Course Prerequisites

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

Course Description:

DSBA-HCIP 6160: Big Data Design, Storage and Provenance

This course covers the modeling, programming, integration, and provenance 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, Spark, Python, 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 like XML and JSON; (3) integration and downstream reporting: ETL, data warehousing, OLAP; (4) big data and cloud approaches; and (5) provenance: data version control, data lifecycle management.
Prerequisite: Graduate standing or permission of instructor.

Student Learning Objectives

The objectives of this course are to learn how to:

- Install and configure RDBMS tools – primarily MySQL
- 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 experience with 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 with visualization tools.

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. be able to explain the role and importance of data provenance in all types of database systems.
5. Demonstrate expertise in data preparation, visualization and data analytics using big data

HCIP Students:

6. 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:

- 25% 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 Provenance (Reading and Discussion Question)
HW 5: Data Visualization: Tableau

● 25% 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. Qwiklabs: Big Query Basics for Data Analysts (NoSQL Databases, Cloud)
TUT 6: Qwiklabs: Big Query for Data Warehousing (NoSQL, Databases, Cloud)

● 25% Murach Chapter Projects: (Individual and Group)

Learning Outcome 2
Average of HW Assignments from Murach DB Exercises from Chapters

● 25% Exam: Final: Online and available May 6th – May 8th

Evaluation of Learning Outcomes 1-5
(Objective (timed, online, terminology and concepts, open notes, includes SQL)

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 student and approved by the professor.

TEXT:

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)
  o 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

Course Syllabus
CLASS MEETING SCHEDULE:
Thursday Jan 21 ONLINE: 5:30 – 6:30 online synchronous via Zoom
Thursday Jan 28 ONLINE: 5:30 – 6:30 online synchronous via Zoom
Thursday Feb 4 ONLINE: 5:30 - 6:30 online synchronous via Zoom
Thursday Feb 11: SPRING BREAK (Feb. 8 – 12)
Thursday Feb 18 ONLINE: 5:30 - 6:30 online synchronous via Zoom
Thursday Feb 25 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104
Thursday Mar 4 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104
Thursday Mar 11 ONLINE: 5:30 - 6:30 online synchronous via Zoom (note: March 12 – midterm progress reports)
Thursday Mar 18 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104
Thursday Mar 25 ONLINE: 5:30 - 6:30 online synchronous via Zoom
Thursday Apr 1 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104
Thursday Apr 8 ONLINE: 5:30 - 6:30 online synchronous via Zoom
Thursday Apr 15 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104
Thursday Apr 22 ONLINE: 5:30 - 6:30 online synchronous via Zoom
Thursday Apr 29 FACE TO FACE 5:30 - 8:15 UNCC Center City Room 1104 (April 30th – no classes, UNCC Day of Remembrance)

May 5th is the last day of classes for UNC Charlotte. May 6th is reading day (no classes or exams)
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.

<table>
<thead>
<tr>
<th>LESSON</th>
<th>TOPIC</th>
<th>NOTES/HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEEK 1</td>
<td>Review Syllabus</td>
<td>Murach Chapter 1: An Introduction to MySQL</td>
</tr>
<tr>
<td>Thursday January 21</td>
<td>Introduction to Databases</td>
<td>Complete Introduction: Discussion Forum</td>
</tr>
<tr>
<td>ONLINE</td>
<td>Video Database Lesson 1: Introduction to Databases</td>
<td>Complete Student Survey: Google Form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUT 1: Datacamp: Introduction to SQL (due 5:30 pm 2/4)</td>
</tr>
<tr>
<td>WEEK 2</td>
<td>Introduction to Relational Databases</td>
<td>Murach Chapter 2: MySQL Workbench</td>
</tr>
<tr>
<td>Thursday January 28th</td>
<td>Video Database Lesson 2: The Relational Model</td>
<td>Murach Chapter 2 Project</td>
</tr>
<tr>
<td></td>
<td><a href="https://youtu.be/kyGVhx5LwXw">https://youtu.be/kyGVhx5LwXw</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional: Video Database Lesson 3: SQL (We are covering this thoroughly in Murach and Datacamp): <a href="https://youtu.be/kqUIoOM3WFs">https://youtu.be/kqUIoOM3WFs</a></td>
<td></td>
</tr>
<tr>
<td>WEEK 3</td>
<td>The ER Model</td>
<td>Murach Chapter 3: How to Retrieve Data from a Single Table</td>
</tr>
<tr>
<td>Thursday February 4th</td>
<td>Video Database Lesson 4:</td>
<td></td>
</tr>
</tbody>
</table>

