

DSBA-HCIP 6160: Database Systems for Data Scientists

Course Information

Course Number/Section **DSBA/HCIP 6160**
Course Title **Fall 2024**
Days & Times *Thursday at 5:30pm*
Location *Dubois, Room1102*

Contact Information

Instructor Dr. Shih-Hsiung Chou
Email Address schou6@charlotte.edu
Office Hours Email or Zoom by appt.
TA Sai Divya Lanka
Email Address slanka2@charlotte.edu
Office Hours Email or Zoom by appt.

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 data within database systems. We will primarily focus on relational, non-relational and semi-structured data and some of the key languages and tools used for each including SQL. 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 – MySQL or 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
5. Apply best practices in the design of new and/or critique of existing real-world MIMIC-IV data.

Grading and Assessment Criteria: Outcomes are assessed by:

- 15% Assignments

Learning Outcomes 1, 2, 4, 5

*All Murach's Exercises will be provided on Canvas, **NOT** the exercises in the book!*

1. Install software and insert data (MySQL, MySQL Workbench, 3 datasets provided by Murach's MySQL; Murach Ch2 Exercise)
2. SQL Basics skills (Murach Ch. 3-4 Exercise)
3. SQL Intermediate (Murach Ch. 5-9 Exercise)
4. Database Design and Implementation (Murach Ch. 10 - 12 Exercise)
5. SQL Advanced (Murach Ch. 13-16 Exercise)
6. Database Administration (Murach Ch. 17-19 Exercise)
7. Host MySQL on AWS (Murach Ch. 20 Exercise)
8. Snowflake Exercise
9. MongoDB Exercise

- 5% Attendance
- 35% Datacamp and Tutorials
Learning Outcomes 1, 2, 3, 5
 - 1: Introduction to SQL
 - 2: Intermediate SQL
 - 3: Joining Data in SQL
 - 4: Data Manipulation in SQL
 - 5: PostgreSQL Summary Stats and Window Functions
 - 6: Functions for Manipulating Data in PostgreSQL
 - 7: Database Design
 - 8: Introduction to NoSQL

- 15% Quizzes
Evaluation of Learning Outcomes 1-4

Quiz 1: SQL Basics

Quiz 2: Intermediate SQL

Quiz 3: Database Design and Implementation

Quiz 4: Advanced SQL, Data Warehouse

Quiz 5: Database Administration, Cloud, NoSQL

- 35% Group Project

Grading Scale for Course:

A 90-100 B 80-89 C 70-79 U 69 and below

Please note, that I will not round up to another grade level. For example, if you get a '89.9', it will be a B.

Late Assignments, Test Grades, and Group Project Grades:

Late Assignments (assignments submitted past the due date) will receive 5 points off for every day it is late without prior written approval with the TA or professor. Assignments over a week late can still receive a 50 so long as it is turned in prior to the final class date. Assignments never submitted or completed will receive a 0. Tests can not be retaken without written approval from the professor.

Group project grades are based on the group leader submitting assignments on time. Participation in group projects and assignments IS REQUIRED! Points can be taken off at the professor's discretion due to lack of participation.

TEXT:

Required: Murach's MySQL 4th Edition, Joel Murach 2019. Murach Press.

OTHER RESOURCES: Access to DataCamp, AWS, Snowflake, and MongoDB will be provided free of charge to students enrolled in the class. Students will create accounts using the @charlotte.edu email.

SOFTWARE:

- Students will be able to use MAC, Windows, or Linux. Students must have access to the internet. (required)
- 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 (optional)

CLASS MEETING SCHEDULE:

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.

Date	Event
August 22, 2024	<p>Week 1</p> <ul style="list-style-type: none"> ▪ Syllabus review ▪ Murach Chap 1 An Introduction to Relational Database ▪ Murach Chap 2 How to use MySQL Workbench and other Development Tools ▪ Installation of MySQL database and MySQL Workbench ▪ HW – <ul style="list-style-type: none"> - Register DataCamp and post your “Class Introduction” on Discussion - Murach Chap 2 Exercise - DataCamp: 1. Introduction to SQL
August 29, 2024	<p>Week 2</p> <ul style="list-style-type: none"> ▪ Murach Chap 3 How to Retrieve Data from a Single Table ▪ Murach Chap 4 – How to Retrieve Data from Two or More Tables—How to work with inner joins ▪ HW- <ul style="list-style-type: none"> - Murach Chap 3 Exercise - DataCamp: 2. Intermediate SQL
September 5, 2024	<p>Week 3</p> <ul style="list-style-type: none"> ▪ Murach Chap 4 – How to Retrieve Data from Two or More Tables—Rest sub-sections

