A close up of a logo

Description automatically generated

**DSE 6210: Big Data: SQL & NoSQL**

**Instructor: Jeremiah Lowhorn**

**Course Term: Fall 2023**

**On-line live meeting: Wednesdays 7:00pm EST**

**Office Hours: Fridays: 12-3 EST**

**Email: lowhornj@merrimack.edu**

**Course Description**: Students will be immersed in a hands-on 8-week course covering PostGres (SQL) and MongoDB (NoSQL). Students will develop an understanding of the design and function of relational and non-relational databases. The SQL portion of the course will cover the PostGres dialect, relational design, database performance & scaling, and common relational database operations. The NoSQL portion of the course will cover the MongoDB Query Language (MQL), non-relational design, database performance & scaling, and common NoSQL database operations. Students will complete two data projects that go from the database design phase to the final ETL. This course is an elective for the MS in Data Science curriculum.

Prerequisites: DSE 5001, DSE 5002

**Required Textbooks**:

1. *Fundamentals of Database Systems 7th edition, Ramez Elmasri and Shamkant Navathe, Pearson, 2017. ISBN: 978-1292097619*

**Software:**

1. PostGreSQL
2. MongoDB Atlas Free Tier
3. LucidChart (or any comparable software that provides ERD and flowchart diagrams)

**Course Objectives/Learning Goals**:

This course fulfills the following program learning goals:

LG2: Collecting and Processing Data. Students use R, Python, and other coding languages to collect, explore, clean, wrangle, and summarize large data sets.

LG3: Presenting and Integrating Results into Action. Use industry-leading software to “tell the story of the data” by creating graphical summaries with Tableau and interactive dashboards with R Shiny and other coding languages.

This course has the following course learning goals:

1. Design relational and non-relational schemas for efficient data processing.
2. Understand the differences and applications for SQL and NoSQL database technologies.
3. Understand basic ETL methods and how/when to apply them (INSERT, UPDATE, DELETE).
4. Write complex queries between multiple tables including but not limited to: sub-queries, joins, unions, aggregation, and windowing.
5. Demonstrate knowledge of both SQL and NoSQL technologies and how/when to use them.

**Course Expectations:** I expect that you will spend a substantial amount of your time outside of class fully mastering the material. In addition to participating in the one-hour live weekly class session, you should plan on spending an average of twenty-two (23) hours a week on work outside of class. This time will be spent re-watching the recording of the class session one or more times to unpack the dense presentation of material, reading the assigned readings, completing exercises, and completing two larger projects (about 184 hours altogether).

|  |  |
| --- | --- |
| **Activity** | **Weekly Time Commitment** |
| Lectures – attending live session and reviewing recording | 3 hours per week |
| Course materials/reading | 5 hours per week |
| Design & Coding Labs | 9 hours per week |
| Projects (2 data projects) | 5 hours per week |
| Discussion – 8 discussion posts | 1 hours per week |

# Grading & Assignments

|  |  |  |  |
| --- | --- | --- | --- |
| **A** | **B** | **C** | **F** |
| **Excellent** | **Good** | **Fair/Poor** | **Unacceptable** |
| 93+ A | 87-89 B+ | 77-79 C+ | Below 70 |
| 90-92 A- | 83-86 B | 73-76 C |  |
|  | 80-82 B- | 70-72 C- |  |

Although the graduate grading policy is similar to the undergraduate policy, it should be noted that the expectations for graduate students are much higher and therefore the grading is more rigorous. Candidates for any graduate-level degree or certificate must attain a final cumulative grade point average of 3.0 before the degree or certificate will be conferred.

The following letter and special grades are used across all graduate programs. Please note that this document sets the minimum standards for the College. Individual degree programs may have stricter GPA and course grade requirements.

A: “A” indicates outstanding work

B: “B” means that the work is satisfactory

C: “C” (2.0) is deemed unsatisfactory at the graduate level. No more than two courses at the C level (2.0 or higher) will be counted as acceptable toward a graduate degree. Students may be permitted to repeat only two courses and may repeat each course only one time. Those who receive more than two C level grades will be automatically dismissed from their program of study.

