Northeastern University
College of Engineering

INFO 7255 Advanced Big-Data Applications and Indexing Techniques

Course Information
Course Title: Advanced Big-Data Applications and Indexing Techniques
Course Number: INFO 7255
Term and Year: Spring 2024
Credit Hour: 4
Course Format:

Instructor Information
Full Name: Marwan Sabbouh
Email Address: m.sabbouh@northeastern.edu

Course Prerequisites
Graduate level CSYE 6220 Minimum Grade of B- or Graduate level INFO 6205 Minimum Grade of B- or Graduate level INFO 6250 Minimum Grade of B-

Course Description
This class covers a use case from its inception to its implementation using big-data architecture design pattern. Covers big-data design and indexing patterns to organize, aggregate, and manipulate huge amounts of data beyond human scale. Offers students an opportunity to learn advanced techniques to improve the performance, security, and robustness of ReST APIs, GraphQL, key/value stores, and Inverted- Index class of solutions. Additionally, we study the application of these technologies to solve real-world use cases in a scalable manner.

Standard Learning Outcomes

Learning outcomes common to all College of Engineering Graduate programs:

1. An ability to identify, formulate, and solve complex engineering problems.

2. An ability to explain and apply engineering design principles, as appropriate to the program's educational objectives.

3. An ability to produce solutions that meet specified end-user needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors.
The Information Systems Program accepts students of different engineering backgrounds with minimum programming skills and produces first class Information Systems engineers that operate at the intersection of real-world complexity, software development, and IT management. Graduating students will be able to construct end-to-end advanced software applications that meet business needs.

Specific Learning Outcomes for the Information Systems program:
  1. Create a strong technical foundation through diverse, high-level courses

2. *Built crucial interpersonal skills needed to succeed in any industry*
3. *Foster a deep level of applied learning through project based case studies*

Course Outcomes and Assesment Standards

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 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.

Standard Learning Outcomes

Learning outcomes common to all College of Engineering Graduate programs:
1. *An ability to create an secured rest APIs*
2. *Become fluent in the application and usage of key value stores, e.g. Redis, Cassandra, etc.*
3. *Become fluent in securing ReST APIs using oAUTH 2.0*
4. *Become fluent in the application of inverted index search engines*
5. *Become fluent in Graph QL*
6. *Become fluent in strongly typed data transfer protocols, OData, etc..*

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.

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	Introduction to Big Data architecture
Week 2	Strongly typed data protocols
Week 3	Data modeling with Json
Week 4	Rest API
Week 5	Prototype demo 1
Week 6	Rest API security
Week 7	OAuth 2.0
Week 8	Search and Discovery
Week 9	Prototype demo 2
Week 10	Data modeling for search
Week 12	Introduction to GraphQL
Week 13	Addressing
Week 14	Final prototype demo
Week 15	Makeup

End-of-Course Evaluation Surveys
Your feedback regarding your educational experience in this class is very important to the College of Professional Studies. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this course, please take the time to complete the evaluation survey at **https://neu.evaluationkit.com**. Your survey responses are **completely anonymous and confidential**. For courses 6 weeks in length or shorter, surveys will be open one week prior to the end of the courses; for courses greater than 6 weeks in length, surveys will be open for two weeks. An email will be sent to your HuskyMail account notifying you when surveys are available.

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/ 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/.

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.

24/7 Blackboard Technical Help

For immediate technical support for Blackboard, call 617-373-4357 or emailhelp@northeastern.edu

Within Blackboard, open a support case via the red support button on the right side of the screen, click Create Case

myNortheastern, e-mail, and basic technical support
Visit the Information Technology Services (ITS) Support Portal
Email: help@northeastern.edu
ITS Customer Service Desk: 617-373-4357

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/ 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.