Course Syllabus
<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topic</th>
<th>Resource 1</th>
<th>Resource 2</th>
<th>Homework 1</th>
<th>Homework 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 4</td>
<td>(Spring Break is February 8th to February 12th)</td>
<td>Data Modeling and the ER Model: <a href="https://youtu.be/IfaqkiHpljo">https://youtu.be/IfaqkiHpljo</a></td>
<td></td>
<td></td>
<td>Spring Break</td>
<td></td>
</tr>
<tr>
<td>WEEK 6</td>
<td>Thursday February 25th</td>
<td>Joining Data Inserting, Updating, Deleting Data</td>
<td></td>
<td></td>
<td>Murach Chapter 4: How to Retrieve Data from Two or More Tables</td>
<td>Murach Chapter 4 Project</td>
</tr>
<tr>
<td>WEEK 7</td>
<td>Thursday March 4th</td>
<td>Advanced SQL: Summary Queries and Subqueries</td>
<td>TUT 4: Intermediate SQL (due 5:30 pm March 18)</td>
<td></td>
<td>Murach Chapter 6: Summary Queries</td>
<td>Murach Chapter 6 Project</td>
</tr>
</tbody>
</table>

**Course Syllabus**

Page 7
<table>
<thead>
<tr>
<th>WEEK 8</th>
<th>Advanced SQL: Datatypes and Functions</th>
<th>Murach 8 and 9: Datatypes and Functions Murach 8 and 9 Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday March 11th</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 9</td>
<td>Indexes and Views</td>
<td>Murach 10, 11, 12: Database Design (review), Indexes, Views Murach 10, 11, 12 Projects</td>
</tr>
<tr>
<td>Thursday March 18th</td>
<td>Video Lesson: <a href="https://youtu.be/Xk3cgUdoieU">https://youtu.be/Xk3cgUdoieU</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 10</td>
<td>Advanced SQL: Stored Procedures, Functions, Triggers, Transactions</td>
<td>Murach 13, 14, 15, 16 Murach 13, 14, 15, 16 Projects</td>
</tr>
<tr>
<td>Thursday March 25th</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 11</td>
<td>Database Administration, Security, Backups and Restoration</td>
<td>Murach 17, 18, 19 Murach 17, 18, 19 Projects</td>
</tr>
<tr>
<td>Thursday April 1st</td>
<td>Video Lesson: <a href="https://youtu.be/dMkwFzRgxZY">https://youtu.be/dMkwFzRgxZY</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 12</td>
<td>Cloud Databases, Big Data</td>
<td>TUT 5. Qwiklabs: Big Query Basics for Data Analysts (NoSQL Databases, Cloud) HCIP Students: Apply best practices in the design of new and/or critique of existing population health data sources. (part of SLO for HCIP)</td>
</tr>
<tr>
<td>Thursday April 8th</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 13</td>
<td>Data Warehousing</td>
<td>TUT 6: Qwiklabs: Big Query for Data Warehousing (NoSQL, Databases, Cloud)</td>
</tr>
<tr>
<td>Thursday April 15th</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 14</td>
<td>Data Provenance</td>
<td>HW 4: Data Provenance</td>
</tr>
<tr>
<td>Thursday April 22nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK 15</td>
<td>Data Visualization, Exploratory Data Analysis</td>
<td>HW 5: Data Visualization with Tableau</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>LAST CLASS</td>
<td>Review of Class Concepts</td>
<td></td>
</tr>
<tr>
<td>FINAL EXAM</td>
<td>Final Exam Online</td>
<td></td>
</tr>
<tr>
<td>Available Online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday May 6th to Saturday May 8th end of day</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**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 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 a project repository such as Github and a video (no longer than 6 minutes) demonstrated 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.
<table>
<thead>
<tr>
<th></th>
<th>Date Range</th>
<th>Topics</th>
<th>Data Camps/Assignments</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5/25 to 5/31</td>
<td>Relational Modeling and Design, Chapter 3: ER Modeling, Chapter 4: Relational Model</td>
<td>Data Camp 2: Intro to Relational Databases in SQL, Data Camp 3: Joining Data in SQL</td>
<td>HW: ER Modeling Exercise, HW: Creating and Accessing a DB in MySQL, MySQLWorkbench</td>
</tr>
<tr>
<td>4</td>
<td>6/8 to 6/14</td>
<td>Chapter 11: Data Warehousing and Data Mining, Chapter 12: Big Data and NoSQL</td>
<td>Data Camp 6: Applying SQL to Real World Problems, Qwiklab: Big Query with GCP</td>
<td>HW: Datawarehousing/OLAP Lab, HW: NoSQL Databases with Mongo DB</td>
</tr>
<tr>
<td>5</td>
<td>6/16 to 6/22</td>
<td>Social and Ethical Issues, Extra Topic: Data Governance, Data Provenance</td>
<td>Datacamp 7: Introduction to Data Science in Python</td>
<td>HW: Working with Data in Python, HW: Visualizing Data with Tableau</td>
</tr>
<tr>
<td>6</td>
<td>6/23 to 6/24</td>
<td>Review, Final Exam</td>
<td></td>
<td>FINAL EXAM ONLINE</td>
</tr>
</tbody>
</table>

**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.