

WAYNE STATE UNIVERSITY

COLLEGE OF ENGINEERING

DSE 6000: **Computing Platforms for Data Science - 3 Credits** Course Syllabus Winter 2019

Instructor: Jayashree Ravi, MS Distributed Systems.
Part Time Faculty, Wayne State University
jravi@wayne.edu

Classroom: 1133 MAIN

Office: 1133 MAIN,, Wayne State University, Detroit, MI 48202

Web Sites¹: <http://canvas.wayne.edu>

Description: Covers an overview of various computing platforms for developing, deploying, configuring a wide range of data science applications for different domains. The programming models, characteristics of supported workload, and management of performance, cost, and scalability will be compared side by side.

Course Learning Outcomes:

- At the end of the course, the successful student will be able to develop an in-depth understanding of the strengths and weaknesses of different computing platforms for data science.
- At the end of the course, the successful student will be able to develop a good understanding for the recent progress made by the scientific and information technology community in data science and understand the pros and cons of using different platforms.
- At the end of the course, the successful student will have acquired hands-on experience in the processing different types and scales of data using various open source platforms and techniques.

Prerequisites: Familiarity with any programming language; familiarity with the Linux operating system

Textbook: As this is an emerging field, there is no single good textbook for it yet. Instead, we will rely on a number of articles, tutorials, and case studies.

References: Apache foundation: <http://www.apache.org/>
Additional tutorials and web resources will be presented throughout the course to supplement classroom learning

Software: Access to the computing environment which is pre-installed with necessary platforms such as HortonWorks, Hadoop, Spark, Cassandra, and others.

Grading:	Semester Project	@ 30 x 1	30 pts
	Three Quizzes	@ 10 x 3	30 pts
	Hands-on Assignments	@ 40 x 3	40 pts

Total:100 pts

Hands-on assignments, quizzes, and semester project might be curved and changed with regard to importance (i.e., in points), at the discretion of the instructor. Project reports and assignment reports have to be typed, and when feasible, results have to be justified and thoroughly summarized (without appending lots of pages of output). Reports have to be submitted at the beginning of the class on the due date. Late reports will receive lower grades.

¹ Canvas website is protected by individual user login names and passwords. The username is the uniquely assigned WSU AccessID. To activate your WSU AccessID or change the password or set an alternate forwarding e-mail address, visit <https://computing.wayne.edu/accessid>. Call the WSU Computing & Information Technology (C&IT) Help Desk at 313-577-4778 for any difficulties.

Assignments:	Students are encouraged to bring some datasets/problems of interest. If requested, the instructor can provide datasets.
Semester Project:	This is a team project with two to three students. Students are encouraged to investigate existing open data sets and discover meaningful insights which were not found earlier.
Attendance Policy:	Students attending any given class are required to join the class within the first five minutes to minimize any class disruptions.
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 Services:	<ul style="list-style-type: none"> o <i>The Academic Success Center</i> (1600 Undergraduate Library) assists students with content in select courses and in strengthening study skills. Visit http://success.wayne.edu for schedules and information on study skills workshops, tutoring and supplemental instruction (primarily in 1000 and 2000 level courses). o <i>The Writing Center</i> 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.
Academic Dishonesty – Plagiarism and Cheating:	<p>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.</p> <ul style="list-style-type: none"> o <u><i>Cheating</i></u>: 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. o <u><i>Fabrication</i></u>: 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. o <u><i>Plagiarism</i></u>: 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 in an attempt to make the thoughts of another appear as your own. o <u><i>Other</i></u> 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 so as 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.
Student Disability 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 located at 1600 David Adamany Undergraduate Library in the Student Academic Success Services department. The SDS telephone number is 313-577-1851 or 313-202-4216 for videophone use. Once you have met with your disability specialist, I will be glad

to meet with you privately during my office hours to discuss your accommodations. 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. You can learn more about the disability office at www.studentdisability.wayne.edu.

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.

**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>

**Deferred
Grade:**

A grade of 'I' can only be assigned if all of the following criteria are met:

1. the student IS NOT currently failing the class and,
2. there is NOT a substantial quantity of work yet to be completed,
3. there is no extra work required of the instructor beyond the normal duties of grading the paper/exam,
4. there is no need for the student to attend the class in subsequent terms.

The final decision to assign an incomplete grade rests with the instructor. An 'I' grade MUST be made up within one year of assignment of the grade.

Tentative Course Schedule:

No. of Classes	Topic
1	Introduction <i>Data Science history, evolution, and current state-of-the-art techniques. What is Big Data, and why Big Data processing is different? Evolution of processing systems. Difference between Data Science/ Data Analysis, Data engineering, and Big Data.</i>
1	Intro to Descriptive Statistics <i>A revision class to get students well prepared for Data Analytics and projects</i>
3	Intro to Data Analytics <i>Data Analysis process. Investigate a DataSet. Apply NumPy, Pandas to an example dataset using Python, Anaconda, conda and Jupyter notebook will be installed for analysis.</i> <i>Hand out lab assignment, Quiz-1</i>
3	Data Wrangling <i>Data Extraction Fundamentals. Data in complex formats: CSV, JSON, XML, SQL Data. Data Acquisition techniques, Data Quality, Identify dirty data. Data curation and storage techniques.</i> <i>NoSQL Databases. Conclude with an example dataset investigation.</i>
2	Document Indexing and Searching <i>Introduce Lucene, Solr -indexing/searching techniques used by document search companies. Install Solr and run a few examples on document indexing.</i> <i>Quiz-2, Hands-on assignment - Due in 2 weeks</i>
3	Distributed / Parallel Computing and methodologies <i>MapReduce paradigm</i> <i>HDFS - Hadoop Distributed File System</i> <i>Hadoop</i> <i>YARN</i> <i>Demonstrate word count problem using the above technique.</i>
1	In Memory processing methodologies <i>Performance advantages of In-Memory techniques</i> <i>Apache Spark</i> <i>Compare the performance side-by-side with Hadoop.</i>
1	Data representations and API's <i>Resilient Distributed Datasets (RDD)</i> <i>Data Frames API</i> <i>SparkSQL</i>
3	Data Analytic platforms <i>Apache Hive</i> <i>Apache Pig</i> <i>Quiz -3</i> <i>Hands-on assignment - Due in 2 weeks</i>
2	Data Visualization and Notebooks <i>Zeppelin, Jupyter etc</i>
3	<i>Study Datasets. Investigate datasets given by the instructor and share your findings with the class. Discuss the final project and review assignment solutions. scattered across semester</i>
2	Cloud Services from Amazon, Google, Microsoft <i>In-depth on Google</i> <i>Dataproc</i>

	<i>Storage</i> <i>Compute Engine</i>
	Popular Hadoop stacks <i>HortonWorks</i> <i>Cloudera</i> <i>MapR</i> 2 <i>Semester Project - Due in 2 weeks</i>
1	Demo Day For Semester Project