

DSC 450: Fall 2023-2024 Syllabus and Course Information

Instructor: Nicholas Scope

Email: nscope@cdm.depaul.edu

Office Hours on Zoom: Wednesday 9:00-10:00pm, Friday 7:00-8:00am, or by appointment

Course homepage: <https://d2l.depaul.edu/d2l/home>

Required Text:

Python for Data Analysis, by Wes McKinney, O'Reilly 2012. ISBN 1449319793 or 978-1449319793

Grading

There are a total of 11 modules in this course (Module 0 through Module 10). There will be a homework assignment associated with each module, except for Module 6 (midterm) and Module 10 (final). Assignments or quizzes will be posted on the course web site and will be due at the end of current module, 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. Late assignments will be accepted up to three days late with a 10% penalty for each day or fraction of a day that the assignment is late; these penalties will be assessed uniformly and in full to all assignments submitted at any point beyond the posted due date and time (including those submitted or re-submitted later the same day).

Three classes will hold an in-person portion of the class on the dates of September 6th, September 27th, and November 8th. September 6th will serve as an opportunity to introduce yourself in person and meet your class while September 27th and November 8th will serve as an opportunity to ask any questions in preparation for the midterm and final.

The homework assignments and quizzes will be worth a total of 45% of the course grade. Weekly discussion posts will be worth 10% of the grade. There will be a take-home midterm exam given in Module 6, and it will be worth 20% of the course grade. There will be a final take-home exam due one week after Module 10 and it will be worth 25% of the grade.

Summary

This is an introductory course in database systems. Specific topics to be covered include:

- * Relational Model
- * Structured Query Language (SQL)
- * Database Design and Normalization
- * RDBMS programming with python and PL/SQL
- * Data analysis in python with NumPy and Pandas
- * Materialized views and indexes

Module 0

- Python tutorial and review
- Introduce yourself

Module 1

- General database intro (What is a database?)
- Relational model
- Functional dependencies
- Schema normalization

Module 2

- Functional dependencies and keys
- Normal forms and normalization
- CREATE, domains, INSERT, CHECK, CONSTRAINT, ALTER, UPDATE, DELETE

Module 3

- SQL DDL with python
- SQLite database
- SELECT, FROM, WHERE, comparisons, LIKE, ORDER BY
- Aggregate functions, GROUP BY, HAVING

Module 4

- SQL queries with python
- Joins (Cartesian product, equi-join, inner join, outer join)
- Aggregate functions
- Uncorrelated nested queries

Module 5

- Correlated nested queries
- Data cleaning and validation
- Data analysis in Python: pandas and numpy
- Loading data into DBMS

Module 6

- Midterm

Module 7

- Twitter data/parsing data from file
- Database Programming: PL/SQL
- Triggers

Module 8

- PL/SQL programming: anonymous blocks, functions, cursors
- Extract-Transform-Load with python
- Transformation in DBMS vs python

Module 9

- Performance considerations in a DBMS
- Views, Materialized views
- Indexes, Clustered indexes
- Regular expressions

Module 10

- Transactions**
- ACID properties**

Regarding Email Communication

Please begin the subject line of any email to me with “DSC 450” so that I can easily identify your messages. 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, you should consult me.