CHARLOTTE SCHOOL OF DATA SCIENCE

DSBA-HCIP 6160: Database Systems for Data Scientists

Instructor -- **Robert Fox** (robert.fox@charlotte.edu)
Teaching Assistant -- **Elliott Kervin** (ekervin@charlotte.edu)

Course Prerequisite(s):
Graduate/Ph.D. student standing or permission of instructor.

Course Description:
This course focuses on the fundamentals of relational databases and SQL, with hands-on practice in MySQL. Students will learn to design clean, efficient schemas and to write advanced analytical queries tailored to real-world analysis workflows. This course covers database integration for analysis, usage with Python and explores the role of embedded databases in local and edge analytics. The course introduces students to specialized data systems, including time series, vector, and graph databases, and discusses where and why they are used. This course also includes a survey of large-scale analytical data infrastructure: data warehouses, data lakes, cloud and distributed databases, MPP systems, PySpark, Apache Iceberg, and related platforms that power big-data analytics in industry. By the end, students will understand not only how to write SQL, but how to apply database design and querying principles in the broader context of contemporary data science practice.

Student Learning Objectives:
- Gain a foundational understanding of Relational Database Management Systems.
- Establish a strong proficiency in Structured Query Language (SQL) to accomplish efficient data analysis.
- Be able to define and implement normalized Entity-Relationship data models.
- Gain experience integrating modern analytics environments (e.g. python-based Jupyter notebooks) and RDBMSs for efficient and effective data science work.
- Gain knowledge of large-scale and distributed data platforms / technologies.
- Gain exposure to non-relational data stores and specialized data stores.
- Understand the basics of database management and data governance.

Student Learning Outcomes. Students will:

1. Demonstrate advanced proficiency in Structured Query Language (SQL)
2. Design, implement, and utilize a relational database to support efficient and effective data analysis.
3. Demonstrate expertise in using relational databases for analytics.
4. Be able to identify, describe, and analyze the key factors and trade-offs of different data storage technologies / platforms with respect to data analysis.

HCIP Students:

5. Apply best practices in the design of new and/or critique of existing population health data sources.

Students in the HCIP section will have the opportunity to work with a health data source and be assessed using a project-based assignment.

Textbook:

Required: Murach's MySQL 4th Edition, Murach and Associates November 2023. Murach Press. ISBN 9781943873104

Other Resources:

Access to DataCamp and access to cloud-based resources / services / products for databases learning will be provided to the students free of charge. (e.g. Snowflake)

Internet:

Students must have reliable access to the Internet.

Laptop:

- Students must have a laptop capable of running MySQL
- 64bit x86 CPU, 4GB RAM (8GB+ recommended), 5GB+ free disk space

Software:

This class will use open-source (and free) software.

- MySQL and MySQL Workbench, a popular open-source database management system (required)
- Python programming language; Jupyter notebooks; and miscellaneous python libraries for databases and data science.
- Potentially other software as needed

Schedule:

This is the approximate weekly schedule for the course. Please note that this schedule may change! Monitor the "Assignments" page in the course Canvas site to understand timing and due dates.

Week 1	**Introduction to Databases; Relational Databases, MySQL Setup**
Week 2	**Data Types & Structured Query Language (SQL) Query Basics**
Week 3	**DML (Data Manipulation Language): Inserts, Updates, & Deletes**
Week 4	**SQL Joins & Subqueries**
Week 5	**Relational Modeling, Entity-Relationship Diagrams, Data Modeling, & Normal Forms**
Week 6	**Relational Structures: DDL, Indexing, Views, & Temporary Tables**
Week 7	**Advanced SQL Aggregation & Functions**
Week 8	**User Defined Functions, Stored Procedures, & Triggers**
Week 9	**Databases & Integration for Analysis; Usage Patterns w/Python; Embedded Databases; Connectors & Database Abstractions; Specialized Databases: Time Series, Vectors, Graphs, and beyond**
Week 10	**Analyzing Big Data; Data Warehouses, Data Lakes, Cloud & Distributed Databases, MPP Databases, PySpark, Apache Iceberg, & other platforms and structures for large-scale data analysis.**
Week 11	**Database Administration & Governance; Skills Review**
Final Week	**Project Presentations**

Grading:

Please reference the "Grades" page on Canvas.

30% Assignments (Murach, Other)

See the "Assignments" section on the Canvas "Assignments" page.

20% Tutorials (DataCamp, Other)

See the "DataCamp" section on the Canvas "Assignments" page.

30% Quizzes

4 Quizzes, 100pts Each:

Quiz 1: SQL Basics

Quiz 2: Intermediate SQL

Quiz 3: Advanced SQL

Quiz 4: Overall: Databases for Data Science

20% Group Project

See the "Project Deliverables" section on the Canvas "Assignments" page.

Grading Scale for Course:

A 90-100 B 80-89 C 70-79 D 60-69 F Below 60

Policy on Late Assignments:
Late Assignments (assignments submitted past the due date) will only be accepted with prior approval (email professor and TA ahead of due date). Assignments more than a week late will have 10 points automatically deducted unless extenuating circumstances are present as documented by the student and approved by the professor.