	<ul style="list-style-type: none"> ▪ Murach Chap 5 How to Insert, Update, and Delete Data ▪ Quiz 1: SQL Basics ▪ HW- <ul style="list-style-type: none"> - DataCamp: 3. Joining Data in SQL - Murach Chap 4 - 5 Exercise
September 12, 2024	<p>Week 4</p> <ul style="list-style-type: none"> ▪ Murach Chap 6 How to Code Summary Queries ▪ Murach Chap 7 How to Code Subqueries ▪ HW- <ul style="list-style-type: none"> - Murach Chap 6 – 7 Exercise
September 19, 2024	<p>Week 5</p> <ul style="list-style-type: none"> ▪ Murach Chap 8 How to Work with Data Types ▪ Murach Chap 9 How to Use Functions ▪ Quiz 2: Intermediate SQL ▪ HW- <ul style="list-style-type: none"> - DataCamp: 4. Data Manipulation in SQL - Murach Chap 8 – 9 Exercise
September 26, 2024	<p>Week 6</p> <ul style="list-style-type: none"> ▪ Murach Chap 10 How to Design a Database ▪ HW- <ul style="list-style-type: none"> - Murach Chap 10 Exercise
October 3, 2024	<ul style="list-style-type: none"> ▪ Week 7 Class ▪ Murach Chap 11 How to Create Databases, tables, and Indexes ▪ Murach Chap 12 How to Create Views ▪ HW- <ul style="list-style-type: none"> - Murach Chap 11- 12 Exercise - DataCamp: 5. PostgreSQL Summary Stats and Window Functions
October 10, 2024	<p>Week 8</p> <ul style="list-style-type: none"> ▪ Murach Chap 13 Language Skills for Writing Stored Programs ▪ Murach Chap 14 How to Use Transactions and Locking ▪ HW- <ul style="list-style-type: none"> - Murach Chap 13-14 Exercise
October 17, 2024	<p>Week 9</p> <ul style="list-style-type: none"> ▪ Murach Chap 15 How to Create Stored Procedures and Functions

	<ul style="list-style-type: none"> ▪ Murach Chap 16 How to Create Triggers and Events ▪ Quiz 3: Database Design and Implementation ▪ HW- <ul style="list-style-type: none"> - DataCamp: 6. Functions for Manipulating Data in PostgreSQL - Murach Chap 15-16 Exercise
October 24, 2024	Week 10 <ul style="list-style-type: none"> ▪ Murach Chap 17 An Introduction to Database Administration ▪ Introduction to Final Project ▪ Register account in AWS and Snowflake ▪ HW- <ul style="list-style-type: none"> - Murach Chap 17 Exercise
October 31, 2024	Week 11 <ul style="list-style-type: none"> - Murach Chap 18 How to Secure a Database - Murach Chap 19 How to Back up and Restore a Database - HW- <ul style="list-style-type: none"> - Murach Chap 18 – 19 Exercise - DataCamp: 7. Database Design
November 7, 2024	Week 12 <ul style="list-style-type: none"> ▪ Murach Chap 20 How to Host a Database With AWS ▪ Introduction to Data Warehouse ▪ HW- <ul style="list-style-type: none"> - Murach Chap 20 Exercise
November 14, 2024	Week 13 Class <ul style="list-style-type: none"> ▪ Introduction to Snowflake ▪ Snowflake Lab ▪ HW- <ul style="list-style-type: none"> - Snowflake exercise
November 21, 2024	Week 14 Class <ul style="list-style-type: none"> ▪ Introduction to NoSQL ▪ Introduction to MongoDB ▪ Quiz 4 Advanced SQL, Data Warehouse ▪ HW- <ul style="list-style-type: none"> - MongoDB exercise ▪ DataCamp: 8. Introduction to NoSQL
November 28, 2024	Week 15 Class

	<ul style="list-style-type: none"> ▪ Zoom: answer question
December 5, 2024	<p>Week 16 Class</p> <ul style="list-style-type: none"> ▪ Zoom: answer question ▪ Quiz 5 Database Administration, Cloud, NoSQL
December 12, 2024	<p>Week 17 No Class</p> <ul style="list-style-type: none"> ▪ Final Project Deliverable
December 19, 2024	<p>Week 18 - Final Grades due by Noon</p>

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 use MIMIC-IV database, which require proper design, development and implementation of a database. Resources required to host your project will typically be a MySQL database and you will extract information for data analysis or building predictive model.

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. **Participation is required. Peer reviews will be collected and made part of the project grade.**

Course Format and Activities

This course is designed based on 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 your TA and 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!