

DSBA-HCIP 6160: Database Systems for Data Scientists

Course Information

Course Number/Section **DSBA-HCIP 6160**

Course Title **Fall 2022**

Days & Times *Thursday at 5:30pm*

Location *City Center, Rm 601*

Instructor Contact Information

Instructor **Stephen Rohrer**

Email Address **srohrer@uncc.edu**

Zoom:

Office Hours 7-8:30pm Thursdays following in-class sessions

email or call/text 252-267-2035 (cell)

TA Sakshi Gupta

Email Address **sgupta46@uncc.edu**

Office Hours 11am-1pm Monday by appointment via Zoom

Course Prerequisites

- Graduate/Ph.D. student standing or permission of instructor.

Course Description:

DSBA-HCIP 6160: Database Systems for Data Scientists

This course covers the design, modeling, programming, aggregation, and analysis of big data. We will primarily focus on relational, non-relational and semi-structured data and some of the key languages and tools used for each – SQL, Python, R, etc. Topics will include: (1) modeling/theory: basics of relational database management systems (RDBMS), database design; (2) programming: SQL and NoSQL query languages as well as languages used to work with semi-structured data; (3) aggregation and functions for reporting: ETL, data warehousing, OLAP; (4) modern cloud approaches; and (5) data acquisition and ingestion for analytics.

Student Learning Objectives

The objectives of this course are to learn how to:

- Install and configure RDBMS tools - primarily MySQL and PostgreSQL
- Define and implement data models based upon Entity-Relationship and normalization concepts while learning to work with data modeling tools
- Leverage advanced understanding of Structured Query Language (SQL) to define data structures and to perform both Create-Read-Update-Delete (CRUD) operations and complex reporting queries
- Develop knowledge of big data/cloud/NoSQL approaches in contrast to more traditional relational systems
- Understand enterprise data lifecycle concepts as data moves downstream from online transaction processing (OLTP) systems, through Extract-Transform-Load (ETL) frameworks and into data warehouses and online analytical processing (OLAP) systems.
- Gain experience loading data into an analytics environment from a database.

Student Learning Outcomes. Students will:

1. demonstrate advanced proficiency in SQL programming
2. design, implement, and utilize a fully normalized relational database system that meets organizational specifications using SQL.
3. be able to describe and analyze key concepts related to SQL, NoSQL, Cloud and Data warehouses and explain when it is best to use a particular DB in order to meet organizational needs.
4. Demonstrate expertise in data preparation and data analytics using big data

HCIP Students:

5. Apply best practices in the design of new and/or critique of existing population health data sources. Assessment embedded within the last of the 3 small group projects.

Students in the HCIP section will have the opportunity to work with a health data source and be assessed using a project-based assignment. This information is a part of the student learning outcomes for HCIP.

Grading and Assessment Criteria: Outcomes are assessed by:

- 20% Assignments

Learning Outcomes 2, 4, 5

HW 1. ER Modeling

HW 2. Database Design: Normalization

- 30% Datacamp and Qwiklabs Tutorials

Learning Outcomes 1, 3, 5

TUT 1: Datacamp: Introduction to SQL

TUT 2: Datacamp: Introduction to Relational DB in SQL

TUT 3: Datacamp: Database Design

TUT 4: Datacamp: Joining Data in SQL

TUT 5: Datacamp: Intermediate SQL

TUT 6: Datacamp: Exploratory Data Analysis in SQL

TUT 7: Datacamp: PostgreSQL Summary Stats and Window Functions

TUT 8: Datacamp: Functions for Manipulating Data in PostgreSQL

TUT 9: Datacamp: Data-Driven Decision Making in SQL

TUT 10: Datacamp: NoSQL Concepts

TUT 11: Datacamp: Reporting in SQL

- 20% Murach Chapter Projects: (Individual and Group)

Learning Outcome 2

Average of HW Assignments from Murach DB Exercises from Chapters

- 15% Exam: Final

Evaluation of Learning Outcomes 1-4

(Objective (timed, online, terminology and concepts, open notes, includes SQL)

- 15% Group Project

Grading Scale for Course:

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

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.

