CDM CSC 455: Fall 2017-2018 Syllabus and Course Information

Section: 701 & 710
Day(s): Tuesday 5:45pm – 9:00pm in CDM 224 (Loop Campus)

Instructor: Aastha Gupta
Email: agupta13@depaul.edu
Office Hours: Wednesdays 9-9:30pm in CDM 224 & Thursdays 4-6pm in CDM 633
Course homepage: https://d2l.depaul.edu/d2l/home
Required Text: Python for Data Analysis, by Wes McKinney, O'Reily 2012. ISBN 1449319793 or 978-1449319793
Additional reading: Oracle Database 12c SQL, by Jason Price. ISBN 9780071799355 (available in DePaul books 24x7 library)

Prerequisites

CSC401 - A solid understanding of beginning level Python. If your requirement has been met through a different programming language, make sure to study the equivalent in Python. You will be expected to know all concepts and syntax covered in CSC401 such as data types, loops, iteration, functions and more. A solid understanding of these concepts prior to the start of the class will be critical to your success.

Grading

There will be a total of 5 homework assignments; assignments (with associated readings) will be posted on the course web site and will be due one week after the day they are posted, unless otherwise noted. It is your responsibility to verify that submitted files are readable, submitted on time, and in the correct locations. All submissions must be made through the course web site. No late assignments will be accepted, but your lowest homework score will be dropped in the computation of your course grade.

The 6 homework assignments will be worth a total of 40% of the course grade. There will be a midterm exam given on Tuesday, Oct 24th (week 7), and it will be worth 30% of the course grade. The final exam will be given as a take-home due on Tuesday, Nov 21st and it will be worth 30% of the grade.

Summary

This is an introductory graduate course in database design and applications. Specific topics to be covered include:

- Relational Model
- Structured Query Language (SQL)
- Database Design and Normalization
- Extract-Transform-Load (ETL) using Python
- Materialized Views and Clustered Indexes

Regarding Email Communication

Please begin the subject line of any email to me with "CSC 455" so that I can easily identify your emails. I will reply to email messages within one business day after the day I receive them; therefore, questions that are only received by me on an assignment's due date are not guaranteed replies before the assignment is due. Please plan accordingly and begin the assignments early enough to ask questions and receive answers. If you are having problems, send me a detailed description of the problems you are having; I will guide you in locating and solving your problems yourself, rather than simply solve your problems for you. For general questions, please consult the syllabus, course announcements, and course discussion forum on the course web site for answers before emailing me. Please do not use the comment field of the assignment submission system to send me questions.

Regarding Academic Integrity

You are expected to be familiar with and to adhere to DePaul's Academic Integrity Policy, which is available on-line at http://academicintegrity.depaul.edu/AcademicIntegrityPolicy.pdf. Violations of the Academic Integrity Policy will be dealt with decisively; penalties may range up to an automatic F in the course and possible expulsion.

Plagiarism includes, but is not limited to: Turning in another person's work as your own (including hiring someone else to complete an assignment for you); Starting with another person's work and modifying it to turn in as your own; Cutting and pasting, or otherwise copying, sections of another person's work into your assignment; Allowing another person (such as a tutor) to write part of your assignment; and so on. Supplying such assistance to another student is considered an equivalent violation of the policy. You may feel free to discuss the

assignments with other students at a general level. However, when it comes to actually completing your assignment, you must work independently. Your assignments must be entirely your own individual work. If you have any questions or doubts about what plagiarism entails, please consult me.

Course Schedule

Week 1 (9/12): Introduction to Databases
Introduction, Relational Model, Python Review, Schema Normalization
Week 2 (9/19): Database Design and SQL DDL
Functional Dependencies, Normalization, SQL DDL and DML, CREATE domains , INSERT, CHECK Constraints, ALTER, UPDATE and DELETE.
Assignment 1: Schema Normalization
Week 3 (9/26): SQL
SQL queries with python, SQL Developer, Use of WHERE, comparison and logical operators, Sorting queries, Aggregated Functions
Assignment 2: SQL I
Week 4 (10/3): Intermediate SQL
SQL queries with python, SQL joins (Cartesian product, equi-join, inner-join, outer-join), Nested queries, parsing data from the file.
Assignment 3: SQL II
Week 5 (10/10): Database Programming
Read/parse data from the web, data cleaning and validation, loading data into a DBMS.
Week 6 (10/17): Advanced SQL
Windowing aggregate function and time series, Customized reporting in SQL
Assignment 4: Database Programming I, due after midterm
Week 7 (10/24): Midterm
Week 8 (10/31): Database Programming
Extract-Transform-Load with Python, transformation in DBMS with python And database integrity.
Assignment 5: Database Programming II
Week 9 (11/7): Query Performance
Performance considerations in DBMS, Views, Materialized Views, Indexes, Clustered Indexes
Week 10 (11/14): NoSQL Databases

Key-value, document, columnar stores, MapReduce/Hadoop, Streaming data processing engines.

Week 11(11/21): Final Exam