



DSBA HCIP 6160 Big Data Design Storage and Provenance

Course Information

Course Number/Section: **DSBA HCIP 6160**
Course Title: **Spring 2022**
Days & Times: *online*

Instructor Contact Information

Instructor: **Dr. Pamela Thompson**
Email Address: plthomps@uncc.edu
Zoom: <https://us06web.zoom.us/j/2694772139>
Office Hours: 7:30 – 8:30 pm Mon, Wed and by appointment
email or call/text 980-234-0042 (cell)

Course Prerequisites

- Graduate student standing or permission of instructor.
- Familiarity with programming language such as PHP, Java or C++

Course Description:

The modeling, programming, integration, and provenance of big data. Focuses on SQL and NoSQL, but may also address other advanced topics. Topics include: (1) modeling/theory: basics of RDBMS and NoSQL, database design; (2) programming: SQL and NoSQL query languages; (3) integration: data warehousing, preprocessing; (4) databases in the Cloud; and (5) provenance: data version control, data lifecycle management.

Student Learning Objectives

The objectives of this course are to learn how to:

- Install and configure RDBMS
- 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
- Apply 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.
- Create effective visualizations from data
- *For students in HCIP:* Apply best practices in the design of new and/or critique of existing population health data sources. Students will specifically work with a health data source.

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. 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. explain the role and importance of data provenance in all types of database systems.
5. Demonstrate data preparation, visualization and data analytics using big data
6. Demonstrate the ability to work with a health data source.

Grading and Assessment Criteria: Outcomes are assessed by:

- 15% HW Assignments (averaged, equal weight)
Learning Outcomes 2, 4, 5
 HW 1. Local Installation of MySQL, MySQL Workbench, Practice Databases
 HW 2. Database Design: Entity Relationship Diagramming
 HW 3. Database Design: Normalization
 HW 4. Data Warehousing
 HW 5: Database Programming, Connectivity, UI
 HW 6: Big Data, Data Lake, Data Provenance
- 15% Datacamp and Qwiklabs Tutorials (averaged, equal weight)
Learning Outcomes 1, 3, 5
 TUT 1: Datacamp: Introduction to SQL

- TUT 2. Datacamp: Introduction to Relational DB in SQL
- TUT 3: Datacamp: Joining Data in SQL
- TUT 4: Datacamp: Intermediate SQL
- TUT 5: Applying SQL to Real World Problems
- TUT 6. Qwiklabs: Big Query Basics for Data Analysts (NoSQL Databases, Cloud)
- TUT 7: Qwiklabs: Big Query for Data Warehousing (NoSQL, Databases, Cloud)
- 15% MURACH EXERCISES (Individual and Small Group)
 - Learning Outcome 2*
 - HW Assignments from Murach DB Exercises (averaged)
- 20% Project (Group)
 - Learning Outcome 2*
 - Project (multiple deliverables, assigned after midterm. HCIP students work with health data source.
- 35% Exams: Midterm and Final: Online

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 up to a week late will have 15 points automatically deducted unless extenuating circumstances are present as documented by student and approved by the professor. Assignments will not be accepted more than a week late.**

TEXT:

Required: Murach's MySQL 3rd or latest 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, Windows or Linux OS. 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
- Tableau: Students can download and install the student version of Tableau for free: Visit <https://www.tableau.com/academic/students>

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 end of day on Sunday.

LESSON	TOPIC	NOTES/HW
WEEK 1	<p>Review Syllabus</p> <p>Introduction to Databases</p> <p>Video Database Lesson 1: Introduction to Databases https://youtu.be/4Z9KEBexzcM</p>	<p>Murach Chapter 1: An Introduction to MySQL</p> <p>Complete Introduction: Discussion Forum</p> <p>Complete Student Survey: Google Form</p> <p>HW 1. Local Installation of MySQL, MySQL Workbench (Appendix A and B, instructions provided in Canvas)</p> <p>TUT 1: Datacamp: Introduction to SQL (due 9/5 end of day)</p>
WEEK 2	<p>Introduction to Relational Databases</p> <p>Video Database Lesson 2: The Relational Model https://youtu.be/kyGVhx5LwXw</p> <p>Optional: Video Database Lesson 3: SQL (We are covering this thoroughly in Murach and Datacamp): https://youtu.be/kqUIoOM3WEs</p>	<p>Murach Chapter 2: MySQL Workbench</p> <p>Murach Chapter 2 Exercise</p>
WEEK 3	<p>The ER Model</p> <p>Video Database Lesson 4:</p>	<p>Murach Chapter 3: How to Retrieve Data from a Single Table</p>