Policy on the Use of AI Tools, Other:
The following materials, equipment, websites, or tools are prohibited for completing course assignments, quizzes, or other assessments unless the professor explicitly permits such use for legitimate pedagogical purposes:
ChatGPT, other artificial intelligence tools, online course material suppliers such as Course Hero or Chegg, etc. If permitted for use on specific assignments, students must carefully follow guidance on appropriate use and citing the sources.

Project:
Our course project will provide you the opportunity to explore and experience database design and programming/analysis in practice. You will collaborate with other students in this course as part of a group. The project will be assigned at the mid- point of the semester and each group will have the chance to choose between several projects provided by the professor. A group can pitch an idea for an original database project as well.

The project will require proper design, development and implementation of a relational database that models a real-world domain. The project has several milestones in the form of project deliverables in order to keep your work progressing. Project deliverables must be met; no late work will be accepted. Students have the chance to correct deficiencies on their deliverables in all but the final project deliverable.

Projects will be supported by SQL code and a presentation demonstrating the project. Students can learn from each other and this is a great opportunity to share ideas and techniques!

Communication:
- Each student is given an email account by UNC-Charlotte ("@charlotte.edu"). This is the account that will be used by your instructor. Canvas announcements about changes to class assignments or other course information will be posted online and may be sent to you. Check your email daily. Do not send email to your instructor from any other account, as it will be considered spam, and deleted.

- Questions, Comments, and Requests:
    - For any questions or clarification of class material, please ask them on the Discussion Board in Canvas whenever possible. Everyone in the class is encouraged to help answer the questions. If satisfactory answers do not

emerge, the instructor and/or TA will answer.

- For any other comments or requests, please send email to the instructor and TA. When emailing your instructor, please use a specific subject line starting with "DSBA-HCIP 6160: " Please copy the TA(s) on assignment related questions.

Attendance:
Attendance is encouraged but not graded in this course. Unless stated otherwise, exams will be administered in-person only. Students who attend generally achieve better academic results. **Keeping Up with the Class:** Make sure to check the "Assignments" area on Canvas regularly to identify what assignments are due (or past due) for you.

Student Responsibilities:
Please refer to University Policy 406 - The Code of Student Responsibility (http://legal.uncc.edu/policies/up-406) for specific information. In addition to the responsibilities specified by the University, for this course, it remains the student's responsibility to be aware of enrollment status, assignment due dates, changes to the syllabus, and deadlines for the UNCC academic calendar. Each student is responsible for his/her attendance and properly withdrawing from the course if necessary.

Class Recordings:
Class sessions may be audio- and/or video-recorded by the instructor for the purposes of student-participant reference and access by other students enrolled in the same course (including students enrolled in different class sections or break-out groups). **Students are not permitted to make their own recordings of class sessions or to share or distribute University recordings of class sessions.** NOTE: Students with specific electronic recording accommodations authorized by the Office of Disability Services may record classes; however, the instructor must be notified of any such accommodation prior to recording. Any distribution of such recordings is prohibited.

Disability Accommodations:
Students in this course seeking accommodations to disabilities must first consult with the Office of Disability Services and follow the instructions of that office for obtaining accommodations.

Non-discrimination:
All students and the instructor are expected to engage with each other respectfully. Unwelcome conduct directed toward another person based upon that person's actual or perceived race, actual or perceived gender, color, religion, age, national origin, ethnicity, disability, or veteran status, or for any other reason, may

constitute a violation of University Policy 406, The Code of Student Responsibility. Any student suspected of engaging in such conduct will be referred to the Office of Student Conduct.

The instructor affirms people of all gender expressions and gender identities. If you prefer to be called a different name than what is indicated on the class roster, please let me know. Feel free to correct me on your preferred gender pronoun. If you have any questions or concerns, please do not hesitate to contact me.

Academic Integrity:
All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code and on the Student Accountability & Conflict Resolution website. The Code is available from the Dean of Students Office or online at legal.charlotte.edu/policies/up-407. Additional resources are available on the Student Accountability & Conflict Resolution website.

Faculty may ask students to produce identification at examinations and may require students to demonstrate that graded assignments completed outside of class are their own work.

Copyright:
Lectures and course materials, including presentations, tests, exams, outlines, and similar materials, are protected by copyright. The creators of the material are the exclusive owners of copyright of those materials. I encourage you to take notes and make copies of course materials for your own educational use. However, you may not, nor may you knowingly allow others to reproduce or distribute lecture notes and course materials publicly without express written consent. This includes providing materials to commercial course material suppliers such as CourseHero, Chegg, and other similar services. Students who publicly distribute or display or help others publicly distribute or display copies or modified copies of an instructor's course materials may be in violation of University Policy 406, The Code of Student Responsibility, or University Policy 407, Code of Student Academic Integrity. Similarly, you own copyright in your original papers and exam essays. If I am interested in posting your answers or papers on the course web site, I will request your written permission.

Disclaimer:
This syllabus is intended to give the student guidance in what may be covered during the semester and will be followed as closely as possible. However, the professor reserves the right to modify, supplement and make changes as needed.