Course	**DSCI 5350.001**
Course Title	**Big Data Analytics**
Professor	Kashif Saeed
Term	Summer I - 2019
Meetings	MW 6:00pm – 9:50pm; Chil 245, FR 161

Professor's Contact Information

Office Phone	940) 565-4769
Other Phone	
Office Location	BLB 312E
Email Address	Kashif.saeed@unt.edu
Office Hours	Wednesdays 5pm-6pm
TA Information	TBD TA Office hours: Will be posted on Canvas

General Course Information

Pre-requisites, Co-requisites, & other restrictions	Some experience with database management systems and SQL is helpful. Some experience with windows software installation is helpful. Some experience with any programming language and using the command line is helpful. **This class requires a laptop for hands-on and assignments. You must have at-least 8GB RAM on your computer You must bring your laptop to every class because of the hands-on nature of the class.**
Course Description	The course covers the following topics: 1. Hadoop Architecture and Ecosystem 2. HDFS and basic LINUX commands 3. Hive, Sqoop, Other Ecosystem tools as needed 4. Spark using Scala
Student Learning Objectives	**SLO1:** Students will be able to describe architecture and methods for storage and provision of structured and unstructured data. **SLO2:** Students will develop competency in using Big Data platform like Hadoop, EMR, or Spark to store and retrieve big data. **SLO3:** Students will demonstrate competency in creating big data pipeline in Hadoop, EMR, or Spark. This also includes familiarity in using streaming systems like Spark Streaming, Flink, Flume, Kafka, or others to ingest streaming data. **SLO4:** Students will learn how to use Python or Scala for programming in Spark – the most widely processing engine for big data platforms.
Key to Success in this course	• Attend classes and pay attention in the class • Take notes and work on hands-on activities • Good understanding of the assignments – there can be questions in the exams from the assignments • If you have doubts, ask questions

Optional Texts & Materials	The books listed below are for reference only. The course is topic based and does not cover a book chapter by chapter. Instructor will provide handouts for each lecture – exams will be from instructor handouts and assignments. 1. O'Reilly Sqoop Cookbook by Ting and Cecho 2. O'Reilly Programming Hive by Rutherglen, Wampler, and Capriolo 3. O'Reilly Programming Pig by Alan Gates 4. O'Reilly Learning Spark by Karau and Zahaia 5. Apache Spark and Scala documentation
Software Used	1. Cloudera VM – will be made available by the instructor 2. VMware https://my.vmware.com/web/vmware/free#desktop_end_user_computing/vmware_player/7_0 The instructor will provide install instructions for the software – it is the responsibility of the student to install the software and work with the TA to resolve any issues with software installation..

Assignments & Academic Calendar

Week	***Date***	***Topics to be covered***	***Assignments***
1	June 3rd	• Syllabus Overview and Expectations • Data Architecture concepts • Cloud computing basis	None
1	June 5th	• Cloudera Setup overview • HDFS Commands • Basic LINUX commands • Hands-on activities	
2	June 10th	• Sqoop Overview • Import and Export commands with Sqoop • Different tools and parameters in Sqoop • Sqoop Hands-on	Assignment 1 posted
2	June 12th	• Hive Overview • Hive Commands and Hands-on • Handling JSON data with Hive	
3	June 17th	• Remaining topics in Hive	Assignment 1 due
3	June 19th	• **Midterm Exam** • Introduction to Scala • Scala Hands-on • Spark Overview • Spark Setup	
4	June 24th	• Spark Additional Topics	
4	June 26th	• Spark Additional Topics	Assignment 2 Due
5	July 1st	• Spark Additional Topics	

5	July 3rd	**Final Exam (Comprehensive & Hands-on)**	

Course Policies

Grading (credit) Criteria	The course uses RANK based grading Top 30% Students – A Next 40% Students– B Next 30% Students – C and Other grades * You must score more than 130 points to avoid a 'D' or 'F'. Grading Points (250 total) Assignments – 80 points (2 assignments @ 40 points each) Midterm – 50 points Final – 80 points Class Activities – 40 Points
Make-up Exams	There will be no make-up exams. However, I will work with you if you have a date conflict and would like to change the date of your exam.
Extra Credit	None
Late Work	Penalty on late assignments will be listed on the Assignment itself.
Software Installation	It is your responsibility to install the software. The instructor and the teaching assistant are available to help, however, you MUST NOT assume that the instructor and/or teaching assistant till install the software for you.
Assignments	It is your responsibility to complete the assignments with or without the teaching assistant help. Remember that the teaching assistant is NOT responsible to solve your assignments for you – he/she can only guide you and provide high-level support to get past the obstacle that you may be experiencing.
Classroom Citizenship	Your behavior interferes with my ability to teach and student's ability to learn; unacceptable behavior will not be tolerated in my class. Students engaging in disruptive behavior will be asked to leave the classroom and will be referred to the center of student rights and responsibilities. Chatting, giggling, laughing, use of cell phone or other hand-held devices, texting,

	using a laptop while the instructor is teaching, making noises, etc. are examples of disruptive behavior. Penalty for Disruptive Behavior: • You will lose 25 Grade Points per offense of disruptive behavior for the first two offenses. The instructor will note down your name and will deduct the points at the end of the semester. • After two offenses of disruptive behavior, you will lose 50 points per offense. • These points will be deducted from your Exam and Assignment total. In addition, the instructor reserves the right to move you to a different seat during exams if the instructor believes that you are involved in cheating, plagiarism, or disrupting others.
Exam Reviews	Exam Reviews DONOT mean that I will provide highlighted text the exam will be from or provide sample question for the exam. Exam Reviews mean that you will have class time to clear any doubts you may have from the previous classes; it is my responsibility to ensure that I explain to clear your doubts, but it is your responsibility to come prepared to the exam review class to ask questions. If you have not attended classes prior to the exam review, do not assume that I can cover the material from all previous classes in the exam review class.
Class Attendance	Strongly recommended – missed class activities cannot be made up.
UNT Policies	Academic Integrity Standards and Sanctions for Violation. According to UNT Policy 18.1.16, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University. [Insert specific sanction or academic penalty for specific academic integrity violation]. ADA Statement. UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one's specific course needs. Students may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website at disability.unt.edu Emergency Notification & Procedures. UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials. Acceptable Student Behavior. Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students

	engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at deanofstudents.unt.edu/conduct. Student Perceptions of Teaching. Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13 and 14 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via *IASystem* Notification" (no-reply@iasystem.org) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the spot website at www.spot.unt.edu or email spot@unt.edu.

The descriptions, timelines, grading policies, or other information contained in this syllabus are subject to change at the discretion of the Professor.