**Syllabus for CSC 6712 Distributed Storage Systems**

Instructor: Dr. RJ Nowling

Email: [nowling@msoe.edu](mailto:nowling@msoe.edu)

Office: DH421

Office hours: by appointment.

**What You'll Learn This Semester**

In some applications, data storage and processing needs have vastly exceeded what can be accomplished using a single computer. A number of database and file systems that use distributed computing techniques to provide enhanced scalability and reliability have become available and seen wide adoption. This course will cover software architectures, algorithms, and practical implications of approaches for scaling storage systems to large data sizes and high read/write throughputs, providing elasticity in the face of changing loads, and reliability in the face of failures. Relevant papers will be reviewed alongside case studies of industry and open-source implementations. Students will complete a term-long project to implement a functional distributed storage system. Previous coursework in or experience with systems programming (including networking and parallel programming) and databases recommended. (prereq: CSC 5201 Microservices and Cloud Computing or instructor consent)

**Outcomes**

On successful completion of this course, the student will:

* Compare and contrast architectures used for high-volume networking in modern databases.
* Describe the implications of at-most once, at-least once, and only-once message delivery and processing.
* Implement a file format for on-disk data storage and associated data structures for indexing.
* Apply hashing algorithms to partition and distribute data in a distributed system.
* Describe the tradeoff between availability and consistency according to the CAP theorem and the practical implications.
* Identify examples of potential failures of a storage system and describe their implications on applications.
* Describe an algorithm for distributed consensus.
* Compare and contrast replication strategies in terms of reliability and latency.
* Provide examples of distributed databases and file systems that are either consistent or highly-available.

**A person smiling for the camera

Description automatically generatedAbout Your Instructor**

I was born in Ft. Lauderdale, FL. In 2010, I moved to South Bend, IN for graduate school. I met my wife during that time, and we moved to Milwaukee in January 2015 for her to pursue an amazing career opportunity. At the time, I was working remotely as a software engineer in industry. From May 2014 – August 2018, I worked in industry as a "big data" software engineer at Red Hat, Inc. and a "data science engineer" at an online advertising company called AdRoll, Inc.

This Fall will be the start of my seventh year of teaching at MSOE. My specialties include data science, machine learning, and algorithms / data structures. I also maintain an active research programs in genomics and ML production systems.

Since moving to Milwaukee, I've became an avid cyclist. More recently, however, I started practicing karate. My wife and I have two dogs that I enjoy taking to the dog park.

**How This Class Will Be Delivered**

This class is structured as a seminar. You will have weekly assigned readings, which we will discuss synchronously. I'll track reading participation through required submission of reading notes. The readings will be supported by multi-week programming projects. The course will culminate with an analysis of existing databases using what we've learned and a final exam that evaluates your knowledge of the readings.

Our class will be delivered completely online. We will meet one night per week from 5 to 7 pm on MS Teams (see Canvas for the link). The first hour of each class will be used for a synchronous discussion. The second hour will be used as a study hall or office hours for you to work on your projects, ask me questions, or engage in some light socialization. We are fortunate that students like you are coming into the class with valuable real-world experience. I am hoping that we will learn as much from each other.

Despite being online, my goal is to be available to you and ensure that all of us will have opportunities to engage and interact just as if class was held in person. I am aiming for everyone to be successful this quarter. Please feel free to reach out with questions, concerns, or feedback about class topics, how the class is running, how to be successful with learning online, or topics outside of class. During my four years of industry work, I was entirely remote. I've learned strategies to help make online and remote work more successful, and I am always happy to brainstorm with you on ways to help with time management, studying, or counteracting the psychological impacts of working online. I also am happy to help with accommodations or offer flexibility as needed to help you be successful as long as you provide notice in advance (24 hours or more).

**Learning Activities and Materials**

**Canvas**

You access the MSOE Canvas instance here:

<https://msoe.instructure.com/>

All assignments will need to be submitted through Canvas.

## Readings

The readings will be drawn from:

* *Database System Concepts* (7th Edition) by Silberschatz, Korth, and Sudarshan.
* *TCP/IP Sockets in C (2nd Edition)* by Donahoo and Calvert.
* *Database Internals* by Petrov. (available through the library)
* *Computer Systems: A Programmer's Perspective (3rd Edition)* by Bryant and O'Hallaron (available online)
* Various papers provided by the instructor.

You will be required to submit notes for every reading. The notes should summarize the main ideas from the readings with enough detail for me to confirm that you read what was assigned. The reading notes will be good study material for the final exam.

**Programming Assignments**

Your solutions to the programming projects will be submitted through Canvas. Within the Canvas section of the class, you should select the “Content” tab, and the labs will be posted as assignments. Notebooks should be saved as PDFs and uploaded through Canvas. We will review solutions together in class, and I'll answer any questions that might come up.

**Database Review**

You will apply the concepts learned in class to evaluate the design and implementation of an existing database. You will give a 15-minute presentation. Through this, you'll demonstrate your ability to apply what you've learned while giving the entire class an opportunity to broaden their knowledge of existing databases.

**Final Exam**

There will be a final exam in the last week of class. The exam will consistent of short answer questions drawn from the readings and discussions. The exam will be provided as a Word document on Canvas (only available during test time). You will download the document, write your answers, and submit the document through Canvas as a normal assignment. I will grade the exams and upload the graded versions with feedback in the assignment comments.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grades**   |  |  | | --- | --- | | **Item** | **Percentage** | | Programming Projects | 50% | | Reading Notes | 10% | | Database Review | 10% | | Final Exam | 30% | | **Grading Scale**   |  |  | | --- | --- | | **Letter Grade** | **Percentage Needed** | | A | >=93% | | AB | >=89% | | B | >=85% | | BC | >=81% | | C | >=77% | | F | <77% | |

**Plagiarism**

All work submitted by you is expected to be your own writing unless explicitly allowed in the assignment (it will be clearly written). **Copying and pasting code from other students, the internet or any other sources other than the professor or textbook is considered plagiarism and is not allowed.**  **If copying and pasting is discovered, it will cause the both the person who submitted copied code and the person they obtained the code from to earn a 0 on the lab/assignment, and depending on the circumstances, it may cause the submitter to earn a 0 in the course.** IF YOU ARE UNSURE if “borrowing” code is allowed, please ask. It is better to ask for permission than forgiveness in this instance.

All cases of suspected plagiarism or cheating will be submitted to the Vice President of Academics and may become part of your permanent academic record as per campus policy.

## Special Accommodations

If you require special accommodations, please notify the instructor within the first three (3) weeks of the quarter to ensure adequate time to provide appropriate accommodations.

**Schedule**

See Canvas for assignment details and due dates.