Course Information

Course Number: DSBA/HCIP-6160 (Fall 2019)
Course Title: Big Data Design, Storage and Provenance
Days & Times:
   Class: Thursdays, August 22 to November 21, 5:30-8:15 pm
   Final Exam: Thursday, December 12, 5:00-7:30 pm
Location: Center City Room 906
Website: https://uncc.instructure.com/courses/110402

Instructor / TA Contact Information

Instructor: Matthew Campbell (mcampb55@uncc.edu / 980.230.2191)
Teaching Assistant: Kaci Allen (kallen70@uncc.edu)
Office Location / Hours: by appointment only

Course Description

The modeling, programming, and implementation of database systems. Focuses on relational database systems, but may also address non-relational databases or other advanced topics. Topics include: (1) modeling: conceptual data modeling, ER diagram, relational data model, schema design and refinement; (2) programming: relational algebra and calculus, SQL, constraints, triggers, views; (3) implementation: data storage, indexing, query execution, query optimization, and transaction management; and (4) advanced: semi-structured data model, XML, and other emerging topics.

Student Learning Objectives

Some of the key learning objectives of this course are:

- Define and implement conceptual data models
- Understand and develop proficiency with languages such as relational algebra and SQL.
- Learn storage layer concepts such as indexing, transactions, concurrency and others
- Develop experience with semi-structured and unstructured data approaches including NoSQL and cloud paradigms
- Explore data lifecycle concepts which highlight how data evolves through various systems. Some examples we might cover include OLTP, ETL, data warehouse and OLAP
- Gain experience with data visualization tools.

Course Prerequisites

- Graduate standing or permission of instructor
- Familiarity with programming language such as Python, Java or C++
Assessment Criteria:

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exams (50%)</td>
<td>Two exams - mid-term (20%) and final exam (30%): total weight of 50% toward final overall grade. Format of exams to be communicated at a later date.</td>
</tr>
<tr>
<td>Labs/Tutorials (20%)</td>
<td>Based upon overall points gained in labs/tutorials. For non-scored labs/tutorials, you will receive points proportional to the percent completed (100% complete = full credit).</td>
</tr>
<tr>
<td>Problem Sets (20%)</td>
<td>6 problem set assignments</td>
</tr>
<tr>
<td>Quizzes / Participation (10%)</td>
<td>Weekly quizzes - we will drop the lowest two scores + Piazza introduction</td>
</tr>
</tbody>
</table>

Course Website

Lecture materials and assignments will all be posted in Canvas.

Textbook and other Resources

Course materials will be primarily based upon lecture slides and content provided by the following sites:

- DataCamp: online tutorials (provided free by instructor)
- Stanford Online – Lagunita: Introduction to Databases – Self Paced Series
- QWIKLABS: online cloud tutorials (provided free by instructor)


Class Preparation

This class is designed in a “flipped” manner so that we can spend our limited, valuable class time answering questions, working out problems, and advancing what you already should have learned on your own while preparing for class. That means that you are expected to learn most of the basic materials on your own before we meet for class (except the first class). It is absolutely critical that you keep up with assigned materials in order to both succeed in the classroom and learn as much as possible.

“Practice is the best of all instructors.” -- Publilius Syrus (42 BCE)
Assignments & Academic Calendar

We will attempt to cover the following topics during the course with the rough sequencing below.

- Intro to Relational Databases and SQL
- Joining Data in SQL, Relational Algebra
- ER Diagrams, Relational Model, Relational Design/Normalization
- Intermediate SQL, Data Modification Statements (CRUD)
- Data Definition Language (DDL), Indexes/Transactions, Constraints/Triggers
- Query Plan Optimization, Views and Authorization
- Data Lifecycle, ETL, Data Warehousing, OLAP
- Data Visualization
- Semi-structured data (XML/JSON), Flat Files / Excel Files
- Big Data/NoSQL – Cloud & Hadoop
- Governance/Ethics

Below is a very tentative schedule. NOTE: deadlines and topic sequencing are subject to change at the discretion of the instructor.

