

DSBA/HCIP 6160: Database Systems for Data Scientists

Course Information

<i>Course Number/Section</i>	DSBA/HCIP 6160
<i>Course Title</i>	Database Systems for Data Scientists
<i>Days & Times</i>	<i>Thursdays, 5:30 PM – 8:15 PM</i>
<i>Location</i>	<i>Dubois Center Room 1102</i>

Instructor Information

<i>Instructor</i>	Aaron Lee
<i>Email Address</i>	ALEE175@charlotte.edu
<i>Office Hours</i>	Tuesdays from 11:00 AM to 12:00 PM
<i>Zoom</i>	Link to Meeting Room , Passcode: 733264

<i>TA</i>	Dezmon Patten
<i>Email Address</i>	DPATTEN@charlotte.edu
<i>Office Hours</i>	Fridays from 10:00 AM to 12:00 PM
<i>Zoom</i>	Link to Meeting Room , Passcode: 540441

Course Prerequisites

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

Course Description:

DSBA/HCIP 6160: Database Systems for Data Scientists

This course covers 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 or NoSQL database that meets organizational specifications.
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 to meet organizational needs.
4. Demonstrate expertise in data preparation and data analytics.

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 (Murach, Other)

Learning Outcomes 1, 2, 3, 4, 5

Student Survey

Assignment 1: Install Software

Assignment 2: Murach Ch. 1-4 Exercise (Introductory SQL, Retrieving Data)

Assignment 3: Murach Ch. 5-9 Exercise (Intermediate SQL and Joins, Subqueries)

Assignment 4: Murach Ch. 5-9 Exercise (Datatypes)

Assignment 5: Relational Database Modeling with Entity Relationship Diagramming

Assignment 6: Normalization

Assignment 7: Murach Ch. 13-16 Exercise (Advanced SQL - Stored Procedures)

Assignment 8: Murach Ch. 17-20 Exercise (Data Warehouse)

- 25% Tutorials (DataCamp, Other)

Learning Outcomes 1, 3, 5

Tutorial 1: DataCamp Introduction to SQL ([SQL Fundamentals Track](#))

Tutorial 2: DataCamp Intermediate SQL

Tutorial 3: DataCamp Joining Data in SQL

Tutorial 4: DataCamp Data Manipulation in SQL (CASE, Subqueries, CTE, etc.)

Tutorial 5: DataCamp: Database Design, Data Warehouse, ETL, OLAP

Tutorial 6: DataCamp Reporting in SQL

Tutorial 7: DataCamp: Introduction to NoSQL

Tutorial 8: Cloud Insights from Data With BigQuery ([Cloud Skills Boost GCP](#))

- 30% Quizzes

Evaluation of Learning Outcomes 1-4

5 quizzes, the last quiz is given during the final exam week

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

- 25% Group Project

Learning Outcomes 1, 2, 3, 4, 5

Assigned at mid-point of semester, staged deliverables.

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 and valid reasoning (email professor and TA ahead of due date and receive express approval from the professor). If prior approval is not obtained, the grade for the late assignment will be 0. Unless otherwise noted, assignments more than 1 week late (7 days after the assignment was due including weekends) will not be accepted (receive a grade of 0) 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 cite all of their sources.

The above list of prohibited resources is not intended to be exhaustive but rather aims to provide illustrative examples. Use of any such resources without express authorization will result in a failing grade on the assignment and potential disciplinary action.

[Code of Student Academic Integrity](#)

TEXT:

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

OTHER RESOURCES: Access to DataCamp, AWS, GCP Cloud Skills Boost (formerly Qwiklabs) will be provided free of charge to students enrolled in the class. Students will create accounts using the @uncc.edu email.

We will use Dr. Daniel Soper's Video Database series as a video supplement on Relational Database Concepts. This is provided free of charge.

SOFTWARE: Students will be able to use MAC and Windows. Students must have access to the internet. One of more of the following RDMS will be used in class:

- MySQL and MySQL Workbench, a popular open-source database management system (required)
- PostgreSQL, an alternative popular open-source database used across multiple cloud-hosted and enterprise data warehouse technologies.

CLASS SCHEDULE: The course meets Thursdays from 5:30 PM to 8:15 PM and will begin on Thursday, 1/11/2024. Students will receive instructions for getting started along with Week 1 assignments on that date. Week 1 for the class officially starts on Monday, January 15th with assignments due by the end of day Sunday January 21st. Students are expected to check Canvas regularly and new modules outlining instructions and assignments for that week will be published on Mondays. Assignments will generally be due by end of day on the following Sunday.

ASSIGNMENTS AND ACADEMIC CALENDAR:

The following class schedule and deadlines are subject to change at the discretion of the instructor and class circumstances.

