

Kennesaw State University
Department of Computer Science
CS 4265: Big Data Analytics
Spring 2021

Credit Hours:	3 Credit Hours																
Pre-Requisites:	CS 3304, CS 3410																
Textbook:	Hands-On Big Data Analytics Using Apache Spark, Dan Chia-Tien Lo and Yong Shi, Jan. 2021. Advanced Analytics with Spark (Patterns for Learning from Data at Scale) By Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills O'Reilly Media, 2015 (Reference) ISBN#: 978-1-491-91276-8																
Course Webpage:	D2L																
Class Design:	Lecturing Time: T/R 3:30 PM – 4:45 PM EST Lecturing Room: J-251 and Microsoft Teams Office: J389 Office Hours: T/R 2:30 PM – 3:30 PM EST via D2L Collaborate Ultra Email: Use D2L email for course communication. Homework assignments, lecture slides, and other materials will be posted on the course webpage in D2L.																
Course Description:	This course covers algorithms and tools that are needed to build MapReduce applications with Hadoop or Spark for processing gigabyte, terabyte, or petabyte-sized datasets on clusters of commodity hardware. A wide range of data algorithms will be discussed in this course.																
Learning Outcomes:	Upon the completion of the course, students will be able to 1. explain MapReduce framework for big data analytics 2. use Hadoop and Spark for big data analytics 3. design and implement a wide range of data algorithms																
Special Dates:	<table> <tr> <td>First Day of Classes</td><td>1/12</td></tr> <tr> <td>Holiday</td><td>1/18</td></tr> <tr> <td>Last Day to Withdraw w/o Academic Penalty</td><td>3/15</td></tr> <tr> <td>Spring Break</td><td>3/8-3/14</td></tr> <tr> <td>Last Day of Class</td><td>5/3</td></tr> <tr> <td>Test 1</td><td>2/11</td></tr> <tr> <td>Test 2</td><td>3/18</td></tr> <tr> <td>Test 3</td><td>4/29</td></tr> </table>	First Day of Classes	1/12	Holiday	1/18	Last Day to Withdraw w/o Academic Penalty	3/15	Spring Break	3/8-3/14	Last Day of Class	5/3	Test 1	2/11	Test 2	3/18	Test 3	4/29
First Day of Classes	1/12																
Holiday	1/18																
Last Day to Withdraw w/o Academic Penalty	3/15																
Spring Break	3/8-3/14																
Last Day of Class	5/3																
Test 1	2/11																
Test 2	3/18																
Test 3	4/29																

Weekly Course Schedule: Subject to change

Week	Topic
1	Discussion of course syllabus and policies, Course webpage Introduction Machine Learning, Data Mining, and Big Data Analytics
2	Spark Scala
3	Hadoop MapReduce Framework
4	Decision trees
5	Linear models Linear regression
6	Regularization and Sparsity on Linear regression
7	Logistic regression
8	Naïve Bayes Classifier

9	Fisher Linear Discriminant Analysis
10	Principal Component Analysis
11	KNN
12	K-Means
13	Spark SQL
14	Project Presentation
15	Project Presentation

Assessment Criteria:	
Attendance/Discussion	5%
Assignments	25%
Test 1, 2, and 3	30%
Project Proposal	10%
Project Development	20%
Project Presentation	10%

Grade Evaluation	
A	90% - 100%
B	89% - 80%
C	79% - 70%
D	69% - 60%
F	59% or below

Course Policies and Other Information:

Attendance: Class attendance is required and very important for successful completion of the course. Students are expected to attend and participate in every class/lab/recitation via F2F or Microsoft Teams, according to the seating schedule determined in a later section. If a student misses more than 4 sessions (class meetings and/or labs), the student's final grade for the course may be reduced by 5% (that is half letter grade). Excused absences must be planned for, when possible, and justified with documentation. The student is responsible for making up missed class/lab sessions. Late arrival that causes disruption, early departure that causes disruption, excessive conversation among students (a disruption in its own right), inappropriate use of electronic devices that cause disruptions and other actions that disrupt the classroom are unacceptable. The instructor reserves formats for checking attendance such as sign-on sheets, checking for understanding, giving back tests, D2L attendance logs, etc.