Any grade lower than a C - will not be acceptable for graduate-level work and cannot be counted as credit towards the degree. However, the grade will be counted toward the graduate student’s GPA. If a student receives a grade lower than a C in a required course, the student will be required to retake the course. A course may only be retaken one time. The most recent grade will replace the first grade and will be factored into the student’s GPA.

Please see “Academic Requirements and Policies” in the Graduate Catalog at <http://catalog.merrimack.edu/content.php?catoid=11&navoid=265> for more information.

Student requirements and their corresponding weights are summarized as follows:

|  |  |
| --- | --- |
| **Assignment** | **Weight** |
| Data Modeling Project | 20% |
| Data Model Implementation Project | 20% |
| Discussions | 20% |
| Labs | 40% |

The course is laid out in two sections, Relational Databases and NoSQL Databases. The first four weeks will focus on SQL and the materials in the labs and discussions prepare the student to complete a data project which will be due at the beginning of the 5th week of class. The second four weeks of class will focus on NoSQL and prepare students to complete a similar, yet distinct data project which will be due at the end of the course. There will be a total of 6 labs and 8 discussions split between SQL and NoSQL.

# Data Modeling Project

The purpose of this project is to showcase your understanding of the design of relational database systems. You will be provided an SRS (software requirements specification) and be asked to build an ERD based upon the requirements of the SRS. You may chose to model your data with the relational model or the non-relational model. However, the non-relational model will not be covered before this portion of the project is due, so you will need to work ahead if you choose that route.

# Data Model Implementation Project

The goal of this assignment is to build the database for the data model you designed in Project 1. You will need to write the DDL statements, build the ETL pipelines, and insert some sample data. Again, if you choose to use MongoDB for the first portion of the assignment your implementation will also need to be in MongoDB.

# Labs

There will be 8 Labs throughout the course, each providing hands-on experience implementing and reinforcing the techniques and methods discussed during the Live Session and Readings. Students are expected to prepare and submit responses to the questions found at the end of each Lab to receive credit.

# Important Deadlines

|  |  |
| --- | --- |
| 8/28/2023 | First day of classes |
| 9/5/2023 | Last day to add/drop a course |
| 10/6/2023 | Last day to withdraw from a course with a “W” |
| 10/20/2023 | Last day of classes |

# Project Assignment Content and Format

As noted earlier, the core part of this course’s learning assessment takes the form of hands-on, applied database project. Students will be required to submit their SQL/MQL code in a text file each week along with their writeups in PDF format.

# Grading-Related Policies for this Course

**Assignments** must be **submitted on Blackboard by 11:59 pm of the due date listed in the Weekly Topics & Assignments** section. Please note that there will be **NO late work accepted**. Final grades will be based on a percentage of possible points earned, but the instructor reserves the right to adjust the course grade.

# Class Policies

**1. Academic Integrity**

See Merrimack College Student Handbook for college’s Academic Integrity Policy. The academic integrity code and policy is also posted on the Office of the Provost’s web site: [http://www.merrimack.edu/about/offices\_services/office-of-the-provost/academic-integrity-](http://www.merrimack.edu/about/offices_services/office-of-the-provost/academic-integrity-code.php) [code.php.](http://www.merrimack.edu/about/offices_services/office-of-the-provost/academic-integrity-code.php) Here is a brief excerpt from that: “…Academic integrity is fundamental to creating and maintaining an atmosphere of cooperation and trust. It is thus a concern for everyone in the college community. The academic integrity code below is designed to help students understand what is not permissible in their academic and intellectual lives at the college. It seeks to protect students from unintentional acts of dishonesty and to preserve the trust inherent in the student- teacher relationship, which is compromised if suspicion arises regarding the integrity of a student’s work. The code is also designed to inform students of the rules which will be used to judge academic integrity infractions…”

# Academic Accommodations from the Accessibility Services Office

Merrimack College provides reasonable accommodations for students with documented disabilities. Students who have, or think they may have, a disability are invited to contact the Accessibility Services Office via the online request form found on the Accessibility Services website: [www.merrimack.edu/aso](http://www.merrimack.edu/aso), email: [accessibilityservices@merrimack.edu](mailto:accessibilityservices@merrimack.edu), or by visiting us on the third floor of McQuade Library (in the event we are open).

