Northeastern University
College of Engineering

INFO 7250 Engineering of Big Data-Systems

Course Information
Course Title: Engineering of Big Data-Systems
Course Number: INFO 7250
Term and Year: Summer 2022
Credit Hour: 4
Course Format: On-Ground

Instructor Information
Full Name: Yusuf Ozbek
Email Address: y.ozbek@northeastern.edu

Course Prerequisites
Graduate level INFO 6205 Minimum Grade of B- or Graduate level INFO 6250 Minimum Grade of B- or Graduate level INFO 7390 Minimum Grade of B- or Graduate level CSYE 6220 Minimum Grade of B-

Course Description
Introduces a general framework for thinking about big data. Services such as Web analytics and intelligent e-commerce have promoted a rapid increase in the volume of data generated, analyzed, and archived. In order to solve the problems related to big data, a newer type of database product has emerged. Covers how to apply technologies like Hadoop, Accumulo, MongoDB, and various NoSQL databases to build simple, robust, and efficient systems to manage and analyze big data. Also describes an easy approach to big data systems that can be built and run by a small team of students. Guides students through the theory of big data systems, how to implement them in practice, and how to deploy and operate them once they are built.

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

Services like social networks, web analytics, and intelligent e-commerce have promoted a rapid increase in the volume of data generated, analyzed and archived. These larger-volumes of data are too big for a traditional database, and are imposing challenges for storing, analyzing, and archiving. In order to solve the problems related to big data, a newer type of database products have emerged. These database products are collectively identified as NoSQL, and are quite varied with their unique features.

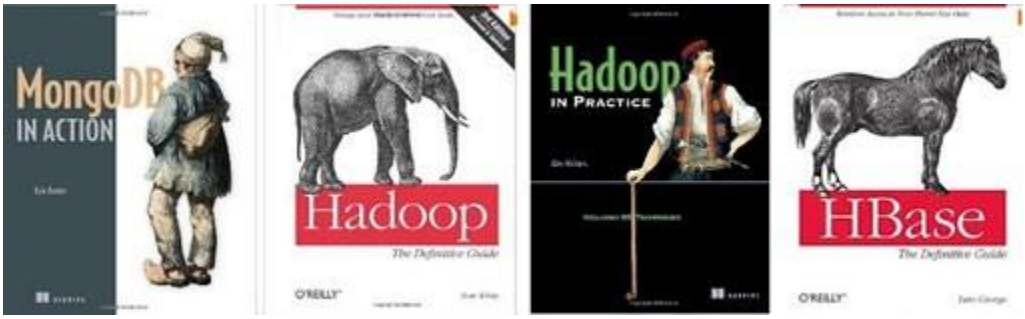
This class introduces a general framework for thinking about big data, and then shows how to apply technologies like Hadoop, HBase, MongoDB, and various NoSQL databases to build simple and efficient systems to manage and analyze big data, and describes an easy approach to big data systems that can be built and run by students. This course serves as an introductory course for Information Systems students who want to explore Big Data storage, processing, analysis, and application issues.

At the end of this course, students will

- understand the capabilities and pitfalls of Big Data systems, and how these issues are addressed.
- become familiar with the fundamental concepts of Big Systems and analytics.
- learn how to implement the MapReduce programming model.
- understand the challenges for Big Data Applications that deal with very large volumes of data.
- propose scalable solutions for problems related to Big Data.
- get hands-on experience on Big Data Analytics.

Reference Books:

Note: Electronic versions are available on the Snell Library website, and Safari Books Online.



Core topics:

The course is divided into three main sections:

- Introduction to the Big Data, exploring challenges, trends, and applications in the Industry,
- Algorithms for Big Data analysis,
- Integrating Hadoop with other technologies.

Course Outline

- Introduction - a new paradigm for Big Data and NoSQL basics
- Getting Initial Hands-on Experience with MongoDB
- Replication in MongoDB
- Sharding in MongoDB
- Language Bindings - interacting and interfacing with NoSQL
- MapReduce
- Cassandra
- Hadoop
- Hadoop Integration

- Apache Pig
- Analyzing Big Data with Hive
- Hbase
- Mahout

Assignments:
Project is the most important learning tool of this class. There will be weekly assignments during the semester, and one final project to apply all the technics learned in the class to make complex analyses on a given dataset.

Labs:
We will have multiple labs during the semester. These labs are based on Apache Hadoop, and we will use a virtual machine for most of the labs.

Final Project:
Students may use a given dataset, or explore a problem of their interest and propose their own solution. The project has the following deliverables:

- 1 or 2 pages of project proposal, to be submitted after the mid semester, explaining why the problem is important, and what analyses are to be performed.
- Presentation of the project at the end of the semester
- A written report, to be submitted before the presentation, to highlight motivation, method, results, and conclusion.

Grading Policy:
Assignments: 10%
Midterm: 20-30%
Final Project: 20-30%
Final Exam: 20-30%

Attendance Policy:
Attendance is required. Students are responsible for any material covered in class. Lots of the materials covered in class will not be in the textbook. Announcements about homework, projects, programming assignments, etc. may be made in class or online or by emails.

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.