Managing Data and Databases

=

Data and Database Management

INSY 437 / INSY 638
Winter 2015
Tue/Thu 4:05pm-5:25pm
Room Bronfman 045

Mahmood Shafeie Zargar
mahmood.shafeiezargar@mcgill.ca
Office: Bronfman 548
Office Hours: Tuesdays and Thursdays 5:30pm-6:00pm (or by appointment)

Ver: 15.01.6

*Note: Syllabus is subject to change; please always consult our database to ensure that you are aware of any updates.

Course Objectives

This course is designed to provide a first hands-on experience with databases as well as a general understanding of data management concepts for B.Com. student completing the concentration/major in Information Systems or interested in data and data management. This course is open to students in their 3rd year. It will count towards the IS concentration or major. It is also open to MBA students in their 2nd

year completing an IS concentration or an additional course to enhance their IT knowledge in the workplace. This course does not assume any specific technical background or programming skill, but it comprises a strong technical component and requires a willingness to invest in technical skills. This course is not a substitute for database design and development courses offered in computer science departments.

Learning Outcomes

By the end of this course, you will be able to:

- Understand the basic concepts of data storage, retrieval and transformation
- Recognize the latest trends in data storage and database use
- Apply modeling techniques to design a database
- Identify the advantages and disadvantages of different storage models (Relational, NoSQL, etc.)
- Use SQL for data definition (DDL) and data manipulation (DML) purposes
- Use a database server (MySQL)
- Design a relational database using Entity Relationship Diagrams (ERD)
- Design a CRUD or reporting interface for a pre-existing relational database
- Use the modern cloud-based app deployment platforms
- Work as a team and in a distributed manner on a database or a database application project
- Analyze the data life cycle in organizations
- Examine how data enables new business models
- Conceptualize the increasing importance of data privacy and information security for citizens and corporations
- Understand the new trends in big data, data science, data-enabled decision making and predictive analytics
- Understand the new trends in open data and release of public data

Materials

We will be using a variety of online articles and selected chapters from three textbooks that offer different perspectives on the topics covered:

- References
  - Data Management by Richard Watson, 6th Edition (Amazon - Kindle Edition)
  - The Data Revolution by Rob Kitchin (Amazon - Kindle Edition)
  - A number of additional book chapters and articles from scientific journals and the press

- Software
  - MySQLWorkbench (Oracle)
  - LibreOffice (The Document Foundation)
  - dbeaver Enterprise Edition (JKISS)

- Other resources
  - MySQL Cookbook by Paul Dubois, 3rd Edition (Safari Books Online)

- SQL for Dummies by Allen G. Taylor (Safari Books Online)
- Any online SQL tutorial (e.g. 1keydata)

Online Resources

We will be using a relational database for documents (slides, assignments, etc.), announcements, class discussion and grades. We will be using several websites (e.g. Heroku) for data hosting and data retrieval as well.

Grading/Evaluation (preliminary)

Final project (team)	40%
2 Quizzes	30%
2 Exercises (individual)	20%
Attendance and Participation	10%

Final project (team) 40%

In groups of 4 or 5 members, students will work on a project of their choice. The project will consist of working on an open dataset (e.g. Open Data from The City of Montreal) or developing a database application.

Project deliverables:

- One-page description of the project (contractual)
- One-page progress report
- The actual software produced (database, database application, visualization, pivot table, etc…)
- Project presentation
- Project documentation (instruction manual, help files, in-line code comments, etc…)

The two workshop sessions (Kickstart session and Capstone session) will also be an opportunity for you to demonstrate your performance and mastery of your project.

Quizzes (2 @ 15% each) 30%

These quizzes will cover all the material not covered in the take-home exercises (see below). This includes non-technical compulsory readings, material discussed during the class and eventual presentations by classmates or guest speakers. If you have attended and actively participated in class discussion, these should be quite straightforward. The quizzes will take about 45 minutes of the class period and will be composed of a few short essay questions.

The quiz questions are crowdsourced. Each student will contribute one or two questions to each quiz. Subsequently the questions will be curated and published online. The curated list will be your only reference for the quiz.

Exercises (2 @ 10% each) 20%

The two take-home exercises cover the technical content of the course, including design principles, SQL codes and concepts. The exercises are designed to assess your practical, rather than theoretical, understanding of the database concepts and techniques. They may require you to prepare SQL files to resolve a problem, create actual databases, run actual queries over datasets or design databases.

Attendance/Participation **10%**

Please e-mail me before class if you will be missing a session. If you miss more than two classes, your participation grade will drop. However, being in class is not sufficient for receiving full participation points. Be active and contribute to class discussion!

Course Schedule

This course schedule is subject to change; the course readings are subject to change. Please check regularly to make sure you have the updated version of the document.

Date	Class Activities	Before Class
1 Jan 6	**Introduction:** **Reconstructing course outline**	Deliverables: Install MySQL Workbench
2 Jan 8	**Databases:** **Saving, retrieving and exchanging data** *Concepts:* Data, Data Formats, Data Storage Systems	Deliverable: Open a Heroku account and add a MySQL instance Read Kitchin Ch. 1 (stop at "Temporally and spatially") Skim Kitchin Ch. 2
3 Jan 13	**Data & Business:** **Information lifecycle of the firm** *Discussing The Project*	Read Watson Section 1 - Ch. 1 & 2
4 Jan 15	**Databases:** **Save Your Lists** *Concepts:* Fields, Data Types, Null Value, Tables, Keys, Indexes, Constraints, Uniqueness and Auto-increment *Commands:* Create, Select, Insert, Update, Delete	Read Watson Ch. 3