TEXT:

Required: Murach's MySQL 3rd Edition, Joel Murach 2019. Murach Press.

Optional: Databases Illuminated by Ricardo, Urban. 3rd Ed. Jones and Bartlett, 2017.

OTHER RESOURCES: Access to Datacamp and Qwiklabs will be provided free of charge to students enrolled in the class. Students will create accounts using their @uncc.edu email.

We will use Dr. Daniel Soper's Video Database series to provide a video supplement on Relational Database Concepts.

SOFTWARE: Students will be able to use MAC and Windows. Students must have access to the internet

- MySQL and MySQL Workbench, a popular open-source database management system (required)
 - Instructions for installing the software will be available as a part of assignments
- PostgreSQL, an alternative popular open-source database used across multiple cloud-hosted and enterprise data warehouse technologies

CLASS MEETING SCHEDULE:

Date	Event
August 25, 2022	Week 1 Class
September 1, 2022	Week 2 Class
September 8, 2022	Week 3 Class
September 15, 2022	Week 4 Class
September 22, 2022	Week 5 Class
September 29, 2022	Week 6 Class
October 6, 2022	Week 7 Class
October 13, 2022	Week 8 Class
October 20, 2022	Week 9 Class
October 27, 2022	Week 10 Class
November 3, 2022	Week 11 Class
November 10, 2022	Week 12 Class
November 17, 2022	Week 13 Class
November 24, 2022	Week 14 Class - Thanksgiving - No Classes
December 1, 2022	Week 15 Class
December 8, 2022	Week 16 Class
December 15, 2022	Week 17 - Final Examinations

ASSIGNMENTS AND ACADEMIC CALENDAR:

The following class schedule and deadlines are subject to change at the discretion of the instructor and class circumstances. All assignments are due by the start of the next class which is generally Thursday 5:30 pm unless otherwise indicated.

<u>LESSON</u>	<u>TOPIC</u>
WEEK 1	Introduction to Data 1) Syllabus review 2) Murach Chap 1 3) HW - DataCamp TUT 1
WEEK 2	SQL Query Basics 1) Murach Chap 2 2) Murach Chap 3 3) Murach Chap 8 4) HW - Murach Chap 2
WEEK 3	ER Modeling and Normalization 1) Lecture 1 2) Murach Chap 10 3) HW - DataCamp TUT 2 4) Group Assignment - Normalization
WEEK 4	SQL Inserts, updates, deletes 1) Murach Chap 5 2) HW - DataCamp TUT 3
WEEK 5	SQL Joins 1) Murach Chap 4 2) Chap 4 Homework
WEEK 6	Subqueries 1) Murach Chap 7 2) HW - DataCamp TUT4
WEEK 7	SQL Functions and Summary Functions 1) Murach Chap 9
WEEK 8	Summary and Analytical Queries 1) Murach Chap 6
WEEK 9	Views 1) Murach Chap 12

WEEK 10		Stored Procedures 1) Chap 13
WEEK 11		User Defined Functions 1) Chap 15
WEEK 12		Data Warehousing 1) Lecture 3
WEEK 13		Data Acquisition and Ingestion for Analytics 1) Lecture 5
Week 14		Thanksgiving - NO CLASS
WEEK 15		NoSQL Concepts 1) Lecture 4
WEEK 16		Review and Test prep 1) Practice quiz
Week 17		Final Exam

Project

Our course project will provide you the opportunity to explore and experience database design and programming 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 database that addresses an opportunity. Front-end and back-end components will be in place to provide common functions (create, read, update, delete) and data queries for browsing, searching and filtering. The project is not trivial but not so complex that it requires more than the time allotted. Resources required to host your project will typically be a MySQL database and web or application interface. Your project can be hosted locally but web or cloud is recommended.

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!

All project details will be provided in the formal project assignment made no later than the 8th week of class.

Course Format and Activities

This course is designed to mirror the curriculum developed by Dr. Pam Thompson. This course will draw materials primarily from the textbook and handouts/materials posted on the course website. Students will study the materials and complete all the course requirements. In order to properly address the assignments for this class, you will need to put in a considerable amount of time and energy. Please log on often to check for announcements, assignments, course documents, news forums, grades daily to stay informed.

