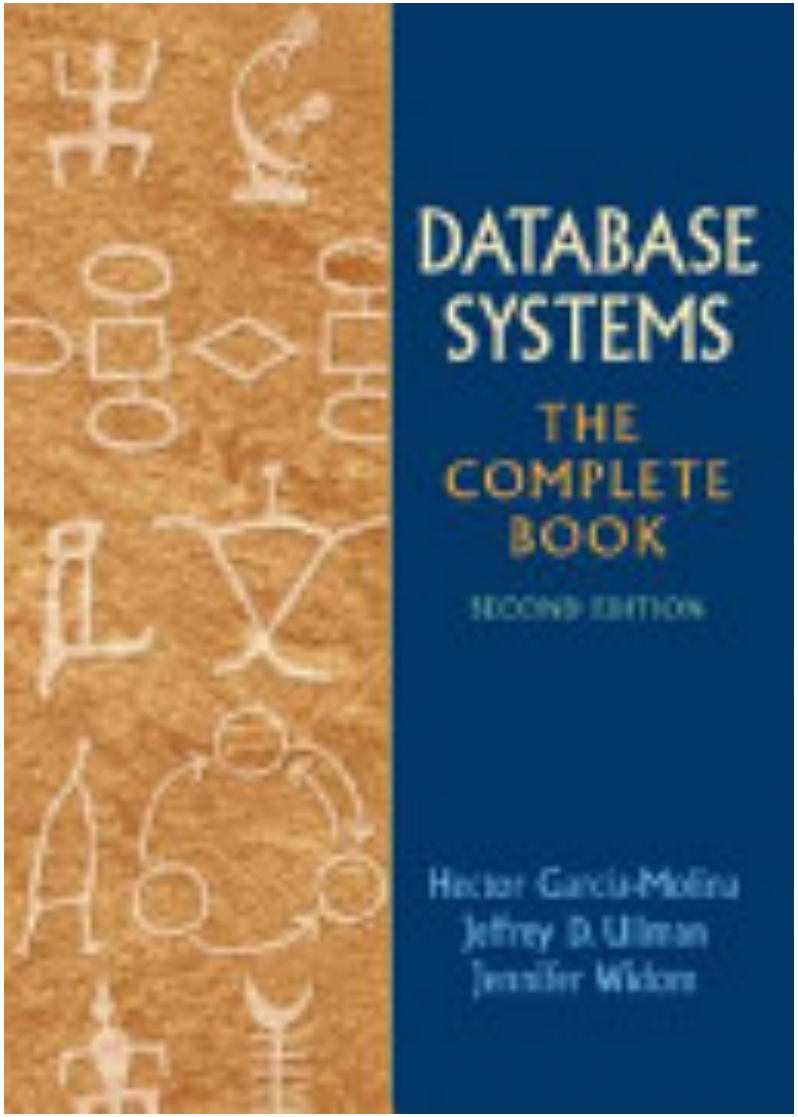
- Lecture Notes
- Books:
  - Course Book: [4 Books available in library on 48 hour loan]
    - Raghu Ramakrishnan and Johannes Gehrke: Database Management Systems; McGraw Hill, 3rd Edition
  - Other Recommendations:
    - Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom: Database Systems: The Complete Book, 2nd edition, 2008.
    - Silberschatz, Korth, Sudarshan: Database System Concepts, McGraw Hill, 5th Edition, 2006

Database Management Systems

Third Edition

Ramakrishnan • Gehrke


Sixth Edition
Database System Concepts
Abraham Silberschatz • Henry F. Korth • S. Sudarshan


DATABASE SYSTEMS
THE COMPLETE BOOK
SECOND EDITION
Hector Garcia-Molina
Jeffrey D. Ullman
Jennifer Widom

- If you are complete nuts about databases
  - Encyclopedia of Database Systems
  - Written by 800+ experts
  - 3800+ pages

Quizzes / Socrative Clicker

- Web App - https://b.socrative.com/login/student/
- iOS App - https://itunes.apple.com/ca/app/socrative-student/id477618130?mt=8
- Android App - https://play.google.com/store/apps/details?id=com.socrative.student

Academic Integrity

McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see http://www.mcgill.ca/integrity for more information).

French/English policy

- In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Active Involvement

- Use my courses for discussions
  - There's a discussion forum for every deliverable.
  - Your classmates might have the information you need already or might be interested in the same information.
  - Helps relevant TA's / lecturers to monitor and answer questions.

- Check your McGill email routinely. There might be important announcements, your team mates might be trying to reach out to you. You are responsible to be actively involved in your project group.

Course Outline

- Introduction
- Entity-Relationship Model
- Relational Model and Data Definition Language
- Relational Algebra
- SQL I: Simple Queries
- SQL II: Advanced Queries
- SQL III: Constraints
- SQL V: Application Programming
- Buffer Management
- Indexing
- Query Execution
- Query Optimization
- Key-value stores and map-reduce
- Transactions
- Concurrency Control
- Trends (Graph Databases)