Data Science Applications Development
(3 Credits)

Course Code: DSE 6300

Class Schedule: Monday: 6:00pm-8:30 pm. (In-class, unless the university decides to go online)

Semester/Year: Winter 2022. Jan 10 to April 25.

Faculty Contact Information

Name: Dr. Kamel Rushaidat

Office hours: After class or by appointment.

E-mail address: Kamel.r@wayne.edu

Communication Policy

Email is the quickest way to reach me. I check my email on daily and I agree to respond to your emails within 24 hours.

Course Description

This course focuses on the software engineering cycle of developing a comprehensive data science application. Agile for Data Science will be used and practiced for the final class project. Other tools for application development will be used in the domain of Data science. Students will have the freedom to choose a computing platform, a SQL or a NoSQL database as the underlying infrastructure for developing a data science application. Students will also choose a domain and problem in which one needs to address one of the big data challenges: volume, velocity, or variety. Students are also encouraged to use scalable data mining and machine learning algorithms to solve the given domain problem.

Course Learning Outcomes

The Data Science Applications Development course has the following course learning outcomes:

- Explain distributed programming paradigms
- Develop distributed computation
- Develop real-time/streaming Applications
- Develop data science integration
- Understand how Agile works in Data Science applications

Course Approach

The students will apply one of the programming paradigms, either a computing platform such as the Hadoop ecosystem a SQL DB such as MySQL, MariaDB or PostgreSQL or NoSQL database such as Apache Cassandra or MongoDB to develop a data science application in a big data domain. Challenges of developing data science software applications, along with different strategies to address those challenges will be discussed in class. Students will use the final class project as an example on developing data science applications.

Prerequisites

The familiarity of high-level programming languages such as Java, Scala or Python. Modern Databases such as MySQL, Computing Platforms for Data Science, such as Hadoop.

Course Format

This is a hands-on course. Reading assignments will be given to students before class to prepare for the practical part and for class discussions. Students can work in groups so that they can learn from each other and can discuss the lessons and pitfalls during their hands-on exercise and experiences for debugging a distributed big data application.

Lecture, presentations, and discussion will focus on the discussions of datasets, their challenges, and problems domains as well as the principles, methodologies, lifecycle of developing a comprehensive big data application.

Tentative Schedule:

The **tentative** schedule is shown below.

Week	Week of	Subject(s)
Week 1	01/10	Syllabus. Course introduction. Distributed Systems architecture.
Week 2	01/17	No class, Holiday (Martin Luther King Jr. Day)
Week 3	01/24	SQL vs NoSQL database.
Week 4	01/31	Version control (Git and GitHub).
Week 5	02/07	Intro to Functional programming. Quiz 1. Assignment 1.
Week 6	02/14	Introduction to Big Data
Week 7	02/21	Introduction to Apache Kafka.
Week 8	02/28	Introduction to the Spark API (I).
Week 9	03/07	Introduction to the SPARK API (II). Quiz 2. Assignment 2.
Week 10	03/14	Spring Break, no classes this week
Week 11	03/21	Spark Topics. Agile Data Science : Introduction. Agile vs waterfall. Agile tools.
Week 12	03/28	Data ingestion (ETL), integrations, and modeling. Quiz 3. Assignment 3.
Week 13	04/04	Final class project discussion and problem statement - Developing a Data Science application. Names of partners for final project due in class.
Week 14	04/11	Work on the final project. (Finish any remaining topics)
Week 15	04/18	Final Project presentation and demo.
Week 16	04/25	Final Project presentation and demo.

Textbook & Reading Material

As this is an emerging field, there is no single textbook that covers all the required material. However, we will be using material from the following textbooks and other published papers.

1. High Performance Spark: Best Practices for Scaling & Optimizing Apache Spark by Holden Karau & Rachel Warren.
2. Development Workflows for Data Scientists: Enabling Fast, Efficient, and Reproducible Results for Data Science Teams, by Ciara Byrne. (Free download).
3. Agile Data Science 2.0. Building Full-Stack Data Analytics Applications with Kafka and Spark by Russell Jurney.
4. Data Science in Production: Building Scalable Model Pipelines with Python by Ben Weber
5. We may add a fifth reference based on the class progress.

Semester Project

Towards the end of the class, students will form teams to tackle a real-world problem they chose for a final class project. The final grade for the project will be over several deliverables during the course. The class project should be a big data application.

Each Student will also need to present a topic of their choosing on a Data Science topic. This will be discussed further in the class.

