## Columbian College of Arts and Sciences

## Data Science


# Spring 2026

---

## Course and contact information:

**Course:** Data Science, DATS6102 - Data Warehousing for Data Science
**Semester:** Spring 2026

**Section:** 10

**Meeting time:** Tuesdays 06:10 pm - 08:40 pm

**Location:** Building: Phillips Hall (PHIL). Room: PHIL 110

## Instructor:

**Name:** Hazim Shatnawi
**Campus address:** 2036 H St NW, Room 308 Washington, DC 20052

**GW E-mail:** hazim.shatnawi@gwu.edu

**Office hours: Tuesdays** 2:30 pm – 5:50 pm

## Course prerequisites

No prerequisites for this course.

## IT Help Desk

***If you need help (installing software, issues with your computer, etc.)***, the Data Science Helpdesk resource for such issues. The Helpdesk team can provide technical support to both undergraduate and graduate students. They can assist with software installations, troubleshooting, and other technical needs.

## Student Resources:

The program developed a website for student resources where you can find the right resources you need. Student Resources

## Course Description:

This course covers practical skills and theory in database management. By the end of the course, students will be proficient in writing basic queries in the structured query language (SQL), creating an RDBMS model, and learning how to create a Star Schema model from the RDBMS one. Star Schema is a simplified model used when working with big data. The emergence of big data storage needs has driven the adoption and development of a new class of non-relational databases commonly referred to as NoSQL databases. This course will also explore the origins and the characteristics of NoSQL that distinguish them from the traditional relational database management systems (RDMS). We will take a closer look at one database from each of the following NoSQL data models (document - MongoDB, and graph - Neo4j).

In addition, this course will also focus on how to connect and manipulate each database via general-purpose programming languages and their drivers for connections. Python is the main language that we will use.

## Course Instructions

- You can contact me anytime via email, online meeting, or in-office meeting (Should be scheduled ahead of time).
- All announcements will be broadcast via email through Blackboard (and for updating the syllabus as well).
    - **It is your responsibility to check announcements and emails when making changes to the syllabus.**
- Please don't come late to class
- If you want to go outside during the class, please ask first
- For makeup, you have to provide a proof document

## Learning Objectives

- **Introduction to relational database management system with SQL**
    - **Understand Database Types and Selection Criteria:**
        - ✓ Differentiate between SQL and NoSQL databases, highlighting characteristics of MySQL (SQL), MongoDB (document-oriented), and Neo4j (graph).
        - ✓ Learn the criteria for selecting different database systems based on organizational needs and data characteristics.
    - **Master SQL Database Operations with MySQL:**
        - ✓ Comprehend the structure and functionality of MySQL.
        - ✓ Execute SQL queries to retrieve, manipulate, and analyze data from single and multiple tables.

- ✓ Understand the use of subqueries and summary queries in data analysis.
- ✓ Develop skills in creating, updating, and deleting data within MySQL tables.
- ✓ Gain insights into basic database design principles.

- **Connecting to MySQL with Programming Languages:**
  - ✓ Learn to connect to MySQL databases using Python and PHP for data manipulation and retrieval.
- **Introduction to advanced database structures:**
  - ✓ Star Schema model – to be built based on the RDBMS model that we will go through in the MySQL part.

❖ **Introduction to document-Oriented Database:**

- **Understand MongoDB's JSON-like Document Structure:**
  - ✓ Grasp the fundamentals of MongoDB's document model, which uses a JSON-like format for data storage and retrieval.
  - ✓ Recognize the advantages of this flexible data model for handling diverse data types and structures.
- **Installation and Initial Setup:**
  - ✓ Install MongoDB and MongoDB Compass, the intuitive GUI for MongoDB.
  - ✓ Learn basic configuration and setup procedures for MongoDB across various operating systems.
- **Performing CRUD Operations in MongoDB:**
  - ✓ Execute Create, Read, Update, and Delete (CRUD) operations.
  - ✓ Understand MongoDB's approach to data manipulation and management.
- **Querying Data with MongoDB Query Language (MQL):**
  - ✓ Utilize MongoDB Query Language for data retrieval, including complex queries and data aggregation.
  - ✓ Acknowledge the flexibility of querying in a document-oriented database environment.
- **Advanced-Data Manipulation Techniques:**
  - ✓ Implement advanced data manipulation techniques, focusing on data aggregation and indexing strategies.
  - ✓ Explore practical scenarios and applications where these techniques are vital.
- **Integration with Programming Languages:**
  - ✓ Connect Python to MongoDB for executing database operations and data manipulations.
  - ✓ Discuss the integration of MongoDB with other programming languages, demonstrating its versatility in application development.
- **Real-world applications and Case Studies:**
  - ✓ Analyze real-world scenarios to understand MongoDB's application in various industries.
  - ✓ Engage with case studies highlighting MongoDB's use in managing large-scale, unstructured data sets.