<table>
<thead>
<tr>
<th>Date</th>
<th>Topics</th>
<th>Homework / Exams</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/22 Relational Databases and Intro to SQL</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8/29 Joining Data in SQL, Relational Algebra</td>
<td>Problem Set 1 assigned (due 9/12)</td>
</tr>
<tr>
<td>3</td>
<td>9/5 ER Diagrams, Relational Model, Relational Design/Normalization,</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9/12 Intermediate SQL, Data Modification Statements (CRUD)</td>
<td>Problem Set 2 assigned (due 9/26)</td>
</tr>
<tr>
<td>5</td>
<td>9/19 Data Definition Language (DDL), Indexes/Transactions, Constraints/Triggers</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9/26 Query Plan Optimization, Views and Authorization</td>
<td>Midterm Exam</td>
</tr>
<tr>
<td>7</td>
<td>10/3 Data Lifecycle, ETL, Data Warehousing, OLAP</td>
<td>Problem Set 3 assigned (due 10/17)</td>
</tr>
<tr>
<td>8</td>
<td>10/10 Data Visualization</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10/17 Flat Files / Excel Files</td>
<td>Problem Set 4 assigned (due 10/31)</td>
</tr>
<tr>
<td>10</td>
<td>10/24 Semi-structured data (XML/JSON)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10/31 Big Data/NoSQL – Cloud</td>
<td>Problem Set 5 assigned (due 11/14)</td>
</tr>
<tr>
<td>12</td>
<td>11/7 Big Data/NoSQL – Cloud</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11/14 Big Data/NoSQL – Hadoop</td>
<td>Problem Set 6 assigned (due 11/28)</td>
</tr>
<tr>
<td>14</td>
<td>11/21 Big Data/NoSQL – Hadoop, Governance/Ethics</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12/12</td>
<td>Final Exam</td>
</tr>
</tbody>
</table>
Computing Requirements

You will need a working computer to use during class to complete labs and quizzes. We will use Windows 10 virtual machines to perform all classwork. We will not have time in class to troubleshoot issues with base machines (i.e., the base setup you use to log into your virtual machine). Please ensure that your computer is working properly so you’re able to participate.

The following free software programs will be necessary during the course of the semester. The instructor will make announcements of any additional software which might be needed throughout the course.

- **MySQL**: popular open-source database management system
- **SQLite Browser**: lightweight database engine which we’ll use for various exercises/labs
- **Tableau**: student version of data visualization tool available for download
- **Anaconda - Jupyter / Python environment**

Group discussion:
The most vital use of the Piazza 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 respond to the topics that interest you.

Submission of Work:
Follow each assignment instruction; all work should be uploaded into the assignment link, or the Discussion board on Canvas or on Piazza as instructed. 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. See Grading Policies below for details on late submission penalties.

Devices/Computers in the Classroom
The use of cell phones, smart phones, or other mobile communication devices is disruptive, and is therefore generally prohibited in class. Except in emergencies, those using such devices must leave the classroom for the remainder of the class period.

Students are permitted to use computers during class for note-taking and other class-related work only. Those using computers during class for work not related to that class must leave the classroom for the remainder of the class period.

Disability Services
Students in this course seeking accommodations to disabilities must first consult with the Office of Disability Services and follow the instructions of that office for obtaining accommodations.

Grading Policies

**Homework Late Penalties:** Homework is due prior to the start of lecture. If it is submitted more than 10 minutes after the start of lecture, it will be considered a full day late. There will be a 10% deduction for homework that is 1 day late and a 20% deduction for homework that is 2-3 days late. We will not accept
homework that is more than 3 days late. Plan your time carefully and give yourself plenty of time to ask questions and obtain help.

Your total points earned for the entire course will be converted to a percentage and a grade will be awarded as per the University grading system:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90 - 100</td>
</tr>
<tr>
<td>B</td>
<td>80 - 89</td>
</tr>
<tr>
<td>C</td>
<td>70 - 79</td>
</tr>
<tr>
<td>D</td>
<td>60 - 69</td>
</tr>
<tr>
<td>F</td>
<td>Below 60</td>
</tr>
</tbody>
</table>

As of the time of publication of this syllabus, there are no plans to offer makeup exams.

**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, always keep a copy of your work.
5. **Communication Protocol**:
   a. **Canvas - Announcements**
      • Announcements will be posted in Canvas. Make sure to check the assignment area frequently enough to stay informed.
   b. **Piazza – Questions and Clarifications**
      • Question and requests for clarifications should be directed to the Piazza discussion forums. Everyone in the class is encouraged to help answer these questions. If satisfactory answers do not emerge, the TA or instructor will answer.
      • Posting of solutions to assignments are not appropriate and considered a violation of honor code.
      • If there is a post which has not received a response for more than 72 hrs, please email the TA or call the instructor.
   c. **Emails - Comments and Requests**
      • 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 to 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 emails for private comments, questions or requests which you can’t ask in the Piazza forums.
      • When emailing your instructor, please use a specific subject line starting with "DSBA-HCIP 6160", e.g., "DSBA-HCIP 6160: Homework 1".
      • The instructor will reply to legitimate email inquiries from students within 48 hours with the exception of weekends or university holidays. If you do not receive a reply
within this period, please resubmit your question(s) or call your instructor. Leave a message if necessary.

Policy on Academic Integrity:
All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online.

Faculty may ask students to produce identification at examinations and may require students to demonstrate that graded assignments completed outside of class are their own work.

Code of Student Responsibility
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
The standards and requirements set forth in this syllabus may be modified at any time by the course instructor. Notice of such changes will be by announcement in class or by written or email notice or by changes to this syllabus posted on Canvas.

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