	<p>Data Modeling and the ER Model: https://youtu.be/IfaqkiHpljo</p>	<p>Murach Chapter 3 Exercise</p> <p>TUT 2: Introduction to Relational DB in SQL (due 9/19 end of day)</p> <p>HW 2: ER Modeling</p> <p>(see Creating ER Diagrams: https://youtu.be/lAtCySGDD48)</p>
Week 4	<p>Database Design</p> <p>Video Database Lesson 5: Database Design: https://youtu.be/B_QU9OGwtbI</p>	<p>HW 3: Normalization</p> <p>TUT 3: Joining Data in SQL (due 9/26 end of day)</p>
WEEK 5	<p>Inserting, Updating, Deleting Data</p> <p>Inserting, Updating Deleting Data</p>	<p>Murach Chapter 4: How to Retrieve Data from Two or More Tables</p> <p>Murach Chapter 4 Exercise</p> <p>Murach Chapter 5: How to Insert, Update and Delete Data</p> <p>Murach Chapter 5 Exercise</p>
WEEK 6	<p>Advanced SQL: Summary Queries and Subqueries</p>	<p>TUT 4: Intermediate SQL (due 10/17 end of day)</p> <p>Murach Chapter 6: Summary Queries</p> <p>Murach Chapter 6 Exercise</p>

WEEK 7	<p>Subqueries</p> <p>Midterm Exam Online</p>	<p>Murach Chapter 7: Subqueries</p> <p>Murach Chapter 7 Exercise</p> <p>Midterm Exam Online</p>
WEEK 8	<p>Advanced SQL: Datatypes and Functions</p>	<p>Murach 8 and 9: Datatypes and Functions</p> <p>Murach 8 and 9 Exercise</p>
WEEK 9	<p>Indexes and Views</p> <p>Video Lesson: https://youtu.be/Xk3cgUdoieU</p>	<p>Murach 10, 11, 12: Database Design (review), Indexes, Views</p> <p>Project Deliverable 1</p>
WEEK 10	<p>Advanced SQL: Stored Procedures, Functions, Triggers, Transactions</p>	<p>Murach 13, 14, 15, 16</p> <p>Murach 13, 14, 15, 16 Exercise</p>
WEEK 11	<p>Database Administration, Security, Backups and Restoration</p> <p>Video Lesson: https://youtu.be/dMkwFzRgxZY</p>	<p>Murach 17, 18, 19</p> <p>Project Deliverable 2</p> <p>TUT 5: Dataamp: Applying SQL to Real World Problems (due 11/21 end of day)</p>
WEEK 12	<p>Cloud Databases, Big Data</p>	<p>TUT 6. Qwiklabs: Big Query Basics for Data Analysts (NoSQL Databases, Cloud)</p>

		TUT 7: Qwiklabs: Big Query for Data Warehousing (NoSQL, Databases, Cloud) Project Deliverable 3
WEEK 13	Data Warehousing	HW 4: Data Warehousing HW 5: Data Visualization with Tableau
WEEK 14	Database Programming, Connectivity, UI and Data Provenance	HW 5: Database Programming, Connectivity and UI
WEEK 15	Data Lake, Big Data, Data Provenance Review of Class Concepts	HW 6: Big Data, Data Lake, Data Provenance Projects, all assignments due by last day of class (end of day, 11:59 pm)
FINAL EXAM Available online scheduled		Final Exam Online

Project

Our course project will provide you the opportunity to explore and experience database design and programming in practice. You will collaborate with *four or five* other students in this course. The project will be assigned at the mid-point of the semester. The project will focus on relational databases or data warehouses.

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 a project repository such as Github and a video (no longer than 6 minutes) or narrated PPT 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 around the midpoint of the semester.

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 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 to 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, insure 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).
- The instructor will respond to email within 48 hours of the email. If you have sent email that has not received a response for more than 72 hrs, please call the instructor (980-234-0042 cell).

(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 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."

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 Spring and sharing my knowledge.