❖ **Introduction to Graph Databases:**

- **Gain a crash course in graph theory and algorithms relevant to Neo4j.**
- **Learn the basics of graph traversal algorithms and their applications in data science.**
- **Get an introduction to other graph-based algorithms for network analysis and data mining.**

**Understanding Neo4j's Graph Database Model:**

- ✓ Grasp the structure and core concepts of Neo4j as a graph database, including nodes, relationships, and properties.
- ✓ Recognize the strengths and use cases of graph databases in modeling complex, **interconnected data.**

- **Installation and Configuration:**
  - ✓ Install Neo4j Desktop and set up the Java Virtual Machine (JVM), which is essential for running Neo4j.
  - ✓ Learn about the initial configuration and basic setup of a Neo4j environment.
- **CRUD Operations and Data Manipulation in Neo4j:**
  - ✓ Perform Create, Read, Update, and Delete (CRUD) operations within the Neo4j platform.
  - ✓ Understand the fundamentals of data manipulation in a graph database context.
- **Querying with Cypher Query Language:**
  - ✓ Master the Cypher query language for querying and manipulating data in Neo4j.
  - ✓ Explore complex query techniques, including pattern matching and graph traversals.
- **Python Integration for Graph Operations:**
  - ✓ Connect Python to Neo4j to facilitate graph creation, data manipulation, and dataset loading.
  - ✓ Discuss practical applications of integrating Neo4j with Python for data analysis and processing.
- **Exploring the Neo4j Graph Data Science Library:**
  - ✓ Delve into the Neo4j Graph Data Science Library, understanding its capabilities in graph analytics and data science.
  - ✓ Explore the library's features for advanced graph analytics, including applying graph algorithms and data extraction techniques.

❖ **Evaluating Database Types for Specific Needs:**

- Understand the key differences between various database types, including SQL databases (like MySQL), document-oriented databases (like MongoDB), and graph databases (like Neo4j).
- Learn to assess the suitability of each database type for handling different kinds of data problems and datasets, considering factors like data structure, scalability, complexity, and specific use cases.
- Advanced Database Applications:
- Explore advanced applications and functionalities of different database systems, focusing on how they address complex data challenges in real-world scenarios.
- Discuss case studies and examples where the choice of the database significantly impacted the outcome of a project or a solution.

## Required textbooks and/or other materials and recommended readings:

| Author | Title | Edition |
|---|---|---|
| Murach, J. **(Required)** | Murach's MySQL | (3rd Edition) |
| Neo4j Inc. (Optional) | Neo4j Documentation | https://neo4j.com/docs/ |
| MongoDB Inc. (Optional) | MongoDB Documentation | https://www.mongodb.com/ |
| Python-w3schools (Optional) | Python tutorial – basics only | https://www.w3schools.com/python/ |

## Week-by-week schedule of topics to be presented and scheduling of final examinations:

| DATE | SUBJECT | ASSIGNMENTS/ACTIVITIES/QUIZZES | LABS |
|---|---|---|---|
| 01/13 | | | |
| 01/20 | SQL - Chapter 1 + Chapter 3 (Part 1) | XAMPP (SQL) Installation (in-class) | Lab1 - Due 01/21 11:59 pm |
| 01/27 | SQL - Chapter 3 (Part 2) + Chapter 4 (Part 1) | | Lab2 - Due 01/28 11:59 pm |
| 02/03 | SQL - Chapter 4 (Part 2) + Chapter 5 | Assignment 1 - Announced | Lab3 - Due 02/04 11:59 pm |
| 02/10 | SQL - Chapter 6 + Chapter 7 (Part 1) | Quiz 1 - at the beginning of the lecture | Lab4 - Due 02/11 11:59 pm |
| 02/17 | SQL - Chapter 7 (Part 2) + SQL with Python | Assignment 1 - Due 11:59 PM | Lab5 - Due 02/18 11:59 pm |
| 02/24 | Midterm (SQL Part) + Intro to Document-Oriented DBs | Midterm (first half)<br>MongoDB Installation & Intro to Document-Oriented DBs (second half) | Lab6 - Due 02/25 11:59 pm |
| 03/03 | MongoDB - Part 1 | Assignment 2 - Announced | Lab7 - Due 03/04 11:59 pm |
| 03/10 | Spring Break (No classes) | | |
| 03/17 | MongoDB - Part 2 | | Lab8 - Due 03/18 11:59 pm |
| 03/24 | MongoDB - Part 3 (Wrap-up) + Graph DBs Intro | Quiz 2 - at the beginning of the lecture<br>Neo4j Installation | Lab9 - Due 03/25 11:59 pm |