Students are encouraged to contact the office as soon as possible to ensure adequate time to meet and create a plan. Students already registered with Accessibility Services are encouraged semesterly to request for their letters to be emailed and students are responsible to then email the PDF to their instructors personally. Accommodations cannot be made retroactively.

# Live On-line Meeting Attendance

Participation in weekly live on-line meetings is highly encouraged, but it is not mandatory. All sessions will be recorded, and recordings will be posted; however, when not attending students forego the ability to directly ask questions about topics being discussed, or other parts of the course.

# 4. Requests for Extensions

The general policy is that, outside of properly verified serious medical emergencies\* (as defined below), extensions are not given, which applies to all labs, projects, and discussions. Missing an assignment without an acceptable reason (to be clear, that means a serious medical emergency, as defined below) will result in 0 points for the exam or a project. The intent here is not to penalize anyone – quite to the contrary, it is to create a level playing field so that no one has a unique and an unfair advantage. All assignment due dates are published (see the Live Meetings, Topics & Assignments section below) and will not change, barring a natural or other emergency – please consider those dates when planning any non-class related activities.

\**Serious medical emergency is defined as an injury or illness that is acute and poses an immediate risk to a person’s life or long term health. To be “properly verified”, the said serious medical condition must be attested to by hospitalization and related medical treatment documentation*.

**Mental Health:** It is important to make your mental health a priority! We will do activities in class to support your mental health and wellness, and I also encourage you to engage in your own self-care habits outside of class. If you want more information or resources, please come see me! If you are struggling with your mental health, or you believe a classmate is, please talk to me so I can put you in touch with qualified and caring support to get you back on track and feeling better.

**Suicide Prevention Lifeline:** We can all help prevent suicide. The Lifeline provides 24/7, free and confidential support for people in distress, prevention and crisis resources for you or your loved ones, and best practices for professionals.

<http://www.ulifeline.org/stay_well>

National Suicide Prevention Lifeline 1-800-273-8255(TALK) or Text START to 741-741.

*Student Success Resources -* It takes a village to progress in your knowledge. The below resources are a sampling of what Merrimack College provides to assist you in meeting your goals. If you have questions, please see me for additional information.

**Counseling Services** are available to Graduate students at the Counseling Center at Merrimack via Uwill for telecounseling services only. Telecounseling includes video, chat, phone, message. Uwill is HIPPA and FERPA compliant which means they follow the strictest privacy guidelines. To access Uwill, register with your Merrimack email address at [app.uwill.com](http://app.uwill.com/). To receive support after hours call 978-837-5444 to be connected to the Uwill 24/7/365 crisis line which offers phone assistance only. The Counseling Center is located on the third floor of the Sakowich Center.

Student Success Resources

It takes a village to progress in your knowledge. The below resources are a sampling of what Merrimack College provides to assist you in meeting your goals. If you have questions, please see us for additional information.

**Success Coach:** As a student of the Master in Computer Science you have access to the success coach Michael Bowler that may help you out with specifics of the program, as well as course issues. To contact the success coach write an email to [bowlerm@merrimack.edu](mailto:bowlerm@merrimack.edu) with your issue.

**Academic Success Center:** The Academic Success Center offers workshops, programs, tutoring, and individual or group meetings to all students. Topics include: academic expectations, dealing with test anxiety, effective reading, preparing for tests, study skills, and time management. Visit

<https://www.merrimack.edu/academics/academic-success-center/>

**The Writing Center:** Because everyone is a writer, the Merrimack College Writing Center provides free consultations for any student, in any major, in any class. Individual or small group consultations are generally done face-to-face in the Writing Center, McQuade Library 3rd floor. Limited online consultations are also available. Visit [www.merrimack.edu/writingcenter](http://www.merrimack.edu/writingcenter) or visit MyMack for a listing of hours and for registration and appointment instructions. Please note that while drop-ins are welcome, appointments take priority. For specific information about availability, schedule, and location of all tutoring available on campus, please go to <http://www.merrimack.edu/tutoring>.

