Course Syllabus

Instructor: Marwan Sabbouh, Dr. of Engineering

marwansabbouh@gmail.com (mailto:marwansabbouh@gmail.com)

Course name: INFO 7255

Course description

New data points are being generated at ever increasing rates. Traditional techniques based on relational databases to ingesting, storing, indexing, and analyzing the data are no longer sufficient to deal with the volume, variety, and velocity of new data points. The volume, variety, and velocity of new data points are creating bottlenecks at every stage of the processing chain. This course will present Big Data architecture for building distributed software systems. At the outer endpoint of the distributed system, there is a need to quickly validate the incoming data so as to maintain data quality. When storing the data, write latency can never exceed the tens of milliseconds for any real world application with a healthy user base. When indexing the data, the indexer throughput rate must be high enough to keep up with velocity increase of the incoming data. The indexing technique must support logical operators, wildcards, geolocation, join, and aggregate queries. Once the data is stored and indexed, we are faced with other challenges related to near real-time predictive analytics. The issue for near real time analytics is how quickly we can take advantage of new data points after they are stored in the system to answer a question. This requires that the duration of the workflow required to ingest, store, index, and analyze the data be kept to a minimum. Even after all these requirements are met, there is one additional requirement. The above system must be schema less. That is, the system must support extensibility of its own data models and the addition of new data models without any new programming.

Teaching methodology

This class will meet once per week. Each weekly session will highlight a big data architectural pattern. A typical weekly session is divided into two halves. The first half features a lecture by the instructor highlighting the design pattern, followed by the second half where each student applies the lesson learned to build a component of the Big Data architecture. By the end of the course, it is expected that each student would have built a prototype distributed system based on the big data design patterns.

Virtual office hours

1. Immediately following the class lecture
2. Each Thursday at 7 PM Eastern time.

Open source packages

You should expect to touch many of the software packages listed below:

1. Json simple for Json parsing
2. Spring Boot for rest API development
3. Elastic Search for search and retrieval capabilities
4. Redis for Cache solutions
5. Json Schema for schema validation
6. Zuul for API Gateway pattern

Schedule

Week 1: Jan 13, 2024	Introduction to Big Data architecture
Week 2	Strongly typed data protocols
Week 3	Key value stores& Rest APIs
Week 4:	Json as Graph
Week 5: Feb 10, 2024	Prototype demo 1
Week 6	Securing rest API
Week 7	Oauth 2.0
Week 8	Search Part 1
Week 9 March 10, 2024	Prototype demo 2
Week 10: March 17, 2024	Spring Break
Week 11	Search part2, Queuing

Week 12:	Introduction to GraphQL
Week 13:	Selected Topics
Week 14: April 13	Final Prototype demo
Week 15 :April 20	Final exam

Grading

Final exam	26 pts
MCQ	0 pts (3 online class only)
Class attendance	3 pts (0 online class only)
3rd demo	41 pts
On-time delivery	-3 pts
Flawless presentation	-2 pts
Parent-child	-10 pts
Patch, including to the indexer	-8 pts
Securing your API	-4 pts

Queue	-6 pts
Basic HTTP methods	-4 pts
ETAG	-4 pts
2nd demo	25 pts
Patch	-6 pts
security	-6 pts
ETag	-5 pts
Basic HTTP methods	-5 pts
Penalty for missing the deadline	-3 pts
First demo	5 pts
Penalty for missing the deadline	-1 pts
Basic HTTP methods	-4 pts

Attendance policy

This only applies to on ground classes. Online students are exempt from this policy.

1. Students can miss up to two classes without any penalties, provided you notify me in writing 12 hours prior to the class.

2. if students need to miss more classes, I will require a medical proof.
3. Absence of the medical proof, I will deduct up to 3 points depending on the number of classes missed.
4. You are considered absent if you are more than 15 minutes late

Academic Integrity

A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to **http://www.northeastern.edu/osccr/academic-integrity-policy/ (http://www.northeastern.edu/osccr/academic-integrity-policy/)** to access the full academic integrity policy.

Student Accommodations

Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit **http://www.northeastern.edu/drc/getting-started-with-the-drc/ (http://www.northeastern.edu/drc/getting-started-with-the-drc/)** .

Library Services

The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for Education specific resources, visit **http://subjectguides.lib.neu.edu/edresearch (http://subjectguides.lib.neu.edu/edresearch)** .

Diversity and Inclusion

Northeastern University is committed to equal opportunity, affirmative action, diversity and social justice while building a climate of inclusion on and beyond campus. In the classroom, member of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration and an awareness of global perspectives on social justice.

Please visit **http://www.northeastern.edu/oidi/ (http://www.northeastern.edu/oidi/)** for complete information on Diversity and Inclusion

TITLE IX

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty and staff.

In case of an emergency, please call 911.

Please visit www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.

Course Summary:

Date	Details	Due
Sat Feb 10, 2024	Demo one (https://northeastern.instructure.com/courses/171942/assignments/2107156)	due by 10am
Sat Mar 9, 2024	Second demo (https://northeastern.instructure.com/courses/171942/assignments/2107157)	due by 10am
Sat Apr 13, 2024	Third demo (https://northeastern.instructure.com/courses/171942/assignments/2107158)	due by 10am
	Attendance (https://northeastern.instructure.com/courses/171942/assignments/2107155)	due by 11:59pm