Reading:

The readings for this course will be taken from the textbook and a variety of other current sources. Students must read the course materials and post any questions that you wish to be discussed on the forum.

Group discussion:

The most vital use of Discussion Forums is to exchange ideas with other classmates. It is important that you check into the forums regularly. You are encouraged to ask

questions regarding the required readings, discuss the unit topics, share information and resources with classmates, and respond to problems posted by your classmates or instructor. You should read everyone's posts and responses to the topics that interest you.

Submission of Work:

- Follow each assignment instruction; all work should be uploaded into the assignment link as requested. Datacamp Assignments are graded in Datacamp. It is the students' responsibility to keep his/her copies of all work submitted to the instructor. All work is to be turned in by the due date, no late work will be accepted.

Policy on Academic Integrity: The university policy 407, the Code of Student Academic Integrity, applies. This policy is available at <http://legal.uncc.edu/policies/up-407>.

Academic honesty is absolutely essential. Cheating, plagiarism or other academic misconduct will not be tolerated. If you are caught cheating, you will not pass this course and will be subject to any and all penalties specified in the code of Student Academic Integrity. **If a student is found cheating, she or he will receive a ZERO for that assignment. If a student is found cheating a second time, she or he will receive an "F" for the course.**

Examples of violation academic integrity include, but are not limited to:

- pretending that somebody else's work is yours so that you can get a higher grade than your own work merits
- falsifying data
- lying in order to extend a deadline or gain some other special advantage
- helping other people to do any of these things
- copying answers on tests
- using prohibited reference materials (such as notes or books) during an exam
- turning in papers that you have not written yourself or that you wrote for a different course
- quoting material without marking it as quoted and without attributing it to its source (or closely paraphrasing material without attributing it to its source)
- misrepresenting a medical or family emergency or other personal contingency in order to delay a scheduled exam or to get extra time on an assignment
- pretending to have a disability you do not have (or exaggerating one you do have) in order to gain an unwarranted advantage unavailable to other students
- modifying graded material and then resubmitting it to "correct the error in grading"

Rules Governing Students with Special Requirements

Students with disabilities which require accommodations should:

1. Register with the Office of Disability Support Services and 504 Compliance to provide documentation
2. Bring the necessary information indicating the need for accommodation and what type of accommodation is needed. This should be done during the first week of classes or as soon as the student receives the information. If the instructor is not notified in a timely manner, retroactive accommodations may not be provided.

Miscellaneous Requirements

1. All requests to change grading of any course work must be submitted in writing within a week after the grades are made available. Requests must be specific and explain why you feel your work deserves additional credit.
2. All requests about missing (or zero) grades must be submitted in writing to the instructor within a week after the grades are announced. After that period the grade stands.
3. Please note that a student will not automatically receive an "I" grade when he/she misses some work, or misses the final exam. An "I" is given to those students who have a passing average at the time of the 'incident'. I grades must go through a formal approval process and must be based on extenuating or emergency circumstances according to UNCC policy.
4. Submission of work: It is the student's responsibility to ensure that the instructor has received work submitted. This is especially important when work is submitted electronically.
 - a. If you use email, ensure that you keep a copy of the sent email, and ask for a 'read receipt'.
 - b. If submitting via our online course site Canvas, always keep a copy of your work.
5. Communication Protocol:
 - (a) 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 comments or requests, please send email to the instructor and TA..
 - (b) Canvas
 - Announcements will be posted in Canvas. Make sure to check the assignment area frequently enough to stay informed.
 - There are obviously things that are not appropriate for the Canvas discussion area, such as solutions for assignments (violation of honor code).
 - (c) Emails
 - Each student is given an email account by UNC-Charlotte. This is the account that will be used by your instructor. 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 be deleted.

- Please use Canvas, not emails, for general questions, unless you wish to keep your questions or comments private.
- When emailing your instructor, please use a specific subject line starting with "DSBA-HCIP 6160: Homework 1 - [Last Name]."

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.

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.

Good luck in class! I am looking forward to working with you this Fall and sharing my knowledge.