**The Tutoring & Math Center:** The Tutoring & Math Center operates both face-to-face and on Blackboard. Students may meet with tutors in a socially-distant, college, state, and CDC approved one-to-one face-to-face session by appointment. For virtual, drop-in assistance, students may log into Blackboard where a peer tutor will assist them in real-time via a Collaborate session – no appointment needed. All students in STEM courses have been given access to the Tutoring \& Math Center Blackboard page, titled "Tutoring \& Math Center Online Tutoring". Visit <https://www.merrimack.edu/academics/academic-success-center/tutoring-and-math-center> or

email MathCenter@Merrimack.edu for more information.

**McQuade Library:** For help citing sources consult the Writing Center or try a citation management tool such as Mendeley or Zotero. McQuade Library LibGuides can help:

<http://libguides.merrimack.edu/mendeley>

<http://libguides.merrimack.edu/zotero>

**Tech Support:** If you are having trouble logging in to your accounts, or with your iPad, please contact the help desk by emailing askit@merrimack.edu. Here are some basics to get you started: <https://www.merrimack.edu/about/offices_services/information-technology-services/students/>

# Weekly Topics & Assignments

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Topic | Readings | Assignments & Discussion |
| Week 1:Intro to DBMS | * When to use DBMS * Advantages of the DBMS Approach * Data models, schemas, and instances * Database languages and interfaces * Classification of DBMS | * Fundamentals of DS Chapter 1, 2 | * Assignment 1   PostGres Setup  Textbook exercises  1.1, 1.9, 1.10, 2.12   * Discussion 1   Textbook Ch1 Review Question 1.3 |
| Week 2: RDBMS Design | * Using high level conceptual data models for DB design * ERD diagrams * Relational model constraints * Update operations, transactions, and dealing with constraint violations | * Fundamentals of DS Chapter 3, 5a | * Assignment 2 (Create ERDs and work with constraint violations)   Textbook lab questions 3.33, 3.34, 5.17, 5.19, 5.20   * Discussion 2   Textbook Ch 5 Review Question 5.8 |
| Week 3: Basic SQL | * Definition and data types * Specifying constraints * Basic retrieval queries * INSERT, DELTE, and UPDATE | * Fundamentals of DS Chapter 6 * Fundamentals of DS Chapter 17 (optional reading) | * Assignment 3 (ETL operations)   Textbook lab questions 6.7, 6.8, 6.10, 6.12, 6.13  Discussion 4  Textbook Ch 6 Review Question 6.3 |
| Week 4: Advanced Queries | * Sub queries (with) * Joins & unions * Triggers * Views * Case statements | * Fundamentals of DS Chapter 7 * Instructor provided reading | * Assignment 4   DVD Rental DB Lab   * Discussion 5   Textbook Ch 7 Review Question 7.2 |
|  | Project 1 – SQL Project | |  |
| Week 5:Intro to  NoSQL/MongoDB | * Overview of non-relational databases * Non relational database types * When to use non-relational databases * MongoDB Architecture * Setting up MongoDB cloud | * MongoDB Handouts: 1, 2, 3, 4, 5, 6, 7 | * Assignment 1 – Setting up MongoDB Atlas Free Tier, Inserting sample collections, performing basic MQL * Discussion 5 – Compare SQL/NoSQL (When and why) |
| Week 6: | * Document models * Collections in the Document model * Data Modeling & Relationships * Inserting and Updating Data in MongoDB * Indexing | * MongoDB Handouts: 13, 14, 17, 18, 19 | * Assignment 3 – Insert & Update, Creating and demonstrating the importance of indexes in NoSQL * Discussion 6 – The importance of understanding how your collection is being used in NoSQL |
| Week 7: | * Querying in Non-Relational Databases * MongoDB Query Language (MQL) * Querying Complex Data in MongoDB with MQL * Querying Data with Operators and Compound Conditions | * MongoDB Handouts: 9, 10, 11, 12 | * Assignment 7 – Querying complex arrays * Discussion 7 – Challenges of moving a relational model to a non-relational. Benefits? |
| Week 8: | * The MongoDB Aggregation Framework * Querying Data in MongoDB with the Aggregation Framework | * MongoDB Handouts: 15, 16, 20 | * Assignment 8 – MongoDB aggregation framework * Discussion 8 – Will NoSQL spell the end of the relational model? |
|  | Project 2 – NoSQL Project | |  |