| | | | |
|---|---|---|---|
| 03/31 | Neo4j - Part 1 | Project – Announced<br>Assignment 2 - Due 11:59 PM | Lab10 - Due 04/01 11:59 pm |
| 04/07 | Neo4j - Part 2 | Assignment 3 - Announced | Lab11 - Due 04/08 11:59 pm |
| 04/14 | Neo4j - Part 3 | Quiz 3 - at the beginning of the lecture | Lab12 - Due 04/15 11:59 pm |
| 04/21 | Neo4j - Part 4 | Assignment 3 - Due 11:59 PM | Lab13 - Due 04/22 11:59 pm |
| 04/28 | Project Presentations | Slides Submission - Due 11:59 pm | |

Additional Deadlines / Final Exam

| DATE | DEADLINE |
|---|---|
| 04/29 | Project (the work) - Due Date 11:59 pm |
| Final Exams | TBA (per university schedule) |

## EXAMS AND QUIZZES SYLLABUS

- The syllabus for the exams and quizzes will be available online on the Blackboard platform, but students must attend the in-class presentation. This session will provide vital insights and preparation tips for effectively approaching these assessments.

- In this course, the exams and quizzes are structured to comprehensively evaluate students' understanding of both theoretical and practical aspects of the subject. They will feature a blend of T/F, MC, and essay-type questions. The T/F questions aim to test fundamental concepts quickly (the smallest portion of the exam), while the MC and essay questions delve deeper into students' grasp of the material, assessing both theoretical knowledge and practical application skills.

- The exams and quizzes are designed to cover a range of areas. They will test theoretical understanding, ensuring students have a strong foundation in the subject matter. Practical application will also be a key focus, with questions that might involve writing queries or interpreting results to demonstrate the real-world applicability of the course content. Additionally, analytical skills will be assessed through questions that require students to read, comprehend, and articulate well-reasoned responses to various scenarios or problems. This approach is intended to evaluate critical thinking and the effective application of knowledge.

- The objective of these assessments is to ensure a well-rounded and comprehensive understanding of the course material, preparing students for academic and professional endeavors in multi-paradigm

databases.

## GRADING

This course will be graded on a 100-point scale. Student grades will be based on the weighted average as follows:

- Assignments_and_Quizzes_out_of_50 - **50%**
- Labs - **10%**
- Midterm Exam - **10%**
- Project work and Presentations - **10%**
- Attendance and Participation - **5%**
- Final Exam - **15%**

### GRADING DETAILS

1. **Exams (25% of total grade)**

    - **Midterm Exam (10%)**
        - On Blackboard. A collection of MC, Essays, and T/F questions. Must be taken in class.
    - **Final Exam (15%)**
        - On Blackboard. A collection of MC, Essays, and T/F questions. Must be taken in class.

2. **Individual Assignments and Quizzes (50% of total grade)**

    - **Three Quizzes and Three Assignments**
        - These will test practical skills and theoretical knowledge. 1 quiz and 1 assignment per database paradigm.
    - **Scoring**
        - The best 5 scores out of the 6 grades will be considered for grading.
        - If you're satisfied with the grades of the first five, the 6th assignment/quiz is optional.

3. **Course Group Project (10% of total grade)**

        - Students **(in groups)** will develop and document the project's code and files.
        - All related materials, including code, data sets (if applicable), and analysis, must be thoroughly documented.
        - Use your style and structure to write it (no specific template is required)

        - **Presentation:**
            - As part of the evaluation, teams will present their projects. Presentations should be clear, concise, and professional, utilizing relevant examples.
                - It's important to avoid overloading slides, reading directly from them,
                - Present your project with screenshots, text, and visuals.
                - Ensure slides are clear and concise, covering the topic, methods, results, and insights.

**Teamwork and Presentation**

- The project will be conducted in groups. Team allocations will be detailed in an Excel file and attached on the proper date, effectively managing presentation time.

- **DELIVERABLES:**
  - The code and all related datasets, etc.
  - The PowerPoint Presentation
  - 1-2 pages explaining the problem, your ideas, challenges, and what you did to do it. It's just simple writing about your work.

4. **Labs (10% of total grade)**
5. **Attendance and Participation (5% of total grade)**

## GRADING SCALE

| | |
|---|---|
| 93 - 100 | A |
| 90 - <93 | A- |
| 87 - <90 | B+ |
| 83 - <87 | B |
| 80 - <83 | B- |
| 77 - <80 | C+ |
| 73 - <77 | C |
| 70 - <73 | C- |
| <70 | F |

## University policies

Information on university policies, academic support, support for students in and outside the classroom, and GW campus emergency information can be found at the following link:
**bulletin.gwu.edu/university-syllabus-policies/**