Assignment Grading Policy:

For all homework assignments, if a student consults any resource (other than the text and class notes) including another individual, this consultation must be documented on the submission. This documentation must include what (or who) was consulted and what information was obtained. Copying or paraphrasing code from another source or failure to provide this documentation will be considered a violation of the Student Code of Conduct.

Assignments will be handed over throughout the term and students will be given sufficient amount of time to complete it with resources. Assignments are intended for completing on individual basis. Students are highly encouraged to submit assignment by due date to avoid penalties in marks. Late assignments will receive penalty based on the formula, $\text{penalty} = (\text{number of days late})^3$, and must be submitted via the late submission dropbox in D2L.

Project: There will be a course project that will be needed to be done in a group of 3 students. The project topics need to be approved by the instructor. Some topics will be suggested. The project has two parts for marking: implementation and presentation (the last 2 weeks reserved for project presentations), and project report. The instructor will provide feedback on proposal and progress with tentative timeline to complete project.

Class Format: Lecturing (PowerPoint Slides) with live broadcasting for all sections. Slides will be posted on the course webpage gradually throughout the semester.

Technology requirements: A webcam with a headset or a microphone are required to attend live classroom broadcasting and take exams with the eastern time zone. Apache Spark and Scala programming language will be required in a PC or Mac. D2L is used for all course materials and work submissions.

Email Policy: "Send emails only to dlo2@kennesaw.view.usg.edu (students must use their D2L email accounts to send to this address)"

- Email messages received by 5:00 PM in weekdays will be replied in the same day (by 11:59 PM).
- Other messages will NOT be guaranteed to be responded in the same day.
- Only email in D2L will be replied.

Withdrawal Policy: The last day to withdraw without academic penalty is 3/15/2021. Ceasing to attend class or oral notice thereof DOES NOT constitute official withdrawal from the course. Students who simply stop attending classes without officially withdrawing usually are assigned failing grades. Students wishing to withdraw after the scheduled change period (add/drop) must obtain and complete a withdrawal form from the Academic Services Department in the Registrar's Office.

Classroom Behavior: Students are reminded to conduct themselves in accordance with the Student Code of Conduct, as published in the Undergraduate and Graduate Catalogs. Every KSU student is responsible for upholding the provision. Students who are in violation of KSU policy will be asked to leave the classroom and may be subject to disciplinary action by the University.

Academic Integrity Statement: Every KSU student is responsible for upholding the provisions of the Student Code of Conduct, as published in the Undergraduate and Graduate Catalogs. Section II of the Student Code of Conduct addresses the University's policy on academic honesty, including provisions regarding plagiarism and cheating, unauthorized access to University materials, misrepresentation/falsification of University records or academic work, malicious removal, retention, or destruction of library materials, malicious/intentional misuse of computer facilities and/or services, and misuse of student identification cards. Incidents of alleged academic misconduct will be handled through the established procedures of the University Judiciary Program, which includes either an "informal" resolution by a faculty member, resulting in a grade adjustment, or a formal hearing procedure, which may subject a student to the Code of Conduct's minimum one semester suspension requirement.

Frequently students will be provided with "take-home" exams or exercises. It is the student's responsibility to ensure they fully understand to what extent they may collaborate or discuss content with other students. No exam work may be performed with the assistance of others or outside material unless specifically instructed as permissible. If an exam or assignment is designated "no outside assistance" this includes, but is not limited to, peers, books, publications, the Internet and the WWW. If a student is instructed to provide citations for sources, proper use of citation support is expected. Additional information can be found at the following locations.

