DATA 351

Data Management with SQL

MW, Smullin 222
Fall 2022

Jed Rembold, PhD
jjrembold@willamette.edu
http://willamette.edu/~jjrembold/classes/data351
Ford 214
Office Hours: MW 2:30-4, TTh 2-4, or anytime online or when my door is open!
Office Phone: (503) 370-6860

This syllabus is subject to change or adaptation as the semester progresses.

Course Description: As "big data" more and more becomes a common facet of everyday life, the bulk of attention has been focused on the analysis and usage of this information. Such a focus ignores the vital fact that, no matter how much data is gathered, little analysis is possible unless that data has been stored and organized in such a way as to be easily accessible. This course focuses on introducing the basic skills of a data engineer tasked with acquiring, storing, and maintaining such repositories of information. While other methods exist, relational databases are one of the industry standards in storing and organizing information, and so the bulk of this class will revolve around learning how to create, manipulate, query, and maintain such databases using SQL. In particular, this class will focus on the Postgresql flavor of SQL, though the majority of learned techniques will be readily applicable to any other SQL variant. Students should leave the course feeling comfortable creating and utilizing a relational database to both store and query information.

Prerequisite(s): CS-151, DS-151
Note: A minimum grade of C- is required for this course to count toward university credit.
Credits: 4.0

Text: *Practical SQL: A Beginner's Guide to Storytelling with Data* (2nd edition)
Author: Anthony DeBarros
ISBN-13: 9781718501065
Availability: New paperback copies are available online for between $30 and 40 dollars, or electronic versions can be found for closer to $30. 1st edition versions of the book are also perfectly acceptable. While not strictly required, the text is highly recommended as a supplementary reference, and the class is structured directly around it.

Course Objectives:
Over the semester, students will gain working knowledge in:

1. The basic tasks of a data engineer, including what relational databases are
2. The fundamentals of working with and querying a relational database using SQL
3. Exploring trends and descriptive statistics using SQL
4. Joining database tables to build up more complicated relationships
5. More advanced database queries utilizing text mining or spatial relationships

As the field of data science grows, so do the methods and options available to data engineers of how to store and organize the data. This introductory course focuses mainly on relational databases, but there are other modern models that may be better suited for particular types of data. The information learned in this course should prove a solid base upon which to explore those other models independently or in a future course.

Grade Weighting:

Homework	40%
Project	20%
Exam 1	10%
Exam 2	10%
Final Exam	20%

Letter Grade Distribution:

>= 92.00	A	72.00 - 77.99	C
90.00 - 91.99	A-	70.00 - 71.99	C-
88.00 - 89.99	B+	68.00 - 69.99	D+
82.00 - 87.99	B	62.00 - 67.99	D
80.00 - 81.99	B-	60.00 - 61.99	D-
78.00 - 79.99	C+	<= 59.99	F

Student Learning Objectives (SLO):
Upon completion of the course, students should be able to:

- Describe what a relational database is and what advantages they have over other forms of storing data (like spreadsheets).
- Create multiple tables in a database and parse and insert data into those tables, including specifying relationships between tables.
- Query data from a database, including using advanced filters, descriptive statistics, and joins to combine information from multiple tables.
- Use SQL for analyzing more complicated types of information, including parsing text using regular expressions and analyzing spatial geometric information.
- Gather and synthesize multiple sources of information into a relational database and then use that database to answer or provided insight into the problem of interest.

Course Assessment:

- **Homework**
  - There will be weekly homework sets which will be due Wednesday nights at 11:59pm. Homework will largely be SQL based, though there will occasionally be more conceptual questions as well. Work can be written in plain-text files originally made available through GitHub Classroom, with completed versions uploaded back to GitHub Classroom before the deadline. Assignments will be posted on the class webpage each week and the provided link at the top of the assignment should be followed to accept the assignment on GitHub Classroom and to download initial file templates and data.
- **Project**
  - There will be project at the end of the semester, in which you will showcase both parsing and writing data from multiple sources into a database, as well as then being able to use queries to extract some information to answer an interesting question. These projects will happen in pairs, and will involve a short presentation on the last day of classes.
- **Tests**
  - There will be 3 tests this semester: two midterms spaced throughout the semester and one final at the end. All will involve a mix of theory, writing correct SQL to achieve a goal, or interpreting what a piece of SQL is doing. There will be both a sample exam and a study guide made available before each exam, so students can feel prepared for the types of questions that may appear.

