MATH 290.2 Fall 2022 (2 credits)
Course Syllabus

Balzs Zombory

Queens College, City University of New York

document last updated Saturday 27th August, 2022 9:37am

Instructor	Professor Balázs Zombory
Contact	`Balazs.Zombory@qc.cuny.edu` and at `#discussions` in our slack workspace
Time / Loc	Thursday 6:00-7:50PM on zoom
Course Homepage	https://github.com/bzombory/QC_Math_290_Fall_2022

Course Overview and Schedule

MATH 290.2 is an introduction to the fundamentals of SQL, relational database management systems, and data visualization. After completing this course you will be able to write moderately complex SQL statements, carry out basic data quality tasks, and visualize your data with the help of Power BI or Tableau. Here is a tentative schedule by week number:

1. Introduction to Relational Database Management Systems (RDBMS) - Data Integrity Concepts (Data Types, Domain and Entity Integrity, Referential and User Defined Function (UDF) Integrity), an overview of SQL implementations landscape (TSQL, Oracle, IBM, MySQL, Postgres).

2. SQL Constraints (NOT NULL, DEFAULT, UNIQUE, PRIMARY KEY, FOREIGN KEY, CHECK, INDEX).

3. SQL Syntax general review DDL vs DML (SELECT, FROM, WHERE, GROUP BY, HAVING, ORDER BY) / Introduction to Power BI setting up Power BI with AWS backend.

4. CRUD operators - An overview of CREATE, READ, UPDATE, DELETE operations, syntax and rules. Introduction to Power BI setting up Power BI with AWS backend.

5. Subqueries and Common Table Expressions - visualizing complex data in Power BI and Tableau.

6. Relationships in an RDBMS - Establishing relationships in Power BI.
7. Join Patterns (Left, Right, Full, Inner).
8. Advanced Join Patterns (Self, recursive, cross).
9. Analytical / Window functions / user defined functions.
10. Practical query optimization (Introduction to group project task).
11. Stored procedures and user defined functions in PostgreSQL.
12. Good and bad charts inin PowerBI and Tableau. We will be reviwing the main characteristics of a good data visualization.
13. Metadata and database profiling.
14. Data Quality concepts and examples.
15. Implementing DQ pipeline with SQL.
16. Summary and where to go from here (Spark, noSQL, Graph database), group project presentations.

Prerequisites

MATH 241 and CSCI 111.

Course Materials

Code snippets, run-book, and knowledge-base will be shared.

Content that will be useful during this class: PostgrSQL - a detailed documentation. PowerBI - getting started.

Computer Software: We will be using DBeaver Community Edition as our SQL IDE, PostgreSQL as our query engine, and PowerBI as our data visualization layer. Make sure that you have a GitHub and an AWS account.

Recomennded Reading

There is no required reading for this course. Having said that, reading the books below would enhance your learning experience.

1. Practical SQL: A Beginner's Guide to Storytelling with Data by Anthony DeBarross
2. How Charts Lie by Alberto Cairo

Optional Reading

The first two books on the list are very easy to read and rather entertaining for people who have data on their mind. The other four books dive deeper into concepts of PostgreSQL and Data Visualization.

1. Humble Pi by Matt Parker
2. The Data Detective by Tim Harford
3. Essential Postgres by Rick Silva
4. Storytelling with Data by Cole Nussbaumer Knaflic
5. The Art of SQL by Stephane Faroult and Peter Robson
6. Mastering PostgreSQL 13 by Hans-Jrgen Schning

Announcements

Course announcements will be made via slack in the #general channel. I will also send communication to your email in our class roaster database.

Lectures on Zoom

Classes are 110 minutes and run from Thursday, August 15th until Thursday, December 24th for a total of 16 class meetings. **Zoom policies: Keep camera on. Backgrounds are appreciated. Questions can be asked in Slack or by raising a hand.**

Lecture Uploads

I will make occasional diagrams on the black board and these images will be shared as well. Each class will have a note taker into a shared document.

Homework

There will be 9-12 homework 8-11 of which to be completed in pairs. You will be responsible to pick a coding buddy to work with during the semester by the end of the second lecture (February 10). The exercises will be posted in the #homeworks Slack channel and will be due a week later. I will grade all assignments and it will count 40% towards your grade. Each homework is equal weight and the exercises will get progressively harder. All coding exercises will be submitted via GitHub. The code will need to run on PostgreSQL 12 (if applicable). Questions are encouraged and I can be reached via slack or email. Questions are encouraged and I can be reached via slack, email, or during my office hour (1/2 hour to be more precise).

Philosophy of Homework

Research is highly encouraged, however, you need to be able to explain your answers. To ensure that we don't just copy and paste a solution from StackOverfow during every lecture a person will be picked to explain the homework they submitted. Please see details on this in the Coding Interview section.

Time Spent on Homework

Since this is 2 credit course, homework shouldn't take longer than 6 hours a week.

Late Homework

Late homework is accepted until 7 days past due date (exactly by the start of Thursday lecture). Every day your work is past due will decrease your score by 5%. For example, if you handed in a homework 5 days late and your score would have been 80% had you submitted the work on time the past due score would be 55.

Coding Interview

It is important to be able to explain your work under moderate pressure in a clear and concise manner. Most technology companies have hour long coding interviews during their candidate screening process. High scores are being awarded to people who are able to think clearly about their code. Every lecture will have a 10-20 minute long coding interview embedded in it. During this a random person will be selected from the class to explain last week's coding assignment. After the interview feedback will be provided. Class participation points are being awarded for the person who is doing the explaining and the person who will provide the feedback.

Group Project

Starting in week 10 you will be assigned to 3-4 groups where your task will be to either create a data visualization project from scratch for a pre-selected data set. Each project will be presented and graded during the final class of the semester.

Class Participation (and attendance)

Attendance will be checked every class. If you need to miss a class, please let me know in advance. Being present during the lectures, asking questions, and participating in class room discussion will give you participation credit.

Grading and Grading Policy

Your course grade will be calculated based on the percentages as follows:

Homework	40%
Group Project	30%
Coding Interview	20%
Class participation	10%

The Grade Distribution

The course has two distinct aims: make sure you can write, comprehend, and describe moderately complex SQL statements and be able to present your findings using PowerBI or Tableau with clear and insightful visualizations. If you are able to do this by the end of the semester you should expect an A.

Auditing

Auditing is not allowed in this course.