NoSQL Database Management

MISM Course 95-737

Carnegie Mellon University

Instructor: Derrick Spooner
E-mail: dspooner@andrew.cmu.edu

Please include both the instructor and the TAs when emailing to get the fastest response.
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 an Intel-based virtual machine. Apple M1 chips are *NOT* natively compatible with the course virtual machine, and you should plan to utilize an Intel-based architecture if possible. See https://www.heinz.cmu.edu/heinz-shared/_files/pdf/laptop-program-guidelines-2021.pdf
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
Week 1	Introduction / NoSQL Database Theory	DS: Ch. 1-2 GH: Ch.1, 3
Week 2	Key-Value Databases *Redis Lab*	DS: Ch. 3-5 GH: Ch. 7
Week 3	Document Stores *MongoDB Lab*	DS: Ch. 6-8 GH: Ch. 4
Week 4	Column Family Stores *Cassandra Lab*	DS: Ch. 9-11 GH: Ch. 6
Week 5	Graph Databases *Neo4j Lab*	DS: Ch. 12-14 GH: Ch. 5
Week 6	The Database Landscape / Choosing a NoSQL Database	DS: Ch. 15
Week 7	**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 due on week 6. 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 due by week 4, and this draft version will be peer reviewed by another student in the class. Final versions of the research reports will be due by week 6. Following is a list of due dates for each assignment:

Item	Due Date
Homework 1 – Distributed Database Management Systems	Week 2 @ 6:30 PM EST
Lab 1 – Redis	Week 2 @ 6:30 PM EST
Lab 2 – MongoDB	Week 3 @ 6:30 PM EST
First Draft of Research Report	Week 4 @ 6:30 PM EST
Lab 3 – Cassandra	Week 4 @ 6:30 PM EST
Peer Reviews of Research Reports	Week 5 @ 6:30 PM EST
Homework 2 - Aggregate Oriented Design	Week 5 @ 6:30 PM EST
Lab 4 – Neo4j	Week 5 @ 6:30 PM EST
Final Version of Research Report	Week 6 @ 6:30 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. You must provide justification for why the specific question(s) on an assignment should be reviewed and updated.

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 closed-notes, closed book, and held in person during the class meeting time of Week 7 using the Respondus LockDown Browser. Please do not schedule anything that might conflict with the final exam. No one will be excused from this date without prior written approval and there will be no make-up exam dates.

Late assignment policy

Homework is due at 6:30 PM EST on the assigned due date. I WILL NOT accept late homework unless the student has made arrangements with me prior to the assignment's due date. PRIOR ARRANGEMENTS MUST BE MADE NO LATER THAN 12 PM ON THE DUE DATE. You have unlimited attempts to re-submit updated copies of your assignments in Canvas until the due date/time, and I will only consider the most recent, on-time submission for grading.

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.