WEEK	TOPIC	ASSIGNMENTS
GETTING STARTED Wednesday 1/10	Introduction to Class Syllabus	Student Survey (Google Form) Student Availability (Doodle Poll) Discord Registration Purchase required textbook, download student files
WEEK 1 Monday 1/15 – Sunday 1/21	Introduction to Relational Databases MySQL, MySQL Workbench	Murach Chapter 1: An Introduction to MySQL Murach Chapter 2: MySQL Workbench Assignment 1: Install Software Tutorial 1: DataCamp Introduction to SQL (Part of the SQL Fundamentals Track)

<p>WEEK 2 Monday 1/22 – Sunday 1/28</p>	<p>Introduction to Relational Databases and SQL</p>	<p>Murach Chapter 3: How to Retrive Data from a Single Table</p> <p>Tutorial 1 contd</p>
<p>WEEK 3 Monday 1/29 – Sunday 2/4</p>	<p>Introduction to Relational Databases and SQL</p> <p>Relational Databases: Joins</p>	<p>Murach Chapter 4: How to retrieve data from two or more tables</p> <p>TUT 2: DataCamp Intermediate SQL</p>

<p>WEEK 4 Monday 2/5 – Sunday 2/11</p>	<p>Relational Databases: Joins</p> <p>Relational Databases: Updating Data</p>	<p>Murach Chapter 4 Continued</p> <p>Murach Exercise 1 Chapters 1-4</p> <p>Tutorial 3: DataCamp Joining Data in SQL</p> <p>Murach Chapter 5: Insert, Update and Delete Data</p>
<p>WEEK 5 Monday 2/12 – Sunday 2/18</p>	<p>More SQL Skills</p>	<p>Quiz 1 Introductory SQL</p> <p>Murach Chapter 6: Summary Queries</p>

<p>WEEK 6 Monday 2/19 – Sunday 2/25</p>	<p>More SQL Skills</p>	<p>Murach Chapter 7: Subqueries</p> <p>Murach Exercise 2 Chapters 6 – 9</p> <p>Tutorial 4: Data Manipulation in SQL</p>
<p>WEEK 7 Monday 2/26 – Sunday 3/3</p>	<p>More SQL Skills</p>	<p>Murach Chapter 8: Datatypes</p> <p>Murach Chapter 9 How to Use Functions</p> <p>Murach Exercise 2 continued</p> <p>Tutorial 4 contd</p>
<p>WEEK 8 Monday 3/4 – Sunday 3/10</p>	<p>SPRING BREAK</p>	

<p>WEEK 9 Monday 3/11 – Sunday 3/17</p>	<p>Database Design and Normalization</p>	<p>Quiz 2 Intermediate SQL (Murach Ch 6-9)</p> <p>Murach Chapter 10 How to Design a Database</p> <p>Assignment 4: Relational Database Modeling with Entity Relationship Diagramming</p> <p>Tutorial 5: DataCamp: Database Design, Data Warehouse, ETL, OLAP</p>
<p>WEEK 10 Monday 3/18 – Sunday 3/24</p>	<p>Database Design and Normalization</p> <p>Group Project Assignment</p>	<p>Murach Chapter 11 How to create databases, tables and indexes</p> <p>Murach Chapter 12 How to Create Views</p> <p>Assignment 5: Normalization</p>
<p>WEEK 11 Monday 3/25 – Sunday 3/31</p>	<p>Advanced Topics: MySQL</p> <p>Students will gain an introductory understanding of the concepts in this section.</p>	<p>Quiz 3: Database Design-Implement.</p> <p>Murach Ch 13 & 14: Stored Programs, Transactions</p> <p>Tutorial 6: DataCamp Reporting in SQL</p>

<p>WEEK 12 Monday 4/1 – Sunday 4/7</p>	<p>Advanced SQL</p>	<p>Murach 15 & 16: Stored Procedure, Triggers</p> <p>Assignment 6: Murach Ch. 13-16 Exercise (Advanced SQL)</p> <p>Tutorial 7: DataCamp: Introduction to NoSQL</p> <p>Quiz 4</p> <p>Project Work</p>
<p>WEEK 13 Monday 4/8 – Sunday 4/14</p>	<p>Data Warehouses, Traditional, Cloud</p>	<p>Murach Ch 17-18: Database Administration</p> <p>Project Work</p> <p>Tutorial 8: Cloud Insights from Data With BigQuery (Cloud Skills Boost GCP)</p>
<p>WEEK 14 Monday 4/15 – Sunday 4/21</p>	<p>Database Administration</p>	<p>Murach Ch 19-20: Back Up, Restore DB and Cloud Hosted DB</p> <p>Project Work</p> <p>Assignment 7: Murach Ch. 17-20 Exercise (DB Admin and Cloud)</p>

WEEK 15 Monday 4/22 – Monday 4/29	Database Provenance, Security, Privacy, Ethics	Project Work
WEEK 16 4/24 – 4/29	Final Project Deliverable	Final Exam: Quiz 5 (available during Final Exam Week)

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 relational or NoSQL 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 or MongoDB Database (NoSQL) and web or application interface. Your project can be hosted locally but web or cloud is strongly 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 and/or NoSQL 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 11th week of class.

Course Format and Activities

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 prior to the class meeting time and post any questions that you wish to be discussed on the forum.

Group discussion:

The most vital use of the course discussion forum 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's instructions; 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 and will be governed by the late submission policy if not.

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 (email) 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 (email) to the instructor within a week after the grades are announced. No requests will be considered after this timeframe.
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'. An "I" grade 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
 - i. For any questions or clarification of class material, please ask them on the course discussion forum 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.
 - ii. 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.
 - i. 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.
 - d. Please use Canvas, not emails, for general questions, unless you wish to keep your questions or comments private.
 - e. When emailing your instructor, please use a specific subject line starting with "DSBA-HCIP 6160: Homework 1 - [Last Name]." Please copy the TA(s) on assignment related questions.

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.