

NoSQL Database Management

MISM Course F21-95-737 A1

Fall 2021

Carnegie Mellon University

Instructor: Dan Costa

E-mail: dlcosta@andrew.cmu.edu

Virtual Office Hours: Sundays and Wednesdays, 8:30-9:30PM EST, and by appointment

Web site: <https://cmu.instructure.com/>

Textbooks

- Dan Sullivan. *NoSQL for Mere Mortals*. Addison-Wesley Professional. 2015. ISBN: 0134023218 (DS)
- Guy Harrison. *Next-Generation Databases*. Apress. 2016. ISBN: 9781484213292 (GH)

Prerequisites and Requirements

Prerequisite: 95-703, Database Management

Requirement: Students *MUST* have a computer with the ability to install a virtual machine.

Note: This course will include labs that involve the installation, configuration and programming of multiple databases. It is strongly recommended that students have some experience with programming languages (e.g. Java, Python, HTML, SQL) and command-line interfaces (Windows Command Prompt, Unix shell).

Course Description

The widespread emergence of big data storage needs has driven the development and adoption of a new class of non-relational databases commonly referred to as NoSQL databases. This course will explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems. Core concepts of NoSQL databases will be presented, followed by an exploration of how different database technologies implement these core concepts. We will take a closer look at 1-2 databases from each of the four main NoSQL data models (key-value, column family, document, and graph), highlighting the business needs that drive the development and use of each database. Finally, we will present criteria that decision makers should consider when choosing between relational and non-relational databases and techniques for selecting the NoSQL database that best addresses specific use cases.

Learning Objectives

| Learning Objective | How Assessed |
|---|--|
| Demonstrate competency in designing NoSQL database management systems. | Final Exam, Labs, Assignments |
| Demonstrate competency in describing how NoSQL databases differ from relational databases from a theoretical perspective. | Final Exam, Labs, Assignments, Research Report |
| Demonstrate competency in selecting a particular NoSQL database for specific use cases. | Final Exam, Labs, Assignments |

Schedule (tentative, subject to change during semester)

| Date | Lecture / Lab | Readings / References |
|--------------|---|----------------------------|
| August 30 | Introduction / NoSQL Database Theory | DS: Ch. 1-2 GH: Ch.1, 3 |
| September 6 | Key-Value Databases <i>Redis Lab</i> | DS: Ch. 3-5 GH: Ch. 7 |
| September 13 | Document Stores <i>MongoDB Lab</i> | DS: Ch. 6-8 GH: Ch. 4 |
| September 20 | Column Family Stores <i>Cassandra Lab</i> | DS: Ch. 9-11 GH: Ch. 6 |
| September 27 | Graph Databases <i>Neo4j Lab</i> | DS: Ch. 12-14 GH: Ch. 5 |
| October 4 | The Database Landscape / Choosing a NoSQL Database | DS: Ch. 15 |
| October 11 | Final Exam | |

Assignments / Research Report

There will be 2 assignments based on topics covered in lectures and your work with the tools in the lab sessions. In addition, there will be a research report. Students will select a topic germane to NoSQL databases for independent research, submit the topic for approval, and develop a 5-8 page report on their chosen topic. A draft version of each report will be peer reviewed by another student in the class. Following is a list of due dates for each assignment:

| Item | Due Date |
|--|-----------------------------|
| Homework 1 – Distributed Database Management Systems | September 6 @ 11:59 PM EST |
| Lab 1 – Redis | September 10 @ 11:59 PM EST |
| Lab 2 – MongoDB | September 14 @ 11:59 PM EST |
| Lab 3 – Cassandra | September 21 @ 11:59PM EST |
| First Draft of Research Report | September 24 @ 11:59PM EST |
| Homework 2 - Aggregate Oriented Design | September 27 @ 11:59 PM EST |
| Lab 4 – Neo4j | September 28 @ 11:59 PM EST |
| Peer Reviews of Research Reports | October 1 @ 11:59 PM EST |
| Final Version of Research Report | October 8 @ 11:59 PM EST |

Evaluation Method

Labs: 15%

Assignments: 25%

Research Report: 25% (80% final version, 10% peer review completion, 10% first draft submission)

Final Exam: 35%

Students will only have 2 weeks after an assignment or exam is returned to question or challenge a grade. After the two-week challenge period, the grade will not be changed. Please contact the instructor if you wish to question a grade.

Grading Scale

100 - 98 A+
97 - 92 A
91 - 90 A-
89 - 88 B+
87 - 82 B
81 - 80 B-
79 - 78 C+
77 - 72 C
71 - 70 C-

Grade Distribution

I plan on using the Heinz School guidelines in deciding on the overall grade distribution. Accordingly, the average grade will be an A-. However, I grade on an absolute scale. If every student does well in the class, each will get an A+ regardless of the recommended grading scale. The same holds true on the other end of the scale.

Final Exam

The final exam will cover material from the entire course. The exam will be held on Monday, May 3 from 6:30PM EST to 9:20PM EST. Please do not schedule anything that might conflict with the final exam. No one will be excused from it and there will be no make-up exam dates.

Late assignment policy

I WILL NOT accept late assignment submissions unless the student has made arrangements with me prior to the assignment's submission deadline. REQUESTS FOR ASSIGNMENT DUE DATE EXTENSIONS MUST BE MADE NO LATER THAN 12 PM ON THE DUE DATE.

Policy on cheating and plagiarism

This course follows Heinz School and Carnegie Mellon policies for student conduct, including policies that address inappropriate student collaboration and plagiarism. Each student is responsible for handing in their own work. For any assignment found to be the partial or complete result of cheating or plagiarism, your grade for that assignment will be zero. Cheating is defined as inappropriate collaboration among students on an assignment. This can include copying someone else's work with or without alteration. When students are found to be collaborating in this way, BOTH will pay the penalty regardless of who originated the work.

Student Wellness

Take care of yourself, and each other! Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress. All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful. If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at <http://www.cmu.edu/counseling/>. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

Accommodations for Students with Disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

How to Succeed in This Class

1. Get started on the research project as soon as possible. Even just spending some time in week 1 thinking about topics you may be interested in learning more about can make a significant difference.
2. Do not get stuck on syntax when working on labs and homework assignments. If you find yourself spending large amounts of time troubleshooting syntax errors for a query, reach out to me or one of the teaching assistants.
3. Complete the assigned readings for each week, if not before the class for which they are assigned, then after that class and before the next.
4. Use office hours productively. Get feedback on outlines of your research reports, suggestions for references, tips on homework or lab questions, and clarifications of lecture materials. Make sure you are comfortable applying the concepts and lessons learned from course case studies to new requirements or use cases.
5. Complete as much of the research project for the first draft as is possible. If you feel your research project would benefit from accessing future week's lecture materials or supplemental information, please don't hesitate to ask for it.