INSTRUCTORS

John Rachlin

Byron Wallace

ABOUT THE COURSE

Catalog Discription

Provides a deep-dive into the design and implementation of enterprise-grade software systems with an emphasis on software architectures for more complex data-driven applications. Covers extensible architectures that support testing, data provenance, reuse, maintainability, scalability, and robustness; building software APIs and libraries for wide-scale adoption and ease-of-use. Students will learn how to design implement and test complex loosely-coupled service-oriented architectures and will learn how to use the python programming language for distributed processing, stream-based data processing, and inter-process communication via message-passing. Homework assignments will center around a semester-long project exploring various topical machine learning themes possibly including but not limited to: library construction for data cleansing, analysis, and data mining, sentiment analysis and natural-language processing, scientific computing problems in bioinformatics, astronomical data analysis, or healthcare. Student will explore the features, capabilities, and underlying design of popular data analysis and visualization frameworks.

4.000 Credit Hours
Prerequisites: DS2000,DS2500

Textbooks

Title	Percival and Gregory (2020): Architecture Patterns with Python: Enabling Test-Driven Development, Domain-Driven Design, and Event-Driven Microservices (O'Reilly)
Buy online	Amazon.com

Description As Python continues to grow in popularity, projects are becoming larger and more complex. Many Python developers are now taking an interest in high-level software design patterns such as hexagonal/clean architecture, event-driven architecture, and the strategic patterns prescribed by domain-driven design (DDD). But translating those patterns into Python isn't always straightforward. With this hands-on guide, Harry Percival and Bob Gregory from MADE.com introduce proven architectural design patterns to help Python developers manage application complexity—and get the most value out of their test suites.

Resources

Anaconda 3.7	https://www.anaconda.com/distribution/#download-section
Atom editor	https://atom.io/
PyCharm	https://www.jetbrains.com/pycharm/
Jupyter Notebooks	https://jupyter.org
Piazza	Coming soon!

COURSE GOALS

This course introduces programming for data and information science...

Lectures	will focus on developing an advanced understanding of enterprise pr germane to Data Science and Data Engineering
Homework	will apply software architecture principles and best practices via pro
4-5 Quizzes	will assess individual understanding of enterprise programming conc

Project will demonstrate proficiency in software architecture best practices.

Week	Topics	Reading
1	Development Environments: PyCharm, Git, Command Line Essentials	Handouts
2	Unit testing and debugging. Domain Modelling and Design Challenges	Ch 1
3	Repository Pattern	Ch 2
4	Abstraction with Coupled and Loosely-coupled designs	Ch 3
5	Service-oriented Architectures	Ch 4
6	Test-Driven Development and Profiling	Ch 5
7	Unit of Work Pattern: Context Managers	Ch 6
8	Aggregates, Consistency Boundaries, Microservices	Ch 7
9	Event-driven architectures, Asyncio and socket connections	Ch 8
10	The message bus, message brokers	Ch 9 & 10
11	Events for integrated microservices	Ch 11

12	CQRS: Command-query responsibility separation	Ch 12
13	Dependency injection and bootstrapping	Ch 13
14	Project science fair	N/A

EVALUATION

The final grade for this course will be weighted as follows...

- Homework: 60%
- Quizzes: 20%
- Class Project: 20%

Final grades will be assigned based on the following scale...

A
95 – 100
A–
90 – <95
B+
87 – <90
B
82 – <87
B–
80 – <82
C+
77 – <80
C
72 – <77
C–
70 – <72
D+
67 – <70
D
62 – <67
D–
60 – <62
F

```
<60
```
Active participation (e.g. contributing in class, Piazza) may be used to adjust the final grade upwards at the discretion of the instructor. At the point of final-grade determination, numeric grades are rounded naturally (e.g. 94.4 is a 94, but 94.5 is a 95).

Homework

Submissions will be made via Blackboard. Homeworks are due on Mondays at 11:59pm. Late work will not be accepted.

Project

The goal of the project is to gain hands-on experience with data science. Project deliverables will include source code, and a research paper written in the format of typical conference proceedings. There will be a class poster session during the last week of class where students will share their projects.

Late Policy

Late assignments will not be accepted. This is because we will be reviewing the (weekly) homeworks in lecture. One missed homework will, however, be forgiven (your lowest homework grade will be dropped).

Students who miss scheduled quizzes will not, as a matter of course, be able to make up those quizzes. If there is a legitimate reason why a student will not be able to complete an assignment on time or not be present for a quiz, then they should contact the instructor beforehand. Under extreme circumstances, as decided on a case-by-case basis by the instructor, students may be allowed to make up assignments or quizzes without first informing the instructor.

Academic Conduct

This class has very strict standards for borrowing code: if you borrow anything for use in your homework/project, you must have a citation. A good guideline is that if you take more than three lines of code from some source, you must include the information on where it came from. A URL or a notation (e.g., "MATLAB help files") is fine. If it is an entire function, note it at the beginning of the code segment and include any original credit information. Provide a qualitative description of what you used, and what you changed/contributed. **If you have a question about what is considered a violation of this policy, ASK!**

Unless stated otherwise, assignments reflect individual work. While you may discuss concepts and ideas with other students, there is to be no direct collaboration. If you steal someone else's work, you fail the class. If someone uses your work, you fail the class. If you are unsure about this policy, **ask the instructor**. The university's academic integrity policy discusses actions regarded as violations and consequences for students: http://www.northeastern.edu/osccr/academic-integrity

CLASSROOM ENVIRONMENT

To create and preserve a classroom atmosphere that optimizes teaching and learning, all participants share a responsibility in creating a civil and non-disruptive forum for the discussion of ideas. Students are expected to conduct themselves at all times in a manner that does not disrupt teaching or learning. Your comments to others should be constructive and free from harassing statements. You are encouraged to disagree with other students and the instructor, but such disagreements need to respectful and be based upon facts and documentation (rather than prejudices and personalities). The instructor reserves the right to interrupt conversations that deviate from these expectations. Repeated unprofessional or disrespectful conduct may result in a lower grade or more severe consequences. Part of the learning process in this course is respectful engagement of ideas with others.

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.

If you or someone you know has been a survivor of a Prohibited Offense, *confidential* support and guidance can be found through **University Health and Counseling Services** staff (http://www.northeastern.edu/uhcs/) and the **Center for Spiritual Dialogue and Service** clergy members (http://www.northeastern.edu/spirituallife/). By law, those employees are not required to report allegations of sex or gender-based discrimination to the University.

Alleged violations can be reported non-confidentially to the Title IX Coordinator within **The Office for Gender Equity and Compliance** at: titleix@northeastern.eduand/or through **NUPD** (Emergency 617.373.3333; Non-Emergency 617.373.2121). Reporting Prohibited Offenses to NUPD does **NOT** commit the victim/affected party to future legal action.

Faculty members are considered "responsible employees" at Northeastern University, meaning they are required to report all allegations of sex or gender-based discrimination to the Title IX Coordinator.

In case of an emergency, please call 911.

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

Students With Disabilities

Students who have disabilities who wish to receive academic services and/or accommodations should visit the Disability Resource Center at 20 Dodge Hall or call (617) 373-2675. If you have already done so, please provide your letter from the DRC to me early in the semester so that I can arrange those accommodations.