<http://www.apa.org/journals/webref.html>

<http://bailiwick.lib.uiowa.edu/journalism/cite.html>

<http://www.indiana.edu/~wts/wts/plagiarism.html>

<http://www.virtualsalt.com/antiplag.htm>

Electronic Devices: In order to minimize the level of distraction, all beepers and cellular phones must be on quiet mode during class meeting times. Students who wish to use a computer/PDA for note taking need prior approval of the instructor since key clicks and other noises can distract other students. Recording of lectures by any method requires prior approval of the instructor. Students using a laptop in class should not check their email, browse the web, or in other way detract from the focus of the class.

Disruption of Campus Life Statement: It is the purpose of the institution to provide a campus environment, which encourages academic accomplishment, personal growth, and a spirit of understanding and cooperation. An important part of maintaining such an environment is the commitment to protect the health and safety of every member of the campus community. Belligerent, abusive, profane, threatening and/or inappropriate behavior on the part of students is a violation of the Kennesaw State University Student Conduct Regulations. Students who are found guilty of such misconduct may be subject to immediate dismissal from the institution. In addition, these violations of state law may also be subject to criminal action beyond the University disciplinary process.

Computer Usage Policy: The Kennesaw State University computer usage policy is posted at https://policy.kennesaw.edu/sites/web.kennesaw.edu.policy/files/computerusagepolicy_11212014_0.pdf. Students are responsible for being familiar with the policy and the penalties authorized in the policy. During lectures, students are not supposed to use computers, unless otherwise being told so.

Covid-19 Policy:

1. While in the corridors, students should keep to the right whenever possible. Observing this “rule of the road” will enable students to keep a maximum distance from one another as they pass in the hallway
2. Students should avoid congregating in the corridors and wait for their classes outside of buildings whenever possible (i.e., not clustering outside the classroom doors).
3. Students should enter the classroom one at a time with social distancing.
4. Classes taught in odd-numbered rooms be released 5 minutes early
5. At the end of class, instructors should dismiss students one group at a time in the case of one doorway, or two groups in the case of two doorways. An exit group might consist of a row of students, a small learning group or whatever grouping makes sense depending on the classroom arrangement. Students should exit through the doorways one at a time to practice social distancing.
6. In a fire or other emergency, faculty and students should disregard the social distancing conventions and make their way to the exits as quickly and calmly as possible
7. Remind students to wear masks when they are in the building, including when in hallways and in classrooms.

Face Masks in the classroom:

As mandated by the University System of Georgia, the university requires the use of face masks in the classroom and in KSU buildings to protect you, your classmates, and instructors. Per the University System of Georgia, anyone not using a face covering when required will be asked to wear one or must leave the area. Repeated refusal to comply with the requirement may result in discipline through the applicable conduct code.

Reasonable accommodations may be made for those who are unable to wear a face covering for documented health reasons. Please contact Student Disability Services at sds@kennesaw.edu for student accommodation requests.

Shifting Modalities:

Please note that the university reserves the right to shift teaching modalities at any time during the semester, if health and safety guidelines require it to do so. Some teaching modalities that may be used are F2F, Hyflex, Hybrid, or online, both synchronous and asynchronous instruction.

Staying Home When Sick:

If you are ill, please stay home and contact your health professional. In that case, please email the instructor to say you are missing class due to illness. Signs of illness include, but are not limited to, the following:

- Cough
- Fever of 100.4 or higher
- Runny nose or new sinus congestion
- Shortness of breath or difficulty breathing
- Chills
- Sore Throat
- New loss of taste and/or smell

Seating Plans:

Students will sit at the same seat for each of your scheduled F2F classes, determined at your scheduled first day of class. So the instructor can use a seating plan for contact tracing if a student infects Covid-19. You are required to register your seating groups 1, 2, or 3 by the deadline. Your scheduled F2F classes (by groups) are as follows:

Group #	Scheduled F2F Classes
1	1/12, 1/21, 2/2, 2/16, 2/25, 3/9, 3/23, 4/1, 4/13
2	1/14, 1/26, 2/4, 2/18, 3/2, 3/11, 2/25, 4/6, 4/15
3	1/19, 1/28, 2/9, 2/23, 3/4, 3/16, 3/30, 4/8, 4/20

Note: to be fair to all students, after 4/20, all students will attend online. So every student will attend 9 F2F lectures. Students may not exchange lecture dates with others. Please mark calendar for your scheduled lectures.