5 Jan 20	**Data & Business:** **Data-driven business models**	Read Hartmann et al. (2014) Big Data for Big Business? A Taxonomy of Data-Driven Business Models Used by Start-Up Firms Check out the following reports and news articles: Apple, Google, Microsoft: Where does the money come from? | ZDNet Facebook: Quarterly Earning Slides Q3 2014.
6 Jan 22	**Databases:** **Retrieve Your Lists** *Concepts:* Sorting, Filtering, Aggregation, Grouping *Commands:* Where, Order By, Distinct, Group By, Aggregate Functions, Having, In, Not In, Like	Skim Watson Ch. 10
7 Jan 27	**Data & Business:** **Cloud Services and Hosted Data**	Read Bughin et al. (2010) Clouds, big data, and smart assets: Ten tech-enabled business trends to watch | McKinsey Quarterly Read Watson Ch. 19 (stop at "Conclusion—paradise postponed") Have a look at the following press article: The Details About the CIA's Deal With Amazon | The Atlantic
8 Jan 29	**Databases:** **Link Your Lists** *Concepts:* Relations, Foreign Keys *Commands:* Join	Read Watson Ch. 4 & 5
9 Feb 3	**Data & Society:** **Open Data** Check out the following websites: http://open.canada.ca http://www.donnees.gouv.qc.ca http://donnees.ville.montreal.qc.ca	Read Kitchin Ch. 3 Skim Open data: Unlocking innovation and performance with liquid information Have a look at: The Open Data Handbook

10 Feb 5	**Databases:** **Let The Links Flourish** *Concepts:* Entity Relationship Diagrams, Complex Relations, Complex Joins *Commands:* Types of join, Union *Distributing Exercise 1*	Read Watson Ch. 6 & 9
11 Feb 10	**Data & Society:** **The Data Deluge and Big Data**	Read Kitchin Ch. 4 Skim Kitchin Ch. 5 Have a look at the following press article: Why "Big Data" Is a Big Deal | Harvard Magazine Big Data: Are We Making a Big Mistake? | Financial Times
12 Feb 12	**Databases:** **ETL** *Concepts:* Data Extraction (inc. Cleaning), Transformation and Loading (ETL) *Commands:* String Functions, Data Casting, Regexp *Practice:* Importing and exporting different file formats, ETL tools	Deliverables: Exercise 1 Due Install dbeaver
13 Feb 17	**Data & Business:** **Data Science and Data-Driven Decision Making**	Deliverable: Submit one or two questions for mid-term quiz Read Kitchin Ch. 6 Have a look at the following articles: McAfee et al. (2012) Big Data: The management revolution | HBR Davenport & Patil (2012) Data Scientist: The Sexiest Job of the 21st Century | HBR
14 Feb 19	**Databases:** **CRUD Operations, Views, Reporting**	Deliverable: Install LibreOffice Check out Base Handbook 4.0 Have a look at Watson Ref. 2 (SQL Playbook)

15 Feb 24	**Workshop 1:** **Project Kickstart** *Discussing The Project:* Project mini-presentations (2-3 minutes per group). No particular preparation needed.	Deliverable: Submit one-page description of your project and register your group
16 Feb 26	**Quiz 1**	
Mar 3	**Reading Week**	
Mar 5	**Reading Week**	
17 Mar 10	**Databases:** **Modeling and Normalization** *Distributing Exercise 2*	Read Watson Ch. 7 & 8
18 Mar 12	Databases: **Dimensional Data for Data Warehousing and BI** **MapReduce, Hadoop and Distributed Data** *Discussing The Project*: Informal progress report (2-3 minutes per group). No particular preparation required.	Read Watson Ch. 13
19 Mar 17	**Data & Business:** **Information Security**	Read Upton & Sadie (2014) The danger from within | HBR Have a look at the following news articles: Has the NSA Been Using the Heartbleed Bug as an Internet Peephole? | WIRED Data Lost on 583,000 Canada Student Loan Borrowers. | Maclean's And one of these two: Stuxnet's Secret Twin | Foreign Policy The Real Story of Stuxnet | IEEE Spectrum
20 Mar 19	**NoSQL:** **Document Databases & Big Data Stacks**	Deliverable: Exercise 2 Due Read: Top 5 Considerations for NoSQL Databases Have a look at: The MongoDB 2.6 Manual

21 Mar 24	**Data & Society:** **Data Ownership and Privacy**	Read Kitchin Ch. 10 Have a look at the following news articles: The Incorporated Woman | The Economist EU Court Backs 'right to Be Forgotten': Google Must Amend Results on Request | The Guardian A Death in the Database | The New Yorker And if you have time: Lenard & Rubin (2013) The Big Data Revolution: Privacy Considerations
22 Mar 26	**NoSQL:** **Graph Databases**	Deliverable: Submit one or two questions for the final quiz Read Robinson & Webber (2013) Graph Databases. O'Reilly Media - Ch. 1
23 Mar 31	**Workshop 2:** **Project Capstone**	
24 Apr 2	**Quiz 2**	
25 Apr 7	**Project Presentations**	
26 Apr 9	**Project Presentations**	

University Policies

1. "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 www.mcgill.ca/students/srr/honest/ for more information)
"L'université McGill attache une haute importance à l'honnêteté académique. Il incombe par conséquent à tous les étudiants de comprendre ce que l'on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l'étudiant et des procédures disciplinaires (pour de plus amples renseignements, veuillez consulter le site www.mcgill.ca/students/srr/honest/)."

2. "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." (approved by Senate on 21 January 2009 - see also the section in this document on Assignments and evaluation.)
"Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté (sauf dans le cas des cours dont l'un des objets est la maîtrise d'une langue)."

Credits

Credits go to my colleagues Hani Safadi and Youngsok Bang for providing the original material. I am also indebted to Samer Faraj and Animesh Animesh for their constructive feedback on earlier drafts of the course design.