Grading

- 3 Short Assignments: 15%
- 3 Short Quizzes: 15%
- Participation in class: 5%
- Data Science Topic Presentation: 15%
- Semester Project: 50%

Grading criteria

Grading scale:
- 90% - 100 %   A
- 85% - 89.99%   A-
- 80% - 84.99%   B+
- 75% - 79.99%   B
- 70% - 74.99%   B-
- 67% - 69.99%   C+
- 63% - 66.99%   C
- 60% - 62.99%   C-
- 57% - 59.99%   D+
- 53% - 56.99%   D
- 50% - 52.99%   D-
- Below 50%   fail

Religious holidays

Because of the extraordinary variety of religious affiliations of the University student body and staff, the Academic Calendar makes no provisions for religious holidays. However, it is University policy to respect the faith and religious obligations of the individual. Students with classes or examinations that conflict with their religious observances are expected to notify their instructors well in advance so that mutually agreeable alternatives may be worked out.

Student Disabilities Services

If you have a documented disability that requires accommodations, you will need to register with Student Disability Services for coordination of your academic accommodations. The Student Disability Services (SDS) office is in the Adamany Undergraduate Library. The SDS telephone number is 313-577-1851 or 313-202-4216 (Videophone use only). Once your accommodation is in place, someone can meet with you privately to discuss your special needs. Student Disability Services' mission is to assist the university in creating an accessible community where students with disabilities have an equal opportunity to fully participate in their educational experience at Wayne State University.

Students who are registered with Student Disability Services and who are eligible for alternate testing accommodations such as extended test time and/or a distraction-reduced environment should present the required test permit to the professor at least one week in advance of the exam. Federal law requires that a student registered with SDS is entitled to the reasonable accommodations specified in the student's accommodation letter, which might include allowing the student to take the final exam on a day different than the rest of the class.

Academic Dishonesty -- Plagiarism and Cheating

Academic misbehavior means any activity that tends to compromise the academic integrity of the institution or subvert the education process. All forms of academic misbehavior are prohibited at Wayne State University, as outlined in the Student Code of Conduct (http://www.doso.wayne.edu/student-conduct-services.html). Students who commit or assist in committing dishonest acts are subject to downgrading (to a failing grade for the test, paper, or other course-related activity in question, or for the entire course) and/or additional sanctions as described in the Student Code of Conduct.

- **Cheating**: Intentionally using or attempting to use, or intentionally providing or attempting to provide, unauthorized materials, information or assistance in any academic exercise. Examples include: (a) copying from another student's test paper; (b) allowing another student to copy from a test paper; (c) using unauthorized material such as a "cheat sheet" during an exam.
- **Fabrication**: Intentional and unauthorized falsification of any information or citation. Examples include: (a) citation of information not taken from the source indicated; (b) listing sources in a bibliography not used in a research paper.
- **Plagiarism**: To take and use another's words or ideas as one's own. Examples include: (a) failure to use appropriate referencing when using the words or ideas of other persons; (b) altering the language, paraphrasing, omitting, rearranging, or forming new combinations of words to make the thoughts of another appear as your own.

Other forms of academic misbehavior include, but are not limited to: (a) unauthorized use of resources, or any attempt to limit another student's access to educational resources, or any attempt to

alter equipment to lead to an incorrect answer for subsequent users; (b) enlisting the assistance of a substitute in the taking of examinations; (c) violating course rules as defined in the course syllabus or other written information provided to the student; (d) selling, buying or stealing all or part of an un-administered test or answers to the test; (e) changing or altering a grade on a test or other academic grade records.

Course Drops and Withdrawals

In the first two weeks of the (full) term, students can drop this class and receive 100% tuition and course fee cancellation. After the end of the second week there is no tuition or fee cancellation. Students who wish to withdraw from the class can initiate a withdrawal request on Pipeline. You will receive a transcript notation of WP (passing), WF (failing), or WN (no graded work) at the time of withdrawal. No withdrawals can be initiated after the end of the tenth week. Students enrolled in the 10th week and beyond will receive a grade. Because withdrawing from courses may have negative academic and financial consequences, students considering course withdrawal should make sure they fully understand all the consequences before taking this step. More information on this can be found at: http://reg.wayne.edu/pdf-policies/students.pdf

Student services

The Academic Success Center (1600 Undergraduate Library) assists students with content in select courses and in strengthening study skills. Visit www.success.wayne.edu for schedules and information on study skills workshops, tutoring and supplemental instruction (primarily in 1000 and 2000 level courses).

The Writing Center is located on the 2nd floor of the Undergraduate Library and provides individual tutoring consultations free of charge. Visit http://clasweb.clas.wayne.edu/writing to obtain information on tutors, appointments, and the type of help they can provide.

Class recordings

Students need prior written permission from the instructor before recording any portion of this class. If permission is granted, the audio and/or video recording is to be used only for the student's personal instructional use. Such recordings are not intended for a wider public audience, such as postings to the internet or sharing with others. Students registered with Student Disabilities Services (SDS) who wish to record class materials must present their specific accommodation to the instructor, who will subsequently comply with the request unless there is some specific reason why s/he cannot, such as discussion of confidential or protected information.