The KSU Library System: The KSU library system assists all students, faculty and staff with their research, including using library databases to find articles, accessing books and other materials in our catalog, and for specialized research needs. Librarians are available for in-person walk-up assistance at library help desks, one-on-one research appointments, and 24x7 via library chat. For more information on library locations, hours, and how to access our services please visit <http://library.kennesaw.edu/>.

The KSU Writing Center: The center helps students in all majors improve their writing. Experienced, friendly writing assistants help with topic development, revision, research, documentation, grammar, and more. For more information or to make an appointment, visit <http://writingcenter.kennesaw.edu> or stop by English Building, Room 242 (Kennesaw campus) or Building A, Room 184 (Marietta campus).

Department or College Policies:

Students are expected to be aware that the Computer Science department has certain policies in place that govern practices within the department including:

1. All requests for course overloads must be made through the College advising office and with the approval of the Program coordinator and department chair. The instructor of any course is not permitted to authorize course overloads.
2. All requests for prerequisite bypasses must be made through the College advising office and with the approval of the Program coordinator and department chair. The instructor of any course is not permitted to authorize course overwrites.
3. All students are encouraged to register their current choice of major using the department major change process. Students who are not recorded under their intended major may find that they may be limited from registering for courses they require to complete their intended program of study.

Institutional Policies

Please visit each of the following links for Institutional policies.

Federal, BOR, & KSU Course Syllabus Policies:

https://curriculum.kennesaw.edu/resources/federal_bor_ksu_student_policies.php

Student Resources:

https://curriculum.kennesaw.edu/resources/ksu_student_resources_for_course_syllabus.php

Academic Integrity Statement:

<https://scai.kennesaw.edu/codes.php>

KSU Student Resources:

This link contains information on help and resources available to students:

https://curriculum.kennesaw.edu/resources/ksu_student_resources_for_course_syllabus.php

Additional Resources:

For CCSE Student resources: <http://ccse.kennesaw.edu/student-resources.php>

KSU Service Desk: The KSU Service Desk is your portal to getting assistance or access to University IT Services. Students call: 470-578-3555 or email studenthelpdesk@kennesaw.edu

Information and links to Resources for Graduate Students: <http://graduate.kennesaw.edu/students/>

Links to frequently used and helpful services: <http://www.kennesaw.edu/myksu/>

Acknowledgment and Acceptance of Academic Integrity Statement

In any academic community, certain standards and ethical behavior are required to ensure the unhindered pursuit of knowledge and the free exchange of ideas. Academic honesty means that you respect the right of other individuals to express their views and opinions, and that you, as a student, not engage in plagiarism, cheating, illegal access, misuse or destruction of college property, or falsification of college records or academic work.

As a member of the Kennesaw State University academic community you are expected to adhere to these ethical standards. You are expected to read, understand and follow the code of conduct as outlined in the KSU graduate and undergraduate catalogs. You need to be aware that if you are found guilty of violating these standards you will be subject to certain penalties as outlined in the college judiciary procedures. These penalties include permanent expulsion from KSU.

Read the Academic Integrity Statement and then sign and date in the space below. You are required to abide by these ethical standards while you are a student at KSU. Your signature indicates that you understand the ethical standards expected of you in this academic community, and that you understand the consequences of violating these standards.

CS 4265: Big Data Analytics
Course Name

Dan Lo
Instructor Name

Print Name

Student ID Number

Signature

Date