Course Policies:

Late Work Policy

I understand that sometimes things come up where you are unable to get an assignment in on time, and I strive to be incredibly flexible and accepting of late work. However, there also comes a point when you get too far behind to realistically keep up with the class. In an effort to compromise between the two, my late policy allots you 3 cumulative days (72 hours) of late work throughout the entire semester. So you can turn 3 assignments in 24 hours late, 6 assignments in 12 hours late, etc. without penalty. Once you have used up your 3 days (72 hours), assignments will lose 20% of their value every 24 hours late (on a continuous scale). **This means that assignments will be worth nothing after being 5 days late.** If you are approaching that point, you just need to turn in what you have so that you can move on and keep up with the class. In the case of extenuating circumstances, please just come talk to me. We'll figure out what can be done.

Incomplete Policy

An incomplete grade will only be granted in the case of prolonged illness or family emergencies that remove the student from the learning environment for an extended time period during the semester. Under no situations will an incomplete be granted due to a student falling behind through lack of motivation, understanding, or time management skills. If you are concerned about your progress and how you are doing in the class, please come visit me! We can sort out where you are struggling and work out a plan to get you back on track.

Willamette Policies:

Academic Honesty

Cheating is defined as any form of intellectual dishonesty or misrepresentation of one's knowledge. Plagiarism, a form of cheating, consists of intentionally *or unintentionally* representing someone else's work as one's own. Integrity is of prime importance in a college setting, and thus cheating, plagiarism, theft, or assisting another to perform any of the previously listed acts is strictly prohibited. I will impose penalties for plagiarism or cheating ranging from a grade reduction on an assignment or exam to failing the course. I can also involve the Office of the Dean of the College of Liberal Arts for further action. For further information, visit: http://www.willamette.edu/cla/catalog/resources/policies/plagiarism_cheating.php.

Time Commitments

Willamette's Credit Hour Policy holds that for every hour of class time there is an expectation of 2-3 hours work outside of class. Thus, for a class meeting three hours a week, you should anticipate spending 6-9 hours outside of class engaged in course-related activities. Examples include study time, reading and homework, assignments, research projects, and group work.

Diversity and Disability

Willamette University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. Our goal is to create learning environments that are usable, equitable, inclusive and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please notify me as soon as possible. Students with disabilities are also encouraged to contact the Accessible Education Services office in Smullin 155 at 503-370-6737 or accessible-info@willamette.edu to discuss a range of options to removing barriers in the course, including accommodations.

Tentative Course Outline:

The weekly coverage may change slightly as the semester progresses, but this should serve as a rough guide.

Week	Date	Chapter	Description	Due
1	Mon, Aug 29	Ch 1	First Database and Table	
	Wed, Aug 31	Ch 2	Using SELECT	
2	Mon, Sep 05		*Labor Day*	
	Wed, Sep 07	Ch 3	Understanding Data Types	HW 1 due
3	Mon, Sep 12	Ch 4	Importing and Exporting Data	
	Wed, Sep 14	Ch 5	Math and Stats with SQL	HW 2 due
4	Mon, Sep 19	Ch 5	Math and Stats with SQL	
	Wed, Sep 21	Ch 6	Joining Tables	HW 3 due
5	Mon, Sep 26	Ch 6	Joining Tables	
	Wed, Sep 28		**Test 1 (Ch 1-5)**	
6	Mon, Oct 03	Ch 7	Designing Tables	
	Wed, Oct 05	Ch 7	Designing Tables	HW 4 due
7	Mon, Oct 10	Ch 8	Grouping and Summarizing	
	Wed, Oct 12	Ch 9	Inspecting and Modifying Data	HW 5 due
8	Mon, Oct 17	Ch 9	Inspecting and Modifying Data	
	Wed, Oct 19	Ch 11	Dates and Times	HW 6 due
9	Mon, Oct 24		Window Functions	
	Wed, Oct 26		**Test 2 (Ch 6-11)**	
10	Mon, Oct 31	Ch 12	Advanced Queries	
	Wed, Nov 02	Ch 12	Advanced Queries	HW 7 due
11	Mon, Nov 07	Ch 13	Mining Text	
	Wed, Nov 09	Ch 13	Mining Text	HW 8 due
12	Mon, Nov 14	Ch 14	Spatial Data with POSTGIS	
	Wed, Nov 16	Ch 14	Spatial Data with POSTGIS	HW 9 due
13	Mon, Nov 21		*Fall Break*	
	Wed, Nov 23		*Fall Break*	
14	Mon, Nov 28		Webscraping	
	Wed, Nov 30	Ch 15	Views, Functions, and Triggers	HW 10 due
15	Mon, Dec 05	Ch 15	Views, Functions, and Triggers	
	Wed, Dec 07	Ch 17	Project Presentations	
16	Wed, Dec 14		**Final**	